*IMA Journal of Numerical Analysis* (2013) Page 1 of 17 doi:10.1093/imanum/drn000

# An Analysis of HDG Methods for the Helmholtz Equation

# JINTAO CUI†

Department of Mathematics and Statistics, University of Arkansas at Little Rock, 620 Dickinson Hall, 2801 University Avenue, Little Rock, AR 72204, U.S.A.

AND

### WUJUN ZHANG‡

Department of Mathematics, University of Maryland, Mathematics Building, University of Maryland, College Park, MD 20742, U.S.A.

# [Received on 18 January 2013]

The finite element method has been widely used to discretize the Helmholtz equation with various types of boundary conditions. The strong indefiniteness of the Helmholtz equation makes it difficult to establish stability estimates for the numerical solution. In particular, discontinuous Galerkin methods for Helmholtz equation with a high wave number result in very large matrices since they typically have more degrees of freedom than conforming methods. However, hybridizable discontinuous Galerkin (HDG) methods offer an attractive alternative because they have build-in stabilization mechanisms and a reduced global linear system. In this paper, we study the HDG methods for the Helmholtz equation with first order absorbing boundary condition in two and three dimensions. We prove that the proposed HDG methods are stable (hence well-posed) without any mesh constraint. The stability constant is independent of the polynomial degree. By using a projection-based error analysis, we also derive the error estimates in  $L_2$ norm for piecewise polynomial spaces with arbitrary degree.

Keywords: Hybridizable discontinuous Galerkin methods; Helmholtz equation.

# 1. Introduction

In this paper, we provide a new *a priori* error analysis of the hybridizable discontinuous Galerkin (HDG) method for solving the Helmholtz equation with first order absorbing boundary condition, namely,

$$-\Delta u - \kappa^2 u = f \quad \text{in } \Omega, \tag{1.1a}$$

$$\frac{\partial u}{\partial n} + i\kappa u = g \quad \text{on } \partial\Omega, \tag{1.1b}$$

where  $\Omega \subset \mathbb{R}^d$ , d = 2, 3, is a bounded and strictly star-shaped polygonal/polyhedral domain with respect to a point  $\mathbf{x}_{\Omega} \in \Omega$ . More precisely, there exists a positive constant  $C_{\Omega}$  such that

$$(\boldsymbol{x} - \boldsymbol{x}_{\Omega}) \cdot \boldsymbol{n}_{\Omega} \geqslant C_{\Omega} \qquad \forall \boldsymbol{x} \in \partial \Omega.$$
(1.2)

We introduce  $q := -\nabla u$ . Then the Helmholtz equation can be rewritten in mixed form as finding

<sup>†</sup>Corresponding author. Email: jxcui1@ualr.edu <sup>‡</sup>Email: wujun@umd.edu

© The author 2013. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.

(u, q) such that

$$\boldsymbol{q} = -\nabla u \quad \text{in } \boldsymbol{\Omega}, \tag{1.3a}$$

$$\nabla \cdot \boldsymbol{q} - \kappa^2 \boldsymbol{u} = f \qquad \text{in } \Omega, \tag{1.3b}$$

$$-\boldsymbol{q}\cdot\boldsymbol{n}+i\boldsymbol{\kappa}\boldsymbol{u}=\boldsymbol{g}\qquad\text{on }\partial\Omega.\tag{1.3c}$$

It is well-known that the quality of numerical approximation of Helmholtz's equation significantly depends on the wave number. In Ihlenburg & Babuška (1995), it was first rigorously proved for the one dimensional case that under the "rule of thumb" constraint,  $\kappa h \leq 1$ , the relative  $H^1$  error between the approximation and true solution satisfies

$$\frac{|u-u_h|_1}{|u|_1} \leqslant C(\kappa h + \kappa^3 h^2)$$

where  $u_h$  is the approximate solution obtained from a standard finite element method and  $|\cdot|_1$  denotes the  $H^1$  semi-norm on domain  $\Omega$ . The second term on the right hand side, which is known as the pollution error, implies that the error grows significantly with increasing wave number, even if the 'rule of thumb' is satisfied. In the past decades, a considerable deal of effort has been devoted to study the explicit influence of wave number on the numerical approximation with the goal of minimizing or eliminating the pollution effect. We refer the reader to Harari (1997); Babuška & Sauter (2000); Deraemaeker et al. (1999); Gerdes & Ihlenburg (1999); Oberai & Pinsky (2000); Ainsworth (2004); Ainsworth et al. (2006); Ainsworth & Wajid (2009) and the references therein. Various methods have been proposed to obtain more stable numerical approximations, which include stabilizing the formulation by Galerkin least square methods (see, e.g., Harari & Hughes, 1992; Babuška et al., 1995); by discontinuous Galerkin methods (see, e.g., Ainsworth et al., 2006; Chung & Engquist, 2006; Feng & Wu, 2009; Feng & Xing, 2010); by using non-polynomial trial and test functions or generalized finite element methods (see, e.g., Babuška et al., 1995; Babuška & Melenk, 1997; Suleau et al., 2000; Farhat et al., 2001; Laghrouche et al., 2002; Bao et al., 2004); and by using DG couplings of plane wave-based methods (see, e.g., Cessenat & Després, 1998, 2003; Huttunen & Monk, 2007; Buffa & Monk, 2008; Gittelson et al., 2009; Luostari et al., 2009; Hiptmair et al., 2011).

The cause of the pollution effect is related to the loss of stability of time harmonic wave equations. On the other hand, one advantage of hybridizable discontinuous Galerkin (HDG) methods is their builtin stabilization mechanisms. Moreover, the HDG methods are attractive because the degrees of freedom in each element can be removed from the global computation and the method reduces to solving only for degrees of freedom on the skeleton of the mesh. In Griesmaier & Monk (2011), the HDG method was introduced for solving the interior Dirichlet problem for the Helmholtz equation. Optimal convergence rates (with respect to *h*) for both *u* and *q* are achieved under the constraint that  $h\kappa^2$  is sufficiently small. The goal of this paper is to investigate the stability and convergence result of the HDG methods for (1.3) without any mesh constraint. We prove that the proposed HDG methods are stable for any wave number  $\kappa$  and mesh size *h*. The stability constant is shown to be independent of the polynomial degree. Moreover, by using a projection-based error analysis, we derive the error estimates in  $L_2$  norm for polynomial spaces with arbitrary degree.

The rest of the paper is organized as follows. In Section 2, we introduce HDG method for the Helmholtz equation (1.3). In Section 3, we state and discuss the stability and error estimates for the proposed HDG method, and in Section 4 we present their proofs.

### 2. The Hybridizable Discontinuous Galerkin Method

### 2.1 Meshes and Notations

Let  $\mathscr{T}_h$  be a shape-regular triangulation of  $\Omega$  which consists of simplex T with faces F in  $\mathbb{R}^3$  (or triangles *T* with edges *F* in  $\mathbb{R}^2$ ). We denote by  $\mathscr{E}_h$  the set of all faces/edges *F* of all tetrahedra/triangle *T* of the triangulation  $\mathscr{T}_h$ , by  $\mathscr{E}_h^I$  the set of all interior faces/edges, and by  $\partial \mathscr{T}_h$  the set of boundaries  $\partial T$  of the elements *T* of  $\mathscr{T}_h$ . Let  $L_2(\mathscr{T}_h) := \prod_{T \in \mathscr{T}_h} L_2(T)$ ,  $L_2(\mathscr{T}_h) := [L_2(\mathscr{T}_h)]^d$ ,  $L_2(\mathscr{E}_h) := \prod_{F \in \mathscr{E}_h} L_2(F)$  and we denote

$$\begin{split} \|v\|_{\mathscr{T}_h}^2 &= \sum_{T \in \mathscr{T}_h} \|v\|_T^2 \qquad \forall v \in L_2(\mathscr{T}_h), \\ \|\mu\|_{\partial \mathscr{T}_h}^2 &= \sum_{T \in \mathscr{T}_h} \|\mu\|_{\partial T}^2 \qquad \forall \mu \in L_2(\mathscr{E}_h). \end{split}$$

We set the bilinear forms

$$(\mathbf{v},\mathbf{w})_{\mathscr{T}_h} := \sum_{T \in \mathscr{T}_h} \int_T \mathbf{v} \cdot \overline{\mathbf{w}} dx, \qquad (v,w)_{\mathscr{T}_h} := \sum_{T \in \mathscr{T}_h} \int_T v \overline{w} dx, \qquad \langle v,w \rangle_{\partial \mathscr{T}_h} := \sum_{T \in \mathscr{T}_h} \int_{\partial T} v \overline{w} ds.$$

We define on an interior edge/face  $F = \partial T^+ \cap \partial T^-$  ( $T^+$  has bigger global labelling) that

$$[w]|_{F} := w|_{T^{+}} - w|_{T^{-}}, \qquad [v]|_{F} := v|_{T^{+}} - v|_{T^{-}}.$$

On an boundary edge/face F, we set  $[w]|_F = w|_F$  and  $[v]|_F = v|_F$ . We define the average  $\{\!\{w\}\!\}$  and **{***v***}** by

$$\{\!\{w\}\!\}|_{F} := \frac{1}{2} (w|_{T^{+}} + w|_{T^{-}}), \quad \{\!\{v\}\!\}|_{F} := \frac{1}{2} (v|_{T^{+}} + v|_{T^{-}}) \quad \text{for any } F = \partial T^{+} \cap \partial T^{-},$$

and  $\{\!\{w\}\!\}|_F := w|_F, \{\!\{v\}\!\}|_F := v|_F$  for any  $F \in \partial \Omega$ .

# 2.2 HDG Method

Define the following discrete spaces:

$$\boldsymbol{V}_h := \{ \boldsymbol{\nu} \in \boldsymbol{L}_2(\mathscr{T}_h) : \boldsymbol{\nu}|_T \in \boldsymbol{P}_p(T) \ \forall \ T \in \mathscr{T}_h \},$$
(2.1a)

$$W_h := \{ w \in L_2(\mathscr{T}_h) : w |_T \in P_p(T) \ \forall \ T \in \mathscr{T}_h \},$$

$$(2.1b)$$

$$M_h := \{ \mu \in L_2(\mathscr{E}_h) : \mu |_T \in P_p(F) \ \forall F \in \mathscr{E}_h \},$$
(2.1c)

 $P_p(T)$  is the space of polynomials of total degree at most p defined on T, and  $P_p(T) = [P_p(T)]^d$ . The HDG method seeks approximations  $u_h \in W_h$  of  $u, q_h \in V_h$  of q, and a numerical trace  $\hat{u}_h \in M_h$ approximating u on  $\mathcal{E}_h$ , which satisfy

$$(\boldsymbol{q}_h,\boldsymbol{\tau}_h)_{\mathscr{T}_h} - (\boldsymbol{u}_h, \nabla \cdot \boldsymbol{\tau}_h)_{\mathscr{T}_h} + \langle \widehat{\boldsymbol{u}}_h, \boldsymbol{\tau}_h \cdot \boldsymbol{n} \rangle_{\partial \mathscr{T}_h} = 0, \qquad (2.2a)$$

$$-(\boldsymbol{q}_{h},\nabla v_{h})_{\mathscr{T}_{h}}-\boldsymbol{\kappa}^{2}(\boldsymbol{u}_{h},\boldsymbol{v}_{h})_{\mathscr{T}_{h}}+\langle \widehat{\boldsymbol{q}}_{h}\cdot\boldsymbol{n},\boldsymbol{v}_{h}\rangle_{\mathscr{T}_{h}}=(f,v_{h})_{\mathscr{T}_{h}},$$
(2.2b)

$$\langle -\widehat{\boldsymbol{q}}_{h} \cdot \boldsymbol{n} + i\kappa\widehat{\boldsymbol{\mu}}_{h}, \boldsymbol{\mu}_{h} \rangle_{\partial\Omega} = \langle g, \boldsymbol{\mu}_{h} \rangle_{\partial\Omega}, \qquad (2.2c)$$

$$\langle \widehat{\boldsymbol{q}}_h \cdot \boldsymbol{n}, \mu_h \rangle_{\partial \mathscr{T}_h \setminus \partial \Omega} = 0,$$
 (2.2d)

for all  $(\boldsymbol{\tau}_h, v_h, \boldsymbol{\mu}_h) \in \boldsymbol{V}_h \times W_h \times M_h$ , where

$$\widehat{\boldsymbol{q}}_{h} = \boldsymbol{q}_{h} + i\tau(\boldsymbol{u}_{h} - \widehat{\boldsymbol{u}}_{h})\boldsymbol{n} \qquad \text{on } \partial \mathscr{T}_{h}, \tag{2.3}$$

for some positive stabilization function  $\tau_e$ , and  $\boldsymbol{n}$  is the outward normal to  $T \in \mathcal{T}_h$ . Throughout the paper, we take  $\tau$  to be constant on each faces and we denote  $\tau_{max} := \max\{\tau, e \in \mathcal{T}_h\}$  and  $\tau_{min} := \min\{\tau, e \in \mathcal{T}_h\}$ . We also denote by  $P_W$ ,  $P_M$  the  $L_2$ -orthogonal projection onto the space  $W_h$  and  $M_h$ , respectively.

REMARK 2.1 The HDG method considered in this work belongs to the class of hybridizable local discontinuous Galerkin (LDG-H) methods (Cockburn *et al.*, 2009). It is similar with the HDG method studied in Griesmaier & Monk (2011) for the interior Dirichlet problem for the Helmholtz equation. Note that in their paper, the flux  $\mathbf{q} = -ik\nabla u$  while in our work  $\mathbf{q} = -\nabla u$ . By a simple comparison, we observe that we essentially recover the method proposed in Griesmaier & Monk (2011) if  $\tau = O(\kappa)$ .

### 3. The Main Results

### 3.1 Stability Estimates

We consider the following problem:

Find  $(\boldsymbol{q}_h, u_h, \widehat{u}_h) \in (\boldsymbol{V}_h, V_h, M_h)$  such that

$$(\boldsymbol{q}_{h},\boldsymbol{\tau}_{h})_{\mathscr{T}_{h}}-(\boldsymbol{u}_{h},\nabla\cdot\boldsymbol{\tau}_{h})_{\mathscr{T}_{h}}+\langle\widehat{\boldsymbol{u}}_{h},\boldsymbol{\tau}_{h}\cdot\boldsymbol{n}\rangle_{\partial\mathscr{T}_{h}}=(\boldsymbol{Q},\boldsymbol{\tau}_{h})_{\mathscr{T}_{h}},$$
(3.1a)

$$-(\boldsymbol{q}_{h},\nabla v_{h})_{\mathscr{T}_{h}}-\boldsymbol{\kappa}^{2}(u_{h},v_{h})_{\mathscr{T}_{h}}+\langle \widehat{\boldsymbol{q}}_{h}\cdot\boldsymbol{n},v_{h}\rangle_{\partial\mathscr{T}_{h}}=(f,v_{h})_{\mathscr{T}_{h}},$$
(3.1b)

$$\langle -\widehat{\boldsymbol{q}}_{h} \cdot \boldsymbol{n} + i\kappa\widehat{\boldsymbol{\mu}}_{h}, \boldsymbol{\mu}_{h} \rangle_{\partial\Omega} = \langle g, \boldsymbol{\mu}_{h} \rangle_{\partial\Omega}, \qquad (3.1c)$$

$$\langle \widehat{\boldsymbol{q}}_h \cdot \boldsymbol{n}, \mu_h \rangle_{\partial \mathscr{T}_h \setminus \partial \Omega} = 0,$$
 (3.1d)

for all  $(\boldsymbol{\tau}_h, v_h, \mu_h) \in \boldsymbol{V}_h \times W_h \times M_h$ . Here  $\widehat{\boldsymbol{q}}_h$  is defined by (2.3).

The HDG method (2.2) can be viewed as a special case of (3.1) with  $\boldsymbol{Q} = \boldsymbol{0}$ . We keep  $\boldsymbol{Q}$  in the numerical scheme (3.1) in order to obtain a general stability estimate, which will facilitate the error analysis of the HDG method.

The following stability result holds for system (3.1). The proof will be given in Section 4.

THEOREM 3.1 Let  $(u_h, q_h, \hat{u}_h)$  be the solution of (3.1). Then there exists a positive constant  $C_{\text{sta}}$  independent of p such that

$$\|\boldsymbol{q}_{h}\|_{\mathscr{T}_{h}}^{2} + \kappa^{2} \|\boldsymbol{u}_{h}\|_{\mathscr{T}_{h}}^{2} \lesssim C_{\text{sta}} M(f, g, \boldsymbol{Q}), \qquad (3.2)$$

where

$$M(f,g,\boldsymbol{Q}) := \|f\|_{\mathscr{T}_h}^2 + \|g\|_{\partial\Omega}^2 + \|\boldsymbol{Q}\|_{\mathscr{T}_h}^2$$

and

$$C_{\text{sta}} := \{\kappa^2 \tau_{\min}^{-1} + \kappa + \tau_{\min}^{-1} h^{-3} + \tau_{\max} h^{-1}\}^2 + 1.$$
(3.3)

To prevent the proliferation of constants, henceforth we also use the notation  $A \leq B$  to represent the inequality  $A \leq (constant) \times B$ , where the positive constant is independent of the mesh size *h* and wave number  $\kappa$ .

REMARK 3.1 The stability estimates for the continuous Helmholtz equation (1.1) were established in Melenk (1995) for two dimensional domains and in Cummings & Feng (2006) for the three dimensional domains. The stability constant in both cases is proved to be  $(1 + 1/\kappa^2 + 1/\kappa^4)$ . In Hetmaniuk (2007) the result was extended to the Helmholtz equation with mixed boundary conditions.

REMARK 3.2 The HDG method (2.2) is stable and well-posed for all wave numbers  $\kappa > 0$  and mesh size h > 0. As a comparison, in Griesmaier & Monk (2011) the HDG method for Helmholtz equation (with Dirichlet boundary condition) is proven to be stable only if *h* satisfies a constraint such that  $h\kappa^2$  is sufficiently small.

REMARK 3.3 The stability estimate in Theorem 3.1 holds for all polynomial degrees. More precisely, for  $p \ge 0$ , if we take  $\tau = O(\kappa)$  in (2.3) and assume  $h\kappa \le 1$ , the stability constant  $C_{\text{sta}}$  in (3.3) behaves as  $(\kappa + \kappa^{-1}h^{-3})^2$ . Note that the IPDG methods introduced in Feng & Wu (2009) and the "LDG" methods introduced in Feng & Xing (2010) have similar stability results only for piecewise linear polynomials. The results of IPDG methods using higher order polynomial spaces can be found in Feng & Wu (2011). Note that the "LDG" methods studied in Feng & Xing (2010) are not local discontinuous Galerkin methods since the numerical trace of u does depend on  $q_h$ . Actually, for some special choice of the stabilization parameters, the "LDG" methods treated therein are HDG methods.

In the rest of this section, we derive the error estimates for the solution of HDG method (2.2). This will be done in two steps. First we introduce a projection operator  $\Pi_h$  inspired by the particular form of the numerical flux  $\hat{q}_h$  from (2.3). Second, we bound the error between the projection and the HDG solution by using the stability results obtained in Theorem 3.1.

#### 3.2 The Projection

Given a function  $(\boldsymbol{q}, u)$  in  $\boldsymbol{H}^1(\mathscr{T}_h) \times H^1(\mathscr{T}_h)$ , we define its projection  $\Pi_h(\boldsymbol{q}, u) := (\boldsymbol{\pi} \boldsymbol{q}, \Pi u) \in \boldsymbol{V}_h \times W_h$  as follows. On an arbitrary element *T* of the triangulation  $\mathscr{T}_h$ , we require that

$$(\boldsymbol{\pi}\boldsymbol{q}-\boldsymbol{q},\boldsymbol{\tau})_T=0 \qquad \forall \, \boldsymbol{\nu} \in \boldsymbol{P}_{p-1}(T),$$
 (3.4a)

$$(\Pi u - u, w)_T = 0 \qquad \forall w \in P_{p-1}(T), \tag{3.4b}$$

$$\langle (\boldsymbol{\pi}\boldsymbol{q} - \boldsymbol{q}) \cdot \boldsymbol{n} + i\tau(\Pi u - u), \mu \rangle_F = 0 \qquad \forall \ \mu \in P_p(F), \tag{3.4c}$$

for all  $T \in \mathcal{T}_h$  and faces/edges F of the element T. The projection defined above is similar to the one used in Cockburn *et al.* (2010).

We denote by  $\|v\|_{H^{\ell}(T)}$  the usual  $H^{\ell}$  norm of v on the domain T. We set  $H^{\ell}(T) := [H^{\ell}(T)]^{d}$  and  $\|v\|_{H^{\ell}(T)} := \sum_{i=1}^{d} \|v_i\|_{H^{\ell}(T)}$ . The next result states that the projection  $\Pi_h$  is well defined and has reasonable approximation properties.

THEOREM 3.2 (Cockburn *et al.*, 2010, Theorem 2.1) Suppose that  $\tau > 0$  in (3.4c) and  $p \ge 0$ . Then the system (3.4) is uniquely solvable for  $\pi q$  and  $\Pi u$ . Furthermore, there is a constant *C* independent of *T*,  $\kappa$  and  $\tau$  such that

$$\begin{aligned} \|\boldsymbol{\pi}\boldsymbol{q} - \boldsymbol{q}\|_{T} &\leq Ch_{T}^{\ell_{\boldsymbol{q}}+1} \, |\boldsymbol{q}|_{\boldsymbol{H}^{\ell_{\boldsymbol{q}}+1}(T)} + Ch_{T}^{\ell_{\boldsymbol{u}}+1} \tau_{T}^{*} |\boldsymbol{u}|_{\boldsymbol{H}^{\ell_{\boldsymbol{u}}+1}(T)}, \\ \|\boldsymbol{\Pi}\boldsymbol{u} - \boldsymbol{u}\|_{T} &\leq Ch_{T}^{\ell_{\boldsymbol{u}}+1} \, |\boldsymbol{u}|_{\boldsymbol{H}^{\ell_{\boldsymbol{u}}+1}(T)} + C\frac{h_{T}^{\ell_{\boldsymbol{q}}+1}}{\tau_{T}^{\max}} |\nabla \cdot \boldsymbol{q}|_{\boldsymbol{H}^{\ell_{\boldsymbol{q}}+1}(T)}, \end{aligned}$$

for  $\ell_{\boldsymbol{q}}, \ell_{\boldsymbol{u}}$  in [0, p]. Here  $\tau_T^{\max} := \max \tau \big|_{\partial T} > 0$ , and  $\tau_T^* := \max \tau \big|_{\partial T \setminus F^*}$ , where  $F^*$  is a face of T at which  $\tau \big|_{\partial T}$  is maximum.

Here, we assume that the exact wave solution *u* satisfies

$$|u|_{H^{p+1}(\Omega)} \lesssim \kappa |u|_{H^p(\Omega)} \quad \text{ for } p \ge 0.$$

Under the assumptions of Theorem 3.1, we arrive at the following result.

COROLLARY 3.1 Assume that  $u \in H^{p+2}(\Omega)$  and  $f \in H^{p+1}(\Omega)$ , there exists constant C such that

$$\|\boldsymbol{\pi}\boldsymbol{q} - \boldsymbol{q}\|_{T} \lesssim (h\kappa)^{p+1} (1 + \tau_{max}\kappa^{-1}) \|\boldsymbol{q}\|_{L_{2}(T)}$$
  
$$\|\Pi u - u\|_{T} \lesssim (h\kappa)^{p+1} (1 + \kappa^{2}\tau_{min}^{-1}) \|u\|_{L_{2}(T)}.$$

# 3.3 Error Estimates

Note that because  $\tau$  is piecewise constant,

$$\langle \tau(P_M u - u), \mu \rangle_{\partial, \mathcal{T}_h} = 0 \qquad \text{for all } \mu \in M_h.$$
 (3.5)

The projection of the errors satisfy the following. The proof will be given in Section 4.

LEMMA 3.1 Define the projection of errors  $\boldsymbol{\varepsilon}_h^{\boldsymbol{q}} := \boldsymbol{\pi} \boldsymbol{q} - \boldsymbol{q}_h$ ,  $\boldsymbol{\varepsilon}_h^{\boldsymbol{u}} := \Pi \boldsymbol{u} - \boldsymbol{u}_h$ , and  $\boldsymbol{\varepsilon}_h^{\widehat{\boldsymbol{u}}} := P_M \boldsymbol{u} - \widehat{\boldsymbol{u}}_h$ . We have

$$(\boldsymbol{\varepsilon}_{h}^{\boldsymbol{q}},\boldsymbol{\tau}_{h})_{\mathcal{T}_{h}}-(\boldsymbol{\varepsilon}_{h}^{\boldsymbol{u}},\nabla\cdot\boldsymbol{\tau}_{h})_{\mathcal{T}_{h}}+\langle\boldsymbol{\varepsilon}_{h}^{\boldsymbol{u}},\boldsymbol{\tau}_{h}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_{h}}=\left((\boldsymbol{\pi}\boldsymbol{q}-\boldsymbol{q}),\boldsymbol{\tau}_{h}\right)_{\mathcal{T}_{h}},\tag{3.6a}$$

$$-\kappa^{2}(\boldsymbol{\varepsilon}_{h}^{u},\boldsymbol{v}_{h})_{\mathcal{T}_{h}}-(\boldsymbol{\varepsilon}_{h}^{\boldsymbol{q}},\nabla\boldsymbol{v}_{h})_{\mathcal{T}_{h}}+\langle\widehat{\boldsymbol{\varepsilon}}_{h}\cdot\boldsymbol{n},\boldsymbol{v}_{h}\rangle_{\partial\mathcal{T}_{h}}=-\big(\kappa^{2}(\boldsymbol{\Pi}\boldsymbol{u}-\boldsymbol{u}),\boldsymbol{v}_{h}\big)_{\mathcal{T}_{h}},$$
(3.6b)

$$\langle -\widehat{\boldsymbol{\varepsilon}}_h \cdot \boldsymbol{n} + i\kappa \varepsilon_h^u, \mu_h \rangle_{\partial\Omega} = 0,$$
 (3.6c)

$$\langle \widehat{\boldsymbol{\varepsilon}}_h \cdot \boldsymbol{n}, \boldsymbol{\mu}_h \rangle_{\partial \mathscr{T}_h \setminus \partial \Omega} = 0, \qquad (3.6d)$$

for all  $(\boldsymbol{\tau}_h, v_h, \mu_h) \in \boldsymbol{V}_h \times W_h \times M_h$ . Here,

$$\widehat{\boldsymbol{\varepsilon}}_{h} \cdot \boldsymbol{n} := \boldsymbol{\varepsilon}_{h}^{\boldsymbol{q}} \cdot \boldsymbol{n} + i\tau(\boldsymbol{\varepsilon}_{h}^{\mu} - \boldsymbol{\varepsilon}_{h}^{\widehat{\mu}}) = P_{M}(\boldsymbol{q} \cdot \boldsymbol{n}) - \widehat{\boldsymbol{q}}_{h} \cdot \boldsymbol{n}. \qquad \text{on } \partial \mathscr{T}_{h} \setminus \partial \Omega.$$
(3.7)

In view of (2.3), (3.1), (3.6) and (3.7), we observe that the projection of errors is tailored to the very structure of the numerical trace. Therefore, we can directly apply Theorem 3.1 with  $\boldsymbol{q}_h = \boldsymbol{\varepsilon}_h^{\boldsymbol{q}}$ ,  $u_h = \boldsymbol{\varepsilon}_h^{\boldsymbol{u}}$ ,  $\hat{u}_h = \boldsymbol{\varepsilon}_h^{\hat{u}}$ ,  $\hat{\boldsymbol{q}}_h = \hat{\boldsymbol{\varepsilon}}_h$ ,  $\boldsymbol{Q} = \boldsymbol{\pi} \boldsymbol{q} - \boldsymbol{q}$ ,  $f = -\kappa^2 (\Pi u - u)$  and g = 0 to obtain the next Theorem.

THEOREM 3.3 Under the assumptions of Theorem 3.1, we have

$$\|\boldsymbol{\pi}\boldsymbol{q}-\boldsymbol{q}_{h}\|_{\mathscr{T}_{h}}^{2}+\kappa^{2}\|\boldsymbol{\Pi}\boldsymbol{u}-\boldsymbol{u}_{h}\|_{\mathscr{T}_{h}}^{2}\lesssim C_{\mathrm{sta}}\big(\|\boldsymbol{\pi}\boldsymbol{q}-\boldsymbol{q}\|_{\mathscr{T}_{h}}^{2}+\kappa^{4}\|\boldsymbol{\Pi}\boldsymbol{u}-\boldsymbol{u}\|_{\mathscr{T}_{h}}^{2}\big),$$

where  $C_{\text{sta}}$  is defined in (3.3).

We are ready to state the explicit bounds for the errors of HDG method (3.1) in terms of wave number  $\kappa$  and mesh size *h*. The proof follows from Corollary 3.1, Theorem 3.3 and an application of the triangle inequality.

COROLLARY 3.2 Let (q, u) and  $(q_h, u_h)$  denote the solutions of (1.3) and (2.2), respectively. Assume that  $u \in H^{p+2}(\Omega)$  and  $f \in H^{p+1}(\Omega)$ . There exist two positive constants  $C_1$  and  $C_2$  such that the following error estimates hold:

$$\|\boldsymbol{q} - \boldsymbol{q}_h\|_{\mathscr{T}_h}^2 \leq \{C_1(h\kappa)^{2p+2}(1 + \tau_{max}\kappa^{-1})^2 + c_1^2 + c_2^2 +$$

$$\|u - u_{h}\|_{\mathscr{T}_{h}}^{2} \leq \{C_{1}(h\kappa)^{2p+2}(1 + \kappa^{2}\tau_{min}^{-1})^{2} + C_{2}C_{\mathrm{sta}}(h\kappa)^{2p+2}[(1 + \tau_{max}\kappa^{-1})^{2} + \kappa^{4}(\kappa^{-1} + \kappa\tau_{min}^{-1})^{2}]\} \|u\|_{\mathscr{T}_{h}}^{2},$$

$$(3.9)$$

where  $C_{\text{sta}}$  is defined in (3.3).

REMARK 3.4 The error estimates above hold for any  $\kappa$  and h. We recall that in Ihlenburg & Babuška (1995) the preasymptotic error estimates for the finite element method solution were proven only in the one dimensional case provided that  $\kappa h \leq 1$ .

REMARK 3.5 The second term on the right-hand side of (3.9) is the pollution term for  $\|\boldsymbol{q} - \boldsymbol{q}_h\|_{\mathcal{P}_h}$ .

REMARK 3.6 Taking  $\tau = \kappa$ , we essentially recover the HDG method proposed in Griesmaier & Monk (2011). In this case, we observe that

$$\|\boldsymbol{q}-\boldsymbol{q}_h\|_{\mathscr{T}_h}^2 \leqslant C_1(h\kappa)^{2p+2} + C_2(h\kappa)^{2p+2}\kappa^4 (\kappa+\kappa^{-1}h^{-3})^2.$$

Note that when  $h\kappa^2$  is sufficiently small, optimal order convergence for  $\mathbf{q}$  (with respect to h) is proved in Griesmaier & Monk (2011) for Dirichlet boundary condition. The stabilization function in the definition of numerical trace therein is taken to be constant on each edge/face  $F \in \mathcal{E}_h$ . Moreover, in the case  $h\kappa^2$  is sufficiently small, one can follow the similar approach in Griesmaier & Monk (2011) to establish the super-convergence of the approximation  $u_h$  to the projection  $\Pi u$ , and further obtain an approximation  $u_h^*$  of improved accuracy by a postprocessing scheme outlined in Cockburn *et al.* (2010). In the case p = 1, the above error estimate is similar with the one obtained by the IPDG method in Feng & Wu (2009) for piecewise linear polynomials.

REMARK 3.7 In Chen *et al.* (2012) a HDG method for Helmholtz equation (1.1) with high wave numbers is proposed and analyzed.  $L_2$ -projections and duality techniques are used therein to prove the stability and convergence properties of the numerical scheme. Numerical experiments are performed with  $\tau = \frac{p}{\kappa h}$ . Numerical examinations of the HDG methods studied in this paper constitute future work.

# 4. Proofs

In this section, we provide the proofs of our main results presented in Section 3. To derive the stability estimate in Theorem 3.1, we mimic the analysis for the Helmholtz equation (Cummings, 2001; Cummings & Feng, 2006; Hetmaniuk, 2007) and that for the discrete problem based on IPDG methods (Feng & Wu, 2009). The key ingredient is to take the test function  $v_h = \boldsymbol{\alpha} \cdot \nabla u_h$  and  $\boldsymbol{\tau}_h = \boldsymbol{\alpha}(\nabla \cdot \boldsymbol{q}_h)$  and to use Rellich identity on each element. Here  $\boldsymbol{\alpha}(\boldsymbol{x}) := \boldsymbol{x} - \boldsymbol{x}_{\Omega}$ . This idea can be traced back to Melenk (1995).

Let us begin by gathering several useful lemmas.

LEMMA 4.1 (TWO INTEGRAL IDENTITIES) It holds that

$$\kappa^{2} \operatorname{Re}(u_{h}, \boldsymbol{\alpha} \cdot \nabla u_{h})_{\mathscr{T}_{h}} = \frac{\kappa^{2}}{2} \sum_{T \in \mathscr{T}_{h}} \int_{\partial T} \boldsymbol{\alpha} \cdot \boldsymbol{n} |u_{h}|^{2} ds - \frac{\kappa^{2} d}{2} ||u_{h}||_{\mathscr{T}_{h}}^{2}.$$

$$(4.1)$$

$$\operatorname{Re}\left\{2\left(\boldsymbol{q}_{h},\boldsymbol{\alpha}(\nabla\cdot\boldsymbol{q}_{h})\right)_{\mathscr{T}_{h}}\right\} = \operatorname{Re}\left\{2\langle\boldsymbol{\alpha}\cdot\boldsymbol{q}_{h},\boldsymbol{q}_{h}\cdot\boldsymbol{n}\rangle_{\partial\mathscr{T}_{h}} - \langle(\boldsymbol{\alpha}\cdot\boldsymbol{n}),|\boldsymbol{q}_{h}|^{2}\rangle_{\partial\mathscr{T}_{h}} + (d-2)\|\boldsymbol{q}_{h}\|_{\mathscr{T}_{h}}^{2} + 2\left(A(\boldsymbol{q}_{h})\boldsymbol{\alpha},\boldsymbol{q}_{h}\right)_{\mathscr{T}_{h}}\right\},\tag{4.2}$$

where  $A(\boldsymbol{q}_h)$  is a  $d \times d$  matrix whose component

$$(A\boldsymbol{q}_h)_{ij} = \frac{\partial \boldsymbol{q}_h^i}{\partial \boldsymbol{x}^j} - \frac{\partial \boldsymbol{q}_h^j}{\partial \boldsymbol{x}^i},\tag{4.3}$$

where  $\boldsymbol{q}_{h}^{i}$  denote the *i*-th component of  $\boldsymbol{q}_{h}$ .

*Proof.* The identity (4.1) can be proved by a direct calculation. Details can be found in Feng & Wu (2009). Next we prove identity (4.2). For each  $T \in \mathcal{T}_h$ ,

$$(2-d)\int_{T}\boldsymbol{q}_{h}\cdot\overline{\boldsymbol{q}}_{h}dx+2\int_{T}\boldsymbol{q}_{h}\cdot[\boldsymbol{\alpha}(\nabla\cdot\overline{\boldsymbol{q}}_{h})]dx$$
$$=(2-d)\int_{T}\boldsymbol{q}_{h}\cdot\overline{\boldsymbol{q}}_{h}dx+2\int_{T}\nabla\cdot[(\boldsymbol{q}_{h}\cdot\boldsymbol{\alpha})\overline{\boldsymbol{q}}_{h}]dx-2\int_{T}[(\nabla\boldsymbol{q}_{h})\boldsymbol{\alpha}]\cdot\overline{\boldsymbol{q}}_{h}dx-2\int_{T}[\boldsymbol{q}_{h}(\nabla\boldsymbol{\alpha})]\cdot\overline{\boldsymbol{q}}_{h}dx.$$

By the definition of  $\boldsymbol{\alpha}$ , we have  $-2\int_T [\boldsymbol{q}_h(\nabla \boldsymbol{\alpha})] \cdot \overline{\boldsymbol{q}}_h dx = -2\int_T \boldsymbol{q}_h \cdot \overline{\boldsymbol{q}}_h dx$ . Therefore, by divergence theorem,

$$\begin{aligned} \operatorname{Re}\{(2-d)\int_{T}\boldsymbol{q}_{h}\cdot\overline{\boldsymbol{q}}_{h}dx+2\int_{T}\boldsymbol{q}_{h}\cdot[\boldsymbol{\alpha}(\nabla\cdot\overline{\boldsymbol{q}}_{h})]dx\} \\ =&\operatorname{Re}\{-d\int_{T}\boldsymbol{q}_{h}\cdot\overline{\boldsymbol{q}}_{h}dx+2\int_{\partial T}(\boldsymbol{q}_{h}\cdot\boldsymbol{\alpha})(\overline{\boldsymbol{q}}_{h}\cdot\boldsymbol{n})ds-2\int_{T}[(\nabla\boldsymbol{q}_{h})\boldsymbol{\alpha}]\cdot\overline{\boldsymbol{q}}_{h}dx\} \\ =&\operatorname{Re}\{-d\int_{T}\boldsymbol{q}_{h}\cdot\overline{\boldsymbol{q}}_{h}dx+2\int_{\partial T}(\boldsymbol{q}_{h}\cdot\boldsymbol{\alpha})(\overline{\boldsymbol{q}}_{h}\cdot\boldsymbol{n})ds-\int_{T}\boldsymbol{\alpha}\cdot(\nabla|\boldsymbol{q}_{h}|^{2})dx+2\int_{T}[A(\boldsymbol{q}_{h})\boldsymbol{\alpha}]\cdot\overline{\boldsymbol{q}}_{h}dx\}.\end{aligned}$$

Note that  $\nabla \cdot \boldsymbol{\alpha} = d$ , hence

$$\begin{aligned} \operatorname{Re}\{(2-d)\int_{T}\boldsymbol{q}_{h}\cdot\overline{\boldsymbol{q}}_{h}dx+2\int_{T}\boldsymbol{q}_{h}\cdot[\boldsymbol{\alpha}(\nabla\cdot\overline{\boldsymbol{q}}_{h})]dx\} \\ =&\operatorname{Re}\{-\int_{T}(\nabla\cdot\boldsymbol{\alpha})\boldsymbol{q}_{h}\cdot\overline{\boldsymbol{q}}_{h}dx-\int_{T}\boldsymbol{\alpha}\cdot(\nabla|\boldsymbol{q}_{h}|^{2})dx+2\int_{\partial T}(\boldsymbol{q}_{h}\cdot\boldsymbol{\alpha})(\overline{\boldsymbol{q}}_{h}\cdot\boldsymbol{n})ds+2\int_{T}[A(\boldsymbol{q}_{h})\boldsymbol{\alpha}]\cdot\overline{\boldsymbol{q}}_{h}dx\} \\ =&\operatorname{Re}\{-\int_{T}\nabla\cdot(\boldsymbol{\alpha}|\boldsymbol{q}_{h}|^{2})dx+2\int_{\partial T}(\boldsymbol{q}_{h}\cdot\boldsymbol{\alpha})(\overline{\boldsymbol{q}}_{h}\cdot\boldsymbol{n})ds+2\int_{T}[A(\boldsymbol{q}_{h})\boldsymbol{\alpha}]\cdot\overline{\boldsymbol{q}}_{h}dx\} \\ =&\operatorname{Re}\{-\int_{\partial T}(\boldsymbol{\alpha}\cdot\boldsymbol{n})|\boldsymbol{q}_{h}|^{2}ds+2\int_{\partial T}(\boldsymbol{q}_{h}\cdot\boldsymbol{\alpha})(\overline{\boldsymbol{q}}_{h}\cdot\boldsymbol{n})ds+2\int_{T}[A(\boldsymbol{q}_{h})\boldsymbol{\alpha}]\cdot\overline{\boldsymbol{q}}_{h}dx\}.\end{aligned}$$

Identity (4.2) can be obtained by summing above equation over all  $T \in \mathscr{T}_h$ .

The following lemma shows that the residuals in the elements is related with the residuals on the boundary.

LEMMA 4.2 (ESTIMATE OF RESIDUALS) It holds

$$\|\boldsymbol{q}_{h}+\nabla u_{h}\|_{\mathscr{T}_{h}} \leqslant Ch^{-\frac{1}{2}}\|u_{h}-\widehat{u}_{h}\|_{\partial\mathscr{T}_{h}}.$$
(4.4)

$$\|\nabla \cdot \boldsymbol{q}_{h} - \kappa^{2} \boldsymbol{u}_{h} - f_{h}\|_{\partial \mathcal{T}_{h}} \leqslant Ch^{-1/2} \|\nabla \cdot \boldsymbol{q}_{h} - \kappa^{2} \boldsymbol{u}_{h} - f_{h}\|_{\mathcal{T}_{h}} \leqslant Ch^{-1} \|\tau(\boldsymbol{u}_{h} - \widehat{\boldsymbol{u}}_{h})\|_{\partial \mathcal{T}_{h}},$$
(4.5)

$$\sum_{e \in \mathscr{E}_h^I} \| [\boldsymbol{q}_h \cdot \boldsymbol{n}] \|_{L_2(e)}^2 \leqslant C \| \tau(u_h - \widehat{u}_h) \|_{\partial \mathscr{F}_h}^2, \tag{4.6}$$

$$\sum_{\boldsymbol{e}\in\mathscr{E}_{h}^{I}} \|[\boldsymbol{q}_{h}\times\boldsymbol{n}]\|_{L_{2}(\boldsymbol{e})}^{2} \leqslant Ch^{-2}\|\boldsymbol{u}_{h}-\widehat{\boldsymbol{u}}_{h}\|_{\partial\mathscr{T}_{h}}^{2}.$$
(4.7)

where  $f_h$  is the  $L^2$ -orthogonal projection of the function f into the space  $W_h$ .

Proof. From (3.1a), a simple application of integration by parts yields

$$(\boldsymbol{q}_h + 
abla u_h, \boldsymbol{\tau}_h)_T = \langle u_h - \widehat{u}_h, \boldsymbol{\tau}_h \cdot \boldsymbol{n} 
angle_{\partial T}$$

for each  $T \in \mathscr{T}_h$ . Note that  $\boldsymbol{q}_h + \nabla u_h$  is in the  $L^2$ -orthogonal complement of  $\boldsymbol{V}_h^0(T)$  where

$$\boldsymbol{V}_h^0(T) := \{\boldsymbol{\tau}_h \in \boldsymbol{V}_h | \boldsymbol{\tau}_h \cdot \boldsymbol{n} = 0 \text{ on } \partial T\},\$$

we have

$$\|(\boldsymbol{q}_h + \nabla u_h) \cdot \boldsymbol{n}\|_{\partial T} \leq Ch^{-1/2} \|\boldsymbol{q}_h + \nabla u_h\|_T$$

Taking  $\boldsymbol{\tau}_h = \boldsymbol{q}_h + \nabla u_h$ , estimate (4.4) follows immediately from the inequality above.

We omit the proof of inequality (4.5) since the inequality can be proved in a similar fashion. To prove (4.6), we obtain, due to the single-valuedness of  $\hat{q}_h \cdot n$  (3.1d),

$$\sum_{e \in \mathscr{E}_h^I} \| [\boldsymbol{q}_h \cdot \boldsymbol{n}] \|_{L_2(e)}^2 = \sum_{e \in \mathscr{E}_h^I} \| [(\boldsymbol{q}_h - \widehat{\boldsymbol{q}}_h) \cdot \boldsymbol{n}] \|_{L_2(e)}^2$$
$$\leq \| \boldsymbol{\tau}(\boldsymbol{u}_h - \widehat{\boldsymbol{u}}_h) \|_{\partial \mathscr{T}_h}^2,$$

by the definition of the numerical flux (2.3).

It remains to show (4.7). Note that by inverse inequality, we have

$$\begin{split} \sum_{e \in \mathscr{E}_{h}^{I}} \| [\boldsymbol{q}_{h} \times \boldsymbol{n}] \|_{L_{2}(e)}^{2} &\leq \sum_{e \in \mathscr{E}_{h}^{I}} \| [(\boldsymbol{q}_{h} + \nabla u_{h}) \times \boldsymbol{n}] \|_{L_{2}(e)}^{2} + \sum_{e \in \mathscr{E}_{h}^{I}} \| [\nabla u_{h} \times \boldsymbol{n}] \|_{L_{2}(e)}^{2} \\ &\lesssim \sum_{T \in \mathscr{T}_{h}} h^{-1} \| \boldsymbol{q}_{h} + \nabla u_{h} \|_{L_{2}(T)}^{2} + \sum_{e \in \mathscr{E}_{h}^{I}} h^{-2} \| [u_{h}] \|_{L_{2}(e)}^{2} \\ &\lesssim h^{-2} \| u_{h} - \widehat{u}_{h} \|_{\mathscr{C}_{\mathcal{T}}}^{2}, \end{split}$$

by (4.4). This completes the proof of the lemma.

Finally, we need to derive a control on the jumps and show the following Gårding-type identity. LEMMA 4.3 (GÅRDING-TYPE IDENTITY) It holds

$$\|\boldsymbol{q}_{h}\|_{\mathcal{T}_{h}}^{2} = \operatorname{Re}\left\{(f, u_{h})_{\mathcal{T}_{h}} + \langle g, \widehat{u}_{h} \rangle_{\partial \Omega}\right\} + \kappa^{2} \|u_{h}\|_{\mathcal{T}_{h}}^{2}.$$
(4.8)

$$\kappa \|\widehat{u}_{h}\|_{\partial\Omega}^{2} + \langle \tau(u_{h} - \widehat{u}_{h}), (u_{h} - \widehat{u}_{h}) \rangle_{\partial\mathcal{T}_{h}} = \operatorname{Im}\left\{ (f, u_{h})_{\mathcal{T}_{h}} + \langle g, \widehat{u}_{h} \rangle_{\partial\Omega} \right\}.$$

$$(4.9)$$

*Proof.* Taking  $\boldsymbol{\tau}_h = \boldsymbol{q}_h$  and  $v_h = u_h$  in (2.2), we obtain the following identity:

$$\begin{split} \|\boldsymbol{q}_{h}\|_{\mathscr{T}_{h}}^{2} &- (\nabla \cdot \boldsymbol{q}_{h}, u_{h})_{\mathscr{T}_{h}} + \langle \boldsymbol{q}_{h} \cdot \boldsymbol{n}, \widehat{u}_{h} \rangle_{\partial \mathscr{T}_{h}} = 0, \\ &- (\boldsymbol{q}_{h}, \nabla u_{h})_{\mathscr{T}_{h}} - \kappa^{2} (u_{h}, u_{h})_{\mathscr{T}_{h}} + \langle \widehat{\boldsymbol{q}}_{h} \cdot \boldsymbol{n}, u_{h} \rangle_{\partial \mathscr{T}_{h}} = (f, u_{h})_{\mathscr{T}_{h}}, \end{split}$$

Adding the equations above, we obtain

$$\|\boldsymbol{q}_{h}\|_{\mathscr{T}_{h}}^{2}-\boldsymbol{\kappa}^{2}(\boldsymbol{u}_{h},\boldsymbol{u}_{h})_{\mathscr{T}_{h}}+\langle(\widehat{\boldsymbol{q}}_{h}-\boldsymbol{q}_{h})\cdot\boldsymbol{n},(\boldsymbol{u}_{h}-\widehat{\boldsymbol{u}}_{h})\rangle_{\partial\mathscr{T}_{h}}+i\boldsymbol{\kappa}\langle\widehat{\boldsymbol{q}}_{h}\cdot\boldsymbol{n},\widehat{\boldsymbol{u}}_{h}\rangle_{\partial\Omega}=(f,\boldsymbol{u}_{h})_{\mathscr{T}_{h}},$$

by a simple calculation. Due to the Robin boundary condition (2.2c) and the definition of the numerical flux (2.3), we get

$$\|\boldsymbol{q}_{h}\|_{\mathscr{T}_{h}}^{2} - \kappa^{2}(u_{h}, u_{h})_{\mathscr{T}_{h}} + i\langle \tau(u_{h} - \widehat{u}_{h}), (u_{h} - \widehat{u}_{h})\rangle_{\partial \mathscr{T}_{h}} + i\kappa\langle \widehat{u}_{h}, \widehat{u}_{h}\rangle_{\partial \Omega} = (f, u_{h})_{\mathscr{T}_{h}} + \langle g, \widehat{u}_{h}\rangle_{\partial \Omega}.$$

The lemma follows from taking the real part and the imaginary part of the above equation.

Now we are ready to prove the stability estimate, Theorem 3.1.

*Proof of Theorem 3.1.* **Step 1.** From integration by parts, equations (3.1a) and (3.1b) defining HDG methods that

$$(\boldsymbol{q}_{h} + \nabla u_{h}, \boldsymbol{\tau}_{h})_{\mathscr{T}_{h}} = \langle u_{h} - \widehat{u}_{h}, \boldsymbol{\tau}_{h} \cdot \boldsymbol{n} \rangle_{\partial \mathscr{T}_{h}} + (\boldsymbol{Q}, \boldsymbol{\tau}_{h})_{\mathscr{T}_{h}}$$
$$(f_{h} - \nabla \cdot \boldsymbol{q}_{h} + \kappa^{2} u_{h}, v_{h})_{\mathscr{T}_{h}} = \langle (\widehat{\boldsymbol{q}}_{h} - \boldsymbol{q}_{h}) \cdot \boldsymbol{n}, v_{h} \rangle_{\partial \mathscr{T}_{h}},$$

Taking the test function  $\boldsymbol{\tau}_h = \boldsymbol{\alpha}(\nabla \cdot \boldsymbol{q}_h)$  and  $v = \boldsymbol{\alpha} \cdot \nabla u_h$ , adding two equations above and taking the real part, we obtain

$$\operatorname{Re}\left\{\left(\boldsymbol{\alpha}(\nabla \cdot \boldsymbol{q}_{h}), \boldsymbol{q}_{h}\right)_{\mathscr{T}_{h}} + \kappa^{2}\left(u_{h}, \boldsymbol{\alpha} \cdot \nabla u_{h}\right)_{\mathscr{T}_{h}} + \langle(\boldsymbol{\alpha} \cdot \boldsymbol{n})(\nabla \cdot \boldsymbol{q}_{h}), (\widehat{u}_{h} - u_{h})\rangle_{\partial \mathscr{T}_{h}} - \langle(\widehat{\boldsymbol{q}}_{h} - \boldsymbol{q}_{h}) \cdot \boldsymbol{n}, \boldsymbol{\alpha} \cdot \nabla u_{h}\rangle_{\partial \mathscr{T}_{h}}\right\} = \operatorname{Re}\left\{\left(\boldsymbol{\mathcal{Q}}, \boldsymbol{\alpha}(\nabla \cdot \boldsymbol{q}_{h})\right)_{\mathscr{T}_{h}} - \left(f_{h}, \boldsymbol{\alpha} \cdot \nabla u_{h}\right)_{\mathscr{T}_{h}}\right\}.$$

$$(4.10)$$

It follows from identities (4.1) and (4.2) that

$$-\frac{d\kappa^{2}}{2}\|\boldsymbol{u}_{h}\|_{\mathcal{T}_{h}}^{2}+\frac{d-2}{2}\|\boldsymbol{q}_{h}\|_{\mathcal{T}_{h}}^{2}+\operatorname{Re}\left\{-\frac{1}{2}\langle(\boldsymbol{\alpha}\cdot\boldsymbol{n}),|\boldsymbol{q}_{h}|^{2}\rangle_{\partial\mathcal{T}_{h}}+\langle\boldsymbol{\alpha}\cdot\boldsymbol{q}_{h},\boldsymbol{q}_{h}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_{h}}\right.\\\left.-\langle(\widehat{\boldsymbol{q}}_{h}-\boldsymbol{q}_{h})\cdot\boldsymbol{n},\boldsymbol{\alpha}\cdot\nabla\boldsymbol{u}_{h}\rangle_{\partial\mathcal{T}_{h}}+\frac{\kappa^{2}}{2}\langle(\boldsymbol{\alpha}\cdot\boldsymbol{n}),|\boldsymbol{u}_{h}|^{2}\rangle_{\partial\mathcal{T}_{h}}+\langle(\widehat{\boldsymbol{u}}_{h}-\boldsymbol{u}_{h}),(\boldsymbol{\alpha}\cdot\boldsymbol{n})(\nabla\cdot\boldsymbol{q}_{h})\rangle_{\partial\mathcal{T}_{h}}\right.\\\left.+\left(A(\boldsymbol{q}_{h})\boldsymbol{\alpha},\boldsymbol{q}_{h}\right)_{\mathcal{T}_{h}}\right\}=\operatorname{Re}\left\{\left(\boldsymbol{Q},\boldsymbol{\alpha}(\nabla\cdot\boldsymbol{q}_{h})\right)_{\mathcal{T}_{h}}-\left(f_{h},\boldsymbol{\alpha}\cdot\nabla\boldsymbol{u}_{h}\right)_{\mathcal{T}_{h}}\right\}.$$

With a simple manipulation, we have

$$\frac{d\kappa^2}{2} \|u_h\|_{\mathscr{T}_h}^2 - \frac{d-2}{2} \|\boldsymbol{q}_h\|_{\mathscr{T}_h}^2 = T_1 + T_2 + T_3 + T_4,$$
(4.11)

where

$$T_{1} := \operatorname{Re}\left\{-\frac{1}{2}\langle(\boldsymbol{\alpha}\cdot\boldsymbol{n}),|\boldsymbol{q}_{h}|^{2}\rangle_{\partial\mathcal{T}_{h}} + \langle\boldsymbol{\alpha}\cdot\boldsymbol{q}_{h},\boldsymbol{q}_{h}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_{h}} - \langle(\widehat{\boldsymbol{q}}_{h}-\boldsymbol{q}_{h})\cdot\boldsymbol{n},\boldsymbol{\alpha}\cdot\nabla\boldsymbol{u}_{h}\rangle_{\partial\mathcal{T}_{h}}\right\},$$

$$T_{2} := \operatorname{Re}\left\{\frac{\kappa^{2}}{2}\langle(\boldsymbol{\alpha}\cdot\boldsymbol{n}),|\boldsymbol{u}_{h}|^{2}\rangle_{\partial\mathcal{T}_{h}} + \langle(\widehat{\boldsymbol{u}}_{h}-\boldsymbol{u}_{h}),(\boldsymbol{\alpha}\cdot\boldsymbol{n})(\nabla\cdot\boldsymbol{q}_{h})\rangle_{\partial\mathcal{T}_{h}}\right\},$$

$$T_{3} := \operatorname{Re}\left\{(f_{h},\boldsymbol{\alpha}\cdot\nabla\boldsymbol{u}_{h})_{\mathcal{T}_{h}} - (\boldsymbol{Q},\boldsymbol{\alpha}(\nabla\cdot\boldsymbol{q}_{h}))_{\mathcal{T}_{h}}\right\},$$

$$T_{4} := \operatorname{Re}\left\{(A(\boldsymbol{q}_{h})\boldsymbol{\alpha},\boldsymbol{q}_{h})_{\mathcal{T}_{h}}\right\}.$$

11 of 17

We estimate the terms on the right-hand side of (4.11) separately. Step 2: Estimate of  $T_1$ . By a straightforward calculation, we can rewrite  $T_1$  as follows.

$$T_{1} = \operatorname{Re}\left\{-\frac{1}{2}\langle(\boldsymbol{\alpha}\cdot\boldsymbol{n}),|\boldsymbol{q}_{h}|^{2}\rangle_{\partial\mathscr{T}_{h}} + \langle\boldsymbol{\alpha}\cdot\boldsymbol{q}_{h},\widehat{\boldsymbol{q}}_{h}\cdot\boldsymbol{n}\rangle_{\partial\mathscr{T}_{h}} - \langle(\widehat{\boldsymbol{q}}_{h}-\boldsymbol{q}_{h})\cdot\boldsymbol{n},\boldsymbol{\alpha}\cdot(\nabla u_{h}+\boldsymbol{q}_{h})\rangle_{\partial\mathscr{T}_{h}}\right\}$$

$$= \operatorname{Re}\left\{-\frac{1}{2}\langle(\boldsymbol{\alpha}\cdot\boldsymbol{n}),|\boldsymbol{q}_{h}|^{2}\rangle_{\partial\mathscr{T}_{h}\setminus\partial\Omega} + \langle\boldsymbol{\alpha}\cdot\boldsymbol{q}_{h},\widehat{\boldsymbol{q}}_{h}\cdot\boldsymbol{n}\rangle_{\partial\mathscr{T}_{h}\setminus\partial\Omega} - \langle(\widehat{\boldsymbol{q}}_{h}-\boldsymbol{q}_{h})\cdot\boldsymbol{n},\boldsymbol{\alpha}\cdot(\nabla u_{h}+\boldsymbol{q}_{h})\rangle_{\partial\mathscr{T}_{h}} - \frac{1}{2}\langle(\boldsymbol{\alpha}\cdot\boldsymbol{n}),|\boldsymbol{q}_{h}|^{2}\rangle_{\partial\Omega} + \langle\boldsymbol{\alpha}\cdot\boldsymbol{q}_{h},\widehat{\boldsymbol{q}}_{h}\cdot\boldsymbol{n}\rangle_{\partial\Omega}\right\}.$$

$$(4.12)$$

We estimate the first two terms of  $T_1$  as follows. Since  $\hat{\boldsymbol{q}}_h \cdot \boldsymbol{n}$  is single-valued (3.1d), we have

$$\begin{aligned} &\operatorname{Re}\left\{-\frac{1}{2}\langle(\boldsymbol{\alpha}\cdot\boldsymbol{n}),|\boldsymbol{q}_{h}|^{2}\rangle_{\partial\mathscr{T}_{h}\setminus\partial\Omega}+\langle\boldsymbol{\alpha}\cdot\boldsymbol{q}_{h},\widehat{\boldsymbol{q}}_{h}\cdot\boldsymbol{n}\rangle_{\partial\mathscr{T}_{h}\setminus\partial\Omega}\right\}\\ &\leqslant \sum_{e\in\mathscr{E}_{h}^{I}}|\boldsymbol{\alpha}\cdot\boldsymbol{n}||\langle[\boldsymbol{q}_{h}]],\{\!\{\boldsymbol{q}_{h}\}\!\}\rangle_{e}|+\sum_{e\in\mathscr{E}_{h}^{I}}\langle[\![\boldsymbol{\alpha}\cdot\boldsymbol{q}_{h}]],\widehat{\boldsymbol{q}}_{h}\cdot\boldsymbol{n}\rangle_{e}\\ &\leqslant (\max_{e\in\mathscr{E}_{h}^{I}}|\boldsymbol{\alpha}\cdot\boldsymbol{n}|)\sum_{e\in\mathscr{E}_{h}^{I}}|\langle[\![\boldsymbol{q}_{h}]],\{\!\{\boldsymbol{q}_{h}\}\!\}\rangle_{e}|+(\max_{e\in\mathscr{E}_{h}^{I}}|\boldsymbol{\alpha}|)\sum_{e\in\mathscr{E}_{h}^{I}}\langle[\![\boldsymbol{q}_{h}]],\boldsymbol{q}_{h}\cdot\boldsymbol{n}+i\tau(u_{h}-\widehat{u}_{h})\rangle_{e},\end{aligned}$$

by (2.3). Moreover, by the definitions of jumps,

$$\|[\boldsymbol{q}_h]\|_{L_2(e)}^2 = \|[\boldsymbol{q}_h \cdot \boldsymbol{n}]\|_{L_2(e)}^2 + \|[\boldsymbol{q}_h \times \boldsymbol{n}]\|_{L_2(e)}^2,$$

we have

$$\operatorname{Re}\left\{-\frac{1}{2}\langle (\boldsymbol{\alpha}\cdot\boldsymbol{n}), |\boldsymbol{q}_{h}|^{2}\rangle_{\partial\mathcal{F}_{h}\setminus\partial\Omega} + \langle \boldsymbol{\alpha}\cdot\boldsymbol{q}_{h}, \widehat{\boldsymbol{q}}_{h}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{F}_{h}\setminus\partial\Omega}\right\}$$

$$\leq \frac{1}{6}\|\boldsymbol{q}_{h}\|_{\mathcal{F}_{h}}^{2} + C_{1}h\|\tau(\boldsymbol{u}_{h}-\widehat{\boldsymbol{u}}_{h})\|_{\partial\mathcal{F}_{h}}^{2}$$

$$+ C_{2}\sum_{e\in\mathcal{E}_{h}^{I}}h^{-1}\|[\boldsymbol{q}_{h}\cdot\boldsymbol{n}]\|_{L_{2}(e)}^{2} + C_{3}\sum_{e\in\mathcal{E}_{h}^{I}}h^{-1}\|[\boldsymbol{q}_{h}\times\boldsymbol{n}]\|_{L_{2}(e)}^{2}$$

$$\leq \frac{1}{6}\|\boldsymbol{q}_{h}\|_{\mathcal{F}_{h}}^{2} + C_{1}h\|\tau(\boldsymbol{u}_{h}-\widehat{\boldsymbol{u}}_{h})\|_{\partial\mathcal{F}_{h}}^{2}$$

$$+ C_{2}h^{-1}\|\tau(\boldsymbol{u}_{h}-\widehat{\boldsymbol{u}}_{h})\|_{\partial\mathcal{F}_{h}}^{2} + C_{3}h^{-3}\|\boldsymbol{u}_{h}-\widehat{\boldsymbol{u}}_{h}\|_{\partial\mathcal{F}_{h}}^{2}.$$

$$(4.13)$$

by (4.6) and (4.7). Next, from the definition of numerical flux (2.3), the third term on the right-hand side of (4.12) satisfies

$$\operatorname{Re}\{-\langle (\widehat{\boldsymbol{q}}_{h}-\boldsymbol{q}_{h})\cdot\boldsymbol{n},\boldsymbol{\alpha}\cdot(\nabla u_{h}+\boldsymbol{q}_{h})\rangle_{\partial\mathcal{T}_{h}}\} \lesssim \|\tau(u_{h}-\widehat{u}_{h})\|_{\partial\mathcal{T}_{h}}^{2} + \|\nabla u_{h}+\boldsymbol{q}_{h}\|_{\partial\mathcal{T}_{h}}^{2}$$

$$\lesssim \|\tau(u_{h}-\widehat{u}_{h})\|_{\partial\mathcal{T}_{h}}^{2} + h^{-1}\|u_{h}-\widehat{u}_{h}\|_{\partial\mathcal{T}_{h}}^{2}.$$
by (4.4)

Finally, by noting that  $\widehat{\boldsymbol{q}}_h \cdot \boldsymbol{n} = i\kappa u_h - P_M g$  on  $e \in \partial \Omega$ , we can estimate the last two terms on the

right-hand side of (4.12) as follows.

Therefore, combining (4.12)–(4.15), we arrive at

**Step 3: Estimate of** *T*<sub>2</sub>**.** We rewrite

$$T_{2} = \operatorname{Re}\left\{\frac{\kappa^{2}}{2}\langle(\boldsymbol{\alpha}\cdot\boldsymbol{n}),|\boldsymbol{u}_{h}|^{2}\rangle_{\partial\mathcal{T}_{h}} + \langle(\widehat{\boldsymbol{u}}_{h}-\boldsymbol{u}_{h}),(\boldsymbol{\alpha}\cdot\boldsymbol{n})(f_{h}+\kappa^{2}\boldsymbol{u}_{h})\rangle_{\partial\mathcal{T}_{h}} \right. \\ \left. + \langle(\widehat{\boldsymbol{u}}_{h}-\boldsymbol{u}_{h}),(\boldsymbol{\alpha}\cdot\boldsymbol{n})(\nabla\cdot\boldsymbol{q}_{h}-f_{h}-\kappa^{2}\boldsymbol{u}_{h})\rangle_{\partial\mathcal{T}_{h}}\right\} \\ = \operatorname{Re}\left\{-\frac{\kappa^{2}}{2}\langle(\boldsymbol{\alpha}\cdot\boldsymbol{n}),|\boldsymbol{u}_{h}-\widehat{\boldsymbol{u}}_{h}|^{2}\rangle_{\partial\mathcal{T}_{h}} + \frac{\kappa^{2}}{2}\langle(\boldsymbol{\alpha}\cdot\boldsymbol{n}),|\widehat{\boldsymbol{u}}_{h}|^{2}\rangle_{\partial\mathcal{T}_{h}} + \langle(\widehat{\boldsymbol{u}}_{h}-\boldsymbol{u}_{h}),(\boldsymbol{\alpha}\cdot\boldsymbol{n})f_{h}\rangle_{\partial\mathcal{T}_{h}} \right. \\ \left. + \langle(\widehat{\boldsymbol{u}}_{h}-\boldsymbol{u}_{h}),(\boldsymbol{\alpha}\cdot\boldsymbol{n})(\nabla\cdot\boldsymbol{q}_{h}-f_{h}-\kappa^{2}\boldsymbol{u}_{h})\rangle_{\partial\mathcal{T}_{h}}\right\}$$

Applying Young's inequality, we obtain

$$T_{2} \leq C_{8}\kappa^{2} \|u_{h} - \widehat{u}_{h}\|_{\partial \mathscr{T}_{h}}^{2} + C_{9}\kappa^{2} \|\widehat{u}_{h}\|_{\partial \Omega}^{2} + C_{10}h^{-1} \|u_{h} - \widehat{u}_{h}\|_{\partial \mathscr{T}_{h}}^{2} + C_{11}h^{-2} \|u_{h} - \widehat{u}_{h}\|_{\partial \mathscr{T}_{h}}^{2} + C_{12} \|f\|_{\mathscr{T}_{h}}^{2} + C_{13}h^{2} \|\nabla \cdot \boldsymbol{q}_{h} - f_{h} - \kappa^{2}u_{h}\|_{\partial \mathscr{T}_{h}}^{2}.$$

By (4.5), we have

$$T_{2} \leq C_{8}\kappa^{2} \|u_{h} - \widehat{u}_{h}\|_{\partial \mathscr{T}_{h}}^{2} + C_{9}\kappa^{2} \|\widehat{u}_{h}\|_{\partial \Omega}^{2} + \{ (C_{10}h^{-1} + C_{11}h^{-2})\tau_{min}^{-1} + C_{14}h\tau_{max} \} \|\tau^{1/2}(u_{h} - \widehat{u}_{h})\|_{\partial \mathscr{T}_{h}}^{2} + C_{12}\|f\|_{\mathscr{T}_{h}}^{2}.$$

Hence, by (4.9),

$$T_{2} \leq \left(C_{8}\kappa^{2}\tau_{\min}^{-1} + C_{9}\kappa + (C_{10}h^{-1} + C_{11}h^{-2})\tau_{\min}^{-1} + C_{14}h\tau_{\max}\right)\operatorname{Im}\left\{(f, u_{h})_{\mathscr{T}_{h}} + \langle g, \widehat{u}_{h} \rangle_{\partial\Omega}\right\} + C_{12}\|f\|_{\mathscr{T}_{h}}^{2}.$$
(4.17)

**Step 4: Estimate of** *T*<sub>3</sub>**.** We rewrite

$$T_{3} = \operatorname{Re}\left\{\left(f_{h}, \boldsymbol{\alpha} \cdot (\nabla u_{h} + \boldsymbol{q}_{h})\right)_{\mathscr{T}_{h}} - \left(f_{h}, \boldsymbol{\alpha} \cdot \boldsymbol{q}_{h}\right) - \left(\boldsymbol{\mathcal{Q}}, \boldsymbol{\alpha}(\nabla \cdot \boldsymbol{q}_{h} - f_{h} - \kappa^{2}u_{h})\right)_{\mathscr{T}_{h}} - \left(\boldsymbol{\mathcal{Q}}, \boldsymbol{\alpha}(f_{h} + \kappa^{2}u_{h})\right)_{\mathscr{T}_{h}}\right\}.$$
(4.18)

Therefore, by applying Young's inequality, we get

$$T_{3} \leq C_{15} \|f\|_{\mathscr{T}_{h}}^{2} + C_{16} \|\nabla u_{h} + \boldsymbol{q}_{h}\|_{\mathscr{T}_{h}}^{2}$$
  
+  $C_{17} \|\nabla \cdot \boldsymbol{q}_{h} - f_{h} - \kappa^{2} u_{h}\|_{\mathscr{T}_{h}}^{2} + \frac{1}{6} \|\boldsymbol{q}_{h}\|_{\mathscr{T}_{h}}^{2} + C_{18} \|\boldsymbol{\mathcal{Q}}\|_{\mathscr{T}_{h}}^{2} + \frac{\kappa^{2}}{6} \|u_{h}\|_{\mathscr{T}_{h}}^{2}.$ 

It then follows from (4.4) and (4.5) that

$$T_{3} \leq C_{16}h^{-1} \|u_{h} - \widehat{u}_{h}\|_{\partial \mathscr{T}_{h}}^{2} + C_{17}h^{-1} \|\tau(u_{h} - \widehat{u}_{h})\|_{\partial \mathscr{T}_{h}}^{2} + C_{15} \|f\|_{\mathscr{T}_{h}}^{2} + \frac{1}{6} \|\boldsymbol{q}_{h}\|_{\mathscr{T}_{h}}^{2} + C_{18} \|\boldsymbol{\mathcal{Q}}\|_{\mathscr{T}_{h}}^{2} + \frac{\kappa^{2}}{6} \|u_{h}\|_{\mathscr{T}_{h}}^{2}.$$

Hence, by (4.9),

$$T_{3} \leq (C_{16}h^{-1}\tau_{min}^{-1} + C_{17}h^{-1}\tau_{max})\operatorname{Im}\left\{(f, u_{h})_{\mathscr{T}_{h}} + \langle g, \widehat{u}_{h} \rangle_{\partial\Omega}\right\}$$

$$+ C_{15}\|f\|_{\mathscr{T}_{h}}^{2} + \frac{1}{6}\|\boldsymbol{q}_{h}\|_{\mathscr{T}_{h}}^{2} + C_{18}\|\boldsymbol{\mathcal{Q}}\|_{\mathscr{T}_{h}}^{2} + \frac{\kappa^{2}}{6}\|u_{h}\|_{\mathscr{T}_{h}}^{2}.$$

$$(4.19)$$

**Step 5: Estimate of**  $T_4$ **.** For the last term  $T_4$ , we have

$$T_{4} = \operatorname{Re}\left\{\left(A(\boldsymbol{q}_{h} + \nabla u_{h})\boldsymbol{\alpha}, \boldsymbol{q}_{h}\right)_{\mathscr{T}_{h}}\right\}$$

$$\leq C_{19}h^{-2}\|\boldsymbol{q}_{h} + \nabla u_{h}\|_{\mathscr{T}_{h}}^{2} + \frac{1}{6}\|\boldsymbol{q}_{h}\|_{\mathscr{T}_{h}}^{2}$$

$$\leq C_{19}h^{-3}\|u_{h} - \widehat{u}_{h}\|_{\mathscr{T}_{h}}^{2} + \frac{1}{6}\|\boldsymbol{q}_{h}\|_{\mathscr{T}_{h}}^{2}$$

$$by (4.4)$$

$$\leq C_{19}h^{-3}\tau_{\min}^{-1}\operatorname{Im}\left\{(f,u_h)_{\mathscr{T}_h}+\langle g,\widehat{u}_h\rangle_{\partial\Omega}\right\}+\frac{1}{6}\|\boldsymbol{q}_h\|_{\mathscr{T}_h}^2.$$
 by (4.9)

Step 6: Conclusion.

Combining (4.11), (4.16), (4.17), (4.19) and (4.20), we arrive at

$$\frac{d}{2}\kappa^{2}\|u_{h}\|_{\mathscr{T}_{h}}^{2} - \frac{d-2}{2}\|\boldsymbol{q}_{h}\|_{\mathscr{T}_{h}}^{2} \leqslant C(\kappa,h,\tau)\operatorname{Im}\left\{(f,u_{h})_{\mathscr{T}_{h}} + \langle g,\widehat{u}_{h}\rangle_{\partial\Omega}\right\} \\ + C_{5}\|g\|_{L_{2}(\partial\Omega)}^{2} + C_{18}\|\boldsymbol{Q}\|_{\mathscr{T}_{h}}^{2} + C_{23}\|f\|_{\mathscr{T}_{h}}^{2} - \frac{C_{\Omega}}{3}\|\boldsymbol{q}_{h}\|_{L_{2}(\partial\Omega)}^{2} + \frac{1}{2}\|\boldsymbol{q}_{h}\|_{\mathscr{T}_{h}}^{2} + \frac{\kappa^{2}}{6}\|u_{h}\|_{\mathscr{T}_{h}}^{2}$$

where

$$C(\kappa,h,\tau) := C_8 \kappa^2 \tau_{\min}^{-1} + C_{20} \kappa + C_{22} \tau_{\min}^{-1} (h^{-3} + h^{-2} + h^{-1}) + C_{21} \tau_{\max} (h^{-1} + h).$$

It follows from (4.8) that

$$\frac{\kappa^{2}}{3} \|u_{h}\|_{\mathscr{T}_{h}}^{2} \leq \{C(\kappa,h,\tau) + \frac{d-1}{2}\} |(f,u_{h})_{\mathscr{T}_{h}} + \langle g,\widehat{u}_{h}\rangle_{\partial\Omega}| 
+ C_{5} \|g\|_{L_{2}(\partial\Omega)}^{2} + C_{18} \|\boldsymbol{\mathcal{Q}}\|_{\mathscr{T}_{h}}^{2} + C_{23} \|f\|_{\mathscr{T}_{h}}^{2} 
\leq \{C(\kappa,h,\tau)^{2} + C_{23}\} M(f,g,\boldsymbol{\mathcal{Q}}) 
+ \frac{\kappa^{2}}{6} \|u_{h}\|_{\mathscr{T}_{h}}^{2} + \frac{\kappa^{2}}{6} \|\widehat{u}_{h}\|_{\partial\Omega}^{2} 
\leq C_{\text{sta}} M(f,g,\boldsymbol{\mathcal{Q}}).$$
(4.21)

where

$$C_{\text{sta}} := \{C_8 \kappa^2 \tau_{\min}^{-1} + C_{20} \kappa + C_{22} \tau_{\min}^{-1} h^{-3} + C_{21} \tau_{\max} h^{-1}\}^2 + C.$$

We have used the fact that  $\kappa^2 \|\widehat{u}_h\|_{\partial\Omega}^2 \leq \kappa^2 \|u_h\|_{\mathscr{T}_h}^2 + CM(f,g,\mathbf{0})$  to derive the last inequality. Finally, by (4.8), we arrive at

$$\|\boldsymbol{q}_h\|_{\mathscr{T}_h}^2 + \kappa^2 \|u_h\|_{\mathscr{T}_h}^2 \lesssim C_{\mathrm{sta}} M(f,g,\boldsymbol{Q}).$$

This completes the proof of the Theorem 3.1.

In the rest of this section, we prove Lemma 3.1. *Proof of Lemma 3.1.* Since the exact solution  $(\mathbf{q}, u)$  of (1.3) satisfies

$$\begin{aligned} (\boldsymbol{q},\boldsymbol{\tau}_{h})_{\mathscr{T}_{h}} - (\boldsymbol{u},\nabla\cdot\boldsymbol{\tau}_{h})_{\mathscr{T}_{h}} + \langle \boldsymbol{u},\boldsymbol{\tau}_{h}\cdot\boldsymbol{n}\rangle_{\partial\mathscr{T}_{h}} &= 0, \\ -\kappa^{2}(\boldsymbol{u},\boldsymbol{v}_{h})_{\mathscr{T}_{h}} - (\boldsymbol{q},\nabla\boldsymbol{v}_{h})_{\mathscr{T}_{h}} + \langle \boldsymbol{q}\cdot\boldsymbol{n},\boldsymbol{v}_{h}\rangle_{\partial\mathscr{T}_{h}} &= (f,\boldsymbol{v}_{h})_{\mathscr{T}_{h}}, \end{aligned}$$

for all  $\boldsymbol{\tau}_h \in \boldsymbol{V}_h$  and  $v_h \in W_h$ , we obtain, by The definitions of  $\Pi_h$  (cf. (3.4)) and  $P_M$ , that

$$(\boldsymbol{\pi}\boldsymbol{q},\boldsymbol{\tau}_{h})_{\mathcal{T}_{h}} - (\boldsymbol{\Pi}\boldsymbol{u},\nabla\cdot\boldsymbol{\tau}_{h})_{\mathcal{T}_{h}} + \langle \boldsymbol{P}_{M}\boldsymbol{u},\boldsymbol{\tau}_{h}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_{h}} = \left((\boldsymbol{\pi}\boldsymbol{q}-\boldsymbol{q}),\boldsymbol{\tau}_{h}\right)_{\mathcal{T}_{h}}, \qquad (4.22a)$$
$$-\kappa^{2}(\boldsymbol{\Pi}\boldsymbol{u},\boldsymbol{v}_{h})_{\mathcal{T}_{h}} - (\boldsymbol{\pi}\boldsymbol{q},\nabla\boldsymbol{v}_{h})_{\mathcal{T}_{h}} + \langle \boldsymbol{\pi}\boldsymbol{q}\cdot\boldsymbol{n} + i\boldsymbol{\tau}(\boldsymbol{\Pi}\boldsymbol{u}-\boldsymbol{u}),\boldsymbol{v}_{h}\rangle_{\partial\mathcal{T}_{h}}$$
$$= \left(f,\boldsymbol{v}_{h}\right)_{\mathcal{T}_{h}} - \left(\kappa^{2}(\boldsymbol{\Pi}\boldsymbol{u}-\boldsymbol{u}),\boldsymbol{v}_{h}\right)_{\mathcal{T}_{h}}, \qquad (4.22b)$$

for all  $\boldsymbol{\tau}_h \in \boldsymbol{V}_h$  and  $v_h \in W_h$ . Now subtracting (2.2a) from (4.22a), we obtain

$$(\boldsymbol{\pi}\boldsymbol{q}-\boldsymbol{q}_h,\boldsymbol{\tau}_h)_{\mathscr{T}_h}-(\boldsymbol{\Pi}\boldsymbol{u}-\boldsymbol{u}_h,\nabla\cdot\boldsymbol{\tau}_h)_{\mathscr{T}_h}+\langle P_M\boldsymbol{u}-\widehat{\boldsymbol{u}}_h,\boldsymbol{\tau}_h\cdot\boldsymbol{n}\rangle_{\partial\mathscr{T}_h}=\left((\boldsymbol{\pi}\boldsymbol{q}-\boldsymbol{q}),\boldsymbol{\tau}_h\right)_{\mathscr{T}_h},$$

which yields (3.6a). Similarly, subtracting (2.2b) from (4.22b) gives

$$-\kappa^{2}(\Pi u - u_{h}, v_{h})_{\mathscr{T}_{h}} - (\boldsymbol{\pi}\boldsymbol{q} - \boldsymbol{q}_{h}, \nabla v_{h})_{\mathscr{T}_{h}} + \langle (\boldsymbol{\pi}\boldsymbol{q} - \widehat{\boldsymbol{q}}_{h}) \cdot \boldsymbol{n} + i\tau(\Pi u - u), v_{h} \rangle_{\partial \mathscr{T}_{h}}$$
$$= - (\kappa^{2}(\Pi u - u), v_{h})_{\mathscr{T}_{h}}$$

It follows from the definition of  $\hat{\boldsymbol{\varepsilon}}_h$  in (3.7) and of the numerical flux  $\hat{\boldsymbol{q}}_h$  in (2.3), we find that

$$\begin{aligned} \widehat{\boldsymbol{\varepsilon}}_h \cdot \boldsymbol{n} &= (\boldsymbol{\pi} \boldsymbol{q} - \boldsymbol{q}_h) \cdot \boldsymbol{n} + i\tau (\Pi u - u_h - P_M u + \widehat{u}_h) \\ &= (\boldsymbol{\pi} \boldsymbol{q} - \widehat{\boldsymbol{q}}_h) \cdot \boldsymbol{n} + i\tau (u_h - \widehat{u}_h) + i\tau (\Pi u - u) - i\tau (u_h + P_M u - \widehat{u}_h - u), \end{aligned}$$

which together with (3.5) yields (3.6b). The equation (3.6c) follows directly from (1.3c), (2.2c), (3.4c) and (3.7).

Furthermore, using (2.3), (3.4c), (3.5) and the definition of  $\hat{\boldsymbol{\varepsilon}}_h \cdot \boldsymbol{n}$  in (3.7) we obtain

$$\begin{split} \langle \widehat{\boldsymbol{\varepsilon}}_h \cdot \boldsymbol{n}, \boldsymbol{\mu} \rangle_{\partial \mathscr{T}_h \setminus \partial \Omega} &= \langle (\boldsymbol{\pi} \boldsymbol{q} - \boldsymbol{q}_h) \cdot \boldsymbol{n} + i\tau (\boldsymbol{\Pi} u - u_h - P_M u + \widehat{u}_h), \boldsymbol{\mu} \rangle_{\partial \mathscr{T}_h \setminus \partial \Omega} \\ &= \langle (\boldsymbol{q} - \boldsymbol{q}_h) \cdot \boldsymbol{n} + i\tau (u - u_h - u + \widehat{u}_h), \boldsymbol{\mu} \rangle_{\partial \mathscr{T}_h \setminus \partial \Omega} \\ &= \langle \boldsymbol{q} \cdot \boldsymbol{n}, \boldsymbol{\mu} \rangle_{\partial \mathscr{T}_h \setminus \partial \Omega} - \langle \widehat{\boldsymbol{q}}_h, \boldsymbol{\mu} \rangle_{\partial \mathscr{T}_h \setminus \partial \Omega}. \end{split}$$

This proves (3.7). Finally, the equation (3.6d) follows from (2.2d) and the fact that  $q \in H(\text{div}; \Omega)$ .

14 of 17

```
REFERENCES
```

### Acknowledgements

The authors would like to thank Prof. Bernardo Cockburn for his helpful comments and suggestions on this work.

### References

- AINSWORTH, M. (2004) Dispersive and dissipative behaviour of high order discontinuous Galerkin finite element methods. *J. Comput. Phys.*, **198**, 106–130.
- AINSWORTH, M., MONK, P. & MUNIZ, W. (2006) Dispersive and dissipative properties of discontinuous Galerkin finite element methods for the second-order wave equation. J. Sci. Comput., 27, 5–40.
- AINSWORTH, M. & WAJID, H. A. (2009) Dispersive and dissipative behavior of the spectral element method. *SIAM J. Numer. Anal.*, **47**, 3910–3937.
- BABUŠKA, I., IHLENBURG, F., PAIK, E. T. & STEFAN, S. A. (1995) A generalized finite element method for solving the Helmholtz equation in two dimensions with minimal pollution. *Comput. Methods Appl. Mech. Engrg.*, **128**, 325–359.
- BABUŠKA, I. & MELENK, J. (1997) The partition of unity method. Int. J. Numer. Meth. Engng., 40, 727–758.
- BABUŠKA, I. & SAUTER, S. A. (2000) Is the pollution effect of the FEM avoidable for the Helmholtz equation considering high wave numbers? *SIAM Rev.*, **42**, 451–484 (electronic).
- BAO, G., WEI, G. & ZHAO, S. (2004) Numerical solution of the Helmholtz equation with high wavenumbers. *Int. J. Numer. Meth. Engng.*, **59**, 389–408.
- BUFFA, A. & MONK, P. (2008) Error estimates for the ultra weak variational formulation of the Helmholtz equation. *M2AN Math. Model. Numer. Anal.*, **42**, 925–940.
- CESSENAT, O. & DESPRÉS, B. (1998) Application of an ultra weak variational formulation of elliptic PDEs to the two-dimensional Helmholtz problem. *SIAM J. Numer. Anal.*, **35**, 255–299 (electronic).
- CESSENAT, O. & DESPRÉS, B. (2003) Using plane waves as base functions for solving time harmonic equations with the ultra weak variational formulation. *J. Comput. Acoust.*, **11**, 227–238. Medium-frequency acoustics.
- CHEN, H., LU, P. & XU, X. (2012) A hybridizable discontinuous Galerkin method for the helmholtz equation with high wave number, arxiv:1207.3419v1.
- CHUNG, E. T. & ENGQUIST, B. (2006) Optimal discontinuous Galerkin methods for wave propagation. *SIAM J. Numer. Anal.*, **44**, 2131–2158 (electronic).
- COCKBURN, B., GOPALAKRISHNAN, J. & LAZAROV, R. (2009) Unified hybridization of discontinuous Galerkin, mixed and continuous Galerkin methods for second order elliptic problems. *SIAM J. Numer. Anal.*, **47**, 1319–1365.
- COCKBURN, B., GOPALAKRISHNAN, J. & SAYAS, F.-J. (2010) A projection-based error analysis of HDG methods. *Math. Comp.*, **79**, 1351–1367.

16 of 17

#### REFERENCES

- CUMMINGS, P. (2001) Analysis of finite element based numerical methods for acoustic waves, elastic waves and fluid-solid interactions in the frequency domain. *Ph.D. thesis, The University of Tennessee*.
- CUMMINGS, P. & FENG, X. (2006) Sharp regularity coefficient estimates for complex-valued acoustic and elastic Helmholtz equations. *Math. Models Methods Appl. Sci.*, **16**, 139–160.
- DERAEMAEKER, A., BABUŠKA, I. & BOUILLARD, P. (1999) Dispersion and pollution of the FEM solution for the Helmholtz equation in one, two and three dimensions. *Int. J. Numer. Meth. Engng.*, **46**, 471–499.
- FARHAT, C., HARARI, I. & FRANCA, L. (2001) The discontinuous enrichment method. Comput. Methods Appl. Mech. Engrg., 190, 6455–6479.
- FENG, X. & WU, H. (2009) Discontinuous Galerkin methods for the Helmholtz equation with large wave number. SIAM J. Numer. Anal., 47, 2872–2896.
- FENG, X. & WU, H. (2011) *hp*-discontinuous Galerkin methods for the Helmholtz equation with large wave number. *Math. Comp.*, **80**, 1997–2024.
- FENG, X. & XING, Y. (2010) Absolutely stable local discontinuous Galerkin methods for the Helmholtz equation with large wave number. *Math. Comp., arXiv:1010.4563.*
- GERDES, K. & IHLENBURG, F. (1999) On the pollution effect in FE solutions of the 3D-Helmholtz equation. *Comput. Methods Appl. Mech. Engrg.*, **170**, 155–172.
- GITTELSON, C. J., HIPTMAIR, R. & PERUGIA, I. (2009) Plane wave discontinuous Galerkin methods: analysis of the *h*-version. *M2AN Math. Model. Numer. Anal.*, **43**, 297–331.
- GRIESMAIER, R. & MONK, P. (2011) Error analysis for a hybridizable discontinuous Galerkin method for the Helmholtz equation. J. Sci. Comput., 49, 291–310.
- HARARI, I. (1997) Reducing spurious dispersion, anisotropy and reflection in finite element analysis of time-harmonic acoustics. *Comput. Methods Appl. Mech. Engrg.*, 140, 39–58.
- HARARI, I. & HUGHES, T. J. (1992) Galerkin/least-squares finite element methods for the reduced wave equation with nonreflecting boundary conditions in unbounded domains. *Comput. Methods Appl. Mech. Engrg.*, 98, 411–454.
- HETMANIUK, U. (2007) Stability estimates for a class of Helmholtz problems. *Commun. Math. Sci.*, **5**, 665–678.
- HIPTMAIR, R., MOIOLA, A. & PERUGIA, I. (2011) Plane wave discontinuous Galerkin methods for the 2D Helmholtz equation: analysis of the *p*-version. *SIAM J. Numer. Anal.*, **49**, 264–284.
- HUTTUNEN, T. & MONK, P. (2007) The use of plane waves to approximate wave propagation in anisotropic media. J. Comput. Math., 25, 350–367.
- IHLENBURG, F. & BABUŠKA, I. (1995) Finite element solution of the Helmholtz equation with high wave number. I. The *h*-version of the FEM. *Comput. Math. Appl.*, **30**, 9–37.
- LAGHROUCHE, O., BETTESS, P. & ASTLEY, R. (2002) Modelling of short wave diffraction problem using approximating systems of plane waves. *Int. J. Numer. Meth. Engng.*, **54**, 1501–1533.

#### REFERENCES

- LUOSTARI, T., HUTTUNEN, T. & MONK, P. (2009) Plane wave methods for approximating the time harmonic wave equation. *Highly oscillatory problems*. London Math. Soc. Lecture Note Ser., vol. 366. Cambridge: Cambridge Univ. Press, pp. 127–153.
- MELENK, J. (1995) On generalized finite element methods. Ph.D. thesis, University of Maryland.
- OBERAI, A. & PINSKY, P. M. (2000) A numerical comparison of finite element methods for the Helmholtz equation. *J. Comput. Acoust.*, **8**, 211–221. Finite elements for wave problems (Trieste, 1999).
- SULEAU, S., DERAEMAEKER, A. & BOUILLARD, P. (2000) Dispersion and pollution of meshless solutions for the Helmholtz equation. *Comput. Methods Appl. Mech. Engrg.*, **190**, 639–657.