

## AN ANALYSIS OF HDG METHODS FOR CONVECTION-DOMINATED DIFFUSION PROBLEMS

GUOSHENG FU<sup>1</sup>, WEIFENG QIU<sup>2</sup> AND WUJUN ZHANG<sup>3</sup>

**Abstract.** We present the first *a priori* error analysis of the *h*-version of the hybridizable discontinuous Galerkin (HDG) methods applied to convection-dominated diffusion problems. We show that, when using polynomials of degree no greater than  $k$ , the  $L^2$ -error of the scalar variable converges with order  $k + 1/2$  on general conforming quasi-uniform simplicial meshes, just as for conventional DG methods. We also show that the method achieves the optimal  $L^2$ -convergence order of  $k + 1$  on special meshes. Moreover, we discuss a new way of implementing the HDG methods for which the spectral condition number of the global matrix is independent of the diffusion coefficient. Numerical experiments are presented which verify our theoretical results.

**Mathematics Subject Classification.** 65N30.

Received October 3, 2013. Revised March 12, 2014.  
Published online January 16, 2015.

### 1. INTRODUCTION

In this paper, we present the first *a priori* error analysis of the *h*-version of the HDG methods for the following convection-dominated diffusion model problem:

$$-\epsilon \Delta u + \boldsymbol{\beta} \cdot \nabla u = f \quad \text{in } \Omega, \quad (1.1a)$$

$$u = g \quad \text{on } \partial\Omega, \quad (1.1b)$$

where  $\Omega \in \mathbb{R}^d$  ( $d = 2, 3$ ) is a polyhedral domain,  $\epsilon \ll |\boldsymbol{\beta}|_{L^\infty(\Omega)}$ ,  $f \in L^2(\Omega)$  and  $g \in H^{1/2}(\partial\Omega)$ . As in [1], we assume that the velocity field  $\boldsymbol{\beta} \in W^{1,\infty}(\Omega)$  has neither closed curves nor stationary points, *i.e.*,

$$\boldsymbol{\beta} \in W^{1,\infty}(\Omega) \text{ has no closed curves, } \boldsymbol{\beta}(\mathbf{x}) \neq \mathbf{0} \quad \forall \mathbf{x} \in \Omega. \quad (1.2)$$

This implies that there exists a smooth function  $\psi$  so that

$$\boldsymbol{\beta} \cdot \nabla \psi(x) \geq b_0 \quad \forall x \in \Omega, \quad (1.3)$$

---

*Keywords and phrases.* HDG, convection-dominated diffusion.

<sup>1</sup> School of Mathematics, University of Minnesota, Minneapolis, MN 55455, USA. [fuxxx165@math.umn.edu](mailto:fuxxx165@math.umn.edu)

<sup>2</sup> Department of Mathematics, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong, P.R. China. [weifeqiu@cityu.edu.hk](mailto:weifeqiu@cityu.edu.hk)

<sup>3</sup> Department of Mathematics, University of Maryland, College Park, MD 20742-4015, USA. [wujun@umd.edu](mailto:wujun@umd.edu)

for some constant  $b_0 > 0$ , see [24] or ([1], Appendix A) for a proof. We also assume that

$$-\nabla \cdot \boldsymbol{\beta}(\mathbf{x}) \geq 0 \quad \forall \mathbf{x} \in \Omega, \quad (1.4)$$

which means that the “effective” reaction is non-negative since

$$\boldsymbol{\beta} \cdot \nabla u = \nabla \cdot (\boldsymbol{\beta}u) - (\nabla \cdot \boldsymbol{\beta})u.$$

Let us remark that assumption (1.2) ensures the well-posedness of the continuous problem in the pure hyperbolic limit ( $\epsilon = 0$ ), see Chapter 3 of [27], for details. It is also well-known [25, 27] that solutions to the problem (1.1) may develop layers, whose approximation is the major difficulty of designing high-order, robust numerical schemes. We refer to [38, 39] for a comprehensive information on different numerical techniques for (1.1).

In the last decade, the discontinuous Galerkin methods [11, 23] have been extensively considered for convection-diffusion equations. For example, see the local discontinuous Galerkin (LDG) methods [8, 12, 13, 22, 30], the method of Baumann and Oden [2], the interior-penalty discontinuous Galerkin (IP-DG) methods [1, 40], the multiscale discontinuous Galerkin method [6, 31], the mixed-hybrid-discontinuous Galerkin (MH-DG) method [26], and the HDG methods [16, 36, 37]. On the other hand, for steady-state problems, the main disadvantage of conventional DG methods, compared to other methods, is that they require a higher number of globally-coupled degrees of freedom for the same mesh. In order to address this issue, the HDG methods were introduced in [19] in the framework of second-order uniformly elliptic problems. The methods are such that the globally-coupled degrees of freedom are only those of the numerical traces on the mesh skeleton. A similar idea was used in [26] to obtain the MH-DG method. Hence, the use of the hybridization technique eliminates the main disadvantage of DG methods to a significant extent.

In [14, 20], it was shown that, for the purely diffusive model problem, the numerical approximation of HDG methods achieves the same order of convergence as that of mixed methods. More precisely, when using polynomials of degree no greater than  $k$ , the  $L^2$ -error for both the scalar and flux approximation converges optimally with order  $k + 1$ , and a postprocessed scalar approximation converges with order  $k + 2$  for  $k \geq 1$ . Recently in [9, 10], similar results have been proven for the convection-diffusion equation when the diffusion coefficient is comparable to the convection coefficient, with variable-degree approximations and nonconforming meshes.

In this work, we focus on the analysis of the convection-dominated case, that is, when  $\epsilon \ll |\boldsymbol{\beta}|_{L^\infty(\Omega)}$ . We show that for the HDG methods using polynomial degree  $k \geq 1$  with a suitably chosen stabilization function, we have, for general meshes, that

$$\|u_h - u\|_{L^2(\Omega)} \leq Ch^k \left( \epsilon h^{-1/2} + \epsilon^{1/2} + h^{1/2} \right) |u|_{H^{k+1}(\Omega)}, \quad (1.5)$$

and, for meshes (almost) aligned with the direction of  $\boldsymbol{\beta}$ , that

$$\|u_h - u\|_{L^2(\Omega)} \leq Ch^k \left( \epsilon h^{-1/2} + \epsilon^{1/2} + h \right) |u|_{H^{k+1}(\Omega)}, \quad (1.6)$$

where  $C$  is a constant independent of  $\epsilon$  and  $h$ . Note that if  $\epsilon \leq \mathcal{O}(h^2)$ , we obtain optimal convergence for  $\|u_h - u\|_{L^2(\Omega)}$  in (1.6), which can be considered as an extension of a similar result for the pure hyperbolic case [15, 17]. We also show that, with a suitably chosen stabilization function, the condition number of the global matrix for the scaled numerical traces is  $\mathcal{O}(h^{-2})$ , independent of  $\epsilon$ .

To prove these estimates, we cannot use the approach used in [9, 10] because the constants in the error estimates in [9, 10] may blow up as  $\epsilon$  approaches 0. For example, the constant  $\Upsilon_K^{\max}$  in Theorem 2.1 of [9], is of order  $\mathcal{O}(\epsilon^{-1})$ . In order to obtain an estimate that is robust with respect to  $\epsilon$ , we need to modify the energy argument used in [9, 10] by using test functions similar to that used in [1, 32]. In [32], a weighted test function was used to obtain the  $L^2$ -stability of the original DG method [35] for the pure hyperbolic equation. In [1], the idea was extended to convection-diffusion-reaction equations using the IP-DG method. We also need to use a new projection to obtain error estimates with less restrictive regularity assumptions.

Next, we would like to compare our results with those obtained for the IP-DG method in [1]. Our convergence result for  $\|u_h - u\|_{L^2(\Omega)}$  on general meshes is the same as that in [1], while the optimal convergence on special meshes is new. Also, our method has less globally-coupled degrees of freedom, and our choice of the stabilization function is determined clearly in the numerical formulation; there is no need to choose it empirically as in the IP-DG method.

Now, let us compare our results with those for the MH-DG method [26]. The MH-DG method uses a combination of upwind techniques used in DG methods for hyperbolic problems with conservative discretizations of mixed methods for elliptic problems. To the best of our knowledge, [26] is the first paper which utilizes hybrid formulations for the mixed and DG methods to make them compatible. We show that our method is quite similar to the MH-DG method. Actually, in Appendix A, we show that, using the same approximation spaces as those of the MH-DG method, the HDG method becomes exactly the same as the MH-DG method by suitably choosing the stabilization function. The new features of our analysis with respect to that of [26] are that we can deal with variable velocity field  $\beta$  and that we have an estimate of  $\|u_h - u\|_{L^2(\Omega)}$ , which is not obtained in [26]. Moreover, we prove that the condition number for the global linear system can be rendered to be independent of  $\epsilon$  and of order  $\mathcal{O}(h^{-2})$ .

A well-known stabilization technique for convection-dominated diffusion problems in the finite element method literature is residual-based stabilization, see the SUPG [5] and residual-free bubbles [3, 4] methods. The main disadvantages of residual-based stabilization are that they are not locally conservative and that the performance of the methods relies heavily on a proper choice of the stabilization parameter, which might be hard to determine or expensive to compute. We refer readers to [29] for a detailed comparison of the hp-version of the DG methods with the SUPG methods in the pure hyperbolic case, and to [26] for a detailed comparison of the MH-DG method with the SUPG methods for the convection-diffusion case.

The rest of the paper is organized as follows. In Section 2, we introduce the HDG method and state and discuss the main theoretical results. In Section 3, we give a characterization of the HDG method, and show that, after scaling, the condition number of the global matrix is independent of  $\epsilon$ . In Section 4, we present the convergence analysis of the HDG method. Finally, in Section 5, we display numerical experiments which verify our theoretical results.

## 2. THE HDG METHOD AND MAIN RESULTS

In this section, we present the HDG method and state and discuss our main theoretical results.

### 2.1. The mesh

Let  $\mathcal{T}_h$  be a conforming, quasi-uniform simplicial triangulation of  $\Omega$ . Given an element (triangle/tetrahedron)  $K \in \mathcal{T}_h$ , which we assume to be an open set,  $\partial K$  denotes the set of its edges in the two dimensional case and of its faces in the three dimensional case. Elements of  $\partial K$  will be generally referred to as faces, regardless of the spatial dimension, and denoted by  $F$ . The set of all (interior) faces of the triangulation will be denoted  $\mathcal{E}_h$  ( $\mathcal{E}_h^o$ ). We distinguish functions defined on the faces of the triangulation (the skeleton) by saying that they are defined on  $\mathcal{E}_h$  from functions defined on the boundaries of the elements (and therefore having the ability to display two different values on interior faces) by saying that they are defined on  $\partial\mathcal{T}_h$ . Hence the spaces  $L^2(\mathcal{E}_h)$  and  $L^2(\partial\mathcal{T}_h)$  have different meanings. For each element  $K \in \mathcal{T}_h$ , we set  $h_K := |K|^{\frac{1}{d}}$ , and for each of face  $F$ ,  $h_F := |F|^{\frac{1}{d-1}}$ , where  $|\cdot|$  denotes the Lebesgue measure in  $d$  or  $d-1$  dimensions. We define  $h = \max_{K \in \mathcal{T}_h} h_K$ .

Moreover, we also consider special meshes that satisfy the following assumption: there exists a constant  $C$  so that

$$\max_{x \in F} (\sup \beta(x) \cdot \mathbf{n}, 0) \leq Ch_K, \quad \forall F \in \partial K \setminus F_K^+, \forall K \in \mathcal{T}_h, \quad (2.1)$$

where  $F_K^+$  is the face of  $K$  such that  $\sup_{x \in F_K^+} \boldsymbol{\beta}(\mathbf{x}) \cdot \mathbf{n} = \max_{F \in \partial K} \sup_{x \in F} \boldsymbol{\beta}(\mathbf{x}) \cdot \mathbf{n}$ . These meshes have been introduced in [15] (see also [17]) for the analysis of the original DG method. In Appendix C, we sketch how to generate a triangulation satisfying assumption (2.1).

## 2.2. The HDG method

In order to define the HDG method, we first rewrite our model problem (1.1) as the following first-order system by introducing  $\mathbf{q} = -\epsilon \nabla u$  as a new unknown:

$$\epsilon^{-1} \mathbf{q} + \nabla u = 0 \text{ in } \Omega, \quad (2.2a)$$

$$\nabla \cdot \mathbf{q} + \boldsymbol{\beta} \cdot \nabla u = f \text{ in } \Omega, \quad (2.2b)$$

$$u = g \text{ on } \partial\Omega. \quad (2.2c)$$

Let us also define the following finite element spaces:

$$\mathbf{V}_h = \{\mathbf{r} \in L^2(\Omega; \mathbb{R}^d) : \mathbf{r}|_K \in P_k(K; \mathbb{R}^d) \quad \forall K \in \mathcal{T}_h\}, \quad (2.3a)$$

$$W_h = \{w \in L^2(\Omega) : w|_K \in P_k(K) \quad \forall K \in \mathcal{T}_h\}, \quad (2.3b)$$

$$M_h = \{\mu \in L^2(\mathcal{E}_h) : \mu|_F \in P_k(F) \quad \forall F \in \mathcal{E}_h\}, \quad (2.3c)$$

$$M_h(g) = \{\mu \in M_h : \langle \mu, \xi \rangle_{\partial\Omega} = \langle g, \xi \rangle_{\partial\Omega} \quad \forall \xi \in M_h\}, \quad (2.3d)$$

where  $P_k(D)$  is the space of polynomials of total degree not larger than  $k \geq 0$  defined on  $D$ , and

$$\langle \xi, \eta \rangle_{\partial\Omega} = \sum_{F \in \partial\Omega} \int_F \xi \eta \, ds.$$

The HDG method seeks an approximation  $(\mathbf{q}_h, u_h, \hat{u}_h) \in \mathbf{V}_h \times W_h \times M_h$  by requiring that

$$(\epsilon^{-1} \mathbf{q}_h, \mathbf{r})_{\mathcal{T}_h} - (u_h, \nabla \cdot \mathbf{r})_{\mathcal{T}_h} + \langle \hat{u}_h, \mathbf{r} \cdot \mathbf{n} \rangle_{\partial\mathcal{T}_h} = 0, \quad (2.4a)$$

$$-(\mathbf{q}_h + \boldsymbol{\beta} u_h, \nabla w)_{\mathcal{T}_h} - (\nabla \cdot \boldsymbol{\beta} u_h, w)_{\mathcal{T}_h} + \left\langle \left( \hat{\mathbf{q}}_h + \widehat{\boldsymbol{\beta}} u_h \right) \cdot \mathbf{n}, w \right\rangle_{\partial\mathcal{T}_h} = (f, w)_{\mathcal{T}_h}, \quad (2.4b)$$

$$\langle \hat{u}_h, \mu \rangle_{\partial\Omega} = \langle g, \mu \rangle_{\partial\Omega}, \quad (2.4c)$$

$$\left\langle \left( \hat{\mathbf{q}}_h + \widehat{\boldsymbol{\beta}} u_h \right) \cdot \mathbf{n}, \mu \right\rangle_{\partial\mathcal{T}_h \setminus \partial\Omega} = 0, \quad (2.4d)$$

for all  $(\mathbf{r}, w, \mu) \in \mathbf{V}_h \times W_h \times M_h$ , where the numerical trace  $(\hat{\mathbf{q}}_h + \widehat{\boldsymbol{\beta}} u_h) \cdot \mathbf{n}$  is given by

$$\left( \hat{\mathbf{q}}_h + \widehat{\boldsymbol{\beta}} u_h \right) \cdot \mathbf{n} = \mathbf{q}_h \cdot \mathbf{n} + \boldsymbol{\beta} \cdot \mathbf{n} \hat{u}_h + \tau(u_h - \hat{u}_h) \text{ on } \partial\mathcal{T}_h, \quad (2.4e)$$

and the stabilization function  $\tau$  is piecewise, nonnegative constant defined on  $\partial\mathcal{T}_h$ . Here we write  $(\eta, \zeta)_{\mathcal{T}_h} := \sum_{K \in \mathcal{T}_h} \int_K \eta \zeta \, dx$ , and  $\langle \eta, \zeta \rangle_{\partial\mathcal{T}_h} := \sum_{K \in \mathcal{T}_h} \int_{\partial K} \eta \zeta \, ds$ . In Section 3, we show that the linear system (2.4) can be efficiently implemented so that the only global unknowns are related to the numerical trace  $\hat{u}_h$ .

The HDG method (2.4) has a unique solution provided that the stabilization function  $\tau$  in (2.4e) satisfies the following assumption:

$$\inf_{\mathbf{x} \in F} \left( \tau - \frac{1}{2} \boldsymbol{\beta}(\mathbf{x}) \cdot \mathbf{n} \right) \geq 0, \quad \forall F \in \partial K, \forall K \in \mathcal{T}_h, \quad (2.5)$$

where in each element, the strict inequality holds at least on one face; see Theorem 3.1 of [36], for a proof.

### 2.3. Assumptions on the stabilization function

Next, we present our assumptions on the stabilization function  $\tau$  verifying the inequality (2.5). We then construct two examples satisfying them.

To do that, we need to introduce some notation. Let  $F_K^*$  be the face of  $K$  on which  $\tau$  attains its maximum, and  $F_K^s$  be the face of  $K$  on which  $\inf_{\mathbf{x} \in F} (\tau - \frac{1}{2}\boldsymbol{\beta}(\mathbf{x}) \cdot \mathbf{n})$  attains its maximum, that is,

$$\tau(F_K^*) := \max_{F \in \partial K} \tau(F) \quad \forall K \in \mathcal{T}_h, \quad (2.6a)$$

$$\inf_{\mathbf{x} \in F_K^s} \left( \tau - \frac{1}{2}\boldsymbol{\beta}(\mathbf{x}) \cdot \mathbf{n} \right) := \max_{F \in \partial K} \inf_{\mathbf{x} \in F} \left( \tau - \frac{1}{2}\boldsymbol{\beta}(\mathbf{x}) \cdot \mathbf{n} \right) \quad \forall K \in \mathcal{T}_h, \quad (2.6b)$$

and set

$$\begin{aligned} \tau_K^w &:= \max_{F \in \partial K \setminus F_K^*} \tau(F), & \tau^w &:= \max_{K \in \mathcal{T}_h} \tau_K^w, \\ \tau_K^v &:= \inf_{\mathbf{x} \in F_K^s} \left( \tau - \frac{1}{2}\boldsymbol{\beta}(\mathbf{x}) \cdot \mathbf{n} \right), & \tau^v &:= \min_{K \in \mathcal{T}_h} \tau_K^v. \end{aligned}$$

We assume that there exists universal positive constants  $C_0, C_1, C_2$  so that

$$\tau_K^w \leq C_0 \quad \forall K \in \mathcal{T}_h, \quad (2.8a)$$

$$\tau_K^v \geq C_1 \min \left( \frac{\epsilon}{h_K}, 1 \right) \quad \forall K \in \mathcal{T}_h, \quad (2.8b)$$

$$\inf_{\mathbf{x} \in F} \left( \tau - \frac{1}{2}\boldsymbol{\beta}(\mathbf{x}) \cdot \mathbf{n} \right) \geq C_2 \max_{\mathbf{x} \in F} |\boldsymbol{\beta}(\mathbf{x}) \cdot \mathbf{n}| \quad \forall F \in \partial K, \forall K \in \mathcal{T}_h. \quad (2.8c)$$

In order to get an improved estimate, we need to replace (2.8a) by the following, more restrictive assumption on  $\tau_K^w$ : assume there exists a positive constant  $C$  so that

$$\tau_K^w \leq Ch_K \quad \forall K \in \mathcal{T}_h. \quad (2.9)$$

We remark this assumption might not be compatible with (2.8c) on general meshes, but it can hold for the meshes that satisfy assumption (2.1).

Now, let us show that it is quite easy to construct  $\tau$  satisfying assumptions (2.8) by displaying two of them. The first example of the stabilization function is

$$\tau_1(F) = \max(\sup_{\mathbf{x} \in F} \boldsymbol{\beta}(\mathbf{x}) \cdot \mathbf{n}, 0), \quad \forall F \in \partial K, \forall K \in \mathcal{T}_h. \quad (2.10)$$

Assumptions (2.8a) and (2.8c) are always satisfied, and assumption (2.8b) holds provided

$$\max_{F \in \partial K} \inf_{\mathbf{x} \in F} (-\boldsymbol{\beta}(\mathbf{x}) \cdot \mathbf{n}) \geq C \quad \forall K \in \mathcal{T}_h,$$

for some positive constant  $C$ ; this is true, for example, for piecewise-constant  $\boldsymbol{\beta}$ . Moreover, assumption (2.9) is also satisfied if the mesh satisfies (2.1).

The second example is

$$\tau_2(F) = \max(\sup_{\mathbf{x} \in F} \boldsymbol{\beta}(\mathbf{x}) \cdot \mathbf{n}, 0) + \min \left( \rho_0 \frac{\epsilon}{h_K}, 1 \right), \quad \forall F \in \partial K, \forall K \in \mathcal{T}_h, \quad (2.11)$$

where  $\rho_0 > 0$  is a constant typically chosen to be less than or equal to 1. Assumptions (2.8) are always satisfied in this case. Moreover, assumption (2.9) is satisfied provided the mesh satisfies (2.1) and we take  $\epsilon \leq \mathcal{O}(h^2)$ .

Let us conclude the discussion on  $\tau$  by remarking that if we replace  $\max(\sup_{\mathbf{x} \in F} \boldsymbol{\beta}(\mathbf{x}) \cdot \mathbf{n}, 0)$  with  $\sup_{\mathbf{x} \in F} |\boldsymbol{\beta}(\mathbf{x}) \cdot \mathbf{n}|$  in the definition of  $\tau$  in (2.10) and (2.11), assumptions (2.8) will be satisfied, while (2.9) is no longer true for the special meshes.

## 2.4. The main theoretical results

From now on, we use  $C$  to denote a generic constant, which may be dependent on the polynomial degree  $k$ , and/or the velocity field  $\boldsymbol{\beta}$ . The value  $C$  at different occurrences may differ.

We proceed to state our main theoretical results. We will show convergence estimates in the following norm

$$\|(\mathbf{r}, w, \mu)\|_e := \left( \|\epsilon^{-1/2} \mathbf{r}\|_{\mathcal{T}_h}^2 + \|w\|_{\mathcal{T}_h}^2 + \left\| \left| \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right|^{1/2} (w - \mu) \right\|_{\partial \mathcal{T}_h}^2 \right)^{1/2},$$

where  $\|\cdot\|_D$  is the standard  $L^2$ -norm in the domain  $D$ .

**Theorem 2.1.** *Let  $(\mathbf{q}, u)$  be the solution to the boundary-value problem (2.2), and let  $(\mathbf{q}_h, u_h, \hat{u}_h)$  be the solution to the HDG method (2.4) where the stabilization function  $\tau$  satisfies assumptions (2.8). Then, there exists  $h_0$ , independent of  $\epsilon$ , such that when  $h < h_0$ , we have*

$$\|(\mathbf{q} - \mathbf{q}_h, u - u_h, u - \hat{u}_h)\|_e \leq C \epsilon^{1/2} h^{s_v+1/2} \left( \epsilon^{1/2} + h^{1/2} \right) |u|_{H^{s_v+2}(\mathcal{T}_h; \mathbb{R}^d)} + Ch^{s_w+1/2} |u|_{H^{s_w+1}(\mathcal{T}_h)},$$

for all  $s_v \in [0, k]$  and  $s_w \in [0, k]$ .

**Theorem 2.2.** *Let  $(\mathbf{q}, u)$  be the solution to the boundary-value problem (2.2), and let  $(\mathbf{q}_h, u_h, \hat{u}_h)$  be the solution to the HDG method (2.4) where the stabilization function  $\tau$  satisfies assumptions (2.8) and (2.9). Then, there exists  $h_0$ , independent of  $\epsilon$ , such that when  $h < h_0$ , we have*

$$\|u - u_h\|_{\mathcal{T}_h} \leq C \epsilon^{1/2} h^{s_v+1/2} \left( \epsilon^{1/2} + h^{1/2} \right) |u|_{H^{s_v+2}(\mathcal{T}_h; \mathbb{R}^d)} + Ch^{s_w+1} |u|_{H^{s_w+1}(\mathcal{T}_h)},$$

for all  $s_v \in [0, k]$  and  $s_w \in [0, k]$ .

**Remark 2.3.** If  $\tau$  satisfies assumptions (2.8),  $k \geq 1$ ,  $u \in H^{k+1}(\Omega)$  and  $\epsilon \leq \mathcal{O}(h)$ , we get

$$\|u - u_h\|_{\mathcal{T}_h} \leq Ch^{k+1/2} |u|_{H^{k+1}(\mathcal{T}_h)}$$

by choosing  $s_v = k - 1$ ,  $s_w = k$  in Theorem 2.1. If  $\epsilon = 0$ , our method collapses to the original DG method [35]. Since the best  $L^2$ -error of the DG method for pure convection problems on general meshes is  $\|u - u_h\|_{\mathcal{T}_h} \leq Ch^{k+1/2} |u|_{H^{k+1}}$ , see [34]; it is reasonable to expect  $\|u - u_h\|_{\mathcal{T}_h}$  to be of order  $h^{k+1/2}$  when  $\epsilon \ll 1$ .

**Remark 2.4.** If  $\tau$  satisfies assumptions (2.8) and (2.9),  $k \geq 1$ ,  $u \in H^{k+1}(\Omega)$  and  $\epsilon \leq \mathcal{O}(h^2)$ , we have

$$\|u - u_h\|_{\mathcal{T}_h} \leq Ch^{k+1} |u|_{H^{k+1}(\mathcal{T}_h)},$$

by choosing  $s_v = k - 1$ ,  $s_w = k$  in Theorem 2.2. Note that we can construct  $\tau$  satisfying (2.8) and (2.9) provided that the mesh satisfies (2.1). Hence, our result can be considered as a generalization of the results in [15, 17] in which the authors obtained optimal  $L^2$ -convergence of the original DG method for convection-reaction equations on special meshes.

**Remark 2.5.** It is shown in Appendix A that we can recover the MH-DG method [26] from our formulation by suitably choosing the stabilization function  $\tau$  and the approximation spaces  $\mathbf{V}_h, W_h, M_h$ . Hence, our results can be directly applied to the MH-DG method. In particular, we gain the  $L^2$ -control of  $u_h$  and obtained optimal order of convergence for  $\|u - u_h\|_{\mathcal{T}_h}$  for special meshes.

## 3. A CHARACTERIZATION OF THE HDG METHOD

Here, we show how to eliminate, in an elementwise manner, the unknowns  $\mathbf{q}_h$  and  $u_h$  from the equations (2.4) and rewrite the original system solely in terms of the unknown  $\hat{u}_h$ , see also [16, 36]. In this way, we do not have to deal with the large linear system generated by (2.4), but with the inversion of a sparser matrix of remarkably smaller size.

### 3.1. The local problems

We begin by showing how to express the unknowns  $\mathbf{q}_h$  and  $u_h$  in terms of the unknown  $\widehat{u}_h$ .

Given  $\lambda \in L^2(\mathcal{E}_h)$  and  $f \in L^2(\mathcal{T}_h)$ , consider the solution to the set of local problems in each  $K \in \mathcal{T}_h$ : find

$$(\mathbf{q}_h, u_h) \in \mathbf{V}(K) \times W(K),$$

where  $\mathbf{V}(K) := P_k(K; \mathbb{R}^d)$  and  $W(K) := P_k(K)$ , such that

$$(\epsilon^{-1} \mathbf{q}_h, \mathbf{r})_K - (u_h, \nabla \cdot \mathbf{r})_K = - \langle \lambda, \mathbf{r} \cdot \mathbf{n} \rangle_{\partial K}, \quad (3.1a)$$

$$(\nabla \cdot \mathbf{q}_h, w)_K - (u_h, \nabla \cdot (\boldsymbol{\beta} w))_K + \langle \tau u_h, w \rangle_{\partial K} = \langle (\tau - \boldsymbol{\beta} \cdot \mathbf{n}) \lambda, w \rangle_{\partial K} + (f, w)_K, \quad (3.1b)$$

for all  $(\mathbf{r}, w) \in \mathbf{V}(K) \times W(K)$ .

We denote by  $(\mathbf{q}_h^f, u_h^f)$  the solution of the above local problem when we take  $\lambda = 0$ . Similarly, we denote  $(\mathbf{q}_h^\lambda, u_h^\lambda)$  the solution when  $f = 0$ . We can thus write that

$$(\mathbf{q}_h, u_h) = (\mathbf{q}_h^\lambda, u_h^\lambda) + (\mathbf{q}_h^f, u_h^f).$$

### 3.2. The global problem

If we now take  $\lambda := \widehat{u}_h$ , we see that  $(\mathbf{q}_h, u_h)$  is expressed in terms of  $\widehat{u}_h$  (and  $f$ ). We can thus eliminate those two unknowns from the equations and solve for  $\widehat{u}_h$  only. The global problem that determines  $\widehat{u}_h$  is not difficult to find.

We have that  $\widehat{u}_h \in M_h(g)$  must satisfy

$$a_h(\widehat{u}_h, \mu) = b_h(\mu) \quad \forall \mu \in M_h(0),$$

where

$$\begin{aligned} a_h(\lambda, \mu) &:= - \langle \mathbf{q}_h^\lambda \cdot \mathbf{n}, \mu \rangle_{\partial \mathcal{T}_h} - \langle \tau (u_h^\lambda - \lambda), \mu \rangle_{\partial \mathcal{T}_h}, \\ b_h(\mu) &:= \langle \mathbf{q}_h^f \cdot \mathbf{n}, \mu \rangle_{\partial \mathcal{T}_h} + \langle \tau u_h^f, \mu \rangle_{\partial \mathcal{T}_h}. \end{aligned}$$

Indeed, note that the definition of  $M_h(g)$  incorporates the boundary condition (2.4c), and that the last equation is nothing but a rewriting of the transmission condition (2.4d) by observing that

$$\langle \boldsymbol{\beta} \cdot \mathbf{n} \lambda, \mu \rangle_{\partial \mathcal{T}_h} = 0, \quad \forall \lambda \in M_h(g), \forall \mu \in M_h(0).$$

### 3.3. A characterization of the approximate solution

The above results suggest the following characterization of the approximate solution of the HDG method. We leave the proof to the interested readers as an exercise, see also [16, 36].

**Theorem 3.1.** *The approximate solution of the HDG method satisfies*

$$(\mathbf{q}_h, u_h) = (\mathbf{q}_h^{\widehat{u}_h}, u_h^{\widehat{u}_h}) + (\mathbf{q}_h^f, u_h^f).$$

Moreover,  $\widehat{u}_h \in M_h(g)$  is the solution of

$$a_h(\widehat{u}_h, \mu) = b_h(\mu) \quad \forall \mu \in M_h(0). \quad (3.2)$$

Also, we have that

$$\begin{aligned} a_h(\lambda, \mu) &= (\epsilon^{-1} \mathbf{q}_h^\lambda, \mathbf{q}_h^\mu)_{\mathcal{T}_h} - (u_h^\lambda, \nabla \cdot (\boldsymbol{\beta} u_h^\mu))_{\mathcal{T}_h} + \langle \boldsymbol{\beta} \cdot \mathbf{n} \lambda, u_h^\mu \rangle_{\partial \mathcal{T}_h} + \langle \tau (u_h^\lambda - \lambda), u_h^\mu - \mu \rangle_{\partial \mathcal{T}_h}, \\ b_h(\mu) &= (f, u_h^\mu)_{\mathcal{T}_h} + \langle \boldsymbol{\beta} \cdot \mathbf{n} u_h^f, \mu \rangle_{\partial \mathcal{T}_h} + (u_h^f, \boldsymbol{\beta} \cdot \nabla u_h^\mu)_{\mathcal{T}_h} - (\nabla u_h^f, \boldsymbol{\beta} u_h^\mu)_{\mathcal{T}_h}. \end{aligned}$$

### 3.4. The conditioning of the HDG method

We note that both examples of stabilization function  $\tau$  in (2.10) and (2.11) on a face  $F$  can be very small if  $\boldsymbol{\beta} \cdot \mathbf{n}|_F$  and  $\epsilon$  are very small. In this case, the condition number of the global matrix generated by  $a_h$  in (3.2) might blow up as  $\epsilon$  goes to zero.

In order to make the condition number independent of  $\epsilon$ , we need a new assumption on  $\tau$ , namely,

$$\inf_{\mathbf{x} \in F} \left( \tau - \frac{1}{2} \boldsymbol{\beta}(\mathbf{x}) \cdot \mathbf{n} \right) \geq C_2 \min \left( \frac{\epsilon}{h_F}, 1 \right) \quad \forall F \in \partial K, \forall K \in \mathcal{T}_h. \quad (3.3)$$

If we introduce

$$\begin{aligned} \tilde{\lambda} &= \Lambda_\epsilon \lambda, \quad \tilde{\mu} = \Lambda_\epsilon \mu, \quad \forall \lambda \in M_h(g), \mu \in M_h(0), \\ \text{where } \Lambda_\epsilon|_F &= \left( \sup_{x \in F} |\boldsymbol{\beta} \cdot \mathbf{n}(x)| + \min \left( \frac{\epsilon}{h_F}, 1 \right) \right)^{1/2}, \quad \forall F \in \mathcal{E}_h, \end{aligned} \quad (3.4)$$

the preferred form for implementation for the HDG method is to find  $\tilde{\lambda} \in M_h(\Lambda_\epsilon g)$  satisfying

$$\tilde{a}_h(\tilde{\lambda}, \tilde{\mu}) = b_h(\Lambda_\epsilon^{-1} \tilde{\mu})$$

for all  $\tilde{\mu} \in M_h(0)$ . Here,

$$\tilde{a}_h(\tilde{\lambda}, \tilde{\mu}) = a_h(\Lambda_\epsilon^{-1} \tilde{\lambda}, \Lambda_\epsilon^{-1} \tilde{\mu}). \quad (3.5)$$

We have the following theorem concerning the condition number of the scaled global matrix in (3.5).

**Theorem 3.2.** *Let the stabilization function  $\tau$  satisfy assumptions (2.8) and (3.3), and let  $\epsilon \leq \mathcal{O}(h)$ . Let  $\kappa$  be the spectral condition number of the global matrix generated by  $\tilde{a}_h$  in (3.5). Then there is  $h_0 > 0$ , which is independent of  $\epsilon$  and  $h$ , such that when  $h < h_0$ ,*

$$\kappa \leq Ch^{-2}.$$

We present a detailed proof of Theorem 3.2 in Appendix B.

**Remark 3.3.** Obviously, assumption (3.3) is satisfied by the second stabilization function (2.11) but not by the first one (2.10) on meshes that aligned with the direction of  $\boldsymbol{\beta}$ . Theorem 3.2 shows that the condition number of the global matrix of the HDG method for convection-dominated diffusion problems is the same as that of HDG methods for elliptic problems in [18].

## 4. CONVERGENCE ANALYSIS

In this section, we prove Theorems 2.1 and 2.2. We begin by introducing the following bilinear form:

$$\begin{aligned} B((\mathbf{q}, u, \lambda), (\mathbf{r}, w, \mu)) &= (\epsilon^{-1} \mathbf{q}, \mathbf{r})_{\mathcal{T}_h} - (u, \nabla \cdot \mathbf{r})_{\mathcal{T}_h} + \langle \lambda, \mathbf{r} \cdot \mathbf{n} \rangle_{\partial \mathcal{T}_h} \\ &\quad - (\mathbf{q} + \boldsymbol{\beta} u, \nabla w)_{\mathcal{T}_h} + \langle (\mathbf{q} + \boldsymbol{\beta} \lambda) \cdot \mathbf{n} + \tau(u - \lambda), w \rangle_{\partial \mathcal{T}_h} \\ &\quad - ((\nabla \cdot \boldsymbol{\beta})u, w)_{\mathcal{T}_h} - \langle (\mathbf{q} + \boldsymbol{\beta} \lambda) \cdot \mathbf{n} + \tau(u - \lambda), \mu \rangle_{\partial \mathcal{T}_h}, \end{aligned} \quad (4.1)$$

for all  $(\mathbf{q}, u, \lambda)$  and  $(\mathbf{r}, w, \mu) \in H^1(\mathcal{T}_h; \mathbb{R}^d) \times H^1(\mathcal{T}_h) \times L^2(\mathcal{E}_h)$ . It's easy to see that the HDG method (2.4) can be recasted in the following compact form: Find  $(\mathbf{q}_h, u_h, \hat{u}_h) \in \mathbf{V}_h \times W_h \times M_h(g)$  so that

$$B((\mathbf{q}_h, u_h, \hat{u}_h), (\mathbf{r}, w, \mu)) = (f, w)_{\mathcal{T}_h}, \quad (4.2)$$

for all  $(\mathbf{r}, w, \mu) \in \mathbf{V}_h \times W_h \times M_h(0)$ .

#### 4.1. Stability property for the HDG method

It is well known that we have the following result regarding the stability of the convection-dominated diffusion problem,

$$\epsilon \|\nabla u\|_{L^2(\Omega)}^2 + \|u\|_{L^2(\Omega)}^2 \leq C \|f\|_{L^2(\Omega)}^2, \quad (4.3)$$

provided that  $\boldsymbol{\beta}$  satisfies assumption (1.3) and  $g = 0$  on  $\partial\Omega$ , see [1]. On the other hand, by taking  $(\mathbf{r}, w, \mu) = (\mathbf{q}_h, u_h, \hat{u}_h)$  in (4.2), the standard energy argument only gives the following estimate:

$$(\epsilon^{-1} \mathbf{q}_h, \mathbf{q}_h)_{\mathcal{T}_h} + \left\langle \left( \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right) (u_h - \hat{u}_h), u_h - \hat{u}_h \right\rangle - \frac{1}{2} ((\nabla \cdot \boldsymbol{\beta}) u_h, u_h)_{\mathcal{T}_h} = (f, u_h)_{\mathcal{T}_h}.$$

Hence, we do not have control of the  $L^2$ -norm of  $u_h$  by the standard energy argument when the velocity field  $\boldsymbol{\beta}$  is divergence-free. The main idea of our stability analysis is to achieve the control of the  $L^2$ -norm of  $u_h$  by mimicking the proof of the stability property (4.3) at the discrete level. We shall proceed in the following three steps.

**Step 1.** In view of assumption (1.2), we define a function

$$\varphi := e^{-\psi} + \chi, \quad (4.4)$$

where  $\chi$  is a positive constant to be determined later. Mimicking the proof of stability results carried out for the continuous problem, we obtain the following lemma.

**Lemma 4.1.** *Let  $\varphi$  be given in (4.4) where  $\chi \geq 1 + 2b_0^{-1} \|e^{-\psi}\|_{L^\infty(\Omega)} \cdot \|\nabla \psi\|_{L^\infty(\Omega)}^2$ . Also, let  $\tau$  satisfy assumption (2.5). Then for all  $(\mathbf{q}_h, u_h, \lambda_h) \in \mathbf{V}_h \times W_h \times M_h(0)$ , the following inequality holds*

$$B((\mathbf{q}_h, u_h, \lambda_h), (\mathbf{q}_\varphi, u_\varphi, \lambda_\varphi)) \geq C \|(\mathbf{q}_h, u_h, \lambda_h)\|_e^2,$$

where  $\mathbf{q}_\varphi = \varphi \mathbf{q}_h$ ,  $u_\varphi = \varphi u_h$  and  $\lambda_\varphi = \varphi \lambda_h$ .

*Proof.* With  $(\mathbf{q}_\varphi, u_\varphi, \lambda_\varphi)$  given above, we have that

$$B((\mathbf{q}_h, u_h, \lambda_h), (\mathbf{q}_\varphi, u_\varphi, \lambda_\varphi)) = (\epsilon^{-1} \mathbf{q}_h, \varphi \mathbf{q}_h)_{\mathcal{T}_h} + T_1 + T_2 + T_3,$$

where

$$\begin{aligned} T_1 &= -(u_h, \nabla \cdot (\varphi \mathbf{q}_h))_{\mathcal{T}_h} + \langle \lambda_h, \varphi \mathbf{q}_h \cdot \mathbf{n} \rangle_{\partial \mathcal{T}_h} - (\mathbf{q}_h, \nabla(\varphi u_h))_{\mathcal{T}_h} + \langle \mathbf{q}_h \cdot \mathbf{n}, \varphi(u_h - \lambda_h) \rangle_{\partial \mathcal{T}_h} \\ T_2 &= -(\boldsymbol{\beta} u_h, \nabla(\varphi u_h))_{\mathcal{T}_h} - ((\nabla \cdot \boldsymbol{\beta}) u_h, \varphi u_h)_{\mathcal{T}_h} + \langle \boldsymbol{\beta} \cdot \mathbf{n} \lambda_h, \varphi u_h \rangle_{\partial \mathcal{T}_h} \\ T_3 &= \langle \tau(u_h - \lambda_h), \varphi(u_h - \lambda_h) \rangle_{\partial \mathcal{T}_h}. \end{aligned}$$

By integration by parts, we obtain

$$\begin{aligned} T_1 &= -(u_h, \nabla \cdot (\varphi \mathbf{q}_h))_{\mathcal{T}_h} + \langle \lambda_h, \varphi \mathbf{q}_h \cdot \mathbf{n} \rangle_{\partial \mathcal{T}_h} - (\mathbf{q}_h, \nabla(\varphi u_h))_{\mathcal{T}_h} + \langle \mathbf{q}_h \cdot \mathbf{n}, \varphi(u_h - \lambda_h) \rangle_{\partial \mathcal{T}_h} \\ &= -(u_h, \nabla \varphi \cdot \mathbf{q}_h)_{\mathcal{T}_h} - (\varphi u_h, \nabla \mathbf{q}_h)_{\mathcal{T}_h} - (\mathbf{q}_h, \nabla(\varphi u_h))_{\mathcal{T}_h} + \langle \mathbf{q}_h \cdot \mathbf{n}, \varphi u_h \rangle_{\partial \mathcal{T}_h} \\ &= -(u_h, \nabla \varphi \cdot \mathbf{q}_h)_{\mathcal{T}_h} \\ &= (u_h, e^{-\psi} \nabla \psi \cdot \mathbf{q}_h)_{\mathcal{T}_h} \\ T_2 &= -(\boldsymbol{\beta} u_h, \nabla(\varphi u_h))_{\mathcal{T}_h} - ((\nabla \cdot \boldsymbol{\beta}) u_h, \varphi u_h)_{\mathcal{T}_h} + \langle \boldsymbol{\beta} \cdot \mathbf{n} \lambda_h, \varphi u_h \rangle_{\partial \mathcal{T}_h} \\ &= -(\boldsymbol{\beta} \cdot \nabla \varphi, u_h^2)_{\mathcal{T}_h} - \left( \boldsymbol{\beta} \varphi, \nabla \frac{u_h^2}{2} \right)_{\mathcal{T}_h} - ((\nabla \cdot \boldsymbol{\beta}) \varphi, u_h^2)_{\mathcal{T}_h} + \langle \boldsymbol{\beta} \cdot \mathbf{n} \lambda_h, \varphi u_h \rangle_{\partial \mathcal{T}_h} \\ &= -\frac{1}{2} (\boldsymbol{\beta} \cdot \nabla \varphi, u_h^2)_{\mathcal{T}_h} - \frac{1}{2} \langle \boldsymbol{\beta} \cdot \mathbf{n} u_h, \varphi u_h \rangle_{\partial \mathcal{T}_h} - \frac{1}{2} ((\nabla \cdot \boldsymbol{\beta}) \varphi, u_h^2)_{\mathcal{T}_h} + \langle \boldsymbol{\beta} \cdot \mathbf{n} \lambda_h, \varphi u_h \rangle_{\partial \mathcal{T}_h} \\ &= \frac{1}{2} (\boldsymbol{\beta} \cdot \nabla \psi, e^{-\psi} u_h^2)_{\mathcal{T}_h} - \frac{1}{2} ((\nabla \cdot \boldsymbol{\beta}) \varphi, u_h^2)_{\mathcal{T}_h} - \frac{1}{2} \langle \boldsymbol{\beta} \cdot \mathbf{n} (u_h - \lambda_h), \varphi(u_h - \lambda_h) \rangle_{\partial \mathcal{T}_h}, \end{aligned}$$

where in the last step, we used  $\langle \boldsymbol{\beta} \cdot \mathbf{n} \lambda_h, \varphi \lambda_h \rangle_{\partial \mathcal{T}_h} = 0$  due to the fact that  $\lambda_h$  is single valued on the interior faces and  $\lambda_h = 0$  on  $\partial \Omega$ .

Combining  $T_1$ ,  $T_2$  and  $T_3$ , we have that

$$\begin{aligned} B((\mathbf{q}_h, u_h, \lambda_h), (\mathbf{q}_\varphi, u_\varphi, \lambda_\varphi)) &= (\epsilon^{-1} \mathbf{q}_h, \varphi \mathbf{q}_h)_{\mathcal{T}_h} + (u_h, e^{-\psi} \nabla \psi \cdot \mathbf{q}_h)_{\mathcal{T}_h} \\ &\quad + \frac{1}{2} ([\boldsymbol{\beta} \cdot \nabla \psi] u_h, e^{-\psi} u_h)_{\mathcal{T}_h} - \frac{1}{2} ((\nabla \cdot \boldsymbol{\beta}) u_h, \varphi u_h)_{\mathcal{T}_h} \\ &\quad + \left\langle \left( \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right) \varphi (u_h - \lambda_h), u_h - \lambda_h \right\rangle_{\partial \mathcal{T}_h}. \end{aligned}$$

Invoking assumptions (1.3) and (1.4), and  $\varphi \geq \chi$ , we obtain

$$\begin{aligned} B((\mathbf{q}_h, u_h, \lambda_h), (\mathbf{q}_\varphi, u_\varphi, \lambda_\varphi)) &\geq (\epsilon^{-1} \mathbf{q}_h, \varphi \mathbf{q}_h)_{\mathcal{T}_h} + (u_h, e^{-\psi} \nabla \psi \cdot \mathbf{q}_h)_{\mathcal{T}_h} + \frac{1}{2} b_0 (u_h, e^{-\psi} u_h)_{\mathcal{T}_h} \\ &\quad + \left\langle \left( \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right) \varphi (u_h - \lambda_h), u_h - \lambda_h \right\rangle_{\partial \mathcal{T}_h} \\ &\geq \chi (\epsilon^{-1} \mathbf{q}_h, \mathbf{q}_h)_{\mathcal{T}_h} + (u_h, e^{-\psi} \nabla \psi \cdot \mathbf{q}_h)_{\mathcal{T}_h} + \frac{1}{2} b_0 (u_h, e^{-\psi} u_h)_{\mathcal{T}_h} \\ &\quad + \chi \left\langle \left( \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right) (u_h - \lambda_h), u_h - \lambda_h \right\rangle_{\partial \mathcal{T}_h}. \end{aligned}$$

Using the Cauchy–Schwartz inequality, we have

$$(u_h, e^{-\psi} \nabla \psi \cdot \mathbf{q}_h)_{\mathcal{T}_h} \leq \frac{1}{2} \left[ \delta^{-1} \|\nabla \psi\|_{L^\infty(\Omega)}^2 (e^{-\psi} \mathbf{q}_h, \mathbf{q}_h)_{\mathcal{T}_h} + \delta (e^{-\psi} u_h, u_h)_{\mathcal{T}_h}^2 \right]$$

for any  $\delta > 0$ . Taking  $\chi \geq 1 + 2b_0^{-1} \|e^{-\psi}\|_{L^\infty(\Omega)} \cdot \|\nabla \psi\|_{L^\infty(\Omega)}^2$  and  $\delta = b_0/2$ , we get

$$\begin{aligned} B((\mathbf{q}_h, u_h, \lambda_h), (\mathbf{q}_\varphi, u_\varphi, \lambda_\varphi)) &\geq \epsilon^{-1} \frac{\chi}{2} (\mathbf{q}_h, \mathbf{q}_h)_{\mathcal{T}_h} + \frac{b_0}{4} (e^{-\psi} u_h, u_h)_{\mathcal{T}_h} \\ &\quad + \chi \left\| \left( \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right)^{1/2} (u_h - \lambda_h) \right\|_{\partial \mathcal{T}_h}^2. \end{aligned}$$

To complete the proof, we simply absorb  $\chi$ ,  $e^{-\psi}$  and  $b_0$  into the generic constant  $C$ .  $\square$

**Step 2.** We note that the test function  $(\mathbf{q}_\varphi, u_\varphi, \lambda_\varphi) = (\varphi \mathbf{q}_h, \varphi u_h, \varphi \lambda_h)$  in Lemma 4.1 is not in the discrete space  $\mathbf{V}_h \times W_h \times M_h(0)$ . To establish a discrete stability property, we shall consider taking the discrete test functions as a projection of  $(\mathbf{q}_\varphi, u_\varphi, \lambda_\varphi)$  onto the spaces  $\mathbf{V}_h \times W_h \times M_h$ , denoted by  $\mathbf{\Pi}_h \mathbf{q}_\varphi, \Pi_h u_\varphi, P_M \lambda_\varphi$ . Here  $P_M$  is the  $L^2$ -projection onto  $M_h$ . And  $\mathbf{\Pi}_h$  and  $\Pi_h$  are the projections from  $H^1(\mathcal{T}_h; \mathbb{R}^d)$  and  $H^1(\mathcal{T}_h)$  onto  $\mathbf{V}_h$  and  $W_h$  respectively satisfying

$$(\mathbf{\Pi}_h \mathbf{q}, \mathbf{v})_K = (\mathbf{q}, \mathbf{v})_K \quad \forall \mathbf{v} \in \mathcal{P}_{k-1}(K), \quad (4.5a)$$

$$\langle \mathbf{\Pi}_h \mathbf{q} \cdot \mathbf{n}, \mu \rangle_F = \langle \mathbf{q} \cdot \mathbf{n}, \mu \rangle_F \quad \forall \mu \in \mathcal{P}_k(F), \quad \forall F \in \partial K \setminus F_K^s, \quad (4.5b)$$

$$(\Pi_h u, w)_K = (u, w)_K \quad \forall w \in \mathcal{P}_{k-1}(K), \quad (4.5c)$$

$$\langle \Pi_h u, \mu \rangle_{F_K^*} = \langle u, \mu \rangle_{F_K^*}, \quad \forall \mu \in \mathcal{P}_k(F_K^*). \quad (4.5d)$$

where  $F_K^s$  and  $F_K^*$  are defined in (2.6). We have the following optimal approximation property for  $\mathbf{\Pi}_h$  and  $\Pi_h$ , whose proof was available in ([15], Prop. 2.1).

**Lemma 4.2.** *Assume that  $\mathbf{q} \in H^{s+1}(K; \mathbb{R}^d)$  for  $s \in [0, k]$  on an element  $K \in \mathcal{T}_h$ . Then*

$$\|\mathbf{\Pi}_h \mathbf{q} - \mathbf{q}\|_K \leq C h^{s+1} |\mathbf{q}|_{H^{s+1}(K; \mathbb{R}^d)}.$$

Assume that  $u \in H^{s+1}(K)$  for  $s \in [0, k]$  on an element  $K \in \mathcal{T}_h$ . Then

$$\|\Pi_h u - u\|_K \leq C h^{s+1} |u|_{H^{s+1}(K)}.$$

We also need to estimate the difference between  $\mathbf{q}$ ,  $u$  and  $\lambda$  and their corresponding projections. Such an estimate is established in the following lemma. We refer the readers to Lemma 4.2 in [1] for a detailed proof.

**Lemma 4.3.** *Let  $K \in \mathcal{T}_h$  and  $\eta \in C^1(\bar{K}) \cap W^{k+1, \infty}(K)$ . Then, for any  $(\mathbf{v}, v) \in P_k(K; \mathbb{R}^d) \times P_k(K)$  and  $\chi \in \mathbb{R}$ ,*

$$\begin{aligned} \|\Pi_h((\eta + \chi)\mathbf{v}) - (\eta + \chi)\mathbf{v}\|_K &\leq Ch_K \|\eta\|_{W^{k+1, \infty}(K)} \|\mathbf{v}\|_K, \\ \|\Pi_h((\eta + \chi)\mathbf{v}) - (\eta + \chi)\mathbf{v}\|_F &\leq Ch_K^{1/2} \|\eta\|_{W^{k+1, \infty}(K)} \|\mathbf{v}\|_K, \quad \forall F \in \partial K, \\ \|\Pi_h((\eta + \chi)v) - (\eta + \chi)v\|_K &\leq Ch_K \|\eta\|_{W^{k+1, \infty}(K)} \|v\|_K, \\ \|\Pi_h((\eta + \chi)v) - (\eta + \chi)v\|_F &\leq Ch_K^{1/2} \|\eta\|_{W^{k+1, \infty}(K)} \|v\|_K, \quad \forall F \in \partial K. \end{aligned}$$

Now, we go back to the stability estimate in Lemma 4.1, and divide the left hand side of the inequality into two terms, namely,

$$\begin{aligned} B((\mathbf{q}_h, u_h, \lambda_h), (\mathbf{q}_\varphi, u_\varphi, \lambda_\varphi)) &= B((\mathbf{q}_h, u_h, \lambda_h), (\Pi_h \mathbf{q}_\varphi, \Pi_h u_\varphi, P_M \lambda_\varphi)) \\ &\quad + B((\mathbf{q}_h, u_h, \lambda_h), ((\text{Id} - \Pi_h) \mathbf{q}_\varphi, (\text{Id} - \Pi_h) u_\varphi, (\text{Id} - P_M) \lambda_\varphi)). \end{aligned}$$

**Step 3.** We define the union of faces to simplify the presentation:

$$\begin{aligned} \partial \mathcal{T}_h^* &:= \cup_{K \in \mathcal{T}_h} \cup_{F \in \partial K \setminus F_K^*} F, \\ \partial \mathcal{T}_h^s &:= \cup_{K \in \mathcal{T}_h} F_K^s, \end{aligned}$$

where  $F_K^*$  and  $F_K^s$  are defined in (2.6). Now, we are ready to derive the discrete stability result for the HDG method.

**Lemma 4.4.** *Let  $\tau$  satisfies assumptions (2.8), then there exists  $h_0$ , independent of  $\epsilon$ , so that for any  $h < h_0$ , we have the following stability estimate: for all  $(\mathbf{q}_h, u_h, \lambda_h) \in \mathbf{V}_h \times W_h \times M_h(0)$ ,*

$$\sup_{0 \neq (\mathbf{r}_h, w_h, \mu_h) \in \mathbf{V}_h \times W_h \times M_h(0)} \frac{B((\mathbf{q}_h, u_h, \lambda_h), (\mathbf{r}_h, w_h, \mu_h))}{\|(\mathbf{r}_h, w_h, \mu_h)\|_e} \geq C \|(\mathbf{q}_h, u_h, \lambda_h)\|_e.$$

*Proof.* For any  $(\mathbf{r}, w, \mu) \in H^1(\mathcal{T}_h; \mathbb{R}^d) \times H^1(\mathcal{T}_h) \times L^2(\mathcal{E}_h)$  with  $\mu = 0$  on  $\partial \Omega$ , define

$$\delta \mathbf{r} := \mathbf{r} - \Pi_h \mathbf{r}, \delta w := w - \Pi_h w, \delta \mu := \mu - P_M \mu.$$

Using integration by parts and the definition of the projections, we get

$$\begin{aligned} B((\mathbf{q}_h, u_h, \lambda_h), (\delta \mathbf{r}, \delta w, \delta \mu)) &= (\epsilon^{-1} \mathbf{q}_h, \delta \mathbf{r})_{\mathcal{T}_h} - (u_h, \nabla \cdot \delta \mathbf{r})_{\mathcal{T}_h} + \langle \lambda_h, \delta \mathbf{r} \cdot \mathbf{n} \rangle_{\partial \mathcal{T}_h} \\ &\quad - (\mathbf{q}_h + \beta u_h, \nabla \delta w)_{\mathcal{T}_h} + \langle (\mathbf{q}_h + \beta \lambda_h) \cdot \mathbf{n} + \tau(u_h - \lambda_h), \delta w \rangle_{\partial \mathcal{T}_h} \\ &\quad - ((\nabla \cdot \beta) u_h, \delta w)_{\mathcal{T}_h} - \langle (\mathbf{q}_h + \beta \lambda_h) \cdot \mathbf{n} + \tau(u_h - \lambda_h), \delta \mu \rangle_{\partial \mathcal{T}_h} \\ &= (\epsilon^{-1} \mathbf{q}_h, \delta \mathbf{r})_{\mathcal{T}_h} + (\nabla u_h, \delta \mathbf{r})_{\mathcal{T}_h} + \langle \lambda_h - u_h, \delta \mathbf{r} \cdot \mathbf{n} \rangle_{\partial \mathcal{T}_h} \\ &\quad + (\nabla \cdot \mathbf{q}_h, \delta w)_{\mathcal{T}_h} + (\beta \cdot \nabla u_h, \delta w)_{\mathcal{T}_h} + \langle (\tau - \beta \cdot \mathbf{n})(u_h - \lambda_h), \delta w \rangle_{\partial \mathcal{T}_h} \\ &\quad - \langle (\mathbf{q}_h + \beta \lambda_h) \cdot \mathbf{n} + \tau(u_h - \lambda_h), \delta \mu \rangle_{\partial \mathcal{T}_h} \\ &= (\epsilon^{-1} \mathbf{q}_h, \delta \mathbf{r})_{\mathcal{T}_h} + \langle \lambda_h - u_h, \delta \mathbf{r} \cdot \mathbf{n} \rangle_{\partial \mathcal{T}_h^s} + ((\beta - \mathbf{P}_{0,h} \beta) \cdot \nabla u_h, \delta w)_{\mathcal{T}_h} \\ &\quad + \langle (\tau - \beta \cdot \mathbf{n})(u_h - \lambda_h), \delta w \rangle_{\partial \mathcal{T}_h^*} - \langle \beta \cdot \mathbf{n}(u_h - \lambda_h), \delta w \rangle_{\partial \mathcal{T}_h \setminus \partial \mathcal{T}_h^*}, \end{aligned}$$

where  $\mathbf{P}_{0,h}$  is the vectorial piecewise-constant projection.

Now, we take  $(\mathbf{r}, w, \mu) = (\mathbf{q}_\varphi, u_\varphi, \lambda_\varphi)$  as in Lemma 4.1. By Cauchy–Schwartz inequality and the approximation results in Lemma 4.3, we have

$$\begin{aligned} (\epsilon^{-1}\mathbf{q}_h, \delta\mathbf{q}_\varphi)_{\mathcal{T}_h} &\leq \|\epsilon^{-1/2}\mathbf{q}_h\|_{\mathcal{T}_h} \|\epsilon^{-1/2}\delta\mathbf{q}_\varphi\|_{\mathcal{T}_h} \\ &\leq Ch\|\epsilon^{-1/2}\mathbf{q}_h\|_{\mathcal{T}_h}^2, \end{aligned}$$

$$\begin{aligned} \langle \lambda_h - u_h, \delta\mathbf{q}_\varphi \cdot \mathbf{n} \rangle_{\partial\mathcal{T}_h^s} &\leq \left\| \left| \tau - \frac{1}{2}\boldsymbol{\beta} \cdot \mathbf{n} \right|^{1/2} (\lambda_h - u_h) \right\|_{\partial\mathcal{T}_h^s} \left\| \left| \tau - \frac{1}{2}\boldsymbol{\beta} \cdot \mathbf{n} \right|^{-1/2} \delta\mathbf{q}_\varphi \right\|_{\partial\mathcal{T}_h^s} \\ &\leq C \left( \frac{\epsilon}{\tau^v} \right)^{1/2} \left\| \left| \tau - \frac{1}{2}\boldsymbol{\beta} \cdot \mathbf{n} \right|^{1/2} (\lambda_h - u_h) \right\|_{\partial\mathcal{T}_h^s} \|\epsilon^{-1/2}\delta\mathbf{q}_\varphi\|_{\partial\mathcal{T}_h^s} \\ &\leq C \left( \frac{\epsilon h}{\tau^v} \right)^{1/2} \left\| \left| \tau - \frac{1}{2}\boldsymbol{\beta} \cdot \mathbf{n} \right|^{1/2} (\lambda_h - u_h) \right\|_{\partial\mathcal{T}_h^s} \|\epsilon^{-1/2}\mathbf{q}_h\|_{\mathcal{T}_h} \\ &\leq C(h^2 + \epsilon h)^{1/2} \left\| \left| \tau - \frac{1}{2}\boldsymbol{\beta} \cdot \mathbf{n} \right|^{1/2} (\lambda_h - u_h) \right\|_{\partial\mathcal{T}_h^s} \|\epsilon^{-1/2}\mathbf{q}_h\|_{\mathcal{T}_h} \quad \text{by (2.8b),} \end{aligned}$$

$$\begin{aligned} ((\boldsymbol{\beta} - \mathbf{P}_{0,h}\boldsymbol{\beta}) \cdot \nabla u_h, \delta u_\varphi)_{\mathcal{T}_h} &\leq Ch\|\nabla u_h\|_{\mathcal{T}_h} \|\delta u_\varphi\|_{\mathcal{T}_h} \\ &\leq Ch\|u_h\|_{\mathcal{T}_h}^2 \end{aligned}$$

$$\begin{aligned} \langle (\tau - \boldsymbol{\beta} \cdot \mathbf{n})(u_h - \lambda_h), \delta u_\varphi \rangle_{\partial\mathcal{T}_h^*} &\leq \left\| |\tau - \boldsymbol{\beta} \cdot \mathbf{n}|^{1/2} (\lambda_h - u_h) \right\|_{\partial\mathcal{T}_h^*} \left\| |\tau - \boldsymbol{\beta} \cdot \mathbf{n}|^{1/2} \delta u_\varphi \right\|_{\partial\mathcal{T}_h^*} \\ &\leq C \left\| \left| \tau - \frac{1}{2}\boldsymbol{\beta} \cdot \mathbf{n} \right|^{1/2} (\lambda_h - u_h) \right\|_{\partial\mathcal{T}_h^*} \left\| |\tau - \boldsymbol{\beta} \cdot \mathbf{n}|^{1/2} \delta u_\varphi \right\|_{\partial\mathcal{T}_h^*} \quad \text{by (2.8c)} \\ &\leq C(h(\tau^w + 1))^{1/2} \left\| \left| \tau - \frac{1}{2}\boldsymbol{\beta} \cdot \mathbf{n} \right|^{1/2} (\lambda_h - u_h) \right\|_{\partial\mathcal{T}_h^*} \|u_h\|_{\mathcal{T}_h} \\ &\leq Ch^{1/2} \left\| \left| \tau - \frac{1}{2}\boldsymbol{\beta} \cdot \mathbf{n} \right|^{1/2} (\lambda_h - u_h) \right\|_{\partial\mathcal{T}_h^*} \|u_h\|_{\mathcal{T}_h} \quad \text{by (2.8a)} \end{aligned}$$

$$\begin{aligned} \langle \boldsymbol{\beta} \cdot \mathbf{n}(u_h - \lambda_h), \delta u_\varphi \rangle_{\partial\mathcal{T}_h \setminus \partial\mathcal{T}_h^*} &\leq \left\| |\boldsymbol{\beta} \cdot \mathbf{n}|^{1/2} (\lambda_h - u_h) \right\|_{\partial\mathcal{T}_h \setminus \partial\mathcal{T}_h^*} \|\delta u_\varphi\|_{\partial\mathcal{T}_h \setminus \partial\mathcal{T}_h^*} \\ &\leq C \left\| \left| \tau - \frac{1}{2}\boldsymbol{\beta} \cdot \mathbf{n} \right|^{1/2} (\lambda_h - u_h) \right\|_{\partial\mathcal{T}_h \setminus \partial\mathcal{T}_h^*} \|\delta u_\varphi\|_{\partial\mathcal{T}_h \setminus \partial\mathcal{T}_h^*} \quad \text{by (2.8c)} \\ &\leq Ch^{1/2} \left\| \left| \tau - \frac{1}{2}\boldsymbol{\beta} \cdot \mathbf{n} \right|^{1/2} (\lambda_h - u_h) \right\|_{\partial\mathcal{T}_h} \|u_h\|_{\mathcal{T}_h}. \end{aligned}$$

Summing the above inequalities all together, we get

$$B((\mathbf{q}_h, u_h, \lambda_h), (\delta\mathbf{q}_\varphi, \delta u_\varphi, \delta\lambda_\varphi)) \leq Ch^{1/2} \|(\mathbf{q}_h, u_h, \lambda_h)\|_e^2.$$

Hence, choosing  $h$  sufficiently small, we can ensure that

$$B((\mathbf{q}_h, u_h, \lambda_h), (\delta\mathbf{q}_\varphi, \delta u_\varphi, \delta\lambda_\varphi)) \leq \frac{1}{2} B((\mathbf{q}_h, u_h, \lambda_h), (\mathbf{q}_\varphi, u_\varphi, \lambda_\varphi)).$$

Consequently, we obtain

$$B((\mathbf{q}_h, u_h, \lambda_h), (\mathbf{\Pi}_h \mathbf{q}_\varphi, P_h u_\varphi, P_M \lambda_\varphi)) \geq C \|\|(\mathbf{q}_h, u_h, \lambda_h)\|\|_e^2.$$

On the other hand, it is easy to obtain the following estimates

$$\|\|(\mathbf{\Pi}_h \mathbf{q}_\varphi, P_h u_\varphi, P_M \lambda_\varphi)\|\|_e \leq C \|\|(\mathbf{q}_h, u_h, \lambda_h)\|\|_e.$$

We conclude the proof by combining these two estimates.  $\square$

## 4.2. The error equation

Here, we obtain the equation satisfied by the errors. Note that by Galerkin-orthogonality, we have

$$B((\mathbf{q} - \mathbf{q}_h, u - u_h, u - \widehat{u}_h), (\mathbf{r}, w, \mu)) = 0 \quad \forall (\mathbf{r}, w, \mu) \in V_h \times W_h \times M_h(0), \quad (4.6)$$

where  $(\mathbf{q}, u)$  is the exact solution of equations (2.2).

We define the following quantities that will be used in the analysis:

$$\begin{aligned} \varepsilon_h^q &:= \mathbf{q}_h - \mathbf{\Pi}_h \mathbf{q}, & \delta \mathbf{q} &:= \mathbf{q} - \mathbf{\Pi}_h \mathbf{q}, \\ \varepsilon_h^u &:= u_h - \Pi_h u, & \delta u &:= u - \Pi_h u, \\ \varepsilon_h^{\widehat{u}} &:= \widehat{u}_h - P_M u, & \widehat{\delta u} &:= u - P_M u. \end{aligned}$$

Recall that  $\mathbf{\Pi}_h$  and  $\Pi_h$  are the projections defined in (4.5), and  $P_M$  is the  $L^2$ -projection from  $L^2(\mathcal{E}_h)$  onto  $M_h$ .

Now, we are ready to present our error equation.

**Lemma 4.5.** *The error equation takes the following form.*

$$\begin{aligned} B((\varepsilon_h^q, \varepsilon_h^u, \varepsilon_h^{\widehat{u}}), (\mathbf{r}, w, \mu)) &= (\varepsilon^{-1} \delta \mathbf{q}, \mathbf{r})_{\mathcal{T}_h} + \langle \delta \mathbf{q} \cdot \mathbf{n}, w - \mu \rangle_{\partial \mathcal{T}_h^s} - (\boldsymbol{\beta} \delta u, \nabla w)_{\mathcal{T}_h} \\ &\quad - ((\nabla \cdot \boldsymbol{\beta}) \delta u, w)_{\mathcal{T}_h} + \langle \boldsymbol{\beta} \cdot \mathbf{n} \widehat{\delta u}, w \rangle_{\partial \mathcal{T}_h} + \langle \tau \delta u, w - \mu \rangle_{\partial \mathcal{T}_h^*}, \end{aligned} \quad (4.7)$$

for all  $(\mathbf{r}, w, \mu) \in V_h \times W_h \times M_h(0)$ .

*Proof.* We use the Galerkin-orthogonality (4.6) and the definition of the projections to prove the result. For all  $(\mathbf{r}, w, \mu) \in V_h \times W_h \times M_h(0)$ , we have

$$\begin{aligned} B((\varepsilon_h^q, \varepsilon_h^u, \varepsilon_h^{\widehat{u}}), (\mathbf{r}, w, \mu)) &= B((\delta \mathbf{q}, \delta u, \widehat{\delta u}), (\mathbf{r}, w, \mu)) \\ &= (\varepsilon^{-1} \delta \mathbf{q}, \mathbf{r})_{\mathcal{T}_h} - (\delta u, \nabla \cdot \mathbf{r})_{\mathcal{T}_h} + \langle \widehat{\delta u}, \mathbf{r} \cdot \mathbf{n} \rangle_{\partial \mathcal{T}_h} \\ &\quad - (\delta \mathbf{q} + \boldsymbol{\beta} \delta u, \nabla w)_{\mathcal{T}_h} + \langle (\delta \mathbf{q} + \boldsymbol{\beta} \widehat{\delta u}) \cdot \mathbf{n} + \tau (\delta u - \widehat{\delta u}), w - \mu \rangle_{\partial \mathcal{T}_h} \\ &\quad - ((\nabla \cdot \boldsymbol{\beta}) \delta u, w)_{\mathcal{T}_h} \\ &= (\varepsilon^{-1} \delta \mathbf{q}, \mathbf{r})_{\mathcal{T}_h} + \langle \delta \mathbf{q} \cdot \mathbf{n}, w - \mu \rangle_{\partial \mathcal{T}_h^s} - (\boldsymbol{\beta} \delta u, \nabla w)_{\mathcal{T}_h} \\ &\quad - ((\nabla \cdot \boldsymbol{\beta}) \delta u, w)_{\mathcal{T}_h} + \langle \boldsymbol{\beta} \cdot \mathbf{n} \widehat{\delta u}, w - \mu \rangle_{\partial \mathcal{T}_h} + \langle \tau \delta u, w - \mu \rangle_{\partial \mathcal{T}_h^*} \\ &= (\varepsilon^{-1} \delta \mathbf{q}, \mathbf{r})_{\mathcal{T}_h} + \langle \delta \mathbf{q} \cdot \mathbf{n}, w - \mu \rangle_{\partial \mathcal{T}_h^s} - (\boldsymbol{\beta} \delta u, \nabla w)_{\mathcal{T}_h} \\ &\quad - ((\nabla \cdot \boldsymbol{\beta}) \delta u, w)_{\mathcal{T}_h} + \langle \boldsymbol{\beta} \cdot \mathbf{n} \widehat{\delta u}, w \rangle_{\partial \mathcal{T}_h} + \langle \tau \delta u, w - \mu \rangle_{\partial \mathcal{T}_h^*}, \end{aligned}$$

where in the last step we used the fact that  $\langle \boldsymbol{\beta} \cdot \mathbf{n} \widehat{\delta u}, \mu \rangle_{\partial \mathcal{T}_h} = 0$  for all  $\mu \in M_h(0)$ .  $\square$

### 4.3. The error analysis

Now, we are ready to prove our main results, Theorems 2.1 and 2.2. In order to prove Theorem 2.1 and Theorem 2.2. We only need to bound the right hand side of the error equation (4.7) to get the error estimates. For all  $(\mathbf{r}, w, \mu) \in V_h \times W_h \times M_h(0)$ , we have

$$\begin{aligned}
(\epsilon^{-1} \delta \mathbf{q}, \mathbf{r})_{\mathcal{T}_h} &\leq \|\epsilon^{-1/2} \delta \mathbf{q}\|_{\mathcal{T}_h} \|\epsilon^{-1/2} \mathbf{r}\|_{\mathcal{T}_h}, \\
\langle \delta \mathbf{q} \cdot \mathbf{n}, w - \mu \rangle_{\partial \mathcal{T}_h^s} &\leq \left\| \left| \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right|^{-1/2} \delta \mathbf{q} \right\|_{\partial \mathcal{T}_h^s} \left\| \left| \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right|^{1/2} (w - \mu) \right\|_{\partial \mathcal{T}_h^s} \\
&\leq C \left( \frac{\epsilon}{\tau^v} \right)^{1/2} \|\epsilon^{-1/2} \delta \mathbf{q}\|_{\partial \mathcal{T}_h^s} \left\| \left| \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right|^{1/2} (w - \mu) \right\|_{\partial \mathcal{T}_h^s}, \\
(\boldsymbol{\beta} \delta u, \nabla w)_{\mathcal{T}_h} &= ((\boldsymbol{\beta} - \mathbf{P}_{0,h} \boldsymbol{\beta}) \delta u, \nabla w)_{\mathcal{T}_h} \\
&\leq Ch \|\delta u\|_{\mathcal{T}_h} \|\nabla w\|_{\mathcal{T}_h} \\
&\leq C \|\delta u\|_{\mathcal{T}_h} \|w\|_{\mathcal{T}_h},
\end{aligned}$$

$$\begin{aligned}
((\nabla \cdot \boldsymbol{\beta}) \delta u, w)_{\mathcal{T}_h} &\leq C \|\delta u\|_{\mathcal{T}_h} \|w\|_{\mathcal{T}_h}, \\
\langle \boldsymbol{\beta} \cdot \mathbf{n} \widehat{\delta u}, w \rangle_{\partial \mathcal{T}_h} &= \langle (\boldsymbol{\beta} - \mathbf{P}_{0,h} \boldsymbol{\beta}) \cdot \mathbf{n} \widehat{\delta u}, w \rangle_{\partial \mathcal{T}_h} \\
&\leq Ch \|\widehat{\delta u}\|_{\partial \mathcal{T}_h} \|w\|_{\partial \mathcal{T}_h} \\
&\leq Ch^{1/2} \|\widehat{\delta u}\|_{\partial \mathcal{T}_h} \|w\|_{\mathcal{T}_h} \\
\langle \tau \delta u, w - \mu \rangle_{\partial \mathcal{T}_h^*} &\leq \|\tau^{1/2} \delta u\|_{\partial \mathcal{T}_h^*} \|\tau^{1/2} (w - \mu)\|_{\partial \mathcal{T}_h^*} \\
&\leq C \|\tau^{1/2} \delta u\|_{\partial \mathcal{T}_h^*} \left\| \left| \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right|^{1/2} (w - \mu) \right\|_{\partial \mathcal{T}_h^*} \quad \text{by (2.8c)} \\
&\leq C (\tau^w)^{1/2} \|\delta u\|_{\partial \mathcal{T}_h^*} \left\| \left| \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right|^{1/2} (w - \mu) \right\|_{\partial \mathcal{T}_h^*}.
\end{aligned}$$

Adding up these estimates all together, we obtain

$$\begin{aligned}
B((\boldsymbol{\varepsilon}_h^q, \boldsymbol{\varepsilon}_h^u, \boldsymbol{\varepsilon}_h^{\widehat{u}}), (\mathbf{r}, w, \mu)) &\leq C (\|\epsilon^{-1/2} \delta \mathbf{q}\|_{\mathcal{T}_h} + \left( \frac{\epsilon}{\tau^v} \right)^{1/2} \|\epsilon^{-1/2} \delta \mathbf{q}\|_{\partial \mathcal{T}_h^s} + \|\delta u\|_{\mathcal{T}_h} \\
&\quad + h^{1/2} \|\widehat{\delta u}\|_{\partial \mathcal{T}_h} + (\tau^w)^{1/2} \|\delta u\|_{\partial \mathcal{T}_h^*}) \|(\mathbf{r}, w, \mu)\|_e.
\end{aligned}$$

Note that  $\boldsymbol{\varepsilon}_h^{\widehat{u}} \in M_h(0)$  because  $\boldsymbol{\varepsilon}_h^{\widehat{u}}|_{\partial \Omega} = 0$ . Using Lemma 4.4, we immediately get

$$\begin{aligned}
\|(\boldsymbol{\varepsilon}_h^q, \boldsymbol{\varepsilon}_h^u, \boldsymbol{\varepsilon}_h^{\widehat{u}})\|_e &\leq C \left( \|\epsilon^{-1/2} \delta \mathbf{q}\|_{\mathcal{T}_h} + \left( \frac{\epsilon}{\tau^v} \right)^{1/2} \|\epsilon^{-1/2} \delta \mathbf{q}\|_{\partial \mathcal{T}_h^s} + \|\delta u\|_{\mathcal{T}_h} \right. \\
&\quad \left. + h^{1/2} \|\widehat{\delta u}\|_{\partial \mathcal{T}_h} + (\tau^w)^{1/2} \|\delta u\|_{\partial \mathcal{T}_h^*} \right),
\end{aligned}$$

The approximation properties of the projections gives the following estimates,

$$\begin{aligned} \|\epsilon^{-1/2}\delta\mathbf{q}\|_{\mathcal{T}_h} &\leq C\epsilon^{-1/2}h^{s+1}|\mathbf{q}|_{H^{s+1}(\mathcal{T}_h;\mathbb{R}^d)} \\ \left(\frac{\epsilon}{\tau^w}\right)^{1/2}\|\epsilon^{-1/2}\delta\mathbf{q}\|_{\partial\mathcal{T}_h^s} &\leq C\left(\frac{h}{\tau^w}\right)^{1/2}h^s|\mathbf{q}|_{H^{s+1}(\mathcal{T}_h;\mathbb{R}^d)} \\ \|\delta u\|_{\mathcal{T}_h} &\leq Ch^{s+1}|u|_{H^{s+1}(\mathcal{T}_h)} \\ h^{1/2}\|\widehat{\delta u}\|_{\partial\mathcal{T}_h} &\leq Ch^{s+1}|u|_{H^{s+1}(\mathcal{T}_h)} \\ (\tau^w)^{1/2}\|\delta u\|_{\partial\mathcal{T}_h^*} &\leq C(\tau^w)^{1/2}h^{s+1/2}|u|_{H^{s+1}(\mathcal{T}_h)}, \end{aligned}$$

for all  $s \in [0, k]$ . Now, using assumption (2.8) on  $\tau$ , we obtain the following estimates,

$$\|(\boldsymbol{\varepsilon}_h^{\mathbf{q}}, \varepsilon_h^u, \varepsilon_h^{\widehat{u}})\|_e \leq C(\epsilon^{-1/2}h^{s_{\mathbf{v}}+1} + h^{s_{\mathbf{v}}+1/2})|\mathbf{q}|_{H^{s_{\mathbf{v}}+1}(\mathcal{T}_h;\mathbb{R}^d)} + Ch^{s_w+1/2}|u|_{H^{s_w+1}(\mathcal{T}_h)},$$

for all  $s_{\mathbf{v}} \in [0, k]$  and  $s_w \in [0, k]$ . Using the fact that  $|\mathbf{q}|_{H^k(\mathcal{T}_h;\mathbb{R}^d)} = \epsilon|u|_{H^{k+1}(\mathcal{T}_h)}$ , we get

$$\|(\boldsymbol{\varepsilon}_h^{\mathbf{q}}, \varepsilon_h^u, \varepsilon_h^{\widehat{u}})\|_e \leq C\left(\epsilon^{1/2}h^{s_{\mathbf{v}}+1} + \epsilon h^{s_{\mathbf{v}}+1/2}\right)|u|_{H^{s_{\mathbf{v}}+2}(\mathcal{T}_h;\mathbb{R}^d)} + Ch^{s_w+1/2}|u|_{H^{s_w+1}(\mathcal{T}_h)}, \quad (4.8)$$

for all  $s_{\mathbf{v}} \in [0, k]$  and  $s_w \in [0, k]$ . Moreover, by approximation properties of the projection, we can easily get

$$\|(\delta\mathbf{q}, \delta u, \widehat{\delta u})\|_e \leq C\epsilon^{1/2}h^{s_{\mathbf{v}}+1}|u|_{H^{s_{\mathbf{v}}+2}(\mathcal{T}_h;\mathbb{R}^d)} + Ch^{s_w+1/2}|u|_{H^{s_w+1}(\mathcal{T}_h)}, \quad (4.9)$$

for all  $s_{\mathbf{v}} \in [0, k]$  and  $s_w \in [0, k]$ . Combining (4.8), (4.9) and using the triangle inequality, we obtain

$$\|(\mathbf{q} - \mathbf{q}_h, u - u_h, u - \widehat{u}_h)\|_e \leq C(\epsilon^{1/2}h^{s_{\mathbf{v}}+1} + \epsilon h^{s_{\mathbf{v}}+1/2})|u|_{H^{s_{\mathbf{v}}+2}(\mathcal{T}_h;\mathbb{R}^d)} + Ch^{s_w+1/2}|u|_{H^{s_w+1}(\mathcal{T}_h)},$$

and

$$\begin{aligned} \|u - u_h\|_{\mathcal{T}_h} &\leq \|(\boldsymbol{\varepsilon}_h^{\mathbf{q}}, \varepsilon_h^u, \varepsilon_h^{\widehat{u}})\|_e + \|\delta u\|_{\mathcal{T}_h} \\ &\leq C\left(\epsilon^{1/2}h^{s_{\mathbf{v}}+1} + \epsilon h^{s_{\mathbf{v}}+1/2}\right)|u|_{H^{s_{\mathbf{v}}+2}(\mathcal{T}_h;\mathbb{R}^d)} + Ch^{s_w+1/2}|u|_{H^{s_w+1}(\mathcal{T}_h)}. \end{aligned}$$

This completes the proof of Theorem 2.1.

For the proof of Theorem 2.2, everything is exactly the same, except that

$$(\tau^w)^{1/2}\|\delta u\|_{\partial\mathcal{T}_h^*} \leq Ch^{s+1}|u|_{H^{s+1}(\mathcal{T}_h)},$$

because assumption (2.9) ensures  $\tau^w \leq \mathcal{O}(h)$ . Hence we get

$$\begin{aligned} \|(\boldsymbol{\varepsilon}_h^{\mathbf{q}}, \varepsilon_h^u, \varepsilon_h^{\widehat{u}})\|_e &\leq C(\epsilon^{1/2}h^{s_{\mathbf{v}}+1} + \epsilon h^{s_{\mathbf{v}}+1/2})|u|_{H^{s_{\mathbf{v}}+2}(\mathcal{T}_h;\mathbb{R}^d)} + Ch^{s_w+1}|u|_{H^{s_w+1}(\mathcal{T}_h)} \\ \|u - u_h\|_{\mathcal{T}_h} &\leq \|(\boldsymbol{\varepsilon}_h^{\mathbf{q}}, \varepsilon_h^u, \varepsilon_h^{\widehat{u}})\|_e + \|\delta u\|_{\mathcal{T}_h} \\ &\leq C(\epsilon^{1/2}h^{s_{\mathbf{v}}+1} + \epsilon h^{s_{\mathbf{v}}+1/2})|u|_{H^{s_{\mathbf{v}}+2}(\mathcal{T}_h;\mathbb{R}^d)} + Ch^{s_w+1}|u|_{H^{s_w+1}(\mathcal{T}_h)}. \end{aligned}$$

This completes the proof of Theorem 2.2.

## 5. NUMERICAL RESULTS

In this section, we present numerical studies using simple model problems in 2D to verify our theoretical results and display the performance of the HDG methods when the exact solution exhibit layers. Our test problems are similar to those studied in [1]. We fix the domain to be the unit square in all the experiments, and run simulations of the HDG methods (2.4) with the following three choices of approximation spaces and stabilization functions:

1. The approximation spaces are

$$\begin{aligned} \mathbf{V}_h &= \{\mathbf{r} \in L^2(\Omega; \mathbb{R}^d) : \mathbf{r}|_K \in P_k(K; \mathbb{R}^d) \quad \forall K \in \mathcal{T}_h\}, \\ W_h &= \{w \in L^2(\Omega) : w|_K \in P_k(K) \quad \forall K \in \mathcal{T}_h\}, \\ M_h &= \{\mu \in L^2(\mathcal{E}_h) : \mu|_F \in P_k(F) \quad \forall F \in \mathcal{E}_h\}, \end{aligned}$$

while the stabilization function is given by

$$\tau(F) = \max \left( \sup_{\mathbf{x} \in F} \boldsymbol{\beta}(\mathbf{x}) \cdot \mathbf{n}, 0 \right), \quad \forall F \in \partial K, \forall K \in \mathcal{T}_h.$$

We denote this choice as  $P_k$ -HDG1.

2. The approximation spaces are the same as in the previous case, while the stabilization function is given by

$$\tau(F) = \max \left( \sup_{\mathbf{x} \in F} \boldsymbol{\beta}(\mathbf{x}) \cdot \mathbf{n}, 0 \right) + \min \left( 0.1 \frac{\epsilon}{h_F}, 1 \right), \quad \forall F \in \partial K, \forall K \in \mathcal{T}_h.$$

We denote this choice as  $P_k$ -HDG2.

3. The approximation spaces are given as follows:

$$\begin{aligned} \mathbf{V}_h &= \{\mathbf{r} \in L^2(\Omega; \mathbb{R}^d) : \mathbf{r}|_K \in P_k(K; \mathbb{R}^d) + \mathbf{x}P_k(K) \quad \forall K \in \mathcal{T}_h\}, \\ W_h &= \{w \in L^2(\Omega) : w|_K \in P_k(K) \quad \forall K \in \mathcal{T}_h\}, \\ M_h &= \{\mu \in L^2(\mathcal{E}_h) : \mu|_F \in P_k(F) \quad \forall F \in \mathcal{E}_h\}, \end{aligned}$$

while the stabilization function is the same as  $P_k$ -HDG1. We denote this method as  $P_k$ -HDG3.

We remark that the method  $P_k$ -HDG3 is exactly the MH-DG method considered in [26] when  $\boldsymbol{\beta}$  is piecewise-constant, which is proven in Appendix A.

### 5.1. A smooth solution test

We take the velocity field  $\boldsymbol{\beta} = [1, 2]^T$ , and the diffusion coefficient  $\epsilon$  as  $1, 10^{-3}, 10^{-9}$ . The source term  $f$  is chosen so that the exact solution is  $u(x, y) = \sin(2\pi x) \sin(2\pi y)$ . We obtain the computational meshes by uniform refinement of a mesh that consists of a structured  $5 \times 5 \times 2$  triangular elements, where the slanted edges are pointing in the northeast direction.

Let us remark that when  $\epsilon = 1$  and  $k \geq 1$ , we can follow [9] to use superconvergence results to locally postprocess the solution to get a new approximation of the scalar variable  $u_h^*$ , which converges faster than  $u_h$ . Here the definition of  $u_h^* \in P_{k+1}(K)$  for each element  $K \in \mathcal{T}_h$  is as follows:

$$\begin{aligned} (\nabla u_h^*, \nabla w)_K &= -(\epsilon^{-1} \mathbf{q}_h, \nabla w)_K && \text{for all } w \in P_{k+1}(K), \\ (u_h^*, 1)_K &= (u_h, 1)_K. \end{aligned}$$

Tables 1 and 2 show the  $L^2$ -convergence results for  $u_h$  and  $u_h^*$  for the three HDG methods when  $\epsilon = 1$ . For all the methods, we observe convergence order of  $k+1$  for  $u_h$  and convergence order of  $k+2$  for  $u_h^*$  (when  $k \geq 1$ ). Also note that the errors for the postprocessing  $u_h^*$  for the three methods are very close to each other.

TABLE 1. History of convergence for  $\|u - u_h\|_{L^2(\mathcal{T}_h)}$  when  $\epsilon = 1$ .

Degree Mesh		$\epsilon = 1$					
$k$	$h^{-1}$	error	order	error	order	error	order
		<i>HDG1</i>		<i>HDG2</i>		<i>HDG3</i>	
0	5	1.74e-0	--	7.60e-1	--	2.06e-1	--
	10	9.41e-1	0.88	3.33e-1	1.20	1.06e-1	0.97
	20	4.83e-1	0.96	1.72e-1	0.95	5.29e-2	1.00
	40	2.44e-1	0.99	8.71e-2	0.98	2.64e-2	1.00
1	5	3.75e-1	--	1.72e-1	--	4.88e-2	--
	10	1.01e-1	1.89	3.88e-2	2.15	1.26e-2	1.95
	20	2.59e-2	1.97	9.96e-3	1.96	3.18e-3	1.99
	40	6.52e-3	1.99	2.51e-3	1.99	7.96e-4	2.00
2	5	6.19e-2	--	2.88e-2	--	8.60e-3	--
	10	8.26e-3	2.90	3.20e-3	3.16	1.12e-3	2.95
	20	1.05e-3	2.97	4.09e-4	2.97	1.41e-4	2.99
	40	1.33e-4	2.99	5.16e-5	2.99	1.77e-5	3.00
3	5	8.35e-3	--	3.90e-3	--	1.21e-3	--
	10	5.53e-4	3.92	2.16e-4	4.18	7.81e-5	3.95
	20	3.52e-5	3.98	1.37e-5	3.97	4.92e-6	3.99
	40	2.21e-6	3.99	8.64e-7	3.99	3.08e-7	4.00

TABLE 2. History of convergence for  $\|u - u_h^*\|_{L^2(\mathcal{T}_h)}$  when  $\epsilon = 1$ .

Degree Mesh		$\epsilon = 1$					
$k$	$h^{-1}$	error	order	error	order	error	order
		<i>HDG1</i>		<i>HDG2</i>		<i>HDG3</i>	
1	5	2.25e-2	--	1.70e-2	--	1.39e-2	--
	10	3.08e-3	2.87	2.14e-3	2.99	1.70e-3	3.04
	20	3.94e-4	2.96	2.65e-4	3.02	2.08e-4	3.03
	40	4.96e-5	2.99	3.28e-5	3.01	2.56e-5	3.02
2	5	2.49e-3	--	2.13e-3	--	1.92e-3	--
	10	1.59e-4	3.97	1.35e-4	3.98	1.23e-4	3.97
	20	9.95e-6	4.00	8.45e-6	4.00	7.71e-6	3.99
	40	6.22e-7	4.00	5.28e-7	4.00	4.82e-7	4.00
3	5	2.78e-4	--	2.43e-4	--	2.20e-4	--
	10	8.87e-6	4.97	7.68e-6	4.98	6.94e-6	4.99
	20	2.78e-7	4.99	2.40e-7	5.00	2.17e-7	5.00
	40	8.70e-9	5.00	7.50e-9	5.00	6.77e-9	5.00

When  $\epsilon \ll 1$ , there is no superconvergence result for the HDG methods. Hence we only show  $\|u - u_h\|_{\mathcal{T}_h}$  in Table 3 when  $\epsilon = 10^{-3}$  and  $\epsilon = 10^{-9}$ . Again, optimal converge rates are recovered; better than the one predicted by the theoretical result in Theorem 2.1 which predict the loss of half order of accuracy. Moreover, our numerical results in Table 3 show that the performance of these three HDG methods are equally good. Hence, we prefer to use  $P_k$ -HDG1 and  $P_k$ -HDG2 rather than  $P_k$ -HDG3 because  $P_k$ -HDG3 has more degrees of freedom for the local problems.

TABLE 3. History of convergence for  $\|u - u_h\|_{L^2(\mathcal{T}_h)}$  when  $\epsilon = 10^{-3}$  and  $\epsilon = 10^{-9}$ .

Degree Mesh		$\epsilon = 10^{-3}$				$\epsilon = 10^{-9}$					
$k$	$h^{-1}$	error	order	error	order	error	order	error	order	error	order
		<i>HDG1</i>	<i>HDG2</i>	<i>HDG3</i>	<i>HDG1</i>	<i>HDG2</i>	<i>HDG3</i>	<i>HDG1</i>	<i>HDG2</i>	<i>HDG3</i>	<i>HDG3</i>
0	5	3.16e-1	--	3.16e-1	--	3.14e-1	--	3.18e-1	--	3.18e-1	--
	10	1.71e-1	0.88	1.71e-1	0.88	1.69e-1	0.89	1.74e-1	0.87	1.74e-1	0.87
	20	8.78e-2	0.96	8.78e-2	0.96	8.60e-2	0.98	9.06e-2	0.94	9.06e-2	0.94
	40	4.37e-2	1.00	4.38e-2	1.00	4.22e-2	1.03	4.63e-2	0.97	4.63e-2	0.97
1	5	7.84e-2	--	7.84e-2	--	7.75e-2	--	7.96e-2	--	7.96e-2	--
	10	2.00e-2	1.97	2.00e-2	1.97	1.95e-2	1.99	2.04e-2	1.85	2.04e-2	1.85
	20	4.95e-3	2.01	4.95e-3	2.01	4.73e-3	2.05	5.13e-3	1.96	5.13e-3	1.96
	40	1.21e-3	2.03	1.21e-3	2.03	1.11e-3	2.09	1.28e-3	1.99	1.28e-3	1.99
2	5	1.32e-2	--	1.32e-2	--	1.31e-2	--	1.35e-2	--	1.35e-2	--
	10	1.72e-3	2.95	1.72e-3	2.95	1.68e-3	2.96	1.77e-3	2.93	1.77e-3	2.93
	20	2.14e-4	3.00	2.14e-4	3.00	2.05e-4	3.03	2.24e-4	2.98	2.24e-4	2.98
	40	2.63e-5	3.02	2.63e-5	3.02	2.45e-5	3.07	2.80e-5	3.00	2.80e-5	3.00
3	5	1.83e-3	--	1.83e-3	--	1.80e-3	--	1.87e-3	--	1.87e-3	--
	10	1.17e-4	3.97	1.17e-4	3.97	1.13e-4	3.99	1.20e-4	3.95	1.20e-4	3.95
	20	7.23e-6	4.01	7.23e-6	4.01	6.82e-6	4.05	7.56e-6	3.99	7.56e-6	3.99
	40	4.43e-7	4.03	4.43e-7	4.03	4.01e-7	4.09	4.73e-7	4.00	4.73e-7	4.00

## 5.2. A rotating flow test

We take  $\epsilon = 10^{-6}$ ,  $\beta = [y - 1/2, 1/2 - x]^T$ , and  $f = 0$ . The solution  $u$  is prescribed along the slip  $1/2 \times [0, 1/2]$ , as follows:

$$u(1/2, y) = \sin^2(2\pi y) \quad y \in [0, 1/2].$$

See [31] for a detailed description of this test.

In Figure 1, we plot  $u_h$  obtained from the three HDG methods for various polynomial degrees in a structured triangular grid of 128 elements. To better compare the results, we plot in Figure 2 extracted data of  $u_h$  along the horizontal center line  $y = 1/2$ . We also plot in Figure 3 a comparison of  $P_0$ -HDG1 in 8192 elements and  $P_3$ -HDG1 in 128 elements. From Figure 1, we find that all the HDG methods produce similar results. Moreover, it is clear that higher order methods lead to better approximation results and are computationally cheaper than lower order methods for qualitatively similar numerical results.

## 5.3. An interior layer test

We take  $\beta = [1/2, \sqrt{3}/2]^T$ ,  $f = 0$ , and the Dirichlet boundary condition as follows:

$$u = \begin{cases} 1 & \text{on } \{y = 0, 0 \leq x \leq 1\}, \\ 1 & \text{on } \{x = 0, 0 \leq y \leq 1/5\}, \\ 0 & \text{elsewhere.} \end{cases}$$

It is clear that for  $\epsilon$  small, the exact solution produces an interior layer along  $\beta$  direction starting from  $(0, 1/5)$ , and boundary layers on the right and top right boundary.

In Figures 4 and 5, we plot the computational results in a structured triangular grid of 128 elements for  $\epsilon = 10^{-3}$  and  $\epsilon = 10^{-9}$  respectively. In order to better see the performance of the HDG method in capturing interior layers, in Figures 6, we plot the contour of  $u_h$  using  $P_k$ -HDG1 with  $0 \leq k \leq 3$  for  $\epsilon = 10^{-3}$  in three consecutive meshes with the coarsest one consists of 200 elements. Again, all the HDG methods produce quite similar results. Note that, as expected, the piecewise-constant approximations are free of oscillations but

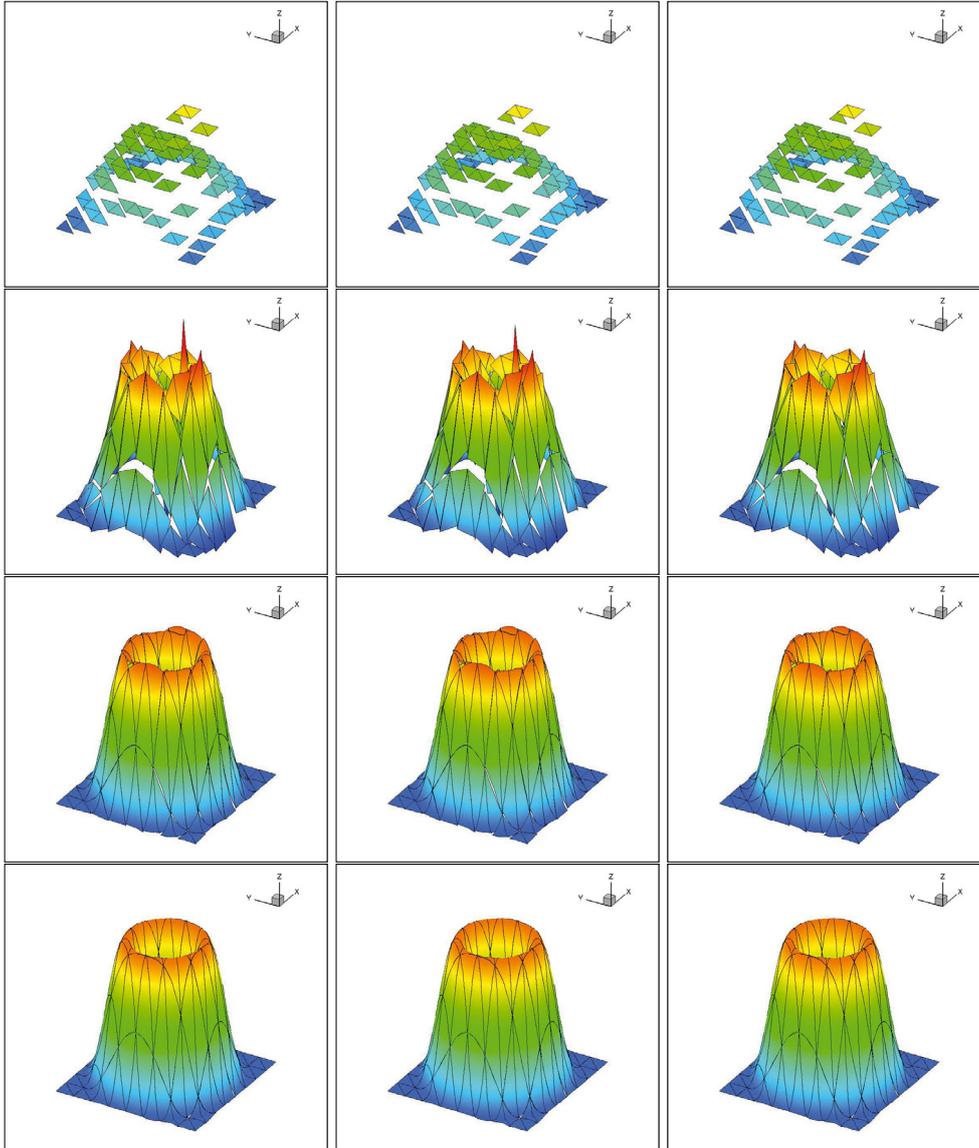


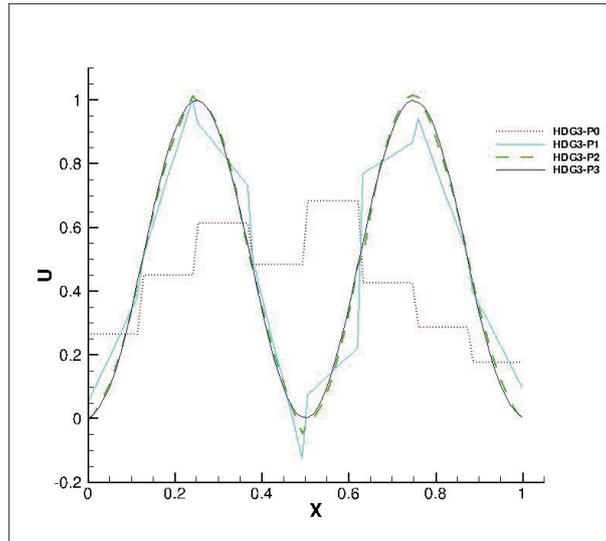
FIGURE 1. 3D plot of  $u_h$  for rotating flow test with  $\epsilon = 10^{-6}$  in 128 elements. Left-right:  $HDG1$ ,  $HDG2$ ,  $HDG3$ . Top-bottom:  $P_0$ - $P_3$ .

extensively smear out the interior layer, while, on the other hand, higher order approximations capture the interior layer within a few elements but produce oscillations within the layer.

#### 5.4. A boundary layer test

Finally, we take  $\beta = [1, 1]^T$ , and choose the source term  $f$  so that the exact solution

$$u(x, y) = \sin \frac{\pi x}{2} + \sin \frac{\pi y}{2} \left( 1 - \sin \frac{\pi x}{2} \right) + \frac{e^{-1/\epsilon} - e^{-(1-x)(1-y)/\epsilon}}{1 - e^{-1/\epsilon}}.$$

FIGURE 2.  $u_h$  along  $y = 1/2$  for  $P_k$ -HDG3 with 128 elements.TABLE 4. History of convergence of HDG1 for  $\|u - u_h\|_{L^2(\tilde{\Omega})}$  when  $\epsilon = 10^{-2}$  and  $\epsilon = 10^{-6}$ .

Degree Mesh		$\epsilon = 10^{-2}$		$\epsilon = 10^{-6}$	
$k$	$h^{-1}$	error	order	error	order
0	10	3.61e-2	--	3.32e-2	--
	20	1.81e-2	0.99	1.67e-2	1.00
	40	9.06e-3	1.00	8.34e-3	1.00
	80	4.52e-3	1.00	4.17e-3	1.00
1	10	4.22e-3	--	1.20e-3	--
	20	8.54e-4	2.30	3.00e-4	2.00
	40	2.13e-4	2.00	7.51e-5	2.00
	80	5.30e-5	2.01	1.88e-5	2.00
2	10	1.48e-3	--	1.90e-5	--
	20	6.66e-5	4.47	2.37e-6	3.00
	40	8.19e-6	3.02	2.96e-7	3.00
	80	1.03e-6	3.00	3.70e-8	3.00
3	10	4.10e-4	--	3.17e-7	--
	20	5.35e-6	6.26	1.99e-8	3.99
	40	3.56e-7	3.91	1.25e-9	4.00
	80	2.27e-8	3.97	7.79e-11	4.00

The solution develops boundary layers along the top and right boundaries for small  $\epsilon$  (see Fig. 7 for  $\epsilon = 10^{-2}$  and Fig. 8 for  $\epsilon = 10^{-6}$ ). We take an exact solution which is a slight modification of that considered in [1] so that, away from the boundary layers, our exact solution behaves not like a quadratic polynomial as in [1]. This modification is useful for us to clearly see the orders of convergence for  $k = 2, 3$ .

In Figures 7 and 8, we plot the exact solution and computational results for  $\epsilon = 10^{-2}$  and  $\epsilon = 10^{-6}$  in a structured 200 elements. We find that all the HDG methods produce similar results. The boundary layers are not resolved since the mesh is too coarse.

In Table 4, we show the convergence of  $u_h$  in  $L^2$ -norm for  $\epsilon = 10^{-2}, 10^{-6}$  in the reduced domain  $\tilde{\Omega} = [0, 0.9] \times [0, 0.9] \subset \Omega$  to exclude the unresolved boundary layers. Just as in the smooth case, the three HDG

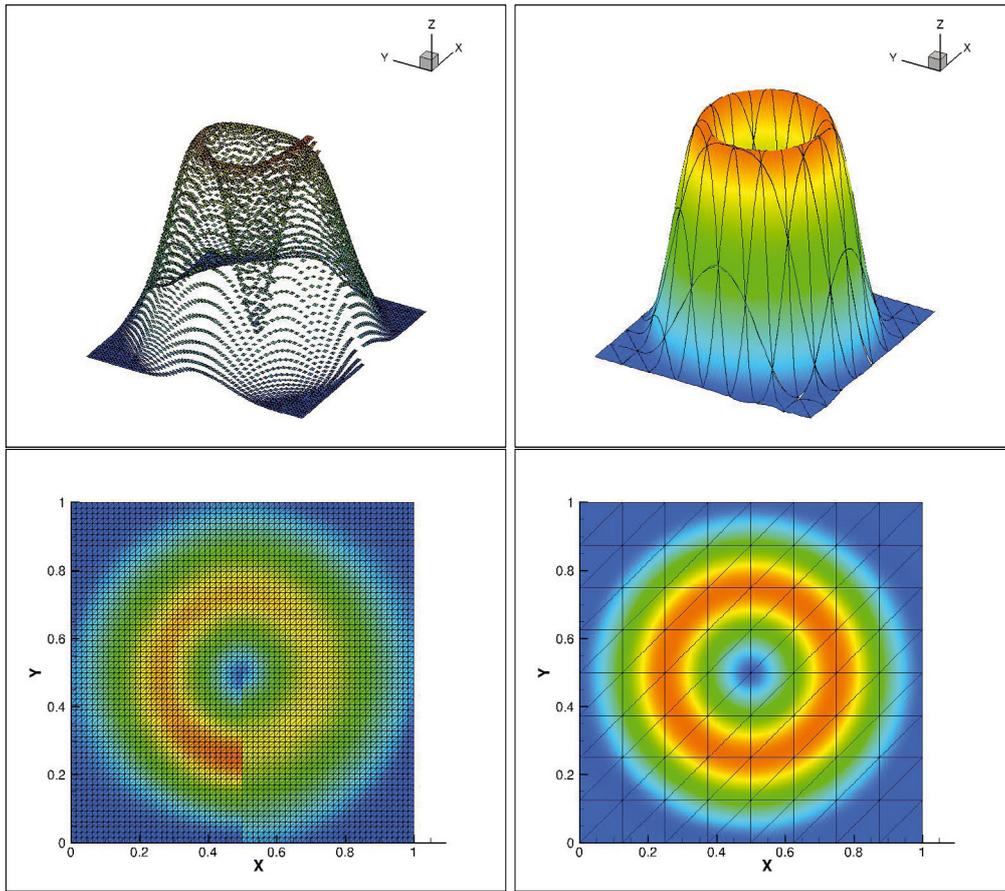


FIGURE 3. A comparison of  $P_0$ -HDG and  $P_3$ -HDG. Left:  $P_0$ -HDG1 with 8192 elements; right  $P_3$ -HDG1 with 128 elements. Top: 3D plot; bottom: 2D contour.

methods produce very similar convergence results. Hence we only show the computed results for  $P_k$ -HDG1 in Table 4. We observe optimal  $L^2$ -convergence rates for  $u_h$ .

### 5.5. The condition number

Now, we present the condition number of the matrix generated by the original bilinear form  $a_h$  in (3.2) and the scaled bilinear form  $\tilde{a}_h$  in (3.5). We use the same setup as that of the smooth test with two choices of  $\beta$ . The first choice of  $\beta$  is  $\beta = [1, 2]^T$ . For this choice, assumption (3.3) is satisfied by both example of  $\tau$  in (2.10) and (2.11). Since the condition numbers of all three HDG methods are very similar in our tests, we only present that for  $P_k$ -HDG1 in Table 5. Notice that  $\mathcal{O}(h^{-2})$  is observed for  $\epsilon = 1, 10^{-3}, 10^{-9}$  for different polynomial degrees, and that the condition number of the scaled system is similar to that of unscaled system. The second choice of  $\beta$  is  $\beta = [1, 1]^T$ . For this choice, assumption (3.3) is satisfied by the second choice of  $\tau$  in (2.11), but not for the choice (2.10) since the mesh is aligned with  $\beta$ . However, one can easily modify  $\tau$  in  $P_k$ -HDG1 and  $P_k$ -HDG3 on the aligned faces so that (3.3) holds. We only present the condition numbers for  $P_k$ -HDG2 in Table 6. The dependence of the condition number on  $\epsilon$  seems to be of order  $\mathcal{O}(\epsilon^{-1})$  for the original system, and we observe a huge improvement of the condition number after scaling for  $\epsilon = 10^{-9}$ . Also, we find that the condition number for the scaled system is of order  $\mathcal{O}(h^{-2})$  for  $\epsilon = 1$ , and of order  $\mathcal{O}(h^{-1})$  for  $\epsilon = 10^{-3}, 10^{-9}$ , which is not predicted by our theory.

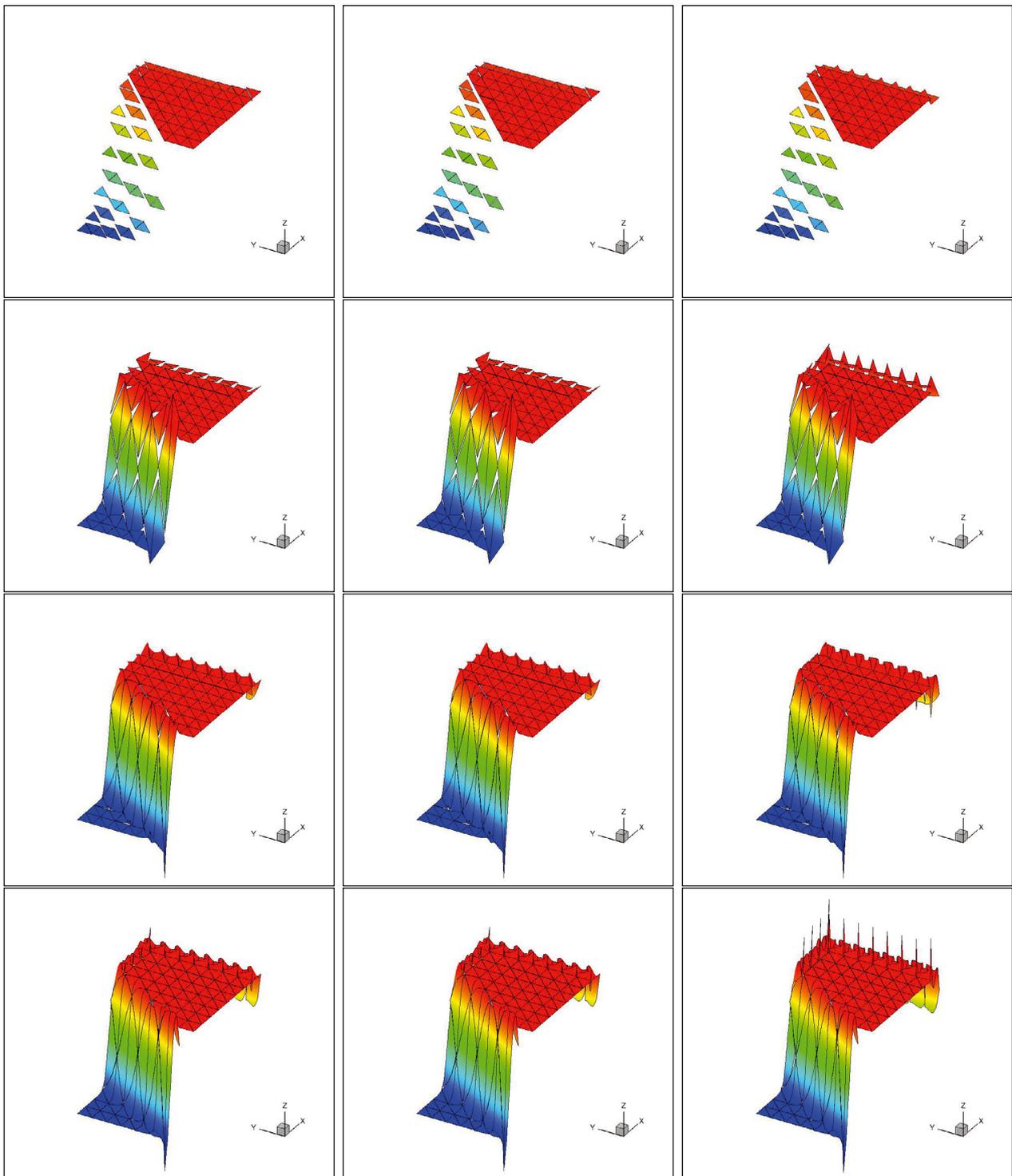


FIGURE 4. 3D plot of  $u_h$  for the interior layer test with  $\epsilon = 10^{-3}$  in 128 elements. Left-right:  $HDG1$ ,  $HDG2$ ,  $HDG3$ . Top-bottom:  $P_0$ - $P_3$ .

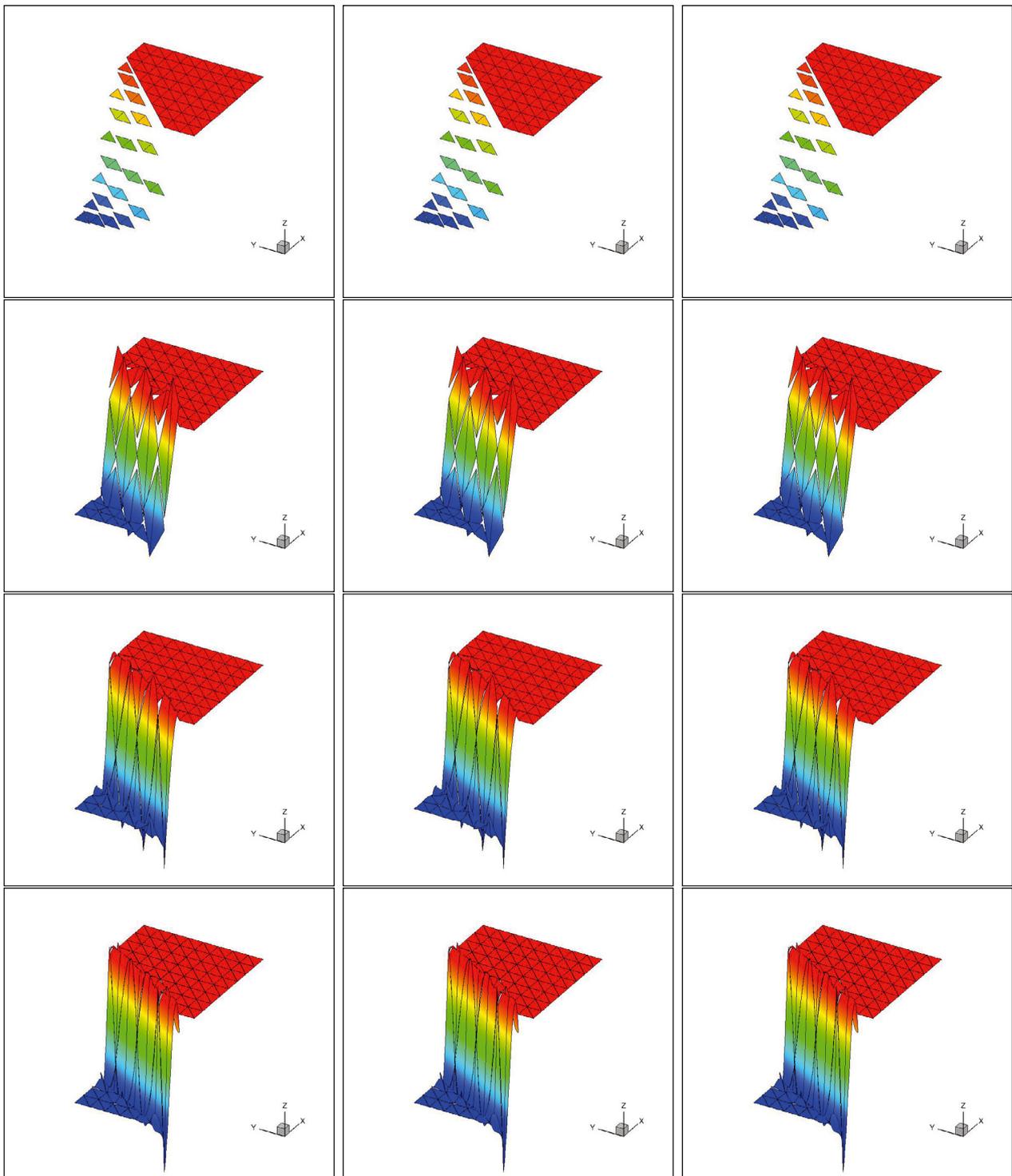


FIGURE 5. 3D plot of  $u_h$  for the interior layer test with  $\epsilon = 10^{-9}$  in 128 elements. Left-right:  $HDG1$ ,  $HDG2$ ,  $HDG3$ . Top-bottom:  $P_0$ - $P_3$ .

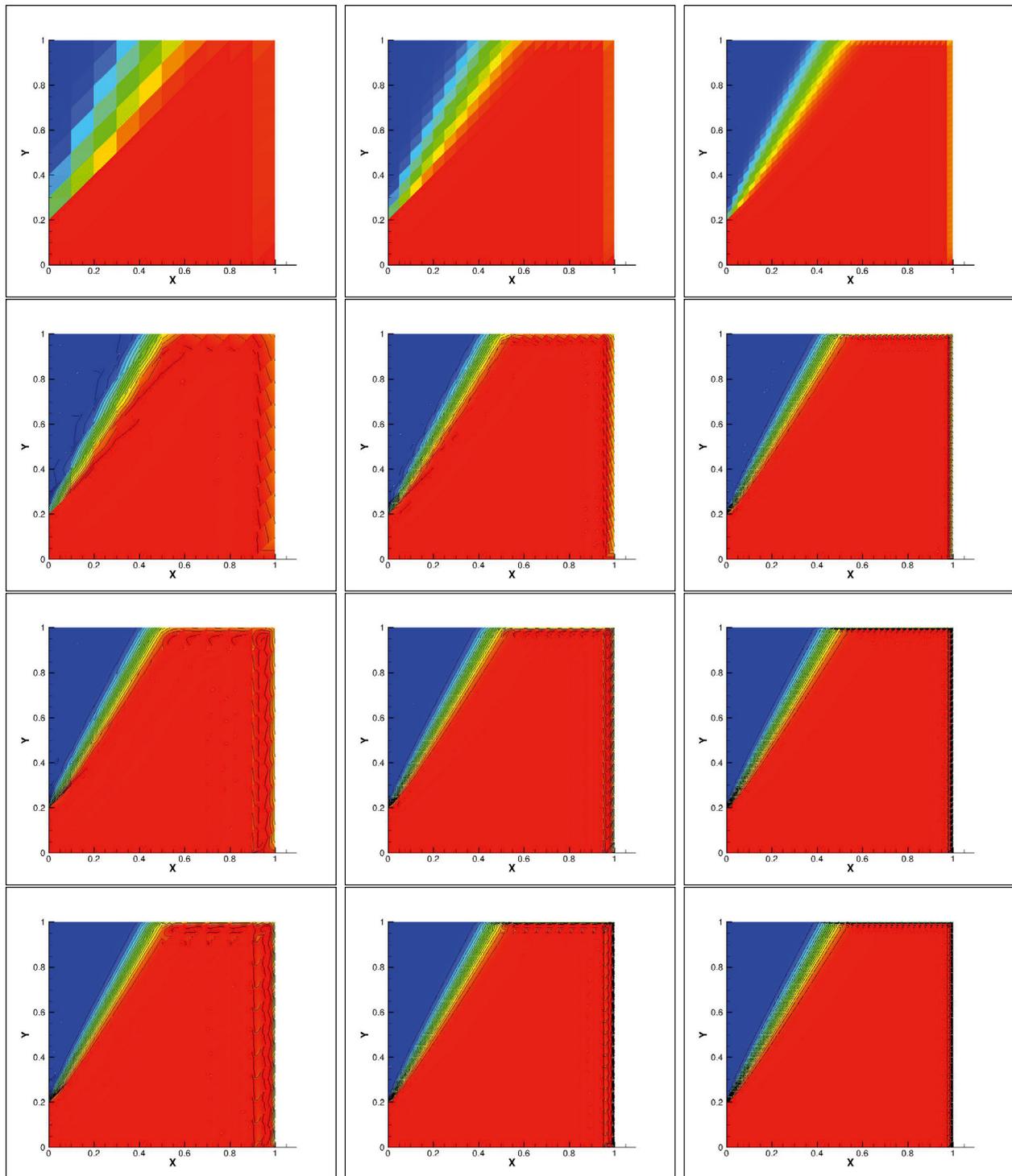


FIGURE 6. Contour plot of  $u_h$  using  $HDG1$  for the interior layer test with  $\epsilon = 10^{-3}$ . Left to right: three consecutive meshes with the left one consists of a uniform triangulation of 200 elements. Top-bottom:  $P_0$ - $P_3$ .

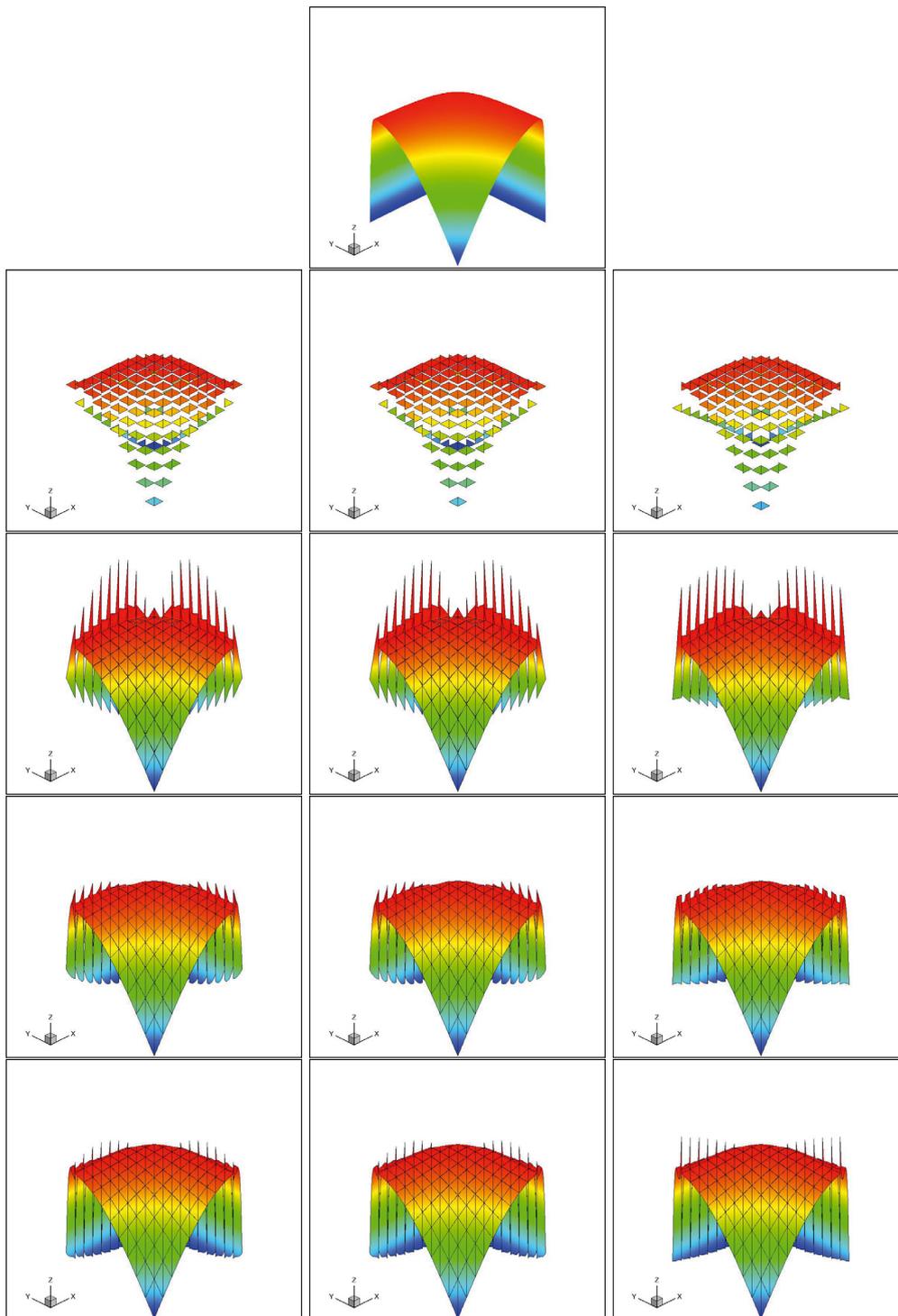


FIGURE 7. 3D plot of the exact solution and  $u_h$  for the boundary layer test with  $\epsilon = 10^{-2}$  in 128 elements. Top center: the exact solution. Left-right:  $HDG1$ ,  $HDG2$ ,  $HDG3$ . Top-bottom:  $P_0$ - $P_3$ .

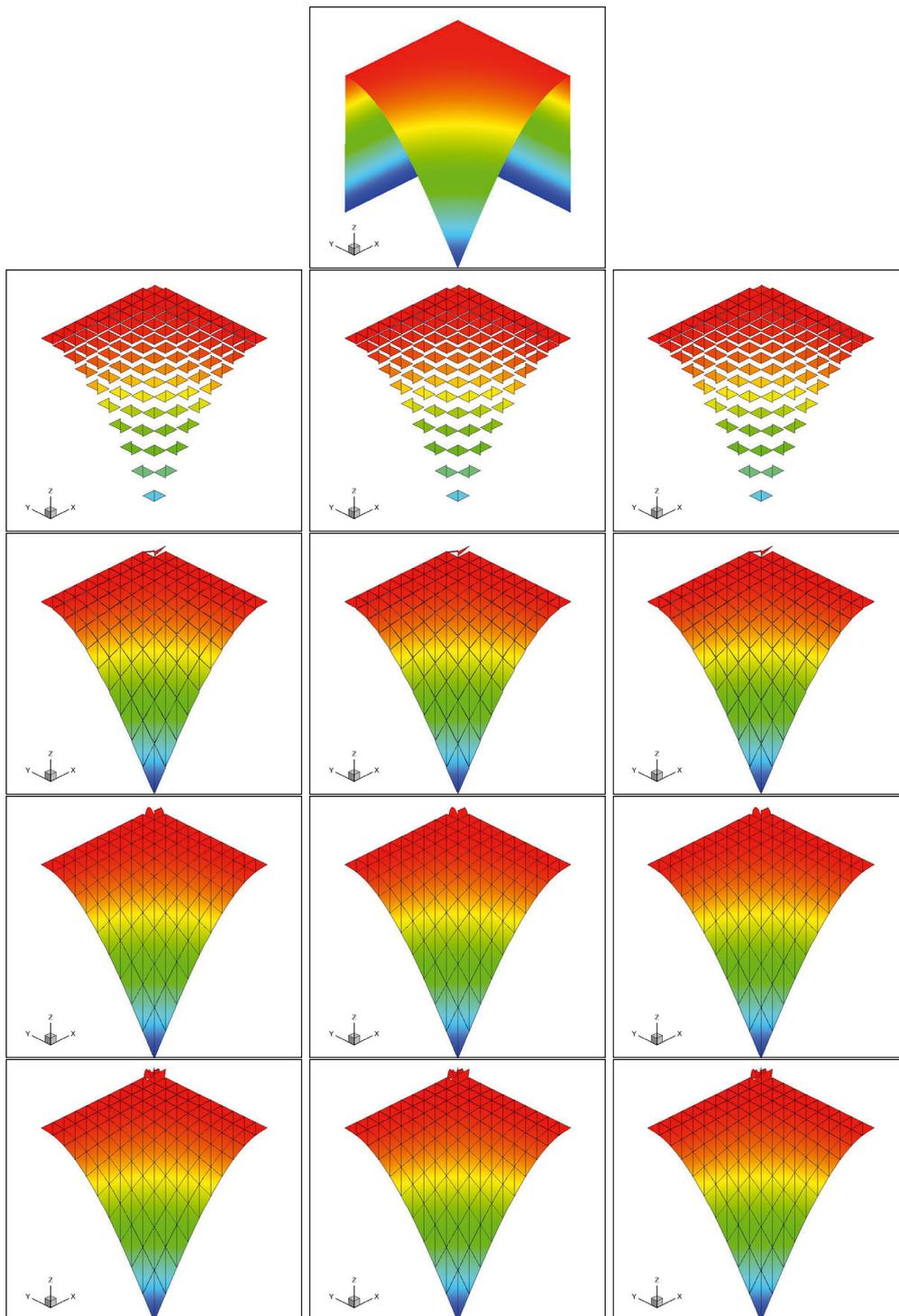


FIGURE 8. 3D plot of the exact solution and  $u_h$  for the boundary layer test with  $\epsilon = 10^{-6}$  in 128 elements. Top center: the exact solution. Left-right:  $HDG1$ ,  $HDG2$ ,  $HDG3$ . Top-bottom:  $P_0$ - $P_3$ .

TABLE 5. Condition numbers for  $HDG1$  when  $\beta = [1, 2]^T$ .

$\epsilon$	Mesh $h^{-1}$	Condition numbers for $a_h(\cdot, \cdot)$				Condition numbers for $\tilde{a}_h(\cdot, \cdot)$			
		k=0	k=1	k=2	k=3	k=0	k=1	k=2	k=3
1e-0	5	1.11e+2	1.42e+2	2.82e+2	3.31e+2	2.49e+2	2.65e+2	5.30e+2	6.29e+2
	10	3.37e+2	5.28e+2	1.06e+3	1.24e+2	6.29e+2	9.83e+2	1.98e+3	2.34e+2
	20	1.37e+3	2.09e+3	4.17e+3	4.89e+2	2.55e+3	3.89e+3	7.78e+3	9.21e+2
	40	5.50e+3	8.32e+3	1.66e+4	1.95e+2	1.03e+4	1.55e+4	3.10e+4	3.67e+2
1e-3	5	4.86e+1	1.60e+2	3.64e+2	5.00e+2	3.71e+1	1.12e+2	2.44e+2	3.40e+2
	10	1.74e+2	5.23e+2	8.79e+2	1.27e+3	1.36e+2	3.75e+2	6.51e+2	8.48e+2
	20	6.48e+2	1.82e+3	2.35e+3	4.06e+3	5.06e+2	1.25e+3	1.17e+3	2.50e+3
	40	2.48e+3	6.94e+3	8.76e+3	1.49e+4	1.94e+3	4.44e+3	5.98e+3	8.96e+3
1e-9	5	4.90e+1	1.69e+2	4.21e+2	5.76e+2	3.73e+1	9.57e+1	2.97e+2	3.80e+2
	10	1.74e+2	5.38e+2	1.17e+3	1.52e+3	1.36e+2	4.04e+2	9.11e+2	1.05e+3
	20	6.49e+2	1.94e+3	3.76e+3	4.73e+3	5.05e+2	1.46e+3	2.94e+3	3.40e+3
	40	2.50e+3	7.06e+3	1.34e+4	1.65e+4	1.93e+3	5.37e+3	9.45e+3	1.25e+4

TABLE 6. Condition numbers for  $HDG2$  when  $\beta = [1, 1]^T$ .

$\epsilon$	Mesh $h^{-1}$	Condition numbers for $a_h(\cdot, \cdot)$				Condition numbers for $\tilde{a}_h(\cdot, \cdot)$			
		k=0	k=1	k=2	k=3	k=0	k=1	k=2	k=3
1e-0	5	9.99e+1	1.41e+2	2.82e+2	3.31e+2	1.16e+2	1.96e+2	3.92e+2	4.64e+2
	10	3.40e+2	5.34e+2	1.07e+3	1.25e+2	4.35e+2	6.82e+2	1.37e+3	1.61e+2
	20	1.38e+3	2.12e+3	4.23e+3	4.96e+2	1.77e+3	2.71e+3	5.41e+3	6.38e+2
	40	5.57e+3	8.44e+3	1.68e+4	1.98e+2	7.12e+3	1.08e+4	2.15e+4	2.54e+2
1e-3	5	2.01e+2	6.63e+2	9.06e+2	1.55e+3	2.48e+2	1.81e+3	7.02e+3	9.83e+3
	10	4.00e+2	1.41e+3	1.62e+3	3.27e+3	5.63e+2	2.34e+3	7.45e+3	1.16e+4
	20	1.25e+3	4.42e+3	4.98e+3	9.57e+3	1.20e+3	3.99e+3	9.01e+3	1.66e+4
	40	4.69e+3	1.63e+4	1.86e+4	3.12e+4	2.60e+3	7.29e+3	1.32e+4	2.49e+4
1e-9	5	1.43e+8	4.02e+8	6.04e+8	7.34e+8	2.38e+2	2.07e+3	1.23e+4	2.49e+4
	10	1.31e+8	3.84e+8	5.68e+8	7.03e+8	5.35e+2	4.66e+3	2.77e+4	5.60e+4
	20	1.25e+8	3.75e+8	5.51e+8	6.87e+8	1.13e+3	9.84e+3	5.85e+4	1.12e+5
	40	1.22e+8	3.70e+8	5.42e+8	6.79e+8	2.32e+3	2.02e+4	1.20e+5	2.30e+5

### APPENDIX A. THE RELATION BETWEEN THE HDG METHOD AND THE MH-DG METHOD IN [26]

In this section, we establish the relation between the HDG method (2.4) and the MH-DG considered in [26]. We first present the MH-DG method for equations (1.1) under the condition that  $g = 0$  and  $\beta \in H(\text{div}; \Omega)$  is constant in each element. Then, we show that this method coincides with the HDG method (2.4) when using the same approximation spaces and choosing the stability function  $\tau$  to be (2.10).

The MH-DG method seeks an approximation  $(\mathbf{q}_h, u_h, \lambda_h) \in \tilde{\mathbf{V}}_h \times W_h \times M_h(0)$  so that

$$B_h((\mathbf{q}_h, u_h, \lambda_h), (\mathbf{r}, w, \mu)) = (f, w)_{\mathcal{T}_h}, \tag{A.1}$$

for all  $(\mathbf{r}, w, \mu) \in \tilde{\mathbf{V}}_h \times W_h \times M_h(0)$ , where  $W_h$  and  $M_h(0)$  is defined in (2.3) and  $\tilde{\mathbf{V}}_h$  is the so called Raviart–Thomas space, slightly larger than  $\mathbf{V}_h$ , defined as follows

$$\tilde{\mathbf{V}}_h = \{ \mathbf{r} \in L^2(\Omega; \mathbb{R}^d) : \mathbf{r}|_K \in P_k(K; \mathbb{R}^d) + \mathbf{x}P_k(K) \quad \forall K \in \mathcal{T}_h \},$$

and

$$\begin{aligned} B_h((\mathbf{q}, u, \lambda), (\mathbf{r}, w, \mu)) &= (\epsilon^{-1} \mathbf{q}, \mathbf{r})_{\mathcal{T}_h} - (u, \nabla \cdot \mathbf{r})_{\mathcal{T}_h} + \langle \lambda, \mathbf{r} \cdot \mathbf{n} \rangle_{\partial \mathcal{T}_h} \\ &\quad - (\mathbf{q} + \beta u, \nabla w)_{\mathcal{T}_h} + \langle \mathbf{q} \cdot \mathbf{n} + \beta \cdot \mathbf{n} \{ \lambda / u \}, w - \mu \rangle_{\partial \mathcal{T}_h}, \end{aligned}$$

where

$$\{ \lambda / u \} := \begin{cases} \lambda, & \text{if } \beta \cdot \mathbf{n} < 0, \\ u, & \text{if } \beta \cdot \mathbf{n} \geq 0. \end{cases}$$

Comparing the bilinear form for the HDG method (4.1) with approximation spaces  $\tilde{\mathbf{V}}_h \times W_h \times M_h(0)$  and the stability function  $\tau$  in (2.10) and that for the MH-DG method in (A.1), we notice that the only difference lies in the definition of the numerical flux  $(\hat{\mathbf{q}}_h + \widehat{\beta u}_h) \cdot \mathbf{n}$  and  $\mathbf{q}_h \cdot \mathbf{n} + \beta \cdot \mathbf{n} \{ \lambda / u \}$ . However, by the following simple calculation, we observe that the two numerical flux are actually the same:

$$\begin{aligned} (\hat{\mathbf{q}}_h + \widehat{\beta u}_h) \cdot \mathbf{n} &= \mathbf{q}_h \cdot \mathbf{n} + \beta \cdot \mathbf{n} \lambda_h + \tau(u_h - \lambda_h) \\ &= \mathbf{q}_h \cdot \mathbf{n} + \beta \cdot \mathbf{n} \lambda_h + \max(\beta \cdot \mathbf{n}, 0)(u_h - \lambda_h) \\ &= \begin{cases} \mathbf{q}_h \cdot \mathbf{n} + \beta \cdot \mathbf{n} \lambda_h, & \text{if } \beta \cdot \mathbf{n} < 0 \\ \mathbf{q}_h \cdot \mathbf{n} + \beta \cdot \mathbf{n} u_h, & \text{if } \beta \cdot \mathbf{n} \geq 0 \end{cases} \\ &= \mathbf{q}_h \cdot \mathbf{n} + \beta \cdot \mathbf{n} \{ \lambda_h / u_h \}. \end{aligned}$$

Hence, these two methods coincide.

## APPENDIX B. CONDITIONING OF THE HDG METHODS

In this section, we give a proof of Theorem 3.2. Again, the key idea is to recover an estimate of the  $L^2$ -norm of  $u_h$ , see the estimate (B.5) below. By similar argument in the proof of Lemma 4.4, we have the following local energy estimate:

**Lemma B.1.** *If  $\epsilon \leq \mathcal{O}(h)$ , then there is  $h_0 > 0$ , which is independent of  $\epsilon$  and  $h$ , such that for any  $\lambda \in M_h(0)$  and  $K \in \mathcal{T}_h$ ,*

$$\epsilon^{-1/2} \|\mathbf{q}_h^\lambda\|_K + \|u_h^\lambda\|_K + \left\| \left| \tau - \frac{1}{2} \beta \cdot \mathbf{n} \right|^{1/2} u_h^\lambda \right\|_{\partial K} \leq C \left\| \left| \tau - \frac{1}{2} \beta \cdot \mathbf{n} \right|^{1/2} \lambda \right\|_{\partial K}, \quad (\text{B.1})$$

if  $h < h_0$ .

For all  $(\boldsymbol{\sigma}, v, \lambda), (\mathbf{r}, w, \mu) \in H^1(\mathcal{T}_h; \mathbb{R}^d) \times H^1(\mathcal{T}_h) \times L^2(\mathcal{E}_h)$ , we define

$$\begin{aligned} b_h((\boldsymbol{\sigma}, v, \lambda), (\mathbf{r}, w, \mu)) &= (\epsilon^{-1} \boldsymbol{\sigma}, \mathbf{r})_{\mathcal{T}_h} + \langle \tau(v - \lambda), w - \mu \rangle_{\partial \mathcal{T}_h} \\ &\quad - (\beta v, \nabla w)_{\mathcal{T}_h} + \langle (\beta \cdot \mathbf{n}) \lambda, w - \mu \rangle_{\partial \mathcal{T}_h} - ((\nabla \cdot \beta) v, w)_{\mathcal{T}_h}, \end{aligned}$$

The next result is similar to Lemma 4.4.

**Lemma B.2.** *If  $\epsilon \leq \mathcal{O}(h)$ , then there is  $h_0 > 0$ , which is independent of  $\epsilon$  and  $h$ , such that for any  $\lambda \in M_h(0)$ ,*

$$\epsilon^{-1} \|\mathbf{q}_h^\lambda\|_{\mathcal{T}_h}^2 + \|u_h^\lambda\|_{\mathcal{T}_h}^2 + \left\| \left| \tau - \frac{1}{2} \beta \cdot \mathbf{n} \right|^{1/2} (u_h^\lambda - \lambda) \right\|_{\partial \mathcal{T}_h}^2 \leq C b_h((\mathbf{q}_h^\lambda, u_h^\lambda, \lambda), (\mathbf{q}_h^{(P_{0,M}\varphi)\lambda}, u_h^{(P_{0,M}\varphi)\lambda}, (P_{0,M}\varphi)\lambda)),$$

if  $h < h_0$ . Here, the weight function  $\varphi = e^{-\psi} + \chi$  is introduced in (4.4).  $P_{0,M}$  is the  $L^2$ -orthogonal projection onto  $P_0(\mathcal{E}_h)$ .

**Remark B.3.** Notice that in general, the space  $\{(\mathbf{q}_h^m, u_h^m, m) : m \in M_h(0)\}$  is a non-trivial subspace of  $\mathbf{V}_h \times W_h \times M_h(0)$ . Then, given  $m \in M_h$ ,  $(\mathbf{II}_h(\varphi \mathbf{q}_h^m), P_h(\varphi u_h^m), P_M(\varphi m))$  is *not* necessarily contained in  $\{(\mathbf{q}_h^m, u_h^m, m) : m \in M_h\}$ . So, the proof of Lemma B.2 can *not* be derived from the stability of HDG methods in Lemma 4.4.

*Proof.* We accomplish the proof in the following steps.

(I) By the same argument in the proof of Lemma 4.1, if  $h$  is small enough (independent of  $\epsilon$ ), then

$$\epsilon^{-1} \chi \|\mathbf{q}_h^\lambda\|_{\mathcal{T}_h}^2 + \|u_h^\lambda\|_{\mathcal{T}_h}^2 + \chi \left\| \left| \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right|^{1/2} (u_h^\lambda - \lambda) \right\|_{\partial \mathcal{T}_h}^2 \leq C b_h((\mathbf{q}_h^\lambda, u_h^\lambda, \lambda), (\varphi \mathbf{q}_h^\lambda, \varphi u_h^\lambda, \varphi \lambda)).$$

(II) By similar argument in the proof of Lemma 4.4, if we choose  $\chi$  big enough and  $h$  small enough (both are independent of  $\epsilon$ ), then

$$\epsilon^{-1} \chi \|\mathbf{q}_h^\lambda\|_{\mathcal{T}_h}^2 + \|u_h^\lambda\|_{\mathcal{T}_h}^2 + \chi \left\| \left| \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right|^{1/2} (u_h^\lambda - \lambda) \right\|_{\partial \mathcal{T}_h}^2 \leq C b_h((\mathbf{q}_h^\lambda, u_h^\lambda, \lambda), ((P_{0,h} \varphi) \mathbf{q}_h^\lambda, P_h(\varphi u_h^\lambda), (P_{0,M} \varphi) \lambda)).$$

Here, we define  $P_{0,h} \varphi$  to be the average of  $\varphi$  on every element  $K \in \mathcal{T}_h$ .

(III) Now, we want to bound  $b_h((\mathbf{q}_h^\lambda, u_h^\lambda, \lambda), ((P_{0,h} \varphi) \mathbf{q}_h^\lambda, (P_{0,h} \varphi) u_h^\lambda, (P_{0,M} \varphi) \lambda))$  from below. Notice that

$$\begin{aligned} b_h((\mathbf{q}_h^\lambda, u_h^\lambda, \lambda), ((P_{0,h} \varphi) \mathbf{q}_h^\lambda, (P_{0,h} \varphi) u_h^\lambda, (P_{0,M} \varphi) \lambda)) &= b_h((\mathbf{q}_h^\lambda, u_h^\lambda, \lambda), ((P_{0,h} \varphi) \mathbf{q}_h^\lambda, P_h(\varphi u_h^\lambda), (P_{0,M} \varphi) \lambda)) \\ &\quad + (\boldsymbol{\beta} \cdot \nabla u_h^\lambda, P_h(\varphi u_h^\lambda) - (P_{0,h} \varphi) u_h^\lambda)_{\mathcal{T}_h} \\ &\quad + \langle (\tau - \boldsymbol{\beta} \cdot \mathbf{n})(u_h^\lambda - \lambda), P_h(\varphi u_h^\lambda) - (P_{0,h} \varphi) u_h^\lambda \rangle_{\partial \mathcal{T}_h}. \end{aligned}$$

By (3.1b), we have

$$\begin{aligned} b_h((\mathbf{q}_h^\lambda, u_h^\lambda, \lambda), ((P_{0,h} \varphi) \mathbf{q}_h^\lambda, (P_{0,h} \varphi) u_h^\lambda, (P_{0,M} \varphi) \lambda)) &= b_h((\mathbf{q}_h^\lambda, u_h^\lambda, \lambda), ((P_{0,h} \varphi) \mathbf{q}_h^\lambda, P_h(\varphi u_h^\lambda), (P_{0,M} \varphi) \lambda)) \\ &\quad - (\nabla \cdot \mathbf{q}_h^\lambda, P_h(\varphi u_h^\lambda) - (P_{0,h} \varphi) u_h^\lambda)_{\mathcal{T}_h}. \end{aligned} \quad (\text{B.2})$$

By inverse inequality to  $\nabla \cdot \mathbf{q}_h^\lambda$  and assumption  $\epsilon \leq \mathcal{O}(h)$ , we have

$$\|\nabla \cdot \mathbf{q}_h^\lambda\|_{\mathcal{T}_h} \leq C \epsilon^{-1/2} h^{-1/2} \|\mathbf{q}_h^\lambda\|_{\mathcal{T}_h}.$$

In addition, we have

$$\|P_h(\varphi u_h^\lambda) - (P_{0,h} \varphi) u_h^\lambda\|_{\mathcal{T}_h} \leq Ch \|u_h^\lambda\|_{\mathcal{T}_h}.$$

So, if  $h$  is small enough (independent of  $\epsilon, \chi$ ), we have

$$\begin{aligned} &\chi \epsilon^{-1} \|\mathbf{q}_h^\lambda\|_{\mathcal{T}_h}^2 + \|u_h^\lambda\|_{\mathcal{T}_h}^2 + \chi \left\| \left| \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right|^{1/2} (u_h^\lambda - \lambda) \right\|_{\partial \mathcal{T}_h}^2 \\ &\leq C b_h((\mathbf{q}_h^\lambda, u_h^\lambda, \lambda), ((P_{0,h} \varphi) \mathbf{q}_h^\lambda, (P_{0,h} \varphi) u_h^\lambda, (P_{0,M} \varphi) \lambda)). \end{aligned} \quad (\text{B.3})$$

(IV) Now, we want to bound  $b_h((\mathbf{q}_h^\lambda, u_h^\lambda, \lambda), ((P_{0,h} \varphi) \mathbf{q}_h^\lambda, u_h^{(P_{0,M} \varphi) \lambda}, (P_{0,M} \varphi) \lambda))$  from below. Similar to (B.2), we have

$$\begin{aligned} b_h((\mathbf{q}_h^\lambda, u_h^\lambda, \lambda), ((P_{0,h} \varphi) \mathbf{q}_h^\lambda, u_h^{(P_{0,M} \varphi) \lambda}, (P_{0,M} \varphi) \lambda)) &= b_h((\mathbf{q}_h^\lambda, u_h^\lambda, \lambda), ((P_{0,h} \varphi) \mathbf{q}_h^\lambda, (P_{0,h} \varphi) u_h^\lambda, (P_{0,M} \varphi) \lambda)) \\ &\quad - (\nabla \cdot \mathbf{q}_h^\lambda, (P_{0,h} \varphi) u_h^\lambda - u_h^{(P_{0,M} \varphi) \lambda})_{\mathcal{T}_h}. \end{aligned} \quad (\text{B.4})$$

Since  $\lambda \rightarrow u_h^\lambda$  is linear and (B.1), then for any  $K \in \mathcal{T}_h$ ,

$$\begin{aligned} \|(P_{0,h}\varphi)u_h^\lambda - u_h^{(P_{0,M}\varphi)\lambda}\|_K &= \|u_h^{(P_{0,h}\varphi)\lambda} - u_h^{(P_{0,M}\varphi)\lambda}\|_K \\ &\leq C \left\| \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right\|^{1/2} \|(P_{0,h}\varphi)\lambda - (P_{0,M}\varphi)\lambda\|_{\partial K} \\ &\leq Ch \left\| \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right\|^{1/2} \lambda \|_{\partial K} \\ &\leq Ch^{1/2} \left( \|u_h^\lambda\|_K^2 + \Sigma_{F \in \mathcal{E}(K)} \left\| \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right\|^{1/2} (u_h^\lambda - \lambda) \right\|_F^2 \right)^{1/2}. \end{aligned}$$

Recall that by an inverse inequality to  $\nabla \cdot \mathbf{q}_h^\lambda$  and assumption  $\epsilon \leq \mathcal{O}(h)$ ,

$$\|\nabla \cdot \mathbf{q}_h^\lambda\|_{\mathcal{T}_h} \leq C \epsilon^{-1/2} h^{-1/2} \|\mathbf{q}_h^\lambda\|_{\mathcal{T}_h}.$$

So, if  $\chi$  is big enough (independent of  $\epsilon, h$ ), we have

$$\epsilon^{-1} \|\mathbf{q}_h^\lambda\|_{\mathcal{T}_h}^2 + \|u_h^\lambda\|_{\mathcal{T}_h}^2 + \left\| \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right\|^{1/2} (u_h^\lambda - \lambda) \Big\|_{\partial \mathcal{T}_h}^2 \leq C b_h((\mathbf{q}_h^\lambda, u_h^\lambda, \lambda), ((P_{0,h}\varphi)\mathbf{q}_h^\lambda, u_h^{(P_{0,M}\varphi)\lambda}, (P_{0,M}\varphi)\lambda)).$$

(V) By (B.1), (B.3), (3.3) and the fact that  $\lambda \rightarrow \mathbf{q}_h^\lambda$  is linear, we have

$$\epsilon^{-1} \|\mathbf{q}_h^\lambda\|_{\mathcal{T}_h}^2 + \|u_h^\lambda\|_{\mathcal{T}_h}^2 + \left\| \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right\|^{1/2} (u_h^\lambda - \lambda) \Big\|_{\partial \mathcal{T}_h}^2 \leq C b_h((\mathbf{q}_h^\lambda, u_h^\lambda, \lambda), (\mathbf{q}_h^{(P_{0,M}\varphi)\lambda}, u_h^{(P_{0,M}\varphi)\lambda}, (P_{0,M}\varphi)\lambda))$$

if  $h$  is small enough (independent of  $\epsilon$ ).

So, we can conclude the proof is complete.  $\square$

Now, we are ready to prove Theorem 3.2. By the definition of  $\tilde{a}_h$  in (3.5),

$$\begin{aligned} \tilde{a}_h(\tilde{\lambda}, \tilde{\mu}) &= a_h \left( \Lambda_\epsilon^{-1} \tilde{\lambda}, \Lambda_\epsilon^{-1} \tilde{\mu} \right) \\ &= b_h((\mathbf{q}_h^{\Lambda_\epsilon^{-1} \tilde{\lambda}}, u_h^{\Lambda_\epsilon^{-1} \tilde{\lambda}}, \Lambda_\epsilon^{-1} \tilde{\lambda}), (\mathbf{q}_h^{\Lambda_\epsilon^{-1} \tilde{\mu}}, u_h^{\Lambda_\epsilon^{-1} \tilde{\mu}}, \Lambda_\epsilon^{-1} \tilde{\mu})), \end{aligned}$$

for all  $\tilde{\lambda}, \tilde{\mu} \in M_h(0)$ .

We recall that  $\Lambda_\epsilon|_F = (\sup_{x \in F} |\boldsymbol{\beta} \cdot \mathbf{n}(x)| + \min(\frac{\epsilon}{h_F}, 1))^{1/2}$ ,  $\forall F \in \mathcal{E}_h$  in (3.4). By assumption (3.3) and Lemma B.2, we have that for any  $\tilde{\lambda} \in M_h(0)$ ,

$$\begin{aligned} \tilde{a}_h(\tilde{\lambda}, (P_{0,M}\varphi)\tilde{\lambda}) &\geq C \left( \|u_h^{\Lambda_\epsilon^{-1} \tilde{\lambda}}\|_{\mathcal{T}_h}^2 + \left\| \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right\|^{1/2} (u_h^{\Lambda_\epsilon^{-1} \tilde{\lambda}} - \Lambda_\epsilon^{-1} \tilde{\lambda}) \Big\|_{\partial \mathcal{T}_h}^2 \right) \\ &\geq Ch \left\| \tau - \frac{1}{2} \boldsymbol{\beta} \cdot \mathbf{n} \right\|^{1/2} \Lambda_\epsilon^{-1} \tilde{\lambda} \Big\|_{\partial \mathcal{T}_h}^2 \quad (\text{by trace inequality and triangle inequality}) \quad (\text{B.5}) \\ &\geq C \|\tilde{\lambda}\|_h^2 \geq C \|\tilde{\lambda}\|_h \cdot \|(P_{0,M}\varphi)\tilde{\lambda}\|_h, \end{aligned}$$

where

$$\|\tilde{\lambda}\|_h = h^{1/2} \|\tilde{\lambda}\|_{\mathcal{E}_h}, \quad \forall \tilde{\lambda} \in L^2(\mathcal{E}_h).$$

According to (B.1) and the definition of  $\Lambda_\epsilon$  in (3.4), we have

$$\tilde{a}_h(\tilde{\lambda}, \tilde{\mu}) \leq Ch^{-2} \|\tilde{\lambda}\|_h \cdot \|\tilde{\mu}\|_h, \quad \forall \tilde{\lambda}, \tilde{\mu} \in M_h(0). \quad (\text{B.6})$$

Using (B.5) and (B.6), we can conclude the proof of Theorem 3.2.

## APPENDIX C. GENERATING SPECIAL MESHES

As in [17], we do not intend to provide a detailed description how to generate meshes satisfying assumption (2.1). We just give the main idea to generate the triangulation in the following, which is similar to the idea in [17].

- (i) Given a positive value  $h$ , we triangulate the outflow boundary  $\Gamma^+ = \{x \in \partial\Omega : \boldsymbol{\beta} \cdot \mathbf{n}(x) > 0\}$  in segments of size no bigger than  $h$ .
- (ii) For each node  $x_0$  on  $\Gamma^+$ , we apply the forward Euler time-marching method to the problem

$$\frac{d}{dt}x(t) = -\boldsymbol{\beta}(x(t)) \quad t > 0, x(0) = x_0,$$

to obtain the set of nodes  $\{x_i\}_{i=1}^{N(x_0)}$  such that the distance between  $x_i$  and  $x_{i-1}$  is of order  $h$  and  $x_{N(x_0)}$  is the point on  $\partial\Omega \setminus \Gamma^+$ .

- (iii) We add the vertices of  $\partial\Omega \setminus \Gamma^+$  to the set of nodes. Then we generate a triangulation.
- (iv) We numerically check assumption (2.1) and modify the simplexes which violate the assumption by using an algorithm similar to that in [33].

*Acknowledgements.* The authors would like to thank Professor Bernardo Cockburn for constructive criticism leading to a better presentation of the material in this paper. The authors would also like to thank one of the referees for pointing out the paper by Egger and Schöberl [26] and for suggesting to comment on the relation of the HDG method to the MH-DG method in [26], as was done in Appendix A. The work of Weifeng Qiu was supported by City University of Hong Kong under start-up grant (No. 7200324) and by the GRF of Hong Kong (Grant No. 9041980).

## REFERENCES

- [1] B. Ayuso and D. Marini, Discontinuous Galerkin Methods for Advection-Diffusion-Reaction Problems. *SIAM J. Numer. Anal.* **47** (2009) 1391–1420.
- [2] C.E. Baumann and J.T. Oden, A discontinuous hp finite element method for convection-diffusion problems. *Comput. Methods Appl. Mech. Engrg.* **175** (1999) 311–341.
- [3] F. Brezzi, T.J.R. Hughes, L.D. Marini, A. Russo and E. Süli, A Priori Error Analysis of Residual-free Bubbles for Advection-Diffusion Problems. *SIAM J. Numer. Anal.* **36** (1999) 1933–1948.
- [4] F. Brezzi, L.D. Marini and E. Süli, Residual-free bubbles for advection-diffusion problems: the general error analysis. *Numer. Math.* **85** (2000) 31–47.
- [5] A.N. Brooks and T.J.R. Hughes, Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations, *Comput. Methods Appl. Mech. Engrg.* **32** (1982) 199–259.
- [6] A. Buffa, T.J.R. Hughes and G. Sangalli, Analysis of a multiscale discontinuous Galerkin method for convection-diffusion problems. *SIAM J. Numer. Anal.* **44** (2006) 1420–1440.
- [7] Erik Burman and Alexandre Ern, Stabilized Galerkin Approximation of Convection-Diffusion-Reaction equations: Discrete Maximum Principle and Convergence. *Math. Comput.* **74** (2005) 1637–1652.
- [8] P. Castillo, B. Cockburn, D. Schötzau and C. Schwab, An optimal a priori error estimate for the hp-version of the local discontinuous Galerkin method for convection-diffusion problems. *Math. Comput.* **71** (2002) 455–478.
- [9] Y. Chen and B. Cockburn, Analysis of variable-degree HDG methods for convection-diffusion equations. Part I: General nonconforming meshes. *IMA J. Num. Anal.* **32** (2012) 1267–1293.
- [10] Y. Chen and B. Cockburn, Analysis of variable-degree HDG methods for convection-diffusion equations. Part II: Semimatching nonconforming meshes. To appear in *Math. Comput.*
- [11] B. Cockburn, A Discontinuous Galerkin methods for convection-dominated problems, In High-order methods for computational physics, vol. 9 of *Lect. Notes Comput. Sci. Eng.* Springer, Berlin (1999) 69–224.
- [12] B. Cockburn and C. Dawson, Some extensions of the local discontinuous Galerkin method for convection-diffusion equations in multidimensions, In *The mathematics of finite elements and applications, X, MAFELAP 1999 (Uxbridge)*. Elsevier, Oxford (2000) 225–238.
- [13] B. Cockburn and B. Dong, An analysis of the minimal dissipation local discontinuous Galerkin method for convection-diffusion problems. *J. Sci. Comput.* **32** (2007) 233–262.
- [14] B. Cockburn, B. Dong and J. Guzmán, A superconvergent LDG-hybridizable Galerkin method for second-order elliptic problems. *Math. Comput.* **77** (2008) 1887–1916.

- [15] B. Cockburn, B. Dong and J. Guzmán, Optimal convergence of the original DG method for the transport-reaction equation on special meshes. *SIAM J. Numer. Anal.* **48** (2008) 1250–1265.
- [16] B. Cockburn, B. Dong, J. Guzmán, M. Restelli and R. Sacco, A hybridizable discontinuous Galerkin method for steady-state convection-diffusion-reaction problems. *J. Sci. Comput.* **31** (2009) 3827–3846.
- [17] B. Cockburn, B. Dong, J. Guzmán and J. Qian, Optimal Convergence of the Original DG Method on Special Meshes for Variable Transport Velocity. *SIAM J. Numer. Anal.* **46** (2010) 133–146.
- [18] B. Cockburn, O. Dubois, J. Gopalakrishnan and S. Tan, Multigrid for an HDG method. To appear in *IMA J. Numer. Anal.*
- [19] B. Cockburn, J. Gopalakrishnan and R. Lazarov, Unified hybridization of discontinuous Galerkin, mixed and continuous Galerkin methods for second order elliptic problems. *SIAM J. Numer. Anal.* **47** (2009) 1319–1365.
- [20] B. Cockburn, J. Gopalakrishnan and F.-J. Sayas, A projection-based error analysis of HDG methods. *Math. Comput.* **79** (2010) 1351–1367.
- [21] B. Cockburn, W. Qiu and K. Shi, Conditions for superconvergence of HDG Methods for second-order elliptic problems. *Math. Comput.* **81** (2012) 1327–1353.
- [22] B. Cockburn and C.W. Shu, The local discontinuous Galerkin method for time-dependent convection-diffusion systems. *SIAM J. Numer. Anal.* **35** (1998) 2440–2463.
- [23] B. Cockburn and C.W. Shu, Runge–Kutta discontinuous Galerkin method for convection-dominated problems. *J. Sci. Comput.* **16** (2001) 173–261.
- [24] A. Devinatz, R. Ellis and A. Friedman, The asymptotic behavior of the first real eigenvalue of second order elliptic operators with a small parameter in the highest derivatives. II. *Indiana Univ. Math. J.* **23** (1973-1974) 991–1011.
- [25] W. Eckhaus, Boundary layers in linear elliptic singular perturbation problems. *SIAM Rev.* **44** (1972).
- [26] H. Egger and J. Schöberl, A hybrid mixed discontinuous Galerkin finite-element method for convection diffusion problems. *IMA J. Num. Anal.* **30** (2010) 1206–1234.
- [27] H. Goering, A. Felgenhauer, G. Lube, H.-G. Roos and L. Tobiska, Singularly perturbed differential equations. Akademie-Verlag, Berlin (1983).
- [28] J. Gopalakrishnan and G. Kanschat, A multilevel discontinuous Galerkin method. *Numer. Math.* **95** (2003) 527–550.
- [29] P. Houston, C. Schwab and E. Süli, Stabilized hp–finite element methods for first order hyperbolic problems. *SIAM J. Numer. Anal.* **37** (2000) 1618–1643.
- [30] P. Houston, C. Schwab and E. Süli, Discontinuous hp-finite element methods for advection-diffusion-reaction problems. *SIAM J. Numer. Anal.* **39** (2002) 2133–2163.
- [31] T.J.R. Hughes, G. Scovazzi, P. Bochev and A. Buffa, A multiscale discontinuous Galerkin method with the computational structure of a continuous Galerkin method. *Comput. Methods Appl. Mech. Engrg.* **195** (2006) 2761–2787.
- [32] C. Johnson and J. Pitkäranta, An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation. *Math. Comput.* **46** (1986) 1–26.
- [33] T. Iliescu, A flow-aligning algorithm for convection-dominated problems. *Int. J. Numer. Methods Engrg.* **46** (1999) 993–1000.
- [34] T.E. Peterson, A note on the convergence of the discontinuous Galerkin method for a scalar hyperbolic equation. *SIAM J. Numer. Anal.* **28** (1991) 133–140.
- [35] W. Reed and T. Hill, *Triangular mesh methods for the neutron transport equation*. Technical Report LA- UR-73-479, Los Alamos Scientific Laboratory (1973).
- [36] N.C. Nguyen, J. Peraire and B. Cockburn, An implicit high-order hybridizable discontinuous Galerkin method for linear convection-diffusion equations. *J. Comput. Phys.* **288** (2009) 3232–3254.
- [37] N.C. Nguyen, J. Peraire and B. Cockburn, An implicit high-order hybridizable discontinuous Galerkin method for nonlinear convection-diffusion equations. *J. Comput. Phys.* **288** (2009) 8841–8855.
- [38] H.-G. Roos, Robust numerical methods for singularly perturbed differential equations: a survey covering 2008–2012. *ISRN Appl. Math.* (2012) 1–30.
- [39] H.-G. Roos, M. Stynes and L. Tobiska, Robust numerical methods for singularly perturbed differential equations. Convection-diffusion and flow problems. 2nd edition. In vol. 24 of *Springer Series Comput. Math.* Springer-Verlag, Berlin (2008).
- [40] H. Zarin and H.G. Roos, Interior penalty discontinuous approximations of convection-diffusion problems with parabolic layers. *Numer. Math.* **100** (2005) 735–759.