Spring 2018

# HDG methods for Dirichlet boundary control of PDEs

Yangwen Zhang

HDG METHODS FOR DIRICHLET BOUNDARY CONTROL OF PDES

by

YANGWEN ZHANG

A DISSERTATION

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

MATHEMATICS

2018

Approved by

John R. Singler, Advisor
Xiaoming He, Co-Advisor
Yanzhi Zhang
Weiwei Hu
Jagannathan Sarangapani

## ABSTRACT

We begin an investigation of hybridizable discontinuous Galerkin (HDG) methods for approximating the solution of Dirichlet boundary control problems for PDEs. These problems can involve atypical variational formulations, and often have solutions with low regularity on polyhedral domains. These issues can provide challenges for numerical methods and the associated numerical analysis. In this thesis, we use an existing HDG method for a Dirichlet boundary control problem for the Poisson equation, and obtain optimal a priori error estimates for the control in the high regularity case. We also propose a new HDG method to approximate the solution of a Dirichlet boundary control problem governed by a linear elliptic convection diffusion PDE. Although there are many works in the literature on Dirichlet boundary control problems for the Poisson equation, we are not aware of any existing theoretical or numerical analysis works for convection diffusion Dirichlet control problems. We obtain well-posedness and regularity results for the Dirichlet control problem, and we prove optimal a priori error estimates in 2D for the control in both the high regularity and low regularity cases. As far as the authors are aware, there are no existing comparable results in the literature. Moreover, we present numerical experiments to demonstrate the performance of the HDG methods and illustrate our numerical analysis results.

# ACKNOWLEDGMENTS

First of all, I want to express my heartfelt gratitude to my advisor, Dr. John R Singler, who led me into the research world. His integrity and enthusiasm inspire me to become better as a scholar, an educator, and a person. I hope that one day I would become as good an advisor to my students as Dr. Singler has been to me.

Also, I am grateful for my co-advisor, Dr. Xiaoming He, for sharing his valuable experience and code. Thanks to Dr. Weiwei Hu, who brought me to the IMA, where I made great progress for my research. Thanks to Dr. Bernardo Cockburn for teaching me the HDG method and for many helpful discussions, especially in finding a postdoc position recently because of his recommendation. Thanks to my other two collaborators, Dr. Mariano Matoes and Dr. Wei Gong, for many helpful discussions. Thanks to my friend, Michael Julian, who helped us to improve the languange. Also thanks to my friends, Mr. Changxin Qiu, Mr. Daren Liu, Ms. Yin Li, and Dr. Yuchuan Chu, for many dinner chattings.

Two of my friends made great contributations for my research. The first one is Dr. Jiguang Shen, he spent many months to advise me how to code the HDG method, which is an important start for my research. The other friend is Dr. Xiaobo Zheng. We made a great development of the HDG method for Dirichlet boundary control of PDEs. Moreover, we built robust connections, such as Mr. Xiao Zhang and Dr. Gang Chen, who are my collaborators and Xiaobo's academic brothers.

I wish to express my sincerest gratitude to my parents. Their constant support provides inspiration and is a driving force for me to pursue down the academic road.

Thanks to my wife, Liangya; without her, this thesis could not have been completed. I will always appreciate her love and support, sometimes even sacrifices, to help me focus on my work. Finally, thanks to my lovely daughter, Christina; she gives me $\infty^\infty$ happiness and without her help, I could have finished this thesis at least three months earlier.

# TABLE OF CONTENTS

Page

# LIST OF ILLUSTRATIONS

Figure Page

# LIST OF TABLES

# 1. INTRODUCTION

Dirichlet boundary control has many applications in fluid flow problems and other fields, and therefore the mathematical study of these control problems has become an important area of research. Major theoretical and computational developments have been made in the recent past; see, e.g., [13, 24, 25, 29, 30, 31, 36, 37, 38, 40, 58, 59, 61]. However, only in the last ten years have researchers developed thorough well-posedness, regularity, and finite element error analysis results for elliptic PDEs; see [2, 11, 26, 49, 62] and the references therein. One difficulty of Dirichlet boundary control problems is that the Dirichlet boundary data does not directly enter a standard variational setting for the PDE; instead, the state equation is understood in a very weak sense. Also, solutions of the optimality system typically do not have high regularity on polyhedral domains; corners cause the normal derivative of the adjoint state in the optimality condition to have limited smoothness. Solutions with limited regularity can lead to complications for numerical methods and numerical analysis.

To avoid the difficulties described above, researchers have considered other approaches including modified cost functionals [19, 35, 37, 55], approximating the Dirichlet boundary condition with a Robin boundary condition [3, 5, 10, 41, 57], and weak boundary penalization [14].

One way to approximate the solution of the original problem without penalization and also avoid the variational difficulty is to use a mixed finite element method. Recently, Gong and Yan [33] considered this approach and obtained

$$\|u - u_h\|_{0,\Omega} = O(h^{1-1/s})$$

when the control $u$ belongs to $H^{1-1/s}(\Gamma)$ and the lowest order Raviart-Thomas elements are used for the computation.

As researchers continue to investigate Dirichlet boundary control problems of increasingly complexity, it may become natural to utilize discontinuous Galerkin methods for the spatial discretization of problems involving strong convection and discontinuities. We have performed preliminary computations using an hybridizable discontinuous Galerkin (HDG) method for an elliptic Dirichlet boundary control problem for the Stokes equations. Our preliminary results for this problem indicate that the optimal control can indeed be discontinuous at the corners of the domain. Before we continue to investigate problems of such complexity, we begin this line of research by considering an HDG method to approximate the solution of a Dirichlet boundary control problem for the Poisson equation.

HDG methods were proposed by Cockburn et al. in [20] as an improvement of tradition discontinuous Galerkin methods and have many applications; see, e.g., [12, 17, 21, 22, 23, 51, 52, 53, 54, 60]. The approximate scalar variable and flux are expressed in an element-by-element fashion in terms of an approximate trace of the scalar variable along the element boundary. Then, a unique value for the trace at inter-element boundaries is obtained by enforcing flux continuity. This leads to a global equation system in terms of the approximate boundary traces only. The high number of globally coupled degrees of freedom is significantly reduced compared to other DG methods and standard mixed methods.

In section 1 of the thesis, we approximate the solution of the Dirichlet boundary control problem for the Poisson equation using an existing HDG method. This method uses polynomials of degree $k + 1$ to approximate the state $y$ and dual state $z$ and polynomials of degree $k \geq 0$ for the fluxes $\boldsymbol{q} = -\nabla y$ and $\boldsymbol{p} = -\nabla z$, respectively. Moreover, we also used polynomials of degree $k$ to approximate the numerical trace of the state and dual state on the edges (or faces) of the spatial mesh, which are the only globally coupled unknowns. In Section 2.3, and in Section 2.4 we prove an optimal superlinear rate of convergence for the control in 2D under certain assumptions on the domain and $y_d$. To give a specific example, for a rectangular 2D domain and $y_d \in H^1(\Omega) \cap L^\infty(\Omega)$, we obtain the following a priori error bounds for the state $y$, adjoint state $z$, their fluxes $\boldsymbol{q} = -\nabla y$ and $\boldsymbol{p} = -\nabla z$, and the optimal control $u$:

$$
\begin{aligned}
\|y - y_h\|_{0,\Omega} &= O(h^{3/2-\varepsilon}), \qquad \|z - z_h\|_{0,\Omega} = O(h^{3/2-\varepsilon}),\\
\|\boldsymbol{q} - \boldsymbol{q}_h\|_{0,\Omega} &= O(h^{1-\varepsilon}), \qquad\quad \|\boldsymbol{p} - \boldsymbol{p}_h\|_{0,\Omega} = O(h^{3/2-\varepsilon}),
\end{aligned}
$$

and

$$
\|u - u_h\|_{0,\Gamma} = O(h^{3/2-\varepsilon}),
$$

for any $\varepsilon > 0$. We demonstrate the performance of the HDG method with numerical experiments in 2D and 3D in Section 2.5. Despite the large amount of existing work on this problem, a similar convergence rate has only very recently been proved for one other numerical method: a finite element method on a special class of meshes [2].

In the last two sections, we study the Dirichlet boundary control of convection diffusion equations. Optimal control problems governed by convection diffusion equations play an important role in many scientific and engineering problems [46]. Efficient and accurate numerical methods are essential to successful applications of such optimal control problems. There exist many contributions [4, 6, 7, 27, 39, 63] to numerical methods and algorithms for this kind of problem. Despite this large amount of existing work on numerical methods for convection diffusion optimal control problems and also Dirichlet boundary control problems for the Poisson equation and other PDEs, we are not aware of any existing work on the analysis and approximation of solutions for the for the convection diffusion Dirichlet boundary control problem considered here. Work on this problem is an important step towards the analysis and approximation of Dirichlet boundary control problems for the Navier-Stokes equations and other fluids models.

However, it is not clear to the authors if existing HDG methods can guarantee the superlinear convergence rate as we obtained in section 1 for elliptic convection diffusion PDEs. Therefore, we devise a new HDG method in Section 3.3 using polynomials of degree $k+1$ to approximate the state $y$, dual state $z$, and the numerical traces. Moreover, we use polynomials of degree $k \geq 0$ for the fluxes $\boldsymbol{q}$ and $\boldsymbol{p}$, respectively. In Section 3.4, we prove the same superlinear rate of convergence as in section 1 for the control in 2D under certain assumptions on the largest angle of the convex polygonal domain and the smoothness of the desired state $y_d$.

In section 3, we remove the assumptions required in section 2 on the convex polygonal domain and the desired state $y_d$ and prove optimal convergence rates for the control. Removing these assumptions on the domain and the desired state lower the regularity of the solution of the optimality system; therefore, regular HDG error analysis techniques are not applicable. We perform a nonstandard HDG error analysis based on techniques from [44] to establish the low regularity convergence results.

We emphasize that this new HDG method may be of primary interest for boundary control problems such as the one considered here. Existing HDG methods use order $k$ polynomials for the numerical traces, which are the only globally coupled unknowns. This new HDG method uses order $k+1$ polynomials for the numerical traces, and therefore it has a higher computational cost compared to existing HDG methods. However, adding one polynomial degree to the space for the numerical traces is the only way we have found to guarantee the optimal convergence rate for the control. The authors are not aware of any other application where this new HDG method will lead to an improved convergence analysis over existing HDG methods.

This thesis consists of material from the three preprints [32, 42, 43]. Some minor changes to the preprints have been made in this thesis in order to increase the readability of the thesis; no fundamental changes to the preprints have been made in this thesis.

# 2. POISSON WITH HIGH REGULARITY

## 2.1. MODEL PROBLEM

In this Section, we consider the following elliptic Dirichlet boundary control problem on a Lipschitz polyhedral domain $\Omega \subset \mathbb{R}^d$, $d \geq 2$, with boundary $\Gamma = \partial\Omega$ :

$$\min J(u), \quad J(u) = \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \frac{\gamma}{2}\|u\|^2_{L^2(\Gamma)}, \tag{2.1}$$

where $\gamma > 0$ and $y$ is the solution of the Poisson equation with nonhomogeneous Dirichlet boundary conditions

$$-\Delta y = f \quad \text{in } \Omega, \tag{2.2}$$

$$y = u \quad \text{on } \Gamma. \tag{2.3}$$

It is well known that the Dirichlet boundary control problem (2.1)-(2.3) is equivalent to the optimality system

$$-\Delta y = f \qquad \text{in } \Omega, \tag{2.4a}$$

$$y = u \qquad \text{on } \Gamma, \tag{2.4b}$$

$$-\Delta z = y - y_d \quad \text{in } \Omega, \tag{2.4c}$$

$$z = 0 \qquad \text{on } \Gamma, \tag{2.4d}$$

$$u = \gamma^{-1}\partial_{\boldsymbol{n}} z \quad \text{on } \Gamma. \tag{2.4e}$$

where $\boldsymbol{n}$ is the unit outer normal to $\Gamma$.

## 2.2. BACKGROUND: THE OPTIMALITY SYSTEM AND REGULARITY

To begin, we review some fundamental results concerning the optimality system for the control problem and the regularity of the solution in 2D polygonal domains.

Throughout the thesis we adopt the standard notation $W^{m,p}(\Omega)$ for Sobolev spaces on $\Omega$ with norm $\|\cdot\|_{m,p,\Omega}$ and seminorm $|\cdot|_{m,p,\Omega}$ . We denote $W^{m,2}(\Omega)$ by $H^m(\Omega)$ with norm $\|\cdot\|_{m,\Omega}$ and seminorm $|\cdot|_{m,\Omega}$. Also, $H_0^1(\Omega) = \{v \in H^1(\Omega) : v = 0 \text{ on } \partial\Omega\}$. We denote the $L^2$-inner products on $L^2(\Omega)$ and $L^2(\Gamma)$ by

$$(v, w) = \int_\Omega vw \quad \forall v, w \in L^2(\Omega),$$
$$\langle v, w\rangle = \int_\Gamma vw \quad \forall v, w \in L^2(\Gamma).$$

Define the space $H(\text{div};\Omega)$ as

$$H(\text{div}, \Omega) = \{\boldsymbol{v} \in [L^2(\Omega)]^2, \nabla\cdot\boldsymbol{v} \in L^2(\Omega)\}.$$

To avoid the the variational difficulty we follow the strategy introduced by Wei Gong and Ningning Yan [33] and consider a mixed formulation of the optimality system. Introduce two flux variables $\boldsymbol{q} = -\nabla y$ and $\boldsymbol{p} = -\nabla z$. The mixed weak form of (2.4a)-(2.4e) is

$$(\boldsymbol{q}, \boldsymbol{r}) - (y, \nabla\cdot\boldsymbol{r}) + \langle u, \boldsymbol{r}\cdot\boldsymbol{n}\rangle = 0, \tag{2.5a}$$

$$(\nabla\cdot\boldsymbol{q}, w) = (f, w), \tag{2.5b}$$

$$(\boldsymbol{p}, \boldsymbol{r}) - (z, \nabla\cdot\boldsymbol{r}) = 0, \tag{2.5c}$$

$$(\nabla\cdot\boldsymbol{p}, w) - (y, w) = (y_d, w), \tag{2.5d}$$

$$\langle \gamma u + \boldsymbol{p}\cdot\boldsymbol{n}, \xi\rangle = 0, \tag{2.5e}$$

for all $(\boldsymbol{r}, w, \xi) \in H(\text{div}, \Omega) \times L^2(\Omega) \times L^2(\Gamma)$.

One of the main reasons that Dirichlet boundary control problem can be challenging numerically is that the solution can have very low regularity, and this restricts the convergence rates of finite element and DG methods. In order to prove a superlinear convergence rate for the optimal control for the HDG method in 2.4, we assume the solution has the following fractional Sobolev regularity:

$$u \in H^{r_u}(\Gamma), \quad y \in H^{r_y}(\Omega), \quad z \in H^{r_z}(\Omega), \quad \boldsymbol{q} \in H^{r_{\boldsymbol{q}}}(\Omega), \quad \boldsymbol{p} \in H^{r_{\boldsymbol{p}}}(\Omega), \tag{2.6}$$

with

$$r_u > 1, \quad r_y > 1, \quad r_z > 2, \quad r_{\boldsymbol{q}} > 1/2, \quad r_{\boldsymbol{p}} > 1. \tag{2.7}$$

We require $r_{\boldsymbol{q}} > 1/2$ in order to guarantee $q$ has a well-defined trace on the boundary $\Gamma$. We note that it may be possible to use the techniques in [44] to lower the regularity requirement on $\boldsymbol{q}$. We leave this to be considered elsewhere.

For a 2D convex polygonal domain and $f = 0$, we use a recent regularity result of Mateos and Neitzel [48] below to give conditions on the domain and $y_d$ to guarantee the solution has the above regularity. For a higher dimensional convex polyhedral domain, the regularity theory is much more complicated, and we do not attempt to provide conditions to guarantee the above regularity in this work.

**Theorem 1 ([48], Lemma 3 and Corollary 1)** *Suppose $f = 0$ and $\Omega \subset \mathbb{R}^2$ is a bounded convex domain with polygonal boundary $\Gamma$. Let $\omega \in [\pi/3, \pi)$ be the largest interior angle of $\Gamma$, and define $p_\Omega$, $r_\Omega$ by*

$$p_\Omega = \frac{2}{2 - \pi/\max\{\omega, \pi/2\}} \in (2, \infty],$$

*and*

$$r_\Omega = 1 + \frac{\pi}{\omega} \in (2, 4].$$

*If* $y_d \in L^p(\Omega) \cap H^{r-2}(\Omega)$ *for all* $p < p_\Omega$ *and* $r < r_\Omega$*, then the solution* $(u, y, z)$ *satisfies*

$$
\begin{aligned}
u &\in H^{r-3/2}(\Gamma) \cap W^{1-1/p,p}(\Gamma),\\
y &\in H^{r-1}(\Omega) \cap W^{1,p}(\Omega),\\
z &\in H_0^1(\Omega) \cap H^r(\Omega) \cap W^{2,p}(\Omega)
\end{aligned}
$$

*for all*

$$p < p_\Omega, \quad r < \min\{3, r_\Omega\}.$$

We also require the regularity for the flux variables $\boldsymbol{q} = -\nabla y$ and $\boldsymbol{p} = -\nabla z$.

**Corollary 1** *Under the assumptions of Theorem 1, the flux variables* $\boldsymbol{q} = -\nabla y$ *and* $\boldsymbol{p} = -\nabla z$ *satisfy*

$$\boldsymbol{q} \in H^{r-2}(\Omega) \cap H(\mathrm{div}, \Omega), \quad \boldsymbol{p} \in H^{r-1}(\Omega) \cap H(\mathrm{div}, \Omega)$$

*for all* $r < \min\{3, r_\Omega\}$.

**Proof:** We treat the optimal control $u$ as known, and then $(y, \boldsymbol{q})$ satisfy the weak mixed formulation (2.5a)-(2.5b). Since $u \in H^{1/2}(\Gamma)$, the standard theory for this mixed problem gives $\boldsymbol{q} \in H(\mathrm{div}, \Omega)$. Taking $\boldsymbol{r}$ smooth and integrating by parts in (2.5a) gives $\boldsymbol{q} = -\nabla y$, and therefore the fractional Sobolev regularity for $\boldsymbol{q}$ follows directly from Theorem 1. The regularity for $\boldsymbol{p}$ follows similarly.

The regularity for the flux variable $\boldsymbol{q} = -\nabla y$ is low; Theorem 1 only guarantees $\boldsymbol{q} \in H^{r_{\boldsymbol{q}}}$ for some $0 < r_{\boldsymbol{q}} < 1$. For the HDG approximation theory, we need the regularity condition $r_{\boldsymbol{q}} > 1/2$. We can guarantee this condition by restricting the maximum interior angle $\omega$. Specifically, if if $y_d$ has the required smoothness and $\omega$ satisfies

$$\omega \in [\pi/3, 2\pi/3),$$

then $r_\Omega \in (5/2, 4]$ and we are guaranteed $\boldsymbol{q} \in H^{r_{\boldsymbol{q}}}$ for some $r_{\boldsymbol{q}} > 1/2$.

Also, when we restrict $\omega \in [\pi/3, 2\pi/3)$ as above, this guarantees $u \in H^{r_u}$ for some $1 < r_u < 3/2$ and furthermore the regularity assumption (2.6)-(2.7) is satisfied. For a rectangular domain, we have $p_\Omega = \infty$ and $r_\Omega = 3$. Therefore if $y_d \in H^1(\Omega) \cap L^\infty(\Omega)$ we are guaranteed the fractional Sobolev regularity

$$r_u = \frac{3}{2} - \varepsilon, \quad r_y = 2 - \varepsilon, \quad r_z = 3 - \varepsilon, \quad r_{\boldsymbol{q}} = 1 - \varepsilon, \quad r_{\boldsymbol{p}} = 2 - \varepsilon$$

for any $\varepsilon > 0$.

## 2.3. HDG FORMULATION AND IMPLEMENTATION

A mixed method can avoid the variational difficulty by the introducing flux variables $\boldsymbol{q}$ and $\boldsymbol{p}$ and the equation for the optimal control (2.5e). However, these two additional vector variables will increase the computational cost, even if the lowest order RT method is used.

We introduce an HDG method for the optimality system (2.4) to take advantage of the mixed formulation and also reduce the computational cost compared to standard mixed methods. Specifically, we introduce the flux variables but eliminate them before we solve the global equation; this significantly reduces the degrees of freedom.

Before we introduce the HDG method, we first set some notation. Let $\{\mathcal{T}_h\}$ be a conforming quasi-uniform polyhedral mesh of $\Omega$. We denote by $\partial\mathcal{T}_h$ the set $\{\partial K : K \in \mathcal{T}_h\}$. For an element $K$ of the collection $\mathcal{T}_h$, let $e = \partial K \cap \Gamma$ denote the boundary face of $K$ if the $d-1$ Lebesgue measure of $e$ is non-zero. For two elements $K^+$ and $K^-$ of the collection $\mathcal{T}_h$, let $e = \partial K^+ \cap \partial K^-$ denote the interior face between $K^+$ and $K^-$ if the $d-1$ Lebesgue measure of $e$ is non-zero. Let $\varepsilon_h^o$ and $\varepsilon_h^\partial$ denote the set of interior and boundary faces, respectively. We denote by $\varepsilon_h$ the union of $\varepsilon_h^o$ and $\varepsilon_h^\partial$. We finally introduce

$$(w, v)_{\mathcal{T}_h} = \sum_{K \in \mathcal{T}_h} (w, v)_K, \qquad \langle \zeta, \rho \rangle_{\partial \mathcal{T}_h} = \sum_{K \in \mathcal{T}_h} \langle \zeta, \rho \rangle_{\partial K} .$$

Let $\mathcal{P}^k(D)$ denote the set of polynomials of degree at most $k$ on a domain $D$. We introduce the discontinuous finite element spaces

$$\boldsymbol{V}_h := \{\boldsymbol{v} \in [L^2(\Omega)]^d : \boldsymbol{v}|_K \in [\mathcal{P}^k(K)]^d, \forall K \in \mathcal{T}_h\}, \tag{2.8}$$

$$W_h := \{w \in L^2(\Omega) : w|_K \in \mathcal{P}^{k+1}(K), \forall K \in \mathcal{T}_h\}, \tag{2.9}$$

$$M_h := \{\mu \in L^2(\varepsilon_h) : \mu|_e \in \mathcal{P}^k(e), \forall e \in \varepsilon_h\}. \tag{2.10}$$

The space $W_h$ is for scalar variables, while $\boldsymbol{V}_h$ is for flux variables and $M_h$ is for boundary trace variables. Note that the polynomial degree for the scalar variables is one order higher than the polynomial degree for the other variables. Also, the boundary trace variables will be used to eliminate the state and flux variables from the coupled global equations, thus substantially reducing the number of degrees of freedom.

Let $M_h(o)$ and $M_h(\partial)$ denote the spaces defined in the same way as $M_h$, but with $\varepsilon_h$ replaced by $\varepsilon_h^o$ and $\varepsilon_h^\partial$, respectively. Note that $M_h$ consists of functions which are continuous inside the faces (or edges) $e \in \varepsilon_h$ and discontinuous at their borders. In addition, for any function $w \in W_h$ we use $\nabla w$ to denote the piecewise gradient on each element $K \in \mathcal{T}_h$. A similar convention applies to the divergence operator $\nabla \cdot \boldsymbol{r}$ for all $\boldsymbol{r} \in \boldsymbol{V}_h$.

**2.3.1. The HDG Formulation.** To approximate the solution of the mixed weak form (2.4a)-(2.4e) of the optimality system, the HDG method seeks approximate fluxes $\boldsymbol{q}_h, \boldsymbol{p}_h \in \boldsymbol{V}_h$, states $y_h, z_h \in W_h$, interior element boundary traces $\widehat{y}_h^o, \widehat{z}_h^o \in M_h(o)$, and boundary control $u_h \in M_h(\partial)$ satisfying

$$(\boldsymbol{q}_h, \boldsymbol{r_1})_{\mathcal{T}_h} - (y_h, \nabla\cdot\boldsymbol{r_1})_{\mathcal{T}_h} + \langle \widehat{y}_h^o, \boldsymbol{r_1}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} + \langle u_h, \boldsymbol{r_1}\cdot\boldsymbol{n}\rangle_{\varepsilon_h^\partial} = 0, \tag{2.11a}$$

$$-(\boldsymbol{q}_h, \nabla w_1)_{\mathcal{T}_h} + \langle \widehat{\boldsymbol{q}}_h\cdot\boldsymbol{n}, w_1\rangle_{\partial\mathcal{T}_h} = (f, w_1)_{\mathcal{T}_h}, \tag{2.11b}$$

for all $(\boldsymbol{r_1}, w_1) \in \boldsymbol{V}_h \times W_h$,

$$(\boldsymbol{p}_h, \boldsymbol{r_2})_{\mathcal{T}_h} - (z_h, \nabla\cdot\boldsymbol{r_2})_{\mathcal{T}_h} + \langle \widehat{z}_h^o, \boldsymbol{r_2}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} = 0, \tag{2.11c}$$

$$-(\boldsymbol{p}_h, \nabla w_2)_{\mathcal{T}_h} + \langle \widehat{\boldsymbol{p}}_h\cdot\boldsymbol{n}, w_2\rangle_{\partial\mathcal{T}_h} - (y_h, w_2)_{\mathcal{T}_h} = -(y_d, w_2)_{\mathcal{T}_h}, \tag{2.11d}$$

for all $(\boldsymbol{r_2}, w_2) \in \boldsymbol{V}_h \times W_h$,

$$\langle \widehat{\boldsymbol{q}}_h\cdot\boldsymbol{n}, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} = 0, \tag{2.11e}$$

for all $\mu_1 \in M_h(o)$,

$$\langle \widehat{\boldsymbol{p}}_h\cdot\boldsymbol{n}, \mu_2\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} = 0, \tag{2.11f}$$

for all $\mu_2 \in M_h(o)$, and

$$\langle u_h, \mu_3\rangle_{\varepsilon_h^\partial} + \langle \gamma^{-1}\widehat{\boldsymbol{p}}_h\cdot\boldsymbol{n}, \mu_3\rangle_{\varepsilon_h^\partial} = 0, \tag{2.11g}$$

for all $\mu_3 \in M_h(\partial)$.

The numerical traces on $\partial\mathcal{T}_h$ are defined as

$$\widehat{\boldsymbol{q}}_h \cdot \boldsymbol{n} = \boldsymbol{q}_h \cdot \boldsymbol{n} + h^{-1}(P_M y_h - \widehat{y}_h^o) \quad \text{on } \partial\mathcal{T}_h \backslash \varepsilon_h^\partial, \tag{2.11h}$$

$$\widehat{\boldsymbol{q}}_h \cdot \boldsymbol{n} = \boldsymbol{q}_h \cdot \boldsymbol{n} + h^{-1}(P_M y_h - u_h) \quad \text{on } \varepsilon_h^\partial, \tag{2.11i}$$

$$\widehat{\boldsymbol{p}}_h \cdot \boldsymbol{n} = \boldsymbol{p}_h \cdot \boldsymbol{n} + h^{-1}(P_M z_h - \widehat{z}_h^o) \quad \text{on } \partial\mathcal{T}_h \backslash \varepsilon_h^\partial, \tag{2.11j}$$

$$\widehat{\boldsymbol{p}}_h \cdot \boldsymbol{n} = \boldsymbol{p}_h \cdot \boldsymbol{n} + h^{-1} P_M z_h \quad \text{on } \varepsilon_h^\partial, \tag{2.11k}$$

where $P_M$ denotes the standard $L^2$-orthogonal projection from $L^2(\varepsilon_h)$ onto $M_h$. This completes the formulation of the HDG method.

**2.3.2. Implementation.** To arrive at the HDG formulation we implement numerically, we insert (2.11h)-(2.11k) into (2.11a)-(2.11g), and find after some simple manipulations that

$$(\boldsymbol{q}_h, \boldsymbol{p}_h, y_h, z_h, \widehat{y}_h^o, \widehat{z}_h^o, u_h) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times W_h \times M_h(o) \times M_h(o) \times M_h(\partial)$$

is the solution of the following weak formulation:

$$(\boldsymbol{q}_h, \boldsymbol{r_1})_{\mathcal{T}_h} - (y_h, \nabla \cdot \boldsymbol{r_1})_{\mathcal{T}_h} + \langle \widehat{y}_h^o, \boldsymbol{r_1} \cdot \boldsymbol{n} \rangle_{\partial\mathcal{T}_h \backslash \varepsilon_h^\partial} + \langle u_h, \boldsymbol{r_1} \cdot \boldsymbol{n} \rangle_{\varepsilon_h^\partial} = 0, \tag{2.12a}$$

$$(\boldsymbol{p}_h, \boldsymbol{r_2})_{\mathcal{T}_h} - (z_h, \nabla \cdot \boldsymbol{r_2})_{\mathcal{T}_h} + \langle \widehat{z}_h^o, \boldsymbol{r_2} \cdot \boldsymbol{n} \rangle_{\partial\mathcal{T}_h \backslash \varepsilon_h^\partial} = 0, \tag{2.12b}$$

$$(\nabla \cdot \boldsymbol{q}_h, w_1)_{\mathcal{T}_h} + \langle h^{-1} P_M y_h, w_1 \rangle_{\partial\mathcal{T}_h} - \langle h^{-1} \widehat{y}_h^o, w_1 \rangle_{\partial\mathcal{T}_h \backslash \varepsilon_h^\partial} \tag{2.12c}$$

$$-\langle h^{-1} u_h, w_1 \rangle_{\varepsilon_h^\partial} = (f, w_1)_{\mathcal{T}_h}, \tag{2.12d}$$

$$(\nabla \cdot \boldsymbol{p}_h, w_2)_{\mathcal{T}_h} + \langle h^{-1} P_M z_h, w_2 \rangle_{\partial\mathcal{T}_h} - \langle h^{-1} \widehat{z}_h^o, w_2 \rangle_{\partial\mathcal{T}_h \backslash \varepsilon_h^\partial} \tag{2.12e}$$

$$-(y_h, w_2)_{\mathcal{T}_h} = -(y_d, w_2)_{\mathcal{T}_h}, \tag{2.12f}$$

$$\langle \boldsymbol{q}_h \cdot \boldsymbol{n}, \mu_1 \rangle_{\partial\mathcal{T}_h \backslash \varepsilon_h^\partial} + \langle h^{-1} y_h, \mu_1 \rangle_{\partial\mathcal{T}_h \backslash \varepsilon_h^\partial} - \langle h^{-1} \widehat{y}_h^o, \mu_1 \rangle_{\partial\mathcal{T}_h \backslash \varepsilon_h^\partial} = 0, \tag{2.12g}$$

$$\langle \boldsymbol{p}_h \cdot \boldsymbol{n}, \mu_2 \rangle_{\partial\mathcal{T}_h \backslash \varepsilon_h^\partial} + \langle h^{-1} z_h, \mu_2 \rangle_{\partial\mathcal{T}_h \backslash \varepsilon_h^\partial} - \langle h^{-1} \widehat{z}_h^o, \mu_2 \rangle_{\partial\mathcal{T}_h \backslash \varepsilon_h^\partial} = 0, \tag{2.12h}$$

$$\langle u_h, \mu_3 \rangle_{\varepsilon_h^\partial} + \langle \gamma^{-1} \boldsymbol{p}_h \cdot \boldsymbol{n}, \mu_3 \rangle_{\varepsilon_h^\partial} + \langle \gamma^{-1} h^{-1} z_h, \mu_3 \rangle_{\varepsilon_h^\partial} = 0, \tag{2.12i}$$

for all $(\boldsymbol{r}_1, \boldsymbol{r}_2, w_1, w_2, \mu_1, \mu_2, \mu_3) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times W_h \times M_h(o) \times M_h(o) \times M_h(\partial)$.

Assume $\boldsymbol{V}_h = \text{span}\{\boldsymbol{\varphi}_i\}_{i=1}^{N_1}$, $W_h = \text{span}\{\phi_i\}_{i=1}^{N_2}$, $M_h^o = \text{span}\{\psi_i\}_{i=1}^{N_3}$, and $M_h^\partial = \text{span}\{\psi_i\}_{i=1+N_3}^{N_4}$. Then

$$\begin{aligned}
\boldsymbol{q}_h &= \sum_{j=1}^{N_1} q_j \boldsymbol{\varphi}_j, \quad \boldsymbol{p}_h = \sum_{j=1}^{N_1} p_j \boldsymbol{\varphi}_j, \quad y_h = \sum_{j=1}^{N_2} y_j \phi_j, \quad z_h = \sum_{j=1}^{N_2} z_j \phi_j, \\
\widehat{y}_h^o &= \sum_{j=1}^{N_3} \alpha_j \psi_j, \quad \widehat{z}_h^o = \sum_{j=1}^{N_3} \gamma_j \psi_j, \quad u_h = \sum_{j=1+N_3}^{N_4} \beta_j \psi_j.
\end{aligned} \tag{2.13}$$

Substitute (2.13) into (2.12a)-(2.12i) and use the corresponding test functions to test (2.12a)-(2.12i), respectively, to obtain the matrix equation

$$\begin{bmatrix}
A_1 & 0 & -A_2 & 0 & A_8 & 0 & A_9 \\
0 & A_1 & 0 & -A_2 & 0 & A_8 & 0 \\
A_2^T & 0 & A_5 & 0 & -A_{10} & 0 & -A_{11} \\
0 & A_2^T & -A_4 & A_5 & 0 & -A_{10} & 0 \\
A_8^T & 0 & A_{10}^T & 0 & A_{11} & 0 & 0 \\
0 & A_8^T & 0 & A_{10}^T & 0 & A_{11} & 0 \\
0 & \gamma^{-1} A_{12} & 0 & \gamma^{-1} A_{13} & 0 & 0 & A_{14}
\end{bmatrix}
\begin{bmatrix}
\mathfrak{q} \\ \mathfrak{p} \\ \mathfrak{y} \\ \mathfrak{z} \\ \widehat{\mathfrak{y}} \\ \widehat{\mathfrak{z}} \\ \mathfrak{u}
\end{bmatrix}
=
\begin{bmatrix}
0 \\ 0 \\ b_1 \\ -b_2 \\ 0 \\ 0 \\ 0
\end{bmatrix}. \tag{2.14}$$

Here, $\mathfrak{q}, \mathfrak{p}, \mathfrak{y}, \mathfrak{z}, \widehat{\mathfrak{y}}, \widehat{\mathfrak{z}}, \mathfrak{u}$ are the coefficient vectors for $\boldsymbol{q}_h, \boldsymbol{p}_h, y_h, z_h, \widehat{y}_h^o, \widehat{z}_h^o, u_h$, respectively, and

$$\begin{gathered}
A_1 = [(\boldsymbol{\varphi}_j, \boldsymbol{\varphi}_i)_{\mathcal{T}_h}], \quad A_2 = [(\phi_j, \nabla \cdot \boldsymbol{\varphi}_i)_{\mathcal{T}_h}], \quad A_3 = [(\psi_j, \boldsymbol{\varphi}_i \cdot \boldsymbol{n})_{\mathcal{T}_h}], \quad A_4 = [(\phi_j, \phi_i)_{\mathcal{T}_h}], \\
A_5 = [\langle h^{-1} P_M \phi_j, \phi_i \rangle_{\partial \mathcal{T}_h}], \quad A_6 = [\langle h^{-1} \psi_j, \psi_i \rangle_{\partial \mathcal{T}_h}], \quad A_7 = [\langle h^{-1} \psi_j, \varphi_i \rangle_{\partial \mathcal{T}_h}], \\
b_1 = [(f, \phi_i)_{\mathcal{T}_h}], \quad b_2 = [(y_d, \phi_i)_{\mathcal{T}_h}].
\end{gathered}$$

The remaining matrices $A_8 - A_{14}$ are constructed by extracting the corresponding rows and columns from $A_3$, $A_6$, and $A_7$. In the actual computation, to save memory we do not assemble the large matrix in equation (2.14).

Equation (2.14) can be rewritten as

$$\begin{bmatrix} B_1 & B_2 & B_3 \\ -B_2^T & B_4 & B_5 \\ B_6 & B_7 & B_8 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{bmatrix} = \begin{bmatrix} 0 \\ b \\ 0 \end{bmatrix}, \tag{2.15}$$

where $\boldsymbol{\alpha} = [\mathfrak{q};\mathfrak{p}]$, $\boldsymbol{\beta} = [\mathfrak{y};\mathfrak{z}]$, $\boldsymbol{\gamma} = [\widehat{\mathfrak{y}};\widehat{\mathfrak{z}};\mathfrak{u}]$, $b = [b_1; -b_2]$, and $\{B_i\}_{i=1}^{8}$ are the corresponding blocks of the coefficient matrix in (2.14).

Due to the discontinuous nature of the approximation spaces $\boldsymbol{V_h}$ and $W_h$, the first two equations of (2.15) can be used to eliminate both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in an element-by-element fashion. As a consequence, we can write system (2.15) as

$$\begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} G_1 & H_1 \\ G_2 & H_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma} \\ b \end{bmatrix} \tag{2.16}$$

and

$$B_6\boldsymbol{\alpha} + B_7\boldsymbol{\beta} + B_8\boldsymbol{\gamma} = 0. \tag{2.17}$$

We provide details on the element-by-element construction of $G_1, G_2$ and $H_1, H_2$ in [43]. Next, we eliminate both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to obtain a reduced globally coupled equation for $\boldsymbol{\gamma}$ only:

$$\mathbb{K}\boldsymbol{\gamma} = \mathbb{F}, \tag{2.18}$$

where

$$\mathbb{K} = B_6 G_1 + B_7 G_2 + B_8 \quad \text{and} \quad \mathbb{F} = B_6 H_1 + B_7 H_2.$$

Once $\boldsymbol{\gamma}$ is available, both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be recovered from (2.16).

**Remark 1** *For HDG methods, the standard approach is to first compute the local solver independently on each element and then assemble the global system. The process we follow here is to first assemble the global system and then reduce its dimension by simple block-diagonal algebraic operations. The two approaches are equivalent.*

Equation (2.16) says we can express the approximate the scalar state variable and corresponding fluxes in terms of the approximate traces on the element boundaries. The global equation (2.18) only involves the approximate traces. Therefore, the high number of globally coupled degrees of freedom in the HDG method is significantly reduced. This is one excellent feature of HDG methods.

## 2.4. ERROR ANALYSIS

Next, we provide a convergence analysis of the above HDG method for the Dirichlet boundary control problem. Throughout this section, we assume $\Omega$ is a bounded convex polyhedral domain and we also assume the regularity condition (2.6)-(2.7) is satisfied. For the 2D case, recall Section 2.2 provides conditions on $\Omega$ and $y_d$ guaranteeing the required regularity.

**2.4.1. Main Result.** First, we present the following main theoretical result of this work. Recall we assume the fractional Sobolev regularity exponents satisfy

$$r_u > 1, \quad r_y > 1, \quad r_z > 2, \quad r_{\boldsymbol{q}} > 1/2, \quad r_{\boldsymbol{p}} > 1.$$

**Theorem 2** *For*

$$s_y = \min\{r_y, k+2\}, \;\; s_z = \min\{r_z, k+2\}, \;\; s_{\boldsymbol{q}} = \min\{r_{\boldsymbol{q}}, k+1\}, \;\; s_{\boldsymbol{p}} = \min\{r_{\boldsymbol{p}}, k+1\},$$

*we have*

$$
\begin{aligned}
\|u-u_h\|_{\varepsilon_h^\partial} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}}\|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega}+h^{s_z-\frac{3}{2}}\|z\|_{s_z,\Omega}+h^{s_{\boldsymbol{q}}+\frac{1}{2}}\|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega}+h^{s_y-\frac{1}{2}}\|y\|_{s_y,\Omega},\\
\|y-y_h\|_{\mathcal{T}_h} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}}\|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega}+h^{s_z-\frac{3}{2}}\|z\|_{s_z,\Omega}+h^{s_{\boldsymbol{q}}+\frac{1}{2}}\|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega}+h^{s_y-\frac{1}{2}}\|y\|_{s_y,\Omega}\\
\|\boldsymbol{q}-\boldsymbol{q}_h\|_{\mathcal{T}_h} &\lesssim h^{s_{\boldsymbol{p}}-1}\|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega}+h^{s_z-2}\|z\|_{s_z,\Omega}+h^{s_{\boldsymbol{q}}}\|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega}+h^{s_y-1}\|y\|_{s_y,\Omega},\\
\|\boldsymbol{p}-\boldsymbol{p}_h\|_{\mathcal{T}_h} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}}\|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega}+h^{s_z-\frac{3}{2}}\|z\|_{s_z,\Omega}+h^{s_{\boldsymbol{q}}+\frac{1}{2}}\|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega}+h^{s_y-\frac{1}{2}}\|y\|_{s_y,\Omega},\\
\|z-z_h\|_{\mathcal{T}_h} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}}\|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega}+h^{s_z-\frac{3}{2}}\|z\|_{s_z,\Omega}+h^{s_{\boldsymbol{q}}+\frac{1}{2}}\|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega}+h^{s_y-\frac{1}{2}}\|y\|_{s_y,\Omega}.
\end{aligned}
$$

Using the regularity results for the 2D case presented in Section 2.2, we obtain the following result.

**Corollary 2** *Suppose $d=2$, $f=0$, and $k=1$. Let $\omega\in[\pi/3,2\pi/3)$ be the largest interior angle of $\Gamma$, and define $p_\Omega$, $r_\Omega$ by*

$$
p_\Omega=\frac{2}{2-\pi/\max\{\omega,\pi/2\}}\in(4,\infty],\qquad r_\Omega=1+\frac{\pi}{\omega}\in(5/2,4].
$$

*If $y_d\in L^p(\Omega)\cap H^{r-2}(\Omega)$ for all $p<p_\Omega$ and $r<r_\Omega$, then for any $r<\min\{3,r_\Omega\}$ we have*

$$
\begin{aligned}
\|u-u_h\|_{\varepsilon_h^\partial} &\lesssim h^{r-\frac{3}{2}}(\|\boldsymbol{p}\|_{H^{r-1}(\Omega)}+\|z\|_{H^r(\Omega)}+\|\boldsymbol{q}\|_{H^{r-2}(\Omega)}+\|y\|_{H^{r-1}(\Omega)}),\\
\|y-y_h\|_{\mathcal{T}_h} &\lesssim h^{r-\frac{3}{2}}(\|\boldsymbol{p}\|_{H^{r-1}(\Omega)}+\|z\|_{H^r(\Omega)}+\|\boldsymbol{q}\|_{H^{r-2}(\Omega)}+\|y\|_{H^{r-1}(\Omega)}),\\
\|\boldsymbol{q}-\boldsymbol{q}_h\|_{\mathcal{T}_h} &\lesssim h^{r-2}(\|\boldsymbol{p}\|_{H^{r-1}(\Omega)}+\|z\|_{H^r(\Omega)}+\|\boldsymbol{q}\|_{H^{r-2}(\Omega)}+\|y\|_{H^{r-1}(\Omega)}),\\
\|\boldsymbol{p}-\boldsymbol{p}_h\|_{\mathcal{T}_h} &\lesssim h^{r-\frac{3}{2}}(\|\boldsymbol{p}\|_{H^{r-1}(\Omega)}+\|z\|_{H^r(\Omega)}+\|\boldsymbol{q}\|_{H^{r-2}(\Omega)}+\|y\|_{H^{r-1}(\Omega)}),\\
\|z-z_h\|_{\mathcal{T}_h} &\lesssim h^{r-\frac{3}{2}}(\|\boldsymbol{p}\|_{H^{r-1}(\Omega)}+\|z\|_{H^r(\Omega)}+\|\boldsymbol{q}\|_{H^{r-2}(\Omega)}+\|y\|_{H^{r-1}(\Omega)}).
\end{aligned}
$$

Note that $\min\{3,r_\Omega\}$ is always greater than 5/2, which guarantees a superlinear convergence rate for all variables except $\boldsymbol{q}$. Also, if $\Omega$ is a rectangle (i.e., $\omega=\pi/2$) and $y_d\in H^1(\Omega)\cap L^\infty(\Omega)$, then $r_\Omega=3$ and we obtain an $O(h^{3/2-\varepsilon})$ convergence rate for $u$, $y$, $z$, and $\boldsymbol{p}$, and an $O(h^{1-\varepsilon})$ convergence rate for $\boldsymbol{q}$ for any $\varepsilon>0$.

**2.4.2. Preliminary Material.** Before we prove the main result, we discuss $L^2$ projections, an HDG operator $\mathscr{B}$, and the well-posedness of the HDG equations.

We first define the standard $L^2$ projections $\boldsymbol{\Pi} : [L^2(\Omega)]^d \to \boldsymbol{V}_h$, $\Pi : L^2(\Omega) \to W_h$, and $P_M : L^2(\varepsilon_h) \to M_h$, which satisfy

$$
\begin{aligned}
(\boldsymbol{\Pi q}, \boldsymbol{r})_K &= (\boldsymbol{q}, \boldsymbol{r})_K, \qquad \forall \boldsymbol{r} \in [\mathcal{P}_k(K)]^d,\\
(\Pi u, w)_K &= (u, w)_K, \qquad \forall w \in \mathcal{P}_{k+1}(K),\\
\langle P_M m, \mu\rangle_e &= \langle m, \mu\rangle_e, \qquad \forall \mu \in \mathcal{P}_k(e).
\end{aligned} \tag{2.19}
$$

In the analysis, we use the following classical results:

$$
\|\boldsymbol{q} - \boldsymbol{\Pi q}\|_{\mathcal{T}_h} \le C h^{s_{\boldsymbol{q}}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega}, \qquad \|y - \Pi y\|_{\mathcal{T}_h} \le C h^{s_y} \|y\|_{s_y,\Omega}, \tag{2.20a}
$$

$$
\|y - \Pi y\|_{\partial\mathcal{T}_h} \le C h^{s_y - \frac{1}{2}} \|y\|_{s_y,\Omega}, \quad \|\boldsymbol{q}\cdot\boldsymbol{n} - \boldsymbol{\Pi q}\cdot\boldsymbol{n}\|_{\partial\mathcal{T}_h} \le C h^{s_{\boldsymbol{q}} - \frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega}, \tag{2.20b}
$$

$$
\|w\|_{\partial\mathcal{T}_h} \le C h^{-\frac{1}{2}} \|w\|_{\mathcal{T}_h}, \qquad \forall w \in W_h, \tag{2.20c}
$$

where $s_{\boldsymbol{q}}$ and $s_y$ are defined in Theorem 2. We have the same projection error bounds for $\boldsymbol{p}$ and $z$.

To shorten lengthy equations, we define the HDG operator $\mathscr{B}$ as follows:

$$
\mathscr{B}(\boldsymbol{q}_h, y_h, \widehat{y}_h^o; \boldsymbol{r}_1, w_1, \mu_1) \tag{2.21}
$$

$$
\begin{aligned}
&= (\boldsymbol{q}_h, \boldsymbol{r}_1)_{\mathcal{T}_h} - (y_h, \nabla\cdot\boldsymbol{r}_1)_{\mathcal{T}_h} + \langle \widehat{y}_h^o, \boldsymbol{r}_1\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\quad - (\boldsymbol{q}_h, \nabla w_1)_{\mathcal{T}_h} + \langle \boldsymbol{q}_h\cdot\boldsymbol{n} + h^{-1} P_M y_h, w_1\rangle_{\partial\mathcal{T}_h}\\
&\quad - \langle h^{-1}\widehat{y}_h^o, w_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} - \langle \boldsymbol{q}_h\cdot\boldsymbol{n} + h^{-1}(P_M y_h - \widehat{y}_h^o), \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}.
\end{aligned} \tag{2.22}
$$

By the definition of $\mathscr{B}$, we can rewrite the HDG formulation of the optimality system (2.11) as follows: find $(\boldsymbol{q}_h, \boldsymbol{p}_h, y_h, z_h, \widehat{y}_h^o, \widehat{z}_h^o, u_h) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times W_h \times M_h(o) \times M_h(o) \times M_h(\partial)$ such that

$$\mathscr{B}(\boldsymbol{q}_h, y_h, \widehat{y}_h^o; \boldsymbol{r}_1, w_1, \mu_1) = -\langle u_h, \boldsymbol{r_1} \cdot \boldsymbol{n} - h^{-1} w_1 \rangle_{\varepsilon_h^\partial} + (f, w_1)_{\mathcal{T}_h}, \tag{2.23a}$$

$$\mathscr{B}(\boldsymbol{p}_h, z_h, \widehat{z}_h^o; \boldsymbol{r}_2, w_2, \mu_2) = (y_h - y_d, w_2)_{\mathcal{T}_h}, \tag{2.23b}$$

$$\gamma^{-1} \langle \boldsymbol{p}_h \cdot \boldsymbol{n} + h^{-1} P_M z_h, \mu_3 \rangle_{\varepsilon_h^\partial} = -\langle u_h, \mu_3 \rangle_{\varepsilon_h^\partial}, \tag{2.23c}$$

for all $(\boldsymbol{r}_1, \boldsymbol{r}_2, w_1, w_2, \mu_1, \mu_2, \mu_3) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times W_h \times M_h(o) \times M_h(o) \times M_h(\partial)$.

Next, we present a basic property of the operator $\mathscr{B}$ and show the HDG equations (2.23) have a unique solution.

**Lemma 1** *For any $(\boldsymbol{v}_h, w_h, \mu_h) \in \boldsymbol{V}_h \times W_h \times M_h$, we have*

$$\begin{aligned} \mathscr{B}(\boldsymbol{v}_h, w_h, \mu_h; \boldsymbol{v}_h, w_h, \mu_h) = (\boldsymbol{v}_h, \boldsymbol{v}_h)_{\mathcal{T}_h} + \langle h^{-1}(P_M w_h - \mu_h), P_M w_h - \mu_h \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} \\ + \langle h^{-1} P_M w_h, P_M w_h \rangle_{\varepsilon_h^\partial}. \end{aligned}$$

**Proof:** By the definition of $\mathscr{B}$ in (2.22), we have

$$\begin{aligned} &\mathscr{B}(\boldsymbol{v}_h, w_h, \mu_h; \boldsymbol{v}_h, w_h, \mu_h) \\ &\quad = (\boldsymbol{v}_h, \boldsymbol{v}_h)_{\mathcal{T}_h} - (w_h, \nabla \cdot \boldsymbol{v}_h)_{\mathcal{T}_h} + \langle \mu_h, \boldsymbol{v}_h \cdot \boldsymbol{n} \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} - (\boldsymbol{v}_h, \nabla w_h)_{\mathcal{T}_h} \\ &\qquad + \langle \boldsymbol{v}_h \cdot \boldsymbol{n} + h^{-1} P_M w_h, w_h \rangle_{\partial \mathcal{T}_h} - \langle h^{-1} \mu_h, w_h \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} \\ &\qquad - \langle \boldsymbol{v}_h \cdot \boldsymbol{n} + h^{-1}(P_M w_h - \mu_h), \mu_h \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} \\ &\quad = (\boldsymbol{v}_h, \boldsymbol{v}_h)_{\mathcal{T}_h} + \langle h^{-1} P_M w_h, w_h \rangle_{\partial \mathcal{T}_h} - \langle h^{-1} \mu_h, w_h \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} \\ &\qquad - \langle h^{-1}(P_M w_h - \mu_h), \mu_h \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} \\ &\quad = (\boldsymbol{v}_h, \boldsymbol{v}_h)_{\mathcal{T}_h} + \langle h^{-1}(P_M w_h - \mu_h, P_M w_h - \mu_h \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} + \langle h^{-1} P_M w_h, P_M w_h \rangle_{\varepsilon_h^\partial}. \end{aligned}$$

**Proposition 1** *There exists a unique solution of the HDG equations* (2.23).

**Proof:** Since the system (2.23) is finite dimensional, we only need to prove the uniqueness. Therefore, we assume $y_d = f = 0$ and we show the system (2.23) only has the trivial solution.

First, by the definition of $\mathscr{B}$, we have

$$
\begin{aligned}
&\mathscr{B}(\boldsymbol{q}_h, y_h, \widehat{y}_h^o; \boldsymbol{p}_h, -z_h, -\widehat{z}_h^o) + \mathscr{B}(\boldsymbol{p}_h, z_h, \widehat{z}_h^o; -\boldsymbol{q}_h, y_h, \widehat{y}_h^o)\\
&\quad = (\boldsymbol{q}_h, \boldsymbol{p}_h)_{\mathcal{T}_h} - (y_h, \nabla\cdot \boldsymbol{p}_h)_{\mathcal{T}_h} + \langle \widehat{y}_h^o, \boldsymbol{p}_h\cdot \boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} + (\boldsymbol{q}_h, \nabla z_h)_{\mathcal{T}_h}\\
&\qquad - \langle \boldsymbol{q}_h\cdot\boldsymbol{n} + h^{-1}P_M y_h, z_h\rangle_{\partial\mathcal{T}_h} + \langle h^{-1}\widehat{y}_h^o, z_h\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\qquad + \langle \boldsymbol{q}_h\cdot\boldsymbol{n} + h^{-1}(P_M y_h - \widehat{y}_h^o), \widehat{z}_h^o\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} - (\boldsymbol{p}_h, \boldsymbol{q}_h)_{\mathcal{T}_h} + (z_h, \nabla\cdot\boldsymbol{q}_h)_{\mathcal{T}_h}\\
&\qquad - \langle \widehat{z}_h^o, \boldsymbol{q}_h\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} - (\boldsymbol{p}_h, \nabla y_h)_{\mathcal{T}_h} + \langle \boldsymbol{p}_h\cdot\boldsymbol{n} + h^{-1}P_M z_h, y_h\rangle_{\partial\mathcal{T}_h}\\
&\qquad - \langle h^{-1}\widehat{z}_h^o, y_h\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} - \langle \boldsymbol{p}_h\cdot\boldsymbol{n} + h^{-1}(P_M z_h - \widehat{z}_h^o), \widehat{y}_h^o\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}.
\end{aligned}
$$

Integrating by parts and using the properties of $P_M$ in (2.19) gives

$$\mathscr{B}(\boldsymbol{q}_h, y_h, \widehat{y}_h^o; \boldsymbol{p}_h, -z_h, -\widehat{z}_h^o) + \mathscr{B}(\boldsymbol{p}_h, z_h, \widehat{z}_h^o; -\boldsymbol{q}_h, y_h, \widehat{y}_h^o) = 0.$$

Next, take $(\boldsymbol{r}_1, w_1, \mu_1) = (\boldsymbol{p}_h, -z_h, -\widehat{z}_h^o)$, $(\boldsymbol{r}_2, w_2, \mu_2) = (-\boldsymbol{q}_h, y_h, \widehat{y}_h^o)$, and $\mu_3 = -\gamma u_h$ in the HDG equations (2.23a), (2.23b), and (2.23c), respectively, and sum to obtain

$$(y_h, y_h)_{\mathcal{T}_h} + \gamma\,\|u_h\|^2_{\varepsilon_h^\partial} = 0.$$

This implies $y_h = 0$ and $u_h = 0$ since $\gamma > 0$.

Next, taking $(\boldsymbol{r}_1, w_1, \mu_1) = (\boldsymbol{q}_h, y_h, \widehat{y}_h^o)$ and $(\boldsymbol{r}_2, w_2, \mu_2) = (\boldsymbol{p}_h, z_h, \widehat{z}_h^o)$ in Lemma 1 gives $\boldsymbol{q}_h = \boldsymbol{p}_h = \boldsymbol{0}, \widehat{y}_h^o = 0$, $P_M z_h = 0$ on $\varepsilon_h^\partial$, and $P_M z_h - \widehat{z}_h^o = 0$ on $\partial\mathcal{T}_h\backslash\varepsilon_h^\partial$. Also, since $\widehat{z}_h = 0$ on $\varepsilon_h^\partial$ we have

$$P_M z_h - \widehat{z}_h = 0. \tag{2.24}$$

Substituting (2.24) into (2.11c), and remembering again $\widehat{z}_h = 0$ on $\varepsilon_h^\partial$, we get

$$-(z_h, \nabla\cdot \boldsymbol{r_2})_{\mathcal{T}_h} + \langle P_M z_h, \boldsymbol{r_2}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h} = 0.$$

Use the property of $P_M$ in (2.19), integrate by parts, and take $\boldsymbol{r}_2 = \nabla z_h$ to obtain

$$(\nabla z_h, \nabla z_h)_{\mathcal{T}_h} = 0.$$

Thus, $z_h$ is constant on each $K \in \mathcal{T}_h$, and also $z_h = P_M z_h = \widehat{z}_h$ on $\partial\mathcal{T}_h$. Since $\widehat{z}_h = 0$ on $\varepsilon_h^\partial$ and single valued on each face, we have $z_h = 0$ on each $K \in \mathcal{T}_h$, and therefore also $\widehat{z}_h^o = 0$.

**2.4.3. Proof of Main Result.** To prove the main result, we follow a similar strategy taken by Gong and Yan [33], see also [18, 45, 50], and introduce an auxiliary problem with the approximate control $u_h$ in (2.23a) replaced by a projection of the exact optimal control. We first bound the error between the solutions of the auxiliary problem and the mixed weak form (2.4a)-(2.4e) of the optimality system. The we bound the error between the solutions of the auxiliary problem and the HDG problem (2.23). A simple application of the triangle inequality then gives a bound on the error between the solutions of the HDG problem and then mixed form of the optimality system.

The precise form of the auxiliary problem is given as follows: find $(\boldsymbol{q}_h(u),$ $\boldsymbol{p}_h(u), y_h(u), z_h(u), \widehat{y}_h^o(u), \widehat{z}_h^o(u)) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times W_h \times M_h(o) \times M_h(o)$ such that

$$\mathscr{B}(\boldsymbol{q}_h(u), y_h(u), \widehat{y}_h^o(u); \boldsymbol{r}_1, w_1, \mu_1) = -\langle P_M u, \boldsymbol{r_1}\cdot\boldsymbol{n} - h^{-1}w_1\rangle_{\varepsilon_h^\partial} + (f, w_1)_{\mathcal{T}_h}, \tag{2.25a}$$

$$\mathscr{B}(\boldsymbol{p}_h(u), z_h(u), \widehat{z}_h^o(u); \boldsymbol{r}_2, w_2, \mu_2) = (y_h(u) - y_d, w_2)_{\mathcal{T}_h}. \tag{2.25b}$$

for all $(\boldsymbol{r}_1, \boldsymbol{r}_2, w_1, w_2, \mu_1, \mu_2) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times W_h \times M_h(o) \times M_h(o)$.

We split the proof of the main result, Theorem 2, in 7 steps. We begin by bounding the error between the solutions of the auxiliary problem and the mixed form (2.4a)-(2.4e) of the optimality system. We split the errors in the variables using the $L^2$ projections. In steps 1-3, we focus on the primary variables, i.e., the state $y$ and the flux $\boldsymbol{q}$, and we use the following notation:

$$
\begin{aligned}
\delta^{\boldsymbol{q}} &= \boldsymbol{q} - \boldsymbol{\Pi}\boldsymbol{q}, & \varepsilon_h^{\boldsymbol{q}} &= \boldsymbol{\Pi}\boldsymbol{q} - \boldsymbol{q}_h(u),\\
\delta^{y} &= y - \Pi y, & \varepsilon_h^{y} &= \Pi y - y_h(u),\\
\delta^{\widehat{y}} &= y - P_M y, & \varepsilon_h^{\widehat{y}} &= P_M y - \widehat{y}_h(u),\\
\widehat{\boldsymbol{\delta}}_1 &= \delta^{\boldsymbol{q}}\cdot\boldsymbol{n} + h^{-1}P_M\delta^{y}, & \widehat{\boldsymbol{\varepsilon}}_1 &= \varepsilon_h^{\boldsymbol{q}}\cdot\boldsymbol{n} + h^{-1}(P_M\varepsilon_h^{y} - \varepsilon_h^{\widehat{y}}),
\end{aligned}
\tag{2.26}
$$

where $\widehat{y}_h(u) = \widehat{y}_h^{o}(u)$ on $\varepsilon_h^{o}$ and $\widehat{y}_h(u) = P_M u$ on $\varepsilon_h^{\partial}$. Note that this implies $\varepsilon_h^{\widehat{y}} = 0$ on $\varepsilon_h^{\partial}$.

Step 1: The error equation for part 1 of the auxiliary problem (2.25a)

**Lemma 2** *We have*

$$
\mathscr{B}(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^{y}, \varepsilon_h^{\widehat{y}}; \boldsymbol{r}_1, w_1, \mu_1) = -\langle\widehat{\boldsymbol{\delta}}_1, w_1\rangle_{\partial\mathcal{T}_h} + \langle\widehat{\boldsymbol{\delta}}_1, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^{\partial}}. \tag{2.27}
$$

**Proof:** By the definition of the operator $\mathscr{B}$ in (2.22), we have

$$
\begin{aligned}
&\mathscr{B}(\boldsymbol{\Pi}\boldsymbol{q}, \Pi y, P_M y; \boldsymbol{r}_1, w_1, \mu_1)\\
&\quad= (\boldsymbol{\Pi}\boldsymbol{q}, \boldsymbol{r}_1)_{\mathcal{T}_h} - (\Pi y, \nabla\cdot\boldsymbol{r}_1)_{\mathcal{T}_h} + \langle P_M y, \boldsymbol{r}_1\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^{\partial}}\\
&\qquad - (\boldsymbol{\Pi}\boldsymbol{q}, \nabla w_1)_{\mathcal{T}_h} + \langle\boldsymbol{\Pi}\boldsymbol{q}\cdot\boldsymbol{n} + h^{-1}P_M\Pi y, w_1\rangle_{\partial\mathcal{T}_h}\\
&\qquad - \langle h^{-1}P_M y, w_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^{\partial}} - \langle\boldsymbol{\Pi}\boldsymbol{q}\cdot\boldsymbol{n} - h^{-1}P_M\delta^{y}, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^{\partial}}.
\end{aligned}
$$

By properties of the $L^2$ projections (2.19), we have

$$\begin{aligned}\mathscr{B}(\boldsymbol{\Pi q}, \Pi y, P_M y; \boldsymbol{r}_1, w_1, \mu_1) = & (\boldsymbol{q}, \boldsymbol{r_1})_{\mathcal{T}_h} - (y, \nabla \cdot \boldsymbol{r_1})_{\mathcal{T}_h} + \langle y, \boldsymbol{r_1} \cdot \boldsymbol{n} \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} \\ & - (\boldsymbol{q}, \nabla w_1)_{\mathcal{T}_h} + \langle \boldsymbol{q} \cdot \boldsymbol{n}, w_1 \rangle_{\partial \mathcal{T}_h} - \langle \delta^{\boldsymbol{q}} \cdot \boldsymbol{n}, w_1 \rangle_{\partial \mathcal{T}_h} \\ & + \langle h^{-1} P_M \Pi y, w_1 \rangle_{\partial \mathcal{T}_h} - \langle h^{-1} P_M y, w_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} \\ & - \langle \boldsymbol{q} \cdot \boldsymbol{n}, \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} + \langle \delta^{\boldsymbol{q}} \cdot \boldsymbol{n}, \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} \\ & + \langle h^{-1} P_M \delta^y, \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial}.\end{aligned}$$

Note that the exact state $y$ and exact flux $\boldsymbol{q}$ satisfy

$$\begin{aligned}(\boldsymbol{q}, \boldsymbol{r}_1)_{\mathcal{T}_h} - (y, \nabla \cdot \boldsymbol{r}_1)_{\mathcal{T}_h} + \langle y, \boldsymbol{r}_1 \cdot \boldsymbol{n} \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} &= - \langle u, \boldsymbol{r}_1 \cdot \boldsymbol{n} \rangle_{\varepsilon_h^\partial}, \\ -(\boldsymbol{q}, \nabla w_1)_{\mathcal{T}_h} + \langle \boldsymbol{q} \cdot \boldsymbol{n}, w_1 \rangle_{\partial \mathcal{T}_h} &= (f, w_1)_{\mathcal{T}_h}, \\ \langle \boldsymbol{q} \cdot \boldsymbol{n}, \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} &= 0,\end{aligned}$$

for all $(\boldsymbol{r}_1, w_1, \mu_1) \in \boldsymbol{V}_h \times W_h \times M_h(o)$. Then we have

$$\begin{aligned}\mathscr{B}(\boldsymbol{\Pi q}, \Pi y, P_M y; \boldsymbol{r}_1, w_1, \mu_1) = & - \langle u, \boldsymbol{r}_1 \cdot \boldsymbol{n} \rangle_{\varepsilon_h^\partial} + (f, w_1)_{\mathcal{T}_h} - \langle \delta^{\boldsymbol{q}} \cdot \boldsymbol{n}, w_1 \rangle_{\partial \mathcal{T}_h} \\ & + \langle h^{-1} P_M \Pi y, w_1 \rangle_{\partial \mathcal{T}_h} - \langle h^{-1} P_M y, w_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} \\ & + \langle \delta^{\boldsymbol{q}} \cdot \boldsymbol{n}, \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} + \langle h^{-1} P_M \delta^y, \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial}.\end{aligned}$$

Subtract part 1 of the auxiliary problem (2.25a) from the above equality to obtain the result:

$$\begin{aligned}\mathscr{B}(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^y, \varepsilon_h^{\widehat{y}}; \boldsymbol{r}_1, w_1, \mu_1) = & - \langle P_M u, h^{-1} w_1 \rangle_{\varepsilon_h^\partial} - \langle \delta^{\boldsymbol{q}} \cdot \boldsymbol{n}, w_1 \rangle_{\partial \mathcal{T}_h} \\ & + \langle h^{-1} P_M \Pi y, w_1 \rangle_{\partial \mathcal{T}_h} - \langle h^{-1} P_M y, w_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} \\ & + \langle \delta^{\boldsymbol{q}} \cdot \boldsymbol{n}, \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} + \langle h^{-1} P_M \delta^y, \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} \\ = & - \langle \widehat{\boldsymbol{\delta}}_1, w_1 \rangle_{\partial \mathcal{T}_h} + \langle \widehat{\boldsymbol{\delta}}_1, \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial}.\end{aligned}$$

Step 2: Estimate for $\varepsilon_h^{\boldsymbol{q}}$ We first provide a key inequality which was proven in [56].

**Lemma 3** *We have*

$$\|\nabla \varepsilon_h^y\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h} \lesssim \|\varepsilon_h^{\boldsymbol{q}}\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|P_M\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h}.$$

**Lemma 4** *We have*

$$\left\|\varepsilon_h^{\boldsymbol{q}}\right\|_{\mathcal{T}_h}^2 + h^{-1}\|P_M\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h}^2 \lesssim h^{2s_{\boldsymbol{q}}}\,\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega}^2 + h^{2s_y-2}\,\|y\|_{s^y,\Omega}^2. \tag{2.28}$$

**Proof:** First, since $\varepsilon_h^{\widehat{y}} = 0$ on $\varepsilon_h^{\partial}$, the basic property of $\mathscr{B}$ in Lemma 1 gives

$$\mathscr{B}(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^y, \varepsilon_h^{\widehat{y}}; \varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^y, \varepsilon_h^{\widehat{y}}) = (\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^{\boldsymbol{q}})_{\mathcal{T}_h} + h^{-1}\|P_M\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h}^2.$$

Then, taking $(\boldsymbol{r}_1, w_1, \mu_1) = (\boldsymbol{\varepsilon}_h^{\boldsymbol{q}}, \varepsilon_h^y, \varepsilon_h^{\widehat{y}})$ in (2.27) in Lemma 2 gives

$$\begin{aligned}
(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^{\boldsymbol{q}})_{\mathcal{T}_h} &+ h^{-1}\|P_M\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h}^2 \\
&= -\langle \widehat{\boldsymbol{\delta}}_1, \varepsilon_h^y - \varepsilon_h^{\widehat{y}}\rangle_{\partial\mathcal{T}_h} \\
&= -\langle \delta^{\boldsymbol{q}}\cdot\boldsymbol{n}, \varepsilon_h^y - \varepsilon_h^{\widehat{y}}\rangle_{\partial\mathcal{T}_h} - h^{-1}\langle \delta^y, P_M\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\rangle_{\partial\mathcal{T}_h} \\
&\le \|\delta^{\boldsymbol{q}}\|_{\partial\mathcal{T}_h}\,\|\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h} + h^{-1}\|\delta^y\|_{\partial\mathcal{T}_h}\|P_M\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h} \\
&\le h^{1/2}\,\|\delta^{\boldsymbol{q}}\|_{\partial\mathcal{T}_h}\,h^{-1/2}\|\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h} + h^{-1}\|\delta^y\|_{\partial\mathcal{T}_h}\|P_M\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h}.
\end{aligned}$$

By Young's inequality and Lemma 3, we obtain

$$\begin{aligned}
\|\varepsilon_h^{\boldsymbol{q}}\|_{\mathcal{T}_h}^2 + h^{-1}\|P_M\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h}^2 &\lesssim h\,\|\delta^{\boldsymbol{q}}\|_{\partial\mathcal{T}_h}^2 + h^{-1}\,\|\delta^y\|_{\partial\mathcal{T}_h}^2 \\
&\lesssim h^{2s_{\boldsymbol{q}}}\,\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega}^2 + h^{2s_y-2}\,\|y\|_{s^y,\Omega}^2.
\end{aligned}$$

Step 3: Estimate for $\varepsilon_h^y$ by a duality argument Next, we introduce the dual problem for any given $\Theta$ in $L^2(\Omega)$ :

$$\begin{aligned} \boldsymbol{\Phi} + \nabla\Psi &= 0 && \text{in } \Omega, \\ \nabla\cdot\boldsymbol{\Phi} &= \Theta && \text{in } \Omega, \\ \Psi &= 0 && \text{on } \Gamma. \end{aligned} \tag{2.29}$$

Since the domain $\Omega$ is convex, we have the following regularity estimate

$$\|\boldsymbol{\Phi}\|_{H^1(\Omega)} + \|\Psi\|_{H^2(\Omega)} \leq C\,\|\Theta\|_{\Omega}\,. \tag{2.30}$$

Before we estimate $\varepsilon_h^y$ we introduce the following notation, which is similar to the earlier notation in (2.26):

$$\delta^{\boldsymbol{\Phi}} = \boldsymbol{\Phi} - \boldsymbol{\Pi}\boldsymbol{\Phi}, \quad \delta^{\Psi} = \Psi - \Pi\Psi, \quad \delta^{\widehat{\Psi}} = \Psi - P_M\Psi. \tag{2.31}$$

By the regularity estimate (2.30), we have the following bounds:

$$\|\delta^{\boldsymbol{\Phi}}\|_{\mathcal{T}_h} \lesssim h\|\Theta\|_{\mathcal{T}_h}, \quad \|\delta^{\Psi}\|_{\mathcal{T}_h} \lesssim h^2\|\Theta\|_{\mathcal{T}_h}, \quad \|\delta^{\widehat{\Psi}}\|_{\partial\mathcal{T}_h} \lesssim h^{\frac{1}{2}}\|\Theta\|_{\mathcal{T}_h}. \tag{2.32}$$

**Lemma 5** *We have*

$$\|\varepsilon_h^y\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{q}}+1}\,\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y}\,\|y\|_{s^y,\Omega}\,. \tag{2.33}$$

**Proof:** Consider the dual problem (2.29) and let $\Theta = \varepsilon_h^y$. In the definition (2.22) of $\mathscr{B}$, take $(\boldsymbol{r}_1, w_1, \mu_1)$ to be $(-\boldsymbol{\Pi}\boldsymbol{\Phi}, \Pi\Psi, P_M\Psi)$ and use $\Psi = 0$ on $\varepsilon_h^\partial$ to obtain

$$\begin{aligned} \mathscr{B}(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^y, \varepsilon_h^{\widehat{y}}; -\boldsymbol{\Pi}\boldsymbol{\Phi}, \Pi\Psi, P_M\Psi) = &-(\varepsilon_h^{\boldsymbol{q}}, \boldsymbol{\Pi}\boldsymbol{\Phi})_{\mathcal{T}_h} + (\varepsilon_h^y, \nabla\cdot\boldsymbol{\Pi}\boldsymbol{\Phi})_{\mathcal{T}_h} - \langle\varepsilon_h^{\widehat{y}}, \boldsymbol{\Pi}\boldsymbol{\Phi}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h} \\ &-(\varepsilon_h^{\boldsymbol{q}}, \nabla\Pi\Psi)_{\mathcal{T}_h} + \langle\widehat{\boldsymbol{\varepsilon}}_1, \Pi\Psi\rangle_{\partial\mathcal{T}_h} - \langle\widehat{\boldsymbol{\varepsilon}}_1, P_M\Psi\rangle_{\partial\mathcal{T}_h}. \end{aligned} \tag{2.34}$$

Next, it is easy to verify that

$$
\begin{aligned}
(\varepsilon_h^y, \nabla\cdot\boldsymbol{\Pi\Phi})_{\mathcal{T}_h} &= \langle \varepsilon_h^y, \boldsymbol{\Pi\Phi}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h} - (\nabla\varepsilon_h^y, \boldsymbol{\Pi\Phi})_{\mathcal{T}_h} \\
&= \langle \varepsilon_h^y, \boldsymbol{\Pi\Phi}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h} - (\nabla\varepsilon_h^y, \boldsymbol{\Phi})_{\mathcal{T}_h} \\
&= -\langle \varepsilon_h^y, \delta^{\boldsymbol{\Phi}}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h} + (\varepsilon_h^y, \nabla\cdot\boldsymbol{\Phi})_{\mathcal{T}_h} \\
&= -\langle \varepsilon_h^y, \delta^{\boldsymbol{\Phi}}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h} + \|\varepsilon_h^y\|_{\mathcal{T}_h}^2.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
-(\varepsilon_h^{\boldsymbol{q}}, \nabla\Pi\Psi)_{\mathcal{T}_h} &= -\langle \varepsilon_h^{\boldsymbol{q}}\cdot\boldsymbol{n}, \Pi\Psi\rangle_{\partial\mathcal{T}_h} + (\nabla\cdot\varepsilon_h^{\boldsymbol{q}}, \Pi\Psi)_{\mathcal{T}_h} \\
&= -\langle \varepsilon_h^{\boldsymbol{q}}\cdot\boldsymbol{n}, \Pi\Psi\rangle_{\partial\mathcal{T}_h} + (\nabla\cdot\varepsilon_h^{\boldsymbol{q}}, \Psi)_{\mathcal{T}_h} \\
&= -\langle \varepsilon_h^{\boldsymbol{q}}\cdot\boldsymbol{n}, \Pi\Psi\rangle_{\partial\mathcal{T}_h} + \langle \varepsilon_h^{\boldsymbol{q}}\cdot\boldsymbol{n}, \Psi\rangle_{\partial\mathcal{T}_h} - (\varepsilon_h^{\boldsymbol{q}}, \nabla\Psi)_{\mathcal{T}_h} \\
&= \langle \varepsilon_h^{\boldsymbol{q}}\cdot\boldsymbol{n}, (P_M\Psi - \Pi\Psi)\rangle_{\partial\mathcal{T}_h} - (\varepsilon_h^{\boldsymbol{q}}, \nabla\Psi)_{\mathcal{T}_h}.
\end{aligned}
$$

Then equation (2.34) becomes

$$
\begin{aligned}
&\mathscr{B}(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^y, \varepsilon_h^{\widehat{y}}; -\boldsymbol{\Pi\Phi}, \Pi\Psi, P_M\Psi) \\
&\quad = -(\varepsilon_h^{\boldsymbol{q}}, \boldsymbol{\Phi})_{\mathcal{T}_h} - \langle \varepsilon_h^y, \delta^{\boldsymbol{\Phi}}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h} + \|\varepsilon_h^y\|_{\mathcal{T}_h}^2 - \langle \varepsilon_h^{\widehat{y}}, \boldsymbol{\Pi\Phi}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h} \\
&\quad + \langle \varepsilon_h^{\boldsymbol{q}}\cdot\boldsymbol{n}, P_M\Psi - \Pi\Psi\rangle_{\partial\mathcal{T}_h} - (\varepsilon_h^{\boldsymbol{q}}, \nabla\Psi)_{\mathcal{T}_h} + \langle \widehat{\boldsymbol{\varepsilon}}_1, \Pi\Psi\rangle_{\partial\mathcal{T}_h} - \langle \widehat{\boldsymbol{\varepsilon}}_1, P_M\Psi\rangle_{\partial\mathcal{T}_h}.
\end{aligned}
$$

The facts $\boldsymbol{\Phi} + \nabla\Psi = 0$, $\langle \varepsilon_h^{\widehat{y}}, \boldsymbol{\Phi}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h} = 0$, and $\langle \widehat{\boldsymbol{\varepsilon}}_1, P_M\Psi\rangle_{\partial\mathcal{T}_h} = \langle \widehat{\boldsymbol{\varepsilon}}_1, \Psi\rangle_{\partial\mathcal{T}_h}$ imply

$$
\begin{aligned}
&\mathscr{B}(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^y, \varepsilon_h^{\widehat{y}}; -\boldsymbol{\Pi\Phi}, \Pi\Psi, P_M\Psi) \\
&\qquad = -\langle \varepsilon_h^y - \varepsilon_h^{\widehat{y}}, \delta^{\boldsymbol{\Phi}}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h} + \|\varepsilon_h^y\|_{\mathcal{T}_h}^2 - h^{-1}\langle P_M\varepsilon_h^y - \varepsilon_h^{\widehat{y}}, \delta^{\Psi}\rangle_{\partial\mathcal{T}_h}.
\end{aligned}
$$

On the other hand, equation (2.27) in Lemma 2 gives

$$\mathscr{B}(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^y, \varepsilon_h^{\widehat{y}}; -\boldsymbol{\Pi\Phi}, \Pi\Psi, P_M\Psi) = -\langle \widehat{\boldsymbol{\delta}}_1, \Pi\Psi\rangle_{\partial\mathcal{T}_h} + \langle \widehat{\boldsymbol{\delta}}_1, P_M\Psi\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}.$$

Moreover,

$$\begin{aligned}
&\langle \widehat{\boldsymbol{\delta}}_1, P_M\Psi\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\quad= \langle \delta^{\boldsymbol{q}} \cdot \boldsymbol{n} + h^{-1}P_M\delta^y, P_M\Psi\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\quad= \langle \boldsymbol{q} \cdot \boldsymbol{n}, P_M\Psi\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} - \langle \boldsymbol{\Pi q} \cdot \boldsymbol{n}, P_M\Psi\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} + \langle h^{-1}P_M\delta^y, P_M\Psi\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\quad= -\langle \boldsymbol{\Pi q} \cdot \boldsymbol{n}, \Psi\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} + \langle h^{-1}P_M\delta^y, \Psi\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\quad= \langle \boldsymbol{q} \cdot \boldsymbol{n}, \Psi\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} - \langle \boldsymbol{\Pi q} \cdot \boldsymbol{n}, \Psi\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} + \langle h^{-1}P_M\delta^y, \Psi\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\quad= \langle \widehat{\boldsymbol{\delta}}_1, \Psi\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\quad= \langle \widehat{\boldsymbol{\delta}}_1, \Psi\rangle_{\partial\mathcal{T}_h},
\end{aligned}$$

where we have used $\langle \boldsymbol{q} \cdot \boldsymbol{n}, P_M\Psi\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} = 0$, $\langle \boldsymbol{q} \cdot \boldsymbol{n}, \Psi\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} = 0$ since $\boldsymbol{q} \in H(\mathrm{div}, \Omega)$ and $\Psi = 0$ on $\varepsilon_h^\partial$.

Comparing the above two equalities gives

$$\begin{aligned}
\|\varepsilon_h^y\|_{\mathcal{T}_h}^2 &= \langle \varepsilon_h^y - \varepsilon_h^{\widehat{y}}, \delta^{\boldsymbol{\Phi}} \cdot \boldsymbol{n}\rangle_{\partial\mathcal{T}_h} + h^{-1}\langle P_M\varepsilon_h^y - \varepsilon_h^{\widehat{y}}, \delta^\Psi\rangle_{\partial\mathcal{T}_h} + \langle \widehat{\boldsymbol{\delta}}_1, \delta^\Psi\rangle_{\partial\mathcal{T}_h}\\
&= \langle \varepsilon_h^y - \varepsilon_h^{\widehat{y}}, \delta^{\boldsymbol{\Phi}} \cdot \boldsymbol{n}\rangle_{\partial\mathcal{T}_h} + h^{-1}\langle P_M\varepsilon_h^y - \varepsilon_h^{\widehat{y}}, \delta^\Psi\rangle_{\partial\mathcal{T}_h}\\
&\quad - \langle \delta^{\boldsymbol{q}} \cdot \boldsymbol{n} + h^{-1}P_M\delta^y, \delta^\Psi\rangle_{\partial\mathcal{T}_h}\\
&\lesssim h^{-\frac{1}{2}}\|\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h} \cdot h^{\frac{1}{2}}\|\delta^{\boldsymbol{\Phi}}\|_{\partial\mathcal{T}_h} + h^{-\frac{1}{2}}\|P_M\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h} \cdot h^{-\frac{1}{2}}\|\delta^\Psi\|_{\partial\mathcal{T}_h}\\
&\quad + \|\delta^{\boldsymbol{q}}\|_{\partial\mathcal{T}_h} \cdot \|\delta^\Psi\|_{\partial\mathcal{T}_h} + h^{-1}\|\delta^y\|_{\partial\mathcal{T}_h} \cdot \|\delta^\Psi\|_{\partial\mathcal{T}_h}\\
&\lesssim (h^{s_{\boldsymbol{q}}+1}\,\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y}\,\|y\|_{s^y,\Omega})\|\varepsilon_h^y\|_{\mathcal{T}_h}.
\end{aligned}$$

As a consequence of Lemma 4 and Lemma 5, a simple application of the triangle inequality gives optimal convergence rates for $\|\boldsymbol{q} - \boldsymbol{q}_h(u)\|_{\mathcal{T}_h}$ and $\|y - y_h(u)\|_{\mathcal{T}_h}$:

**Lemma 6**

$$\|\boldsymbol{q}-\boldsymbol{q}_h(u)\|_{\mathcal{T}_h} \le \|\delta^{\boldsymbol{q}}\|_{\mathcal{T}_h} + \|\varepsilon_h^{\boldsymbol{q}}\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{q}}}\,\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y-1}\,\|y\|_{s^y,\Omega}, \tag{2.35a}$$

$$\|y-y_h(u)\|_{\mathcal{T}_h} \le \|\delta^{y}\|_{\mathcal{T}_h} + \|\varepsilon_h^{y}\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{q}}+1}\,\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y}\,\|y\|_{s^y,\Omega}. \tag{2.35b}$$

Step 4: The error equation for part 2 of the auxiliary problem (2.25b) We continue to bound the error between the solutions of the auxiliary problem and the mixed form (2.4a)-(2.4e) of the optimality system. In steps 4-5, we focus on the dual variables, i.e., the state $z$ and the flux $\boldsymbol{p}$. We split the errors in the variables using the $L^2$ projections, and we use the following notation.

$$\begin{aligned}
\delta^{\boldsymbol{p}} &= \boldsymbol{p} - \boldsymbol{\Pi p}, & \varepsilon_h^{\boldsymbol{p}} &= \boldsymbol{\Pi p} - \boldsymbol{p}_h(u),\\
\delta^{z} &= z - \Pi z, & \varepsilon_h^{z} &= \Pi z - z_h(u),\\
\delta^{\widehat{z}} &= z - P_M z, & \varepsilon_h^{\widehat{z}} &= P_M z - \widehat{z}_h(u),\\
\widehat{\boldsymbol{\delta}}_2 &= \delta^{\boldsymbol{p}}\cdot\boldsymbol{n} + h^{-1}P_M\delta^{z}.
\end{aligned} \tag{2.36}$$

where $\widehat{z}_h(u) = \widehat{z}_h^o(u)$ on $\varepsilon_h^o$ and $\widehat{z}_h(u) = 0$ on $\varepsilon_h^\partial$. Note that this implies $\varepsilon_h^{\widehat{z}} = 0$ on $\varepsilon_h^\partial$.

The derivation of the error equation for part 2 of the auxiliary problem (2.25b) is similar to the analysis for part 1 of the auxiliary problem in step 1 in 2.4.3; the only difference is there is one more term $(y-y_h(u), w_2)_{\mathcal{T}_h}$ in the right hand side. Therefore, we state the result and omit the proof.

**Lemma 7** *We have*

$$\mathscr{B}(\varepsilon_h^{\boldsymbol{p}}, \varepsilon_h^{z}, \varepsilon_h^{\widehat{z}}, \boldsymbol{r}_2, w_2, \mu_2) = -\langle\widehat{\boldsymbol{\delta}}_2, w_2\rangle_{\partial\mathcal{T}_h} + \langle\widehat{\boldsymbol{\delta}}_2, \mu_2\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} + (y-y_h(u), w_2)_{\mathcal{T}_h}. \tag{2.37}$$

Step 5: Estimate for $\varepsilon_h^{\boldsymbol{p}}$ and $\varepsilon_h^{z}$ Before we estimate $\varepsilon_h^{\boldsymbol{p}}$, we give the following discrete Poincaré inequality from [56].

**Lemma 8** *Since $\varepsilon_h^{\widehat{z}} = 0$ on $\varepsilon_h^{\partial}$, we have*

$$\|\varepsilon_h^z\|_{\mathcal{T}_h} \lesssim \|\nabla \varepsilon_h^z\|_{\mathcal{T}_h} + h^{-\frac{1}{2}} \|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial \mathcal{T}_h}. \tag{2.38}$$

**Lemma 9** *We have*

$$\begin{aligned}
&\left\|\varepsilon_h^{\boldsymbol{p}}\right\|_{\mathcal{T}_h} + h^{-\frac{1}{2}} \|P_M \varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial \mathcal{T}_h} \\
&\qquad \lesssim h^{s_{\boldsymbol{p}}} \left\|\boldsymbol{p}\right\|_{s^{\boldsymbol{p}},\Omega} + h^{s_z - 1} \left\|z\right\|_{s^z,\Omega} + h^{s_{\boldsymbol{q}}+1} \left\|\boldsymbol{q}\right\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y} \left\|y\right\|_{s^y,\Omega}, \\
&\|\varepsilon_h^z\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}} \left\|\boldsymbol{p}\right\|_{s^{\boldsymbol{p}},\Omega} + h^{s_z - 1} \left\|z\right\|_{s^z,\Omega} + h^{s_{\boldsymbol{q}}+1} \left\|\boldsymbol{q}\right\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y} \left\|y\right\|_{s^y,\Omega}.
\end{aligned}$$

**Proof:** First, we note the key inequality in Lemma 3 is valid with $(z, \boldsymbol{p}, \hat{z})$ in place of $(y, \boldsymbol{q}, \hat{y})$. This gives

$$\|\nabla \varepsilon_h^z\|_{\mathcal{T}_h} + h^{-\frac{1}{2}} \|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial \mathcal{T}_h} \lesssim \|\varepsilon_h^{\boldsymbol{p}}\|_{\mathcal{T}_h} + h^{-\frac{1}{2}} \|P_M \varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial \mathcal{T}_h},$$

which we use below. Next, since $\varepsilon_h^{\widehat{z}} = 0$ on $\varepsilon_h^{\partial}$, the basic property of $\mathscr{B}$ in Lemma 1 gives

$$\mathscr{B}(\varepsilon_h^{\boldsymbol{p}}, \varepsilon_h^z, \varepsilon_h^{\widehat{z}}, \varepsilon_h^{\boldsymbol{p}}, \varepsilon_h^z, \varepsilon_h^{\widehat{z}}) = (\varepsilon_h^{\boldsymbol{p}}, \varepsilon_h^{\boldsymbol{p}})_{\mathcal{T}_h} + h^{-1} \|P_M \varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial \mathcal{T}_h}^2.$$

Then taking $(\boldsymbol{r}_2, w_2, \mu_2) = (\boldsymbol{\varepsilon}_h^{\boldsymbol{p}}, \varepsilon_h^z, \varepsilon_h^{\widehat{z}})$ in (2.37) in Lemma 7 gives

$$\begin{aligned}
&(\varepsilon_h^{\boldsymbol{p}}, \varepsilon_h^{\boldsymbol{p}})_{\mathcal{T}_h} + h^{-1} \|P_M \varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial \mathcal{T}_h}^2 \\
&\quad = -\langle \widehat{\boldsymbol{\delta}}_2, \varepsilon_h^z - \varepsilon_h^{\widehat{z}} \rangle_{\partial \mathcal{T}_h} + (y - y_h(u), \varepsilon_h^z)_{\mathcal{T}_h} \\
&\quad = -\langle \delta^{\boldsymbol{p}} \cdot \boldsymbol{n}, \varepsilon_h^z - \varepsilon_h^{\widehat{z}} \rangle_{\partial \mathcal{T}_h} - h^{-1} \langle \delta^z, P_M \varepsilon_h^z - \varepsilon_h^{\widehat{z}} \rangle_{\partial \mathcal{T}_h} + (y - y_h(u), \varepsilon_h^z)_{\mathcal{T}_h} \\
&\quad \le \|\delta^{\boldsymbol{p}}\|_{\partial \mathcal{T}_h} \|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial \mathcal{T}_h} + h^{-1} \|\delta^z\|_{\partial \mathcal{T}_h} \left\|P_M \varepsilon_h^z - \varepsilon_h^{\widehat{z}}\right\|_{\partial \mathcal{T}_h} \\
&\qquad + \|y - y_h(u)\|_{\mathcal{T}_h} \|\varepsilon_h^z\|_{\mathcal{T}_h}.
\end{aligned}$$

Hence

$$
\begin{aligned}
(\varepsilon_h^{\boldsymbol{p}}, \varepsilon_h^{\boldsymbol{p}})_{\mathcal{T}_h} &+ h^{-1}\|P_M\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}^2 \\
&\le h^{1/2}\left\|\delta^{\boldsymbol{p}}\right\|_{\partial\mathcal{T}_h} h^{-1/2}\|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h} + h^{-\frac{1}{2}}\|\delta^z\|_{\partial\mathcal{T}_h} h^{-\frac{1}{2}}\left\|P_M\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\right\|_{\partial\mathcal{T}_h} \\
&\quad + \|y - y_h(u)\|_{\mathcal{T}_h}\|\varepsilon_h^z\|_{\mathcal{T}_h} \\
&\le h^{1/2}\left\|\delta^{\boldsymbol{p}}\right\|_{\partial\mathcal{T}_h}(\|\varepsilon_h^{\boldsymbol{p}}\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|P_M\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}) \\
&\quad + h^{-\frac{1}{2}}\|\delta^z\|_{\partial\mathcal{T}_h} h^{-\frac{1}{2}}\left\|P_M\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\right\|_{\partial\mathcal{T}_h} \\
&\quad + C\|y - y_h(u)\|_{\mathcal{T}_h}(\|\nabla\varepsilon_h^z\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}) \\
&\le h^{1/2}\left\|\delta^{\boldsymbol{p}}\right\|_{\partial\mathcal{T}_h}(\|\varepsilon_h^{\boldsymbol{p}}\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|P_M\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}) \\
&\quad + h^{-\frac{1}{2}}\|\delta^z\|_{\partial\mathcal{T}_h} h^{-\frac{1}{2}}\left\|P_M\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\right\|_{\partial\mathcal{T}_h} \\
&\quad + C\|y - y_h(u)\|_{\mathcal{T}_h}(\|\varepsilon_h^{\boldsymbol{p}}\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|P_M\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}).
\end{aligned}
$$

Applying Young's inequality and Lemma 6 gives

$$
\begin{aligned}
(\varepsilon_h^{\boldsymbol{p}}, \varepsilon_h^{\boldsymbol{p}})_{\mathcal{T}_h} &+ h^{-1}\|P_M\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}^2 \\
&\lesssim h\left\|\delta^{\boldsymbol{p}}\right\|_{\partial\mathcal{T}_h}^2 + h^{-1}\|\delta^z\|_{\partial\mathcal{T}_h}^2 + \|y_h(u) - y\|_{\mathcal{T}_h}^2 \\
&\lesssim h^{2s_{\boldsymbol{p}}}\left\|\boldsymbol{p}\right\|_{s^{\boldsymbol{p}},\Omega}^2 + h^{2s_z-2}\left\|z\right\|_{s^z,\Omega}^2 + h^{2s_{\boldsymbol{q}}+2}\left\|\boldsymbol{q}\right\|_{s^{\boldsymbol{q}},\Omega}^2 + h^{2s_y}\left\|y\right\|_{s^y,\Omega}^2.
\end{aligned}
$$

This gives

$$
\begin{aligned}
\|\varepsilon_h^{\boldsymbol{p}}\|_{\mathcal{T}_h} &+ h^{-\frac{1}{2}}\|P_M\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h} \\
&\lesssim h^{s_{\boldsymbol{p}}}\left\|\boldsymbol{p}\right\|_{s^{\boldsymbol{p}},\Omega} + h^{s_z-1}\left\|z\right\|_{s^z,\Omega} + h^{s_{\boldsymbol{q}}+1}\left\|\boldsymbol{q}\right\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y}\left\|y\right\|_{s^y,\Omega}, \\
\|\varepsilon_h^z\|_{\mathcal{T}_h} &\lesssim \|\nabla\varepsilon_h^z\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h} \\
&\lesssim \|\varepsilon_h^{\boldsymbol{p}}\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|P_M\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h} \\
&\lesssim h^{s_{\boldsymbol{p}}}\left\|\boldsymbol{p}\right\|_{s^{\boldsymbol{p}},\Omega} + h^{s_z-1}\left\|z\right\|_{s^z,\Omega} + h^{s_{\boldsymbol{q}}+1}\left\|\boldsymbol{q}\right\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y}\left\|y\right\|_{s^y,\Omega}.
\end{aligned}
$$

As a consequence, a simple application of the triangle inequality gives optimal convergence rates for $\|\boldsymbol{p} - \boldsymbol{p}_h(u)\|_{\mathcal{T}_h}$ and $\|z - z_h(u)\|_{\mathcal{T}_h}$:

**Lemma 10**

$$\begin{aligned}
\|\boldsymbol{p} - \boldsymbol{p}_h(u)\|_{\mathcal{T}_h} &\le \|\delta^{\boldsymbol{p}}\|_{\mathcal{T}_h} + \|\varepsilon_h^{\boldsymbol{p}}\|_{\mathcal{T}_h} \\
&\lesssim h^{s_{\boldsymbol{p}}}\, \|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega} + h^{s_z-1}\, \|z\|_{s^z,\Omega} + h^{s_{\boldsymbol{q}}+1}\, \|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y}\, \|y\|_{s^y,\Omega}, \qquad (2.39\text{a}) \\
\|z - z_h(u)\|_{\mathcal{T}_h} &\le \|\delta^z\|_{\mathcal{T}_h} + \|\varepsilon_h^z\|_{\mathcal{T}_h} \\
&\lesssim h^{s_{\boldsymbol{p}}}\, \|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega} + h^{s_z-1}\, \|z\|_{s^z,\Omega} + h^{s_{\boldsymbol{q}}+1}\, \|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y}\, \|y\|_{s^y,\Omega}. \qquad (2.39\text{b})
\end{aligned}$$

Step 6: Estimate for $\|u - u_h\|_{\varepsilon_h^\partial}$ and $\|y - y_h\|_{\mathcal{T}_h}$

Next, we bound the error between the solutions of the auxiliary problem and the HDG problem (2.23). We use these error bounds and the error bounds in Lemma 6, Lemma 9, and Lemma 10 to obtain the main result.

For the remaining steps, we denote

$$\begin{aligned}
\zeta_{\boldsymbol{q}} &= \boldsymbol{q}_h(u) - \boldsymbol{q}_h, \quad \zeta_y = y_h(u) - y_h, \quad \zeta_{\widehat{y}} = \widehat{y}_h(u) - \widehat{y}_h, \\
\zeta_{\boldsymbol{p}} &= \boldsymbol{p}_h(u) - \boldsymbol{p}_h, \quad \zeta_z = z_h(u) - z_h, \quad \zeta_{\widehat{z}} = \widehat{z}_h(u) - \widehat{z}_h,
\end{aligned}$$

where $\widehat{y}_h = \widehat{y}_h^o$ on $\varepsilon_h^o$, $\widehat{y}_h = u_h$ on $\varepsilon_h^\partial$, $\widehat{z}_h = \widehat{z}_h^o$ on $\varepsilon_h^o$, and $\widehat{z}_h = 0$ on $\varepsilon_h^\partial$. This gives $\zeta_{\widehat{z}} = 0$ on $\varepsilon_h^\partial$.

Subtracting the auxiliary problem and the HDG problem gives the following error equations

$$\mathscr{B}(\zeta_{\boldsymbol{q}}, \zeta_y, \zeta_{\widehat{y}}; \boldsymbol{r}_1, w_1, \mu_1) = -\langle P_M u - u_h, \boldsymbol{r_1} \cdot \boldsymbol{n} - h^{-1} w_1 \rangle_{\varepsilon_h^\partial}, \qquad (2.40\text{a})$$

$$\mathscr{B}(\zeta_{\boldsymbol{p}}, \zeta_z, \zeta_{\widehat{z}}; \boldsymbol{r}_2, w_2, \mu_2) = (\zeta_y, w_2)_{\mathcal{T}_h}, \qquad (2.40\text{b})$$

for all $(\boldsymbol{r}_1, \boldsymbol{r}_2, w_1, w_2, \mu_1, \mu_2) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times W_h \times M_h(o) \times M_h(o)$.

**Lemma 11** *We have*

$$\begin{aligned}\|u-u_h\|^2_{\varepsilon_h^\partial}+\gamma^{-1}\left\|\zeta_y\right\|^2_{\mathcal{T}_h}\\ &=\langle u+\gamma^{-1}\boldsymbol{p}_h(u)\cdot\boldsymbol{n}+\gamma^{-1}h^{-1}P_M z_h(u),u-u_h\rangle_{\varepsilon_h^\partial}\\ &\quad-\langle u_h+\gamma^{-1}\boldsymbol{p}_h\cdot\boldsymbol{n}+\gamma^{-1}h^{-1}P_M z_h,u-u_h\rangle_{\varepsilon_h^\partial}.\end{aligned}$$

**Proof:** First, we have

$$\begin{aligned}&\langle u+\gamma^{-1}\boldsymbol{p}_h(u)\cdot\boldsymbol{n}+\gamma^{-1}h^{-1}P_M z_h(u),u-u_h\rangle_{\varepsilon_h^\partial}\\ &\quad-\langle u_h+\gamma^{-1}\boldsymbol{p}_h\cdot\boldsymbol{n}+\gamma^{-1}h^{-1}P_M z_h,u-u_h\rangle_{\varepsilon_h^\partial}\\ &=\|u-u_h\|^2_{\varepsilon_h^\partial}+\gamma^{-1}\langle\zeta_{\boldsymbol{p}}\cdot\boldsymbol{n}+h^{-1}P_M\zeta_z,u-u_h\rangle_{\varepsilon_h^\partial}.\end{aligned}$$

As in the proof of Lemma 1, it can be shown that

$$\mathscr{B}(\zeta_{\boldsymbol{q}},\zeta_y,\zeta_{\widehat{y}};\zeta_{\boldsymbol{p}},-\zeta_z,-\zeta_{\widehat{z}})+\mathscr{B}(\zeta_{\boldsymbol{p}},\zeta_z,\zeta_{\widehat{z}};-\zeta_{\boldsymbol{q}},\zeta_y,\zeta_{\widehat{y}})=0.$$

One the other hand, we have

$$\begin{aligned}&\mathscr{B}(\zeta_{\boldsymbol{q}},\zeta_y,\zeta_{\widehat{y}};\zeta_{\boldsymbol{p}},-\zeta_z,-\zeta_{\widehat{z}})+\mathscr{B}(\zeta_{\boldsymbol{p}},\zeta_z,\zeta_{\widehat{z}};-\zeta_{\boldsymbol{q}},\zeta_y,\zeta_{\widehat{y}})\\ &\qquad=(\zeta_y,\zeta_y)_{\mathcal{T}_h}-\langle P_M u-u_h,\zeta_{\boldsymbol{p}}\cdot\boldsymbol{n}+h^{-1}\zeta_z\rangle_{\varepsilon_h^\partial}\\ &\qquad=(\zeta_y,\zeta_y)_{\mathcal{T}_h}-\langle u-u_h,\zeta_{\boldsymbol{p}}\cdot\boldsymbol{n}+h^{-1}P_M\zeta_z\rangle_{\varepsilon_h^\partial}.\end{aligned}$$

Comparing the above two equalities gives

$$(\zeta_y,\zeta_y)_{\mathcal{T}_h}=\langle u-u_h,\zeta_{\boldsymbol{p}}\cdot\boldsymbol{n}+h^{-1}P_M\zeta_z\rangle_{\varepsilon_h^\partial}.$$

**Theorem 3** *We have*

$$\|u - u_h\|_{\varepsilon_h^\partial} \lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega},$$
$$\|y - y_h\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}.$$

**Proof:** Since $u + \gamma^{-1}\boldsymbol{p}\cdot\boldsymbol{n} = 0$ on $\varepsilon_h^\partial$ and $u_h + \gamma^{-1}\boldsymbol{p}_h\cdot\boldsymbol{n} + \gamma^{-1}h^{-1}P_M z_h = 0$ on $\varepsilon_h^\partial$ we have

$$\begin{aligned}
\|u - u_h\|_{\varepsilon_h^\partial}^2 + \gamma^{-1}\left\|\zeta_y\right\|_{\mathcal{T}_h}^2 &= \langle u + \gamma^{-1}\boldsymbol{p}_h(u)\cdot\boldsymbol{n} + \gamma^{-1}h^{-1}P_M z_h(u), u - u_h\rangle_{\varepsilon_h^\partial}\\
&= \langle \gamma^{-1}(\boldsymbol{p}_h(u) - \boldsymbol{p})\cdot\boldsymbol{n} + \gamma^{-1}h^{-1}P_M z_h(u), u - u_h\rangle_{\varepsilon_h^\partial}\\
&\lesssim (\|\boldsymbol{p}_h(u) - \boldsymbol{p}\|_{\partial\mathcal{T}_h} + h^{-1}\|P_M z_h(u)\|_{\varepsilon_h^\partial})\,\|u - u_h\|_{\varepsilon_h^\partial}.
\end{aligned}$$

Next, since $\widehat{z}_h(u) = z = 0$ on $\varepsilon_h^\partial$ we have

$$\begin{aligned}
\|\boldsymbol{p}_h(u) - \boldsymbol{p}\|_{\partial\mathcal{T}_h} &\le \|\boldsymbol{p}_h(u) - \boldsymbol{\Pi}\boldsymbol{p}\|_{\partial\mathcal{T}_h} + \|\boldsymbol{\Pi}\boldsymbol{p} - \boldsymbol{p}\|_{\partial\mathcal{T}_h}\\
&\lesssim h^{-\frac{1}{2}}\|\boldsymbol{p}_h(u) - \boldsymbol{\Pi}\boldsymbol{p}\|_{\mathcal{T}_h} + h^{s_{\boldsymbol{p}}-\frac{1}{2}}\|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega}\\
&\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}}\|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}}\|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}}\|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega}\\
&\quad + h^{s_y-\frac{1}{2}}\|y\|_{s_y,\Omega},\\
\|P_M z_h(u)\|_{\varepsilon_h^\partial} &= \|P_M z_h(u) - P_M\Pi z + P_M\Pi z - P_M z + P_M z - \widehat{z}_h(u)\|_{\varepsilon_h^\partial}\\
&\le (\|P_M\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\varepsilon_h^\partial} + \|\Pi z - z\|_{\varepsilon_h^\partial})\\
&\le (\|P_M\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h} + \|\Pi z - z\|_{\partial\mathcal{T}_h}).
\end{aligned}$$

Lemma 9 and properties of the $L^2$ projection gives

$$\|u - u_h\|_{\varepsilon_h^\partial} \lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}.$$

Moreover, we have

$$\left\|\zeta_y\right\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}.$$

Then, by the triangle inequality and Lemma 6 we obtain

$$\|y - y_h\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}.$$

Step 7: Estimates for $\|\boldsymbol{q} - \boldsymbol{q}_h\|_{\mathcal{T}_h}$, $\|\boldsymbol{p} - \boldsymbol{p}_h\|_{\mathcal{T}_h}$ and $\|z - z_h\|_{\mathcal{T}_h}$

**Lemma 12** *We have*

$$\begin{aligned}
\left\|\zeta_{\boldsymbol{q}}\right\|_{\mathcal{T}_h} &\lesssim h^{s_{\boldsymbol{p}}-1} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-2} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-1} \|y\|_{s_y,\Omega}, \\
\left\|\zeta_{\boldsymbol{p}}\right\|_{\mathcal{T}_h} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}, \\
\|\zeta_z\|_{\mathcal{T}_h} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}.
\end{aligned}$$

**Proof:** By Lemma 1 and the error equation (2.40a), we have

$$\begin{aligned}
&\mathscr{B}(\zeta_{\boldsymbol{q}}, \zeta_y, \zeta_{\widehat{y}}; \zeta_{\boldsymbol{q}}, \zeta_y, \zeta_{\widehat{y}}) \\
&\quad = (\zeta_{\boldsymbol{q}}, \zeta_{\boldsymbol{q}})_{\mathcal{T}_h} + \langle h^{-1}(P_M\zeta_y - \zeta_{\widehat{y}}), P_M\zeta_y - \zeta_{\widehat{y}}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} + \langle h^{-1}P_M\zeta_y, P_M\zeta_y\rangle_{\varepsilon_h^\partial} \\
&\quad = -\langle P_M u - u_h, \zeta_{\boldsymbol{q}}\cdot\boldsymbol{n} - h^{-1}\zeta_y\rangle_{\varepsilon_h^\partial} = -\langle u - u_h, \zeta_{\boldsymbol{q}}\cdot\boldsymbol{n} - h^{-1}P_M\zeta_y\rangle_{\varepsilon_h^\partial} \\
&\quad \lesssim \|u - u_h\|_{\varepsilon_h^\partial} (\left\|\boldsymbol{\zeta}_{\boldsymbol{q}}\right\|_{\varepsilon_h^\partial} + h^{-1}\left\|P_M\zeta_y\right\|_{\varepsilon_h^\partial}) \\
&\quad \lesssim h^{-\frac{1}{2}} \|u - u_h\|_{\varepsilon_h^\partial} (\left\|\boldsymbol{\zeta}_{\boldsymbol{q}}\right\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\left\|P_M\zeta_y\right\|_{\varepsilon_h^\partial}),
\end{aligned}$$

which gives

$$\begin{aligned}
\left\|\zeta_{\boldsymbol{q}}\right\|_{\mathcal{T}_h} &\lesssim h^{-\frac{1}{2}} \|u - u_h\|_{\varepsilon_h^\partial} \\
&\lesssim h^{s_{\boldsymbol{p}}-1} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-2} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-1} \|y\|_{s_y,\Omega}.
\end{aligned}$$

Next, we estimate $\zeta_{\boldsymbol{p}}$. By Lemma 1, the error equation (2.40b), and since $\zeta_{\widehat{z}} = 0$ on $\varepsilon_h^\partial$, we have

$$
\begin{aligned}
&\mathscr{B}(\zeta_{\boldsymbol{p}}, \zeta_z, \zeta_{\widehat{z}}; \zeta_{\boldsymbol{p}}, \zeta_z, \zeta_{\widehat{z}}) \\
&\quad = (\zeta_{\boldsymbol{p}}, \zeta_{\boldsymbol{p}})_{\mathcal{T}_h} + \langle h^{-1}(P_M \zeta_z - \zeta_{\widehat{z}}), P_M \zeta_z - \zeta_{\widehat{z}} \rangle_{\partial\mathcal{T}_h \backslash \varepsilon_h^\partial} + \langle h^{-1} P_M \zeta_z, P_M \zeta_z \rangle_{\varepsilon_h^\partial} \\
&\quad = (\zeta_{\boldsymbol{p}}, \zeta_{\boldsymbol{p}})_{\mathcal{T}_h} + \langle h^{-1}(P_M \zeta_z - \zeta_{\widehat{z}}), P_M \zeta_z - \zeta_{\widehat{z}} \rangle_{\partial\mathcal{T}_h} \\
&\quad = (\zeta_y, \zeta_z)_{\mathcal{T}_h} \\
&\quad \le \left\|\zeta_y\right\|_{\mathcal{T}_h} \|\zeta_z\|_{\mathcal{T}_h} \\
&\quad \lesssim \left\|\zeta_y\right\|_{\mathcal{T}_h} (\|\nabla \zeta_z\|_{\mathcal{T}_h} + h^{-\frac{1}{2}} \|\zeta_z - \zeta_{\widehat{z}}\|_{\partial\mathcal{T}_h}) \\
&\quad \lesssim \left\|\zeta_y\right\|_{\mathcal{T}_h} (\|\zeta_{\boldsymbol{p}}\|_{\mathcal{T}_h} + h^{-\frac{1}{2}} \|P_M \zeta_z - \zeta_{\widehat{z}}\|_{\partial\mathcal{T}_h}),
\end{aligned}
$$

where we used the discrete Poincaré inequality in Lemma 8 and also Lemma 3. This implies

$$
\begin{aligned}
&\left\|\zeta_{\boldsymbol{p}}\right\|_{\mathcal{T}_h} + h^{-\frac{1}{2}} \|P_M \zeta_z - \zeta_{\widehat{z}}\|_{\partial\mathcal{T}_h} \\
&\qquad \lesssim h^{s_{\boldsymbol{p}} - \frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z - \frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}} + \frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y - \frac{1}{2}} \|y\|_{s_y,\Omega}.
\end{aligned}
$$

The discrete Poincaré inequality in Lemma 8 also gives

$$
\begin{aligned}
\|\zeta_z\|_{\mathcal{T}_h} &\lesssim \|\nabla \zeta_z\|_{\mathcal{T}_h} + h^{-\frac{1}{2}} \|\zeta_z - \zeta_{\widehat{z}}\|_{\partial\mathcal{T}_h} \\
&\lesssim \|\zeta_{\boldsymbol{p}}\|_{\mathcal{T}_h} + h^{-\frac{1}{2}} \|P_M \zeta_z - \zeta_{\widehat{z}}\|_{\partial\mathcal{T}_h} \\
&\lesssim h^{s_{\boldsymbol{p}} - \frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z - \frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}} + \frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y - \frac{1}{2}} \|y\|_{s_y,\Omega}.
\end{aligned}
$$

The above lemma along with the triangle inequality, Lemma 6, and Lemma 10 complete the proof of the main result:

**Theorem 4** *We have*

$$\|\boldsymbol{q}-\boldsymbol{q}_h\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}-1}\|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-2}\|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}}\|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-1}\|y\|_{s_y,\Omega},$$

$$\|\boldsymbol{p}-\boldsymbol{p}_h\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}}\|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}}\|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}}\|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}}\|y\|_{s_y,\Omega},$$

$$\|z-z_h\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}}\|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}}\|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}}\|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}}\|y\|_{s_y,\Omega}.$$

## 2.5. NUMERICAL EXPERIMENTS

For our numerical experiments, we test problems similar to the examples considered in [33]; see also [11, 49, 55]. We chose $k = 1$ for all computations; i.e., quadratic polynomials are used for the scalar variables, and linear polynomials are used for the flux variables and the boundary trace variables.

We begin with a 2D example on a square domain $\Omega = [0, 1/4] \times [0, 1/4] \subset \mathbb{R}^2$. The largest interior angle is $\omega = \pi/2$, and so $r_\Omega = 3$ and $p_\Omega = \infty$. The data is chosen as

$$f = 0, \quad y_d = (x^2 + y^2)^s \quad \text{and} \quad \gamma = 1,$$

where $s = 10^{-5}$. Then $y_d \in H^1(\Omega) \cap L^\infty(\Omega)$, and Corollary 2 in Section 2.4 gives the convergence rates

$$\|y - y_h\|_{0,\Omega} = O(h^{3/2-\varepsilon}), \qquad \|z - z_h\|_{0,\Omega} = O(h^{3/2-\varepsilon}),$$

$$\|\boldsymbol{q} - \boldsymbol{q}_h\|_{0,\Omega} = O(h^{1-\varepsilon}), \qquad \|\boldsymbol{p} - \boldsymbol{p}_h\|_{0,\Omega} = O(h^{3/2-\varepsilon}),$$

and

$$\|u - u_h\|_{0,\Gamma} = O(h^{3/2-\varepsilon}).$$

Since we do not have an explicit expression for the exact solution, we solved the problem numerically for a triangulation with 262144 elements, i.e., $h = 2^{-12}\sqrt{2}$ and compared this reference solution against other solutions computed on meshes with larger $h$. The numerical results are shown in Table 2.1. The convergence rates observed for $\|\boldsymbol{q} - \boldsymbol{q}_h\|_{0,\Omega}$ and $\|u - u_h\|_{0,\Gamma}$ are in agreement with our theoretical results, while the convergence rates for $\|\boldsymbol{p} - \boldsymbol{p}_h\|_{0,\Omega}$, $\|y - y_h\|_{0,\Omega}$, and $\|z - z_h\|_{0,\Omega}$ are higher than our theoretical results. A similar phenomena can be observed in [33, 49, 55].

Table 2.1. Error of control $u$, state $y$, adjoint state $z$, and their fluxes $\boldsymbol{q}$ and $\boldsymbol{p}$

| $h/\sqrt{2}$ | $2^{-4}$ | $1/2^{-5}$ | $2^{-6}$ | $2^{-7}$ | $2^{-8}$ |
|---|---|---|---|---|---|
| $\|\boldsymbol{q} - \boldsymbol{q}_h\|_{0,\Omega}$ | 4.1343e-02 | 2.1025e-02 | 1.0677e-02 | 5.3865e-03 | 2.6959e-03 |
| order | - | 0.9756 | 0.9776 | 0.9871 | 0.9986 |
| $\|\boldsymbol{p} - \boldsymbol{p}_h\|_{0,\Omega}$ | 1.3463e-03 | 3.8638e-04 | 1.0849e-04 | 2.9862e-05 | 8.0969e-06 |
| order | - | 1.8009 | 1.8325 | 1.8612 | 1.8828 |
| $\|y - y_h\|_{0,\Omega}$ | 5.4609e-04 | 1.3647e-04 | 3.4763e-05 | 8.8037e-06 | 2.2236e-06 |
| order | - | 2.0005 | 1.9730 | 1.9814 | 1.9852 |
| $\|z - z_h\|_{0,\Omega}$ | 1.9671e-05 | 2.6887e-06 | 3.7026e-07 | 5.0372e-08 | 6.7767e-09 |
| order | - | 2.8711 | 2.8603 | 2.8778 | 2.8940 |
| $\|u - u_h\|_{0,\Gamma}$ | 7.3053e-03 | 2.6902e-03 | 9.7764e-04 | 3.5178e-04 | 1.2569e-04 |
| order | - | 1.4412 | 1.4603 | 1.4746 | 1.4849 |

For illustration, we plots the states of $y$ and the boundary control $u$. The low regularity of the primary flux $\boldsymbol{q}$ is apparent for this example due to the corner singularities. For illustration, we plot the state $y$, adjoint state $z$, and their fluxes $\boldsymbol{q}$ and $\boldsymbol{p}$ in Figure 2.1 and the control was plotted in Figure 2.2. The 2D regularity result in Section 2.2 indicate that the primary flux $\boldsymbol{q}$ can have low regularity. In this example, it does indeed appear that $\boldsymbol{q}$ has singularities at the corners of the domain. These figures can be compared to similar plots in [11, 55].

Next, we consider a 3D extension of the 2D example above. The domain is a cube $\Omega = [0, 1/32] \times [0, 1/32] \times [0, 1/32]$, and the data is chosen as

$$f = 0, \quad y_d = (x^2 + y^2 + z^2)^s \quad \text{and} \quad \gamma = 1,$$

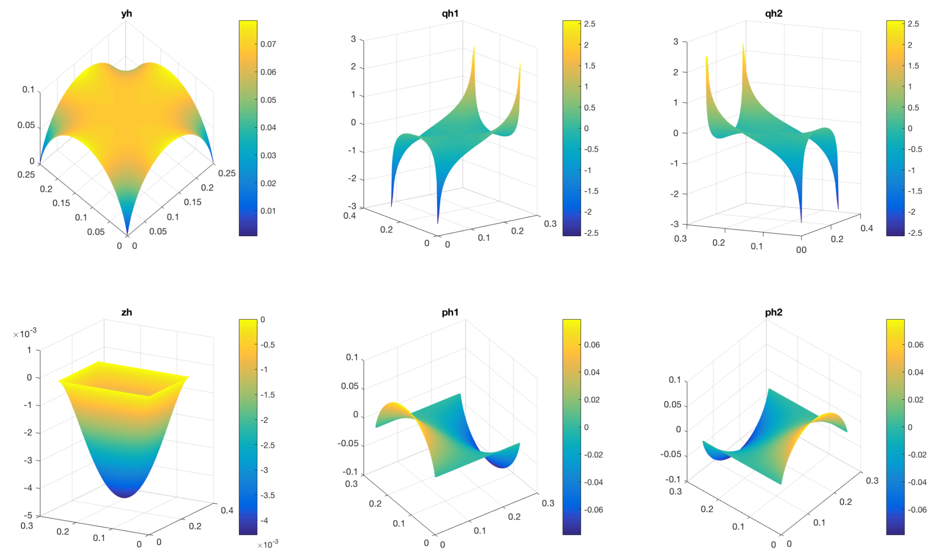


Figure 2.1. The primary state $y_h$, the primary flux $\boldsymbol{q}_h$, the dual state $z_h$, and the dual flux $\boldsymbol{p}_h$ for the 2D example

where $s = -1/4 + 10^{-5}$, so that $y_d \in H^1(\Omega)$. In this case, we did not attempt to determine the regularity of the control and other variables; we simply present the numerical results here.

As in the 2D example above, we do not have an explicit expression for the exact solution. Therefore, we solved the problem numerically for a triangulation with 196608 tetrahedrons, i.e., $h = 2^{-12}\sqrt{3}$ and compared this reference solution against other solutions computed on meshes with larger $h$. The numerical results are shown in Table 2.2. The observed convergence rates for all variables are similar to the results for the 2D example above.

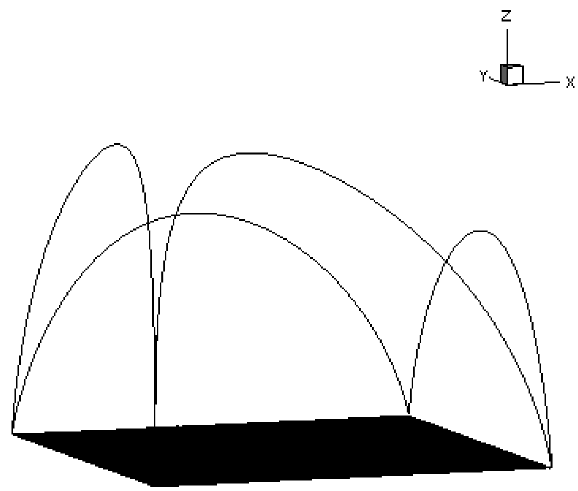


Figure 2.2. The optimal control $u_h$ for the 2D example

Table 2.2. Error of control $u$, state $y$, adjoint state $z$, and their fluxes $\boldsymbol{q}$ and $\boldsymbol{p}$

| $h/\sqrt{3}$ | $2^{-6}$ | $2^{-7}$ | $2^{-8}$ | $2^{-9}$ |
|---|---|---|---|---|
| $\|\boldsymbol{q}-\boldsymbol{q}_h\|_{0,\Omega}$ | 9.2640e-03 | 5.2580e-03 | 2.7462e-03 | 1.2475e-03 |
| order | - | 0.81712 | 0.93706 | 1.1384 |
| $\|\boldsymbol{p}-\boldsymbol{p}_h\|_{0,\Omega}$ | 3.5425e-05 | 1.2283e-05 | 3.8463e-06 | 1.1022e-06 |
| order | - | 1.5281 | 1.6751 | 1.8032 |
| $\|y-y_h\|_{0,\Omega}$ | 1.6040e-05 | 4.5070e-06 | 1.2191e-06 | 2.9781e-07 |
| order | - | 1.8314 | 1.8864 | 2.0333 |
| $\|z-z_h\|_{0,\Omega}$ | 7.8545e-08 | 1.3058e-08 | 2.0042e-09 | 2.8775e-10 |
| order | - | 2.5886 | 2.7039 | 2.8001 |
| $\|u-u_h\|_{0,\Gamma}$ | 4.5932e-04 | 1.8934e-04 | 7.1955e-05 | 2.4123e-05 |
| order | - | 1.2785 | 1.3958 | 1.5767 |

# 3. CONVECTION DIFFUSION WITH HIGH REGULARITY

## 3.1. MODEL PROBLEM

In this section, we consider the following Dirichlet boundary control problem. Let $\Omega \subset \mathbb{R}^d$ $(d \geq 2)$ be a Lipschitz polyhedral domain with boundary $\Gamma = \partial\Omega$. The goal is to find the optimal control $u \in L^2(\Gamma)$ that minimizes the cost function

$$J(u) = \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \frac{\gamma}{2}\|u\|^2_{L^2(\Gamma)}, \quad \gamma > 0, \tag{3.1}$$

subject to the elliptic convection diffusion equation

$$\begin{aligned} -\Delta y + \boldsymbol{\beta} \cdot \nabla y &= f \quad \text{in } \Omega, \\ y &= u \quad \text{on } \partial\Omega, \end{aligned} \tag{3.2}$$

where $f \in L^2(\Omega)$ and the vector field $\boldsymbol{\beta}$ satisfies

$$\nabla \cdot \boldsymbol{\beta} \leq 0. \tag{3.3}$$

We make other smoothness assumptions on $\boldsymbol{\beta}$ for our analysis.

Formally, the optimal control $u \in L^2(\Gamma)$ and the optimal state $y \in L^2(\Omega)$ minimizing the cost functional satisfy the optimality system

$$-\Delta y + \boldsymbol{\beta} \cdot \nabla y = f \qquad \text{in } \Omega, \tag{3.4a}$$

$$y = u \qquad \text{on } \partial\Omega, \tag{3.4b}$$

$$-\Delta z - \nabla \cdot (\boldsymbol{\beta} z) = y - y_d \qquad \text{in } \Omega, \tag{3.4c}$$

$$z = 0 \qquad \text{on } \partial\Omega, \tag{3.4d}$$

$$\nabla z \cdot \boldsymbol{n} - \gamma u = 0 \qquad \text{on } \partial\Omega. \tag{3.4e}$$

## 3.2. ANALYSIS OF THE DIRICHLET CONTROL PROBLEM

To begin, we set notation and prove some fundamental results concerning the optimality system for the control problem in the 2D case.

Duality between $H^1(\Omega)^*$ and $H^1(\Omega)$ will be denoted $[q, r]_\Omega$, while duality between $H^{-\varepsilon}(\Gamma)$ and $H^{\varepsilon}(\Gamma)$ for $0 \leq \varepsilon \leq 1/2$ will be denoted $[u, v]_\Gamma$.

Throughout this section, we consider $\Omega$ a polygonal domain, not necessarily convex, and denote $\omega$ its biggest interior angle. Notice that $1/2 < \pi/\omega < 1$ for nonconvex domains and $1 < \pi/\omega \leq 3$ for convex domains. Furthermore, we assume in this section $\boldsymbol{\beta}$ satisfies the following conditions:

$$\begin{aligned} \boldsymbol{\beta} \in [L^\infty(\Omega)]^d, \quad \nabla \cdot \boldsymbol{\beta} \in L^\infty(\Omega), \quad \nabla \cdot \boldsymbol{\beta} \leq 0 \quad & \text{for any } \Omega, \text{ and also} \\ \nabla\nabla \cdot \boldsymbol{\beta} \in [L^2(\Omega)]^d \quad & \text{if } \Omega \text{ is convex.} \end{aligned} \tag{3.5}$$

Moreover, we assume the forcing $f$ is identically zero in this section. If this is not the case, then a simple change of variable as in [1, pg. 3623] can be used to eliminate the forcing.

**3.2.1. Study of the State Equation.** Notice that for data $u \in L^2(\Gamma)$, we cannot expect to have a variational solution of the state equation (3.2). Therefore, we need a suitable concept of solution that makes the control-to-state operator continuous and that coincides with the variational solution for regular data. Moreover, since HDG is based on a mixed formulation, it is also important to see how this concept of very weak solution extends to mixed formulations.

To define the concept of a very weak solution, we first introduce the adjoint problem and recall its regularity properties.

**Lemma 13** *For every $g \in L^2(\Omega)$ there exists a unique $z_g \in H_0^1(\Omega) \cap H^t(\Omega)$ for all $t \leq 2$ with $t < 1 + \pi/\omega$ such that*

$$-\Delta z_g - \nabla \cdot (\boldsymbol{\beta} z_g) = g \text{ in } \Omega, \;\; z_g = 0 \text{ on } \Gamma. \tag{3.6}$$

*Moreover, $\partial_{\boldsymbol{n}} z_g \in H^s(\Gamma)$ for all $s \leq 1/2$ such that $s < \pi/\omega - 1/2$.*

*If, further, $g \in H^{t^*}(\Omega)$ for some $0 \leq t^* < 1$, then $z_g \in H_0^1(\Omega) \cap H^t(\Omega)$ for all $t \leq 2 + t^*$ with $t < \min\{3, 1 + \pi/\omega\}$ and $\partial_{\boldsymbol{n}} z_g \in H^s(\Gamma)$ for all $s \leq 1/2 + t^*$ such that $s < \min\{3/2, \pi/\omega - 1/2\}$.*

**Proof:** Existence and uniqueness of the solution is standard. The regularity of $\boldsymbol{\beta}$ implies that $\nabla \cdot (\boldsymbol{\beta} z_g) \in L^2(\Omega)$, and hence $z_g \in H^t(\Omega)$ for all $t \leq 2$ such that $t < 1 + \pi/\omega$. For the regularity of the normal derivative in the case $s < 1/2$, apply [1, Corollary 2.3] and trace theory in [34]. For $s = 1/2$, apply [9, Lemma (A2)].

For the extra regularity result, we use that $z_g \in H^t(\Omega)$ for all $t \leq 2$ such that $t < 1 + \pi/\omega$, $\nabla \cdot \boldsymbol{\beta} \in L^\infty(\Omega)$, and $\nabla\nabla \cdot \boldsymbol{\beta} \in [L^2(\Omega)]^d$ to obtain that $\nabla \cdot (\boldsymbol{\beta} z_g) \in H^{t^*}(\Omega)$. Now we have that $-\Delta z_g \in H^{t^*}(\Omega)$ and standard regularity results in [34] lead to $z_g \in H_0^1(\Omega) \cap H^t(\Omega)$ for all $t \leq 2 + t^*$ with $t < \min\{3, 1 + \pi/\omega\}$. The normal trace then satisfies that $\partial_n z_g \in \Pi_{i=1}^m H^s(\Gamma_i)$ for all $s \leq 1/2 + t^*$ such that $s < \min\{3/2, \pi/\omega - 1/2\}$, where $\Gamma_i$ denotes side $i$ of the boundary of $\Omega$. If $\pi/\omega < 1$, then $s < 1/2$ and $\Pi_{i=1}^m H^s(\Gamma_i) = H^s(\Gamma)$. If $\pi/\omega > 1$, we use that $z_g = 0$ on $\Gamma$ as in [8, Section 4] to prove that $\partial_n z_g = 0$ on the corners of the domain. This implies that $\partial_n z_g$ is continuous and hence it belongs to $H^s(\Gamma)$.

**Definition 1** *Let* $\varepsilon$ *be a real number such that* $0 \leq \varepsilon \leq 1/2$ *and* $\varepsilon < \pi/\omega - 1/2$. *For* $u \in H^{-\varepsilon}(\Gamma)$, *we say that* $y \in L^2(\Omega)$ *is a very weak solution of*

$$-\Delta y + \boldsymbol{\beta} \cdot \nabla y = 0 \text{ in } \Omega, \ \ y = u \text{ on } \Gamma \tag{3.7}$$

*if and only if*

$$(y, g)_\Omega + [u, \partial_n z_g]_\Gamma = 0, \tag{3.8}$$

*for all* $g \in L^2(\Omega)$, *where* $z_g$ *is the unique solution of* (3.6).

**Remark 2** *The definition is meaningful thanks to the regularity of the normal derivative of* $z_g$ *provided in 14. Eventually the case* $\varepsilon = 1/2$ *must be discarded, but we keep it while we can cope with it. For our problem we need only to consider the case* $\varepsilon = 0$. *We include the other cases for the sake of completeness and because the definition may be useful for problems with control or state constraints; see e.g., [47, Section 6.2]*

**Lemma 14** *Let* $s$ *be a real number such that* $-1/2 \leq s < 3/2$ *and* $s > 1/2 - \pi/\omega$. *For every* $u \in H^s(\Gamma)$, *there exists a unique* very weak *solution* $y \in H^{1/2+s}(\Omega)$ *of* (3.7) *and*

$$\|y\|_{H^{1/2+s}(\Omega)} \leq C\|u\|_{H^s(\Gamma)}.$$

**Proof:** The case $s = -1/2$ can only happen in a convex domain and we can use the classic transposition method. The proof for $-1/2 < s < 0$ is as the the proof of Lemma 2.5 in [1].

For $1/2 \leq s < 3/2$ we have that (3.7) has a unique variational solution $y$ and that it belongs to $H^{s+1/2}(\Omega)$; see [1, Proof of Corollary 4.2]. Integration by parts shows that $y$ is also a very weak solution.

For $0 \leq s < 1/2$ the result follows from interpolation.

Next we do the same for the mixed formulation. From now on, we assume the polygonal domain $\Omega$ is convex so that $1 < \pi/\omega \leq 3$. First we state an existence and regularity result for the mixed formulation of the convection diffusion equation with regular data.

**Lemma 15** *For every $g \in L^2(\Omega)$, there exists a unique pair $(z_g, \boldsymbol{p}_g) \in H_0^1(\Omega) \times H(\text{div}, \Omega)$ such that*

$$(\boldsymbol{p}_g, \boldsymbol{r})_\Omega - (z_g, \nabla \cdot \boldsymbol{r})_\Omega = 0, \tag{3.9a}$$

$$(\nabla \cdot (\boldsymbol{p}_g - \boldsymbol{\beta} z_g), w)_\Omega = (g, w)_\Omega, \tag{3.9b}$$

*for all $(\boldsymbol{r}, w) \in H(\text{div}, \Omega) \times L^2(\Omega)$. Moreover, $(z_g, \boldsymbol{p}_g) \in (H^2(\Omega) \cap H_0^1(\Omega)) \times [H^1(\Omega)]^d$, $\partial_{\boldsymbol{n}} z_g = \boldsymbol{p}_g \cdot \boldsymbol{n} \in H^{1/2}(\Gamma)$ and*

$$-\Delta z_g - \nabla \cdot (\boldsymbol{\beta} z_g) = g \text{ in } \Omega, z_g = 0 \text{ on } \Gamma \text{ and } \boldsymbol{p}_g = -\nabla z_g. \tag{3.10}$$

*If, further, $g \in H^{t^*}(\Omega)$ for some $0 \leq t^* < 1$, then $(z_g, \boldsymbol{p}_g) \in (H^t(\Omega) \cap H_0^1(\Omega)) \times [H^{t-1}(\Omega)]^d$ for all $t \leq 2 + t^*$ with $t < \min\{3, 1 + \pi/\omega\}$ and $\partial_{\boldsymbol{n}} z_g = \boldsymbol{p}_g \cdot \boldsymbol{n} \in H^s(\Gamma)$ for all $s \leq 1/2 + t^*$ such that $s < \min\{3/2, \pi/\omega - 1/2\}$.*

Notice that the notation $z_g$ is not contradictory. If $z_g$ is the solution of (3.6), then $(z_g, -\nabla z_g)$ is the solution of (3.9a)–(3.9b). Also, if $(z_g, \boldsymbol{p}_g)$ is the solution of (3.9a)–(3.9b), then $z_g$ is the solution of (3.6). Therefore this lemma is a straightforward consequence of Lemma 13

Now we must drop the case $\varepsilon = 1/2$.

**Definition 2** *Let $\varepsilon$ be a real number such that*

$$0 \leq \varepsilon < 1/2.$$

*For $u \in H^{-\varepsilon}(\Gamma)$, we say that $(y, \boldsymbol{q}) \in L^2(\Omega) \times [H^1(\Omega)^*]^d$ is a very weak solution of*

$$-\Delta y + \boldsymbol{\beta} \cdot \nabla y = 0 \text{ in } \Omega, \;\; y = u \text{ in } \Gamma, \;\; \boldsymbol{q} = -\nabla y \tag{3.11}$$

*if and only if*

$$[\boldsymbol{q}, \boldsymbol{r}]_\Omega - (y, \nabla \cdot \boldsymbol{r})_\Omega + [u, \boldsymbol{r} \cdot \boldsymbol{n}]_\Gamma = 0, \tag{3.12a}$$

$$[\boldsymbol{q} + \boldsymbol{\beta} y, \boldsymbol{p}_g]_\Omega - (y \nabla \cdot \boldsymbol{\beta}, z_g)_\Omega = 0, \tag{3.12b}$$

*for all $(\boldsymbol{r}, g) \in [H^1(\Omega)]^d \times L^2(\Omega)$, and $(z_g, \boldsymbol{p}_g) \in (H^2(\Omega) \cap H_0^1(\Omega)) \times [H^1(\Omega)]^d$ is the unique solution of* (3.9a)–(3.9b).

**Remark 3** *For $\varepsilon = 1/2$, the expression $[u, \boldsymbol{r} \cdot \boldsymbol{n}]_\Gamma$ is meaningless even for $\boldsymbol{r} \in [C^\infty(\Omega)]^d$, since $\boldsymbol{n}$ has jump derivatives, and hence $\boldsymbol{r} \cdot \boldsymbol{n} \notin H^{1/2}(\Gamma)$.*

**Theorem 5** *For every $u \in H^{-\varepsilon}(\Gamma)$, there exists a unique* very weak *solution $(y, \boldsymbol{q}) \in L^2(\Omega) \times [H^1(\Omega)^*]^d$ of* (3.11). *Moreover, $(y, \boldsymbol{q}) \in H^{1/2-\varepsilon}(\Omega) \times [H^{1/2+\varepsilon}(\Omega)^*]^d$ and $y$ is the very weak solution of* (3.7).

**Proof:** Let us first prove uniqueness of solution in the space $L^2(\Omega) \times [H^1(\Omega)^*]^d$. Take $u = 0$ and let $(y, \boldsymbol{q}) \in L^2(\Omega) \times [H^1(\Omega)^*]^d$ be functions satisfying:

$$[\boldsymbol{q}, \boldsymbol{r}]_\Omega = (y, \nabla \cdot \boldsymbol{r})_\Omega, \tag{3.13a}$$

$$[\boldsymbol{q} + \boldsymbol{\beta} y, \boldsymbol{p}_g]_\Omega - (y \nabla \cdot \boldsymbol{\beta}, z_g)_\Omega = 0, \tag{3.13b}$$

for all $(\boldsymbol{r}, g) \in [H^1(\Omega)]^d \times L^2(\Omega)$. Consider $g = y$. From equation (3.9b) and taking into account that $\boldsymbol{p}_y = -\nabla z_y$, we have that

$$(y, \nabla \cdot \boldsymbol{p}_y)_\Omega + (\boldsymbol{\beta} y, \boldsymbol{p}_y)_\Omega - (y \nabla \cdot \boldsymbol{\beta}, z_y)_\Omega = (y, y)_\Omega. \tag{3.14}$$

Take $\boldsymbol{r} = \boldsymbol{p}_y$ in (3.13a). We obtain

$$[\boldsymbol{q}, \boldsymbol{p}_y]_\Omega = (y, \nabla \cdot \boldsymbol{p}_y)_\Omega.$$

Substitute this in (3.13b)

$$(y, \nabla \cdot \boldsymbol{p}_y)_\Omega + (\boldsymbol{\beta} y, \boldsymbol{p}_y)_\Omega - (y \nabla \cdot \boldsymbol{\beta}, z_y)_\Omega = 0. \tag{3.15}$$

From (3.14) and (3.15), it is clear that $y = 0$. From (3.13a) we have that $\boldsymbol{q} = 0$ and uniqueness is proved.

Existence is as follows. Take $y \in L^2(\Omega)$ the unique *very weak* solution of (3.7) and define $\boldsymbol{q} \in [H^1(\Omega)^*]^d$ by

$$[q_i, r]_\Omega = (y, \partial_{x_i} r)_\Omega - [u, r n_i]_\Gamma,$$

for all $r \in H^1(\Omega)$, and $n_i$ is the $i-$th component of the vector $\boldsymbol{n}$. Again, this is well defined because we have made sure that $\varepsilon < 1/2$, and the functions in $H^\varepsilon(\Gamma)$ now can have jump discontinuities.

**Corollary 3** *If* $u \in H^{1/2+t^*}(\Gamma)$ *for some* $0 \le t^* < 1$, *then* $(y, \boldsymbol{q}) \in H^{1+t^*}(\Omega) \times ([H^{t^*}(\Omega)]^d \cap H(\mathrm{div}, \Omega))$ *and*

$$(\boldsymbol{q}, \boldsymbol{r})_\Omega - (y, \nabla \cdot \boldsymbol{r})_\Omega + \langle u, \boldsymbol{r} \cdot \boldsymbol{n} \rangle_\Gamma = 0, \tag{3.16a}$$

$$(\nabla \cdot (\boldsymbol{q} + \boldsymbol{\beta} y), w)_\Omega - (y \nabla \cdot \boldsymbol{\beta}, w)_\Omega = 0, \tag{3.16b}$$

*for all* $(\boldsymbol{r}, w) \in H(\mathrm{div}, \Omega) \times L^2(\Omega)$.

For the sake of completeness, we say that $(y, \boldsymbol{q})$ is a very weak solution of the mixed formulation (3.16a)–(3.16b) if it is a very weak solution of (3.11) in the sense of Lemma2.

**3.2.2. Study of the Control Problem.** Now we are ready to study the control problem. Let us first formulate it using the mixed formulation.

$$(\mathrm{P}) \min_{u\in L^2(\Gamma)} J(u) = \frac{1}{2}\|y_u - y_d\|^2_{L^2(\Omega)} + \frac{\gamma}{2}\|u\|^2_{L^2(\Gamma)},$$

where $(y_u, \boldsymbol{q}_u) \in L^2(\Omega) \times [H^1(\Omega)^*]^d$ is the very weak solution of

$$(\boldsymbol{q}_u, \boldsymbol{r})_\Omega - (y_u, \nabla\cdot\boldsymbol{r})_\Omega + \langle u, \boldsymbol{r}\cdot\boldsymbol{n}\rangle_\Gamma = 0, \tag{3.17a}$$

$$(\nabla\cdot(\boldsymbol{q}_u + \boldsymbol{\beta} y_u), w)_\Omega - (y_u \nabla\cdot\boldsymbol{\beta}, w)_\Omega = 0, \tag{3.17b}$$

for all $(\boldsymbol{r}, w) \in H(\mathrm{div}, \Omega) \times L^2(\Omega)$.

**Theorem 6** *Assume $\Omega$ is convex. If $y_d \in H^{t^*}(\Omega)$ for some $0 \le t^* < 1$, then problem (P) has a unique solution $\bar{u} \in L^2(\Gamma)$. Moreover, for any $s \ge 1/2$ satisfying $s \le \frac{1}{2} + t^*$ and $s < \min\{\frac{3}{2}, \frac{\pi}{\omega} - \frac{1}{2}\}$, we have $\bar{u} \in H^s(\Gamma)$,*

$$(\bar{\boldsymbol{p}}, \bar{y}, \bar{z}) \in [H^{s+\frac{1}{2}}(\Omega)]^d \times H^{s+\frac{1}{2}}(\Omega) \times (H^{s+\frac{3}{2}}(\Omega) \cap H^1_0(\Omega)),$$
$$\bar{\boldsymbol{q}} \in [H^{s-\frac{1}{2}}(\Omega)]^d \cap H(\mathrm{div}, \Omega),$$

*and $\partial_{\boldsymbol{n}}\bar{z} = \bar{\boldsymbol{p}}\cdot\boldsymbol{n} \in H^s(\Gamma)$ such that*

$$(\bar{\boldsymbol{q}}, \boldsymbol{r})_\Omega - (\bar{y}, \nabla\cdot\boldsymbol{r})_\Omega + \langle \bar{u}, \boldsymbol{r}\cdot\boldsymbol{n}\rangle_\Gamma = 0, \tag{3.18a}$$

$$(\nabla\cdot(\bar{\boldsymbol{q}} + \boldsymbol{\beta}\bar{y}), w)_\Omega - (\bar{y}\nabla\cdot\boldsymbol{\beta}, w)_\Omega = 0, \tag{3.18b}$$

$$(\bar{\boldsymbol{p}}, \boldsymbol{r})_\Omega - (\bar{z}, \nabla\cdot\boldsymbol{r})_\Omega = 0, \tag{3.18c}$$

$$(\nabla\cdot(\bar{\boldsymbol{p}} - \boldsymbol{\beta}\bar{z}), w)_\Omega = (\bar{y} - y_d, w)_\Omega, \tag{3.18d}$$

$$\langle \gamma\bar{u} + \bar{\boldsymbol{p}}\cdot\boldsymbol{n}, v\rangle_\Gamma = 0, \tag{3.18e}$$

*for all $(\boldsymbol{r}, w, v) \in H(\mathrm{div}, \Omega) \times L^2(\Omega) \times L^2(\Gamma)$.*

**Proof:** The functional $J(u)$ is bounded from below and strictly convex, because thanks to 5 and 14, the control-to-state mapping is linear continuous. Using that it is also coercive, existence of solution follows from the standard argument of taking a minimizing sequence. Uniqueness of solution follows from the strict convexity.

Since the equation is linear, the functional is differentiable (it is $C^\infty$ indeed) and a standard argument leads to the necessary optimality conditions (3.18a)–(3.18e), where the first two equations must be understood in the very weak sense of Lemma 2. Since the problem is strictly convex, these conditions are also sufficient, and therefore the optimality system has a unique solution.

Let us study the regularity of the solution. We already have that $\bar{y} \in L^2(\Omega)$, so 15 leads in a first step to $(\bar{z}, \bar{\boldsymbol{p}}) \in (H^2(\Omega) \cap H_0^1(\Omega)) \times [H^1(\Omega)]^d$, $\partial_{\boldsymbol{n}} \bar{z} = \bar{\boldsymbol{p}} \cdot \boldsymbol{n} \in H^{1/2}(\Gamma)$. Noticing that from (3.18e) we have that $\bar{u} = -\gamma^{-1} \bar{\boldsymbol{p}} \cdot \boldsymbol{n}$, it is clear that the optimal control satisfies $\bar{u} \in H^{1/2}(\Gamma)$. From Lemma 14 we deduce $\bar{y} \in H^{3/2}(\Omega)$.

Using the just deduced regularity of $\bar{y}$ and bootstrapping the argument once, we achieve the desired result.

## 3.3. HDG FORMULATION AND IMPLEMENTATION

Throughout this section, we assume $\Omega$ is a polyhedral domain, not necessarily convex, with $d \geq 2$. We introduce the following discontinuous finite element spaces

$$\boldsymbol{V}_h := \{\boldsymbol{v} \in [L^2(\Omega)]^d : \boldsymbol{v}|_K \in [\mathcal{P}^k(K)]^d, \forall K \in \mathcal{T}_h\}, \tag{3.19}$$

$$W_h := \{w \in L^2(\Omega) : w|_K \in \mathcal{P}^{k+1}(K), \forall K \in \mathcal{T}_h\}, \tag{3.20}$$

$$M_h := \{\mu \in L^2(\varepsilon_h) : \mu|_e \in \mathcal{P}^{k+1}(e), \forall e \in \varepsilon_h\} \tag{3.21}$$

for the flux variables, scalar variables, and boundary trace variables, respectively.

**3.3.1. The HDG Formulation.** To approximate the solution of the mixed weak form (3.18a)-(3.18e) of the optimality system, the HDG method seeks approximate fluxes $\boldsymbol{q}_h, \boldsymbol{p}_h \in \boldsymbol{V}_h$, states $y_h, z_h \in W_h$, interior element boundary traces $\widehat{y}_h^o, \widehat{z}_h^o \in M_h(o)$, and boundary control $u_h \in M_h(\partial)$ satisfying

$$(\boldsymbol{q}_h, \boldsymbol{r}_1)_{\mathcal{T}_h} - (y_h, \nabla\cdot\boldsymbol{r}_1)_{\mathcal{T}_h} + \langle \widehat{y}_h^o, \boldsymbol{r}_1\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} + \langle u_h, \boldsymbol{r}_1\cdot\boldsymbol{n}\rangle_{\varepsilon_h^\partial} = 0, \tag{3.22a}$$

$$\begin{aligned}-(\boldsymbol{q}_h + \boldsymbol{\beta} y_h, \nabla w_1)_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta} y_h, w_1)_{\mathcal{T}_h} + \langle \widehat{\boldsymbol{q}}_h\cdot\boldsymbol{n}, w_1\rangle_{\partial\mathcal{T}_h}& \\ +\langle \boldsymbol{\beta}\cdot\boldsymbol{n} u_h, w_1\rangle_{\varepsilon_h^\partial} = (f, w_1)_{\mathcal{T}_h},&\end{aligned} \tag{3.22b}$$

for all $(\boldsymbol{r}_1, w_1) \in \boldsymbol{V}_h \times W_h$,

$$(\boldsymbol{p}_h, \boldsymbol{r}_2)_{\mathcal{T}_h} - (z_h, \nabla\cdot\boldsymbol{r}_2)_{\mathcal{T}_h} + \langle \widehat{z}_h^o, \boldsymbol{r}_2\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} = 0, \tag{3.22c}$$

$$\begin{aligned}-(\boldsymbol{p}_h - \boldsymbol{\beta} z_h, \nabla w_2)_{\mathcal{T}_h} + \langle \widehat{\boldsymbol{p}}_h\cdot\boldsymbol{n}, w_2\rangle_{\partial\mathcal{T}_h} - \langle \boldsymbol{\beta}\cdot\boldsymbol{n}\widehat{z}_h^o, w_2\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}& \\ -(y_h, w_2)_{\mathcal{T}_h} = -(y_d, w_2)_{\mathcal{T}_h},&\end{aligned} \tag{3.22d}$$

for all $(\boldsymbol{r}_2, w_2) \in \boldsymbol{V}_h \times W_h$,

$$\langle \widehat{\boldsymbol{q}}_h\cdot\boldsymbol{n} + \boldsymbol{\beta}\cdot\boldsymbol{n}\widehat{y}_h^o, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} = 0, \tag{3.22e}$$

for all $\mu_1 \in M_h(o)$,

$$\langle \widehat{\boldsymbol{p}}_h\cdot\boldsymbol{n} - \boldsymbol{\beta}\cdot\boldsymbol{n}\widehat{z}_h^o, \mu_2\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} = 0, \tag{3.22f}$$

for all $\mu_2 \in M_h(o)$, and the optimality condition

$$\langle u_h, \mu_3\rangle_{\varepsilon_h^\partial} + \langle \gamma^{-1}\widehat{\boldsymbol{p}}_h\cdot\boldsymbol{n}, \mu_3\rangle_{\varepsilon_h^\partial} = 0, \tag{3.22g}$$

for all $\mu_3 \in M_h(\partial)$.

The numerical traces on $\partial\mathcal{T}_h$ are defined as

$$\widehat{\boldsymbol{q}}_h \cdot \boldsymbol{n} = \boldsymbol{q}_h \cdot \boldsymbol{n} + h^{-1}(y_h - \widehat{y}_h^o) + \tau_1(y_h - \widehat{y}_h^o) \qquad \text{on } \partial\mathcal{T}_h\backslash\varepsilon_h^\partial, \tag{3.22h}$$

$$\widehat{\boldsymbol{q}}_h \cdot \boldsymbol{n} = \boldsymbol{q}_h \cdot \boldsymbol{n} + h^{-1}(y_h - u_h) + \tau_1(y_h - u_h) \qquad \text{on } \varepsilon_h^\partial, \tag{3.22i}$$

$$\widehat{\boldsymbol{p}}_h \cdot \boldsymbol{n} = \boldsymbol{p}_h \cdot \boldsymbol{n} + h^{-1}(z_h - \widehat{z}_h^o) + \tau_2(z_h - \widehat{y}_h^o) \qquad \text{on } \partial\mathcal{T}_h\backslash\varepsilon_h^\partial, \tag{3.22j}$$

$$\widehat{\boldsymbol{p}}_h \cdot \boldsymbol{n} = \boldsymbol{p}_h \cdot \boldsymbol{n} + h^{-1}z_h + \tau_2 z_h \qquad \text{on } \varepsilon_h^\partial, \tag{3.22k}$$

where $\tau_1$ and $\tau_2$ are stabilization functions defined on $\partial\mathcal{T}_h$. This completes the formulation of the HDG method.

To guarantee the stability for existing HDG methods, the stabilization functions $\tau_1$ and $\tau_2$ for the (uncoupled) convection diffusion equation and the dual problem are chosen to satisfy

$$\tau_1 \geq \frac{1}{2}\boldsymbol{\beta} \cdot \boldsymbol{n}, \quad \tau_2 \geq -\frac{1}{2}\boldsymbol{\beta} \cdot \boldsymbol{n}$$

on $\partial\mathcal{T}_h$; see, e.g., [15, 16, 28, 56]. However, in our convergence analysis in Section 3.4 for the fully coupled optimality system we require the stabilization functions to be chosen very specifically. These requirements on the stabilization functions arise naturally in our analysis.

**3.3.2. Implementation.** For the HDG implementation, we proceed similarly to our earlier work [43]. A fundamental aspect of the HDG method is the local solver, which reduces the number of globally coupled unknowns. The standard approach is to implement the local solver element-by-element independently and then assemble the global system. As in [43], here we first assemble a large global system and then reduce the size of the system using simple block-diagonal matrix operations. This process is equivalent to the standard approach.

Substitute (3.22h)-(3.22k) into (3.22a)-(3.22g) and perform some simple manipulations to obtain

$$(\boldsymbol{q}_h, \boldsymbol{p}_h, y_h, z_h, \widehat{y}_h^o, \widehat{z}_h^o, u_h) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times W_h \times M_h(o) \times M_h(o) \times M_h(\partial)$$

is the solution of the following weak formulation:

$$(\boldsymbol{q}_h, \boldsymbol{r_1})_{\mathcal{T}_h} - (y_h, \nabla \cdot \boldsymbol{r_1})_{\mathcal{T}_h} + \langle \widehat{y}_h^o, \boldsymbol{r_1} \cdot \boldsymbol{n} \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} + \langle u_h, \boldsymbol{r_1} \cdot \boldsymbol{n} \rangle_{\varepsilon_h^\partial} = 0, \tag{3.23a}$$

$$(\boldsymbol{p}_h, \boldsymbol{r_2})_{\mathcal{T}_h} - (z_h, \nabla \cdot \boldsymbol{r_2})_{\mathcal{T}_h} + \langle \widehat{z}_h^o, \boldsymbol{r_2} \cdot \boldsymbol{n} \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} = 0, \tag{3.23b}$$

$$\begin{aligned} &(\nabla \cdot \boldsymbol{q}_h, w_1)_{\mathcal{T}_h} - (\boldsymbol{\beta} y_h, \nabla w_1)_{\mathcal{T}_h} - (\nabla \cdot \boldsymbol{\beta} y_h, w_1)_{\mathcal{T}_h} \\ &+\langle (h^{-1} + \tau_1) y_h, w_1 \rangle_{\partial \mathcal{T}_h} + \langle (\boldsymbol{\beta} \cdot \boldsymbol{n} - \tau_1 - h^{-1}) \widehat{y}_h^o, w_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} \\ &+\langle (\boldsymbol{\beta} \cdot \boldsymbol{n} - \tau_1 - h^{-1}) u_h, w_1 \rangle_{\varepsilon_h^\partial} = (f, w_1)_{\mathcal{T}_h}, \end{aligned} \tag{3.23c}$$

$$\begin{aligned} &(\nabla \cdot \boldsymbol{p}_h, w_2)_{\mathcal{T}_h} - (y_h, w_2)_{\mathcal{T}_h} + (\boldsymbol{\beta} z_h, \nabla w_2)_{\mathcal{T}_h} + \langle (h^{-1} + \tau_2) z_h, w_2 \rangle_{\partial \mathcal{T}_h} \\ &-\langle (h^{-1} + \tau_2 + \boldsymbol{\beta} \cdot \boldsymbol{n}) \widehat{z}_h^o, w_2 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} = -(y_d, w_2)_{\mathcal{T}_h}, \end{aligned} \tag{3.23d}$$

$$\begin{aligned} &\langle \boldsymbol{q}_h \cdot \boldsymbol{n}, \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} + \langle (h^{-1} + \tau_1) y_h, \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} \\ &+\langle (\boldsymbol{\beta} \cdot \boldsymbol{n} - \tau_1 - h^{-1}) \widehat{y}_h^o, \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} = 0, \end{aligned} \tag{3.23e}$$

$$\begin{aligned} &\langle \boldsymbol{p}_h \cdot \boldsymbol{n}, \mu_2 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} + \langle (h^{-1} + \tau_2) z_h, \mu_2 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} \\ &-\langle (\boldsymbol{\beta} \cdot \boldsymbol{n} + \tau_2 + h^{-1}) \widehat{z}_h^o, \mu_2 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} = 0, \end{aligned} \tag{3.23f}$$

$$\langle \boldsymbol{p}_h \cdot \boldsymbol{n}, \mu_3 \rangle_{\varepsilon_h^\partial} + \gamma \langle u_h, \mu_3 \rangle_{\varepsilon_h^\partial} + \langle (h^{-1} + \tau_2) z_h, \mu_3 \rangle_{\varepsilon_h^\partial} = 0, \tag{3.23g}$$

for all $(\boldsymbol{r}_1, \boldsymbol{r}_2, w_1, w_2, \mu_1, \mu_2, \mu_3) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times W_h \times M_h(o) \times M_h(o) \times M_h(\partial)$.

For $\boldsymbol{V}_h = \operatorname{span}\{\boldsymbol{\varphi}_i\}_{i=1}^{N_1}$, $W_h = \operatorname{span}\{\phi_i\}_{i=1}^{N_2}$, $M_h^o = \operatorname{span}\{\psi_i\}_{i=1}^{N_3}$, and $M_h^\partial = \operatorname{span}\{\psi_i\}_{i=1+N_3}^{N_4}$, assume

$$\begin{aligned}
&\boldsymbol{q}_h = \sum_{j=1}^{N_1} q_j \boldsymbol{\varphi}_j, \quad \boldsymbol{p}_h = \sum_{j=1}^{N_1} p_j \boldsymbol{\varphi}_j \quad y_h = \sum_{j=1}^{N_2} y_j \phi_j \quad z_h = \sum_{j=1}^{N_2} z_j \phi_j, \\
&\widehat{y}_h^o = \sum_{j=1}^{N_3} \alpha_j \psi_j, \quad \widehat{z}_h^o = \sum_{j=1}^{N_3} \gamma_j \psi_j, \quad u_h = \sum_{j=1+N_3}^{N_4} \beta_j \psi_j.
\end{aligned} \tag{3.24}$$

Substitute (3.24) into (3.23a)-(3.23f) and use the corresponding test functions to test (3.23a)-(3.23f), respectively, to obtain the matrix equation

$$\begin{bmatrix}
A_1 & 0 & -A_2 & 0 & A_{20} & 0 & A_{21} \\
0 & A_1 & 0 & -A_2 & 0 & A_{20} & 0 \\
A_2^T & 0 & A_{18} & 0 & A_{22} & 0 & A_{23} \\
0 & A_2^T & -A_{12} & A_{19} & 0 & A_{24} & 0 \\
A_{20}^T & 0 & A_{25} & 0 & A_{26} & 0 & 0 \\
0 & A_{20}^T & 0 & A_{24} & 0 & A_{25} & 0 \\
0 & A_{26} & 0 & A_{27} & 0 & 0 & \gamma A_{28}
\end{bmatrix}
\begin{bmatrix}
\mathfrak{q} \\ \mathfrak{p} \\ \mathfrak{y} \\ \mathfrak{z} \\ \widehat{\mathfrak{y}} \\ \widehat{\mathfrak{z}} \\ \mathfrak{u}
\end{bmatrix}
=
\begin{bmatrix}
0 \\ 0 \\ b_1 \\ -b_2 \\ 0 \\ 0 \\ 0
\end{bmatrix}, \tag{3.25}$$

where $\mathfrak{q}, \mathfrak{p}, \mathfrak{y}, \mathfrak{z}, \widehat{\mathfrak{y}}, \widehat{\mathfrak{z}}, \mathfrak{u}$ are the coefficient vectors for $\boldsymbol{q}_h, \boldsymbol{p}_h, y_h, z_h, \widehat{y}_h^o, \widehat{z}_h^o, u_h$, respectively, and

$$
\begin{aligned}
&A_1 = [(\boldsymbol{\varphi}_j, \boldsymbol{\varphi}_i)_{\mathcal{T}_h}], \quad A_2 = [(\phi_j, \nabla\cdot\boldsymbol{\varphi}_i)_{\mathcal{T}_h}], \quad A_3 = [(\psi_j, \boldsymbol{\varphi}_i\cdot\boldsymbol{n})_{\mathcal{T}_h}],\\
&A_4 = [(\boldsymbol{\beta}\phi_j, \nabla\phi_i)_{\mathcal{T}_h}], \quad A_5 = [(\nabla\cdot\boldsymbol{\beta}\phi_j, \phi_i)_{\mathcal{T}_h}], \quad A_6 = [(\boldsymbol{\beta}\phi_j, \nabla\phi_i)_{\mathcal{T}_h}],\\
&A_7 = [\langle h^{-1}\phi_j, \phi_i\rangle_{\partial\mathcal{T}_h}], \quad A_8 = [\langle \tau_1\phi_j, \phi_i\rangle_{\partial\mathcal{T}_h}], \quad A_9 = [\langle \tau_2\phi_j, \phi_i\rangle_{\partial\mathcal{T}_h}],\\
&A_{10} = [\langle \boldsymbol{\beta}\cdot\boldsymbol{n}\psi_j, \phi_i\rangle_{\partial\mathcal{T}_h}], \ A_{11} = [\left\langle \tau_1\psi_j, \phi_i\right\rangle_{\partial\mathcal{T}_h}], \ A_{12} = [\left\langle h^{-1}\psi_j, \phi_i\right\rangle_{\partial\mathcal{T}_h}],\\
&A_{13} = [(\phi_j, \phi_i)_{\mathcal{T}_h}], \ A_{14} = [\left\langle \tau_1\psi_j, \psi_i\right\rangle_{\partial\mathcal{T}_h}], \ A_{15} = [\left\langle h^{-1}\psi_j, \psi_i\right\rangle_{\partial\mathcal{T}_h}],\\
&A_{16} = [\left\langle \boldsymbol{\beta}\cdot\boldsymbol{n}\psi_j, \psi_i\right\rangle_{\partial\mathcal{T}_h}], \quad A_{17} = [\left\langle \tau_2\psi_j, \psi_i\right\rangle_{\partial\mathcal{T}_h}],\\
&A_{18} = A_7 + A_8 - A_4 - A_5, \ A_{19} = A_4 + A_7 + A_9,\\
&b_1 = [(f, \phi_i)_{\mathcal{T}_h}], \quad b_2 = [(y_d, \phi_i)_{\mathcal{T}_h}].
\end{aligned}
$$

The remaining matrices are constructed by extracting the corresponding rows and columns from linear combinations of $A_3$ to $A_{17}$.

Equation (3.25) can be rewritten as

$$
\begin{bmatrix} B_1 & B_2 & B_3 \\ -B_2^T & B_4 & B_5 \\ B_6 & B_7 & B_8 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{bmatrix} = \begin{bmatrix} 0 \\ b \\ 0 \end{bmatrix}, \tag{3.26}
$$

where $\boldsymbol{\alpha} = [\mathfrak{q}; \mathfrak{p}]$, $\boldsymbol{\beta} = [\mathfrak{y}; \mathfrak{z}]$, $\boldsymbol{\gamma} = [\widehat{\mathfrak{y}}; \widehat{\mathfrak{z}}; \mathfrak{u}]$, $b = [b_1; -b_2]$, and $\{B_i\}_{i=1}^8$ are the corresponding blocks of the coefficient matrix in (3.25).

As in Section one, we use the first two equations of (3.25) to solve for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ using simple and efficient block-diagonal matrix computations. Eliminating $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ gives a reduced globally coupled equation for $\boldsymbol{\gamma}$ only:

$$
\mathbb{K}\boldsymbol{\gamma} = \mathbb{F}. \tag{3.27}
$$

Note that the globally coupled system only involves the vector $\boldsymbol{\gamma}$, which contains the coefficients of the approximate boundary traces. Therefore, the number of globally coupled degrees of freedom is much smaller than the total number of degrees of freedom for all variables.

The details of the above procedure are similar to Section one. We only need to show the following result.

**Proposition 3.3.1** *If* $\min{(\tau_1 - \frac{1}{2}\boldsymbol{\beta}\cdot\boldsymbol{n})}|_{\partial K} > 0$ *and* $\min{(\tau_2 + \frac{1}{2}\boldsymbol{\beta}\cdot\boldsymbol{n})}|_{\partial K} > 0$ *for any* $K \in \mathcal{T}_h$*, then the matrices* $A_{18}$ *and* $A_{19}$ *in* (3.25) *are positive definite.*

**Proof:** We only prove $A_{18}$ is positive definite; a similar argument applies to $A_{19}$. The matrix $A_{18}$ is positive definite if and only if $\boldsymbol{x}^T A_{18}\boldsymbol{x} > 0$ for any $\boldsymbol{x} = [x_1, x_2, \cdots, x_{N_2}] \in \mathbb{R}^{N_2}$. For $x = \sum_{j=1}^{N_2} x_j\phi_j$, we have

$$\boldsymbol{x}^T A_{18}\boldsymbol{x} = \langle h^{-1}x, x\rangle_{\partial\mathcal{T}_h} + \langle \tau_1 x, x\rangle_{\partial\mathcal{T}_h} - (\boldsymbol{\beta}x, \nabla x)_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta}x, x)_{\mathcal{T}_h}.$$

Moreover,

$$\begin{aligned}(\boldsymbol{\beta}x, \nabla x)_{\mathcal{T}_h} &= (\boldsymbol{\beta}\cdot\nabla x, x)_{\mathcal{T}_h} = (\nabla\cdot(\boldsymbol{\beta}x), x)_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta}x, x)_{\mathcal{T}_h}\\ &= \langle \boldsymbol{\beta}\cdot\boldsymbol{n}x, x\rangle_{\partial\mathcal{T}_h} - (\boldsymbol{\beta}x, \nabla x)_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta}x, x)_{\mathcal{T}_h},\end{aligned}$$

which implies

$$(\boldsymbol{\beta}x, \nabla x)_{\mathcal{T}_h} = \frac{1}{2}\langle \boldsymbol{\beta}\cdot\boldsymbol{n}x, x\rangle_{\partial\mathcal{T}_h} - \frac{1}{2}(\nabla\cdot\boldsymbol{\beta}x, x)_{\mathcal{T}_h}.$$

Therefore,

$$\boldsymbol{x}^T A_{18}\boldsymbol{x} = \langle (h^{-1} + \tau_1 - \frac{1}{2}\boldsymbol{\beta}\cdot\boldsymbol{n})x, x\rangle_{\partial\mathcal{T}_h} - \frac{1}{2}(\nabla\cdot\boldsymbol{\beta}x, x)_{\mathcal{T}_h} > 0,$$

by the assumption concerning $\tau_1$ and the condition $\nabla\cdot\boldsymbol{\beta} \leq 0$.

## 3.4. ERROR ANALYSIS

Next, we provide a convergence analysis of the above HDG method for the Dirichlet boundary control problem. We assume the solution of the optimality system has certain regularity properties. In 2D, due to the theoretical results in Section 3.2, we can give simple conditions that guarantee the unique solution has the necessary regularity. In 3D, we lack the necessary regularity theory; however, our convergence results still apply if there exists a unique solution of the optimality system with the required regularity.

We begin with a precise statement of our assumptions and the main convergence result.

**3.4.1. Assumptions and Main Result.** Throughout this section, we assume $\Omega$ is a bounded convex polyhedral domain. We assume throughout that $\boldsymbol{\beta}$ satisfies

$$\boldsymbol{\beta} \in [C(\overline{\Omega})]^d, \quad \nabla \cdot \boldsymbol{\beta} \in L^\infty(\Omega), \quad \nabla \cdot \boldsymbol{\beta} \leq 0, \quad \nabla\nabla \cdot \boldsymbol{\beta} \in [L^2(\Omega)]^d. \tag{3.28}$$

Note that this condition is slightly stronger than the condition (3.5) made for the analysis in 3.2. Here, we assume $\boldsymbol{\beta}$ is continuous on $\overline{\Omega}$, while before we assumed $\boldsymbol{\beta} \in [L^\infty(\Omega)]^d$.

For our theoretical results, we choose the stabilization functions $\tau_1$ and $\tau_2$ to satisfy

**(A1)** $\tau_2$ is piecewise constant on $\partial\mathcal{T}_h$.

**(A2)** $\tau_1 = \tau_2 + \boldsymbol{\beta} \cdot \boldsymbol{n}$.

**(A3)** For any $K \in \mathcal{T}_h$, $\min\,(\tau_2 + \frac{1}{2}\boldsymbol{\beta} \cdot \boldsymbol{n})|_{\partial K} > 0$.

We note that **(A2)** and **(A3)** imply

$$\min\,(\tau_1 - \frac{1}{2}\boldsymbol{\beta} \cdot \boldsymbol{n})|_{\partial K} > 0 \quad \text{for any } K \in \mathcal{T}_h. \tag{3.29}$$

In our analysis, we use the conditions **(A3)** and (3.29) frequently and therefore we rarely mention them explicitly. However, we use **(A1)** and **(A2)** less frequently, and therefore we typically mention these conditions when we use them.

We also assume throughout that there exists a unique solution of the optimality system (3.18a)–(3.18e) that satisfies

$$y \in H^{r_y}(\Omega), \quad z \in H^{r_z}(\Omega) \cap H_0^1(\Omega), \quad \boldsymbol{q} \in [H^{r_{\boldsymbol{q}}}(\Omega)]^d, \quad \boldsymbol{p} \in [H^{r_{\boldsymbol{p}}}(\Omega)]^d, \tag{3.30}$$

where

$$r_y > 1, \quad r_z > 2, \quad r_{\boldsymbol{q}} > 1/2, \quad r_{\boldsymbol{p}} > 1. \tag{3.31}$$

This regularity condition ensures that the convergence rates in Theorem 7 below are positive for all variables.

We note that we require $r_{\boldsymbol{q}} > 1/2$ (instead of $r_{\boldsymbol{q}} > 0$) here in order to guarantee $\boldsymbol{q}$ has a well-defined boundary trace in $L^2(\Gamma)$. We use this property in our analysis. As mentioned in the introduction, we relax this assumption in the second part of this work and only require $r_{\boldsymbol{q}} > 0$. Dealing with the very low regularity of $\boldsymbol{q}$ requires entirely different HDG analysis techniques than we use here.

In the 2D case, simple conditions on the desired state $y_d$ and the domain $\Omega$ guarantee that the solution has the above regularity; see Corollary 4 below. In the 3D case, we do not have theory that gives simple conditions guaranteeing such solutions exist.

We now state our main convergence result.

**Theorem 7** *Let*

$$\begin{aligned} s_{\boldsymbol{q}} &= \min\{r_{\boldsymbol{q}}, k+1\}, \qquad & s_y &= \min\{r_y, k+2\}, \\ s_{\boldsymbol{p}} &= \min\{r_{\boldsymbol{p}}, k+1\}, \qquad & s_z &= \min\{r_z, k+2\}. \end{aligned} \tag{3.32}$$

*If the above assumptions hold, then*

$$\begin{aligned}
\|u - u_h\|_{\varepsilon_h^\partial} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}\,,\\
\|y - y_h\|_{\mathcal{T}_h} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}\,,\\
\|\boldsymbol{p} - \boldsymbol{p}_h\|_{\mathcal{T}_h} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}\,,\\
\|z - z_h\|_{\mathcal{T}_h} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}\,.
\end{aligned}$$

*If in addition $k \geq 1$, then*

$$\|\boldsymbol{q} - \boldsymbol{q}_h\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}-1} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-2} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-1} \|y\|_{s_y,\Omega}\,.$$

Now we specialize to the 2D case. For a convex polygonal domain $\Omega$, let $\omega$ denote its largest interior angle. As mentioned before, $\omega$ must satisfy $1 < \pi/\omega \leq 3$, i.e., $\omega \in [\pi/3, \pi)$. The limiting regularity condition is $r_{\boldsymbol{q}} > 1/2$. Therefore, to guarantee the regularity condition (3.30)-(3.31), by Theorem 6 we need two conditions:

1. $\pi/\omega - 1/2 > 1$, i.e., $\omega < 2\pi/3$, and
2. $y_d \in H^{t^*}(\Omega)$ for some $t^* \in (1/2, 1)$.

As mentioned earlier, we remove these restrictions in the second part of this work.

Applying Theorem 6 and the main theorem above gives the following result.

**Corollary 4** *Suppose $d = 2$, $f = 0$, and $y_d \in H^{t^*}(\Omega)$ for some $t^* \in (1/2, 1)$. Let $\omega \in (\pi/3, 2\pi/3)$ be the largest interior angle of $\Gamma$, and define $r_\Omega$ by*

$$r_\Omega = \min\left\{\frac{3}{2}, \frac{\pi}{\omega} - \frac{1}{2}, t^* + \frac{1}{2}\right\} \in (1, 3/2).$$

*Then the regularity condition* (3.30)-(3.31) *is satisfied. Also, if* $k = 1$*, then for any* $r < r_\Omega$ *we have*

$$
\begin{aligned}
\|u - u_h\|_{\mathcal{E}_h^\partial} &\lesssim h^r(\|\boldsymbol{p}\|_{H^{r+1/2}(\Omega)} + \|z\|_{H^{r+3/2}(\Omega)} + \|\boldsymbol{q}\|_{H^{r-1/2}(\Omega)} + \|y\|_{H^{r+1/2}(\Omega)}),\\
\|y - y_h\|_{\mathcal{T}_h} &\lesssim h^r(\|\boldsymbol{p}\|_{H^{r+1/2}(\Omega)} + \|z\|_{H^{r+3/2}(\Omega)} + \|\boldsymbol{q}\|_{H^{r-1/2}(\Omega)} + \|y\|_{H^{r+1/2}(\Omega)}),\\
\|\boldsymbol{q} - \boldsymbol{q}_h\|_{\mathcal{T}_h} &\lesssim h^{r-1/2}(\|\boldsymbol{p}\|_{H^{r+1/2}(\Omega)} + \|z\|_{H^{r+3/2}(\Omega)} + \|\boldsymbol{q}\|_{H^{r-1/2}(\Omega)} + \|y\|_{H^{r+1/2}(\Omega)}),\\
\|\boldsymbol{p} - \boldsymbol{p}_h\|_{\mathcal{T}_h} &\lesssim h^r(\|\boldsymbol{p}\|_{H^{r+1/2}(\Omega)} + \|z\|_{H^{r+3/2}(\Omega)} + \|\boldsymbol{q}\|_{H^{r-1/2}(\Omega)} + \|y\|_{H^r+1/2(\Omega)}),\\
\|z - z_h\|_{\mathcal{T}_h} &\lesssim h^r(\|\boldsymbol{p}\|_{H^{r+1/2}(\Omega)} + \|z\|_{H^{r+3/2}(\Omega)} + \|\boldsymbol{q}\|_{H^{r-1/2}(\Omega)} + \|y\|_{H^{r+1/2}(\Omega)}).
\end{aligned}
$$

*Furthermore, if* $k = 0$*, then for any* $r \in (1, r_\Omega)$ *we have*

$$
\begin{aligned}
\|u - u_h\|_{\mathcal{E}_h^\partial} &\lesssim h^{1/2}(\|\boldsymbol{p}\|_{H^1(\Omega)} + \|z\|_{H^2(\Omega)} + \|\boldsymbol{q}\|_{H^{r-1/2}(\Omega)} + \|y\|_{H^{r+1/2}(\Omega)}),\\
\|y - y_h\|_{\mathcal{T}_h} &\lesssim h^{1/2}(\|\boldsymbol{p}\|_{H^1(\Omega)} + \|z\|_{H^2(\Omega)} + \|\boldsymbol{q}\|_{H^{r-1/2}(\Omega)} + \|y\|_{H^{r+1/2}(\Omega)}),\\
\|\boldsymbol{p} - \boldsymbol{p}_h\|_{\mathcal{T}_h} &\lesssim h^{1/2}(\|\boldsymbol{p}\|_{H^1(\Omega)} + \|z\|_{H^2(\Omega)} + \|\boldsymbol{q}\|_{H^{r-1/2}(\Omega)} + \|y\|_{H^{r+1/2}(\Omega)}),\\
\|z - z_h\|_{\mathcal{T}_h} &\lesssim h^{1/2}(\|\boldsymbol{p}\|_{H^1(\Omega)} + \|z\|_{H^2(\Omega)} + \|\boldsymbol{q}\|_{H^{r-1/2}(\Omega)} + \|y\|_{H^{r+1/2}(\Omega)}).
\end{aligned}
$$

Theorem 6 gives $u \in H^r(\Gamma)$, and so the convergence rate for the control is optimal for $k = 1$. Similarly, the convergence rate for the flux $\boldsymbol{q}$ is optimal for $k = 1$. The convergence rates are suboptimal for the other variables when $k = 1$ and for all variables when $k = 0$.

Since $r_\Omega \in (1, 3/2)$, when $k = 1$ this result guarantees a superlinear convergence rate for all variables except $\boldsymbol{q}$. Also, if $\Omega$ is a rectangle (i.e., $\omega = \pi/2$), $y_d \in H^{1-\varepsilon}(\Omega)$, and $k = 1$, then $r_\Omega = 3/2 - \varepsilon$ and therefore for any $\varepsilon > 0$ all variables except $\boldsymbol{q}$ converge at the rate $O(h^{3/2-\varepsilon})$, and $\boldsymbol{q}$ converges at the rate $O(h^{1-\varepsilon})$.

**3.4.2. Preliminary Material.** Next, we discuss $L^2$ projections, HDG operators $\mathscr{B}_1$ and $\mathscr{B}_2$, and the well-posedness of the HDG equations.

We first define the standard $L^2$ projections $\mathbf{\Pi} : [L^2(\Omega)]^d \to \boldsymbol{V}_h$, $\Pi : L^2(\Omega) \to W_h$, and $P_M : L^2(\varepsilon_h) \to M_h$, which satisfy

$$
\begin{aligned}
(\mathbf{\Pi}\boldsymbol{q}, \boldsymbol{r})_K &= (\boldsymbol{q}, \boldsymbol{r})_K, \qquad \forall \boldsymbol{r} \in [\mathcal{P}_k(K)]^d,\\
(\Pi y, w)_K &= (y, w)_K, \qquad \forall w \in \mathcal{P}_{k+1}(K),\\
\langle P_M m, \mu\rangle_e &= \langle m, \mu\rangle_e, \qquad \forall \mu \in \mathcal{P}_{k+1}(e).
\end{aligned} \tag{3.33}
$$

In the analysis, we use the following classical results:

$$
\|\boldsymbol{q} - \mathbf{\Pi}\boldsymbol{q}\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{q}}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega}, \quad \|y - \Pi y\|_{\mathcal{T}_h} \lesssim h^{s_y} \|y\|_{s_y,\Omega}, \tag{3.34a}
$$

$$
\|y - \Pi y\|_{\partial\mathcal{T}_h} \lesssim h^{s_y - \frac{1}{2}} \|y\|_{s_y,\Omega}, \quad \|\boldsymbol{q}\cdot\boldsymbol{n} - \mathbf{\Pi}\boldsymbol{q}\cdot\boldsymbol{n}\|_{\partial\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{q}} - \frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega}, \tag{3.34b}
$$

$$
\|w\|_{\partial\mathcal{T}_h} \lesssim h^{-\frac{1}{2}} \|w\|_{\mathcal{T}_h}, \quad \forall w \in W_h, \tag{3.34c}
$$

We have the same projection error bounds for $\boldsymbol{p}$ and $z$.

We define the following HDG operators $\mathscr{B}_1$ and $\mathscr{B}_2$.

$$
\begin{aligned}
&\mathscr{B}_1(\boldsymbol{q}_h, y_h, \widehat{y}_h^o; \boldsymbol{r}_1, w_1, \mu_1)\\
&\quad= (\boldsymbol{q}_h, \boldsymbol{r}_1)_{\mathcal{T}_h} - (y_h, \nabla\cdot\boldsymbol{r}_1)_{\mathcal{T}_h} + \langle \widehat{y}_h^o, \boldsymbol{r}_1\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\qquad - (\boldsymbol{q}_h + \boldsymbol{\beta} y_h, \nabla w_1)_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta} y_h, w_1)_{\mathcal{T}_h}\\
&\qquad + \langle \boldsymbol{q}_h\cdot\boldsymbol{n} + (h^{-1} + \tau_1) y_h, w_1\rangle_{\partial\mathcal{T}_h} + \langle (\boldsymbol{\beta}\cdot\boldsymbol{n} - h^{-1} - \tau_1)\widehat{y}_h^o, w_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\qquad - \langle \boldsymbol{q}_h\cdot\boldsymbol{n} + \boldsymbol{\beta}\cdot\boldsymbol{n}\widehat{y}_h^o + (h^{-1} + \tau_1)(y_h - \widehat{y}_h^o), \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial},
\end{aligned} \tag{3.35}
$$

$$
\begin{aligned}
&\mathscr{B}_2(\boldsymbol{p}_h, z_h, \widehat{z}_h^o; \boldsymbol{r}_2, w_2, \mu_2)\\
&\quad= (\boldsymbol{p}_h, \boldsymbol{r}_2)_{\mathcal{T}_h} - (z_h, \nabla\cdot\boldsymbol{r}_2)_{\mathcal{T}_h} + \langle \widehat{z}_h^o, \boldsymbol{r}_2\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} - (\boldsymbol{p}_h - \boldsymbol{\beta} z_h, \nabla w_2)_{\mathcal{T}_h}\\
&\qquad + \langle \boldsymbol{p}_h\cdot\boldsymbol{n} + (h^{-1} + \tau_2) z_h, w_2\rangle_{\partial\mathcal{T}_h} - \langle (\boldsymbol{\beta}\cdot\boldsymbol{n} + h^{-1} + \tau_2)\widehat{z}_h^o, w_2\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\qquad - \langle \boldsymbol{p}_h\cdot\boldsymbol{n} - \boldsymbol{\beta}\cdot\boldsymbol{n}\widehat{z}_h^o + (h^{-1} + \tau_2)(z_h - \widehat{z}_h^o), \mu_2\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}.
\end{aligned} \tag{3.36}
$$

By the definition of $\mathscr{B}_1$ and $\mathscr{B}_2$, we can rewrite the HDG formulation of the optimality system (4.1), as follows: find $(\boldsymbol{q}_h, \boldsymbol{p}_h, y_h, z_h, \widehat{y}_h^o, \widehat{z}_h^o, u_h) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times W_h \times M_h(o) \times M_h(o) \times M_h(\partial)$ such that

$$
\begin{aligned}
\mathscr{B}_1(\boldsymbol{q}_h, y_h, \widehat{y}_h^o; \boldsymbol{r}_1, w_1, \mu_1) &= (f, w_1)_{\mathcal{T}_h} - \langle u_h, \boldsymbol{r}_1 \cdot \boldsymbol{n} \rangle_{\varepsilon_h^\partial} \\
&\quad - \langle (\boldsymbol{\beta} \cdot \boldsymbol{n} - h^{-1} - \tau_1) u_h, w_1 \rangle_{\varepsilon_h^\partial}, && \text{(3.37a)} \\
\mathscr{B}_2(\boldsymbol{p}_h, z_h, \widehat{z}_h^o; \boldsymbol{r}_2, w_2, \mu_2) &= (y_h - y_d, w_2)_{\mathcal{T}_h}, && \text{(3.37b)} \\
\gamma^{-1} \langle \boldsymbol{p}_h \cdot \boldsymbol{n} + h^{-1} z_h + \tau_2 z_h, \mu_3 \rangle_{\varepsilon_h^\partial} &= -\langle u_h, \mu_3 \rangle_{\varepsilon_h^\partial}, && \text{(3.37c)}
\end{aligned}
$$

for all $(\boldsymbol{r}_1, \boldsymbol{r}_2, w_1, w_2, \mu_1, \mu_2, \mu_3) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times W_h \times M_h(o) \times M_h(o) \times M_h(\partial)$.

Next, we present a basic property of the operators $\mathscr{B}_1$ and $\mathscr{B}_2$, and show the HDG equations (3.37) have a unique solution.

**Lemma 16** *For any $(\boldsymbol{v}_h, w_h, \mu_h) \in \boldsymbol{V}_h \times W_h \times M_h(o)$, we have*

$$
\begin{aligned}
&\mathscr{B}_1(\boldsymbol{v}_h, w_h, \mu_h; \boldsymbol{v}_h, w_h, \mu_h) \\
&\quad = (\boldsymbol{v}_h, \boldsymbol{v}_h)_{\mathcal{T}_h} + \langle (h^{-1} + \tau_1 - \frac{1}{2}\boldsymbol{\beta} \cdot \boldsymbol{n})(w_h - \mu_h), w_h - \mu_h \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} \\
&\qquad - \frac{1}{2}(\nabla \cdot \boldsymbol{\beta} w_h, w_h)_{\mathcal{T}_h} + \langle (h^{-1} + \tau_1 - \frac{1}{2}\boldsymbol{\beta} \cdot \boldsymbol{n}) w_h, w_h \rangle_{\varepsilon_h^\partial}, \\
&\mathscr{B}_2(\boldsymbol{v}_h, w_h, \mu_h; \boldsymbol{v}_h, w_h, \mu_h) \\
&\quad = (\boldsymbol{v}_h, \boldsymbol{v}_h)_{\mathcal{T}_h} + \langle (h^{-1} + \tau_2 + \frac{1}{2}\boldsymbol{\beta} \cdot \boldsymbol{n})(w_h - \mu_h), w_h - \mu_h \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} \\
&\qquad - \frac{1}{2}(\nabla \cdot \boldsymbol{\beta} w_h, w_h)_{\mathcal{T}_h} + \langle (h^{-1} + \tau_2 + \frac{1}{2}\boldsymbol{\beta} \cdot \boldsymbol{n}) w_h, w_h \rangle_{\varepsilon_h^\partial}.
\end{aligned}
$$

**Proof:** We only prove the first identity; the second can be obtained by the same argument.

$$
\begin{aligned}
&\mathscr{B}_1(\boldsymbol{v}_h, w_h, \mu_h; \boldsymbol{v}_h, w_h, \mu_h)\\
&\qquad = (\boldsymbol{v}_h, \boldsymbol{v}_h)_{\mathcal{T}_h} - (w_h, \nabla\cdot\boldsymbol{v}_h)_{\mathcal{T}_h} + \langle \mu_h, \boldsymbol{v}_h\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\qquad\quad - (\boldsymbol{v}_h + \boldsymbol{\beta} w_h, \nabla w_h)_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta} w_h, w_h)_{\mathcal{T}_h}\\
&\qquad\quad + \langle \boldsymbol{v}_h\cdot\boldsymbol{n} + (h^{-1}+\tau_1) w_h, w_h\rangle_{\partial\mathcal{T}_h} + \langle(\boldsymbol{\beta}\cdot\boldsymbol{n} - h^{-1} - \tau_1)\mu_h, w_h\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\qquad\quad - \langle \boldsymbol{v}_h\cdot\boldsymbol{n} + \boldsymbol{\beta}\cdot\boldsymbol{n}\mu_h + (h^{-1}+\tau_1)(w_h-\mu_h), \mu_h\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\qquad = (\boldsymbol{v}_h, \boldsymbol{v}_h)_{\mathcal{T}_h} - (\boldsymbol{\beta} w_h, \nabla w_h)_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta} w_h, w_h)_{\mathcal{T}_h}\\
&\qquad\quad + \langle (h^{-1}+\tau_1) w_h, w_h\rangle_{\partial\mathcal{T}_h} + \langle(\boldsymbol{\beta}\cdot\boldsymbol{n} - h^{-1} - \tau_1)\mu_h, w_h\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\qquad\quad - \langle \boldsymbol{\beta}\cdot\boldsymbol{n}\mu_h + (h^{-1}+\tau_1)(w_h-\mu_h), \mu_h\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}.
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
(\boldsymbol{\beta} w_h, \nabla w_h)_{\mathcal{T}_h} &= (\boldsymbol{\beta}\cdot\nabla w_h, w_h)_{\mathcal{T}_h} = (\nabla\cdot(\boldsymbol{\beta} w_h), w_h)_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta} w_h, w_h)_{\mathcal{T}_h}\\
&= \langle \boldsymbol{\beta}\cdot\boldsymbol{n} w_h, w_h\rangle_{\partial\mathcal{T}_h} - (\boldsymbol{\beta} w_h, \nabla w_h)_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta} w_h, w_h)_{\mathcal{T}_h},
\end{aligned}
$$

which implies

$$
(\boldsymbol{\beta} w_h, \nabla w_h)_{\mathcal{T}_h} = \frac{1}{2}\langle \boldsymbol{\beta}\cdot\boldsymbol{n} w_h, w_h\rangle_{\partial\mathcal{T}_h} - \frac{1}{2}(\nabla\cdot\boldsymbol{\beta} w_h, w_h)_{\mathcal{T}_h}.
$$

Then we obtain

$$
\begin{aligned}
&\mathscr{B}_1(\boldsymbol{v}_h, w_h, \mu_h; \boldsymbol{v}_h, w_h, \mu_h)\\
&\quad = (\boldsymbol{v}_h, \boldsymbol{v}_h)_{\mathcal{T}_h} + \langle (h^{-1} + \tau_1 - \frac{1}{2}\boldsymbol{\beta}\cdot\boldsymbol{n})(w_h - \mu_h), w_h - \mu_h\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\quad\quad - \frac{1}{2}(\nabla\cdot\boldsymbol{\beta} w_h, w_h)_{\mathcal{T}_h} + \langle (h^{-1} + \tau_1 - \frac{1}{2}\boldsymbol{\beta}\cdot\boldsymbol{n}) w_h, w_h\rangle_{\varepsilon_h^\partial} - \frac{1}{2}\langle \boldsymbol{\beta}\cdot\boldsymbol{n}\mu_h, \mu_h\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}.
\end{aligned}
$$

Since $\mu_h$ is single-valued across the interfaces, we have

$$-\frac{1}{2}\langle \boldsymbol{\beta}\cdot\boldsymbol{n}\mu_h, \mu_h\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} = 0.$$

This completes the proof.

Next, we give a property of the HDG operators $\mathscr{B}_1$ and $\mathscr{B}_2$ that is critical to our error analysis of the method.

**Lemma 17** *If (A2) holds, then*

$$\mathscr{B}_1(\boldsymbol{q}_h, y_h, \widehat{y}_h^o; \boldsymbol{p}_h, -z_h, -\widehat{z}_h^o) + \mathscr{B}_2(\boldsymbol{p}_h, z_h, \widehat{z}_h^o; -\boldsymbol{q}_h, y_h, \widehat{y}_h^o) = 0.$$

**Proof:** By the definition of $\mathscr{B}_1$ and $\mathscr{B}_2$,

$$\begin{aligned}
&\mathscr{B}_1(\boldsymbol{q}_h, y_h, \widehat{y}_h^o; \boldsymbol{p}_h, -z_h, -\widehat{z}_h^o) + \mathscr{B}_2(\boldsymbol{p}_h, z_h, \widehat{z}_h^o; -\boldsymbol{q}_h, y_h, \widehat{y}_h^o)\\
&\quad= (\boldsymbol{q}_h, \boldsymbol{p}_h)_{\mathcal{T}_h} - (y_h, \nabla\cdot\boldsymbol{p}_h)_{\mathcal{T}_h} + \langle \widehat{y}_h^o, \boldsymbol{p}_h\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\qquad+ (\boldsymbol{q}_h + \boldsymbol{\beta} y_h, \nabla z_h)_{\mathcal{T}_h} + (\nabla\cdot\boldsymbol{\beta} y_h, z_h)_{\mathcal{T}_h} - \langle \boldsymbol{q}_h\cdot\boldsymbol{n} + (h^{-1}+\tau_1) y_h, z_h\rangle_{\partial\mathcal{T}_h}\\
&\qquad- \langle (\boldsymbol{\beta}\cdot\boldsymbol{n} - \tau_1 - h^{-1})\widehat{y}_h^o, z_h\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\qquad+ \langle \boldsymbol{q}_h\cdot\boldsymbol{n} + \boldsymbol{\beta}\cdot\boldsymbol{n}\widehat{y}_h^o + (h^{-1}+\tau_1)(y_h - \widehat{y}_h^o), \widehat{z}_h^o\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\qquad- (\boldsymbol{p}_h, \boldsymbol{q}_h)_{\mathcal{T}_h} + (z_h, \nabla\cdot\boldsymbol{q}_h)_{\mathcal{T}_h} - \langle \widehat{z}_h^o, \boldsymbol{q}_h\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} - (\boldsymbol{p}_h - \boldsymbol{\beta} z_h, \nabla y_h)_{\mathcal{T}_h}\\
&\qquad+ \langle \boldsymbol{p}_h\cdot\boldsymbol{n} + (h^{-1}+\tau_2) z_h, y_h\rangle_{\partial\mathcal{T}_h} - \langle (\boldsymbol{\beta}\cdot\boldsymbol{n} + \tau_2 + h^{-1})\widehat{z}_h^o, y_h\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\qquad- \langle \boldsymbol{p}_h\cdot\boldsymbol{n} - \boldsymbol{\beta}\cdot\boldsymbol{n}\widehat{z}_h^o + (h^{-1}+\tau_2)(z_h - \widehat{z}_h^o), \widehat{y}_h^o\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}.
\end{aligned}$$

Integration by parts gives

$$\begin{aligned}
&\mathscr{B}_1(\boldsymbol{q}_h, y_h, \widehat{y}_h^o; \boldsymbol{p}_h, -z_h, -\widehat{z}_h^o) + \mathscr{B}_2(\boldsymbol{p}_h, z_h, \widehat{z}_h^o; -\boldsymbol{q}_h, y_h, \widehat{y}_h^o)\\
&\quad= \langle (\tau_2 + \boldsymbol{\beta}\cdot\boldsymbol{n} - \tau_1) y_h, z_h\rangle_{\partial\mathcal{T}_h} + \langle (\tau_2 + \boldsymbol{\beta}\cdot\boldsymbol{n} - \tau_1)\widehat{y}_h^o, \widehat{z}_h^o\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}.
\end{aligned}$$

The proof is complete by assumption **(A2)**.

**Proposition 3.4.1** *If (A2) holds, there exists a unique solution of the HDG equations* (3.37).

**Proof:** Since the system (3.37) is finite dimensional, we only need to prove the uniqueness. Therefore, we assume $y_d = f = 0$ and we show the system (3.37) only has the trivial solution.

First, take $(\boldsymbol{r}_1, w_1, \mu_1) = (\boldsymbol{p}_h, -z_h, -\widehat{z}_h^o)$, $(\boldsymbol{r}_2, w_2, \mu_2) = (-\boldsymbol{q}_h, y_h, \widehat{y}_h^o)$, and $w_3 = -\gamma u_h$ in the HDG equations (3.37a), (3.37b), and (3.37c), respectively, by 17 and sum to obtain

$$\begin{aligned}
&\mathscr{B}_1(\boldsymbol{q}_h, y_h, \widehat{y}_h^o; \boldsymbol{p}_h, -z_h, -\widehat{z}_h^o) + \mathscr{B}_2(\boldsymbol{p}_h, z_h, \widehat{z}_h^o; -\boldsymbol{q}_h, y_h, \widehat{y}_h^o)\\
&\quad = (y_h, y_h)_{\mathcal{T}_h} + \gamma \|u_h\|^2_{\varepsilon_h^\partial}\\
&\quad = 0.
\end{aligned}$$

This implies $y_h = u_h = 0$ since $\gamma > 0$.

Next, taking $(\boldsymbol{r}_1, w_1, \mu_1) = (\boldsymbol{q}_h, y_h, \widehat{y}_h^o)$ and $(\boldsymbol{r}_2, w_2, \mu_2) = (\boldsymbol{p}_h, z_h, \widehat{z}_h^o)$ in Lemma 30 gives $\boldsymbol{q}_h = \boldsymbol{p}_h = \boldsymbol{0}$, $\widehat{y}_h^o = 0$, $z_h = 0$ on $\varepsilon_h^\partial$, and $z_h - \widehat{z}_h^o = 0$ on $\partial\mathcal{T}_h \backslash \varepsilon_h^\partial$. Also, since $\widehat{z}_h = 0$ on $\varepsilon_h^\partial$ we have

$$z_h - \widehat{z}_h = 0. \tag{3.38}$$

Substituting (3.38) into (3.22c), and remembering again $\widehat{z}_h = 0$ on $\varepsilon_h^\partial$, we get

$$-(z_h, \nabla \cdot \boldsymbol{r_2})_{\mathcal{T}_h} + \langle z_h, \boldsymbol{r_2} \cdot \boldsymbol{n} \rangle_{\partial \mathcal{T}_h} = 0.$$

Integrate by parts, and take $\boldsymbol{r}_2 = \nabla z_h$ to obtain

$$(\nabla z_h, \nabla z_h)_{\mathcal{T}_h} = 0.$$

Thus, $z_h$ is constant on each $K \in \mathcal{T}_h$, and also $z_h = \widehat{z}_h$ on $\partial\mathcal{T}_h$. Since $\widehat{z}_h = 0$ on $\varepsilon_h^\partial$ and single valued on each face, we have $z_h = 0$ on each $K \in \mathcal{T}_h$, and therefore also $\widehat{z}_h^o = 0$.

**3.4.3. Proof of Main Result.** To prove the main result, we follow the strategy of Section one and split the proof into seven steps. We consider the following auxiliary problem: find

$$(\boldsymbol{q}_h(u), \boldsymbol{p}_h(u), y_h(u), z_h(u), \widehat{y}_h^o(u), \widehat{z}_h^o(u)) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times W_h \times M_h(o) \times M_h(o)$$

such that

$$\begin{aligned}\mathscr{B}_1(\boldsymbol{q}_h(u), y_h(u), \widehat{y}_h(u); \boldsymbol{r}_1, w_1, \mu_1) = (f, w_1)_{\mathcal{T}_h} - \langle P_M u, \boldsymbol{r}_1 \cdot \boldsymbol{n}\rangle_{\varepsilon_h^\partial} \\ - \langle(\boldsymbol{\beta} \cdot \boldsymbol{n} - h^{-1} - \tau_1) P_M u, w_1\rangle_{\varepsilon_h^\partial}, \qquad (3.39a)\end{aligned}$$

$$\mathscr{B}_2(\boldsymbol{p}_h(u), z_h(u), \widehat{z}_h(u); \boldsymbol{r}_2, w_2, \mu_2) = (y_h(u) - y_d, w_2)_{\mathcal{T}_h}, \qquad (3.39b)$$

for all $(\boldsymbol{r}_1, \boldsymbol{r}_2, w_1, w_2, \mu_1, \mu_2) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times W_h \times M_h(o) \times M_h(o)$. We first bound the error between the solutions of the auxiliary problem and the mixed form (3.18a) - (3.18d) of the optimality system. We use the following notation:

$$\begin{aligned}\delta^{\boldsymbol{q}} &= \boldsymbol{q} - \boldsymbol{\Pi}\boldsymbol{q}, & \varepsilon_h^{\boldsymbol{q}} &= \boldsymbol{\Pi}\boldsymbol{q} - \boldsymbol{q}_h(u),\\ \delta^y &= y - \Pi y, & \varepsilon_h^y &= \Pi y - y_h(u),\\ \delta^{\widehat{y}} &= y - P_M y, & \varepsilon_h^{\widehat{y}} &= P_M y - \widehat{y}_h(u),\\ \widehat{\boldsymbol{\delta}}_1 &= \delta^{\boldsymbol{q}} \cdot \boldsymbol{n} + \boldsymbol{\beta} \cdot \boldsymbol{n}\delta^{\widehat{y}} + (h^{-1} + \tau_1)(\delta^y - \delta^{\widehat{y}}),\end{aligned} \qquad (3.40)$$

where $\widehat{y}_h(u) = \widehat{y}_h^o(u)$ on $\varepsilon_h^o$ and $\widehat{y}_h(u) = P_M u$ on $\varepsilon_h^\partial$. Note that this implies $\varepsilon_h^{\widehat{y}} = 0$ on $\varepsilon_h^\partial$.

Step 1: The error equation for part 1 of the auxiliary problem (3.39a)

**Lemma 18** *We have*

$$\begin{aligned}\mathscr{B}_1(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^{y}, \varepsilon_h^{\widehat{y}}, \boldsymbol{r}_1, w_1, \mu_1) &= (\boldsymbol{\beta}\delta^y, \nabla w_1)_{\mathcal{T}_h} + (\nabla\cdot\boldsymbol{\beta}\delta^y, w_1)_{\mathcal{T}_h} \\ &\quad - \langle \widehat{\boldsymbol{\delta}}_1, w_1\rangle_{\partial\mathcal{T}_h} + \langle \widehat{\boldsymbol{\delta}}_1, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}. \end{aligned} \tag{3.41}$$

**Proof:** By the definition of the operator $\mathscr{B}_1$ in (3.35), we have

$$\begin{aligned}&\mathscr{B}_1(\boldsymbol{\Pi q}, \Pi y, P_M y, \boldsymbol{r}_1, w_1, \mu_1) \\ &= (\boldsymbol{\Pi q}, \boldsymbol{r_1})_{\mathcal{T}_h} - (\Pi y, \nabla\cdot\boldsymbol{r_1})_{\mathcal{T}_h} + \langle P_M y, \boldsymbol{r_1}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} \\ &\quad - (\boldsymbol{\Pi q} + \boldsymbol{\beta}\Pi y, \nabla w_1)_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta}\Pi y, w_1)_{\mathcal{T}_h} \\ &\quad + \langle \boldsymbol{\Pi q}\cdot\boldsymbol{n} + (h^{-1}+\tau_1)\Pi y, w_1\rangle_{\partial\mathcal{T}_h} + (\boldsymbol{\beta}\cdot\boldsymbol{n} - h^{-1} - \tau_1)P_M y, w_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} \\ &\quad - \langle \boldsymbol{\Pi q}\cdot\boldsymbol{n} + \boldsymbol{\beta}\cdot\boldsymbol{n}P_M y + (h^{-1}+\tau_1)(\Pi y - P_M y), \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}. \end{aligned}$$

By properties of the $L^2$ projections (3.33), we have

$$\begin{aligned}&\mathscr{B}_1(\boldsymbol{\Pi q}, \Pi y, P_M y, \boldsymbol{r}_1, w_1, \mu_1) \\ &= (\boldsymbol{q}, \boldsymbol{r_1})_{\mathcal{T}_h} - (y, \nabla\cdot\boldsymbol{r_1})_{\mathcal{T}_h} + \langle y, \boldsymbol{r_1}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} \\ &\quad - (\boldsymbol{q} + \boldsymbol{\beta}y, \nabla w_1)_{\mathcal{T}_h} + (\boldsymbol{\beta}\delta^y, \nabla w_1)_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta}y, w_1)_{\mathcal{T}_h} + (\nabla\cdot\boldsymbol{\beta}\delta^y, w_1)_{\mathcal{T}_h} \\ &\quad + \langle \boldsymbol{q}\cdot\boldsymbol{n}, w_1\rangle_{\partial\mathcal{T}_h} - \langle \delta^{\boldsymbol{q}}\cdot\boldsymbol{n}, w_1\rangle_{\partial\mathcal{T}_h} + \langle (h^{-1}+\tau_1)\Pi y, w_1\rangle_{\partial\mathcal{T}_h} \\ &\quad + \langle \boldsymbol{\beta}\cdot\boldsymbol{n}y, w_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} - \langle \boldsymbol{\beta}\cdot\boldsymbol{n}\delta^{\widehat{y}}, w_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} - \langle (h^{-1}+\tau_1)P_M y, w_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} \\ &\quad - \langle \boldsymbol{q}\cdot\boldsymbol{n}, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} + \langle \delta^{\boldsymbol{q}}\cdot\boldsymbol{n}, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} - \langle \boldsymbol{\beta}\cdot\boldsymbol{n}y, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} \\ &\quad + \langle \boldsymbol{\beta}\cdot\boldsymbol{n}\delta^{\widehat{y}}, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} + \langle (h^{-1}+\tau_1)(\delta^y - \delta^{\widehat{y}}), \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}. \end{aligned}$$

Note that the exact state $y$ and exact flux $\boldsymbol{q}$ satisfy

$$
\begin{aligned}
(\boldsymbol{q}, \boldsymbol{r}_1)_{\mathcal{T}_h} - (y, \nabla \cdot \boldsymbol{r}_1)_{\mathcal{T}_h} + \langle y, \boldsymbol{r}_1 \cdot \boldsymbol{n} \rangle_{\partial \mathcal{T}_h} &= 0, \\
-(\boldsymbol{q} + \boldsymbol{\beta} y, \nabla w_1)_{\mathcal{T}_h} - (\nabla \cdot \boldsymbol{\beta} y, w_1)_{\mathcal{T}_h} + \langle (\boldsymbol{q} + \boldsymbol{\beta} y) \cdot \boldsymbol{n}, w_1 \rangle_{\partial \mathcal{T}_h} &= (f, w_1)_{\mathcal{T}_h}, \\
\langle (\boldsymbol{q} + \boldsymbol{\beta} y) \cdot \boldsymbol{n}, \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} &= 0,
\end{aligned}
$$

for all $(\boldsymbol{r}_1, w_1, \mu_1) \in \boldsymbol{V}_h \times W_h \times M_h(o)$. Then we have

$$
\begin{aligned}
&\mathscr{B}_1(\boldsymbol{\Pi q}, \Pi y, P_M y, \boldsymbol{r}_1, w_1, \mu_1) \\
&\quad = -\langle u, \boldsymbol{r}_1 \cdot \boldsymbol{n} \rangle_{\varepsilon_h^\partial} - \langle \boldsymbol{\beta} \cdot \boldsymbol{n} u, w_1 \rangle_{\varepsilon_h^\partial} + (f, w_1)_{\mathcal{T}_h} + (\boldsymbol{\beta} \delta^y, \nabla w_1)_{\mathcal{T}_h} \\
&\qquad + (\nabla \cdot \boldsymbol{\beta} \delta^y, w_1)_{\mathcal{T}_h} - \langle \delta^{\boldsymbol{q}} \cdot \boldsymbol{n}, w_1 \rangle_{\partial \mathcal{T}_h} + \langle (h^{-1} + \tau_1) \Pi y, w_1 \rangle_{\partial \mathcal{T}_h} \\
&\qquad - \langle \boldsymbol{\beta} \cdot \boldsymbol{n} \delta^{\widehat{y}}, w_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} - \langle (h^{-1} + \tau_1) P_M y, w_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} + \langle \delta^{\boldsymbol{q}} \cdot \boldsymbol{n}, \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} \\
&\qquad + \langle \boldsymbol{\beta} \cdot \boldsymbol{n} \delta^{\widehat{y}}, \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} + \langle (h^{-1} + \tau_1)(\delta^y - \delta^{\widehat{y}}), \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial}.
\end{aligned}
$$

Subtract part 1 of the auxiliary problem (3.39a) from the above equality to obtain the result:

$$
\begin{aligned}
&\mathscr{B}_1(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^y, \varepsilon_h^{\widehat{y}}, \boldsymbol{r}_1, w_1, \mu_1) \\
&\quad = (\boldsymbol{\beta} \delta^y, \nabla w_1)_{\mathcal{T}_h} + (\nabla \cdot \boldsymbol{\beta} \delta^y, w_1)_{\mathcal{T}_h} - \langle \delta^{\boldsymbol{q}} \cdot \boldsymbol{n}, w_1 \rangle_{\partial \mathcal{T}_h} \\
&\qquad + \langle (h^{-1} + \tau_1) \Pi y, w_1 \rangle_{\partial \mathcal{T}_h} - \langle \boldsymbol{\beta} \cdot \boldsymbol{n} \delta^{\widehat{y}}, w_1 \rangle_{\partial \mathcal{T}_h} - \langle (h^{-1} + \tau_1) P_M y, w_1 \rangle_{\partial \mathcal{T}_h} \\
&\qquad + \langle \delta^{\boldsymbol{q}} \cdot \boldsymbol{n}, \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} + \langle \boldsymbol{\beta} \cdot \boldsymbol{n} \delta^{\widehat{y}}, \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} + \langle (h^{-1} + \tau_1)(\delta^y - \delta^{\widehat{y}}), \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial} \\
&\quad = (\boldsymbol{\beta} \delta^y, \nabla w_1)_{\mathcal{T}_h} + (\nabla \cdot \boldsymbol{\beta} \delta^y, w_1)_{\mathcal{T}_h} - \langle \widehat{\boldsymbol{\delta}}_1, w_1 \rangle_{\partial \mathcal{T}_h} + \langle \widehat{\boldsymbol{\delta}}_1, \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial}.
\end{aligned}
$$

Step 2: Estimate for $\varepsilon_h^{\boldsymbol{q}}$ We begin with a key inequality that has been proved for an existing HDG method in [56].

**Lemma 19** *We have*

$$
\|\nabla \varepsilon_h^y\|_{\mathcal{T}_h} \le \|\varepsilon_h^{\boldsymbol{q}}\|_{\mathcal{T}_h} + C h^{-\frac{1}{2}} \|\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial \mathcal{T}_h}.
$$

In the HDG method in [56], degree $k$ polynomials are used for the space $M_h$ instead of degree $k+1$ here. Increasing this degree does not lead to any change in the proof of the above lemma; therefore, we omit the proof.

**Lemma 20** *We have*

$$\left\|\varepsilon_h^{\boldsymbol{q}}\right\|_{\mathcal{T}_h}^2 + h^{-1}\|\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h}^2 \lesssim h^{2s_{\boldsymbol{q}}}\,\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega}^2 + h^{2s_y-2}\,\|y\|_{s^y,\Omega}^2\,. \tag{3.42}$$

**Proof:** First, since $\varepsilon_h^{\widehat{y}} = 0$ on $\varepsilon_h^{\partial}$, the basic property of $\mathscr{B}_1$ in Lemma 16 gives

$$\begin{aligned}
&\mathscr{B}(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^y, \varepsilon_h^{\widehat{y}}, \varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^y, \varepsilon_h^{\widehat{y}})\\
&\quad = (\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^{\boldsymbol{q}})_{\mathcal{T}_h} + \|(h^{-1} + \tau_1 - \frac{1}{2}\boldsymbol{\beta}\cdot\boldsymbol{n})^{\frac{1}{2}}(\varepsilon_h^y - \varepsilon_h^{\widehat{y}})\|_{\partial\mathcal{T}_h}^2 + \frac{1}{2}\|(-\nabla\cdot\boldsymbol{\beta})^{\frac{1}{2}}\varepsilon_h^y\|_{\mathcal{T}_h}^2.
\end{aligned}$$

Then, taking $(\boldsymbol{r}_1, w_1, \mu_1) = (\boldsymbol{\varepsilon}_h^{\boldsymbol{q}}, \varepsilon_h^y, \varepsilon_h^{\widehat{y}})$ in (3.41) in Lemma 18 gives

$$\begin{aligned}
(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^{\boldsymbol{q}})_{\mathcal{T}_h} &+ \|(h^{-1} + \tau_1 - \frac{1}{2}\boldsymbol{\beta}\cdot\boldsymbol{n})^{\frac{1}{2}}(\varepsilon_h^y - \varepsilon_h^{\widehat{y}})\|_{\partial\mathcal{T}_h}^2 + \frac{1}{2}\|(-\nabla\cdot\boldsymbol{\beta})^{\frac{1}{2}}\varepsilon_h^y\|_{\mathcal{T}_h}^2\\
&= (\boldsymbol{\beta}\delta^y, \nabla\varepsilon_h^y)_{\mathcal{T}_h} + (\nabla\cdot\boldsymbol{\beta}\delta^y, \varepsilon_h^y)_{\mathcal{T}_h} - \langle\widehat{\boldsymbol{\delta}}_1, \varepsilon_h^y - \varepsilon_h^{\widehat{y}}\rangle_{\partial\mathcal{T}_h}\\
&=: T_1 + T_2 + T_3.
\end{aligned} \tag{3.43}$$

For the terms $T_1$ and $T_2$, simply applying Lemma 19 and Young's inequality gives

$$\begin{aligned}
T_1 &= (\boldsymbol{\beta}\delta^y, \nabla\varepsilon_h^y)_{\mathcal{T}_h} \le C\|\delta^y\|_{\mathcal{T}_h}^2 + \frac{1}{4}\|\varepsilon_h^{\boldsymbol{q}}\|_{\mathcal{T}_h}^2 + \frac{1}{4h}\|\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h}^2,\\
T_2 &= (\nabla\cdot\boldsymbol{\beta}\delta^y, \varepsilon_h^y)_{\mathcal{T}_h} \le C\|\delta^y\|_{\mathcal{T}_h}^2 + \frac{1}{2}\|(-\nabla\cdot\boldsymbol{\beta})^{\frac{1}{2}}\varepsilon_h^y\|_{\mathcal{T}_h}^2,\\
T_3 &= -\langle\widehat{\boldsymbol{\delta}}_1, \varepsilon_h^y - \varepsilon_h^{\widehat{y}}\rangle_{\partial\mathcal{T}_h} \le 4h\|\boldsymbol{\delta}_1\|_{\partial\mathcal{T}_h}^2 + \frac{1}{4h}\|\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h}^2.
\end{aligned}$$

Sum all the estimates for $\{T_i\}_{i=1}^3$ to obtain

$$\begin{aligned}
\|\varepsilon_h^{\boldsymbol{q}}\|_{\mathcal{T}_h}^2 + h^{-1}\|\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h}^2 &\lesssim h\|\boldsymbol{\delta}_1\|_{\partial\mathcal{T}_h}^2 + \|\delta^y\|_{\mathcal{T}_h}^2\\
&\lesssim h^{2s_{\boldsymbol{q}}}\,\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega}^2 + h^{2s_y-2}\,\|y\|_{s^y,\Omega}^2\,.
\end{aligned}$$

Step 3: Estimate for $\varepsilon_h^y$ by a duality argument.

Next, we introduce the dual problem for any given $\Theta$ in $L^2(\Omega)$ :

$$\begin{aligned} \boldsymbol{\Phi} - \nabla\Psi &= 0 \qquad \text{in } \Omega, \\ \nabla\cdot\boldsymbol{\Phi} + \nabla\cdot(\boldsymbol{\beta}\Psi) &= \Theta \qquad \text{in } \Omega, \\ \Psi &= 0 \qquad \text{on } \partial\Omega. \end{aligned} \tag{3.44}$$

Since the domain $\Omega$ is convex, we have the regularity estimate

$$\|\boldsymbol{\Phi}\|_{1,\Omega} + \|\Psi\|_{2,\Omega} \leq C_{\text{reg}} \|\Theta\|_{\Omega}\,. \tag{3.45}$$

Before we estimate $\varepsilon_h^y$, we introduce the following notation, which is similar to the earlier notation in (3.40):

$$\delta^{\boldsymbol{\Phi}} = \boldsymbol{\Phi} - \boldsymbol{\Pi\Phi}, \quad \delta^{\Psi} = \Psi - \Pi\Psi, \quad \delta^{\widehat{\Psi}} = \Psi - P_M\Psi. \tag{3.46}$$

**Lemma 21** *We have*

$$\|\varepsilon_h^y\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{q}}+1}\,\|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y}\,\|y\|_{s^y,\Omega}\,.$$

**Proof:** Consider the dual problem (3.44) and let $\Theta = -\varepsilon_h^y$. Take $(\boldsymbol{r}_1, w_1, \mu_1) = (\boldsymbol{\Pi\Phi}, \Pi\Psi, P_M\Psi)$ in (3.41) in Lemma 18, and since $\Psi = 0$ on $\varepsilon_h^\partial$, we have

$$
\begin{aligned}
&\mathscr{B}_1(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^y, \varepsilon_h^{\widehat{y}}; \boldsymbol{\Pi\Phi}, \Pi\Psi, P_M\Psi) \\
&\quad = (\varepsilon_h^{\boldsymbol{q}}, \boldsymbol{\Pi\Phi})_{\mathcal{T}_h} - (\varepsilon_h^y, \nabla\cdot\boldsymbol{\Pi\Phi})_{\mathcal{T}_h} + \langle \varepsilon_h^{\widehat{y}}, \boldsymbol{\Pi\Phi}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} \\
&\qquad - (\varepsilon_h^{\boldsymbol{q}} + \boldsymbol{\beta}\varepsilon_h^y, \nabla\Pi\Psi)_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta}\varepsilon_h^y, \Pi\Psi)_{\mathcal{T}_h} + \langle \varepsilon_h^{\boldsymbol{q}}\cdot\boldsymbol{n} + (h^{-1}+\tau_1)\varepsilon_h^y, \Pi\Psi\rangle_{\partial\mathcal{T}_h} \\
&\qquad + \langle (\boldsymbol{\beta}\cdot\boldsymbol{n} - h^{-1} - \tau_1)\varepsilon_h^{\widehat{y}}, \Pi\Psi\rangle_{\partial\mathcal{T}_h} \\
&\qquad - \langle \varepsilon_h^{\boldsymbol{q}}\cdot\boldsymbol{n} + \boldsymbol{\beta}\cdot\boldsymbol{n}\varepsilon_h^{\widehat{y}} + (h^{-1}+\tau_1)(\varepsilon_h^y - \varepsilon_h^{\widehat{y}}), P_M\Psi\rangle_{\partial\mathcal{T}_h} \\
&\quad = (\varepsilon_h^{\boldsymbol{q}}, \boldsymbol{\Phi})_{\mathcal{T}_h} - (\varepsilon_h^y, \nabla\cdot\boldsymbol{\Phi})_{\mathcal{T}_h} + (\varepsilon_h^y, \nabla\cdot\delta^{\boldsymbol{\Phi}})_{\mathcal{T}_h} - \langle \varepsilon_h^{\widehat{y}}, \delta^{\boldsymbol{\Phi}}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h} - (\varepsilon_h^{\boldsymbol{q}} + \boldsymbol{\beta}\varepsilon_h^y, \nabla\Psi)_{\mathcal{T}_h} \\
&\qquad + (\varepsilon_h^{\boldsymbol{q}} + \boldsymbol{\beta}\varepsilon_h^y, \nabla\delta^\Psi)_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta}\varepsilon_h^y, \Psi)_{\mathcal{T}_h} + (\nabla\cdot\boldsymbol{\beta}\varepsilon_h^y, \delta^\Psi)_{\mathcal{T}_h} \\
&\qquad - \langle \varepsilon_h^{\boldsymbol{q}}\cdot\boldsymbol{n} + \boldsymbol{\beta}\cdot\boldsymbol{n}\varepsilon_h^{\widehat{y}} + (h^{-1}+\tau_1)(\varepsilon_h^y - \varepsilon_h^{\widehat{y}}), \delta^\Psi - \delta^{\widehat{\Psi}}\rangle_{\partial\mathcal{T}_h}.
\end{aligned}
$$

Here we used $\langle \varepsilon_h^{\widehat{y}}, \boldsymbol{\Phi}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h} = 0$, which holds since $\varepsilon_h^{\widehat{y}}$ is single-valued function on interior edges and $\varepsilon_h^{\widehat{y}} = 0$ on $\varepsilon_h^\partial$.

Next, integration by parts gives

$$
\begin{aligned}
(\varepsilon_h^y, \nabla\cdot\delta^{\boldsymbol{\Phi}})_{\mathcal{T}_h} &= \langle \varepsilon_h^y, \delta^{\boldsymbol{\Phi}}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h} - (\nabla\varepsilon_h^y, \delta^{\boldsymbol{\Phi}})_{\mathcal{T}_h} = \langle \varepsilon_h^y, \delta^{\boldsymbol{\Phi}}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h}, \\
(\varepsilon_h^{\boldsymbol{q}}, \nabla\delta^\Psi)_{\mathcal{T}_h} &= \langle \varepsilon_h^{\boldsymbol{q}}\cdot\boldsymbol{n}, \delta^\Psi\rangle_{\partial\mathcal{T}_h} - (\nabla\cdot\varepsilon_h^{\boldsymbol{q}}, \delta^\Psi)_{\mathcal{T}_h} = \langle \varepsilon_h^{\boldsymbol{q}}\cdot\boldsymbol{n}, \delta^\Psi\rangle_{\partial\mathcal{T}_h}, \\
(\boldsymbol{\beta}\varepsilon_h^y, \nabla\delta^\Psi)_{\mathcal{T}_h} &= \langle \boldsymbol{\beta}\cdot\boldsymbol{n}\varepsilon_h^y, \delta^\Psi\rangle_{\partial\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta}\varepsilon_h^y, \delta^\Psi)_{\mathcal{T}_h} - (\boldsymbol{\beta}\nabla\varepsilon_h^y, \delta^\Psi)_{\mathcal{T}_h}.
\end{aligned}
\tag{3.47}
$$

We have

$$
\begin{aligned}
&\mathscr{B}_1(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^y, \varepsilon_h^{\widehat{y}}; \boldsymbol{\Pi\Phi}, \Pi\Psi, P_M\Psi) \\
&\qquad = \|\varepsilon_h^y\|_{\mathcal{T}_h}^2 + \langle \varepsilon_h^y - \varepsilon_h^{\widehat{y}}, \delta^{\boldsymbol{\Phi}}\cdot\boldsymbol{n} + \boldsymbol{\beta}\cdot\boldsymbol{n}\delta^\Psi\rangle_{\partial\mathcal{T}_h} - (\nabla\varepsilon_h^y, \boldsymbol{\beta}\delta^\Psi)_{\mathcal{T}_h} \\
&\qquad\quad - \langle (h^{-1}+\tau_1)(\varepsilon_h^y - \varepsilon_h^{\widehat{y}}), \delta^\Psi - \delta^{\widehat{\Psi}}\rangle_{\partial\mathcal{T}_h}.
\end{aligned}
$$

On the other hand, $\Psi = 0$ on $\varepsilon_h^{\partial}$ and (3.41) in 18 give

$$
\begin{aligned}
&\mathscr{B}_1(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^{y}, \varepsilon_h^{\widehat{y}}; \boldsymbol{\Pi\Phi}, \Pi\Psi, P_M\Psi)\\
&\qquad = (\boldsymbol{\beta}\delta^{y}, \nabla\Pi\Psi)_{\mathcal{T}_h} + (\nabla\cdot\boldsymbol{\beta}\delta^{y}, \Pi\Psi)_{\mathcal{T}_h} + \langle \widehat{\boldsymbol{\delta}}_1, \delta^{\Psi} - \delta^{\widehat{\Psi}}\rangle_{\partial\mathcal{T}_h}.
\end{aligned}
$$

Comparing the above two equalities, we get

$$
\begin{aligned}
\|\varepsilon_h^{y}\|_{\mathcal{T}_h}^2 &= -\langle \varepsilon_h^{y} - \varepsilon_h^{\widehat{y}}, \delta^{\boldsymbol{\Phi}}\cdot\boldsymbol{n} + \boldsymbol{\beta}\cdot\boldsymbol{n}\delta^{\Psi}\rangle_{\partial\mathcal{T}_h}\\
&\quad + (\nabla\varepsilon_h^{y}, \boldsymbol{\beta}\delta^{\Psi})_{\mathcal{T}_h} + (\boldsymbol{\beta}\delta^{y}, \nabla\Pi\Psi)_{\mathcal{T}_h} + (\nabla\cdot\boldsymbol{\beta}\delta^{y}, \Pi\Psi)_{\mathcal{T}_h}\\
&\quad + \langle (h^{-1} + \tau_1)(\varepsilon_h^{y} - \varepsilon_h^{\widehat{y}})\widehat{\boldsymbol{\delta}}_1, \delta^{\Psi} - \delta^{\widehat{\Psi}}\rangle_{\partial\mathcal{T}_h}\\
&=: R_1 + R_2 + R_3 + R_4 + R_5.
\end{aligned}
$$

For the terms $R_1$ and $R_2$, Lemma 19 and Lemma 20 give

$$
\begin{aligned}
R_1 &= -\langle \varepsilon_h^{y} - \varepsilon_h^{\widehat{y}}, \delta^{\boldsymbol{\Phi}}\cdot\boldsymbol{n} + \boldsymbol{\beta}\cdot\boldsymbol{n}\delta^{\Psi}\rangle_{\partial\mathcal{T}_h}\\
&\le h^{-\frac{1}{2}}\|\varepsilon_h^{y} - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h}\, h^{\frac{1}{2}}\|\delta^{\boldsymbol{\Phi}}\cdot\boldsymbol{n} + \boldsymbol{\beta}\cdot\boldsymbol{n}\delta^{\Psi}\|_{\partial\mathcal{T}_h}\\
&\le h^{-\frac{1}{2}}\|\varepsilon_h^{y} - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h}\|\delta^{\boldsymbol{\Phi}}\cdot\boldsymbol{n} + \boldsymbol{\beta}\cdot\boldsymbol{n}\delta^{\Psi}\|_{\mathcal{T}_h}\\
&\le Ch^{-\frac{1}{2}}\|\varepsilon_h^{y} - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h}(\|\delta^{\boldsymbol{\Phi}}\|_{\mathcal{T}_h} + \|\delta^{\Psi}\|_{\mathcal{T}_h})\\
&\le C(h^{s_{\boldsymbol{q}}+1}\,\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y}\,\|y\|_{s^{y},\Omega})\|\varepsilon_h^{y}\|_{\mathcal{T}_h},\\
R_2 &= (\nabla\varepsilon_h^{y}, \boldsymbol{\beta}\delta^{\Psi})_{\mathcal{T}_h} \le C\|\nabla\varepsilon_h^{y}\|_{\mathcal{T}_h}\|\delta^{\Psi}\|_{\mathcal{T}_h}\\
&\le C(h^{s_{\boldsymbol{q}}+1}\,\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y}\,\|y\|_{s^{y},\Omega})\|\varepsilon_h^{y}\|_{\mathcal{T}_h}.
\end{aligned}
$$

By a simple triangle inequality for terms $R_3$ and $R_4$, we have

$$
\begin{aligned}
R_3 &= (\boldsymbol{\beta}\delta^y, \nabla\Pi\Psi)_{\mathcal{T}_h} \le C\|\delta^y\|_{\mathcal{T}_h}\|\nabla\Pi\Psi\|_{\mathcal{T}_h}\\
&\le C\|\delta^y\|_{\mathcal{T}_h}(\|\nabla\delta^\Psi\|_{\mathcal{T}_h} + \|\nabla\Psi\|_{\mathcal{T}_h})\\
&\le C\|\delta^y\|_{\mathcal{T}_h}(h\|\Psi\|_{2,\Omega} + \|\Psi\|_{1,\Omega}) \le C\|\delta^y\|_{\mathcal{T}_h}\|\Psi\|_{2,\Omega}\\
&\le C(h^{s_{\boldsymbol{q}}+1}\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y}\|y\|_{s^y,\Omega})\|\varepsilon_h^y\|_{\mathcal{T}_h},\\
R_4 &= (\nabla\cdot\boldsymbol{\beta}\delta^y, \Pi\Psi)_{\mathcal{T}_h} \le C\|\delta^y\|_{\mathcal{T}_h}\|\Pi\Psi\|_{\mathcal{T}_h}\\
&\le C\|\delta^y\|_{\mathcal{T}_h}(\|\delta^\Psi\|_{\mathcal{T}_h} + \|\Psi\|_{\mathcal{T}_h})\\
&\le C\|\delta^y\|_{\mathcal{T}_h}(h^2\|\Psi\|_{2,\Omega} + \|\Psi\|_{\Omega}) \le C\|\delta^y\|_{\mathcal{T}_h}\|\Psi\|_{2,\Omega}\\
&\le C(h^{s_{\boldsymbol{q}}+1}\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y}\|y\|_{s^y,\Omega})\|\varepsilon_h^y\|_{\mathcal{T}_h}.
\end{aligned}
$$

For the term $R_5$, we have

$$
\begin{aligned}
R_5 &= \langle (h^{-1}+\tau_1)(\varepsilon_h^y - \varepsilon_h^{\widehat{y}}) + \widehat{\boldsymbol{\delta}}_1, \delta^\Psi - \delta^{\widehat{\Psi}}\rangle_{\partial\mathcal{T}_h}\\
&\le C(h^{-1}\|(\varepsilon_h^y - \varepsilon_h^{\widehat{y}})\|_{\partial\mathcal{T}_h} + \|\widehat{\boldsymbol{\delta}}_1\|_{\partial\mathcal{T}_h})\|\delta^\Psi - \delta^{\widehat{\Psi}}\|_{\partial\mathcal{T}_h}\\
&\le C(h^{s_{\boldsymbol{q}}+1}\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y}\|y\|_{s^y,\Omega})\|\varepsilon_h^y\|_{\mathcal{T}_h}.
\end{aligned}
$$

Finally, we complete the proof by summing the estimates for $R_1$ to $R_5$.

As a consequence, a simple application of the triangle inequality gives optimal convergence rates for $\|\boldsymbol{q} - \boldsymbol{q}_h(u)\|_{\mathcal{T}_h}$ and $\|y - y_h(u)\|_{\mathcal{T}_h}$:

**Lemma 22**

$$
\|\boldsymbol{q} - \boldsymbol{q}_h(u)\|_{\mathcal{T}_h} \le \|\delta^{\boldsymbol{q}}\|_{\mathcal{T}_h} + \|\varepsilon_h^{\boldsymbol{q}}\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{q}}}\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y-1}\|y\|_{s^y,\Omega}, \tag{3.48a}
$$

$$
\|y - y_h(u)\|_{\mathcal{T}_h} \le \|\delta^y\|_{\mathcal{T}_h} + \|\varepsilon_h^y\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{q}}+1}\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y}\|y\|_{s^y,\Omega}. \tag{3.48b}
$$

Step 4: The error equation for part 2 of the auxiliary problem (3.39b). Next, we focus on the dual variables, i.e., the state $z$ and the flux $\boldsymbol{p}$, and estimate the error between the solutions of the auxiliary problem and the mixed form (3.18a) - (3.18d) of the optimality system. Define

$$\begin{aligned}
\delta^{\boldsymbol{p}} &= \boldsymbol{p} - \boldsymbol{\Pi p}, & \varepsilon_h^{\boldsymbol{p}} &= \boldsymbol{\Pi p} - \boldsymbol{p}_h(u),\\
\delta^{z} &= z - \Pi z, & \varepsilon_h^{z} &= \Pi z - z_h(u),\\
\delta^{\widehat{z}} &= z - P_M z, & \varepsilon_h^{\widehat{z}} &= P_M z - \widehat{z}_h(u),\\
\widehat{\boldsymbol{\delta}}_2 &= \delta^{\boldsymbol{p}} \cdot \boldsymbol{n} + \boldsymbol{\beta} \cdot \boldsymbol{n} \delta^{\widehat{z}} + (h^{-1} + \tau_2)(\delta^{z} - \delta^{\widehat{z}}).
\end{aligned} \tag{3.49}$$

**Lemma 23** *We have*

$$\begin{aligned}
&\mathscr{B}_2(\varepsilon_h^{\boldsymbol{p}}, \varepsilon_h^{z}, \varepsilon_h^{\widehat{z}}, \boldsymbol{r}_2, w_2, \mu_2)\\
&\qquad = (\boldsymbol{\beta}\delta^{z}, \nabla w_2)_{\mathcal{T}_h} - \langle \widehat{\boldsymbol{\delta}}_2, w_2\rangle_{\partial\mathcal{T}_h} + \langle \widehat{\boldsymbol{\delta}}_2, \mu_2\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^{\partial}} + (y - y_h(u), w_2)_{\mathcal{T}_h}.
\end{aligned} \tag{3.50}$$

The proof is similar to the proof of Lemma 23 and is omitted.

Step 5: Estimate for $\varepsilon_h^{\boldsymbol{p}}$ Before we estimate $\varepsilon_h^{\boldsymbol{p}}$, we give the following discrete Poincaré inequality from [56].

**Lemma 24** *We have*

$$\|\varepsilon_h^{z}\|_{\mathcal{T}_h} \le C(\|\nabla\varepsilon_h^{z}\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\varepsilon_h^{z} - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}). \tag{3.51}$$

**Lemma 25** *We have*

$$\begin{aligned}
&\left\|\varepsilon_h^{\boldsymbol{p}}\right\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\varepsilon_h^{z} - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}\\
&\qquad \lesssim h^{s_{\boldsymbol{p}}}\|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega} + h^{s_z - 1}\|z\|_{s^{z},\Omega} + h^{s_{\boldsymbol{q}}+1}\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y}\|y\|_{s^{y},\Omega},
\end{aligned} \tag{3.52a}$$

$$\left\|\varepsilon_h^{z}\right\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}}\|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega} + h^{s_z - 1}\|z\|_{s^{z},\Omega} + h^{s_{\boldsymbol{q}}+1}\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y}\|y\|_{s^{y},\Omega}. \tag{3.52b}$$

**Proof:** First, we note the key inequality in Lemma 19 is valid with $(z, \boldsymbol{p}, \hat{z})$ in place of $(y, \boldsymbol{q}, \hat{y})$. This gives

$$\|\nabla \varepsilon_h^z\|_{\mathcal{T}_h} \leq \|\varepsilon_h^{\boldsymbol{p}}\|_{\mathcal{T}_h} + Ch^{-\frac{1}{2}}\|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}, \tag{3.53}$$

which we use below. Next, since $\varepsilon_h^{\widehat{z}} = 0$ on $\varepsilon_h^{\partial}$, the basic property of $\mathscr{B}_2$ in Lemma 16 gives

$$\mathscr{B}_2(\varepsilon_h^{\boldsymbol{p}}, \varepsilon_h^z, \varepsilon_h^{\widehat{z}}, \varepsilon_h^{\boldsymbol{p}}, \varepsilon_h^z, \varepsilon_h^{\widehat{z}}) = (\varepsilon_h^{\boldsymbol{p}}, \varepsilon_h^{\boldsymbol{p}})_{\mathcal{T}_h} + \|(h^{-1} + \tau_2 + \frac{1}{2}\boldsymbol{\beta}\cdot\boldsymbol{n})^{\frac{1}{2}}(\varepsilon_h^z - \varepsilon_h^{\widehat{z}})\|_{\partial\mathcal{T}_h}^2.$$

Then taking $(\boldsymbol{r}_2, w_2, \mu_2) = (\boldsymbol{\varepsilon}_h^{\boldsymbol{p}}, \varepsilon_h^z, \varepsilon_h^{\widehat{z}})$ in Lemma (3.50) in Lemma 23 gives

$$\begin{aligned}
&(\varepsilon_h^{\boldsymbol{p}}, \varepsilon_h^{\boldsymbol{p}})_{\mathcal{T}_h} + \|(h^{-1} + \tau_2 + \frac{1}{2}\boldsymbol{\beta}\cdot\boldsymbol{n})^{\frac{1}{2}}(\varepsilon_h^z - \varepsilon_h^{\widehat{z}})\|_{\partial\mathcal{T}_h}^2\\
&\qquad = (\boldsymbol{\beta}\delta^z, \nabla\varepsilon_h^z)_{\mathcal{T}_h} - \langle\widehat{\boldsymbol{\delta}}_2, \varepsilon_h^z - \varepsilon_h^{\widehat{z}}\rangle_{\partial\mathcal{T}_h} + (y - y_h(u), \varepsilon_h^z)_{\mathcal{T}_h}\\
&\qquad =: T_1 + T_2 + T_3.
\end{aligned}$$

By the same argument as in Lemma 20, simply applying Lemma (3.53) and Young's inequality gives

$$\begin{aligned}
T_1 &= (\boldsymbol{\beta}\delta^z, \nabla\varepsilon_h^z)_{\mathcal{T}_h} \leq C\|\delta^z\|_{\mathcal{T}_h}^2 + \frac{1}{4}\|\varepsilon_h^{\boldsymbol{p}}\|_{\mathcal{T}_h}^2 + \frac{1}{4h}\|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}^2,\\
T_2 &= -\langle\widehat{\boldsymbol{\delta}}_2, \varepsilon_h^z - \varepsilon_h^{\widehat{z}}\rangle_{\partial\mathcal{T}_h} \leq 4h\|\widehat{\boldsymbol{\delta}}\|_{\partial\mathcal{T}_h}^2 + \frac{1}{4h}\|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}^2.
\end{aligned}$$

Finally, for the term $T_3$, we have

$$\begin{aligned}
T_3 &= (y - y_h(u), \varepsilon_h^z)_{\mathcal{T}_h} \leq \|y - y_h(u)\|_{\mathcal{T}_h}\|\varepsilon_h^z\|_{\mathcal{T}_h}\\
&\leq C\|y - y_h(u)\|_{\mathcal{T}_h}(\|\nabla\varepsilon_h^z\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h})\\
&\leq C\|y - y_h(u)\|_{\mathcal{T}_h}(\|\varepsilon_h^{\boldsymbol{p}}\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h})\\
&\leq C\|y - y_h(u)\|_{\mathcal{T}_h}^2 + \frac{1}{4}\|\varepsilon_h^{\boldsymbol{p}}\|_{\mathcal{T}_h}^2 + \frac{1}{4h}\|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}^2.
\end{aligned}$$

Summing $T_1$ to $T_3$ gives

$$\left\|\varepsilon_h^{\boldsymbol{p}}\right\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}} \|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega} + h^{s_z-1} \|z\|_{s^z,\Omega} + h^{s_{\boldsymbol{q}}+1} \|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y} \|y\|_{s^y,\Omega}.$$

Finally, (3.51), (3.52a), and (3.53) together imply (3.52b).

As a consequence, a simple application of the triangle inequality gives optimal convergence rates for $\|\boldsymbol{p} - \boldsymbol{p}_h(u)\|_{\mathcal{T}_h}$ and $\|z - z_h(u)\|_{\mathcal{T}_h}$:

**Lemma 26** *We have*

$$\|\boldsymbol{p} - \boldsymbol{p}_h(u)\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}} \|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega} + h^{s_z-1} \|z\|_{s^z,\Omega} + h^{s_{\boldsymbol{q}}+1} \|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y} \|y\|_{s^y,\Omega}, \tag{3.54a}$$

$$\|z - z_h(u)\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}} \|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega} + h^{s_z-1} \|z\|_{s^z,\Omega} + h^{s_{\boldsymbol{q}}+1} \|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y} \|y\|_{s^y,\Omega}. \tag{3.54b}$$

Step 6: Estimate for $\|u - u_h\|_{\varepsilon_h^\partial}$ and $\|y - y_h\|_{\mathcal{T}_h}$

Next, we bound the error between the solutions of the auxiliary problem and the HDG problem (3.37). We use these error bounds and the error bounds in Lemma 22, Lemma 25, and Lemma 26 to obtain the main results.

For the remaining steps, we denote

$$\begin{aligned}
\zeta_{\boldsymbol{q}} &= \boldsymbol{q}_h(u) - \boldsymbol{q}_h, \quad \zeta_y = y_h(u) - y_h, \quad \zeta_{\widehat{y}} = \widehat{y}_h(u) - \widehat{y}_h,\\
\zeta_{\boldsymbol{p}} &= \boldsymbol{p}_h(u) - \boldsymbol{p}_h, \quad \zeta_z = z_h(u) - z_h, \quad \zeta_{\widehat{z}} = \widehat{z}_h(u) - \widehat{z}_h.
\end{aligned}$$

Subtracting the auxiliary problem and the HDG problem gives the following error equations

$$\mathscr{B}_1(\zeta_{\boldsymbol{q}}, \zeta_y, \zeta_{\widehat{y}}; \boldsymbol{r}_1, w_1, \mu_1) = -\langle P_M u - u_h, \boldsymbol{r}_1 \cdot \boldsymbol{n} + (\boldsymbol{\beta} \cdot \boldsymbol{n} - h^{-1} - \tau_1) w_1\rangle_{\varepsilon_h^\partial}, \tag{3.55a}$$

$$\mathscr{B}_2(\zeta_{\boldsymbol{p}}, \zeta_z, \zeta_{\widehat{z}}; \boldsymbol{r}_2, w_2, \mu_2) = (\zeta_y, w_2)_{\mathcal{T}_h}. \tag{3.55b}$$

**Lemma 27** *If **(A1)** and **(A2)** hold, then*

$$\gamma \|u - u_h\|^2_{\varepsilon_h^\partial} + \left\|\zeta_y\right\|^2_{\mathcal{T}_h} = \langle \gamma u + \boldsymbol{p}_h(u) \cdot \boldsymbol{n} + h^{-1} z_h(u) + \tau_2 z_h(u), u - u_h\rangle_{\varepsilon_h^\partial}$$
$$- \langle \gamma u_h + \boldsymbol{p}_h \cdot \boldsymbol{n} + h^{-1} z_h + \tau_2 z_h, u - u_h\rangle_{\varepsilon_h^\partial}.$$

**Proof:** First, we have

$$\langle \gamma u + \boldsymbol{p}_h(u) \cdot \boldsymbol{n} + h^{-1} z_h(u) + \tau_2 z_h(u), u - u_h\rangle_{\varepsilon_h^\partial} - \langle \gamma u_h + \boldsymbol{p}_h \cdot \boldsymbol{n} + h^{-1} z_h + \tau_2 z_h, u - u_h\rangle_{\varepsilon_h^\partial}$$
$$= \gamma \|u - u_h\|^2_{\varepsilon_h^\partial} + \langle \zeta_{\boldsymbol{p}} \cdot \boldsymbol{n} + h^{-1}\zeta_z + \tau_2\zeta_z, u - u_h\rangle_{\varepsilon_h^\partial}.$$

Next, 17 gives

$$\mathscr{B}_1(\zeta_{\boldsymbol{q}}, \zeta_y, \zeta_{\widehat{y}}; \zeta_{\boldsymbol{p}}, -\zeta_z, -\zeta_{\widehat{z}}) + \mathscr{B}_2(\zeta_{\boldsymbol{p}}, \zeta_z, \zeta_{\widehat{z}}; -\zeta_{\boldsymbol{q}}, \zeta_y, \zeta_{\widehat{y}}) = 0.$$

On the other hand, since $\tau_2$ is piecewise constant on $\partial\mathcal{T}_h$, we have

$$\begin{aligned}
&\mathscr{B}_1(\zeta_{\boldsymbol{q}}, \zeta_y, \zeta_{\widehat{y}}; \zeta_{\boldsymbol{p}}, -\zeta_z, -\zeta_{\widehat{z}}) + \mathscr{B}_2(\zeta_{\boldsymbol{p}}, \zeta_z, \zeta_{\widehat{z}}; -\zeta_{\boldsymbol{q}}, \zeta_y, \zeta_{\widehat{y}}) \\
&\quad = (\zeta_y, \zeta_y)_{\mathcal{T}_h} - \langle P_M u - u_h, \zeta_{\boldsymbol{p}} \cdot \boldsymbol{n} + (h^{-1} + \tau_1 - \boldsymbol{\beta} \cdot \boldsymbol{n})\zeta_z\rangle_{\varepsilon_h^\partial} \\
&\quad = (\zeta_y, \zeta_y)_{\mathcal{T}_h} - \langle P_M u - u_h, \zeta_{\boldsymbol{p}} \cdot \boldsymbol{n} + h^{-1}\zeta_z + \tau_2\zeta_z\rangle_{\varepsilon_h^\partial} \\
&\quad = (\zeta_y, \zeta_y)_{\mathcal{T}_h} - \langle u - u_h, \zeta_{\boldsymbol{p}} \cdot \boldsymbol{n} + h^{-1}\zeta_z + \tau_2\zeta_z\rangle_{\varepsilon_h^\partial}.
\end{aligned}$$

Comparing the above two equalities gives

$$(\zeta_y, \zeta_y)_{\mathcal{T}_h} = \langle u - u_h, \zeta_{\boldsymbol{p}} \cdot \boldsymbol{n} + h^{-1}\zeta_z + \tau_2\zeta_z\rangle_{\varepsilon_h^\partial}.$$

**Theorem 8** *We have*

$$\|u - u_h\|_{\varepsilon_h^\partial} \lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega},$$
$$\|y - y_h\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}.$$

**Proof:** Since $\gamma u + \boldsymbol{p}\cdot\boldsymbol{n} = 0$ on $\varepsilon_h^\partial$ and $\gamma u_h + \boldsymbol{p}_h\cdot\boldsymbol{n} + h^{-1}z_h + \tau_2 z_h = 0$ on $\varepsilon_h^\partial$ we have

$$\begin{aligned}\gamma\|u - u_h\|_{\varepsilon_h^\partial}^2 + \left\|\zeta_y\right\|_{\mathcal{T}_h}^2 &= \langle \gamma u + \boldsymbol{p}_h(u)\cdot\boldsymbol{n} + h^{-1}z_h(u) + \tau_2 z_h(u), u - u_h\rangle_{\varepsilon_h^\partial}\\ &= \langle (\boldsymbol{p}_h(u) - \boldsymbol{p})\cdot\boldsymbol{n} + h^{-1}z_h(u) + \tau_2 z_h(u), u - u_h\rangle_{\varepsilon_h^\partial}.\end{aligned}$$

Next, since $\widehat{z}_h(u) = z = 0$ on $\varepsilon_h^\partial$ we have

$$\begin{aligned}\|\boldsymbol{p}_h(u) - \boldsymbol{p}\|_{\partial\mathcal{T}_h} &\le \|\boldsymbol{p}_h(u) - \boldsymbol{\Pi}\boldsymbol{p}\|_{\partial\mathcal{T}_h} + \|\boldsymbol{\Pi}\boldsymbol{p} - \boldsymbol{p}\|_{\partial\mathcal{T}_h}\\ &\lesssim h^{-\frac{1}{2}}\left\|\varepsilon_h^{\boldsymbol{p}}\right\|_{\mathcal{T}_h} + h^{s_{\boldsymbol{p}}-\frac{1}{2}}\|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega},\\ \|z_h(u)\|_{\varepsilon_h^\partial} &= \|z_h(u) - \Pi z + P_M z - \widehat{z}_h(u)\|_{\varepsilon_h^\partial} = \|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}.\end{aligned}$$

Some simple manipulations gives

$$\|u - u_h\|_{\varepsilon_h^\partial} + \|\zeta_y\|_{\mathcal{T}_h} \lesssim h^{-\frac{1}{2}}\left\|\varepsilon_h^{\boldsymbol{p}}\right\|_{\mathcal{T}_h} + h^{s_{\boldsymbol{p}}-\frac{1}{2}}\|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega} + h^{-1}\|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}.$$

By Lemma 25 and properties of the $L^2$ projection, we have

$$\begin{aligned}&\|u - u_h\|_{\varepsilon_h^\partial} + \|\zeta_y\|_{\mathcal{T}_h}\\ &\qquad\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}}\|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}}\|z\|_{s^z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}}\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}}\|y\|_{s^y,\Omega}.\end{aligned}$$

Then, by the triangle inequality and Lemma 22 we obtain

$$\|y - y_h\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s^z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s^y,\Omega}.$$

Step 7: Estimates for $\|\boldsymbol{p} - \boldsymbol{p}_h\|_{\mathcal{T}_h}$, $\|z - z_h\|_{\mathcal{T}_h}$ and $\|\boldsymbol{q} - \boldsymbol{q}_h\|_{\mathcal{T}_h}$

**Lemma 28** *We have*

$$\begin{aligned}
\left\|\zeta_{\boldsymbol{p}}\right\|_{\mathcal{T}_h} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega},\\
\|\zeta_z\|_{\mathcal{T}_h} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}.
\end{aligned}$$

**Proof:** By Lemma 16, the error equation (3.55b), and since $\zeta_{\widehat{z}} = 0$ on $\varepsilon_h^{\partial}$, we have

$$\begin{aligned}
&\mathscr{B}_2(\zeta_{\boldsymbol{p}}, \zeta_z, \zeta_{\widehat{z}}; \zeta_{\boldsymbol{p}}, \zeta_z, \zeta_{\widehat{z}})\\
&\quad= (\zeta_{\boldsymbol{p}}, \zeta_{\boldsymbol{p}})_{\mathcal{T}_h} + \langle (h^{-1} + \tau_2 + \frac{1}{2}\boldsymbol{\beta}\cdot\boldsymbol{n})(\zeta_z - \zeta_{\widehat{z}}), \zeta_z - \zeta_{\widehat{z}}\rangle_{\partial\mathcal{T}_h}\\
&\quad= (\zeta_y, \zeta_z)_{\mathcal{T}_h}\\
&\quad\le \left\|\zeta_y\right\|_{\mathcal{T}_h} \|\zeta_z\|_{\mathcal{T}_h}\\
&\quad\lesssim \left\|\zeta_y\right\|_{\mathcal{T}_h} (\|\nabla\zeta_z\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\zeta_z - \zeta_{\widehat{z}}\|_{\partial\mathcal{T}_h})\\
&\quad\lesssim \left\|\zeta_y\right\|_{\mathcal{T}_h} (\|\zeta_{\boldsymbol{p}}\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\zeta_z - \zeta_{\widehat{z}}\|_{\partial\mathcal{T}_h}),
\end{aligned}$$

where we used the discrete Poincaré inequality in Lemma 24 and also Lemma 19. This implies

$$\begin{aligned}
&\left\|\zeta_{\boldsymbol{p}}\right\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\zeta_z - \zeta_{\widehat{z}}\|_{\partial\mathcal{T}_h}\\
&\qquad\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}.
\end{aligned}$$

The discrete Poincaré inequality in Lemma 24 also gives

$$
\begin{aligned}
\|\zeta_z\|_{\mathcal{T}_h} &\lesssim \|\nabla \zeta_z\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\zeta_z - \zeta_{\widehat{z}}\|_{\partial\mathcal{T}_h} \\
&\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega} .
\end{aligned}
$$

**Lemma 29** *If **(A1)** and $k \geq 1$ hold, then*

$$
\left\|\zeta_{\boldsymbol{q}}\right\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}-1} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-2} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-1} \|y\|_{s_y,\Omega} .
$$

**Proof:** By Lemma 16, the error equation (3.55a), and since $\tau_2$ is piecewise constant on $\partial\mathcal{T}_h$, we have

$$
\begin{aligned}
&\mathscr{B}_1(\zeta_{\boldsymbol{q}}, \zeta_y, \zeta_{\widehat{y}}; \zeta_{\boldsymbol{q}}, \zeta_y, \zeta_{\widehat{y}}) \\
&\quad= (\zeta_{\boldsymbol{q}}, \zeta_{\boldsymbol{q}})_{\mathcal{T}_h} + \langle (h^{-1} + \tau_1 - \frac{1}{2}\boldsymbol{\beta}\cdot\boldsymbol{n})(\zeta_y - \zeta_{\widehat{y}}), \zeta_y - \zeta_{\widehat{y}}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} - (\nabla\cdot\boldsymbol{\beta}\zeta_y, \zeta_y)_{\mathcal{T}_h} \\
&\qquad + \langle (h^{-1} + \tau_1 - \frac{1}{2}\boldsymbol{\beta}\cdot\boldsymbol{n})\zeta_y, \zeta_y\rangle_{\varepsilon_h^\partial} \\
&\quad= -\langle P_M u - u_h, \zeta_{\boldsymbol{q}}\cdot\boldsymbol{n} + (\boldsymbol{\beta}\cdot\boldsymbol{n} - h^{-1} - \tau_1)\zeta_y\rangle_{\varepsilon_h^\partial} \\
&\quad= -\langle P_M u - u_h, \zeta_{\boldsymbol{q}}\cdot\boldsymbol{n} - (h^{-1} + \tau_2)\zeta_y\rangle_{\varepsilon_h^\partial} \\
&\quad= -\langle u - u_h, \zeta_{\boldsymbol{q}}\cdot\boldsymbol{n} - (h^{-1} + \tau_2)\zeta_y\rangle_{\varepsilon_h^\partial} \\
&\quad\lesssim \|u - u_h\|_{\varepsilon_h^\partial} (\left\|\zeta_{\boldsymbol{q}}\right\|_{\varepsilon_h^\partial} + h^{-1}\left\|\zeta_y\right\|_{\varepsilon_h^\partial}) \\
&\quad\lesssim h^{-\frac{1}{2}} \|u - u_h\|_{\varepsilon_h^\partial} (\left\|\zeta_{\boldsymbol{q}}\right\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\left\|\zeta_y\right\|_{\varepsilon_h^\partial}),
\end{aligned}
$$

which gives

$$
\begin{aligned}
\left\|\zeta_{\boldsymbol{q}}\right\|_{\mathcal{T}_h} &\lesssim h^{-\frac{1}{2}} \|u - u_h\|_{\varepsilon_h^\partial} \\
&\lesssim h^{s_{\boldsymbol{p}}-1} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-2} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-1} \|y\|_{s_y,\Omega} .
\end{aligned}
$$

The above lemma along with the triangle inequality, Lemma 22, and Lemma 26 complete the proof of the main result:

**Theorem 3.4.2** *We have*

$$\|\boldsymbol{p}-\boldsymbol{p}_h\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}}\|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}}\|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}}\|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}}\|y\|_{s_y,\Omega},$$
$$\|z-z_h\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}}\|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}}\|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}}\|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}}\|y\|_{s_y,\Omega}.$$

*If in addition* $k \geq 1$, *then*

$$\|\boldsymbol{q}-\boldsymbol{q}_h\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}-1}\|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-2}\|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}}\|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-1}\|y\|_{s_y,\Omega}.$$

## 3.5. NUMERICAL EXPERIMENTS

We present numerical results for a 2D example problem on a square domain $\Omega = [0, 1/8]\times[0, 1/8]$. For the results presented below, we chose $\tau_1 = \tau_2 = 1$ for the stabilization functions. In other numerical experiments not reported here, we also chose $\tau_1$ and $\tau_2$ to satisfy the conditions **(A1)-(A3)** and we obtained similar results.

The problem data is chosen as

$$f = 0, \quad y_d = (x^2+y^2)^s, \quad \boldsymbol{\beta} = [1,1], \quad \text{and} \quad \gamma = 1,$$

where $s = -10^{-5}$. Therefore $y_d$ has a singularity, but $y_d \in H^{1-\varepsilon}(\Omega)$ for all $\varepsilon > 2\times 10^{-5}$. Since the largest interior angle is $\omega = \pi/2$, we have $r_\Omega = 3/2 - \varepsilon$ for all $\varepsilon > 2\times 10^{-5}$.

An exact solution for this problem is not known, and therefore we compare the approximate solutions computed using various values of $h$ and a reference solution computed on a fine mesh with 524288 elements and $h = 2^{-12}\sqrt{2}$.

Table 3.1. 2D Example with $k = 1$: Errors for the control $u$, state $y$, adjoint state $z$, and the fluxes $\boldsymbol{q}$ and $\boldsymbol{p}$

| $h/\sqrt{2}$ | $2^{-4}$ | $2^{-5}$ | $2^{-6}$ | $2^{-7}$ | $2^{-8}$ |
|---|---|---|---|---|---|
| $\|\boldsymbol{q} - \boldsymbol{q}_h\|_{0,\Omega}$ | 2.57e-2 | 1.32e-2 | 6.66e-3 | 3.35e-3 | 1.68e-3 |
| order | - | 0.96 | 0.98 | 1.00 | 1.00 |
| $\|\boldsymbol{p} - \boldsymbol{p}_h\|_{0,\Omega}$ | 5.01e-4 | 1.57e-4 | 4.57e-5 | 1.29e-5 | 3.55e-6 |
| order | - | 1.68 | 1.78 | 1.83 | 1.86 |
| $\|y - y_h\|_{0,\Omega}$ | 2.00e-4 | 5.04e-5 | 1.29e-5 | 3.26e-6 | 8.21e-7 |
| order | - | 1.99 | 1.99 | 1.98 | 1.99 |
| $\|z - z_h\|_{0,\Omega}$ | 3.41e-6 | 5.20e-7 | 7.60e-8 | 1.08e-8 | 1.49e-9 |
| order | - | 2.71 | 2.77 | 2.82 | 2.85 |
| $\|u - u_h\|_{0,\Gamma}$ | 2.40e-3 | 9.04e-4 | 3.27e-4 | 1.17e-4 | 4.18e-5 |
| order | - | 1.41 | 1.47 | 1.48 | 1.49 |

For $k = 1$, Corollary 4 in Section 3.4 gives the convergence rates

$$\|y - y_h\|_{0,\Omega} = O(h^{3/2-\varepsilon}), \qquad \|z - z_h\|_{0,\Omega} = O(h^{3/2-\varepsilon}),$$

$$\|\boldsymbol{q} - \boldsymbol{q}_h\|_{0,\Omega} = O(h^{1-\varepsilon}), \qquad \|\boldsymbol{p} - \boldsymbol{p}_h\|_{0,\Omega} = O(h^{3/2-\varepsilon}),$$

and

$$\|u - u_h\|_{0,\Gamma} = O(h^{3/2-\varepsilon}).$$

Table 3.1 shows the computed errors for a variety of mesh sizes. The convergence rates for the optimal control $u$ and the flux $\boldsymbol{q}$ are precisely predicted by the convergence theory presented here. The convergence rates for the other variables are higher than predicted by the theory; this phenomenon has been observed numerically in other works on numerical methods for boundary control problems [33, 49, 55].

For the case $k = 0$, Corollary 4 in Section 3.4 gives the convergence rates

$$\|y - y_h\|_{0,\Omega} = O(h^{1/2}), \quad \|z - z_h\|_{0,\Omega} = O(h^{1/2}), \quad \|\boldsymbol{p} - \boldsymbol{p}_h\|_{0,\Omega} = O(h^{1/2}),$$

Table 3.2. 2D Example with $k = 0$: Errors for the control $u$, state $y$, adjoint state $z$, and the fluxes $\boldsymbol{q}$ and $\boldsymbol{p}$

| $h/\sqrt{2}$ | $2^{-4}$ | $2^{-5}$ | $2^{-6}$ | $2^{-7}$ | $2^{-8}$ |
|---|---|---|---|---|---|
| $\Vert\boldsymbol{q} - \boldsymbol{q}_h\Vert_{0,\Omega}$ | 4.67e-2 | 3.27e-2 | 2.01e-2 | 1.18e-2 | 6.66e-3 |
| order | - | 0.51 | 0.70 | 0.76 | 0.82 |
| $\Vert\boldsymbol{p} - \boldsymbol{p}_h\Vert_{0,\Omega}$ | 1.61e-3 | 9.22e-4 | 4.81e-4 | 2.44e-4 | 1.22e-4 |
| order | - | 0.80 | 0.94 | 0.98 | 1.00 |
| $\Vert y - y_h\Vert_{0,\Omega}$ | 6.54e-4 | 2.77e-4 | 9.10e-5 | 2.83e-5 | 8.80e-6 |
| order | - | 1.24 | 1.60 | 1.68 | 1.69 |
| $\Vert z - z_h\Vert_{0,\Omega}$ | 6.54e-5 | 1.82e-5 | 4.74e-6 | 1.20e-6 | 3.02e-7 |
| order | - | 1.85 | 1.94 | 2.00 | 2.00 |
| $\Vert u - u_h\Vert_{0,\Gamma}$ | 5.88e-3 | 3.50e-3 | 1.92e-3 | 1.01e-3 | 5.21e-4 |
| order | - | 0.75 | 0.87 | 0.92 | 0.96 |

and

$$\Vert u - u_h\Vert_{0,\Gamma} = O(h^{1/2}).$$

As mentioned in Section 3.4, the convergence rates for $k = 0$ obtained in Corollary 4 are suboptimal. Numerical results are reported in Table 3.2, and all convergence rates are higher than predicted by the theory. Obtaining optimal convergence rates for the $k = 0$ case is an interesting topic for future work.

# 4. CONVECTION DIFFUSION WITH LOW REGULARITY

## 4.1. BACKGROUND: REGULARITY AND HDG FORMULATION

To begin, we briefly review the regularity results for the optimal control problem and the new HDG method from Section 2.

**4.1.1. Optimal Control Problem: Regularity.** Theorem 6 in Section 2 implies the regularity of the solution of the optimality system (3.18a)-(3.18e) depends on the desired state $y_d$ and the domain $\Omega$. As is known, solutions to Dirichlet boundary control problems can have low regularity; this causes difficulty for numerical analysis.

In Section 2, for the numerical analysis of the new HDG method we assumed $\Omega$ is convex, $y_d \in H^{t^*}(\Omega)$ for some $t^* \in (1/2, 1)$, and $\pi/3 < \omega < 2\pi/3$. These assumptions give high regularity for the optimal control, i.e., $u \in H^{r_u}(\Gamma)$ for some $r_u \in (1, 3/2)$. Furthermore, the assumptions give $\boldsymbol{q} \in H^{r_{\boldsymbol{q}}}(\Omega)$ with $r_{\boldsymbol{q}} > 1/2$, which guarantees $\boldsymbol{q}$ has a well-defined trace on the boundary $\Gamma$. We used this property in the HDG convergence analysis.

In this Section we again assume $\Omega$ is convex, but we remove the restrictions on the desired state and the largest interior angle for the numerical analysis; i.e., we only require $t^* \in [0, 1)$ and $\pi/3 \leq \omega < \pi$. In this case, the regularity of the optimal control can be low, i.e., $u \in H^{r_u}(\Gamma)$ for some $r_u \in [1/2, 1)$, and $\boldsymbol{q}$ is no longer guaranteed to have a well-defined $L^2$ boundary trace; however, the optimality system (3.18a)-(3.18e) can be understood in a standard weak sense.

**4.1.2. The HDG Formulation.** To approximate the solution of the mixed weak form (3.18a)-(3.18e) of the optimality system, the HDG formulation considered here is modified from Section 2 to avoid the estimation of $\boldsymbol{q}$ on the boundary. In the 2D case, recall from Section 4.1.1 that $\boldsymbol{q}$ is not guaranteed to have a well-defined $L^2$ boundary trace since we consider a solution of the optimal control problem with low regularity.

The HDG method seeks approximate fluxes $\boldsymbol{q}_h, \boldsymbol{p}_h \in \boldsymbol{V}_h$, states $y_h, z_h \in W_h$, interior element boundary traces $\widehat{y}_h^o, \widehat{z}_h^o \in M_h(o)$, and boundary control $u_h \in M_h(\partial)$ satisfying

$$(\boldsymbol{q}_h, \boldsymbol{r}_1)_{\mathcal{T}_h} - (y_h, \nabla\cdot\boldsymbol{r}_1)_{\mathcal{T}_h} + \langle \widehat{y}_h^o, \boldsymbol{r}_1\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} + \langle u_h, \boldsymbol{r}_1\cdot\boldsymbol{n}\rangle_{\varepsilon_h^\partial} = 0, \tag{4.1a}$$

$$\begin{aligned}(\nabla\cdot\boldsymbol{q}_h, w_1)_{\mathcal{T}_h} - (\boldsymbol{\beta} y_h, \nabla w_1)_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta} y_h, w_1)_{\mathcal{T}_h} \\ +\langle (h^{-1}+\tau_1) y_h, w_1\rangle_{\partial\mathcal{T}_h} + \langle (\boldsymbol{\beta}\cdot\boldsymbol{n} - \tau_1 - h^{-1})\widehat{y}_h^o, w_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} \\ +\langle (\boldsymbol{\beta}\cdot\boldsymbol{n} - \tau_1 - h^{-1}) u_h, w_1\rangle_{\varepsilon_h^\partial} = (f, w_1)_{\mathcal{T}_h},\end{aligned} \tag{4.1b}$$

for all $(\boldsymbol{r}_1, w_1) \in \boldsymbol{V}_h \times W_h$,

$$(\boldsymbol{p}_h, \boldsymbol{r}_2)_{\mathcal{T}_h} - (z_h, \nabla\cdot\boldsymbol{r}_2)_{\mathcal{T}_h} + \langle \widehat{z}_h^o, \boldsymbol{r}_2\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} = 0, \tag{4.1c}$$

$$\begin{aligned}(\nabla\cdot\boldsymbol{p}_h, w_2)_{\mathcal{T}_h} - (y_h, w_2)_{\mathcal{T}_h} + (\boldsymbol{\beta} z_h, \nabla w_2)_{\mathcal{T}_h} \\ +\langle (h^{-1}+\tau_2) z_h, w_2\rangle_{\partial\mathcal{T}_h} - \langle (h^{-1}+\tau_2+\boldsymbol{\beta}\cdot\boldsymbol{n})\widehat{z}_h^o, w_2\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} = -(y_d, w_2)_{\mathcal{T}_h},\end{aligned} \tag{4.1d}$$

for all $(\boldsymbol{r}_2, w_2) \in \boldsymbol{V}_h \times W_h$,

$$\begin{aligned}\langle \boldsymbol{q}_h\cdot\boldsymbol{n}, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} + \langle (h^{-1}+\tau_1) y_h, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} \\ +\langle (\boldsymbol{\beta}\cdot\boldsymbol{n} - \tau_1 - h^{-1})\widehat{y}_h^o, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} = 0,\end{aligned} \tag{4.1e}$$

for all $\mu_1 \in M_h(o)$,

$$\begin{aligned}\langle \boldsymbol{p}_h\cdot\boldsymbol{n}, \mu_2\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} + \langle (h^{-1}+\tau_2) z_h, \mu_2\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} \\ -\langle (\boldsymbol{\beta}\cdot\boldsymbol{n} + \tau_2 + h^{-1})\widehat{z}_h^o, \mu_2\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} = 0,\end{aligned} \tag{4.1f}$$

for all $\mu_2 \in M_h(o)$, and the optimality condition

$$\langle \boldsymbol{p}_h\cdot\boldsymbol{n}, \mu_3\rangle_{\varepsilon_h^\partial} + \gamma\langle u_h, \mu_3\rangle_{\varepsilon_h^\partial} + \langle (h^{-1}+\tau_2) z_h, \mu_3\rangle_{\varepsilon_h^\partial} = 0, \tag{4.1g}$$

for all $\mu_3 \in M_h(\partial)$.

Here, $\tau_1$ and $\tau_2$ are stabilization functions defined on $\partial\mathcal{T}_h$ that satisfy the same conditions as in Section 2 :

**(A1)** $\tau_2$ is piecewise constant on $\partial\mathcal{T}_h$.

**(A2)** $\tau_1 = \tau_2 + \boldsymbol{\beta} \cdot \boldsymbol{n}$.

**(A3)** For any $K \in \mathcal{T}_h$, $\min{(\tau_2 + \frac{1}{2}\boldsymbol{\beta} \cdot \boldsymbol{n})|_{\partial K}} > 0$.

Conditions **(A2)** and **(A3)** imply

$$\min{(\tau_1 - \frac{1}{2}\boldsymbol{\beta} \cdot \boldsymbol{n})|_{\partial K}} > 0 \quad \text{for any } K \in \mathcal{T}_h. \tag{4.2}$$

This completes the formulation of the HDG method.

Notice that formulation (4.1) is slightly different from formulation (3.4) in Section 2; specifically, equations (b) and (d) are modified. A straightforward computation shows that both are equivalent; see, Section 3.2. Formulation (4.1) above allows us to achieve error estimates in the low regularity case considered here.

## 4.2. ERROR ANALYSIS

As mentioned in Section 4.1.1, the regularity of $\boldsymbol{q}$ can be low and therefore $\boldsymbol{q}$ may not have a $L^2$ boundary trace. The $H(\text{div}, \Omega)$ regularity of $\boldsymbol{q}$ is critically important for the numerical analysis.

We also require the family of meshes $\{\mathcal{T}_h\}$ is a conforming quasi-uniform triangulation of $\Omega$. This assumption on the meshes is stronger than in Section 2; there we assumed $\{\mathcal{T}_h\}$ is a conforming quasi-uniform polyhedral mesh. Therefore, the analysis in Section 2 allows for a more general family of meshes; however, the analysis here allows us to treat the low regularity case.

We now state our main convergence result.

**Theorem 9** *Let*

$$\begin{aligned} s_{\boldsymbol{q}} &= \min\{r_{\boldsymbol{q}}, k+1\}, \qquad & s_y &= \min\{r_y, k+2\}, \\ s_{\boldsymbol{p}} &= \min\{r_{\boldsymbol{p}}, k+1\}, \qquad & s_z &= \min\{r_z, k+2\}. \end{aligned} \tag{4.3}$$

*If the above assumptions hold and* $s_{\boldsymbol{q}} \in [0,1]$*, then*

$$\begin{aligned} \|u - u_h\|_{\varepsilon_h^\partial} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}, \\ \|y - y_h\|_{\mathcal{T}_h} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}, \\ \|\boldsymbol{p} - \boldsymbol{p}_h\|_{\mathcal{T}_h} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}, \\ \|z - z_h\|_{\mathcal{T}_h} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}. \end{aligned}$$

*If in addition the inequalities in* (3.31) *are strict and* $k \geq 1$*, then*

$$\|\boldsymbol{q} - \boldsymbol{q}_h\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}-1} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-2} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-1} \|y\|_{s_y,\Omega}.$$

**Remark 4** *Note that we assume* $s_{\boldsymbol{q}} \in [0,1]$*. This is not a restriction since the case* $s_{\boldsymbol{q}} > 1$ *is treated in Section 2 on a more general family of meshes.*

Specializing to the 2D case gives the following result:

**Corollary 5** *Suppose* $d = 2$*,* $f = 0$*,* $s_{\boldsymbol{q}} \in [0,1]$*, and* $y_d \in H^{t^*}(\Omega)$ *for some* $t^* \in [0,1)$*. Let* $\pi/3 \leq \omega < \pi$ *be the largest interior angle of* $\Gamma$*, and let* $r > 0$ *satisfy*

$$r \leq r_d := \frac{1}{2} + t^* \in [1/2, 3/2), \quad \text{and} \quad r < r_\Omega := \min\left\{\frac{3}{2}, \frac{\pi}{\omega} - \frac{1}{2}\right\} \in (1/2, 3/2].$$

*If $k = 1$, then*

$$\|u - u_h\|_{\varepsilon_h^\partial} \lesssim h^r(\|\boldsymbol{p}\|_{H^{r+1/2}(\Omega)} + \|z\|_{H^{r+3/2}(\Omega)} + \|\boldsymbol{q}\|_{H^{r-1/2}(\Omega)} + \|y\|_{H^{r+1/2}(\Omega)}),$$
$$\|y - y_h\|_{\mathcal{T}_h} \lesssim h^r(\|\boldsymbol{p}\|_{H^{r+1/2}(\Omega)} + \|z\|_{H^{r+3/2}(\Omega)} + \|\boldsymbol{q}\|_{H^{r-1/2}(\Omega)} + \|y\|_{H^{r+1/2}(\Omega)}),$$
$$\|\boldsymbol{p} - \boldsymbol{p}_h\|_{\mathcal{T}_h} \lesssim h^r(\|\boldsymbol{p}\|_{H^{r+1/2}(\Omega)} + \|z\|_{H^{r+3/2}(\Omega)} + \|\boldsymbol{q}\|_{H^{r-1/2}(\Omega)} + \|y\|_{H^{r}+1/2(\Omega)}),$$
$$\|z - z_h\|_{\mathcal{T}_h} \lesssim h^r(\|\boldsymbol{p}\|_{H^{r+1/2}(\Omega)} + \|z\|_{H^{r+3/2}(\Omega)} + \|\boldsymbol{q}\|_{H^{r-1/2}(\Omega)} + \|y\|_{H^{r+1/2}(\Omega)}).$$

*If in addition $r > 1/2$, then*

$$\|\boldsymbol{q} - \boldsymbol{q}_h\|_{\mathcal{T}_h} \lesssim h^{r-1/2}(\|\boldsymbol{p}\|_{H^{r+1/2}(\Omega)} + \|z\|_{H^{r+3/2}(\Omega)} + \|\boldsymbol{q}\|_{H^{r-1/2}(\Omega)} + \|y\|_{H^{r+1/2}(\Omega)}).$$

*Furthermore, if $k = 0$ then*

$$\|u - u_h\|_{\varepsilon_h^\partial} \lesssim h^{1/2}(\|\boldsymbol{p}\|_{H^1(\Omega)} + \|z\|_{H^2(\Omega)} + \|\boldsymbol{q}\|_{H^{r-1/2}(\Omega)} + \|y\|_{H^{r+1/2}(\Omega)}),$$
$$\|y - y_h\|_{\mathcal{T}_h} \lesssim h^{1/2}(\|\boldsymbol{p}\|_{H^1(\Omega)} + \|z\|_{H^2(\Omega)} + \|\boldsymbol{q}\|_{H^{r-1/2}(\Omega)} + \|y\|_{H^{r+1/2}(\Omega)}),$$
$$\|\boldsymbol{p} - \boldsymbol{p}_h\|_{\mathcal{T}_h} \lesssim h^{1/2}(\|\boldsymbol{p}\|_{H^1(\Omega)} + \|z\|_{H^2(\Omega)} + \|\boldsymbol{q}\|_{H^{r-1/2}(\Omega)} + \|y\|_{H^{r+1/2}(\Omega)}),$$
$$\|z - z_h\|_{\mathcal{T}_h} \lesssim h^{1/2}(\|\boldsymbol{p}\|_{H^1(\Omega)} + \|z\|_{H^2(\Omega)} + \|\boldsymbol{q}\|_{H^{r-1/2}(\Omega)} + \|y\|_{H^{r+1/2}(\Omega)}).$$

As in Section two, when $k = 1$ the convergence rates are optimal for the control and the flux $\boldsymbol{q}$ and suboptimal for the other variables. When $k = 0$ the convergence rates for all variables are suboptimal with one exception: If $y_d \in L^2(\Omega)$ only so that $t^* = 0$, then $u \in H^{1/2}(\Gamma)$ only and the convergence rate for the control is optimal. Also, if $r_d$ or $r_\Omega$ is near $1/2$, then the convergence rate is nearly optimal for the control in the $k = 0$ case.

**4.2.1. Preliminary Material I.** We split the preliminary material required for the proof into two parts. First, we give a brief overview of material closely related to the preliminary material in Section 2: $L^2$ projections, HDG operators $\mathscr{B}_1$ and $\mathscr{B}_2$, and the well-posedness of the HDG equations.

As in Section 2, we use the standard $L^2$ projections $\boldsymbol{\Pi} : [L^2(\Omega)]^d \to \boldsymbol{V}_h$, $\Pi : L^2(\Omega) \to W_h$, and $P_M : L^2(\varepsilon_h) \to M_h$, which satisfy

$$
\begin{aligned}
(\boldsymbol{\Pi q}, \boldsymbol{r})_K &= (\boldsymbol{q}, \boldsymbol{r})_K, \qquad \forall \boldsymbol{r} \in [\mathcal{P}_k(K)]^d,\\
(\Pi y, w)_K &= (y, w)_K, \qquad \forall w \in \mathcal{P}_{k+1}(K),\\
\langle P_M m, \mu\rangle_e &= \langle m, \mu\rangle_e, \qquad \forall \mu \in \mathcal{P}_{k+1}(e).
\end{aligned} \tag{4.4}
$$

We have the following bounds:

$$
\|\boldsymbol{q} - \boldsymbol{\Pi q}\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{q}}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega}, \qquad \|y - \Pi y\|_{\mathcal{T}_h} \lesssim h^{s_y} \|y\|_{s_y,\Omega}, \tag{4.5a}
$$

$$
\|y - \Pi y\|_{\partial\mathcal{T}_h} \lesssim h^{s_y - \frac{1}{2}} \|y\|_{s_y,\Omega}, \quad \|w\|_{\partial\mathcal{T}_h} \lesssim h^{-\frac{1}{2}} \|w\|_{\mathcal{T}_h}, \ \forall w \in W_h, \tag{4.5b}
$$

and similar projection error bounds for $\boldsymbol{p}$ and $z$.

In this Section, we do not use the same HDG formulation for the analysis that we used in Section 2. We define the HDG operators $\mathscr{B}_1$ and $\mathscr{B}_2$ by

$$
\begin{aligned}
&\mathscr{B}_1(\boldsymbol{q}_h, y_h, \widehat{y}_h^o; \boldsymbol{r}_1, w_1, \mu_1)\\
&\qquad = (\boldsymbol{q}_h, \boldsymbol{r}_1)_{\mathcal{T}_h} - (y_h, \nabla\cdot \boldsymbol{r}_1)_{\mathcal{T}_h} + \langle \widehat{y}_h^o, \boldsymbol{r}_1 \cdot \boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} + (\nabla\cdot\boldsymbol{q}_h, w_1)_{\mathcal{T}_h}\\
&\qquad\quad - (\boldsymbol{\beta} y_h, \nabla w_1)_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta} y_h, w_1)_{\mathcal{T}_h} + \langle h^{-1} y_h + \tau_1 y_h, w_1\rangle_{\partial\mathcal{T}_h}\\
&\qquad\quad + \langle (\boldsymbol{\beta}\cdot\boldsymbol{n} - h^{-1} - \tau_1)\widehat{y}_h^o, w_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\qquad\quad - \langle \boldsymbol{q}_h\cdot\boldsymbol{n} + \boldsymbol{\beta}\cdot\boldsymbol{n}\widehat{y}_h^o + h^{-1}(y_h - \widehat{y}_h^o) + \tau_1(y_h - \widehat{y}_h^o), \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial},
\end{aligned} \tag{4.6}
$$

$$
\begin{aligned}
&\mathscr{B}_2(\boldsymbol{p}_h, z_h, \widehat{z}_h^o; \boldsymbol{r}_2, w_2, \mu_2)\\
&\qquad = (\boldsymbol{p}_h, \boldsymbol{r}_2)_{\mathcal{T}_h} - (z_h, \nabla\cdot \boldsymbol{r}_2)_{\mathcal{T}_h} + \langle \widehat{z}_h^o, \boldsymbol{r}_2 \cdot \boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} + (\nabla\cdot\boldsymbol{p}_h, w_2)\\
&\qquad\quad + (\boldsymbol{\beta} z_h, \nabla w_2)_{\mathcal{T}_h} + \langle h^{-1} z_h + \tau_2 z_h, w_2\rangle_{\partial\mathcal{T}_h}\\
&\qquad\quad - \langle (\boldsymbol{\beta}\cdot\boldsymbol{n} + h^{-1} + \tau_2)\widehat{z}_h^o, w_2\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\qquad\quad - \langle \boldsymbol{p}_h\cdot\boldsymbol{n} - \boldsymbol{\beta}\cdot\boldsymbol{n}\widehat{z}_h^o + h^{-1}(z_h - \widehat{z}_h^o) + \tau_2(z_h - \widehat{z}_h^o), \mu_2\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}.
\end{aligned} \tag{4.7}
$$

We emphasize that this is an equivalent definition to the one given in Section 2 that is more appropriate to obtain error estimates in the low regularity case.

We rewrite the HDG formulation of the optimality system (4.1) in terms of the HDG operators $\mathscr{B}_1$ and $\mathscr{B}_2$: find $(\boldsymbol{q}_h, \boldsymbol{p}_h, y_h, z_h, \widehat{y}_h^o, \widehat{z}_h^o, u_h) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times W_h \times M_h(o) \times M_h(o) \times M_h(\partial)$ satisfying

$$\begin{aligned}\mathscr{B}_1(\boldsymbol{q}_h, y_h, \widehat{y}_h^o; \boldsymbol{r}_1, w_1, \mu_1) &= (f, w_1)_{\mathcal{T}_h} - \langle u_h, \boldsymbol{r}_1 \cdot \boldsymbol{n}\rangle_{\varepsilon_h^\partial} \\ &\quad - \langle (\boldsymbol{\beta} \cdot \boldsymbol{n} - h^{-1} - \tau_1) u_h, w_1\rangle_{\varepsilon_h^\partial}, \qquad (4.8a)\end{aligned}$$

$$\mathscr{B}_2(\boldsymbol{p}_h, z_h, \widehat{z}_h^o; \boldsymbol{r}_2, w_2, \mu_2) = (y_h - y_d, w_2)_{\mathcal{T}_h}, \qquad (4.8b)$$

$$\gamma^{-1}\langle \boldsymbol{p}_h \cdot \boldsymbol{n} + h^{-1} z_h + \tau_2 z_h, \mu_3\rangle_{\varepsilon_h^\partial} = -\langle u_h, \mu_3\rangle_{\varepsilon_h^\partial}, \qquad (4.8c)$$

for all $(\boldsymbol{r}_1, \boldsymbol{r}_2, w_1, w_2, \mu_1, \mu_2, \mu_3) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times W_h \times M_h(o) \times M_h(o) \times M_h(\partial)$.

For the convenience of the reader, we recall three results proven in Section 2.

**Lemma 30** *For any $(\boldsymbol{v}_h, w_h, \mu_h) \in \boldsymbol{V}_h \times W_h \times M_h$, we have*

$$\begin{aligned}&\mathscr{B}_1(\boldsymbol{v}_h, w_h, \mu_h; \boldsymbol{v}_h, w_h, \mu_h) \\ &\quad = (\boldsymbol{v}_h, \boldsymbol{v}_h)_{\mathcal{T}_h} + \langle (h^{-1} + \tau_1 - \frac{1}{2}\boldsymbol{\beta} \cdot \boldsymbol{n})(w_h - \mu_h), w_h - \mu_h\rangle_{\partial\mathcal{T}_h \backslash \varepsilon_h^\partial} \\ &\quad\quad - \frac{1}{2}(\nabla \cdot \boldsymbol{\beta} w_h, w_h)_{\mathcal{T}_h} + \langle (h^{-1} + \tau_1 - \frac{1}{2}\boldsymbol{\beta} \cdot \boldsymbol{n}) w_h, w_h\rangle_{\varepsilon_h^\partial}, \\ &\mathscr{B}_2(\boldsymbol{v}_h, w_h, \mu_h; \boldsymbol{v}_h, w_h, \mu_h) \\ &\quad = (\boldsymbol{v}_h, \boldsymbol{v}_h)_{\mathcal{T}_h} + \langle (h^{-1} + \tau_2 + \frac{1}{2}\boldsymbol{\beta} \cdot \boldsymbol{n})(w_h - \mu_h), w_h - \mu_h\rangle_{\partial\mathcal{T}_h \backslash \varepsilon_h^\partial} \\ &\quad\quad - \frac{1}{2}(\nabla \cdot \boldsymbol{\beta} w_h, w_h)_{\mathcal{T}_h} + \langle (h^{-1} + \tau_2 + \frac{1}{2}\boldsymbol{\beta} \cdot \boldsymbol{n}) w_h, w_h\rangle_{\varepsilon_h^\partial}.\end{aligned}$$

**Lemma 31** *If **(A2)** holds, then*

$$\mathscr{B}_1(\boldsymbol{q}_h, y_h, \widehat{y}_h^o; \boldsymbol{p}_h, -z_h, -\widehat{z}_h^o) + \mathscr{B}_2(\boldsymbol{p}_h, z_h, \widehat{z}_h^o; -\boldsymbol{q}_h, y_h, \widehat{y}_h^o) = 0.$$

**Proposition 2** *If **(A2)** holds, there exists a unique solution of the HDG equations* (4.8).

**4.2.2. Preliminary Material II.** Next, we discuss preliminary material that is directly related to the low regularity case considered in this Section: the interpolation operators $\mathcal{I}_h^0$, $\mathcal{I}_h^1$, $\mathcal{I}_h$ and their properties.

Recall we assume the primary flux $\boldsymbol{q}$ only satisfies $\boldsymbol{q} \in [H^{r_{\boldsymbol{q}}}(\Omega)]^d \cap H(\text{div}, \Omega)$, where $r_{\boldsymbol{q}} \geq 0$. Therefore, the quantity $\|\boldsymbol{q} \cdot \boldsymbol{n} - \boldsymbol{\Pi q} \cdot \boldsymbol{n}\|_{\partial\mathcal{T}_h}$ is not well defined and the HDG analysis technique used in Section 2 is not applicable. We use analysis techniques from [44] to avoid using the $L^2$ boundary trace of $\boldsymbol{q}$. Let us introduce some notation first.

Define the $H^1$-conforming piecewise linear finite element space $W_h^c$ by

$$W_h^c := \{w_h^c \in H_0^1(\Omega) : w_h^c|_K \in \mathcal{P}_1(K),\ \forall K \in \mathcal{T}_h\}.$$

For any $K \in \mathcal{T}_h$, let $\lambda_1, \lambda_2, \ldots, \lambda_{d+1}$ denote the standard barycentric coordinate functions defined on the simplex $K$. Define

$$\mathbb{S}(K) := S_1(K) + S_2(K) + \cdots + S_{d+1}(K), \tag{4.9}$$

where

$$S_i(K) := \Big(\prod_{j \neq i} \lambda_j\Big) \text{span}\Big\{\prod_j \lambda_j^{\alpha_j} : \sum_j \alpha_j = k,\ \alpha_i = 0\Big\},\ \ i = 1, 2, \ldots, d+1.$$

Now we define the interpolations operators $\mathcal{I}_h^0, \mathcal{I}_h^1, \mathcal{I}_h$. First, define $m_K : L^2(\partial K) \to \mathbb{R}$ by

$$m_K(\mu) := \frac{1}{d+1} \sum_{e \in \partial K} \frac{1}{|e|} \int_e \mu, \tag{4.10}$$

where $|e|$ denotes the $d-1$ dimensional Hausdorff measure of $e$. Next, the interpolation operator $\mathcal{I}_h^0 : L^2(\varepsilon_h) \to W_h^c$ is defined as follows:

$$\mathcal{I}_h^0 \mu(a) = \begin{cases} \dfrac{1}{\#\omega_a} \displaystyle\sum_{K \in \omega_a} m_K(\mu) & \text{if } a \text{ is an interior node of } \mathcal{T}_h, \\ 0 & \text{if } a \text{ is a boundary node of } \mathcal{T}_h, \end{cases}$$

where $\omega_a := \{K \in \mathcal{T}_h : a \text{ is a vertex of } K\}$ and $\#\omega_a$ denotes the number of elements in $\omega_a$.

Next, the interpolation operator $\mathcal{I}_h^1$ on $L^2(\Omega) \times L^2(\varepsilon_h)$ is defined elementwise as follows: for each $K$,

$$\mathcal{I}_h^1(w, \mu)|_K := \mathcal{I}_K^1(w, \mu) = w_1 + w_2,$$

where $(w_1, w_2) \in \mathbb{S}(K) \times (\prod_j \lambda_j)\mathcal{P}_k(K)$ is uniquely determined by

$$\langle w_1, m\rangle_e = \langle \mu, m\rangle_e,$$

$$(w_2, n)_K = (w - w_1, n)_K,$$

for all $(m, n) \in \mathcal{P}_k(e) \times \mathcal{P}_k(K)$ and $e \in \partial K$.

Finally, for $(w, \mu) \in L^2(\Omega) \times L^2(\varepsilon_h)$, we define the third interpolation operator $\mathcal{I}_h$ by

$$\mathcal{I}_h(w, \mu) := \mathcal{I}_h^0 \mu + \mathcal{I}_h^1(w - \mathcal{I}_h^0 \mu, \mu - \mathcal{I}_h^0 \mu).$$

It is straightforward to verify that $\mathcal{I}_h$ and $\mathcal{I}_h^1$ have the following properties; see [44].

**Lemma 32** *For any $(w, \mu) \in L^2(\Omega) \times L^2(\varepsilon_h)$ and $K \in \mathcal{T}_h$, we have*

$$(\mathcal{I}_h(w, \mu), n)_K = (w, n)_K, \tag{4.11a}$$

$$\langle \mathcal{I}_h(w, \mu), m\rangle_{\partial K} = \langle \mu, m\rangle_{\partial K}, \tag{4.11b}$$

*for all* $(m, n) \in \mathcal{P}_k(e) \times \mathcal{P}_k(K)$ *and* $e \in \partial K$*, and*

$$\|\mathcal{I}_h^1(w, \mu)\|_K \lesssim \|w\|_K + h^{\frac{1}{2}} \|\mu\|_{\partial K}. \tag{4.12}$$

*Moreover, if* $\mu|_\Gamma = 0$*, we have*

$$\mathcal{I}_h(w, \mu) \in H_0^1(\Omega). \tag{4.13}$$

In the next three lemmas, we assume $(\boldsymbol{v}_h, w_h, \mu_h) \in \boldsymbol{V}_h \times W_h \times M_h$ satisfy

$$(\boldsymbol{v}_h, \boldsymbol{r})_{\mathcal{T}_h} - (w_h, \nabla \cdot \boldsymbol{r})_{\mathcal{T}_h} + \langle \mu_h, \boldsymbol{r} \cdot \boldsymbol{n} \rangle_{\partial \mathcal{T}_h} = 0, \tag{4.14}$$

for all $\boldsymbol{r} \in \boldsymbol{V}_h$.

We begin with a key inequality; see Section 2 Lemma 4.7 and also [56].

**Lemma 33** *If* $(\boldsymbol{v}_h, w_h, \mu_h) \in \boldsymbol{V}_h \times W_h \times M_h$ *satisfy* (4.14)*, then*

$$\|\nabla w_h\|_{\mathcal{T}_h} \lesssim \|\boldsymbol{v}_h\|_{\mathcal{T}_h} + h^{-\frac{1}{2}} \|w_h - \mu_h\|_{\partial \mathcal{T}_h}. \tag{4.15}$$

The next two results are similar to Lemma 3.4 and Lemma 3.6 in [44]. Here, we have a different space $M_h$ (with polynomials of degree $k+1$ instead of $k$) and we do not have a variable diffusion coefficient. However, the proofs of the next two results are very similar to the proofs in [44] and are omitted.

**Lemma 34** *If* $(\boldsymbol{v}_h, w_h, \mu_h) \in \boldsymbol{V}_h \times W_h \times M_h$ *satisfy* (4.14)*, then*

$$\begin{aligned} h^{-1} \sum_{K \in \mathcal{T}_h} \|w_h - m_K(\mu_h)\|_K + h^{-\frac{1}{2}} \sum_{K \in \mathcal{T}_h} \|\mu_h - m_K(\mu_h)\|_{\partial K} \\ \lesssim \|\boldsymbol{v}_h\|_{\mathcal{T}_h} + h^{-\frac{1}{2}} \|w_h - \mu_h\|_{\partial \mathcal{T}_h}. \end{aligned} \tag{4.16}$$

**Lemma 35** *If* $(\boldsymbol{v}_h, w_h, \mu_h) \in \boldsymbol{V}_h \times W_h \times M_h$ *satisfy* (4.14)*, then*

$$\|\nabla \mathcal{I}_h(w_h, \mu_h)\|_{\mathcal{T}_h} \lesssim \|\boldsymbol{v}_h\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|w_h - \mu_h\|_{\partial\mathcal{T}_h}, \tag{4.17a}$$

$$h^{-1}\|w_h - \mathcal{I}_h(w_h, \mu_h)\|_{\mathcal{T}_h} \lesssim \|\boldsymbol{v}_h\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|w_h - \mu_h\|_{\partial\mathcal{T}_h}. \tag{4.17b}$$

**4.2.3. Proof of Main Result.** Now we move to the proof of the error estimates. We follow the strategy of Section 2 and split the proof into seven steps. In the first five steps we use the rewriting of operators $\mathscr{B}_1$ and $\mathscr{B}_2$ in an explicit way and the proofs are different from the corresponding ones of Section 2. Steps 6 and 7 use the properties of $\mathscr{B}_1$ and $\mathscr{B}_2$ recalled in 30 and 31 and are very similar to Steps 6 and 7 in the high regularity case in Section 2. We include these proofs here to make this Section self-contained.

We first bound the error between the solution of the mixed form (3.18a)-(3.18d) of the optimality system and the solution

$$(\boldsymbol{q}_h(u), \boldsymbol{p}_h(u), y_h(u), z_h(u), \widehat{y}_h^o(u), \widehat{z}_h^o(u)) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times W_h \times M_h(o) \times M_h(o)$$

of the auxiliary problem

$$\begin{aligned}\mathscr{B}_1(\boldsymbol{q}_h(u), y_h(u), \widehat{y}_h^o(u); \boldsymbol{r}_1, w_1, \mu_1) &= (f, w_1)_{\mathcal{T}_h} - \langle P_M u, \boldsymbol{r}_1 \cdot \boldsymbol{n}\rangle_{\varepsilon_h^\partial} \\ &\quad - \langle(\boldsymbol{\beta} \cdot \boldsymbol{n} - h^{-1} - \tau_1) P_M u, w_1\rangle_{\varepsilon_h^\partial},\end{aligned} \tag{4.18a}$$

$$\mathscr{B}_2(\boldsymbol{p}_h(u), z_h(u), \widehat{z}_h^o(u); \boldsymbol{r}_2, w_2, \mu_2) = (y_h(u) - y_d, w_2)_{\mathcal{T}_h}, \tag{4.18b}$$

for all $(\boldsymbol{r}_1, \boldsymbol{r}_2, w_1, w_2, \mu_1, \mu_2) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times W_h \times M_h(o) \times M_h(o)$. As in Part I, we use the notation

$$\begin{aligned}
\delta^{\boldsymbol{q}} &= \boldsymbol{q} - \boldsymbol{\Pi q}, & \varepsilon_h^{\boldsymbol{q}} &= \boldsymbol{\Pi q} - \boldsymbol{q}_h(u),\\
\delta^y &= y - \Pi y, & \varepsilon_h^y &= \Pi y - y_h(u),\\
\delta^{\widehat{y}} &= y - P_M y, & \varepsilon_h^{\widehat{y}} &= P_M y - \widehat{y}_h(u),\\
\widehat{\boldsymbol{\delta}}_1 &= \boldsymbol{\beta} \cdot \boldsymbol{n} \delta^{\widehat{y}} + (h^{-1} + \tau_1)(\delta^y - \delta^{\widehat{y}}),
\end{aligned} \tag{4.19}$$

where $\widehat{y}_h(u) = \widehat{y}_h^o(u)$ on $\varepsilon_h^o$ and $\widehat{y}_h(u) = P_M u$ on $\varepsilon_h^\partial$. This definition gives $\varepsilon_h^{\widehat{y}} = 0$ on $\varepsilon_h^\partial$.

Step 1: The error equation for part 1 of the auxiliary problem (4.18a)

**Lemma 36** *We have*

$$\begin{aligned}
&\mathscr{B}_1(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^y, \varepsilon_h^{\widehat{y}}, \boldsymbol{r}_1, w_1, \mu_1)\\
&\quad = -(\nabla \cdot \delta^{\boldsymbol{q}}, w_1)_{\mathcal{T}_h} - \langle \boldsymbol{\Pi q} \cdot \boldsymbol{n}, \mu_1 \rangle_{\mathcal{T}_h \backslash \varepsilon_h^\partial} + (\boldsymbol{\beta} \delta^y, \nabla w_1)_{\mathcal{T}_h}\\
&\qquad + (\nabla \cdot \boldsymbol{\beta} \delta^y, w_1)_{\mathcal{T}_h} - \langle \widehat{\boldsymbol{\delta}}_1, w_1 \rangle_{\partial \mathcal{T}_h} + \langle \widehat{\boldsymbol{\delta}}_1, \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial}.
\end{aligned} \tag{4.20}$$

**Proof:** Using the definition of $\mathscr{B}_1$ in (4.6) gives

$$\begin{aligned}
&\mathscr{B}_1(\boldsymbol{\Pi q}, \Pi y, P_M y, \boldsymbol{r}_1, w_1, \mu_1)\\
&\quad = (\boldsymbol{\Pi q}, \boldsymbol{r_1})_{\mathcal{T}_h} - (\Pi y, \nabla \cdot \boldsymbol{r_1})_{\mathcal{T}_h} + \langle P_M y, \boldsymbol{r_1} \cdot \boldsymbol{n} \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial}\\
&\qquad + (\nabla \cdot \boldsymbol{\Pi q}, w_1)_{\mathcal{T}_h} - (\boldsymbol{\beta} \Pi y, \nabla w_1)_{\mathcal{T}_h} - (\nabla \cdot \boldsymbol{\beta} \Pi y, w_1)_{\mathcal{T}_h}\\
&\qquad + \langle (h^{-1} + \tau_1) \Pi y, w_1 \rangle_{\partial \mathcal{T}_h} + (\boldsymbol{\beta} \cdot \boldsymbol{n} - h^{-1} - \tau_1) P_M y, w_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial}\\
&\qquad - \langle \boldsymbol{\Pi q} \cdot \boldsymbol{n} + \boldsymbol{\beta} \cdot \boldsymbol{n} P_M y + (h^{-1} + \tau_1)(\Pi y - P_M y), \mu_1 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^\partial}.
\end{aligned}$$

Using properties of the $L^2$ projections (4.4) gives

$$
\begin{aligned}
&\mathscr{B}_1(\boldsymbol{\Pi q}, \Pi y, P_M y, \boldsymbol{r}_1, w_1, \mu_1)\\
&\qquad = (\boldsymbol{q}, \boldsymbol{r_1})_{\mathcal{T}_h} - (y, \nabla\cdot \boldsymbol{r_1})_{\mathcal{T}_h} + \langle y, \boldsymbol{r_1}\cdot \boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\qquad\quad + (\nabla\cdot\boldsymbol{q}, w_1)_{\mathcal{T}_h} - (\nabla\cdot\delta^{\boldsymbol{q}}, w_1)_{\mathcal{T}_h} - (\boldsymbol{\beta} y, \nabla w_1)_{\mathcal{T}_h} + (\boldsymbol{\beta}\delta^y, \nabla w_1)_{\mathcal{T}_h}\\
&\qquad\quad - (\nabla\cdot\boldsymbol{\beta} y, w_1)_{\mathcal{T}_h} + (\nabla\cdot\boldsymbol{\beta}\delta^y, w_1)_{\mathcal{T}_h} + \langle (h^{-1}+\tau_1)\Pi y, w_1\rangle_{\partial\mathcal{T}_h}\\
&\qquad\quad + \langle \boldsymbol{\beta}\cdot\boldsymbol{n} y, w_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} - \langle \boldsymbol{\beta}\cdot\boldsymbol{n}\delta^{\widehat{y}}, w_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} - \langle (h^{-1}+\tau_1) P_M y, w_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\qquad\quad - \langle \boldsymbol{\Pi q}\cdot\boldsymbol{n}, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} - \langle \boldsymbol{\beta}\cdot\boldsymbol{n} y, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} + \langle \boldsymbol{\beta}\cdot\boldsymbol{n}\delta^{\widehat{y}}, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\qquad\quad + \langle (h^{-1}+\tau_1)(\delta^y - \delta^{\widehat{y}}), \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}.
\end{aligned}
$$

The exact state $y$ and flux $\boldsymbol{q}$ satisfy

$$
\begin{aligned}
(\boldsymbol{q}, \boldsymbol{r}_1)_{\mathcal{T}_h} - (y, \nabla\cdot\boldsymbol{r}_1)_{\mathcal{T}_h} + \langle y, \boldsymbol{r}_1\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} &= -\langle u, \boldsymbol{r}_1\cdot\boldsymbol{n}\rangle_{\varepsilon_h^\partial},\\
(\nabla\cdot\boldsymbol{q}, w_1)_{\mathcal{T}_h} - (\boldsymbol{\beta} y, \nabla w_1)_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta} y, w_1)_{\mathcal{T}_h} &\\
+ \langle \boldsymbol{\beta}\cdot\boldsymbol{n} y, w_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} &= -\langle \boldsymbol{\beta}\cdot\boldsymbol{n} u, w_1\rangle_{\varepsilon_h^\partial} + (f, w_1)_{\mathcal{T}_h},
\end{aligned}
$$

for all $(\boldsymbol{r}_1, w_1) \in \boldsymbol{V}_h \times W_h$. This gives

$$
\begin{aligned}
&\mathscr{B}_1(\boldsymbol{\Pi q}, \Pi y, P_M y, \boldsymbol{r}_1, w_1, \mu_1)\\
&\quad = -\langle u, \boldsymbol{r}_1\cdot\boldsymbol{n}\rangle_{\varepsilon_h^\partial} - \langle \boldsymbol{\beta}\cdot\boldsymbol{n} u, w_1\rangle_{\varepsilon_h^\partial} + (f, w_1)_{\mathcal{T}_h} - (\nabla\cdot\delta^{\boldsymbol{q}}, w_1)_{\mathcal{T}_h} + (\boldsymbol{\beta}\delta^y, \nabla w_1)_{\mathcal{T}_h}\\
&\qquad + (\nabla\cdot\boldsymbol{\beta}\delta^y, w_1)_{\mathcal{T}_h} + \langle (h^{-1}+\tau_1)\Pi y, w_1\rangle_{\partial\mathcal{T}_h}\\
&\qquad - \langle \boldsymbol{\beta}\cdot\boldsymbol{n}\delta^{\widehat{y}}, w_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} - \langle (h^{-1}+\tau_1) P_M y, w_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} - \langle \boldsymbol{\Pi q}\cdot\boldsymbol{n}, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\qquad + \langle \boldsymbol{\beta}\cdot\boldsymbol{n}\delta^{\widehat{y}}, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} + \langle (h^{-1}+\tau_1)(\delta^y - \delta^{\widehat{y}}), \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}.
\end{aligned}
$$

Here we used $\langle \boldsymbol{\beta}\cdot\boldsymbol{n}y, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} = 0$, which holds since $\mu_1$ is a single-valued function on the interior edges. Subtracting part 1 of the auxiliary problem (4.18a) from the above equality gives the result:

$$
\begin{aligned}
&\mathscr{B}_1(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^{y}, \varepsilon_h^{\widehat{y}}, \boldsymbol{r}_1, w_1, \mu_1)\\
&\quad= -(\nabla\cdot\delta^{\boldsymbol{q}}, w_1)_{\mathcal{T}_h} + (\boldsymbol{\beta}\delta^y, \nabla w_1)_{\mathcal{T}_h} + (\nabla\cdot\boldsymbol{\beta}\delta^y, w_1)_{\mathcal{T}_h}\\
&\qquad + \langle (h^{-1}+\tau_1)\Pi y, w_1\rangle_{\partial\mathcal{T}_h} - \langle \boldsymbol{\beta}\cdot\boldsymbol{n}\delta^{\widehat{y}}, w_1\rangle_{\partial\mathcal{T}_h} - \langle (h^{-1}+\tau_1)P_M y, w_1\rangle_{\partial\mathcal{T}_h}\\
&\qquad - \langle \boldsymbol{\Pi q}\cdot\boldsymbol{n}, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} + \langle \boldsymbol{\beta}\cdot\boldsymbol{n}\delta^{\widehat{y}}, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} + \langle (h^{-1}+\tau_1)(\delta^y-\delta^{\widehat{y}}), \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\quad= -(\nabla\cdot\delta^{\boldsymbol{q}}, w_1)_{\mathcal{T}_h} + (\boldsymbol{\beta}\delta^y, \nabla w_1)_{\mathcal{T}_h} + (\nabla\cdot\boldsymbol{\beta}\delta^y, w_1)_{\mathcal{T}_h}\\
&\qquad - \langle \boldsymbol{\Pi q}\cdot\boldsymbol{n}, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial} - \langle \widehat{\boldsymbol{\delta}}_1, w_1\rangle_{\partial\mathcal{T}_h} + \langle \widehat{\boldsymbol{\delta}}_1, \mu_1\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}.
\end{aligned}
$$

Step 2: Estimate for $\varepsilon_h^{\boldsymbol{q}}$

**Lemma 37** *We have*

$$
\|\varepsilon_h^{\boldsymbol{q}}\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\varepsilon_h^{y} - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{q}}}\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y-1}\|y\|_{s^y,\Omega}. \tag{4.21}
$$

**Proof:** First, take $(\boldsymbol{v}_h, w_h, \mu_h) = (\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^{y}, \varepsilon_h^{\widehat{y}})$ in the key inequality in Lemma 33 to obtain

$$
\|\nabla\varepsilon_h^{y}\|_{\mathcal{T}_h} \lesssim \|\varepsilon_h^{\boldsymbol{q}}\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\varepsilon_h^{y} - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h}. \tag{4.22}
$$

Next, since $\varepsilon_h^{\widehat{y}} = 0$ on $\varepsilon_h^\partial$, the energy identity for $\mathscr{B}_1$ in 30 gives

$$
\begin{aligned}
&\mathscr{B}(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^{y}, \varepsilon_h^{\widehat{y}}, \varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^{y}, \varepsilon_h^{\widehat{y}})\\
&\quad= (\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^{\boldsymbol{q}})_{\mathcal{T}_h} + \|(h^{-1}+\tau_1 - \frac{1}{2}\boldsymbol{\beta}\cdot\boldsymbol{n})^{\frac{1}{2}}(\varepsilon_h^{y} - \varepsilon_h^{\widehat{y}})\|_{\partial\mathcal{T}_h}^2 + \frac{1}{2}\|(-\nabla\cdot\boldsymbol{\beta})^{\frac{1}{2}}\varepsilon_h^{y}\|_{\mathcal{T}_h}^2.
\end{aligned}
$$

Take $(\boldsymbol{r}_1, w_1, \mu_1) = (\boldsymbol{\varepsilon}_h^{\boldsymbol{q}}, \varepsilon_h^y, \varepsilon_h^{\widehat{y}})$ in the error equation (4.20) in Lemma 36 to obtain

$$\begin{aligned}
(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^{\boldsymbol{q}})_{\mathcal{T}_h} &+ \|(h^{-1} + \tau_1 - \frac{1}{2}\boldsymbol{\beta}\cdot\boldsymbol{n})^{\frac{1}{2}}(\varepsilon_h^y - \varepsilon_h^{\widehat{y}})\|_{\partial\mathcal{T}_h}^2 + \frac{1}{2}\|(-\nabla\cdot\boldsymbol{\beta})^{\frac{1}{2}}\varepsilon_h^y\|_{\mathcal{T}_h}^2 \\
&= -(\nabla\cdot\delta^{\boldsymbol{q}}, \varepsilon_h^y)_{\mathcal{T}_h} - \langle \boldsymbol{\Pi q}\cdot\boldsymbol{n}, \varepsilon_h^{\widehat{y}}\rangle_{\partial\mathcal{T}_h} \\
&\quad + (\boldsymbol{\beta}\delta^y, \nabla\varepsilon_h^y)_{\mathcal{T}_h} + (\nabla\cdot\boldsymbol{\beta}\delta^y, \varepsilon_h^y)_{\mathcal{T}_h} - \langle \widehat{\boldsymbol{\delta}}_1, \varepsilon_h^y - \varepsilon_h^{\widehat{y}}\rangle_{\partial\mathcal{T}_h} \\
&=: T_1 + T_2 + T_3 + T_4.
\end{aligned} \tag{4.23}$$

We rewrite the term $T_1$ using the interpolation operator $\mathcal{I}_h$:

$$\begin{aligned}
T_1 &= -(\nabla\cdot\delta^{\boldsymbol{q}}, \varepsilon_h^y)_{\mathcal{T}_h} - \langle \boldsymbol{\Pi q}\cdot\boldsymbol{n}, \varepsilon_h^{\widehat{y}}\rangle_{\partial\mathcal{T}_h} \\
&= -(\nabla\cdot\boldsymbol{q}, \varepsilon_h^y)_{\mathcal{T}_h} + (\nabla\cdot\boldsymbol{\Pi q}, \varepsilon_h^y)_{\mathcal{T}_h} - \langle \boldsymbol{\Pi q}\cdot\boldsymbol{n}, \varepsilon_h^{\widehat{y}}\rangle_{\partial\mathcal{T}_h} \\
&= -(\nabla\cdot\boldsymbol{q}, \varepsilon_h^y - \mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}}))_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{q}, \mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}}))_{\mathcal{T}_h} \\
&\quad + (\nabla\cdot\boldsymbol{\Pi q}, \varepsilon_h^y) - \langle \boldsymbol{\Pi q}\cdot\boldsymbol{n}, \varepsilon_h^{\widehat{y}}\rangle_{\partial\mathcal{T}_h} \\
&= -(\nabla\cdot\boldsymbol{q}, \varepsilon_h^y - \mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}}))_{\mathcal{T}_h} + (\boldsymbol{q}, \nabla\mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}}))_{\mathcal{T}_h} \\
&\quad + (\nabla\cdot\boldsymbol{\Pi q}, \varepsilon_h^y) - \langle \boldsymbol{\Pi q}\cdot\boldsymbol{n}, \varepsilon_h^{\widehat{y}}\rangle_{\partial\mathcal{T}_h} \\
&= -(\nabla\cdot\boldsymbol{q}, \varepsilon_h^y - \mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}}))_{\mathcal{T}_h} + (\delta^{\boldsymbol{q}}, \nabla\mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}}))_{\mathcal{T}_h} \\
&\quad + (\boldsymbol{\Pi q}, \nabla\mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}}))_{\mathcal{T}_h} + (\nabla\cdot\boldsymbol{\Pi q}, \varepsilon_h^y) - \langle \boldsymbol{\Pi q}\cdot\boldsymbol{n}, \varepsilon_h^{\widehat{y}}\rangle_{\partial\mathcal{T}_h} \\
&= -(\nabla\cdot\boldsymbol{q}, \varepsilon_h^y - \mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}}))_{\mathcal{T}_h} + (\delta^{\boldsymbol{q}}, \nabla\mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}}))_{\mathcal{T}_h}.
\end{aligned}$$

The last step holds since

$$\begin{aligned}
(\boldsymbol{\Pi q}, \nabla\mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}}))_{\mathcal{T}_h} &= \langle \boldsymbol{\Pi q}\cdot\boldsymbol{n}, \mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}})\rangle_{\partial\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\Pi q}, \mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}}))_{\mathcal{T}_h} \\
&= \langle \boldsymbol{\Pi q}\cdot\boldsymbol{n}, \varepsilon_h^{\widehat{y}}\rangle_{\partial\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\Pi q}, \varepsilon_h^y)_{\mathcal{T}_h}.
\end{aligned}$$

This implies

$$
\begin{aligned}
T_1 &\leq \|\nabla\cdot\boldsymbol{q}\|_{\mathcal{T}_h}\|\varepsilon_h^y - \mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}})\|_{\mathcal{T}_h} + \|\delta^{\boldsymbol{q}}\|_{\mathcal{T}_h}\|\nabla\mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}})\|_{\mathcal{T}_h} \\
&\lesssim h(\|\varepsilon_h^{\boldsymbol{q}}\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h}) + h^{s_{\boldsymbol{q}}}\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega}(\|\varepsilon_h^{\boldsymbol{q}}\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h}) \\
&\lesssim h^{s_{\boldsymbol{q}}}\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega}(\|\varepsilon_h^{\boldsymbol{q}}\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h}).
\end{aligned}
$$

Note that we used $s_{\boldsymbol{q}} \in [0,1]$.

For the terms $T_2$, $T_3$, and $T_4$, apply (4.22) and Young's inequality to obtain

$$
\begin{aligned}
T_2 &= (\boldsymbol{\beta}\delta^y, \nabla\varepsilon_h^y)_{\mathcal{T}_h} \leq C\|\delta^y\|_{\mathcal{T}_h}^2 + \frac{1}{4}\|\varepsilon_h^{\boldsymbol{q}}\|_{\mathcal{T}_h}^2 + \frac{1}{4h}\|\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h}^2, \\
T_3 &= (\nabla\cdot\boldsymbol{\beta}\delta^y, \varepsilon_h^y)_{\mathcal{T}_h} \leq C\|\delta^y\|_{\mathcal{T}_h}^2 + \frac{1}{2}\|(-\nabla\cdot\boldsymbol{\beta})^{\frac{1}{2}}\varepsilon_h^y\|_{\mathcal{T}_h}^2, \\
T_4 &= -\langle\widehat{\boldsymbol{\delta}}_1, \varepsilon_h^y - \varepsilon_h^{\widehat{y}}\rangle_{\partial\mathcal{T}_h} \leq 4h\|\widehat{\boldsymbol{\delta}}_1\|_{\partial\mathcal{T}_h}^2 + \frac{1}{4h}\|\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h}^2.
\end{aligned}
$$

Summing the estimates for $\{T_i\}_{i=1}^4$ gives the result.

**Remark 5** *In Section 2, we defined* $\widehat{\boldsymbol{\delta}}_1 = \delta^{\boldsymbol{q}}\cdot\boldsymbol{n} + \boldsymbol{\beta}\cdot\boldsymbol{n}\delta^{\widehat{y}} + (h^{-1}+\tau_1)(\delta^y - \delta^{\widehat{y}})$. *It is not meaningful to estimate* $\|\widehat{\boldsymbol{\delta}}_1\|_{\partial\mathcal{T}_h}$ *if we only assume* $r_{\boldsymbol{q}} \geq 0$. *In this Section, we have* $\widehat{\boldsymbol{\delta}}_1 = \boldsymbol{\beta}\cdot\boldsymbol{n}\delta^{\widehat{y}} + (h^{-1}+\tau_1)(\delta^y - \delta^{\widehat{y}})$*, and we can estimate* $\|\widehat{\boldsymbol{\delta}}_1\|_{\partial\mathcal{T}_h}$.

Step 3: Estimate for $\varepsilon_h^y$ by a duality argument

Next, for any $\Theta$ in $L^2(\Omega)$ we consider the dual problem

$$
\begin{aligned}
\boldsymbol{\Phi} - \nabla\Psi &= 0 && \text{in } \Omega, \\
\nabla\cdot\boldsymbol{\Phi} + \nabla\cdot(\boldsymbol{\beta}\Psi) &= \Theta && \text{in } \Omega, \\
\Psi &= 0 && \text{on } \partial\Omega.
\end{aligned}
\tag{4.24}
$$

Since the domain $\Omega$ is convex, we have the regularity estimate

$$
\|\boldsymbol{\Phi}\|_{1,\Omega} + \|\Psi\|_{2,\Omega} \leq C_{\text{reg}}\|\Theta\|_{\Omega}. \tag{4.25}
$$

We use the following notation in the next proof for the estimate of $\varepsilon_h^y$:

$$\delta^{\boldsymbol{\Phi}} = \boldsymbol{\Phi} - \boldsymbol{\Pi\Phi}, \quad \delta^{\Psi} = \Psi - \Pi\Psi, \quad \delta^{\widehat{\Psi}} = \Psi - P_M\Psi. \tag{4.26}$$

**Lemma 38** *We have*

$$\|\varepsilon_h^y\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{q}}+1} \|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y} \|y\|_{s^y,\Omega} .$$

**Proof:** We take $\Theta = -\varepsilon_h^y$ in the dual problem (4.24) and $(\boldsymbol{r}_1, w_1, \mu_1) = (\boldsymbol{\Pi\Phi}, \Pi\Psi, P_M\Psi)$ in the error equation (4.20) in Lemma 36. Since $\Psi = 0$ on $\varepsilon_h^\partial$, we have

$$\begin{aligned}
&\mathscr{B}_1(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^y, \varepsilon_h^{\widehat{y}}; \boldsymbol{\Pi\Phi}, \Pi\Psi, P_M\Psi)\\
&\quad = (\varepsilon_h^{\boldsymbol{q}}, \boldsymbol{\Pi\Phi})_{\mathcal{T}_h} - (\varepsilon_h^y, \nabla\cdot\boldsymbol{\Pi\Phi})_{\mathcal{T}_h} + \langle \varepsilon_h^{\widehat{y}}, \boldsymbol{\Pi\Phi}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^\partial}\\
&\qquad + (\nabla\cdot\varepsilon_h^{\boldsymbol{q}}, \Pi\Psi)_{\mathcal{T}_h} - (\boldsymbol{\beta}\varepsilon_h^y, \nabla\Pi\Psi)_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta}\varepsilon_h^y, \Pi\Psi)_{\mathcal{T}_h} + \langle (h^{-1}+\tau_1)\varepsilon_h^y, \Pi\Psi\rangle_{\partial\mathcal{T}_h}\\
&\qquad + \langle (\boldsymbol{\beta}\cdot\boldsymbol{n} - h^{-1} - \tau_1)\varepsilon_h^{\widehat{y}}, \Pi\Psi\rangle_{\partial\mathcal{T}_h}\\
&\qquad - \langle \varepsilon_h^{\boldsymbol{q}}\cdot\boldsymbol{n} + \boldsymbol{\beta}\cdot\boldsymbol{n}\varepsilon_h^{\widehat{y}} + (h^{-1}+\tau_1)(\varepsilon_h^y - \varepsilon_h^{\widehat{y}}), P_M\Psi\rangle_{\partial\mathcal{T}_h}\\
&\quad = (\varepsilon_h^{\boldsymbol{q}}, \boldsymbol{\Phi})_{\mathcal{T}_h} - (\varepsilon_h^y, \nabla\cdot\boldsymbol{\Phi})_{\mathcal{T}_h} + (\varepsilon_h^y, \nabla\cdot\delta^{\boldsymbol{\Phi}})_{\mathcal{T}_h} + \langle \varepsilon_h^{\widehat{y}}, \boldsymbol{\Pi\Phi}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h} - (\varepsilon_h^{\boldsymbol{q}}, \nabla\Psi)_{\mathcal{T}_h}\\
&\qquad + \langle \varepsilon_h^{\boldsymbol{q}}\cdot\boldsymbol{n}, \Psi\rangle_{\partial\mathcal{T}_h} - (\boldsymbol{\beta}\varepsilon_h^y, \nabla\Psi)_{\mathcal{T}_h} + (\boldsymbol{\beta}\varepsilon_h^y, \nabla\delta^{\Psi})_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta}\varepsilon_h^y, \Psi)_{\mathcal{T}_h}\\
&\qquad + (\nabla\cdot\boldsymbol{\beta}\varepsilon_h^y, \delta^{\Psi})_{\mathcal{T}_h} - \langle \varepsilon_h^{\boldsymbol{q}}\cdot\boldsymbol{n}, P_M\Psi\rangle_{\partial\mathcal{T}_h} - \langle \boldsymbol{\beta}\cdot\boldsymbol{n}\varepsilon_h^{\widehat{y}}, \delta^{\Psi}\rangle_{\partial\mathcal{T}_h}\\
&\qquad - \langle (h^{-1}+\tau_1)(\varepsilon_h^y - \varepsilon_h^{\widehat{y}}), \delta^{\Psi} - \delta^{\widehat{\Psi}}\rangle_{\partial\mathcal{T}_h}.
\end{aligned}$$

Here we used $\langle \boldsymbol{\beta}\cdot\boldsymbol{n}\varepsilon_h^{\widehat{y}}, \Psi\rangle_{\partial\mathcal{T}_h} = 0$ and $\langle \boldsymbol{\beta}\cdot\boldsymbol{n}\varepsilon_h^{\widehat{y}}, P_M\Psi\rangle_{\partial\mathcal{T}_h} = 0$, which both hold since $\varepsilon_h^{\widehat{y}}$ is a single-valued function on interior edges and $\varepsilon_h^{\widehat{y}} = 0$ on $\varepsilon_h^\partial$.

By the same argument as in the proof of Lemma 37 for the term $T_1$, we have

$$\begin{aligned}
&(\varepsilon_h^y, \nabla\cdot\delta^{\boldsymbol{\Phi}})_{\mathcal{T}_h} + \langle \varepsilon_h^{\widehat{y}}, \boldsymbol{\Pi\Phi}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h}\\
&\qquad\qquad = (\varepsilon_h^y - \mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}}), \nabla\cdot\boldsymbol{\Phi})_{\mathcal{T}_h} - (\nabla\mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}}), \delta^{\boldsymbol{\Phi}})_{\mathcal{T}_h}.
\end{aligned}$$

Next, integration by parts gives

$$(\boldsymbol{\beta}\varepsilon_h^y, \nabla\delta^\Psi)_{\mathcal{T}_h} = \langle \boldsymbol{\beta}\cdot\boldsymbol{n}\varepsilon_h^y, \delta^\Psi\rangle_{\partial\mathcal{T}_h} - (\nabla\cdot\boldsymbol{\beta}\varepsilon_h^y, \delta^\Psi)_{\mathcal{T}_h} - (\boldsymbol{\beta}\cdot\nabla\varepsilon_h^y, \delta^\Psi)_{\mathcal{T}_h}.$$

This implies

$$\begin{aligned}
&\mathscr{B}_1(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^y, \varepsilon_h^{\widehat{y}}; \boldsymbol{\Pi\Phi}, \Pi\Psi, P_M\Psi)\\
&\quad = \|\varepsilon_h^y\|_{\mathcal{T}_h}^2 + \langle \boldsymbol{\beta}\cdot\boldsymbol{n}(\varepsilon_h^y - \varepsilon_h^{\widehat{y}}), \delta^\Psi\rangle_{\partial\mathcal{T}_h} - (\nabla\varepsilon_h^y, \boldsymbol{\beta}\delta^\Psi)_{\mathcal{T}_h}\\
&\qquad + (\varepsilon_h^y - \mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}}), \nabla\cdot\delta^{\boldsymbol{\Phi}})_{\mathcal{T}_h} - (\nabla\mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}}), \delta^{\boldsymbol{\Phi}})_{\mathcal{T}_h}\\
&\qquad - \langle h^{-1}(\varepsilon_h^y - \varepsilon_h^{\widehat{y}}) + \tau_1(\varepsilon_h^y - \varepsilon_h^{\widehat{y}}), \delta^\Psi - \delta^{\widehat{\Psi}}\rangle_{\partial\mathcal{T}_h}.
\end{aligned}$$

Also, since $\Psi = 0$ on $\varepsilon_h^\partial$, the error equation (4.20) in Lemma 36 gives

$$\begin{aligned}
&\mathscr{B}_1(\varepsilon_h^{\boldsymbol{q}}, \varepsilon_h^y, \varepsilon_h^{\widehat{y}}; \boldsymbol{\Pi\Phi}, \Pi\Psi, P_M\Psi)\\
&\quad = -(\nabla\cdot\delta^{\boldsymbol{q}}, \Pi\Psi)_{\mathcal{T}_h} - \langle \boldsymbol{\Pi q}\cdot\boldsymbol{n}, P_M\Psi\rangle_{\mathcal{T}_h}\\
&\qquad + (\boldsymbol{\beta}\delta^y, \nabla\Pi\Psi)_{\mathcal{T}_h} + (\nabla\cdot\boldsymbol{\beta}\delta^y, \Pi\Psi)_{\mathcal{T}_h} - \langle \widehat{\boldsymbol{\delta}}_1, \Pi\Psi - P_M\Psi\rangle_{\partial\mathcal{T}_h}\\
&\quad = -(\nabla\cdot\boldsymbol{q}, \Pi\Psi)_{\mathcal{T}_h} + (\nabla\cdot\boldsymbol{\Pi q}, \Psi)_{\mathcal{T}_h} - \langle \boldsymbol{\Pi q}\cdot\boldsymbol{n}, \Psi\rangle_{\mathcal{T}_h}\\
&\qquad + (\boldsymbol{\beta}\delta^y, \nabla\Pi\Psi)_{\mathcal{T}_h} + (\nabla\cdot\boldsymbol{\beta}\delta^y, \Pi\Psi)_{\mathcal{T}_h} - \langle \widehat{\boldsymbol{\delta}}_1, \Pi\Psi - P_M\Psi\rangle_{\partial\mathcal{T}_h},\\
&\quad = (\nabla\cdot\boldsymbol{q}, \delta^\Psi)_{\mathcal{T}_h} - (\nabla\cdot\boldsymbol{q}, \Psi)_{\mathcal{T}_h} + (\nabla\cdot\boldsymbol{\Pi q}, \Psi)_{\mathcal{T}_h} - \langle \boldsymbol{\Pi q}\cdot\boldsymbol{n}, \Psi\rangle_{\mathcal{T}_h}\\
&\qquad + (\boldsymbol{\beta}\delta^y, \nabla\Pi\Psi)_{\mathcal{T}_h} + (\nabla\cdot\boldsymbol{\beta}\delta^y, \Pi\Psi)_{\mathcal{T}_h} - \langle \widehat{\boldsymbol{\delta}}_1, \Pi\Psi - P_M\Psi\rangle_{\partial\mathcal{T}_h},\\
&\quad = (\nabla\cdot\boldsymbol{q}, \delta^\Psi)_{\mathcal{T}_h} + (\boldsymbol{q}, \nabla\Psi)_{\mathcal{T}_h} - (\boldsymbol{\Pi q}, \nabla\Psi)_{\mathcal{T}_h}\\
&\qquad + (\boldsymbol{\beta}\delta^y, \nabla\Pi\Psi)_{\mathcal{T}_h} + (\nabla\cdot\boldsymbol{\beta}\delta^y, \Pi\Psi)_{\mathcal{T}_h} - \langle \widehat{\boldsymbol{\delta}}_1, \Pi\Psi - P_M\Psi\rangle_{\partial\mathcal{T}_h},\\
&\quad = (\nabla\cdot\boldsymbol{q}, \delta^\Psi) + (\delta^{\boldsymbol{q}}, \nabla\delta^\Psi)_{\mathcal{T}_h} + (\boldsymbol{\beta}\delta^y, \nabla\Pi\Psi)_{\mathcal{T}_h}\\
&\qquad + (\nabla\cdot\boldsymbol{\beta}\delta^y, \Pi\Psi)_{\mathcal{T}_h} - \langle \widehat{\boldsymbol{\delta}}_1, \Pi\Psi - P_M\Psi\rangle_{\partial\mathcal{T}_h}.
\end{aligned}$$

The two equalities above give

$$
\begin{aligned}
\|\varepsilon_h^y\|_{\mathcal{T}_h}^2 &= -\langle \boldsymbol{\beta}\cdot\boldsymbol{n}(\varepsilon_h^y - \varepsilon_h^{\widehat{y}}), \delta^\Psi\rangle_{\partial\mathcal{T}_h} + (\nabla\varepsilon_h^y, \boldsymbol{\beta}\delta^\Psi)_{\mathcal{T}_h} + (\boldsymbol{\beta}\delta^y, \nabla\Pi\Psi)_{\mathcal{T}_h} \\
&\quad + (\nabla\cdot\boldsymbol{\beta}\delta^y, \Pi\Psi)_{\mathcal{T}_h} + \langle (h^{-1}+\tau_1)(\varepsilon_h^y - \varepsilon_h^{\widehat{y}}) + \widehat{\boldsymbol{\delta}}_1, \delta^\Psi - \delta^{\widehat{\Psi}}\rangle_{\partial\mathcal{T}_h} \\
&\quad - (\varepsilon_h^y - \mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}}), \nabla\cdot\delta^{\boldsymbol{\Phi}})_{\mathcal{T}_h} + (\nabla\mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}}), \delta^{\boldsymbol{\Phi}})_{\mathcal{T}_h} \\
&\quad + (\nabla\cdot\boldsymbol{q}, \delta^\Psi) + (\delta^{\boldsymbol{q}}, \nabla\delta^\Psi)_{\mathcal{T}_h} \\
&=: \sum_{i=1}^{9} R_i.
\end{aligned}
$$

Bounds for $R_1$ to $R_5$ have been obtained in Section two; we have

$$
\sum_{i=1}^{5} R_i \lesssim (h^{s_{\boldsymbol{q}}+1}\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y}\|y\|_{s^y,\Omega})\|\varepsilon_h^y\|_{\mathcal{T}_h}.
$$

For the terms $R_6$ and $R_7$, Lemma 35 and Lemma 37 give

$$
\begin{aligned}
R_6 &= -(\varepsilon_h^y - \mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}}), \nabla\cdot\boldsymbol{\Phi})_{\mathcal{T}_h} \\
&\le \|\varepsilon_h^y - \mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}})\|_{\mathcal{T}_h}\|\nabla\cdot\boldsymbol{\Phi}\|_{\mathcal{T}_h} \\
&\lesssim h(\left\|\varepsilon_h^{\boldsymbol{q}}\right\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h})\|\nabla\cdot\boldsymbol{\Phi}\|_{\mathcal{T}_h} \\
&\lesssim (h^{s_{\boldsymbol{q}}+1}\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y}\|y\|_{s^y,\Omega})\|\varepsilon_h^y\|_{\mathcal{T}_h}, \\
R_7 &= (\nabla\mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}}), \delta^{\boldsymbol{\Phi}})_{\mathcal{T}_h} \\
&\le \|\nabla\mathcal{I}_h(\varepsilon_h^y, \varepsilon_h^{\widehat{y}})\|_{\mathcal{T}_h}\|\delta^{\boldsymbol{\Phi}}\|_{\mathcal{T}_h} \\
&\lesssim h(\left\|\varepsilon_h^{\boldsymbol{q}}\right\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\varepsilon_h^y - \varepsilon_h^{\widehat{y}}\|_{\partial\mathcal{T}_h})\|\delta^{\boldsymbol{\Phi}}\|_{\mathcal{T}_h} \\
&\lesssim (h^{s_{\boldsymbol{q}}+1}\|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y}\|y\|_{s^y,\Omega})\|\varepsilon_h^y\|_{\mathcal{T}_h}.
\end{aligned}
$$

For $R_8$, we have

$$R_8 \le \|\nabla \cdot \boldsymbol{q}\|_{\mathcal{T}_h} \|\delta^{\Psi}\|_{\mathcal{T}_h} \lesssim h^2 \|\Psi\|_{2,\Omega} \\ \lesssim h^2 \|\varepsilon_h^y\|_{\mathcal{T}_h}.$$

Applying the triangle inequality for $R_9$ gives

$$R_9 \le \|\delta^{\boldsymbol{q}}\|_{\mathcal{T}_h} \|\nabla \delta^{\Psi}\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{q}}+1} \|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} \|\varepsilon_h^y\|_{\mathcal{T}_h}.$$

Using $s_{\boldsymbol{q}} \in [0,1]$ and summing the estimates for $R_1$ to $R_9$ completes the proof.

The triangle inequality gives optimal convergence rates for $\|\boldsymbol{q} - \boldsymbol{q}_h(u)\|_{\mathcal{T}_h}$ and $\|y - y_h(u)\|_{\mathcal{T}_h}$:

**Lemma 39**

$$\|\boldsymbol{q} - \boldsymbol{q}_h(u)\|_{\mathcal{T}_h} \le \|\delta^{\boldsymbol{q}}\|_{\mathcal{T}_h} + \|\varepsilon_h^{\boldsymbol{q}}\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{q}}} \|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y - 1} \|y\|_{s^y,\Omega}, \tag{4.27a}$$

$$\|y - y_h(u)\|_{\mathcal{T}_h} \le \|\delta^y\|_{\mathcal{T}_h} + \|\varepsilon_h^y\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{q}}+1} \|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y} \|y\|_{s^y,\Omega}. \tag{4.27b}$$

Step 4: The error equation for part 2 of the auxiliary problem (4.18b)

Next, we estimate the error between the exact state $z$ and flux $\boldsymbol{p}$ satisfying the mixed form (3.18a)-(3.18d) of the optimality system and the solutions $z_h(u)$ and $\boldsymbol{p}_h(u)$ of the auxiliary problem. Define

$$\begin{aligned} \delta^{\boldsymbol{p}} &= \boldsymbol{p} - \boldsymbol{\Pi} \boldsymbol{p}, & \varepsilon_h^{\boldsymbol{p}} &= \boldsymbol{\Pi} \boldsymbol{p} - \boldsymbol{p}_h(u), \\ \delta^z &= z - \Pi z, & \varepsilon_h^z &= \Pi z - z_h(u), \\ \delta^{\widehat{z}} &= z - P_M z, & \varepsilon_h^{\widehat{z}} &= P_M z - \widehat{z}_h(u), \\ \widehat{\boldsymbol{\delta}}_2 &= -\boldsymbol{\beta} \cdot \boldsymbol{n} \delta^{\widehat{z}} + (h^{-1} + \tau_2)(\delta^z - \delta^{\widehat{z}}), \end{aligned} \tag{4.28}$$

where $\widehat{z}_h(u) = \widehat{z}_h^o(u)$ on $\varepsilon_h^o$ and $\widehat{z}_h(u) = 0$ on $\varepsilon_h^\partial$. This gives $\varepsilon_h^{\widehat{z}} = 0$ on $\varepsilon_h^\partial$.

**Lemma 40** *We have*

$$
\begin{aligned}
&\mathscr{B}_2(\varepsilon_h^{\boldsymbol{p}}, \varepsilon_h^z, \varepsilon_h^{\widehat{z}}, \boldsymbol{r}_2, w_2, \mu_2) \\
&\qquad = -(\nabla \cdot \delta^{\boldsymbol{p}}, w_2)_{\mathcal{T}_h} - \langle \boldsymbol{\Pi} \boldsymbol{p} \cdot \boldsymbol{n}, \mu_2 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^{\partial}} - (\boldsymbol{\beta} \delta^z, \nabla w_2)_{\mathcal{T}_h} \\
&\qquad\quad + (y - y_h(u), w_2)_{\mathcal{T}_h} - \langle \widehat{\boldsymbol{\delta}}_2, w_2 \rangle_{\partial \mathcal{T}_h} + \langle \widehat{\boldsymbol{\delta}}_2, \mu_2 \rangle_{\partial \mathcal{T}_h \backslash \varepsilon_h^{\partial}}.
\end{aligned} \tag{4.29}
$$

The proof is similar to the proof of Lemma 40 and is omitted.

Step 5: Estimate for $\varepsilon_h^{\boldsymbol{p}}$ We use the following discrete Poincaré inequality to estimate $\varepsilon_h^{\boldsymbol{p}}$.

**Lemma 41** *We have*

$$
\|\varepsilon_h^z\|_{\mathcal{T}_h} \le C(\|\nabla \varepsilon_h^z\|_{\mathcal{T}_h} + h^{-\frac{1}{2}} \|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial \mathcal{T}_h}). \tag{4.30}
$$

**Lemma 42** *We have*

$$
\begin{aligned}
&\left\|\varepsilon_h^{\boldsymbol{p}}\right\|_{\mathcal{T}_h} + h^{-\frac{1}{2}} \|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial \mathcal{T}_h} \\
&\qquad \lesssim h^{s_{\boldsymbol{p}}} \|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega} + h^{s_z - 1} \|z\|_{s^z,\Omega} + h^{s_{\boldsymbol{q}}+1} \|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y} \|y\|_{s^y,\Omega},
\end{aligned} \tag{4.31a}
$$

$$
\left\|\varepsilon_h^z\right\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}} \|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega} + h^{s_z - 1} \|z\|_{s^z,\Omega} + h^{s_{\boldsymbol{q}}+1} \|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y} \|y\|_{s^y,\Omega}. \tag{4.31b}
$$

**Proof:** First, take $(\boldsymbol{v}_h, w_h, \mu_h) = (\varepsilon_h^{\boldsymbol{p}}, \varepsilon_h^z, \varepsilon_h^{\widehat{z}})$ in the key inequality in Lemma 33 to get

$$
\|\nabla \varepsilon_h^z\|_{\mathcal{T}_h} \lesssim \|\varepsilon_h^{\boldsymbol{p}}\|_{\mathcal{T}_h} + h^{-\frac{1}{2}} \|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial \mathcal{T}_h}. \tag{4.32}
$$

Next, since $\varepsilon_h^{\widehat{z}} = 0$ on $\varepsilon_h^{\partial}$, the energy identity for $\mathscr{B}_2$ in Lemma 30 gives

$$\begin{aligned}
&\mathscr{B}_2(\varepsilon_h^{\boldsymbol{p}}, \varepsilon_h^z, \varepsilon_h^{\widehat{z}}, \varepsilon_h^{\boldsymbol{p}}, \varepsilon_h^z, \varepsilon_h^{\widehat{z}}) \\
&\quad = (\varepsilon_h^{\boldsymbol{p}}, \varepsilon_h^{\boldsymbol{p}})_{\mathcal{T}_h} + \|(h^{-1} + \tau_2 + \frac{1}{2}\boldsymbol{\beta} \cdot \boldsymbol{n})^{\frac{1}{2}}(\varepsilon_h^z - \varepsilon_h^{\widehat{z}})\|_{\partial\mathcal{T}_h}^2 + \frac{1}{2}\|(-\nabla \cdot \boldsymbol{\beta})^{\frac{1}{2}}\varepsilon_h^z\|_{\mathcal{T}_h}^2.
\end{aligned}$$

Take $(\boldsymbol{r}_2, w_2, \mu_2) = (\boldsymbol{\varepsilon}_h^{\boldsymbol{p}}, \varepsilon_h^z, \varepsilon_h^{\widehat{z}})$ in the error equation (4.29) in Lemma 40 to obtain

$$\begin{aligned}
(\varepsilon_h^{\boldsymbol{p}}, \varepsilon_h^{\boldsymbol{p}})_{\mathcal{T}_h} &+ \|(h^{-1} + \tau_2 + \frac{1}{2}\boldsymbol{\beta} \cdot \boldsymbol{n})^{\frac{1}{2}}(\varepsilon_h^z - \varepsilon_h^{\widehat{z}})\|_{\partial\mathcal{T}_h}^2 + \frac{1}{2}\|(-\nabla \cdot \boldsymbol{\beta})^{\frac{1}{2}}\varepsilon_h^z\|_{\mathcal{T}_h}^2 \\
&= -(\nabla \cdot \delta^{\boldsymbol{p}}, \varepsilon_h^z)_{\mathcal{T}_h} - \langle \boldsymbol{\Pi}\boldsymbol{p} \cdot \boldsymbol{n}, \varepsilon_h^{\widehat{z}} \rangle_{\partial\mathcal{T}_h} \\
&\quad - (\boldsymbol{\beta}\delta^z, \nabla\varepsilon_h^z)_{\mathcal{T}_h} - \langle \widehat{\boldsymbol{\delta}}_2, \varepsilon_h^z - \varepsilon_h^{\widehat{z}} \rangle_{\partial\mathcal{T}_h} + (y - y_h(u), \varepsilon_h^z)_{\mathcal{T}_h} \\
&=: T_1 + T_2 + T_3 + T_4.
\end{aligned}$$

By the same argument as in the proof of Lemma 37, apply (4.32) and Young's inequality to obtain

$$\begin{aligned}
T_1 &= -(\nabla \cdot \delta^{\boldsymbol{p}}, \varepsilon_h^z)_{\mathcal{T}_h} - \langle \boldsymbol{\Pi}\boldsymbol{p} \cdot \boldsymbol{n}, \varepsilon_h^{\widehat{z}} \rangle_{\partial\mathcal{T}_h} \\
&= -(\nabla \cdot \boldsymbol{p}, \varepsilon_h^z - \mathcal{I}_h(\varepsilon_h^z, \varepsilon_h^{\widehat{z}}))_{\mathcal{T}_h} + (\delta^{\boldsymbol{p}}, \nabla\mathcal{I}_h(\varepsilon_h^z, \varepsilon_h^{\widehat{z}}))_{\mathcal{T}_h} \\
&= -(\nabla \cdot \delta^{\boldsymbol{p}}, \varepsilon_h^z - \mathcal{I}_h(\varepsilon_h^z, \varepsilon_h^{\widehat{z}}))_{\mathcal{T}_h} + (\delta^{\boldsymbol{p}}, \nabla\mathcal{I}_h(\varepsilon_h^z, \varepsilon_h^{\widehat{z}}))_{\mathcal{T}_h} \\
&\le h\|\nabla \cdot \delta^{\boldsymbol{p}}\|_{\mathcal{T}_h} h^{-1}\|\varepsilon_h^z - \mathcal{I}_h(\varepsilon_h^z, \varepsilon_h^{\widehat{z}})\|_{\mathcal{T}_h} + \|\delta^{\boldsymbol{p}}\|_{\mathcal{T}_h}\|\nabla\mathcal{I}_h(\varepsilon_h^z, \varepsilon_h^{\widehat{z}})\|_{\mathcal{T}_h} \\
&\le Ch^2\|\nabla \cdot \delta^{\boldsymbol{p}}\|_{\mathcal{T}_h}^2 + C\|\delta^{\boldsymbol{p}}\|_{\mathcal{T}_h}^2 + \frac{1}{8}\|\varepsilon_h^{\boldsymbol{p}}\|_{\mathcal{T}_h}^2 + \frac{1}{8h}\|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}^2, \\
T_2 &= -(\boldsymbol{\beta}\delta^z, \nabla\varepsilon_h^z)_{\mathcal{T}_h} \le C\|\delta^z\|_{\mathcal{T}_h}^2 + \frac{1}{8}\|\varepsilon_h^{\boldsymbol{p}}\|_{\mathcal{T}_h}^2 + \frac{1}{8h}\|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}^2, \\
T_3 &= -\langle \widehat{\boldsymbol{\delta}}_2, \varepsilon_h^z - \varepsilon_h^{\widehat{z}} \rangle_{\partial\mathcal{T}_h} \le 8h\|\widehat{\boldsymbol{\delta}}_2\|_{\partial\mathcal{T}_h}^2 + \frac{1}{8}\|\varepsilon_h^{\boldsymbol{p}}\|_{\mathcal{T}_h}^2 + \frac{1}{8h}\|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}^2.
\end{aligned}$$

For the term $T_4$, we have

$$
\begin{aligned}
T_4 = (y - y_h(u), \varepsilon_h^z)_{\mathcal{T}_h} &\le \|y - y_h(u)\|_{\mathcal{T}_h} \|\varepsilon_h^z\|_{\mathcal{T}_h} \\
&\le C\|y - y_h(u)\|_{\mathcal{T}_h} (\|\nabla \varepsilon_h^z\|_{\mathcal{T}_h} + h^{-\frac{1}{2}} \|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}) \\
&\le C\|y - y_h(u)\|_{\mathcal{T}_h} (\|\varepsilon_h^{\boldsymbol{p}}\|_{\mathcal{T}_h} + h^{-\frac{1}{2}} \|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}) \\
&\le C\|y - y_h(u)\|_{\mathcal{T}_h}^2 + \frac{1}{8}\|\varepsilon_h^{\boldsymbol{p}}\|_{\mathcal{T}_h}^2 + \frac{1}{8h}\|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h}^2.
\end{aligned}
$$

Summing $T_1$ to $T_4$ gives (4.31a); then (4.30), (4.31a), and (4.32) together imply (4.31b).

The triangle inequality gives optimal convergence rates for $\|\boldsymbol{p} - \boldsymbol{p}_h(u)\|_{\mathcal{T}_h}$ and $\|z - z_h(u)\|_{\mathcal{T}_h}$:

**Lemma 43**

$$
\|\boldsymbol{p} - \boldsymbol{p}_h(u)\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}} \|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega} + h^{s_z - 1} \|z\|_{s^z,\Omega} + h^{s_{\boldsymbol{q}}+1} \|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y} \|y\|_{s^y,\Omega}, \tag{4.33a}
$$

$$
\|z - z_h(u)\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}} \|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega} + h^{s_z - 1} \|z\|_{s^z,\Omega} + h^{s_{\boldsymbol{q}}+1} \|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y} \|y\|_{s^y,\Omega}. \tag{4.33b}
$$

Step 6: Estimates for $\|u - u_h\|_{\varepsilon_h^\partial}$ and $\|y - y_h\|_{\mathcal{T}_h}$

To obtain the main result, we estimate the error between the solution of the auxiliary problem and the HDG discretized optimality system (4.8). Define

$$
\begin{aligned}
&\zeta_{\boldsymbol{q}} = \boldsymbol{q}_h(u) - \boldsymbol{q}_h, \quad \zeta_y = y_h(u) - y_h, \quad \zeta_{\widehat{y}} = \widehat{y}_h(u) - \widehat{y}_h, \\
&\zeta_{\boldsymbol{p}} = \boldsymbol{p}_h(u) - \boldsymbol{p}_h, \quad \zeta_z = z_h(u) - z_h, \quad \zeta_{\widehat{z}} = \widehat{z}_h(u) - \widehat{z}_h,
\end{aligned}
$$

where $\widehat{y}_h = \widehat{y}_h^o$ on $\varepsilon_h^o$, $\widehat{y}_h = u_h$ on $\varepsilon_h^\partial$, $\widehat{z}_h = \widehat{z}_h^o$ on $\varepsilon_h^o$, and $\widehat{z}_h = 0$ on $\varepsilon_h^\partial$. This gives $\zeta_{\widehat{z}} = 0$ on $\varepsilon_h^\partial$.

Subtracting the two problems gives the error equations

$$\mathscr{B}_1(\zeta_{\boldsymbol{q}}, \zeta_y, \zeta_{\widehat{y}}; \boldsymbol{r}_1, w_1, \mu_1) = -\langle P_M u - u_h, \boldsymbol{r}_1 \cdot \boldsymbol{n} + (\boldsymbol{\beta} \cdot \boldsymbol{n} - h^{-1} - \tau_1) w_1 \rangle_{\varepsilon_h^\partial}, \tag{4.34a}$$

$$\mathscr{B}_2(\zeta_{\boldsymbol{p}}, \zeta_z, \zeta_{\widehat{z}}; \boldsymbol{r}_2, w_2, \mu_2) = (\zeta_y, w_2)_{\mathcal{T}_h}. \tag{4.34b}$$

**Lemma 44** *If **(A1)** and **(A2)** hold, then*

$$\gamma \|u - u_h\|^2_{\varepsilon_h^\partial} + \left\|\zeta_y\right\|^2_{\mathcal{T}_h} = \langle \gamma u + \boldsymbol{p}_h(u) \cdot \boldsymbol{n} + h^{-1} z_h(u) + \tau_2 z_h(u), u - u_h \rangle_{\varepsilon_h^\partial}$$
$$- \langle \gamma u_h + \boldsymbol{p}_h \cdot \boldsymbol{n} + h^{-1} z_h + \tau_2 z_h, u - u_h \rangle_{\varepsilon_h^\partial}.$$

**Proof:** We have

$$\langle \gamma u + \boldsymbol{p}_h(u) \cdot \boldsymbol{n} + h^{-1} z_h(u) + \tau_2 z_h(u), u - u_h \rangle_{\varepsilon_h^\partial} - \langle \gamma u_h + \boldsymbol{p}_h \cdot \boldsymbol{n} + h^{-1} z_h + \tau_2 z_h, u - u_h \rangle_{\varepsilon_h^\partial}$$
$$= \gamma \|u - u_h\|^2_{\varepsilon_h^\partial} + \langle \zeta_{\boldsymbol{p}} \cdot \boldsymbol{n} + h^{-1} \zeta_z + \tau_2 \zeta_z, u - u_h \rangle_{\varepsilon_h^\partial}.$$

Next, Lemma 31 gives

$$\mathscr{B}_1(\zeta_{\boldsymbol{q}}, \zeta_y, \zeta_{\widehat{y}}; \zeta_{\boldsymbol{p}}, -\zeta_z, -\zeta_{\widehat{z}}) + \mathscr{B}_2(\zeta_{\boldsymbol{p}}, \zeta_z, \zeta_{\widehat{z}}; -\zeta_{\boldsymbol{q}}, \zeta_y, \zeta_{\widehat{y}}) = 0.$$

Also, since $\tau_2$ is piecewise constant on $\partial\mathcal{T}_h$ we have

$$\begin{aligned}
&\mathscr{B}_1(\zeta_{\boldsymbol{q}}, \zeta_y, \zeta_{\widehat{y}}; \zeta_{\boldsymbol{p}}, -\zeta_z, -\zeta_{\widehat{z}}) + \mathscr{B}_2(\zeta_{\boldsymbol{p}}, \zeta_z, \zeta_{\widehat{z}}; -\zeta_{\boldsymbol{q}}, \zeta_y, \zeta_{\widehat{y}}) \\
&\qquad = (\zeta_y, \zeta_y)_{\mathcal{T}_h} - \langle P_M u - u_h, \zeta_{\boldsymbol{p}} \cdot \boldsymbol{n} + (h^{-1} + \tau_1 - \boldsymbol{\beta} \cdot \boldsymbol{n}) \zeta_z \rangle_{\varepsilon_h^\partial} \\
&\qquad = (\zeta_y, \zeta_y)_{\mathcal{T}_h} - \langle P_M u - u_h, \zeta_{\boldsymbol{p}} \cdot \boldsymbol{n} + h^{-1} \zeta_z + \tau_2 \zeta_z \rangle_{\varepsilon_h^\partial} \\
&\qquad = (\zeta_y, \zeta_y)_{\mathcal{T}_h} - \langle u - u_h, \zeta_{\boldsymbol{p}} \cdot \boldsymbol{n} + h^{-1} \zeta_z + \tau_2 \zeta_z \rangle_{\varepsilon_h^\partial}.
\end{aligned}$$

The above equalities yield

$$(\zeta_y, \zeta_y)_{\mathcal{T}_h} = \langle u - u_h, \zeta_{\boldsymbol{p}} \cdot \boldsymbol{n} + h^{-1}\zeta_z + \tau_2\zeta_z \rangle_{\varepsilon_h^\partial}.$$

**Theorem 10** *We have*

$$\|u - u_h\|_{\varepsilon_h^\partial} \lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega},$$

$$\|y - y_h\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}.$$

**Proof:** The optimality conditions yield $\gamma u + \boldsymbol{p} \cdot \boldsymbol{n} = 0$ and $\gamma u_h + \boldsymbol{p}_h \cdot \boldsymbol{n} + h^{-1} z_h + \tau_2 z_h = 0$ on $\varepsilon_h^\partial$. Therefore, the above lemma gives

$$\begin{aligned} \gamma \|u - u_h\|_{\varepsilon_h^\partial}^2 + \left\|\zeta_y\right\|_{\mathcal{T}_h}^2 &= \langle \gamma u + \boldsymbol{p}_h(u) \cdot \boldsymbol{n} + h^{-1} z_h(u) + \tau_2 z_h(u), u - u_h \rangle_{\varepsilon_h^\partial} \\ &= \langle (\boldsymbol{p}_h(u) - \boldsymbol{p}) \cdot \boldsymbol{n} + h^{-1} z_h(u) + \tau_2 z_h(u), u - u_h \rangle_{\varepsilon_h^\partial}. \end{aligned}$$

Since $\widehat{z}_h(u) = z = 0$ on $\varepsilon_h^\partial$, we have

$$\begin{aligned} \|\boldsymbol{p}_h(u) - \boldsymbol{p}\|_{\partial\mathcal{T}_h} &\le \|\boldsymbol{p}_h(u) - \boldsymbol{\Pi}\boldsymbol{p}\|_{\partial\mathcal{T}_h} + \|\boldsymbol{\Pi}\boldsymbol{p} - \boldsymbol{p}\|_{\partial\mathcal{T}_h} \\ &\lesssim h^{-\frac{1}{2}} \left\|\varepsilon_h^{\boldsymbol{p}}\right\|_{\mathcal{T}_h} + h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega}, \\ \|z_h(u)\|_{\varepsilon_h^\partial} &= \|z_h(u) - \Pi z + \Pi z - z + P_M z - \widehat{z}_h(u)\|_{\varepsilon_h^\partial} \\ &\le \|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h} + \|\Pi z - z\|_{\varepsilon_h^\partial}. \end{aligned}$$

This implies

$$\begin{aligned} \|u - u_h\|_{\varepsilon_h^\partial} + \|\zeta_y\|_{\mathcal{T}_h} &\lesssim h^{-\frac{1}{2}} \left\|\varepsilon_h^{\boldsymbol{p}}\right\|_{\mathcal{T}_h} + h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega} \\ &\quad + h^{-1}\|\varepsilon_h^z - \varepsilon_h^{\widehat{z}}\|_{\partial\mathcal{T}_h} + h^{-\frac{3}{2}}\|\delta^z\|_{\mathcal{T}_h}. \end{aligned}$$

Lemma 42 and approximation properties of the $L^2$ projection give

$$\begin{aligned} &\|u - u_h\|_{\varepsilon_h^\partial} + \|\zeta_y\|_{\mathcal{T}_h} \\ &\qquad \lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s^z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s^y,\Omega}. \end{aligned}$$

The triangle inequality and Lemma 39 yield

$$\|y - y_h\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s^{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s^z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s^{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s^y,\Omega}.$$

Step 7: Estimates for $\|\boldsymbol{p} - \boldsymbol{p}_h\|_{\mathcal{T}_h}$, $\|z - z_h\|_{\mathcal{T}_h}$, and $\|\boldsymbol{q} - \boldsymbol{q}_h\|_{\mathcal{T}_h}$

**Lemma 45** *We have*

$$\begin{aligned} \left\|\zeta_{\boldsymbol{p}}\right\|_{\mathcal{T}_h} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}, \\ \|\zeta_z\|_{\mathcal{T}_h} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}} \|y\|_{s_y,\Omega}. \end{aligned}$$

**Proof:** By the energy identity for $\mathscr{B}_2$ in Lemma 30, the second error equation (4.34b), and since $\zeta_{\widehat{z}} = 0$ on $\varepsilon_h^\partial$, we have

$$\begin{aligned} &\mathscr{B}_2(\zeta_{\boldsymbol{p}}, \zeta_z, \zeta_{\widehat{z}}; \zeta_{\boldsymbol{p}}, \zeta_z, \zeta_{\widehat{z}}) \\ &\qquad = (\zeta_{\boldsymbol{p}}, \zeta_{\boldsymbol{p}})_{\mathcal{T}_h} + \langle (h^{-1} + \tau_2 + \frac{1}{2}\boldsymbol{\beta}\cdot\boldsymbol{n})(\zeta_z - \zeta_{\widehat{z}}), \zeta_z - \zeta_{\widehat{z}}\rangle_{\partial\mathcal{T}_h} \\ &\qquad = (\zeta_y, \zeta_z)_{\mathcal{T}_h} \\ &\qquad \le \left\|\zeta_y\right\|_{\mathcal{T}_h} \|\zeta_z\|_{\mathcal{T}_h} \\ &\qquad \lesssim \left\|\zeta_y\right\|_{\mathcal{T}_h} (\|\nabla\zeta_z\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\zeta_z - \zeta_{\widehat{z}}\|_{\partial\mathcal{T}_h}) \\ &\qquad \lesssim \left\|\zeta_y\right\|_{\mathcal{T}_h} (\|\zeta_{\boldsymbol{p}}\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\zeta_z - \zeta_{\widehat{z}}\|_{\partial\mathcal{T}_h}). \end{aligned}$$

Here, for the last two inequalities we used the discrete Poincaré inequality in Lemma 41 and Lemma 33. This gives

$$\begin{aligned}&\left\|\zeta_{\boldsymbol{p}}\right\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\zeta_z - \zeta_{\widehat{z}}\|_{\partial\mathcal{T}_h}\\&\qquad\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}}\left\|\boldsymbol{p}\right\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}}\left\|z\right\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}}\left\|\boldsymbol{q}\right\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}}\left\|y\right\|_{s_y,\Omega}.\end{aligned}$$

Using the discrete Poincaré inequality and Lemma 33 again yields

$$\begin{aligned}\|\zeta_z\|_{\mathcal{T}_h} &\lesssim \|\nabla\zeta_z\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\|\zeta_z - \zeta_{\widehat{z}}\|_{\partial\mathcal{T}_h}\\&\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}}\left\|\boldsymbol{p}\right\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}}\left\|z\right\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}}\left\|\boldsymbol{q}\right\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}}\left\|y\right\|_{s_y,\Omega}.\end{aligned}$$

To obtain a positive convergence rate for $\boldsymbol{q}$, we need

$$r_y > 1, \quad r_z > 2, \quad r_{\boldsymbol{q}} > 0, \quad r_{\boldsymbol{p}} > 1. \tag{4.35}$$

**Lemma 46** *If (A1),* (4.35)*, and $k \geq 1$ hold, then*

$$\left\|\zeta_{\boldsymbol{q}}\right\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}-1}\left\|\boldsymbol{p}\right\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-2}\left\|z\right\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}}\left\|\boldsymbol{q}\right\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-1}\left\|y\right\|_{s_y,\Omega}.$$

**Proof:** By the energy identity in Lemma 30, the first error equation (4.34a), and since $\tau_2$ is piecewise constant on $\partial\mathcal{T}_h$, we have

$$
\begin{aligned}
&\mathscr{B}_1(\zeta_{\boldsymbol{q}}, \zeta_y, \zeta_{\widehat{y}}; \zeta_{\boldsymbol{q}}, \zeta_y, \zeta_{\widehat{y}}) \\
&\quad = (\zeta_{\boldsymbol{q}}, \zeta_{\boldsymbol{q}})_{\mathcal{T}_h} + \langle (h^{-1} + \tau_1 - \frac{1}{2}\boldsymbol{\beta}\cdot\boldsymbol{n})(\zeta_y - \zeta_{\widehat{y}}), \zeta_y - \zeta_{\widehat{y}}\rangle_{\partial\mathcal{T}_h\backslash\varepsilon_h^{\partial}} - (\nabla\cdot\boldsymbol{\beta}\zeta_y, \zeta_y)_{\mathcal{T}_h} \\
&\qquad + \langle (h^{-1} + \tau_1 - \frac{1}{2}\boldsymbol{\beta}\cdot\boldsymbol{n})\zeta_y, \zeta_y\rangle_{\varepsilon_h^{\partial}} \\
&\quad = -\langle P_M u - u_h, \zeta_{\boldsymbol{q}}\cdot\boldsymbol{n} + (\boldsymbol{\beta}\cdot\boldsymbol{n} - h^{-1} - \tau_1)\zeta_y\rangle_{\varepsilon_h^{\partial}} \\
&\quad = -\langle P_M u - u_h, \zeta_{\boldsymbol{q}}\cdot\boldsymbol{n} - (h^{-1} + \tau_2)\zeta_y\rangle_{\varepsilon_h^{\partial}} \\
&\quad = -\langle u - u_h, \zeta_{\boldsymbol{q}}\cdot\boldsymbol{n} - (h^{-1} + \tau_2)\zeta_y\rangle_{\varepsilon_h^{\partial}} \\
&\quad \lesssim \|u - u_h\|_{\varepsilon_h^{\partial}} (\left\|\zeta_{\boldsymbol{q}}\right\|_{\varepsilon_h^{\partial}} + h^{-1}\left\|\zeta_y\right\|_{\varepsilon_h^{\partial}}) \\
&\quad \lesssim h^{-\frac{1}{2}}\|u - u_h\|_{\varepsilon_h^{\partial}} (\left\|\zeta_{\boldsymbol{q}}\right\|_{\mathcal{T}_h} + h^{-\frac{1}{2}}\left\|\zeta_y\right\|_{\varepsilon_h^{\partial}}).
\end{aligned}
$$

This gives

$$
\begin{aligned}
\left\|\zeta_{\boldsymbol{q}}\right\|_{\mathcal{T}_h} &\lesssim h^{-\frac{1}{2}}\|u - u_h\|_{\varepsilon_h^{\partial}} \\
&\lesssim h^{s_{\boldsymbol{p}}-1}\|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-2}\|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}}\|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-1}\|y\|_{s_y,\Omega}.
\end{aligned}
$$

The above lemma, the triangle inequality, Lemma 39, and Lemma 43 complete the proof of the main result:

**Theorem 11** *We have*

$$
\begin{aligned}
\|\boldsymbol{p} - \boldsymbol{p}_h\|_{\mathcal{T}_h} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}}\|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}}\|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}}\|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}}\|y\|_{s_y,\Omega}, \\
\|z - z_h\|_{\mathcal{T}_h} &\lesssim h^{s_{\boldsymbol{p}}-\frac{1}{2}}\|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-\frac{3}{2}}\|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}+\frac{1}{2}}\|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-\frac{1}{2}}\|y\|_{s_y,\Omega}.
\end{aligned}
$$

*If in addition* (4.35) *is satisfied and* $k \geq 1$*, then*

$$\|\boldsymbol{q} - \boldsymbol{q}_h\|_{\mathcal{T}_h} \lesssim h^{s_{\boldsymbol{p}}-1} \|\boldsymbol{p}\|_{s_{\boldsymbol{p}},\Omega} + h^{s_z-2} \|z\|_{s_z,\Omega} + h^{s_{\boldsymbol{q}}} \|\boldsymbol{q}\|_{s_{\boldsymbol{q}},\Omega} + h^{s_y-1} \|y\|_{s_y,\Omega} .$$

## 4.3. NUMERICAL EXPERIMENTS

We present numerical results for a 2D example problem similar to examples from [11, 33] with $\boldsymbol{\beta} = \mathbf{0}$. We consider a square domain $\Omega = [0, 1/8] \times [0, 1/8] \subset \mathbb{R}^2$, and choose the problem data

$$f = 0, \quad y_d = (x^2 + y^2)^{-1/3}, \quad \boldsymbol{\beta} = [1, 1], \quad \text{and} \quad \gamma = 1.$$

The largest interior angle is $\omega = \pi/2$, and therefore $r_\Omega = 3/2$. Also, we have $y_d \in H^{1/3-\varepsilon}(\Omega)$ for any $\varepsilon > 0$, and therefore $r_d = 5/6 - \varepsilon$ for any $\varepsilon > 0$. For this example, the value of $r_d$ restricts the guaranteed regularity of the solution.

We do not have an exact solution for this problem; therefore, we generate numerical convergence rates by computing errors between approximate solutions computed on different meshes. Specifically, we compare approximate solutions computed on various meshes with the approximate solution on a fine mesh with 524288 elements, i.e., $h = 2^{-12}\sqrt{2}$. For all computations, we take $\tau_2 = 1$ and $\tau_1 = \tau_2 + \boldsymbol{\beta} \cdot \boldsymbol{n}$ so that **(A1)-(A3)** are satisfied.

When $k = 1$, the guaranteed theoretical convergence rates are given by Corollary 5 in Section 4.2:

$$\|y - y_h\|_{0,\Omega} = O(h^{5/6-\varepsilon}), \qquad \|z - z_h\|_{0,\Omega} = O(h^{5/6-\varepsilon}),$$

$$\|\boldsymbol{q} - \boldsymbol{q}_h\|_{0,\Omega} = O(h^{1/3-\varepsilon}), \qquad \|\boldsymbol{p} - \boldsymbol{p}_h\|_{0,\Omega} = O(h^{5/6-\varepsilon}),$$

Table 4.1. 2D Example with $k = 1$: Errors for the control $u$, state $y$, adjoint state $z$, and the fluxes $\boldsymbol{q}$ and $\boldsymbol{p}$

| $h/\sqrt{2}$ | $2^{-4}$ | $2^{-5}$ | $2^{-6}$ | $2^{-7}$ | $2^{-8}$ |
|---|---|---|---|---|---|
| $\|\boldsymbol{q} - \boldsymbol{q}_h\|_{0,\Omega}$ | 1.45e-1 | 1.00e-1 | 7.41e-2 | 5.63e-2 | 4.30e-2 |
| order | - | 0.53 | 0.44 | 0.40 | 0.39 |
| $\|\boldsymbol{p} - \boldsymbol{p}_h\|_{0,\Omega}$ | 2.67e-3 | 9.65e-4 | 3.55e-4 | 1.35e-4 | 5.20e-5 |
| order | - | 1.47 | 1.44 | 1.40 | 1.37 |
| $\|y - y_h\|_{0,\Omega}$ | 1.00e-3 | 3.32e-4 | 1.21e-4 | 4.60e-5 | 1.80e-5 |
| order | - | 1.60 | 1.46 | 1.39 | 1.35 |
| $\|z - z_h\|_{0,\Omega}$ | 5.91e-5 | 1.21e-5 | 2.43e-6 | 4.84e-7 | 9.63e-8 |
| order | - | 2.29 | 2.32 | 2.33 | 2.33 |
| $\|u - u_h\|_{0,\Gamma}$ | 1.31e-2 | 6.38e-3 | 3.32e-3 | 1.81e-3 | 1.00e-3 |
| order | - | 1.03 | 0.94 | 0.88 | 0.85 |

and

$$\|u - u_h\|_{0,\Gamma} = O(h^{5/6-\varepsilon}).$$

Table 4.1 shows numerical results for this case. As in Section 2, the numerically observed convergence rates in Table 4.2 match the theory for the control $u$ and the primary flux $\boldsymbol{q}$, but are higher than the theoretical rates for the other variables. As mentioned in Section 2, similar convergence behavior has been observed in other works [33, 43, 49, 55].

Next, for $k = 0$, Lemma 5 gives the suboptimal convergence rates

$$\|y - y_h\|_{0,\Omega} = O(h^{1/2-\varepsilon}), \quad \|z - z_h\|_{0,\Omega} = O(h^{1/2-\varepsilon}), \quad \|\boldsymbol{p} - \boldsymbol{p}_h\|_{0,\Omega} = O(h^{1/2-\varepsilon}),$$

and

$$\|u - u_h\|_{0,\Gamma} = O(h^{1/2-\varepsilon}).$$

Table 4.2. 2D Example with $k = 0$: Errors for the control $u$, state $y$, adjoint state $z$, and the fluxes $\boldsymbol{q}$ and $\boldsymbol{p}$

| $h/\sqrt{2}$ | $2^{-4}$ | $2^{-5}$ | $2^{-6}$ | $2^{-7}$ | $2^{-8}$ |
|---|---|---|---|---|---|
| $\|\boldsymbol{q} - \boldsymbol{q}_h\|_{0,\Omega}$ | 2.22e-1 | 1.69e-1 | 1.22e-1 | 8.92e-2 | 6.56e-2 |
| order | - | 0.39 | 0.47 | 0.46 | 0.44 |
| $\|\boldsymbol{p} - \boldsymbol{p}_h\|_{0,\Omega}$ | 8.60e-3 | 5.10e-3 | 2.75e-3 | 1.43e-3 | 7.31e-4 |
| order | - | 0.75 | 0.90 | 0.94 | 0.97 |
| $\|y - y_h\|_{0,\Omega}$ | 2.96e-3 | 1.33e-3 | 4.91e-4 | 1.82e-4 | 6.97e-5 |
| order | - | 1.15 | 1.44 | 1.43 | 1.39 |
| $\|z - z_h\|_{0,\Omega}$ | 3.82e-4 | 1.08e-4 | 2.89e-5 | 7.48e-6 | 1.90e-6 |
| order | - | 1.82 | 1.91 | 1.95 | 1.97 |
| $\|u - u_h\|_{0,\Gamma}$ | 2.83e-2 | 1.79e-2 | 1.07e-2 | 6.14e-3 | 3.47e-3 |
| order | - | 0.66 | 0.75 | 0.80 | 0.82 |

As in Section 2, we observe much larger numerical convergence rates for all variables. Improving the analysis for the $k = 0$ case is again an interesting topic we leave to be considered elsewhere.

# 5. CONCLUSIONS

We proposed HDG methods to approximate the solution of optimal Dirichlet boundary control problems for the Poisson and convection diffusion equations. For the Poisson equation, we used an existing HDG method and obtained a superlinear rate of convergence for the control in 2D under certain assumptions on the domain and the target state $y_d$. We also considered a Dirichlet boundary control problem for a convection diffusion equation and made two contributions. First, for a polygonal domain we considered a very weak mixed formulation of the PDE, and established well-posedness and regularity results for the PDE and the optimal control problem. Next, we proposed a new HDG method to approximate the solution of the optimality system and used established optimal superlinear convergence rates for the control under certain assumptions on the domain and the desired state. Finally, we removed the restrictions on the domain and the desired state and used very different analysis techniques to prove optimal convergence rates for the control. As far as we are aware, this is the first work to explore the analysis of this Dirichlet control problem and the numerical analysis of a computational method for this problem. We presented numerical results to demonstrate the performance of the method.

Our results indicate HDG methods have potential for solving more complex Dirichlet boundary control problems. We plan to investigate HDG methods for Dirichlet boundary control of other PDEs, including convection dominated diffusion problems and fluid flows. These problems may involve solutions with large gradients or shocks, and it is natural to consider HDG methods for such problems.

# REFERENCES

[1] Th. Apel, M. Mateos, J. Pfefferer, and A. Rösch. On the regularity of the solutions of Dirichlet optimal control problems in polygonal domains. *SIAM J. Control Optim.*, 53(6):3620–3641, 2015.

[2] Thomas Apel, Mariano Mateos, Johannes Pfefferer, and Arnd Rösch. Error estimates for dirichlet control problems in polygonal domains: Quasi-uniform meshes, 2018.

[3] N. Arada and J.-P. Raymond. Dirichlet boundary control of semilinear parabolic equations. I. Problems with no state constraints. *Appl. Math. Optim.*, 45(2):125–143, 2002.

[4] Roland Becker and Boris Vexler. Optimal control of the convection-diffusion equation using stabilized finite element methods. *Numer. Math.*, 106(3):349–367, 2007.

[5] Faker Ben Belgacem, Henda El Fekih, and Hejer Metoui. Singular perturbation for the Dirichlet boundary control of elliptic problems. *M2AN Math. Model. Numer. Anal.*, 37(5):883–850, 2003.

[6] M. Braack. Optimal control in fluid mechanics by finite elements with symmetric stabilization. *SIAM J. Control Optim.*, 48(2):672–687, 2009.

[7] Erik Burman and Peter Hansbo. Edge stabilization for Galerkin approximations of convection-diffusion-reaction problems. *Comput. Methods Appl. Mech. Engrg.*, 193(15-16):1437–1453, 2004.

[8] E. Casas, A. Günther, and M. Mateos. A paradox in the approximation of Dirichlet control problems in curved domains. *SIAM J. Control Optim.*, 49(5):1998–2007, 2011.

[9] E. Casas, M. Mateos, and J.-P. Raymond. Penalization of Dirichlet optimal control problems. *ESAIM Control Optim. Calc. Var.*, 15(4):782–809, 2009.

[10] Eduardo Casas, Mariano Mateos, and Jean-Pierre Raymond. Penalization of Dirichlet optimal control problems. *ESAIM Control Optim. Calc. Var.*, 15(4):782–809, 2009.

[11] Eduardo Casas and Jean-Pierre Raymond. Error estimates for the numerical approximation of Dirichlet boundary control for semilinear elliptic equations. *SIAM J. Control Optim.*, 45(5):1586–1611, 2006.

[12] Aycil Cesmelioglu, Bernardo Cockburn, and Weifeng Qiu. Analysis of a hybridizable discontinuous Galerkin method for the steady-state incompressible Navier-Stokes equations. *Math. Comp.*, 86(306):1643–1670, 2017.

[13] Nagaiah Chamakuri, Christian Engwer, and Karl Kunisch. Boundary control of bidomain equations with state-dependent switching source functions in the ionic model. *J. Comput. Phys.*, 273:227–242, 2014.

[14] Lili Chang, Wei Gong, and Ningning Yan. Weak boundary penalization for Dirichlet boundary control problems governed by elliptic equations. *J. Math. Anal. Appl.*, 453(1):529–557, 2017.

[15] Yanlai Chen and Bernardo Cockburn. Analysis of variable-degree HDG methods for convection-diffusion equations. Part I: general nonconforming meshes. *IMA J. Numer. Anal.*, 32(4):1267–1293, 2012.

[16] Yanlai Chen and Bernardo Cockburn. Analysis of variable-degree HDG methods for convection-diffusion equations. Part II: Semimatching nonconforming meshes. *Math. Comp.*, 83(285):87–111, 2014.

[17] Yanlai Chen, Bernardo Cockburn, and Bo Dong. Superconvergent HDG methods for linear, stationary, third-order equations in one-space dimension. *Math. Comp.*, 85(302):2715–2742, 2016.

[18] Yanping Chen. Superconvergence of mixed finite element methods for optimal control problems. *Math. Comp.*, 77(263):1269–1291, 2008.

[19] Sudipto Chowdhury, Thirupathi Gudi, and A. K. Nandakumaran. Error bounds for a Dirichlet boundary control problem based on energy spaces. *Math. Comp.*, 86(305):1103–1126, 2017.

[20] Bernardo Cockburn, Jayadeep Gopalakrishnan, and Raytcho Lazarov. Unified hybridization of discontinuous Galerkin, mixed, and continuous Galerkin methods for second order elliptic problems. *SIAM J. Numer. Anal.*, 47(2):1319–1365, 2009.

[21] Bernardo Cockburn, Jayadeep Gopalakrishnan, Ngoc Cuong Nguyen, Jaume Peraire, and Francisco-Javier Sayas. Analysis of HDG methods for Stokes flow. *Math. Comp.*, 80(274):723–760, 2011.

[22] Bernardo Cockburn and Kassem Mustapha. A hybridizable discontinuous Galerkin method for fractional diffusion problems. *Numer. Math.*, 130(2):293–314, 2015.

[23] Bernardo Cockburn and Jiguang Shen. A hybridizable discontinuous Galerkin method for the $p$-Laplacian. *SIAM J. Sci. Comput.*, 38(1):A545–A566, 2016.

[24] A. R. Danilin. Asymptotics of the solution of a problem of optimal boundary control of a flow through a part of the boundary. *Tr. Inst. Mat. Mekh.*, 20(4):116–127, 2014.

[25] J. C. de los Reyes and K. Kunisch. A semi-smooth Newton method for control constrained boundary optimal control of the Navier-Stokes equations. *Nonlinear Anal.*, 62(7):1289–1316, 2005.

[26] Klaus Deckelnick, Andreas Günther, and Michael Hinze. Finite element approximation of Dirichlet boundary control for elliptic PDEs on two- and three-dimensional curved domains. *SIAM J. Control Optim.*, 48(4):2798–2819, 2009.

[27] Luca Dede' and Alfio Quarteroni. Optimal control and numerical adaptivity for advection-diffusion equations. *M2AN Math. Model. Numer. Anal.*, 39(5):1019–1040, 2005.

[28] Guosheng Fu, Weifeng Qiu, and Wujun Zhang. An analysis of HDG methods for convection-dominated diffusion problems. *ESAIM Math. Model. Numer. Anal.*, 49(1):225–256, 2015.

[29] A. V. Fursikov, M. D. Gunzburger, and L. S. Hou. Boundary value problems and optimal boundary control for the Navier-Stokes system: the two-dimensional case. *SIAM J. Control Optim.*, 36(3):852–894, 1998.

[30] A. V. Fursikov, M. D. Gunzburger, and L. S. Hou. Optimal Dirichlet control and inhomogeneous boundary value problems for the unsteady Navier-Stokes equations. In *Control and partial differential equations (Marseille-Luminy, 1997)*, volume 4 of *ESAIM Proc.*, pages 97–116. Soc. Math. Appl. Indust., Paris, 1998.

[31] A. V. Fursikov, M. D. Gunzburger, and L. S. Hou. Optimal boundary control for the evolutionary Navier-Stokes system: the three-dimensional case. *SIAM J. Control Optim.*, 43(6):2191–2232, 2005.

[32] W. Gong, W. Hu, M. Mateos, J. R. Singler, and Y. Zhang. A New HDG Method for Dirichlet Boundary Control of Convection Diffusion PDEs II: Low Regularity. Submitted.

[33] Wei Gong and Ningning Yan. Mixed finite element method for Dirichlet boundary control problem governed by elliptic PDEs. *SIAM J. Control Optim.*, 49(3):984–1014, 2011.

[34] P. Grisvard. *Elliptic problems in nonsmooth domains*, volume 24 of *Monographs and Studies in Mathematics*. Pitman (Advanced Publishing Program), Boston, MA, 1985.

[35] M. D. Gunzburger, L. S. Hou, and Th. P. Svobodny. Analysis and finite element approximation of optimal control problems for the stationary Navier-Stokes equations with Dirichlet controls. *RAIRO Modél. Math. Anal. Numér.*, 25(6):711–748, 1991.

[36] M. D. Gunzburger and S. Manservisi. The velocity tracking problem for Navier-Stokes flows with boundary control. *SIAM J. Control Optim.*, 39(2):594–634, 2000.

[37] Max D. Gunzburger, LiSheng Hou, and Thomas P. Svobodny. Boundary velocity control of incompressible flow with an application to viscous drag reduction. *SIAM J. Control Optim.*, 30(1):167–181, 1992.

[38] Max D. Gunzburger and R. A. Nicolaides. An algorithm for the boundary control of the wave equation. *Appl. Math. Lett.*, 2(3):225–228, 1989.

[39] Matthias Heinkenschloss and Dmitriy Leykekhman. Local error estimates for SUPG solutions of advection-dominated elliptic linear-quadratic optimal control problems. *SIAM J. Numer. Anal.*, 47(6):4607–4638, 2010.

[40] Michael Hinze and Karl Kunisch. Second order methods for boundary control of the instationary Navier-Stokes system. *ZAMM Z. Angew. Math. Mech.*, 84(3):171–187, 2004.

[41] L. S. Hou and S. S. Ravindran. A penalized Neumann control approach for solving an optimal Dirichlet control problem for the Navier-Stokes equations. *SIAM J. Control Optim.*, 36(5):1795–1814, 1998.

[42] W. Hu, M. Mateos, J. R. Singler, and Y. Zhang. A New HDG Method for Dirichlet Boundary Control of Convection Diffusion PDEs I: High Regularity. Submitted.

[43] Weiwei Hu, Jiguang Shen, John R. Singler, Yangwen Zhang, and Xiabo Zheng. A Superconvergent HDG Method for Dirichlet Boundary Control of Elliptic PDEs. Submitted.

[44] Binjie Li and Xiaoping Xie. Analysis of a family of HDG methods for second order elliptic problems. *J. Comput. Appl. Math.*, 307:37–51, 2016.

[45] Wenbin Liu, Danping Yang, Lei Yuan, and Chaoqun Ma. Finite element approximations of an optimal control problem with integral state constraint. *SIAM J. Numer. Anal.*, 48(3):1163–1185, 2010.

[46] A. Martínez, C. Rodríguez, and M. E. Vázquez-Méndez. Theoretical and numerical analysis of an optimal control problem related to wastewater treatment. *SIAM J. Control Optim.*, 38(5):1534–1553, 2000.

[47] M. Mateos and I. Neitzel. Dirichlet control of elliptic state constrained problems. *Comput. Optim. Appl.*, 63(3):825–853, 2016.

[48] Mariano Mateos and Ira Neitzel. Dirichlet control of elliptic state constrained problems. *Comput. Optim. Appl.*, 63(3):825–853, 2016.

[49] S. May, R. Rannacher, and B. Vexler. Error analysis for a finite element approximation of elliptic Dirichlet boundary control problems. *SIAM J. Control Optim.*, 51(3):2585–2611, 2013.

[50] C. Meyer and A. Rösch. Superconvergence properties of optimal control problems. *SIAM J. Control Optim.*, 43(3):970–985, 2004.

[51] Kassem Mustapha, Maher Nour, and Bernardo Cockburn. Convergence and superconvergence analyses of HDG methods for time fractional diffusion problems. *Adv. Comput. Math.*, 42(2):377–393, 2016.

[52] N. C. Nguyen, J. Peraire, and B. Cockburn. An implicit high-order hybridizable discontinuous Galerkin method for linear convection-diffusion equations. *J. Comput. Phys.*, 228(9):3232–3254, 2009.

[53] N. C. Nguyen, J. Peraire, and B. Cockburn. An implicit high-order hybridizable discontinuous Galerkin method for nonlinear convection-diffusion equations. *J. Comput. Phys.*, 228(23):8841–8855, 2009.

[54] N. C. Nguyen, J. Peraire, and B. Cockburn. A hybridizable discontinuous Galerkin method for Stokes flow. *Comput. Methods Appl. Mech. Engrg.*, 199(9-12):582–597, 2010.

[55] G. Of, T. X. Phan, and O. Steinbach. An energy space finite element approach for elliptic Dirichlet boundary control problems. *Numer. Math.*, 129(4):723–748, 2015.

[56] Weifeng Qiu and Ke Shi. An HDG method for convection diffusion equation. *J. Sci. Comput.*, 66(1):346–357, 2016.

[57] Sivaguru S Ravindran. Finite element approximation of Dirichlet control using boundary penalty method for unsteady Navier–Stokes equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 51(3):825–849, 2017.

[58] Sivaguru S. Ravindran. Finite element approximation of Dirichlet control using boundary penalty method for unsteady Navier-Stokes equations. *ESAIM Math. Model. Numer. Anal.*, 51(3):825–849, 2017.

[59] Yulian Spasov and Karl Kunisch. Dynamical system based optimal control of incompressible fluids. Boundary control. *Eur. J. Mech. B Fluids*, 25(2):153–163, 2006.

[60] M. Stanglmeier, N. C. Nguyen, J. Peraire, and B. Cockburn. An explicit hybridizable discontinuous Galerkin method for the acoustic wave equation. *Comput. Methods Appl. Mech. Engrg.*, 300:748–769, 2016.

[61] Ruxandra Stavre. A boundary control problem for the blood flow in venous insufficiency. The general case. *Nonlinear Anal. Real World Appl.*, 29:98–116, 2016.

[62] B. Vexler. Finite element approximation of elliptic Dirichlet optimal control problems. *Numer. Funct. Anal. Optim.*, 28(7-8):957–973, 2007.

[63] Ningning Yan and Zhaojie Zhou. A RT mixed FEM/DG scheme for optimal control governed by convection diffusion equations. *J. Sci. Comput.*, 41(2):273–299, 2009.

# VITA

Yangwen Zhang was born in Songsu, Anhui, eastsouth of China. He lived there for 18 years and finished elementary school, middle school and high school there. After that, he moved to Yunnan province, and received his bachelor degree in mathematics at Honghe College at June 2009. Next, he received a master's degree in June, 2012 from Sichuan University. In September 2013, he moved to the US to study at Missouri University of Science and Technology. In May 2018, he received a PhD in mathematics from Missouri S&T.