

ANALYSIS OF A DG METHOD FOR SINGULARLY PERTURBED CONVECTION-DIFFUSION PROBLEMS*

Runchang Lin¹, Xiu Ye², Shangyou Zhang³ and Peng Zhu^{4,†}

Abstract In this article, we studied a discontinuous Galerkin finite element method for convection-diffusion-reaction problems with singular perturbation. Our approach is highly flexible by allowing the use of discontinuous approximating function on polytopal mesh without imposing extra conditions on the convection coefficient. *A priori* error estimate is devised in a suitable energy norm on general polytopal mesh. Numerical examples are provided.

Keywords Discontinuous Galerkin, finite element method, convection diffusion problem, singular perturbation, polyhedral mesh.

MSC(2010) 65N15, 65N30.

1. Introduction

In this paper, we consider *a priori* error estimation of a discontinuous Galerkin (DG) method for the stationary state convection-diffusion problem

$$-\epsilon \Delta u + \nabla \cdot (\mathbf{b}u) + cu = f \quad \text{in } \Omega \subset \mathbb{R}^d, \quad (1.1)$$

$$u = 0 \quad \text{on } \partial\Omega, \quad (1.2)$$

where $\epsilon > 0$ is the diffusion parameter, $d = 2$ or 3 , and Ω is a polygonal (when $d = 2$) or polyhedral (when $d = 3$) domain with boundary $\partial\Omega$. It is usually assumed that $\mathbf{b}$, c , and f are sufficiently smooth, $\mathbf{b} \in [W^{1,\infty}(\Omega)]^d$, and $c + \frac{1}{2}\nabla \cdot \mathbf{b} \geq c_0 > 0$ for some constant c_0 , so that problem (1.1)-(1.2) has a unique solution in $H_0^1(\Omega) \cap H^2(\Omega)$ (cf., e.g., [11]).

It is well-known that, when convection dominates diffusion (i.e. $\epsilon \ll 1$, which is also referred to as the *singular perturbation parameter* in this case), the solution of

[†]the corresponding author. Email address: zhupeng.hnu@gmail.com (P. Zhu)

¹Department of Mathematics and Physics, Texas A&M International University, Laredo, Texas 78041, USA

²Department of Mathematics, University of Arkansas at Little Rock, Little Rock, Arkansas 72204, USA

³Department of Mathematical Sciences, University of Delaware, Newark, Delaware, 19716, USA

⁴School of Mathematics, Physics and Information Engineering, Jiaxing University, Jiaxing, Zhejiang 314001, China

*This research of the first author was partially supported by a University Research Grant of Texas A&M International University. This research of the second author was supported in part by National Science Foundation under Grant DMS-1620016. This research of the third author was supported in part by NSF of China (No.11571023). This research of the fourth author was supported in part by Natural Science Foundation of Zhejiang (No.LY19A010008).

the boundary value problem typically possesses *layers*, which are thin regions where the solution and/or its derivatives change rapidly. Standard numerical methods will produce solutions with nonphysical oscillations in this case, unless the computational mesh has a size of the magnitude of the layers. Numerical stabilization techniques, including *fitted mesh methods* and *fitted operator methods*, have been developed to resolve the difficulty; cf. the books [8, 9, 11] and the references therein. In particular, DG methods with interior penalty have been proved to be an effective approach for solving convection-diffusion problems [3, 6, 7].

In this paper, we are interested in using discontinuous Galerkin methods with interior penalty to solve convection-diffusion-reaction problems with singular perturbation. In [2], Ayuso and Marini apply weighted-residual approach to recover discontinuous Galerkin formulations for advection-diffusion-reaction problems with singular perturbation. An optimal error estimation of the order $O((\epsilon^{1/2} + h^{1/2})h^k)$ has been obtained on triangular mesh under a set of assumptions for the convection coefficient $\mathbf{b}$. Analysis of finite element method on polygonal mesh is a new trend in the field of numerical PDE, such as the works [1, 5, 10, 12] and the references therein. The goal of this article is to analyze an interior penalty discontinuous Galerkin method for the problem (1.1)-(1.2) on polygonal mesh. Compare to [2], our analysis is simple and allows the use of general mesh such as polytopal mesh, hybrid mesh and mesh with hanging nodes. In addition, a set of strict assumptions for the convection coefficient in [2] is removed in our analysis.

The DG methods in [2] and the DG methods proposed in this paper share the same finite element space. For the DG formulations in [2], integration by parts is used for the convection term and as a consequence, derivative of trial function turns to derivative of the test function. In this paper, we handle the convection term differently without applying integration by parts on it. As a result, we can relax the restriction of the convection coefficient.

The rest of this article is organized as follows. In Section 2, the DG FEM will be introduced. Analysis of the DG method is found in Section 3. Numerical experiments are presented in Section 4 to support the theoretical results.

2. Discontinuous Galerkin Finite Element Schemes

In this paper, standard definitions and notations of Sobolev spaces are used. For a polyhedron $D \subseteq \Omega$, we denote $H^s(D) = W^{s,2}(D)$ the Hilbertian Sobolev space of index $s \geq 0$ defined on D . The associated inner product, norm, and semi-norms in $H^s(D)$ are denoted by $(\cdot, \cdot)_{s,D}$, $\|\cdot\|_{s,D}$, and $|\cdot|_{s,D}$, respectively. When $s = 0$, $H^0(D)$ coincides with the space of square integrable functions $L^2(D)$. In this case, the subscript s is suppressed from the notations of norm, semi-norm, and inner product. Furthermore, the subscript D is also suppressed when $D = \Omega$. Throughout this article, we use C for generic constants independent of ϵ , mesh size, and the solution to equation (1.1)-(1.2), which may not necessarily be the same at each occurrence. We will use plain and bold fonts for scalars and vectors, respectively.

Let $\mathcal{T}_h$ be a shape regular quasi-uniform triangulation of the domain Ω which consists of polygons in two dimension or polyhedra in three dimension satisfying a set of conditions specified in [13]. Denote by $\mathcal{E}_h$ the set of all edges or flat faces in $\mathcal{T}_h$, and let $\mathcal{E}_h^0 = \mathcal{E}_h \setminus \partial\Omega$ be the set of all interior edges or flat faces. For every element $T \in \mathcal{T}_h$, we denote by h_T its diameter. The mesh size for $\mathcal{T}_h$ is $h = \max_{T \in \mathcal{T}_h} h_T$.

For a given integer $k \geq 1$, let V_h be a discontinuous Galerkin finite element space associated with $\mathcal{T}_h$ defined as

$$V_h = \{v \in L^2(\Omega) : v|_T \in \mathbb{P}_k(T), T \in \mathcal{T}_h\}, \quad (2.1)$$

where $\mathbb{P}_k$ is the space of polynomials of total degree up to k .

Let T_1 and T_2 be elements sharing a common edge e , for which $\mathbf{n}_1$ and $\mathbf{n}_2$ are the unit outward normal vectors with respect to T_1 and T_2 , respectively. We define the jump and the average of a scalar valued function v on e as

$$\begin{aligned} \llbracket v \rrbracket_e &= \begin{cases} v|_{T_1} \mathbf{n}_1 + v|_{T_2} \mathbf{n}_2, & e \in \mathcal{E}_h^0, \\ v \mathbf{n}, & e \subset \partial\Omega, \end{cases} \\ \{v\}_e &= \begin{cases} \frac{1}{2}(v|_{T_1} + v|_{T_2}), & e \in \mathcal{E}_h^0, \\ v, & e \subset \partial\Omega. \end{cases} \end{aligned}$$

We introduce the following inner products for broken Sobolev spaces

$$\begin{aligned} (v, w)_{\mathcal{T}_h} &= \sum_{T \in \mathcal{T}_h} (v, w)_T = \sum_{T \in \mathcal{T}_h} \int_T v w d\mathbf{x}, \\ \langle v, w \rangle_{\partial\mathcal{T}_h} &= \sum_{T \in \mathcal{T}_h} \langle v, w \rangle_{\partial T} = \sum_{T \in \mathcal{T}_h} \int_{\partial T} v w ds. \end{aligned}$$

Define

$$\partial_+ T = \{\mathbf{x} \in \partial T : \mathbf{b}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) \geq 0\}, \quad \partial_- T = \{\mathbf{x} \in \partial T : \mathbf{b}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) \leq 0\}.$$

For any $T \in \mathcal{T}_h$, let v^o represent the value of v at the element adjacent to T and let $v^o = 0$ on e if $e \subset \partial\Omega \cap \partial T$. We may introduce some forms on V_h as follows:

$$\begin{aligned} s_d(v, w) &= \alpha \sum_{e \in \mathcal{E}_h} \epsilon h_e^{-1} \int_e \llbracket v \rrbracket \cdot \llbracket w \rrbracket ds, \\ s_c(v, w) &= \sum_{T \in \mathcal{T}_h} \langle (\mathbf{b} \cdot \mathbf{n})(v^o - v), w \rangle_{\partial_- T}, \\ a_d(v, w) &= \epsilon(\nabla v, \nabla w)_{\mathcal{T}_h} - \epsilon \langle \{\nabla v\}, \llbracket w \rrbracket \rangle_{\mathcal{E}_h} - \epsilon \langle \{\nabla w\}, \llbracket v \rrbracket \rangle_{\mathcal{E}_h} + s_d(v, w), \\ a_c(v, w) &= (\nabla \cdot (\mathbf{b}v), w)_{\mathcal{T}_h} + (cv, w) + s_c(v, w), \\ a(v, w) &= a_d(v, w) + a_c(v, w). \end{aligned}$$

Algorithm 2.1 (Discontinuous Galerkin Method). A numerical approximation for (1.1) and (1.2) can be obtained by seeking $u_h \in V_h$ satisfying the following equation:

$$a(u_h, v) = (f, v), \quad \forall v \in V_h. \quad (2.2)$$

3. Analysis

First for any $v \in V_h$, we define

$$\begin{aligned}\|v\|_a^2 &= a_d(v, v) = \epsilon \sum_{T \in \mathcal{T}_h} \|\nabla v\|_T^2 + \alpha \sum_{e \in \mathcal{E}_h} \epsilon h_e^{-1} \llbracket v \rrbracket_e^2, \\ \|v\|_c^2 &= \|v\|^2 + \sum_{e \in \mathcal{E}_h^0} \| |\mathbf{b} \cdot \mathbf{n}|^{1/2} \llbracket v \rrbracket_e \|^2_e, \\ \|v\|^2 &= \|v\|_a^2 + \|v\|_c^2.\end{aligned}$$

It is easy to verify that $\|\cdot\|$ is a norm in V_h .

For any function $\varphi \in H^1(T)$, the following trace inequality holds true,

$$\|\varphi\|_e^2 \leq C (h_T^{-1} \|\varphi\|_T^2 + h_T \|\nabla \varphi\|_T^2). \quad (3.1)$$

The following lemmas are useful to establish the coercivity of the bilinear form $a(\cdot, \cdot)$.

Lemma 3.1. For $v, w \in V_h \cap H^1(\Omega)$,

$$\begin{aligned}(\nabla \cdot (\mathbf{b}v), w)_{\mathcal{T}_h} &= ((\nabla \cdot \mathbf{b})v, w)_{\mathcal{T}_h} - (v, \nabla \cdot (\mathbf{b}w))_{\mathcal{T}_h} \\ &\quad + \langle (\mathbf{b} \cdot \mathbf{n})v, w \rangle_{\partial \mathcal{T}_h},\end{aligned} \quad (3.2)$$

and

$$(\nabla \cdot (\mathbf{b}v), v)_{\mathcal{T}_h} = \frac{1}{2}((\nabla \cdot \mathbf{b})v, v)_{\mathcal{T}_h} + \frac{1}{2}\langle (\mathbf{b} \cdot \mathbf{n})v, v \rangle_{\partial \mathcal{T}_h}, \quad (3.3)$$

Proof. It follows from the definition of integration by parts that

$$\begin{aligned}(\nabla \cdot (\mathbf{b}v), w)_{\mathcal{T}_h} &= ((\nabla \cdot \mathbf{b})v, w)_{\mathcal{T}_h} + (\nabla v, \mathbf{b}w)_{\mathcal{T}_h} \\ &= ((\nabla \cdot \mathbf{b})v, w)_{\mathcal{T}_h} - (v, \nabla \cdot (\mathbf{b}w))_{\mathcal{T}_h} + \langle (\mathbf{b} \cdot \mathbf{n})w, v \rangle_{\partial \mathcal{T}_h},\end{aligned}$$

which implies (3.2). The equation (3.3) is a direct consequence of (3.2). $\square$

Lemma 3.2. For $v \in V_h$, then

$$\frac{1}{2} \sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n}v, v \rangle_{\partial T} + \sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n}(v^o - v), v \rangle_{\partial_- T} \geq \frac{1}{4} \sum_{e \in \mathcal{E}_h^0} \| |\mathbf{b} \cdot \mathbf{n}|^{1/2} \llbracket v \rrbracket_e \|^2_e. \quad (3.4)$$

Proof. It follows from the definitions of $\partial_+ T$ and $\partial_- T$,

$$\begin{aligned}& \frac{1}{2} \sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n}v, v \rangle_{\partial T} + \sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n}(v^o - v), v \rangle_{\partial_- T} \\ &= \sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n}v^o, v \rangle_{\partial_- T} - \frac{1}{2} \sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n}v, v \rangle_{\partial_- T} + \frac{1}{2} \sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n}v, v \rangle_{\partial_+ T} \\ &= \sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n}v^o, v \rangle_{\partial_- T \setminus \partial \Omega} - \frac{1}{2} \sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n}v, v \rangle_{\partial_- T \setminus \partial \Omega} + \frac{1}{2} \sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n}v, v \rangle_{\partial_+ T \setminus \partial \Omega} \\ &\quad - \frac{1}{2} \sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n}v, v \rangle_{\partial_- T \cap \partial \Omega} + \frac{1}{2} \sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n}v, v \rangle_{\partial_+ T \cap \partial \Omega}\end{aligned}$$

$$\begin{aligned}
&= \sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n} v^o, v \rangle_{\partial_- T \setminus \partial \Omega} - \frac{1}{2} \sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n} v, v \rangle_{\partial_- T \setminus \partial \Omega} - \frac{1}{2} \sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n} v^o, v^o \rangle_{\partial_- T \setminus \partial \Omega} \\
&\quad - \frac{1}{2} \sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n} v, v \rangle_{\partial_- T \cap \partial \Omega} + \frac{1}{2} \sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n} v, v \rangle_{\partial_+ T \cap \partial \Omega} \\
&\geq -\frac{1}{2} \sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n} (v^o - v), (v^o - v) \rangle_{\partial_- T \setminus \partial \Omega} \\
&= -\frac{1}{4} \left(\sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n} (v^o - v), (v^o - v) \rangle_{\partial_- T \setminus \partial \Omega} - \sum_{T \in \mathcal{T}_h} \langle \mathbf{b} \cdot \mathbf{n} (v^o - v), (v^o - v) \rangle_{\partial_+ T \setminus \partial \Omega} \right) \\
&\geq \frac{1}{4} \sum_{e \in \mathcal{E}_h^0} \| |\mathbf{b} \cdot \mathbf{n}|^{1/2} \llbracket v \rrbracket \|_e^2,
\end{aligned}$$

which completes the proof of the lemma. $\square$

The following coercivity result holds.

Lemma 3.3. *For $v \in V_h$, then for α large enough,*

$$C \|v\|^2 \leq a(v, v). \quad (3.5)$$

Therefore, the DG formulation (2.2) has a unique solution.

Proof. It is well known that for α large enough

$$C \|v\|_d^2 \leq a_d(v, v). \quad (3.6)$$

Using (3.3) and (3.4), we have

$$\begin{aligned}
a_c(v, v) &= (\nabla \cdot (\mathbf{b}v), v)_{\mathcal{T}_h} + (cv, v) + s_c(v, v) \\
&= \left(\left(c + \frac{1}{2} \nabla \cdot \mathbf{b} \right) v, v \right) + \frac{1}{2} \langle (\mathbf{b} \cdot \mathbf{n})v, v \rangle_{\partial \mathcal{T}_h} + s_c(v, v) \\
&\geq c_0 \|v\|^2 + \frac{1}{4} \sum_{e \in \mathcal{E}_h^0} \| |\mathbf{b} \cdot \mathbf{n}|^{1/2} \llbracket v \rrbracket \|_e^2 \\
&\geq C \|v\|_c^2.
\end{aligned}$$

Combining the above two estimates implies (3.5). $\square$

We next establish *a priori* error estimation of the DG method. Let Q_h be a element-wise defined L^2 projections such that for each element $T \in \mathcal{T}_h$, $Q_h : H^1(T) \rightarrow \mathbb{P}_k(T)$.

Lemma 3.4. *Let u be the solution of the problem (1.1)-(1.2). Then for any $v \in V_h$,*

$$|a_d(u - Q_h u, v)| \leq C \epsilon^{1/2} h^k |u|_{k+1} \|v\|, \quad (3.7)$$

$$|a_c(u - Q_h u, v)| \leq C h^{k+1/2} |u|_{k+1} \|v\|. \quad (3.8)$$

Proof. Using the Cauchy-Schwarz inequality, the trace inequality and the definition of Q_h , we have

$$\begin{aligned}
|a_d(u - Q_h u, v)| &\leq |\epsilon (\nabla(u - Q_h u), \nabla v)_{\mathcal{T}_h}| + |\epsilon \langle \{\nabla(u - Q_h u)\}, \llbracket v \rrbracket \rangle_{\mathcal{E}_h}| \\
&\quad + |\epsilon \langle \{\nabla v\}, \llbracket u - Q_h u \rrbracket \rangle_{\mathcal{E}_h}| + |s_d(u - Q_h u, v)| \\
&\leq C \epsilon^{1/2} h^k |u|_{k+1} \|v\|.
\end{aligned}$$

It follows from the definition of $a_c(\cdot, \cdot)$ that

$$\begin{aligned} a_c(u - Q_h u, v) &= (\nabla \cdot (\mathbf{b}(u - Q_h u), v)_{\mathcal{T}_h} + (c(u - Q_h u), v) + s_c(u - Q_h u, v) \\ &= -(u - Q_h u, \mathbf{b} \cdot \nabla v)_{\mathcal{T}_h} + (c(u - Q_h u), v) \\ &\quad + [(\mathbf{b} \cdot \mathbf{n})(u - Q_h u, v)_{\partial \mathcal{T}_h} + s_c(u - Q_h u, v)]. \end{aligned} \quad (3.9)$$

We will bound the three terms on the right hand side of the above equation.

The inverse inequality and the definition of Q_h imply

$$\begin{aligned} |(u - Q_h u, \mathbf{b} \cdot \nabla v)_{\mathcal{T}_h}| &= |(u - Q_h u, (\mathbf{b} - \bar{\mathbf{b}}) \cdot \nabla v)_{\mathcal{T}_h}| \\ &\leq Ch^{k+1} |u|_{k+1} \|v\|, \end{aligned}$$

where $\bar{\mathbf{b}}$ is the average of $\mathbf{b}$ on each element $T \in \mathcal{T}_h$.

Obviously, we have

$$|(c(u - Q_h u), v)| \leq Ch^{k+1} |u|_{k+1} \|v\|.$$

It follows from the definitions of $\partial_+ T$ and $\partial_- T$,

$$\begin{aligned} &\sum_{T \in \mathcal{T}_h} \langle (\mathbf{b} \cdot \mathbf{n})w, v \rangle_{\partial T} + s_c(w, v) \\ &= \sum_{T \in \mathcal{T}_h} \langle (\mathbf{b} \cdot \mathbf{n})w, v \rangle_{\partial T} + \sum_{T \in \mathcal{T}_h} \langle (\mathbf{b} \cdot \mathbf{n})(w^o - w), v \rangle_{\partial_- T} \\ &= \sum_{T \in \mathcal{T}_h} \langle (\mathbf{b} \cdot \mathbf{n})w^o, v \rangle_{\partial_- T} + \sum_{T \in \mathcal{T}_h} \langle (\mathbf{b} \cdot \mathbf{n})w, v \rangle_{\partial_+ T} \\ &= \sum_{T \in \mathcal{T}_h} \langle (\mathbf{b} \cdot \mathbf{n})w^o, v \rangle_{\partial_- T \setminus \partial \Omega} + \sum_{T \in \mathcal{T}_h} \langle (\mathbf{b} \cdot \mathbf{n})w, v \rangle_{\partial_+ T \setminus \partial \Omega} \\ &\quad + \sum_{T \in \mathcal{T}_h} \langle (\mathbf{b} \cdot \mathbf{n})w^o, v \rangle_{\partial_- T \cap \partial \Omega} + \sum_{T \in \mathcal{T}_h} \langle (\mathbf{b} \cdot \mathbf{n})w, v \rangle_{\partial_+ T \cap \partial \Omega} \\ &= - \sum_{T \in \mathcal{T}_h} \langle (\mathbf{b} \cdot \mathbf{n})w, v^o \rangle_{\partial_+ T \setminus \partial \Omega} + \sum_{T \in \mathcal{T}_h} \langle (\mathbf{b} \cdot \mathbf{n})w, v \rangle_{\partial_+ T \setminus \partial \Omega} \\ &\quad + \sum_{T \in \mathcal{T}_h} \langle (\mathbf{b} \cdot \mathbf{n})w, v \rangle_{\partial_+ T \cap \partial \Omega} \\ &= \sum_{T \in \mathcal{T}_h} \langle (\mathbf{b} \cdot \mathbf{n})w, (v - v^o) \rangle_{\partial_+ T \setminus \partial \Omega} + \sum_{T \in \mathcal{T}_h} \langle (\mathbf{b} \cdot \mathbf{n})w, v \rangle_{\partial_+ T \cap \partial \Omega}. \end{aligned}$$

Using the equation above with $w = u - Q_h u$ and trace inequality (3.1), we have

$$\begin{aligned} &\sum_{T \in \mathcal{T}_h} \langle (\mathbf{b} \cdot \mathbf{n})(u - Q_h u), v \rangle_{\partial T} + s_c(u - Q_h u, v) \\ &= \sum_{T \in \mathcal{T}_h} \langle (\mathbf{b} \cdot \mathbf{n})(u - Q_h u), (v - v^o) \rangle_{\partial_+ T \setminus \partial \Omega} + \sum_{T \in \mathcal{T}_h} \langle (\mathbf{b} \cdot \mathbf{n})(u - Q_h u), v \rangle_{\partial_+ T \cap \partial \Omega} \\ &\leq Ch^{k+1/2} |u|_{k+1} \|v\|. \end{aligned}$$

Combining the above estimates with (3.9) gives (3.8). We complete the proof of the lemma. $\square$

Theorem 3.1. *Let $u_h \in V_h$ be the DG finite element solution of the problem (1.1)-(1.2) arising from (2.2). Then there exists a constant C such that*

$$\|u - u_h\| \leq C(\epsilon^{1/2} + h^{1/2})h^k |u|_{k+1}. \quad (3.10)$$

Proof. Let u be the solution of the problem (1.1)-(1.2). Then it is obvious that for any $v \in V_h$,

$$a(u, v) = (f, v).$$

Subtracting (2.2) from the equation above yields

$$a(u - u_h, v) = 0.$$

Adding and subtracting $Q_h u$ give

$$a(Q_h u - u_h, v) = -a(u - Q_h u, v), \quad \forall v \in V_h.$$

Letting $v = Q_h u - u_h := e_h$, and using (3.5), (3.7) and (3.8), we have

$$\begin{aligned} \|e_h\|^2 &\leq Ca(e_h, e_h) \leq C|a(u - Q_h u, e_h)| \leq C(|a_d(u - Q_h u, e_h)| + |a_c(u - Q_h u, e_h)|) \\ &\leq C(\epsilon^{1/2} + h^{1/2})h^k |u|_{k+1} \|e_h\|. \end{aligned}$$

Using the triangle inequality, we have proved the theorem. $\square$

4. Numerical example

In this section, we present numerical results for the DG formulation (2.2). Examples of different types are used to confirm numerically the theoretical estimates in Section 3. $\mathbb{P}_k$ elements ($k = 1, 2, 3, 4$) on either uniform rectangular meshes, or uniform pentagonal (hybrid with rectangles) meshes, shown in Figure 1, are used in all the examples. Note that as $\mathbb{Q}_k$ elements fit rectangular meshes, the rectangular meshes are still considered as general polygonal meshes for the $\mathbb{P}_k$ elements.

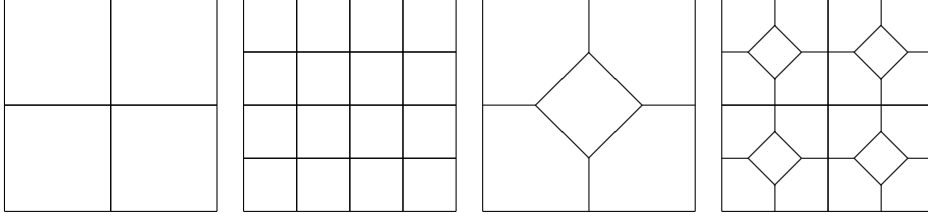


Figure 1. The uniform rectangular meshes of $h = 1/2, 1/4$, and the uniform pentagonal meshes of $h = 1/2, 1/4$.

Example 4.1 (Smooth solution). Let $\Omega = (0, 1)^2$, $\mathbf{b} = (1, 1)^T$, and $c = 1$. The source term f is determined such that the exact solution of (1.1) is

$$u(x, y) = x(1 - x)y(1 - y). \quad (4.1)$$

We computed approximate solutions to the boundary value problem for $\epsilon = 10^{-3}$ and 10^{-9} to confirm the error estimates of the DG scheme. Here, $\alpha = 10$ is used for

(2.2) in all computations. Numerical error $e_h = Q_h u - u_h$ is measured in both the L^2 norm $\|\cdot\|$ and the energy norm $\|\cdot\|_E$. The corresponding convergence rates for linear, quadratic, and cubic elements have been collected in Tables 1-3, respectively. Each convergence rate r is computed under the presumption that one has convergence of order $O(h^r)$. The optimal L^2 estimate of order $O(h^{k+1})$ is observed. On the other hand, numerical errors measured in the energy norm converges at $O(h^{k+1/2})$, which matches the estimate in Theorem 3.1.

Table 1. Example 4.1: The errors $e_h = Q_h u - u_h$ and the orders of convergence for (4.1) with $k = 1$.

h	$\epsilon = 10^{-3}$				$\epsilon = 10^{-9}$			
	$\ e_h\ $	Rate	$\ e_h\ _E$	Rate	$\ e_h\ $	Rate	$\ e_h\ _E$	Rate
Rectangular meshes								
1/4	0.002652	—	0.00865	—	0.002680	—	0.00857	—
1/8	0.000620	2.1	0.00311	1.5	0.000637	2.1	0.00304	1.5
1/16	0.000144	2.1	0.00115	1.4	0.000152	2.1	0.00109	1.5
1/32	0.000034	2.1	0.00043	1.4	0.000037	2.0	0.00039	1.5
Pentagonal meshes								
1/4	0.002047	—	0.00828	—	0.002081	—	0.00826	—
1/8	0.000483	2.1	0.00287	1.5	0.000508	2.0	0.00288	1.5
1/16	0.000115	2.1	0.00101	1.5	0.000128	2.0	0.00102	1.5
1/32	0.000027	2.1	0.00035	1.5	0.000032	2.0	0.00036	1.5

Table 2. Example 4.1: The errors $e_h = Q_h u - u_h$ and the orders of convergence for (4.1) with $k = 2$.

h	$\epsilon = 10^{-3}$				$\epsilon = 10^{-9}$			
	$\ e_h\ $	Rate	$\ e_h\ _E$	Rate	$\ e_h\ $	Rate	$\ e_h\ _E$	Rate
Rectangular meshes								
1/4	0.000606	—	0.00300	—	0.000609	—	0.00290	—
1/8	0.000074	3.0	0.00054	2.5	0.000075	3.0	0.00051	2.5
1/16	0.000009	3.0	0.00010	2.4	0.000009	3.0	0.00009	2.5
1/32	0.000001	3.0	0.00002	2.4	0.000001	3.0	0.00001	2.5
Pentagonal meshes								
1/4	0.000431	—	0.00163	—	0.000433	—	0.00158	—
1/8	0.000050	3.1	0.00029	2.5	0.000051	3.1	0.00027	2.5
1/16	0.000006	3.1	0.00005	2.4	0.000006	3.0	0.00005	2.5
1/32	0.000001	3.0	0.00001	2.4	0.000001	3.0	0.00001	2.5

Example 4.2 (Two boundary layers). In this example, we examine the performance of the proposed DG method in the occurrence of boundary layer. Let $\Omega = (0, 1)^2$, $\mathbf{b} = (1, 0)^T$, and $c = 1$. We choose f and homogeneous Dirichlet boundary conditions so that the exact solution to (1.1)-(1.2) is

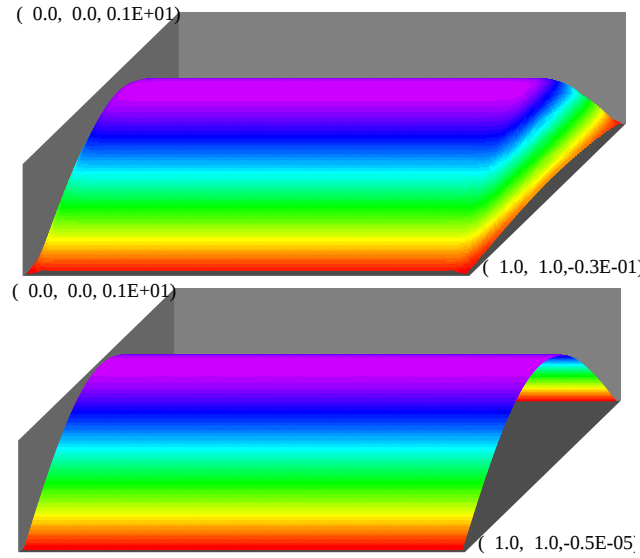
$$u(x, y) = \sin(\pi x) \left(1 - e^{-y/\sqrt{\epsilon}}\right) \left(1 - e^{(y-1)/\sqrt{\epsilon}}\right) \left(1 - e^{-1/\sqrt{\epsilon}}\right)^{-1}. \quad (4.2)$$

Table 3. Example 4.1: The errors $e_h = Q_h u - u_h$ and the orders of convergence for (4.1) with $k = 3$.

h	$\epsilon = 10^{-3}$				$\epsilon = 10^{-9}$			
	$\ e_h\ $	Rate	$\ e_h\ $	Rate	$\ e_h\ $	Rate	$\ e_h\ $	Rate
Rectangular meshes								
1/2	0.0019459	—	0.007840	—	0.0019487	—	0.007699	—
1/4	0.0001230	4.0	0.000690	3.5	0.0001234	4.0	0.000666	3.5
1/8	0.0000077	4.0	0.000062	3.5	0.0000077	4.0	0.000058	3.5
1/16	0.0000004	4.0	0.000005	3.4	0.0000005	3.9	0.000005	3.5
Pentagonal meshes								
1/2	0.0007408	—	0.002477	—	0.0007447	—	0.002426	—
1/4	0.0000477	4.0	0.000220	3.5	0.0000483	3.9	0.000212	3.5
1/8	0.0000030	4.0	0.000020	3.5	0.0000031	4.0	0.000019	3.5
1/16	0.0000002	4.0	0.000002	3.4	0.0000002	3.6	0.000002	3.5

This problem has two characteristic boundary layers with the width of $\mathcal{O}(\sqrt{\epsilon})$ near boundaries $y = 0$ and $y = 1$.

$\mathbb{P}_1$ finite element space is used on rectangular meshes. The numerical solution behaves as expected when $\epsilon = 10^{-4}$, (cf. Table 4), where it converges before resolving the layer and after resolving the layer. The numerical solutions for the boundary value problem with $\epsilon = 10^{-4}$ and $\epsilon = 10^{-10}$ are plotted in Figure 2. We can see the first solution resolves the boundary layer while the second does not. When $\epsilon = 10^{-10}$, the strongly convection-dominated case, optimal convergence rates in both L^2 and energy norms are obtained as shown in Table 4.

**Figure 2.** Example 4.2. The $\mathbb{P}_1$ numerical solution for $\epsilon = 10^{-4}$ (top) and $\epsilon = 10^{-10}$, on a 64×64 rectangle mesh.

Example 4.3 (Internal layer–discontinuous boundary conditions). Let $\Omega = (0, 1)^2$,

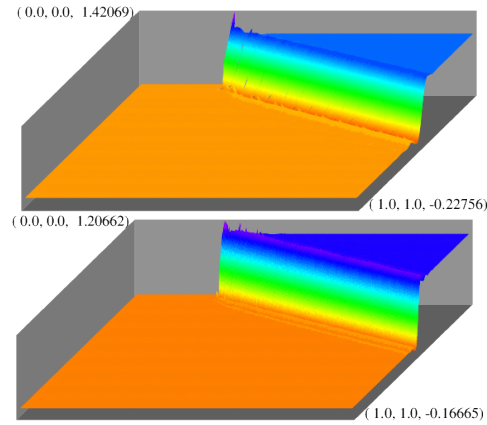
Table 4. Example 4.2: The errors $e_h = Q_h u - u_h$ and the orders of convergence for (4.2) on square grids with $k = 1$.

h	$\epsilon = 10^{-4}$				$\epsilon = 10^{-10}$			
	$\ e_h\ _0$	Rate	$\ e_h\ $	Rate	$\ e_h\ _0$	Rate	$\ e_h\ $	Rate
1/2	0.7645E-01	2.3	0.2297E+00	1.4	0.763E-01	2.3	0.231E+00	1.4
1/4	0.2100E-01	1.9	0.9099E-01	1.3	0.201E-01	1.9	0.934E-01	1.3
1/8	0.1300E-01	0.7	0.3287E-01	1.5	0.513E-02	2.0	0.351E-01	1.4
1/16	0.1657E-01	0.0	0.2436E-01	0.4	0.130E-02	2.0	0.127E-01	1.5
1/32	0.1529E-01	0.1	0.3172E-01	0.0	0.326E-03	2.0	0.452E-02	1.5
1/64	0.7213E-02	1.1	0.2369E-01	0.4	0.819E-04	2.0	0.159E-02	1.5
1/128	0.2399E-02	1.6	0.1108E-01	1.1	0.247E-04	1.7	0.473E-03	1.7

$\epsilon = 10^{-9}$, $\mathbf{b} = (1/2, \sqrt{3}/2)^T$, and $c = 0$ for (1.1). We choose a non-homogeneous Dirichlet boundary condition

$$u = \begin{cases} 1 & \text{on } \{0\} \times (1/4, 1), \\ 0 & \text{elsewhere on } \partial\Omega. \end{cases} \quad (4.3)$$

Due to the discontinuities in boundary conditions, the solution of this problem is not in $H^1(\Omega)$. Numerical oscillations occur near the interior layer caused by the joints of the conflicting Dirichlet boundary conditions. This phenomenon has been reported in the literature for DG methods; see, e.g., [2, 4]. We use $\mathbb{P}_2$ and $\mathbb{P}_4$ elements over rectangular meshes. The numerical solutions on a mesh of 1,024 uniform squares are plotted in Figure 3. The solutions do not oscillate outside of the internal layer.

**Figure 3.** Example 4.3. The P_2 and P_4 numerical solutions for $\epsilon = 10^{-9}$ over a mesh of 1,024 elements.

Example 4.4 (A rotational flow). We solve the following boundary value problem

$$\begin{cases} -\epsilon \Delta u + \nabla \cdot (\mathbf{b}u) + cu = 0, & \text{in } \Omega = (0, 1)^2 \setminus \{1/2\} \times (0, 1/2), \\ u = x(1-x)y(1-y)(y-1/2)^2 & \text{on } \partial\Omega, \end{cases} \quad (4.4)$$

where $\epsilon = 10^{-9}$, $\mathbf{b} = (1/2 - y, x - 1/2)^T$, and $c = 10^{-3}$.

The numerical solutions for $\mathbb{P}_1$ and $\mathbb{P}_4$ elements over a mesh of 1,024 uniform squares are plotted in Figure 4. The numerical results indicate that the discontinuous Galerkin discretization is stable, and accurate.

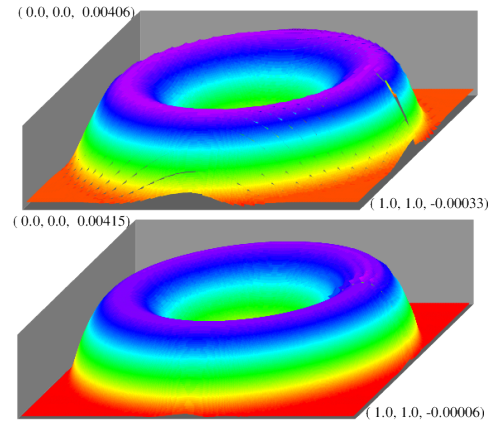


Figure 4. Example 4. The numerical $\mathbb{P}_1$ and $\mathbb{P}_4$ solutions for $\epsilon = 10^{-9}$ over a mesh of 1,024 elements.

Acknowledgements

The authors gratefully acknowledge the editor and two anonymous referees for their time and invaluable suggestions, which improve the quality of the paper to a large extent.

References

- [1] P. F. Antonietti, P. Houston, X. Hu, et al, *Multigrid algorithms for hp-version interiori penalty discontinuous Galerkin methods on polygonal and polyhedral meshes*, Calcolo, 2017, 54, 1169–1198.
- [2] B. Ayuso and L. D. Marini, *Discontinuous Galerkin methods for advection-diffusion-reaction problems*, SIAM J. Numer. Anal., 2009, 47(2), 1391–1420.
- [3] C. E. Baumann and J. T. Oden, *A discontinuous hp finite element method for convection-diffusion problems*, Comput. Methods Appl. Mech. Engrg., 1999, 175(3–4), 311–341.
- [4] A. Buffa, T. J. R. Hughes, and G. Sangalli, *Analysis of a multiscale discontinuous Galerkin method for convection-diffusion problems*, SIAM J. Numer. Anal., 2006, 44, 1420–1440.
- [5] A. Cangiani, E. H. Georgoulis, T. Pryer and O. J. Sutton, *A posteriori error estimates for the virtual element method*, Numer. Math., 2017, 137, 857–893.
- [6] J. Du and E. Chung, *An adaptive staggered discontinuous Galerkin method for steady steady state convection-diffusion equation*, J. Sci. Comput., 2018. DOI: 10.1007/s10915-018-0695-9.

- [7] P. Houston, C. Schwab, and E. Süli, *Discontinuous hp-finite element methods for advection-diffusion-reaction problems*, SIAM J. Numer. Anal., 2002, 39(6), 2133–2163.
- [8] T. Linß, *Layer-adapted meshes for reaction-convection-diffusion problems*, Springer-Verlag, Berlin, 2010.
- [9] J. J. H. Miller, E. O’Riordan and G. I. Shishkin, *Fitted numerical methods for singular perturbation problems. Error estimates in the maximum norm for linear problems in one and two dimensions*. World Scientific Publishing Co., Inc., River Edge, NJ, 1996.
- [10] D. A. Di Pietro and A. Ern, *Hybrid high-order methods for variable-diffusion problems on general meshes*, C.R. Math. Acad. Sci. Paris, 2015, 353(1), 31–34.
- [11] H.-G. Roos, M. Stynes, and L. Tobiska, *Robust numerical methods for singularly perturbed differential equations. Convection-diffusion-reaction and flow problems(Second edition)*, Springer-Verlag, Berlin, 2008.
- [12] C. Talischi, *A family of $H(\text{div})$ finite element approximations on polygonal meshes*, SIAM J. Sci. Comput., 2015, 37, 1067–1088.
- [13] J. Wang and X. Ye, *A weak Galerkin mixed finite element method for second-order elliptic problems*, Math. Comp., 2014, 83, 2101–2126.