SIAM J. NUMER. ANAL.
Vol. 57, No. 6, pp. 2785–2811

PRIMAL DUAL MIXED FINITE ELEMENT METHODS FOR INDEFINITE ADVECTION-DIFFUSION EQUATIONS*

ERIK BURMAN† AND CUIYU HE†

Abstract. We consider primal dual mixed finite element methods for the advection-diffusion equation. For the primal variable we use standard continuous finite element space and for the flux we use the Raviart-Thomas space. We prove optimal a priori error estimates in the H^1- and the L^2-norms for the primal variable in the low Péclet regime. In the high Péclet regime we also prove optimal error estimates for the primal variable in the $H(div)$ norm for smooth solutions. Numerically we observe that the method eliminates the spurious oscillations close to interior layers that pollute the solution of the standard Galerkin method when the local Péclet number is high. This method, however, does produce spurious oscillations when outflow boundary layers are present in the solution. In the last section we propose two simple strategies to remove such numerical artifacts caused by the outflow boundary layer and validate them numerically.

Key words. advection-diffusion, primal dual method, mixed finite element method

AMS subject classification. 65N30

DOI. 10.1137/18M1221473

1. Introduction.

Advection-diffusion problems have been extensively studied in the last decades for their wide applications in the areas of weather forecasting, oceanography, gas dynamics, and contaminant transportation in porous media, to name a few. Many numerical methods for advection-diffusion equations have been explored in the literature. The two main concerns when designing a numerical method for advection-diffusion problems are robustness in the advection dominated limit and satisfaction of local conservation. The standard Galerkin method, using globally continuous approximation, is known to fail on both points, and therefore much effort has been devoted to the design of alternative formulations. Typically to make the method stable in the limit of dominating advection some stabilizing operator is introduced to provide sufficient control of fine scale fluctuations. The most well known stabilized method is the streamline upwind Petrov–Galerkin method introduced by Brooks and Hughes [7] and first analyzed by Johnson, Nävert, and Pitkäranta [37]. In order to avoid disadvantages associated to the Petrov–Galerkin character, for instance related to time discretization, the discontinuous Galerkin method was introduced, first in the context of hyperbolic transport [38, 26]. In this case the stabilizing mechanism is due to the upwind flux, which controls the solution jump over element faces and adds a dissipation proportional to this jump. In the context of finite element methods using H^1-conforming approximation several stabilized methods using symmetric stabilization have been proposed, for instance, the subgrid viscosity method by Guermond [35], the orthogonal subscale method by Codina [27], and the continuous interior penalty method, introduced by Douglas and Dupont [31] and analyzed by Burman and Hansbo [14]. It is well known that for cases of both discontinuous and

*Received by the editors October 18, 2018; accepted for publication (in revised form) September 20, 2019; published electronically November 21, 2019.
https://doi.org/10.1137/18M1221473
Funding: The work of the authors was supported by the EPSRC grant EP/P01576X/1.
†Department of Mathematics, University College London, London, UK WC1E 6BT (e.burman@ucl.ac.uk, c.he@ucl.ac.uk).

continuous approximation spaces a local conservative numerical flux can be defined. In the continuous case, however, it must be reconstructed using postprocessing [36, 17].

In this work, to ensure local conservation of the computed flux we design a method in the mixed setting: we approximate the primal variable in the standard conforming finite element space and the flux in the Raviart–Thomas space. Recall that mixed formulations for convection-diffusion equations were first introduced in [46], with an analysis of the method in the diffusion dominated regime. More recently mixed methods with Galerkin least squares stabilization has been proposed in [42, 4]. Contrary to these works our numerical scheme is based on a constrained minimization problem in which the difference between the flux variable and the flux evaluated using the primal variable is minimized under the constraint of the conservation law. The method is very robust and was initially introduced for the approximation of ill-posed problems, such as the elliptic Cauchy problem; see [15]. Herein we consider well-posed but possibly indefinite advection-diffusion equations. Compared to [15] we propose a full analysis of the well-posed problem in both the diffusion dominated regime and the advection dominated regime. This also bridges the gap between the analyses of [15] and [16]. Through the latter reference the results herein may be extended to ill-posed advection-diffusion equations and to the best of our knowledge represent the first approach for such problems with a local conservation property. The method is also a close relative to the First order system least square (FOSLS) methods of [18, 19, 23], which, however, do not have local conservation. Another difference compared to these works is that we herein choose to represent the total flux, i.e., both diffusive and advective flux, using the flux variable, which appears to be natural when indefinite problems are considered and the total flux has to be imposed on the boundary. If a stabilization term is introduced for the multiplier, the present method becomes a FOSLS type method with a different flux [15].

Finite element methods for indefinite, or noncoercive, elliptic problems with Neumann boundary conditions were considered first in [20] and more recently in [21, 39, 9] using finite volume and finite element methods. The method proposed herein is a mixed variant of the primal dual stabilized finite element method introduced in [9, 12] for the respective indefinite elliptic and hyperbolic problems, drawing on earlier ideas on H^{-1}-least square methods from [5]. Contrary to those works we herein consider a formulation where the approximation spaces are chosen so that it is inf-sup stable. Hence no stabilizing terms are required. Primal dual methods without stabilization were proposed for the advection-diffusion problem in [22] and for second order PDE in [3, 2], inspired by previous work on discontinuous Petrov–Galerkin methods [30, 29]. Similar ideas have recently been exploited successfully in the context of weak Galerkin methods for elliptic problems on nondivergence form [48], Fokker–Planck equations [47], and the ill-posed elliptic Cauchy problem in [49]. In [40] a method was introduced which is reminiscent of the lowest order version of the method we propose herein. The case of high Péclet number was, however, not considered in [40], so our analysis is likely to be useful for the understanding of the method in [40] in this regime.

1.1. Main results. For the error analysis, in the low Péclet regime, we prove optimal convergence orders for the L^2- and H^1- norms for the primal variable for all polynomial orders. For the analysis we do not use coercivity, but only the stability of the solution, showing the interest of the method for indefinite (or T-coercive [24]) problems. In the high Péclet regime we assume that the data of the adjoint operator satisfies a certain positivity criterion, which is different from the classical one for coercivity. We then prove an error estimate in negative norm and optimal order convergence of the error in the streamline derivative of the primal variable measured in the L^2- norm for smooth solutions.

Numerical results for both the diffusion and advection dominated problems are presented. Optimal convergence is verified on smooth problems and on a problem with reduced regularity due to a corner singularity. We note that for problems with an internal layer only mild and localized oscillations are observed (see Figure 1). However, for problems with underresolved outflow boundary layers the effect of the layer causes global pollution of the solution (see Figure 3). In section 6 we propose two simple strategies to improve the method in this case. More specifically, one method imposes the boundary condition weakly, and the second approach introduces a weighting of the stabilizer such that the oscillation is more "costly" closer to the inflow boundary. This latter variant introduces a notion of upwind direction.

This paper is organized as follows. In section 2, the model problem is presented. The numerical scheme is proposed, and its stability and continuity are analyzed in section 3. In section 4, we prove the error estimation results for both problems with either low or high Péclet numbers. Numerical results are presented in section 5. In section 6 we propose two strategies to improve accuracy in the presence of underresolved outflow boundary layers. Numerical results are also presented to test their effectiveness.

2. The model problem. Let $\Omega \in \mathbb{R}^d$, $d \in \{2,3\}$, be a polygonal/polyhedral domain with boundary $\partial\Omega$ and outward pointing unit normal $\boldsymbol{n}$. We consider the following advection-diffusion equation:

$$\nabla \cdot (\boldsymbol{\beta} u - A\nabla u) + \mu u = f \tag{2.1}$$

with the boundary conditions

$$\begin{aligned} u = g \text{ on } \Gamma_D \quad \text{and} \\ (\boldsymbol{\beta} u - A\nabla u)\cdot \boldsymbol{n} = \psi \text{ on } \Gamma_N, \end{aligned} \tag{2.2}$$

where $\Gamma_D, \Gamma_N \subset \partial\Omega$, $\Gamma_D \cap \Gamma_N = \emptyset$, and $\bar\Gamma_D \cup \bar\Gamma_N = \partial\Omega$. For simplicity, we assume that $\Gamma_D \neq \emptyset$. The data is given by $f \in L^2(\Omega)$, $g \in H^{\frac12}(\Gamma_D)$, $\psi \in H^{-\frac12}(\Gamma_N)$, $A \in \mathbb{R}^{d\times d}$, $\mu \in \mathbb{R}$, and $\boldsymbol{\beta} \in [L^\infty(\Omega)]^d$ with $\boldsymbol{\beta}_\infty := \|\boldsymbol{\beta}\|_{L^\infty(\Omega)}$. For the analysis in the advection dominated case we will strengthen the assumptions on the parameters. Furthermore, we assume that the matrix A is symmetric positive definite. With the smallest eigenvalue $\lambda_{min,A} > 0$ and the largest eigenvalue $\lambda_{\max,A}$, we assume that $\lambda_{\max,A}/\lambda_{min,A}$ is bounded by a moderate constant. The analysis below also holds in the case of variable A and μ that are piecewise differentiable on polyhedral subdomains, provided adjustments are made for loss of regularity in the exact solution.

To write the equations on weak form we introduce the following function spaces:

$$V_{g,D} = \left\{v \in H^1(\Omega) : v = g \text{ on } \Gamma_D\right\} \quad \text{and} \quad V_{0,D} = \left\{v \in H^1(\Omega) : v = 0 \text{ on } \Gamma_D\right\}. \tag{2.3}$$

Consider the weak form: find $u \in V_{g,D}$ such that

$$a(u,v) = l(v) \quad \forall\, v \in V_{0,D} \tag{2.4}$$

with

$$a(u,v) := (\mu u, v)_\Omega + (A\nabla u - \boldsymbol{\beta} u, \nabla v)_\Omega$$

and

$$l(v) := (f,v)_\Omega + \langle \psi, v\rangle_{\Gamma_N},$$

where $(\cdot,\cdot)_w$ and $\langle\cdot,\cdot\rangle_\Gamma$ denote the L^2 inner product on $w \subset \mathbb{R}^d$ and $\Gamma \subset \mathbb{R}^{d-1}$, respectively. When w coincides with the domain Ω the subscript is omitted below. We will only assume that the problem satisfies the Babuska–Lax–Milgram theorem [1], which, in the case of homogenous Dirichlet condition, implies the existence and uniqueness and the following stability estimate:

$$\|u\|_V \le \alpha^{-1}\|l\|_{V'},$$

where $\|\cdot\|_V$ is the H^1-norm, α is the constant of the inf-sup condition, and the dual norm is defined by

$$\|l\|_{V'} := \sup_{\substack{v\in V\\ \|v\|_V=1}} l(v).$$

The constant α is problem dependent, but for the sake of discussion we will here assume that $\alpha = O(\lambda_{min,A})$. Observe that in the case of nonhomogeneous Dirichlet condition we may write $u = u_0 + u_g$, where $u_0 \in V_{0,D}$ is unknown and $u_g \in V_{g,D}$ is a chosen lifting of the boundary data such that $\|u_g\|_V \le \|g\|_{H^{\frac{1}{2}}(\Gamma_D)}$. In that case the stability may be written as

$$\|u_0\|_V \le \alpha^{-1}\|l_g\|_{V'}, \tag{2.5}$$

where $l_g(v) = l(v) - a(u_g, v)$. Clearly the form a satisfies the continuity

$$a(u,v) \le \underbrace{(|\mu| + \lambda_{max,A} + \boldsymbol{\beta}_\infty)}_{=:c_a}\|u\|_V\|v\|_V \tag{2.6}$$

and $1 \le c_a/\alpha$.

3. The mixed finite element framework.

3.1. Some preliminary results. Let $\{\mathcal{T}\}_h$ be a family of conforming, quasi-uniform triangulations of Ω consisting of shape regular simplexes $\mathcal{T} = \{K\}$. The diameter of a simplex K will be denoted by h_K, and the family index h is the mesh parameter defined as the largest diameter of all elements, i.e, $h = \max_{K\in\mathcal{T}}\{h_K\}$. We denote by $\mathcal{F}$ the set of all faces in $\mathcal{T}$, by $\mathcal{F}_I$ the set of all interior faces in $\mathcal{T}$, and by $\mathcal{F}_D$ and $\mathcal{F}_N$ the sets of faces on the respective Γ_D and Γ_N. We assume that the mesh is fitted to the boundary domains Γ_D and Γ_N so that these coincide with element faces. For each $F \in \mathcal{F}$ denote by h_F the diameter of F and by $\boldsymbol{n}_F$ a unit vector normal to F. When F is a boundary face, $\boldsymbol{n}_F$ is fixed to be outer normal to $\partial\Omega$.

Frequently, we will use the notation $a \lesssim b$ meaning $a \le Cb$, where C is a non-essential constant, independent of h. Significant properties of the hidden constant will be highlighted.

We denote the standard H^1-conforming finite element space of order k by

$$V_h^k := \left\{v_h \in H^1(\Omega) : v|_K \in \mathbb{P}_k(K) \quad \forall\, K \in \mathcal{T}\right\},$$

where $\mathbb{P}_k(K)$ denotes the set of polynomials of degree less than or equal to k in the simplex K. Let $i_h : C^0(\bar{\Omega}) \mapsto V_h^k$ be the nodal interpolation. The following approximation estimate is satisfied by i_h; see, e.g., [33]. For $v \in H^\varsigma(\Omega)$, $\varsigma \ge 1$, there holds

$$\|v - i_h v\| + h\|\nabla(v - i_h v)\| \lesssim h^s |v|_{H^{k+1}(\Omega)}, \quad k \ge 1, \tag{3.1}$$

where $s = \min(\varsigma, k+1)$ and $\|\cdot\| = \|\cdot\|_{L^2(\Omega)}$.

For the primal variable we introduce the following spaces:

$$V_{g,D}^k := \left\{ v_h \in V_h^k \,:\, v_h = g_h \text{ on } \Gamma_D \right\} \quad \text{and} \quad V_{0,D}^k := \left\{ v_h \in V_h^k \,:\, v_h = 0 \text{ on } \Gamma_D \right\},$$

where g_h is the nodal interpolation of g (or if g has insufficient smoothness, some other optimal approximation of g) on the trace of Γ_D so that g_h is piecewise polynomial of order k with respect to $\mathcal{F}_D$.

For the flux variable we use the Raviart–Thomas space

$$RT^l := \left\{ \boldsymbol{q}_h \in H(div, \Omega) \,:\, \boldsymbol{q}_h|_K \in \mathbb{P}_l(K)^d \oplus \boldsymbol{x}(\mathbb{P}_l(K) \setminus \mathbb{P}_{l-1}(K)) \quad \forall\, K \in \mathcal{T} \right\}$$

with $\boldsymbol{x} \in \mathbb{R}^d$ being the spatial variable, $l \geq 0$, and $\mathbb{P}_{-1}(K) \equiv \emptyset$. We recall the Raviart–Thomas interpolant $\boldsymbol{R}_h : H^\vartheta(div, \Omega) \mapsto RT^l$, where

$$H^\vartheta(div, \Omega) := \left\{ \boldsymbol{w} \in [H^\vartheta(\Omega)]^d \,:\, \nabla \cdot \boldsymbol{w} \in H^\vartheta(\Omega) \right\},$$

and its approximation properties [33]. For $\boldsymbol{q} \in H^\vartheta(div, \Omega)$, $\vartheta \geq 1$ and $\boldsymbol{R}_h \boldsymbol{q} \in RT^l$, there holds

$$\|\boldsymbol{q} - \boldsymbol{R}_h \boldsymbol{q}\| + h\|\nabla \cdot (\boldsymbol{q} - \boldsymbol{R}_h \boldsymbol{q})\| \lesssim h^r \left(|\nabla \cdot \boldsymbol{q}|_{H^{r-1}(\Omega)} + |\boldsymbol{q}|_{H^r(\Omega)} \right) \lesssim h^r |\boldsymbol{q}|_{H^r(\Omega)}, \tag{3.2}$$

where $r = \min(\vartheta,\, l+1)$.

We also introduce the L^2-projection on the face F of some simplex $K \in \mathcal{T}$,

$$\pi_{F,l} : L^2(F) \mapsto \mathbb{P}_l(F),$$

such that for any $\phi \in L^2(F)$

$$\langle \phi - \pi_{F,l}(\phi),\, p_h \rangle_F = 0 \quad \forall\, p_h \in \mathbb{P}_l(F).$$

Then by assuming that the Neumann data ψ is in $L^2(\Gamma_N)$ we define the discretized Neumann boundary data by its L^2-projection such that for each $F \in \mathcal{F}_N$ we have $\psi_h|_F := \pi_{F,l}(\psi)$. With the satisfaction of the Neumann condition built in, we define

$$RT^l_{\psi,N} = \{ \boldsymbol{q}_h \in RT^l : \boldsymbol{q}_h \cdot \boldsymbol{n} = \psi_h \text{ on } \Gamma_N \}$$

and

$$RT^l_{0,N} = \{ \boldsymbol{q}_h \in RT^l : \boldsymbol{q}_h \cdot \boldsymbol{n} = 0 \text{ on } \Gamma_N \}.$$

For the Lagrange multiplier variable, we introduce the space of functions in $L^2(\Omega)$ that are piecewise polynomial of order m in each element by

$$X_h^m := \{ x_h \in L^2(\Omega) : x_h|_K \in \mathbb{P}_m(K) \quad \forall\, K \in \mathcal{T} \}.$$

We define the L^2-projection $\pi_{X,m} : L^2(\Omega) \mapsto X_h^m$ such that for any $y \in L^2(\Omega)$

$$(y - \pi_{X,m}(y), x_h) = 0 \quad \forall\, x_h \in X_h^m.$$

For functions in X_h^m we define the broken norms,

$$\|x_h\|_h := \left(\sum_{K \in \mathcal{T}} \|x_h\|_K^2 \right)^{\frac{1}{2}} \text{ and } \|x_h\|_{1,h} := \left(\|\nabla x\|_h^2 + \|h^{-\frac{1}{2}} [\![x_h]\!] \|^2_{\mathcal{F}_I \cup \mathcal{F}_D} \right)^{\frac{1}{2}}, \tag{3.3}$$

where $\|h^{-1/2}x_h\|^2_{\mathcal{F}_I\cup\mathcal{F}_D} := \sum_{F\in\mathcal{F}_I\cup\mathcal{F}_D} h_F^{-1}\|x_h\|_F^2$ and

$$[\![x_h]\!]|_F(z) := \begin{cases} \lim_{\epsilon\to 0^+}(x_h(z-\epsilon\boldsymbol{n}_F) - x_h(z+\epsilon\boldsymbol{n}_F)) & \text{for } F\in\mathcal{F}_I,\\ x_h(z) & \text{for } F\in\mathcal{F}_D\cup\mathcal{F}_N. \end{cases}$$

Also recall the discrete Poincaré inequality [6]: there exists $c_P > 0$ such that

$$\|x_h\| \le c_P\|x_h\|_{1,h} \quad \forall\, x_h \in X_h^m, \tag{3.4}$$

which guarantees that $\|\cdot\|_{1,h}$ is a norm.

Given a function $x_h \in X_h^m$ we define a reconstruction, $\boldsymbol{\eta}_h(x_h) \in RT^l_{0,N}$, of the gradient of x_h such that for all $F\in\mathcal{F}_I\cup\mathcal{F}_D$,

$$\langle \boldsymbol{\eta}_h(x_h)\cdot\boldsymbol{n}_F,\, p_h\rangle_F = \left\langle h_F^{-1}[\![x_h]\!],\, p_h\right\rangle_F \quad \forall\, p_h\in\mathbb{P}_l(F), \tag{3.5}$$

and, if $l\ge 1$, for all $K\in\mathcal{T}$,

$$(\boldsymbol{\eta}_h(x_h),\, \boldsymbol{q}_h)_K = -(\nabla x_h,\, \boldsymbol{q}_h)_K \quad \forall\, \boldsymbol{q}_h \in [\mathbb{P}_{l-1}(K)]^d. \tag{3.6}$$

We provide the stability of $\boldsymbol{\eta}_h$ with respect to the data in the following proposition.

PROPOSITION 3.1. *There exists a unique $\boldsymbol{\eta}_h \in RT^l_{0,N}$ such that* (3.5)–(3.6) *hold. Moreover, $\boldsymbol{\eta}_h$ satisfies the following stability estimate:*

$$\|\boldsymbol{\eta}_h\| \le C_{ds}\left(\|\pi_{X,l-1}\nabla x_h\|_h^2 + \|h^{-\frac{1}{2}}\pi_{F,l}([\![x_h]\!])\|^2_{\mathcal{F}_I\cup\mathcal{F}_D}\right)^{\frac{1}{2}}, \tag{3.7}$$

where $C_{ds} > 0$ is a constant depending only on the element shape regularity.

Proof. We refer to [15] for the proof. □

We will also frequently use the following inverse and trace inequalities:

$$\|\nabla v\|_K \lesssim h^{-1}\|v\|_K \quad \forall\, v\in\mathbb{P}_k(K) \tag{3.8}$$

and

$$\|v\|_{\partial K} \lesssim h^{-\frac{1}{2}}\|v\|_K + h^{\frac{1}{2}}\|\nabla v\|_K \quad \forall\, v\in H^1(K). \tag{3.9}$$

For a proof of (3.8) we refer to Ciarlet [25], and for (3.9) see, e.g., Monk and Süli [44].

3.2. The finite element method. The problem takes the form of finding the critical point of a Lagrangian $\mathcal{L} : (v_h, \boldsymbol{q}_h, x_h) \in V^k_{g,D}\times RT^l_{\psi,N}\times X_h^m \mapsto \mathbb{R}$ defined by

$$\mathcal{L}[v_h, \boldsymbol{q}_h, x_h] := \frac{1}{2}s[(v_h,\boldsymbol{q}_h),(v_h,\boldsymbol{q}_h)] + b(\boldsymbol{q}_h, v_h, x_h) - (f, x_h). \tag{3.10}$$

Here $x_h\in X_h^m$ is the Lagrange multiplier, $s(\cdot,\cdot)$ denotes the constitutive law on least squares form, here the equation for the flux,

$$s[(u,\boldsymbol{p}),(v,\boldsymbol{q})] := (\boldsymbol{\beta}u - A\nabla u - \boldsymbol{p},\, \boldsymbol{\beta}v - A\nabla v - \boldsymbol{q}), \tag{3.11}$$

and $b(\cdot,\cdot)$ is the linear form defining the partial differential equation, in our case the conservation law,

$$b(\boldsymbol{q}, v, x) := (\nabla\cdot\boldsymbol{q} + \mu v,\, x).$$

By computing the Euler–Lagrange equations of (3.10) we obtain the following linear system: find $(u_h, \boldsymbol{p}_h, z_h) \in V^k_{g,D} \times RT^l_{\psi,N} \times X^m_h$ such that

$$s[(u_h, \boldsymbol{p}_h), (v_h, \boldsymbol{q}_h)] + b(\boldsymbol{q}_h, v_h, z_h) = 0, \tag{3.12}$$

$$b(\boldsymbol{p}_h, u_h, x_h) - (f, x_h) = 0 \tag{3.13}$$

for all $(v_h, \boldsymbol{q}_h, x_h) \in V^k_{0,D} \times RT^l_{0,N} \times X^m_h$. The system (3.12)–(3.13) is of the same form as that proposed in [10, 13] but without adjoint stabilization. Therefore, to ensure that the system is well-posed the spaces $V^k_h \times RT^l \times X^m_h$ must be carefully balanced. Herein we will restrict the discussion to the equal order case $k = l = m$ that is stable without further stabilization. The arguments can be extended to other choices of spaces provided suitable extra stabilizing terms are added (see [15] for details).

Observe that the s-operator in (3.11) connects the flux and the primal variables and, more precisely, brings $\boldsymbol{p}_h$ and $\boldsymbol{\beta} u_h - A\nabla u_h$ to be close. In the low Péclet regime this introduces an effect similar to the penalty on the gradient of the primal variable used in [9]. In the high Péclet regime, on the other hand, the stability of the conservation form of the convective derivative is obtained by the strong control of the conservation law obtained through (3.13) through an inf-sup argument.

Remark 3.1. The constrained-minimization problem introduces an auxiliary variable, i.e., the Lagrange multiplier, which for stability reasons must be chosen as the discontinuous counterpart of the discretization space for the primal variable (unless stabilization is applied; see [15]). This results in a system with a substantially larger number of degrees of freedom than that of the standard Galerkin, the classical mixed method [32, 45], and the least square method [18]. Nevertheless, it is possible to reduce the system used in the iterative solver to a positive definite symmetric matrix where the Lagrange multiplier has been eliminated. This is achieved by iterating on a least square formulation, the solution of which is not locally mass conservative but has similar approximation properties. The number of degrees of freedom of the reduced system consists only of those of the primal and flux variables. For a detailed discussion of this approach we refer to [15].

3.3. Approximation, inf-sup condition, and continuity. For the analysis we introduce the following triple norms on $H^1(\Omega) \times H(div, \Omega)$:

$$|||(v, \boldsymbol{q})|||_{-1} := \left(s[(v, \boldsymbol{q}), (v, \boldsymbol{q})] + \|h(\nabla \cdot \boldsymbol{q} + \mu v)\|^2\right)^{\frac{1}{2}}, \tag{3.14}$$

$$|||(v, \boldsymbol{q})|||_\sharp := |||(v, \boldsymbol{q})|||_{-1} + \|\mu v\| + \|h^{\frac{1}{2}} \boldsymbol{q}\|_{\mathcal{F}} + \|\boldsymbol{q}\|. \tag{3.15}$$

To quantify the dependence of the physical parameters in the bounds below we introduce the factor $c_u := \boldsymbol{\beta}_\infty h + \|A\|_\infty + c_P|\mu|h$. Here c_P is the constant of inequality (3.4).

LEMMA 3.1 (approximation). *For any $v \in H^\varsigma(\Omega)$ and $\boldsymbol{q} \in [H^\vartheta(\Omega)]^d$ the following approximation property holds:*

$$|||(v - i_h v, \boldsymbol{q} - R_h \boldsymbol{q})|||_{-1} \le |||(v - i_h v, \boldsymbol{q} - R_h \boldsymbol{q})|||_\sharp \lesssim c_u h^{s-1}|v|_{H^s(\Omega)} + h^r|\boldsymbol{q}|_{H^r(\Omega)}, \tag{3.16}$$

where $s = \min(\varsigma, k+1)$ and $r = \min(\vartheta, l+1)$.

Proof. Applying the triangle inequality and the approximation properties (3.1) and (3.2) gives

$$|||(v - i_h v, \boldsymbol{q} - R_h \boldsymbol{q})|||_{-1} \lesssim \left(\boldsymbol{\beta}_\infty h + \|A\|_\infty + |\mu| h^2\right) h^{s-1} |v|_{H^s(\Omega)} + h^r |\boldsymbol{q}|_{H^r(\Omega)}. \tag{3.17}$$

To estimate the remaining terms note that the trace inequality (3.9) implies

$$\left\| h^{1/2} (\boldsymbol{q} - R_h \boldsymbol{q}) \right\|_{\mathcal{F}} \lesssim \|\boldsymbol{q} - R_h \boldsymbol{q}\| + h \|\nabla (\boldsymbol{q} - R_h \boldsymbol{q})\|,$$

which, combined with the approximation properties, gives

$$\|\mu(v - i_h v)\| + \left\| h^{\frac{1}{2}} (\boldsymbol{q} - R_h \boldsymbol{q}) \right\|_{\mathcal{F}} + \|\boldsymbol{q} - R_h \boldsymbol{q}\| \lesssim |\mu| h^s |v|_{H^s(\Omega)} + h^r |\boldsymbol{q}|_{H^r(\Omega)}. \tag{3.18}$$

(3.16) is then a direct consequence of (3.17) and (3.18). This completes the proof of the lemma. ∎

To facilitate the analysis we rewrite the system (3.12)–(3.13) in the following compact form: find $(u_h, \boldsymbol{p}_h, z_h) \in V^k_{g,D} \times RT^l_{\psi,N} \times X^m_h$ such that

$$\mathcal{A}[(u_h, \boldsymbol{p}_h, z_h), (v_h, \boldsymbol{q}_h, x_h)] = l_h(x_h) \quad \forall\, (v_h, \boldsymbol{q}_h, x_h) \in V^k_{0,D} \times RT^l_{0,N} \times X^m_h, \tag{3.19}$$

where

$$\mathcal{A}[(u_h, \boldsymbol{p}_h, z_h), (v_h, \boldsymbol{q}_h, x_h)] = s[(u_h, \boldsymbol{p}_h), (v_h, \boldsymbol{q}_h)] + b(\boldsymbol{q}_h, v_h, z_h) + b(\boldsymbol{p}_h, u_h, x_h)$$

and

$$l_h(x_h) = (f, x_h).$$

Note that for the exact solution, $(u, \boldsymbol{p})$, there holds

$$\mathcal{A}[(u, \boldsymbol{p}, 0), (v_h, \boldsymbol{q}_h, x_h)] = l_h(x_h) \quad \forall\, (v_h, \boldsymbol{q}_h, x_h) \in V^k_{0,D} \times RT^l_{0,N} \times X^m_h. \tag{3.20}$$

Proposition 3.2 (inf-sup condition). *Let $k = l = m$ in* (3.19)*. Then there exists $\alpha_c > 0$ such that, for all $(v_h, \boldsymbol{q}_h, x_h) \in V^k_{0,D} \times RT^k_{0,N} \times X^k_h$, there exists $(\tilde{v}_h, \tilde{\boldsymbol{q}}_h, \tilde{x}_h) \in V^k_{0,D} \times RT^k_{0,N} \times X^k_h$ satisfying*

$$\alpha_c \left(|||(v_h, \boldsymbol{q}_h)|||^2_{-1} + \|x_h\|^2_{1,h} \right) \le \mathcal{A}[(v_h, \boldsymbol{q}_h, x_h), (\tilde{v}_h, \tilde{\boldsymbol{q}}_h, \tilde{x}_h)] \tag{3.21}$$

and

$$|||(\tilde{v}_h, \tilde{\boldsymbol{q}}_h)|||_{-1} + \|\tilde{x}_h\|_{1,h} \lesssim |||(v_h, \boldsymbol{q}_h)|||_{-1} + \|x_h\|_{1,h}. \tag{3.22}$$

Proof. Define $\boldsymbol{\eta}_h = \boldsymbol{\eta}_h(x_h) \in RT^k_{0,N}$ by taking $l = m = k$ in (3.5)–(3.6) and $\xi_h := h^2(\nabla \cdot \boldsymbol{q}_h + \mu v_h) \in X^k_h$. We claim that, by choosing $\tilde{v}_h = v_h \in V^k_{0,D}$, $\tilde{\boldsymbol{q}}_h = \boldsymbol{q}_h + \epsilon \boldsymbol{\eta}_h \in RT^k_{0,N}$ and $\tilde{x}_h = -x_h + \xi_h \in X^k_h$, there holds (3.21) and (3.22), where ϵ is to be determined later.

By the above definitions, we have

$$\begin{aligned}
&\mathcal{A}[(v_h, \boldsymbol{q}_h, x_h), (v_h, \boldsymbol{q}_h + \epsilon \boldsymbol{\eta}_h, -x_h + \xi_h)] \\
&\quad = (\boldsymbol{\beta} v_h - A \nabla v_h - \boldsymbol{q}_h, \boldsymbol{\beta} v_h - A \nabla v_h - \boldsymbol{q}_h - \epsilon \boldsymbol{\eta}_h) + (\nabla \cdot (\boldsymbol{q}_h + \epsilon \boldsymbol{\eta}_h) + \mu v_h, x_h) \\
&\qquad + (\nabla \cdot \boldsymbol{q}_h + \mu v_h, -x_h + \xi_h) \\
&\quad = \|\boldsymbol{\beta} v_h - A \nabla v_h - \boldsymbol{q}_h\|^2 + \|h(\nabla \cdot \boldsymbol{q}_h + \mu v_h)\|^2 \\
&\qquad - \epsilon (\boldsymbol{\beta} v_h - A \nabla v_h - \boldsymbol{q}_h, \boldsymbol{\eta}_h) + \epsilon (\nabla \cdot \boldsymbol{\eta}_h, x_h).
\end{aligned} \tag{3.23}$$

For the last term, it follows from integration by parts, (3.5), (3.6), and the facts that $\boldsymbol{\eta}_h \cdot \boldsymbol{n} = 0$ on Γ_N, $\nabla x_h|_K \in \mathbb{P}_{k-1}(K)^d$, and $x_h|_F \in \mathbb{P}_k(F)$ that

$$(\nabla \cdot \boldsymbol{\eta}_h, x_h) = \sum_{K \in \mathcal{T}} (-(\boldsymbol{\eta}_h, \nabla x_h)_K + \langle \boldsymbol{\eta}_h \cdot \boldsymbol{n}_K, x_h \rangle_{\partial K}) = \|\nabla x_h\|^2 + \sum_{F \in \mathcal{F}_I \cup \mathcal{F}_D} \|h^{-\frac{1}{2}} [\![x_h]\!]\|_F^2,$$

which, combined with (3.23), the Cauchy–Schwarz inequality, and (3.7), gives

$$\begin{aligned} &\mathcal{A}[(v_h, \boldsymbol{q}_h, x_h), (v_h, \boldsymbol{q}_h + \epsilon \boldsymbol{\eta}_h, -x_h + \xi_h)] \\ &\quad \geq \|\boldsymbol{\beta} v_h - A\nabla v_h - \boldsymbol{q}_h\|^2 + \|h(\nabla \cdot \boldsymbol{q}_h + \mu v_h)\|^2 - \frac{1}{4}\|\boldsymbol{\beta} v_h - A\nabla v_h - \boldsymbol{q}_h\|^2 \\ &\qquad - \epsilon^2 \|\boldsymbol{\eta}_h\|^2 + \epsilon \left(\|\nabla x_h\|^2 + \sum_{F \in \mathcal{F}_I \cup \mathcal{F}_D} \|h^{-\frac{1}{2}} [\![x_h]\!]\|_F^2 \right) \\ &\quad \geq \frac{3}{4}\|\boldsymbol{\beta} v_h - A\nabla v_h - \boldsymbol{q}_h\|^2 + \|h(\nabla \cdot \boldsymbol{q}_h + \mu v_h)\|^2 + \epsilon(1 - \epsilon C_{ds}^2)\|x_h\|_{1,h}^2. \end{aligned} \tag{3.24}$$

(3.21) is then a direct result of (3.24) by choosing $\epsilon = \frac{1}{2} C_{ds}^{-2}$ and $\alpha_c = \min(\frac{3}{4}, \frac{1}{2}\epsilon)$.

To prove (3.22), first applying the triangle inequality gives

$$|||(\tilde{v}_h, \tilde{\boldsymbol{q}}_h)|||_{-1} + \|\tilde{x}_h\|_{1,h} \leq |||(v_h, \boldsymbol{q}_h)|||_{-1} + \|x_h\|_{1,h} + |||(0, \epsilon \boldsymbol{\eta}_h)|||_{-1} + \|\xi_h\|_{1,h}. \tag{3.25}$$

Then applying the trace and inverse inequalities and (3.7) yields

$$|||(0, \epsilon \boldsymbol{\eta}_h)|||_{-1} = \epsilon\, (\|\boldsymbol{\eta}_h\| + \|h\nabla \cdot \boldsymbol{\eta}_h\|) \lesssim \|\boldsymbol{\eta}_h\| \lesssim \|x_h\|_{1,h} \tag{3.26}$$

and

$$\|\xi_h\|_{1,h} \lesssim h^{-1}\|\xi_h\| = \|h(\nabla \cdot \boldsymbol{q}_h + \mu v_h)\| \leq |||(v_h, \boldsymbol{q}_h)|||_{-1}. \tag{3.27}$$

Finally, combining (3.25)–(3.27) results in (3.22). This completes the proof of the proposition. ∎

Proposition 3.3 (existence and uniqueness). *The linear system defined by* (3.19) *admits a unique solution* $(u_h, \boldsymbol{p}_h, z_h) \in V_{g,D}^k \times RT_{\psi,N}^k \times X_h^k$.

Proof. In order to prove the invertibility of the square linear system it is equivalent to prove the uniqueness. Assume that there exist two sets of solutions, $(u_{1,h}, \boldsymbol{p}_{1,h}, z_{1,h})$ and $(u_{2,h}, \boldsymbol{p}_{2,h}, z_{2,h})$, both in $V_{g,D}^k \times RT_{\psi,N}^k \times X_h^k$. We then have that for all $(v_n, \boldsymbol{q}_h, x_h)$ in the space of $V_{0,D}^k \times RT_{0,N}^k \times X_h^k$ there holds

$$\mathcal{A}[(u_{1,h} - u_{2,h}, \boldsymbol{p}_{1,h} - \boldsymbol{p}_{2,h}, z_{1,h} - z_{2,h}), (v_h, \boldsymbol{q}_h, x_h)] = 0.$$

By Proposition 3.2, the following must be true:

$$\|(u_{1,h} - u_{2,h}, \boldsymbol{p}_{1,h} - \boldsymbol{p}_{2,h})\|_{-1} + \|z_{1,h} - z_{2,h}\|_{1,h} = 0,$$

which immediately implies

$$z_{1,h} = z_{2,h} \quad \text{and} \quad \nabla \cdot (\boldsymbol{\beta}(u_{1,h} - u_{2,h}) - A\nabla(u_{1,h} - u_{2,h})) + \mu(u_{1,h} - u_{2,h}) = 0.$$

Since (2.1)–(2.2) admits a unique trivial solution for zero datum we conclude that $u_{1,h} = u_{2,h}$ and, hence, $\boldsymbol{p}_{1,h} = \boldsymbol{p}_{2,h}$. This completes the proof of the proposition. ∎

We end this section by proving the continuity of the bilinear form. Define

$$H_{0,N}(div,\Omega) = \{\boldsymbol{q} \in H(div,\Omega),\ \boldsymbol{q}\cdot\boldsymbol{n} = 0 \text{ on } \Gamma_N\}$$

and

$$H_{\psi,N}(div,\Omega) = \{\boldsymbol{q} \in H(div,\Omega),\ \boldsymbol{q}\cdot\boldsymbol{n} = \psi \text{ on } \Gamma_N\}.$$

PROPOSITION 3.4 (continuity). *For all $(v,\boldsymbol{q}) \in H^1(\Omega)\times H_{0,N}(div,\Omega)$ and for all $(v_h,\boldsymbol{q}_h,x_h) \in V_h^k \times RT^l \times X_h^m$ there holds*

$$\mathcal{A}[(v,\boldsymbol{q},0),(v_h,\boldsymbol{q}_h,x_h)] \le |||(v,\boldsymbol{q})|||_\sharp\,(|||(v_h,\boldsymbol{q}_h)|||_{-1} + \|x_h\|_{1,h}). \tag{3.28}$$

Proof. The inequality (3.28) follows by first using the Cauchy–Schwarz inequality in the symmetric part of the formulation,

$$s[(v,\boldsymbol{q}),(v_h,\boldsymbol{q}_h)] \lesssim s[(v,\boldsymbol{q}),(v,\boldsymbol{q})]^{\frac{1}{2}} s[(v_h,\boldsymbol{q}),(v_h,\boldsymbol{q})]^{\frac{1}{2}}.$$

For the remaining term we use the divergence formula elementwise to obtain

$$(\nabla\cdot\boldsymbol{q} + \mu v, x_h) = \sum_{K\in\mathcal{T}} -(\boldsymbol{q},\nabla x_h)_K + \sum_{F\in\mathcal{F}_I\cup\mathcal{F}_D} \langle \boldsymbol{q}\cdot\boldsymbol{n}_F, [\![x_h]\!]\rangle_F + (\mu v, x_h).$$

(3.28) then follows by applying the Cauchy–Schwarz inequality and (3.4). This completes the proof of the proposition. ∎

4. Error estimation. In this section we will prove optimal error estimates for smooth solutions, both in the diffusion dominated and the advection dominated regimes. When the diffusion dominates we prove optimal error estimates in both the H^1- and L^2-norms under very mild stability assumptions on the continuous problem. In this analysis constants may blow up as the Péclet number becomes large.

For dominating advection we need to make an assumption on the problem data to prove an error estimate in the H^{-1}-norm. This is then used to prove an estimate that is optimal for the error in the divergence of the flux, computed using the primal variable. In this case this corresponds to the convective derivative on conservation form. However, we cannot improve on the order for the L^2-error as for typical residual based stabilized finite element methods. In this part constants remain bounded as the Péclet number becomes large.

4.1. Error estimate for the residual. First we prove the optimal convergence result for the residual, i.e., the optimal convergence for the triple norm (3.14). This estimate will then be of use in both the high and low Péclet regimes.

LEMMA 4.1 (estimate of residual). *Assume that $(u,\boldsymbol{p})$ is the solution to (2.4) with $u \in H^\varsigma \cap V_{g,D}(\Omega)$, $\boldsymbol{p}\in[H^\vartheta(\Omega)]^d \cap H_{\psi,N}(div,\Omega)$ and that $(u_h,\boldsymbol{p}_h,z_h)\in V_{g,D}^k \times RT_{\psi,N}^k \times X_h^k$ is the solution of (3.19). Then there holds*

$$|||(u-u_h,\boldsymbol{p}-\boldsymbol{p}_h)|||_{-1} + \|z_h\|_{1,h} \lesssim c_u h^{s-1}|u|_{H^s(\Omega)} + h^r|\boldsymbol{p}|_{H^r(\Omega)}, \tag{4.1}$$

where $s = \min(\varsigma,\, k+1) =$ and $r = \min(\vartheta,\, k+1)$.

Proof. Firstly, applying the triangle inequality gives

$$|||(u-u_h,\boldsymbol{p}-\boldsymbol{p}_h)|||_{-1} \le |||(u-i_h u,\boldsymbol{p}-R_h\boldsymbol{p})|||_{-1} + |||(u_h-i_h u,\boldsymbol{p}_h-R_h\boldsymbol{p})|||_{-1}. \tag{4.2}$$

Note that $u_h - i_h u \in V_{0,D}^k$ and $\boldsymbol{p}_h - R_h\boldsymbol{p} \in RT_{0,N}^k$. Then by Proposition 3.2 there exists $(v_h, \boldsymbol{q}_h, w_h) \in V_{0,D}^k \times RT_{0,N}^k \times X_h^k$ such that

$$\begin{aligned}
&|||(u_h - i_h u, \boldsymbol{p}_h - R_h\boldsymbol{p})|||_{-1}^2 + \|z_h\|_{1,h}^2 \\
&\quad\lesssim \mathcal{A}[(u_h - i_h u, \boldsymbol{p}_h - R_h\boldsymbol{p}, z_h)), (v_h, \boldsymbol{q}_h, w_h)] = \mathcal{A}[(u - i_h u, \boldsymbol{p} - R_h\boldsymbol{p}, 0), (v_h, \boldsymbol{q}_h, w_h)] \\
&\quad\lesssim |||(u - i_h u, \boldsymbol{p} - R_h\boldsymbol{p})|||_\sharp \left(|||(u_h - i_h u, \boldsymbol{p}_h - R_h\boldsymbol{p})|||_{-1} + \|z_h\|_{1,h}\right).
\end{aligned}$$

The last equality and inequality follows from (3.19), (3.20), and Proposition 3.4. Therefore, we immediately have that

$$|||(u_h - i_h u, \boldsymbol{p}_h - R_h\boldsymbol{p})|||_{-1} \lesssim |||(u - i_h u, \boldsymbol{p} - R_h\boldsymbol{p})|||_\sharp$$

which, combining with (3.16) and (4.2), implies (4.1). This completes the proof of the lemma. ∎

Observe that the hidden constant in (4.1) has no inverse powers of the diffusivity. Hence, recalling the definition of c_u, we have the following corollary.

COROLLARY 4.1. *Under the same assumptions as in Lemma* 4.1*, if* $\|A\|_\infty << h$*,* $\boldsymbol{\beta}_\infty = O(1)$*,* $|\mu| = O(1)$*,* $s = r = k+1$*, there holds*

$$|||(u - u_h, \boldsymbol{p} - \boldsymbol{p}_h)|||_{-1} + \|z_h\|_{1,h} \lesssim h^{k+1}|u|_{H^{k+1}(\Omega)} + h^{k+1}|\boldsymbol{p}|_{H^{k+1}(\Omega)} \tag{4.3}$$

with hidden constant $O(1)$.

4.2. Error estimates in the diffusion dominated regime. In this subsection we provide results for the error estimation in the diffusion dominated regime, i.e., $\frac{\boldsymbol{\beta}_\infty}{\lambda_{min,A}}$ is of order 1, where $\lambda_{min,A}$ is the smallest eigenvalue of A.

PROPOSITION 4.1 (H^1-norm estimate). *Assume that* $(u, \boldsymbol{p})$ *is the solution to* (2.4)*,* $u \in H^\varsigma(\Omega) \cap V_{g,D}(\Omega)$*,* $\varsigma > 1$*, and* $\boldsymbol{p} \in [H^\vartheta(\Omega)]^d \cap H_{\psi,N}(div, \Omega)$*,* $\vartheta > 0$*, and that* $(u_h, \boldsymbol{p}_h, z_h) \in V_{g,D}^k \times RT_{\psi,N}^k \times X_h^k$ *is the solution of* (3.19)*. Then the following estimate holds:*

$$\|u - u_h\|_V \lesssim \alpha^{-1}\left((c_a + c_u)h^{s-1}|u|_{H^s(\Omega)} + h^r\left(|\boldsymbol{p}|_{H^r(\Omega)} + |\psi|_{H^{r-1/2}(\Gamma_N)}\right)\right), \tag{4.4}$$

where $s = \min(\varsigma, k+1)$ *and* $r = \min(\vartheta, k+1)$.

Remark 4.1. Using the assumption $\alpha = O(\lambda_{min,A})$ we see that the constant in the above estimate satisfies

$$\alpha^{-1}(c_a + c_u) \sim \lambda_{min,A}^{-1}(|\mu| + \lambda_{max,A} + \boldsymbol{\beta}_\infty).$$

Since the expression of the right-hand side blows up as $\lambda_{min,A}$ goes to zero, the above estimation is valid only for the diffusion dominated problem.

Proof. To avoid using coercivity arguments, our starting point for the error analysis below is the stability estimate (2.5). Let $e = u - u_h$. We note that e is a solution to (2.4) with the right-hand side linear operator being $r(v) := l(v) - a(u_h, v)$, i.e.,

$$a(e, v) = r(v).$$

Now apply the decomposition $e = e_0 + e_g$ such that $e_g|_{\Gamma_D} = e|_{\Gamma_D}$ and $\|e_g\|_V \lesssim \|e\|_{H^{1/2}(\Gamma_D)}$. It then follows from (2.5) that

$$\|e_0\|_V \le \alpha^{-1} \sup_{v \in V} \frac{r(v) - a(e_g, v)}{\|v\|_V} \le \alpha^{-1}(\|r\|_{V'} + c_a\|e_g\|_V).$$

Hence

$$\|e\|_V \le \|e_0\|_V + \|e_g\|_V \le \alpha^{-1}(\|r\|_{V'} + c_a\|e_g\|_V) + \|e_g\|_V.$$

For the term $\|e_g\|_V$, by definition and a standard trace inequality, we have

$$\|e_g\|_V \le C\|u - i_h u\|_{H^{\frac{1}{2}}(\Gamma_D)} \le C\|u - i_h u\|_V \le Ch^{s-1}|u|_{H^s(\Omega)}. \tag{4.5}$$

To prove the bound on $\|r\|_{V'}$ we recall that

$$\|r\|_{V'} = \sup_{\substack{v\in V\\ v=0 \text{ on } \Gamma_D}} \frac{a(u-u_h, v)}{\|v\|_V}.$$

Then by integration by parts, (3.13), and Cauchy–Schwarz inequality, we have

(4.6)

$$\begin{aligned}
a(u-u_h, v) &= l(v) - a(u_h, v)\\
&= (f, v) + \langle \psi, v\rangle_{\Gamma_N} - (A\nabla u_h - \beta u_h, \nabla v) - (\mu u_h, v)\\
&= (f - \nabla\cdot \boldsymbol{p}_h - \mu u_h, v) + \langle \psi + \boldsymbol{p}_h\cdot n, v\rangle_{\Gamma_N} - (\boldsymbol{p}_h - \beta u_h + A\nabla u_h, \nabla v)\\
&= (f - \nabla\cdot \boldsymbol{p}_h - \mu u_h, v - \pi_{X,0} v) + (\boldsymbol{p} - \boldsymbol{p}_h - \beta(u-u_h) + A\nabla(u-u_h), \nabla v)\\
&\quad + \langle \psi - \psi_h, v - \pi_{F,0} v\rangle_{\Gamma_N}\\
&\lesssim |||(\boldsymbol{p}-\boldsymbol{p}_h, u-u_h)|||_{-1}\|\nabla v\| + \|h^{\frac{1}{2}}(\psi - \psi_h)\|_{\Gamma_N}\|\nabla v\|
\end{aligned}$$

which, combined with (4.5), (4.1), and the following observation (see, e.g., Lemma 5.2 of [34]),

$$\left\|h^{\frac{1}{2}}(\psi - \psi_h)\right\|_{\Gamma_N} \lesssim h^r|\psi|_{H^{r-1/2}(\Gamma_N)},$$

gives (4.4). This completes the proof of the proposition. ∎

In the remaining part of this subsection we will focus on the convergence of the L^2-norm error in the primal variable. For simplicity we here restrict the discussion to the case of a convex polygonal domain Ω, smooth solutions ($s = k+1$, $r = k$ in Lemma 4.1), and homogeneous Dirichlet condition. We first prove the convergence result for the L^2-norm of the Lagrange multiplier.

Proposition 4.2. *Assume that Ω is convex polygonal, $u \in H^1_0(\Omega)\cap H^{k+1}(\Omega)$, and $\boldsymbol{p} \in [H^k(\Omega)]^d$. Let z_h be the Lagrange multiplier of the system* (3.19). *We have the following error estimate:*

$$\|z_h\| \lesssim h^{k+1}\left(|u|_{H^{k+1}(\Omega)} + |\boldsymbol{p}|_{H^k(\Omega)}\right). \tag{4.7}$$

Proof. Let ϕ be the solution such that

$$\nabla\cdot(\boldsymbol{\beta}\phi - A\nabla\phi) + \mu\phi = z_h$$

with boundary condition $\phi = 0$ on $\partial\Omega$. Then by the well-posedness assumption on (2.1) and the assumption on Ω we have the following stability result:

$$\|\phi\|_{H^2(\Omega)} \lesssim \|z_h\|. \tag{4.8}$$

Let $\boldsymbol{q} = \boldsymbol{\beta}\phi - A\nabla\phi$. By adding and subtracting suitable interpolates we have

$$\|z_h\|^2 = (z_h, \nabla\cdot(\boldsymbol{q} - R_h\boldsymbol{q}) + \mu(\phi - i_h\phi)) + (z_h, \nabla\cdot R_h\boldsymbol{q} + \mu i_h\phi). \tag{4.9}$$

For the first term in (4.9) using the elementwise divergence theorem, the facts that

$$\langle z_h,\, (\boldsymbol{q} - R_h\boldsymbol{q}) \cdot \boldsymbol{n}_K \rangle_F = 0 \quad \forall K \in \mathcal{T}, \forall F \subset \partial K$$

and that

$$\|\boldsymbol{q} - R_h\boldsymbol{q}\| \lesssim h\|\phi\|_{H^2(\Omega)} \lesssim h\|z_h\| \quad \text{and} \quad \|\phi - i_h\phi\| \lesssim h^2\|\phi\|_{H^2(\Omega)} \lesssim h^2\|z_h\|,$$

and (3.4) gives

$$(z_h, \nabla \cdot (\boldsymbol{q} - R_h\boldsymbol{q}) + \mu(\phi - i_h\phi)) \lesssim h(1 + |\mu|h)\|z_h\|_{1,h}\|z_h\|. \tag{4.10}$$

For the second term in (4.9) we first apply (3.12) with $\boldsymbol{q}_h = R_h(\boldsymbol{q}) \in RT^k$ and $v_h = i_h\varphi \in V^k_{0,D}$ with $\Gamma_N = \emptyset$; then applying the Cauchy–Schwarz inequality, (3.1), and (3.2), we have that

$$\begin{aligned}(z_h, \nabla \cdot R_h\boldsymbol{q} + \mu i_h\phi) &= -(\boldsymbol{\beta} u_h - A\nabla u_h - \boldsymbol{p}_h, \boldsymbol{\beta} i_h\phi - A\nabla(i_h\phi) - R_h\boldsymbol{q})\\ &\lesssim \|\boldsymbol{\beta} u_h - A\nabla u_h - \boldsymbol{p}_h\|\,(\boldsymbol{\beta}_\infty\|\phi - i_h\phi\| + \|A\|_\infty\|\nabla(\phi - i_h\phi)\| + \|\boldsymbol{q} - R_h\boldsymbol{q}\|)\\ &\lesssim \|(\boldsymbol{\beta} u_h - A\nabla u_h - \boldsymbol{p}_h)\|\left(\boldsymbol{\beta}_\infty h^2 + \|A\|_\infty h + h\right)\|\phi\|_{H^2(\Omega)}\\ &\lesssim h\|(\boldsymbol{\beta} u_h - A\nabla u_h - \boldsymbol{p}_h)\|\|z_h\|.\end{aligned} \tag{4.11}$$

(4.7) is then a direct consequence of (4.10), (4.11), and (4.1). This completes the proof of the proposition. ∎

We now proceed to prove the error estimation of the primal variable in the L^2-norm. To estimate the error of the primal variable in the L^2-norm we require that the adjoint problem is well-posed and satisfies a shift theorem for the H^2-norm.

Assumption 4.1. Consider the adjoint problem for (2.1). For each $\psi \in L^2(\Omega)$, we assume that the data is such that the following adjoint problem admits a unique solution, using Fredholm's alternative,

$$-\nabla \cdot A\nabla\varphi - \boldsymbol{\beta} \cdot \nabla\varphi + \mu\varphi = \psi \text{ in } \Omega \tag{4.12}$$

with $\varphi|_{\partial\Omega} = 0$. Furthermore, the following regularity result holds true:

$$\|\varphi\|_{H^2(\Omega)} \lesssim \|\psi\|. \tag{4.13}$$

PROPOSITION 4.3. *Let* $u \in H^{k+1}(\Omega) \cap H^1_0(\Omega)$, $\boldsymbol{p} \in [H^k(\Omega)]^d$, *and* $(u_h, \boldsymbol{p}_h, z_h)$ *be the solution of* (3.12)–(3.13). *Under Assumption* 4.1 *we have*

$$\|u - u_h\| \lesssim h^{k+1}\left(|u|_{H^{k+1}(\Omega)} + |\boldsymbol{p}|_{H^k(\Omega)}\right). \tag{4.14}$$

Proof. Let φ be the solution of the dual problem (4.12) with the right-hand side being $e := u - u_h$. Then by integration by parts and the assumption that $\varphi = 0$ on $\partial\Omega$, we have

$$\begin{aligned}\|e\|^2 &= (f, \varphi) + (u_h, \nabla \cdot A\nabla\varphi + \boldsymbol{\beta} \cdot \nabla\varphi - \mu\varphi)\\ &= (f - \nabla \cdot \boldsymbol{p}_h - \mu u_h, \varphi) + (\boldsymbol{\beta} u_h - A\nabla u_h - \boldsymbol{p}_h, \nabla\varphi).\end{aligned} \tag{4.15}$$

The first term can be estimated by applying (3.13) and the Cauchy–Schwarz inequality:

$$\begin{aligned}(f - \nabla \cdot \boldsymbol{p}_h - \mu u_h, \varphi) &= (f - \nabla \cdot \boldsymbol{p}_h - \mu u_h, \varphi - \pi_{X,k}\varphi)\\ &\lesssim h|||(u - u_h, \boldsymbol{p} - \boldsymbol{p}_h)|||_{-1}\|\varphi\|_{H^2(\Omega)}.\end{aligned} \tag{4.16}$$

To estimate the second term we apply (3.12) with $\boldsymbol{q}_h = R_h(\nabla\varphi) \in RT^k$ and the fact that $\nabla \cdot (R_h(\nabla\varphi)) = \pi_{X,k}\triangle\varphi$:

$$\begin{aligned}
(4.17)\qquad &(\boldsymbol{\beta} u_h - A\nabla u_h - \boldsymbol{p}_h, \nabla\varphi) \\
&= -(z_h, \nabla\cdot(R_h(\nabla\varphi))) + (\boldsymbol{\beta} u_h - A\nabla u_h - \boldsymbol{p}_h, (\nabla\varphi - R_h(\nabla\varphi))) \\
&\lesssim \|z_h\|\|\pi_{X,k}\triangle\varphi\| + h\|\boldsymbol{\beta} u_h - A\nabla u_h - \boldsymbol{p}_h\|\|\varphi\|_{H^2(\Omega)} \\
&\lesssim (\|z_h\| + h\|\boldsymbol{\beta} u_h - A\nabla u_h - \boldsymbol{p}_h\|)\,\|\varphi\|_{H^2(\Omega)}.
\end{aligned}$$

Combing (4.15)–(4.17) and (4.13) gives

$$\|e\| \lesssim h|||(u-u_h, \boldsymbol{p}-\boldsymbol{p}_h)|||_{-1} + \|z_h\|. \tag{4.18}$$

(4.14) is then a direct consequence of (4.18), (4.1), and (4.7). This completes the proof of the proposition. ◻

Remark 4.2. Note that the hidden constants in (4.8) and (4.13) blow up in the advection dominated regime. Therefore the above L^2-analysis is relevant only in the low Péclet regime.

Remark 4.3 (the role of z_h). The multiplier variable z_h encodes all information necessary for the a posteriori error estimation. We will show this in the case of homogeneous Dirichlet conditions, i.e., $g = 0$ and $\Gamma_N = \emptyset$. Note that, in this case, from (4.6) there holds (neglecting for simplicity the dependence of α)

$$\|u-u_h\|_{H^1(\Omega)} \lesssim |||(u-u_h, \boldsymbol{p}-\boldsymbol{p}_h)|||_{-1} \lesssim \|\boldsymbol{p}_h - \boldsymbol{\beta} u_h + A\nabla u_h\| + h\|f - \pi_{X,m} f\|.$$

From (3.12) we deduce that

$$\|\boldsymbol{p}_h - \boldsymbol{\beta} u_h + A\nabla u_h\|^2 = -(\nabla\cdot\boldsymbol{p}_h + \mu u_h, z_h) = -(f, z_h). \tag{4.19}$$

Combing the above two inequalities gives

$$\|u-u_h\|_{H^1(\Omega)} \lesssim |(f, z_h)|^{1/2} + h\|f - \pi_{X,m} f\|. \tag{4.20}$$

This shows that the error only depends on the stability of the continuous problem, the right-hand side data, and z_h.

4.3. Error estimates in the advection dominated regime. In this section, we consider error estimates in the advection dominated regime. We consider estimates for smooth solutions $s = r = k+1$ in Lemma 4.1 and $\|A\|_\infty << h$, $\boldsymbol{\beta}_\infty = O(1)$, $|\mu| = O(1)$, so that Corollary 4.1 holds.

For the stability we make the following assumption on the data that ensures stability of the adjoint equation independent of the diffusivity; see [28].

Assumption 4.2. We assume that the the domain Ω is convex, that the diffusivity A is a scalar, and $\boldsymbol{\beta}_\infty = O(1)$. Let $\mathcal{I}$ denote the identity matrix and $\nabla_S\boldsymbol{\beta} := 1/2(\nabla\boldsymbol{\beta} + (\nabla\boldsymbol{\beta})^T)$, i.e., the symmetric part of $\nabla\boldsymbol{\beta}$. Then assume that $\mu\mathcal{I} - (\nabla_S\boldsymbol{\beta} - 1/2\nabla\cdot\boldsymbol{\beta}\mathcal{I})$ is symmetric positive definite, and denote by Λ_{min} its smallest eigenvalue. Moreover we assume that $\boldsymbol{\beta}\cdot\boldsymbol{n} = 0$ on $\partial\Omega$.

We now prove the following inverse inequality regarding the $H^{-1}(\Omega)$ norm.

LEMMA 4.2. *For any $v_h \in V_h^k$ the following inverse inequality holds:*

$$\|v_h\| \lesssim h^{-1}\|v_h\|_{H^{-1}(\Omega)}. \tag{4.21}$$

Proof. Let $E \in H_0^1(\Omega)$ be the weak solution to

$$-\triangle E + E = v_h \text{ in } \Omega.$$

Then by the definition and duality inequality we have

$$\begin{aligned} \|E\|_{H^1(\Omega)} &= \sup_{\substack{w \in H_0^1(\Omega) \\ \|w\|_{H^1(\Omega)}=1}} ((\nabla E, \nabla w) + (E, w)) \\ &= \sup_{\substack{w \in H_0^1(\Omega) \\ \|w\|_{H^1(\Omega)}=1}} (-\triangle E + E, w) = \|v_h\|_{H^{-1}(\Omega)}. \end{aligned} \tag{4.22}$$

By integration by parts we also have

$$\|v_h\|^2 = (v_h, -\triangle E + E) = (v_h, E) + (\nabla v_h, \nabla E) \le \|v_h\|_{H^1(\Omega)} \|E\|_{H^1(\Omega)},$$

which, combined with (4.22) and the inverse inequality, gives (4.21). This completes the proof of the lemma. ∎

Assumption 4.2 allows us to show that the H^1-seminorm of the solution is bounded uniformly in the diffusion coefficient.

LEMMA 4.3. *Let $\phi \in H_0^1(\Omega)$ be the solution to (4.12) with the right-hand side $\psi \in H_0^1(\Omega)$. Then under the Assumption 4.2 the following stability result holds:*

$$\Lambda_{min} \|\nabla \phi\| \le \|\nabla \psi\|. \tag{4.23}$$

Proof. By the definition and integration by parts we have

$$(\psi, -\triangle \phi) = (\mu \nabla \phi, \nabla \phi) + (\boldsymbol{\beta} \cdot \nabla \phi, \triangle \phi) + (A \triangle \phi, \triangle \phi)$$

Using the relation of [11, equation (3.6)] we have for the second term of the right-hand side

$$(\boldsymbol{\beta} \cdot \nabla \phi, \triangle \phi) = \left(\left(\frac{1}{2} \nabla \cdot \boldsymbol{\beta} \mathcal{I} - \nabla_S \boldsymbol{\beta} \right) \nabla \phi, \nabla \phi \right). \tag{4.24}$$

Combining similar terms we then have

$$\left(\left(\mu \mathcal{I} - \left(\nabla_S \boldsymbol{\beta} - \frac{1}{2} \nabla \cdot \boldsymbol{\beta} \mathcal{I} \right) \right) \nabla \phi, \nabla \phi \right) + (A \triangle \phi, \triangle \phi) = (\psi, -\triangle \phi) = (\nabla \psi, \nabla \phi), \tag{4.25}$$

and, therefore,

$$\Lambda_{min} \|\nabla \phi\| \le \|\nabla \psi\| \tag{4.26}$$

and, as a byproduct,

$$\left\| A^{1/2} D^2 \phi \right\| \lesssim \left\| A^{1/2} \triangle \phi \right\| \le \Lambda_{min}^{-1/2} \|\nabla \psi\|.$$

This completes the proof of the lemma. ∎

Proposition 4.4. *Let u and u_h be the solutions of* (2.4) *and* (3.19), *respectively. Then under Assumption* 4.2 *we have the following estimate:*

$$\|u-u_h\|_{H^{-1}(\Omega)} \le C_P \Lambda_{min}^{-1} |||(u-u_h, \boldsymbol{p}-\boldsymbol{p}_h)|||_{-1}, \tag{4.27}$$

where C_P is the constant of the Poincaré inequality

$$\sum_K \|h_K^{-1}(\phi - \pi_{X,0}\phi)\|_K^2 \le C_P^2 \|\nabla \phi\|^2.$$

Proof. By definition we have

$$\|u-u_h\|_{H^{-1}(\Omega)} = \sup_{\substack{w\in H_0^1(\Omega)\\ \|w\|_{H^1(\Omega)}=1}} (u-u_h, w).$$

Let $\varphi \in H_0^1(\Omega)$ be the solution of (4.12) with the right-hand side an arbitrary function $\psi \in H_0^1(\Omega)$ with $\|\psi\|_{H^1(\Omega)} = 1$. Applying the integration by parts, (3.13), and the Cauchy–Schwarz inequality gives

$$\begin{aligned}
&(u-u_h,\psi) = (u-u_h, -\boldsymbol{\beta}\cdot\nabla\varphi - A\triangle\varphi + \mu\varphi)\\
&\quad = (\mu(u-u_h) + \nabla\cdot(\boldsymbol{p}-\boldsymbol{p}_h), \varphi) - (\boldsymbol{\beta}(u-u_h) - A\nabla(u-u_h) - (\boldsymbol{p}-\boldsymbol{p}_h), \nabla\varphi)\\
&\quad = (\mu(u-u_h) + \nabla\cdot(\boldsymbol{p}-\boldsymbol{p}_h), \varphi - \pi_{X,0}\varphi) - (\boldsymbol{\beta}(u-u_h) - A\nabla(u-u_h) - (\boldsymbol{p}-\boldsymbol{p}_h), \nabla\varphi)\\
&\quad \le (C_P\|h(\mu(u-u_h) + \nabla\cdot(\boldsymbol{p}-\boldsymbol{p}_h))\| + \|\boldsymbol{\beta}(u-u_h) - A\nabla(u-u_h) - (\boldsymbol{p}-\boldsymbol{p}_h)\|)\,\|\nabla\varphi\|\\
&\quad \le C_P |||(u-u_h, \boldsymbol{p}-\boldsymbol{p}_h)|||_{-1}\|\nabla\varphi\| \le C_P\Lambda_{min}^{-1}|||(u-u_h,\boldsymbol{p}-\boldsymbol{p}_h)|||_{-1},
\end{aligned}$$

where in the last inequality we also applied the stability result of Lemma 4.3. This completes the proof of the proposition, since the bound is valid for arbitrary $\psi \in H_0^1(\Omega)$ with $\|\psi\|_{H^1(\Omega)} = 1$. ∎

Corollary 4.2 (negative norm, a posteriori, and a priori bounds). *Under the same hypothesis as Proposition* 4.4 *the following a posteriori and a priori error estimates hold:*

$$\begin{aligned}
\|u-u_h\|_{H^{-1}(\Omega)} &\le C_P\Lambda_{min}^{-1}(\|h(f-\mu u_h - \nabla\cdot\boldsymbol{p}_h)\| + \|\boldsymbol{\beta}u_h - A\nabla u_h - \boldsymbol{p}_h\|)\\
&\lesssim C_P\Lambda_{min}^{-1}(h^{k+1}|u|_{H^{k+1}(\Omega)} + h^{k+1}|\boldsymbol{p}|_{H^{k+1}(\Omega)}).
\end{aligned} \tag{4.28}$$

Proof. The proof is immediate using Proposition 4.4 and Corollary 4.1. ∎

We are now ready to prove the main result.

Theorem 4.1. *Let $u \in H^{k+1}(\Omega)\cap H_0^1(\Omega)$, $\boldsymbol{p} \in [H^{k+1}(\Omega)]^d$, and $(u_h, \boldsymbol{p}_h, z_h)$ be the solution of* (3.12)–(3.13). *Assume that $\|A\|_\infty \lesssim h^2$. Then under Assumption* 4.2 *we have the following error estimates:*

$$\|u-u_h\| + \|\nabla\cdot(\boldsymbol{p}-\boldsymbol{p}_h)\| \lesssim h^k\left(|u|_{H^{k+1}(\Omega)} + |\boldsymbol{p}|_{H^{k+1}(\Omega)}\right) \tag{4.29}$$

and

$$\|\nabla\cdot(\boldsymbol{\beta}(u-u_h))\| \lesssim h^k\left(|u|_{H^{k+1}(\Omega)} + |\boldsymbol{p}|_{H^{k+1}(\Omega)}\right). \tag{4.30}$$

Here the hidden constants are bounded in the limit as $A \to 0$.

Proof. Applying the triangle inequality, (4.21), and Corollary 4.2 gives

$$
\begin{aligned}
\|u-u_h\| &\le \|u-i_h u\| + h^{-1}\|u_h - i_h u\|_{H^{-1}(\Omega)} \\
&\lesssim h^{k+1}|u|_{H^{k+1}(\Omega)} + h^{-1}\|u-u_h\|_{H^{-1}(\Omega)} + h^{-1}\|u-i_h u\|_{H^{-1}(\Omega)} \\
&\lesssim h^{k+1}|u|_{H^{k+1}(\Omega)} + h^{-1}\|u-u_h\|_{H^{-1}(\Omega)} + h^{-1}\|u-i_h u\| \\
&\lesssim h^k \left(|u|_{H^{k+1}(\Omega)} + |\boldsymbol{p}|_{H^{k+1}(\Omega)}\right).
\end{aligned} \tag{4.31}
$$

Applying the triangle inequality, (4.31), and (4.1) gives

$$
\begin{aligned}
\|\nabla\cdot(\boldsymbol{p}-\boldsymbol{p}_h)\| &\le \|\nabla\cdot(\boldsymbol{p}-\boldsymbol{p}_h) + \mu(u-u_h)\| + \|\mu(u-u_h)\| \\
&\lesssim h^{-1}|||(u-u_h,\boldsymbol{p}-\boldsymbol{p}_h)|||_{-1} + \|\mu(u-u_h)\| \\
&\lesssim h^k \left(|u|_{H^{k+1}(\Omega)} + |\boldsymbol{p}|_{H^{k+1}(\Omega)}\right).
\end{aligned} \tag{4.32}
$$

(4.29) is then a direct result of (4.31) and (4.32).

To prove (4.30) we first apply the triangle inequality,

$$
\|\nabla\cdot(\boldsymbol{\beta}(u-u_h))\| \le \|\nabla\cdot(\boldsymbol{\beta}(u-i_h u))\| + \|\nabla\cdot(\boldsymbol{\beta}(u_h-i_h u))\|. \tag{4.33}
$$

The first term in (4.33) can be easily estimated using (3.1). For the second term in (4.33) applying the triangle and inverse inequalities gives

$$
\begin{aligned}
\|\nabla\cdot(\boldsymbol{\beta}(u_h-i_h u))\| &\le h^{-1}\|\boldsymbol{\beta}(u_h-i_h u) - (R_h\boldsymbol{p}-\boldsymbol{p}_h) - A\nabla(u_h-i_h u)\| \\
&\quad + h^{-2}\|A\|_\infty \left(\|(u_h-u)\| + \|(i_h u - u)\|\right) \\
&\quad + \|\nabla\cdot(\boldsymbol{p}-\boldsymbol{p}_h)\| + \|\nabla\cdot(\boldsymbol{p}-R_h\boldsymbol{p})\|.
\end{aligned} \tag{4.34}
$$

By the triangle inequality, Corollary 4.1, and (3.17) we have

$$
h^{-1}\|\boldsymbol{\beta}(u_h-i_h u)-(R_h\boldsymbol{p}-\boldsymbol{p}_h)-A\nabla(u_h-i_h u)\| \lesssim h^k(|u|_{H^{k+1}(\Omega)} + |\boldsymbol{p}|_{H^{k+1}(\Omega)}). \tag{4.35}
$$

By the assumption $\|A\|_\infty \lesssim h^2$, (3.1), (3.2), and (4.29), the remaining terms in (4.34) can be estimated as follows:

$$
\begin{aligned}
&h^{-2}\|A\|_\infty \left(\|(u_h-u)\| + \|(i_h u-u)\|\right) + \|\nabla\cdot(\boldsymbol{p}-\boldsymbol{p}_h)\| + \|\nabla\cdot(\boldsymbol{p}-R_h\boldsymbol{p})\| \\
&\quad \lesssim h^k \left(|u|_{H^{k+1}(\Omega)} + |\boldsymbol{p}|_{H^{k+1}(\Omega)} + |\nabla\cdot\boldsymbol{p}|_{H^k(\Omega)}\right).
\end{aligned} \tag{4.36}
$$

Finally, (4.30) is a direct consequence of (4.33)–(4.36). This completes the proof of the lemma. ∎

Remark 4.4. It is possible to prove Theorem 4.1 under the standard coercivity assumption for advection-diffusion problems, but not Proposition 4.4. Also note that we need the diffusivity to be $O(h^2)$ to ensure that the high Péclet result holds. This is a stronger assumption than usual for convection-diffusion equations, but a similar condition was introduced in the analysis of the FOSLS method in [23].

Remark 4.5. Observe that the results in the above section are both suboptimal compared to approximation (except (4.30)) and somewhat academic due to the strong assumptions on the velocity field. The interest of the results resides in the fact that the bounds of Proposition 4.4 and Corollary 4.2 appear to use the stability of the continuous problem and the stability of the numerical method in an optimal way. Since the estimate on the residuals is optimal, it is difficult to see how to improve on the estimate. In the numerical section below, we explore the performance of the method on a less restrictive set of physical parameters. The error bounds in Corollary 4.2 and (4.30) are similar to the corresponding results for classical stabilized methods.

5. Numerical experiments. In this section we present results for numerical experiments in both the diffusion dominated and convection dominated regimes. The numerical results are produced using the FEniCS software [41].

Example 5.1 (diffusion problem with singularity). In this example we test a pure diffusion problem, i.e., $\epsilon = 1$, $\boldsymbol{\beta} = 0$, and $\mu = 0$, on the L-shaped domain $\Omega = (-1,1)^2 \setminus (-1,-1)^2$. We consider the problem with solution being

$$u(r,\theta) = r^{2/3}\sin(2\theta/3), \quad \theta \in [0, 3\pi/2]$$

in polar coordinates. It is well known that the solution satisfies

$$-\triangle u = 0 \quad \text{in}\, \Omega$$

and belongs to $H^{5/3-\epsilon}(\Omega)$ for $\epsilon > 0$ with the singularity located at the reentrant corner, i.e., $(0,0)$. The numerical scheme takes the pure Dirichlet boundary condition.

The magnitude of errors and their corresponding convergence rates are presented in Table 1. For this pure diffusion problem, where the solution has a singularity and limited smoothness, we observe optimal convergence for both the primal and flux variables. The flux variable is a superior approximation of the fluxes, but only by a factor two.

Example 5.2 (indefinite problem). In this example we consider the indefinite problem used in [21, 9]; the parameters are chosen such that A is the identity matrix, $\mu = 0$, and

$$\boldsymbol{\beta} = (-100(x+y), -100(y-x))$$

with the domain $\Omega = [0,1]^2$. The problem is set to satisfy the homogeneous Dirichlet boundary condition and has $\|u\| = 1$. It is easy to check that $\nabla \cdot \boldsymbol{\beta} = -200$ which makes the problem highly noncoercive with a medium high Péclet number. The optimal convergence rates are verified in Table 2 and Table 3 for orders $k = 1$ and $k = 2$, respectively.

Example 5.3 (pure convection problem with internal layer). In this example we consider a pure convection problem with an internal layer. See [33, section 5.2.3]. The solution has the following representation:

$$u(x,y) = exp(-\sigma\rho(x,y)\text{arccos}\left(\frac{y+1}{\rho(x,y)}\right)\arctan\left(\frac{\rho(x,y)-1.5}{\delta}\right), \tag{5.1}$$

where $\sigma = 0.1$, $\rho(x,y) = \sqrt{x^2 + (y+1)^2}$. It is easy to verify that

$$\nabla \cdot \boldsymbol{\beta} = 0, \quad \boldsymbol{\beta} \cdot \nabla u + \sigma u = 0$$

Table 1
Errors and convergence rates for Example 5.1 *with* $k = 1$.

h	$\|u-u_h\|$	Rate	$\|u-u_h\|_{H^1(\Omega)}$	Rate	$\|\boldsymbol{p}-\boldsymbol{p}_h\|$	Rate
1/16	3.025E-3	1.35	7.790E-2	0.65	4.110E-2	0.67
1/32	1.189E-3	1.35	4.949E-2	0.65	2.589E-2	0.67
1/64	4.689E-4	1.34	3.315E-2	0.66	1.631E-2	0.67

TABLE 2
Errors and convergence rates for Example 5.2 with $k = 1$.

h	$\|u - u_h\|$	Rate	$\|u - u\|_{H^1(\Omega)}$	Rate	$\|\boldsymbol{p} - \boldsymbol{p}_h\|$	Rate
1/16	9.469E-3	1.79	4.631E-1	1.01	8.281E-1	1.86
1/32	2.736E-3	1.90	2.295E-1	1.00	2.274E-1	1.92
1/64	7.317E-4	1.96	1.143E-1	1.00	6.025E-2	1.96
1/128	1.876E-4	1.99	5.708E-2	1.00	1.546E-2	1.99

h	$\|\nabla \cdot (\boldsymbol{p} - \boldsymbol{p}_h)\|$	Rate	$\|z_h\|$	Rate
1/16	1.513E-0	2.00	3.494E-3	2.42
1/32	3.789E-1	2.00	6.524E-4	2.02
1/64	9.478E-2	2.00	1.599E-4	1.99
1/128	2.369E-2	2.00	4.036E-5	1.99

TABLE 3
Errors and convergence rates for Example 5.2 with $k = 2$.

h	$\|u - u_h\|$	Rate	$\|u - u\|_{H^1(\Omega)}$	Rate	$\|\boldsymbol{p} - \boldsymbol{p}_h\|$	Rate
1/16	1.585E-4	3.19	1.597E-2	2.00	1.796E-2	3.19
1/32	1.733E-5	3.15	3.986E-3	2.00	1.969E-3	3.15
1/64	1.958E-6	3.05	9.965E-4	2.00	2.220E-4	3.06
1/128	2.358E-7	3.01	2.491E-4	2.00	2.671E-5	3.01

h	$\|\nabla \cdot (\boldsymbol{p} - \boldsymbol{p}_h)\|$	Rate	$\|z_h\|$	Rate
1/16	4.141E-2	3.00	7.311E-5	3.32
1/32	5.181E-3	3.00	7.283E-6	3.12
1/64	6.478E-4	3.00	8.352E-7	3.03
1/128	8.098E-5	3.00	1.018E-7	3.00

TABLE 4
Errors and convergence rates for Example 5.3 with $\delta = 1$ and $k = 1$.

h	$\|u - u_h\|$	Rate	$\|u - u\|_{H^1(\Omega)}$	Rate	$\|\boldsymbol{p} - \boldsymbol{p}_h\|$	Rate
1/32	6.021E-5	2.04	8.591E-3	1.00	6.939E-5	2.03
1/64	1.475E-5	2.03	4.281E-3	1.00	1.711E-5	2.02
1/128	3.638E-6	2.02	2.135E-3	1.00	4.235E-6	2.01

h	$\|\nabla \cdot (\boldsymbol{p} - \boldsymbol{p}_h)\|$	Rate	$\|\nabla \cdot (\boldsymbol{\beta}(u - u_h))\|$	Rate	$\|z_h\|$	Rate
1/32	6.021E-6	2.04	5.343E-3	1.00	1.084E-7	2.99
1/64	1.475E-6	2.03	2.669E-3	1.00	1.360E-8	3.00
1/128	3.638E-7	2.02	1.333E-3	1.00	1.703E-9	3.00

for $\boldsymbol{\beta} = \frac{1}{\rho(x,y)}(y+1, -x)$ and that the inflow boundary, $\Gamma^- = \{x \in \partial\Omega, \boldsymbol{\beta}(x) \cdot \boldsymbol{n} < 0\}$, is $x = 0$ and $y = 1$. The solution possesses an internal layer when δ is small. The finite element scheme we use for this problem is to find $u \in V^k_{g,\Gamma^-}$, $\boldsymbol{p} \in RT^k$, and $z_h \in X^k_h$ such that (3.12) and (3.13) hold.

We first test the case when $\delta = 1$ to test the performance of our method on smooth problems (see performance results in Table 4). We further test the case when $\delta = 0.01$ in which case the solution has a sharp internal layer (see performance results in Table 5).

To test the robustness of our method for the pure convection problem, in Figure 1 we show the numerical solutions with $k = 1$ and $\delta = 0.001$ on structured meshes with various mesh sizes. We observe that, even for the highly sharp internal layer problem

TABLE 5
Errors and convergence rates for Example 5.3 *with* $\delta = 0.01$.

h	$\|u-u_h\|$	Rate	$\|u-u\|_{H^1(\Omega)}$	Rate	$\|\boldsymbol{p}-\boldsymbol{p}_h\|$	Rate
1/128	2.616E-2	1.17	3.801E-0	0.64	2.615E-2	1.17
1/256	9.421E-3	1.47	2.012E-0	0.91	9.421E-3	1.47
1/512	2.515E-3	1.91	8.461E-1	1.25	2.515E-3	1.91

h	$\|\nabla\cdot(\boldsymbol{p}-\boldsymbol{p}_h)\|$	Rate	$\|\nabla\cdot(\boldsymbol{\beta}(u-u_h))\|$	Rate	$\|z_h\|$	Rate
1/128	2.616E-3	1.17	3.435E-1	0.36	1.375E-6	2.25
1/256	9.421E-4	1.47	2.362E-1	0.54	2.367E-7	2.54
1/512	2.515E-4	1.91	1.423E-1	0.73	3.200E-8	2.89

(a) $k = 1$

h	$\|u-u_h\|$	Rate	$\|u-u\|_{H^1(\Omega)}$	Rate	$\|\boldsymbol{p}-\boldsymbol{p}_h\|$	Rate
1/128	4.470E-3	1.77	1.123E-0	1.25	4.470E-3	1.77
1/256	8.402E-4	2.41	3.103E-1	1.86	8.402E-4	2.41
1/512	8.442E-5	3.31	5.018E-2	2.63	8.441E-5	3.32

h	$\|\nabla\cdot(\boldsymbol{p}-\boldsymbol{p}_h)\|$	Rate	$\|\nabla\cdot(\boldsymbol{\beta}(u-u_h))\|$	Rate	$\|z_h\|$	Rate
1/128	4.470E-4	1.77	6.839E-2	1.06	8.267E-8	2.91
1/256	8.402E-5	2.41	2.610E-3	1.39	7.847E-9	3.39
1/512	8.442E-6	3.32	7.817E-3	1.74	5.088E-10	3.94

(b) $k = 2$

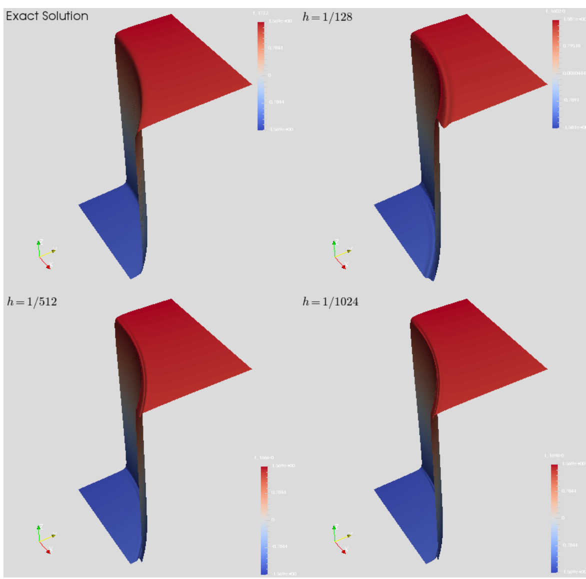


FIG. 1. *Various numerical solutions for Example* 5.3 *with* $\delta = 0.001$.

on relatively coarse meshes, the numerical solutions show no signs of global spurious oscillations. When the mesh size does not resolve the layer only mild and localized oscillations present around the internal layer.

Example 5.4 (boundary layer). In this example we consider the boundary layer problem [43]

TABLE 6
Errors and convergence performance for Example 5.4 *with* $\epsilon = 1$.

h	$\|u-u_h\|$	Rate	$\|u-u_h\|_{H^1(\Omega)}$	Rate	$\|\boldsymbol{p}-\boldsymbol{p}_h\|$	Rate
1/32	5.084E-4	1.99	4.240E-2	0.99	1.213E-3	2.00
1/64	1.273E-4	2.00	2.123E-2	1.00	3.035E-4	2.00
1/128	3.184E-5	2.00	1.062E-2	1.00	7.592E-5	2.00

(a) $k = 1$

h	$\|u-u_h\|$	Rate	$\|u-u_h\|_{H^1(\Omega)}$	Rate	$\|\boldsymbol{p}-\boldsymbol{p}_h\|$	Rate
1/32	5.129E-6	3.00	1.231E-3	2.00	3.602E-5	2.58
1/64	6.415E-7	3.00	3.081E-4	2.00	6.166E-6	2.55
1/128	8.021E-8	3.00	7.705E-5	2.00	1.071E-6	2.52

(b) $k = 2$

$$-\epsilon\triangle u + 2\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} = f$$

on the domain $\Omega = [0,1]^2$, where the true solution has the following representation:

$$u = (1 - \exp(-(1-x)/\epsilon)) * (1 - \exp(-(1-y)/\epsilon)) * \cos(\pi(x+y))$$

and $\epsilon \in \mathbb{R}$. The solution has a $O(\epsilon)$ boundary layer along the right top sides of the domain, and the value of ϵ determines the strength of the boundary layer.

We first test the value $\epsilon = 1$ in which case the solution is smooth. The magnitude of the errors and their corresponding convergence rates are listed in Table 6 for the first and second orders, i.e., $k = 1$ and $k = 2$. For both orders we observe optimal convergence of the errors for the primal variable in both the L^2- and H^1- norms. For the flux variable we observe the optimal rate for the linear order and a slightly suboptimal rate for the second order, compared to interpolation. Nevertheless the flux variable provides an approximation of the flux that is more accurate than that using the primal variable by two orders of magnitude. Note that the convergence orders observed are in all cases consistent with the theoretical results.

We now test optimal convergence rate by letting $\epsilon = 0.01$ (see performance results in Table 7). For both the first and second orders, the method produces the optimal convergence rate for the streamline derivative. For the flux variable we observe the optimal convergence for both orders 1 and 2. For the primal variable we observe the optimal convergence rates both in the L^2- and H^1- norms.

To test the robustness of our method, we compute with $\epsilon = 0.002$ in which case the boundary layer is extremely sharp. The numerics, however, produce global spurious oscillations when the layer is not resolved. We also compared our method with the FOSLS method in [18, 19]. The results by FOSLS also produce global oscillations while our method shows slightly better performance.

In Figure 2 we compare the convergence performance between FOSLS and the primal dual method in the linear case when $\epsilon = 0.01$. We observe that both methods yield optimal convergence results for the primal variable. However, for the flux variable the primal dual variable converges one order faster than the FOSLS method.

6. Outflow boundary layers. From numerics we have seen that the current method does not handle outflow boundary well because of its lack of upstream mechanism. In this section we propose two simple modifications of the method based on the

TABLE 7
Errors and convergence rates for Example 5.4 with $\epsilon = 0.01$.

h	$\|u-u_h\|$	Rate	$\|u-u\|_{H^1(\Omega)}$	Rate	$\|\boldsymbol{p}-\boldsymbol{p}_h\|$	Rate
1/32	9.393E-2	1.42	6.2560E-0	0.97	2.066E-1	1.42
1/64	3.502E-2	1.79	3.1999E-0	1.00	7.724E-2	1.79
1/128	1.010E-2	1.94	1.5916E-0	1.00	2.233E-2	1.94
1/256	2.633E-3	1.98	7.9304E-1	1.00	5.823E-3	1.98

h	$\|\nabla\cdot(\boldsymbol{p}-\boldsymbol{p}_h)\|$	Rate	$\|\boldsymbol{\beta}\cdot\nabla(u-u_h)\|$	Rate	$\|z_h\|$	Rate
1/32	8.988E-1	1.55	9.916E-0	0.97	1.267E-3	1.71
1/64	3.072E-1	1.85	5.055E-0	1.01	3.850E-4	1.92
1/128	8.518E-2	1.96	2.508E-0	1.00	1.019E-4	1.98
1/256	2.190E-2	1.99	1.248E-0	1.00	2.586E-5	1.99

(a) $k = 1$

h	$\|u-u_h\|$	Rate	$\|u-u\|_{H^1(\Omega)}$	Rate	$\|\boldsymbol{p}-\boldsymbol{p}_h\|$	Rate
1/32	1.704E-2	2.78	1.836E-0	1.67	3.708E-2	2.83
1/64	2.475E-3	3.28	5.768E-1	1.88	5.223E-3	3.39
1/128	2.544E-4	3.26	1.569E-1	1.96	4.979E-4	3.38
1/256	2.659E-5	3.10	4.024E-2	1.99	4.762E-5	3.13

h	$\|\nabla\cdot(\boldsymbol{p}-\boldsymbol{p}_h)\|$	Rate	$\|\boldsymbol{\beta}\cdot\nabla(u-u_h)\|$	Rate	$\|z_h\|$	Rate
1/32	2.242E-1	2.45	2.909E-0	1.69	1.756E-4	2.95
1/64	4.103E-2	2.82	9.076E-1	1.88	2.274E-5	3.51
1/128	5.814E-3	2.95	2.463E-1	1.96	1.993E-6	3.60
1/256	7.521E-4	2.99	6.311E-2	1.99	1.634E-7	3.38

(b) $k = 2$

current setting that aim to remove the global spurious oscillation. More specifically, one method imposes the boundary condition weakly, whereas the other takes the approach of weighting the stabilizer such that the oscillation is more "costly" closer to the inflow boundary and, hence, introduces a notion of upwind direction.

6.1. Weakly imposed boundary conditions. In this approach we weakly impose the Dirichlet boundary conditions, giving different weight to the inflow and outflow boundary. The idea is similar to [23] for the FOSLS method. The modified weak formulation is to find $(u_h, \boldsymbol{p}_h, z_h) \in V_h^k \times RT^k \times X_h^k$ such that

$$\mathcal{A}_1[(u_h, \boldsymbol{p}_h, z_h), (v_h, \boldsymbol{q}_h, x_h)] = l_h(x_h) \quad \forall\, (v_h, \boldsymbol{q}_h, x_h) \in V_h^k \times RT^k \times X_h^k, \tag{6.1}$$

where

$$\begin{aligned}\mathcal{A}_1[(u_h, \boldsymbol{p}_h, z_h), (v_h, \boldsymbol{q}_h, x_h)] &= b(\boldsymbol{q}_h, v_h, z_h) + b(\boldsymbol{p}_h, u_h, x_h) + s[(u_h, \boldsymbol{p}_h), (v_h, \boldsymbol{q}_h)] \\ &\quad + \left\langle (h[\beta\cdot\boldsymbol{n}]_-^2 + \gamma\epsilon^2/h)u, v\right\rangle_{\partial\Omega}\end{aligned}$$

and

$$l_h(x_h) = (f, x_h) + \left\langle (h[\beta\cdot\boldsymbol{n}]_-^2 + \gamma\epsilon^2/h)g, v\right\rangle_{\partial\Omega}.$$

In the above formulation $[\beta\cdot n]_- = \min(0, \beta\cdot n)$ and $\epsilon = \min(\lambda_A)$, i.e., the smallest eigenvalue of A. The dimensionless parameter γ is free and can be varied to determine when the Dirichlet condition should come into effect; if set too large the spurious behavior will appear in the transition regime from dominating convection to dominating diffusion.

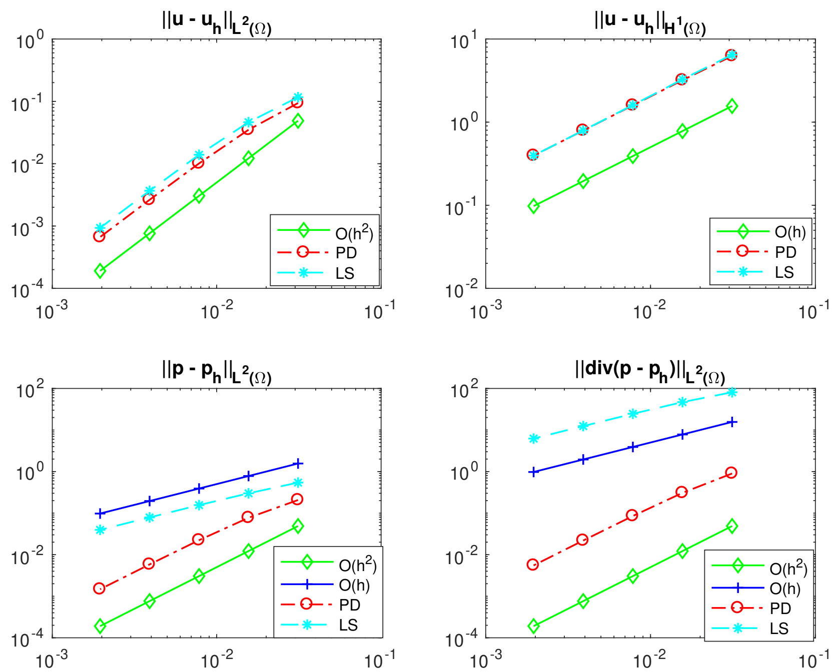


FIG. 2. *Comparison between primal dual method and least square Method.*

Remark 6.1. Note that in the above method the Dirichlet boundary condition is enforced weakly everywhere. Alternatively one may impose the Dirichlet condition strongly on the inflow boundary. The outcome is similar.

We test the method on a commonly used benchmark problem with both an internal and outflow boundary layers [8].

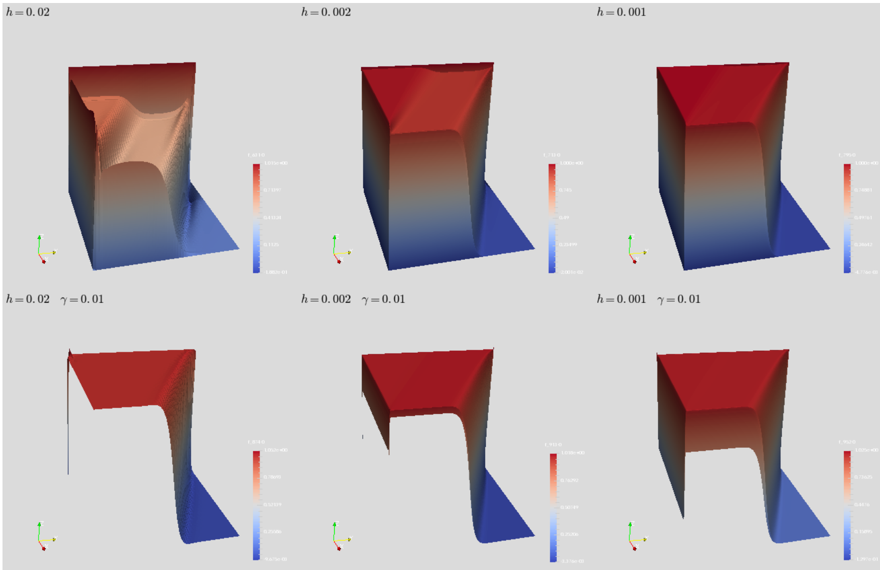
Example 6.1. Let u be the solution that satisfies

$$\begin{aligned} \nabla \cdot (\boldsymbol{\beta} u - \epsilon \nabla u) &= 0 \quad \text{on } \Omega, \\ u &= 1 \quad \text{on } \Gamma_L, \\ u &= 0 \quad \text{on } \partial\Omega \setminus \Gamma_L, \end{aligned}$$

where $\Omega = [0,1]^2$, $\beta = (1, -0.5)$, and Γ_L is the left boundary of the square, i.e., $x = 0$. ϵ is the diffusion coefficient, and in our test we choose $\epsilon = 0.001$ in which case the internal and boundary layers are very sharp.

In Figure 3 we compare the results between the original method (see figures on the top) and the method of (6.1) (see figures at the bottom). We observe that the weak boundary condition method results in an accurate solution in the bulk, with unresolved layers, that are resolved as the mesh-size is small enough, whereas the approximation with strongly imposed conditions has a globally large error.

6.2. Weighted stabilization method. In this subsection we explore how a notion of upwinding can be introduced in the present framework. We propose to introduce a nondimensional weight function in the stabilizing term s. The motivation here is to change the stabilization making oscillations more "costly" closer to the inflow boundary; this way a notion of flow direction is introduced, mimicking the upwind



(a) (b) (c)

FIG. 3. *Numerical performance of the weak boundary method for Example* 6.1.

behavior of classical stabilized methods. More precisely, we consider $\eta : \Omega \to \mathbb{R}$ such that

$$\eta > 0 \quad \text{and} \quad \boldsymbol{\beta} \cdot (\nabla \eta) < 0. \tag{6.2}$$

Defining $\boldsymbol{e}_{\boldsymbol{\beta}} = |\boldsymbol{\beta}|^{-1}\boldsymbol{\beta}$, we could choose, e.g.,

$$\eta = 3 - \boldsymbol{e}_{\boldsymbol{\beta}} \cdot (x, y)$$

for Example 5.4, and, for Example 6.1,

$$\eta = 2 - \boldsymbol{e}_{\boldsymbol{\beta}} \cdot (x, y).$$

It is easy to check that (6.2) holds for both problems. We then introduce η^p, for some $p > 0$ to be specified, as a weight in s. Note that the power p is introduced to modify the decay of η along the characteristic.

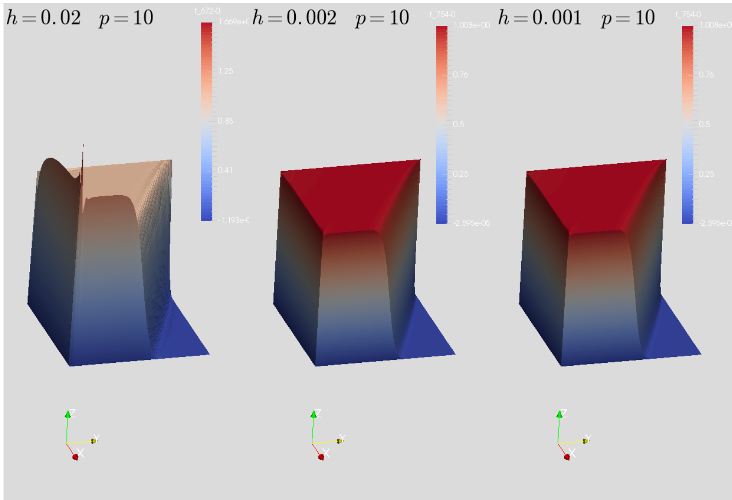
The finite element setting is then to find $(u_h, \boldsymbol{p}_h, z_h) \in V_{g,D}^k \times RT^k \times X_h^k$ such that

$$\mathcal{A}_2[(u_h, \boldsymbol{p}_h, z_h), (v_h, \boldsymbol{q}_h, x_h)] = l_h(x_h) \quad \forall (v_h, \boldsymbol{q}_h, x_h) \in V^k \times RT^k \times X_h^k, \tag{6.3}$$

where

$$\begin{aligned} \mathcal{A}_2[(u_h, \boldsymbol{p}_h, z_h), (v_h, \boldsymbol{q}_h, x_h)] &= s_\eta[(u_h, \boldsymbol{p}_h), (v_h, \boldsymbol{q}_h)] + b(\boldsymbol{p}_h, u_h, x_h) + b(\boldsymbol{q}_h, v_h, z_h), \\ s_\eta[(u_h, \boldsymbol{p}_h), (v_h, \boldsymbol{q}_h)] &= (\eta^p(\boldsymbol{p} + A\nabla u - \boldsymbol{\beta} u), (\boldsymbol{q} + A\nabla v - \boldsymbol{\beta} v)) \end{aligned}$$

and



(a) (b) (c)

FIG. 4. *Numerical performance of the weighted stabilization method.*

$$l_h(x_h) = (f, x_h).$$

Figure 4 shows the numerical solutions solved by (6.3) on the same meshes as in Figure 3. We observe that the global spurious oscillation has been eliminated even for very coarse mesh. Local oscillations along the outflow boundary appear when the layer is not fully resolved.

Acknowledgments. The authors wish to thank the reviewers for constructive remarks that helped improve the manuscript.

REFERENCES

[1] I. Babuška, *Error-bounds for finite element method*, Numer. Math., 16 (1971), pp. 322–333, https://doi.org/10.1007/BF02165003.

[2] C. Bacuta and J. Jacavage, *A non-conforming saddle point least squares approach for elliptic interface problems*, Comput. Methods Appl. Math., 19 (2019), pp. 399–415, https://doi.org/10.1515/cmam-2018-0202.

[3] C. Bacuta and K. Qirko, *A saddle point least squares approach for primal mixed formulations of second order PDEs*, Comput. Math. Appl., 73 (2017), pp. 173–186, https://doi.org/10.1016/j.camwa.2016.11.014.

[4] G. R. Barrenechea, A. H. Poza, and H. Yorston, *A stabilised finite element method for the convection-diffusion-reaction equation in mixed form*, Comput. Methods Appl. Mech. Engrg., 339 (2018), pp. 389–415, https://doi.org/10.1016/j.cma.2018.04.019.

[5] J. H. Bramble, R. D. Lazarov, and J. E. Pasciak, *A least-squares approach based on a discrete minus one inner product for first order systems*, Math. Comp., 66 (1997), pp. 935–955, https://doi.org/10.1090/S0025-5718-97-00848-X.

[6] S. C. Brenner, *Poincaré–Friedrichs inequalities for piecewise $H(1)$ functions*, SIAM J. Numer. Anal., 41 (2003), pp. 306–324, https://doi.org/10.1137/S0036142902401311.

[7] 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), pp. 199–259, https://doi.org/10.1016/0045-7825(82)90071-8.

[8] E. Burman, *A unified analysis for conforming and nonconforming stabilized finite element methods using interior penalty*, SIAM J. Numer. Anal., 43 (2005), pp. 2012–2033, https://doi.org/10.1137/S0036142903437374.

[9] E. Burman, *Stabilized finite element methods for nonsymmetric, noncoercive, and ill-posed problems. Part* I*: Elliptic equations*, SIAM J. Sci. Comput., 35 (2013), pp. A2752–A2780, https://doi.org/10.1137/130916862.

[10] E. Burman, *Error estimates for stabilized finite element methods applied to ill-posed problems*, C. R. Math. Acad. Sci. Paris, 352 (2014), pp. 655–659, https://doi.org/10.1016/j.crma.2014.06.008.

[11] E. Burman, *Robust error estimates in weak norms for advection dominated transport problems with rough data*, Math. Models Methods Appl. Sci., 24 (2014), pp. 2663–2684, https://doi.org/10.1142/S021820251450033X.

[12] E. Burman, *Stabilized finite element methods for nonsymmetric, noncoercive, and ill-posed problems. Part* II*: Hyperbolic equations*, SIAM J. Sci. Comput., 36 (2014), pp. A1911–A1936, https://doi.org/10.1137/130931667.

[13] E. Burman, *Stabilised finite element methods for ill-posed problems with conditional stability*, in Building Bridges: Connections and Challenges in Modern Approaches to Numerical Partial Differential Equations, G. R. Barrenechea, F. Brezzi, A. Cangiani, and E. H. Georgoulis, eds., Lect. Notes Comput. Sci. Eng. 114, 2016, pp. 93–127.

[14] E. Burman and P. Hansbo, *Edge stabilization for Galerkin approximations of convection-diffusion-reaction problems*, Comput. Methods Appl. Mech. Engrg., 193 (2004), pp. 1437–1453, https://doi.org/10.1016/j.cma.2003.12.032.

[15] E. Burman, M. G. Larson, and L. Oksanen, *Primal-dual mixed finite element methods for the elliptic Cauchy problem*, SIAM J. Numer. Anal., 56 (2018), pp. 3480–3509, https://doi.org/10.1137/17M1163335.

[16] E. Burman, L. Oksanen, and M. Nechita, *A Stabilized Finite Element Method for Inverse Problems Subject to the Convection-Diffusion Equation.* I*: Diffusion-Dominated Regime*, e-print, arXiv:1881.00431 [math.NA], 2018.

[17] E. Burman, A. Quarteroni, and B. Stamm, *Interior penalty continuous and discontinuous finite element approximations of hyperbolic equations*, J. Sci. Comput., 43 (2010), pp. 293–312, https://doi.org/10.1007/s10915-008-9232-6.

[18] Z. Cai, R. Lazarov, T. A. Manteuffel, and S. F. McCormick, *First-order system least squares for second-order partial differential equations: Part* I, SIAM J. Numer. Anal., 31 (1994), pp. 1785–1799, https://doi.org/10.1137/0731091.

[19] Z. Cai, T. A. Manteuffel, and S. F. McCormick, *First-order system least squares for second-order partial differential equations: Part* II, SIAM J. Numer. Anal., 34 (1997), pp. 425–454, https://doi.org/10.1137/S0036142994266066.

[20] I. Capuzzo-Dolcetta and S. Finzi Vita, *Finite element approximation of some indefinite elliptic problems*, Calcolo, 25 (1988), pp. 379–395, https://doi.org/10.1007/BF02575842.

[21] C. Chainais-Hillairet and J. Droniou, *Finite-volume schemes for noncoercive elliptic problems with Neumann boundary conditions*, IMA J. Numer. Anal., 31 (2009), pp. 61–85, https://doi.org/10.1093/imanum/drp009.

[22] J. Chan, J. A. Evans, and W. Qiu, *A dual Petrov-Galerkin finite element method for the convection-diffusion equation*, Comput. Math. Appl., 68 (2014), pp. 1513–1529, https://doi.org/10.1016/j.camwa.2014.07.008.

[23] H. Chen, G. Fu, J. Li, and W. Qiu, *First order least squares method with weakly imposed boundary condition for convection dominated diffusion problems*, Comput. Math. Appl., 68 (2014), pp. 1635–1652, https://doi.org/10.1016/j.camwa.2014.11.001.

[24] P. Ciarlet, Jr., *T-coercivity: application to the discretization of Helmholtz-like problems*, Comput. Math. Appl., 64 (2012), pp. 22–34, https://doi.org/10.1016/j.camwa.2012.02.034.

[25] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam, 1978.

[26] B. Cockburn and C.-W. Shu, *TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws.* II*. General framework*, Math. Comp., 52 (1989), pp. 411–435, https://doi.org/10.2307/2008474.

[27] R. Codina, *Stabilization of incompressibility and convection through orthogonal sub-scales in finite element methods*, Comput. Methods Appl. Mech. Engrg., 190 (2000), pp. 1579–1599, https://doi.org/10.1016/S0045-7825(00)00254-1.

[28] H. B. Da Veiga, *On a stationary transport equation*, Ann. Univ. Ferrara Sez. VII Sci. Math., 32 (1986), pp. 79–91, https://doi.org/10.1007/BF02825238.

[29] W. Dahmen, C. Huang, C. Schwab, and G. Welper, *Adaptive Petrov–Galerkin methods for first order transport equations*, SIAM J. Numer. Anal., 50 (2012), pp. 2420–2445, https://doi.org/10.1137/110823158.

[30] L. Demkowicz and J. Gopalakrishnan, *A class of discontinuous Petrov-Galerkin methods. Part I: The transport equation*, Comput. Methods Appl. Mech. Engrg., 199 (2010), pp. 1558–1572, https://doi.org/10.1016/j.cma.2010.01.003.

[31] J. Douglas, Jr., and T. Dupont, *Interior penalty procedures for elliptic and parabolic Galerkin methods*, in Computing Methods in Applied Sciences: Second International Symposium, December 15–19, 1975, R. Glowinski and J. L. Lions, eds., Springer-Verlag, Berlin, 1976, pp. 207–216.

[32] J. Douglas Jr., and J. E. Roberts, *Mixed finite element methods for second order elliptic problems*, Mat. Apl. Comput. (São Paulo), 1 (1982), pp. 91–103.

[33] A. Ern and J.-L. Guermond, *Theory and Practice of Finite Elements*, Appl. Math. Sci. 159, Springer-Verlag, New York, 2004, https://doi.org/10.1007/978-1-4757-4355-5.

[34] A. Ern and J.-L. Guermond, *Finite element quasi-interpolation and best approximation*, ESAIM Math. Model. Numer. Anal., 51 (2017), pp. 1367–1385, https://doi.org/10.1051/m2an/2016066.

[35] J.-L. Guermond, *Stabilization of Galerkin approximations of transport equations by subgrid modeling*, ESAIM Math. Model. Numer. Anal., 33 (1999), pp. 1293–1316, https://doi.org/10.1051/m2an:1999145.

[36] T. J. R. Hughes, G. Engel, L. Mazzei, and M. G. Larson, *The continuous Galerkin method is locally conservative*, J. Comput. Phys., 163 (2000), pp. 467–488, https://doi.org/10.1006/jcph.2000.6577.

[37] C. Johnson, U. Nävert, and J. Pitkäranta, *Finite element methods for linear hyperbolic problems*, Comput. Methods Appl. Mech. Engrg., 45 (1984), pp. 285–312, https://doi.org/10.1016/0045-7825(84)90158-0.

[38] C. Johnson and J. Pitkäranta, *An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation*, Math. Comp., 46 (1986), pp. 1–26, https://doi.org/10.2307/2008211.

[39] K. Kavaliou and L. Tobiska, *A finite element method for a noncoercive elliptic problem with Neumann boundary conditions*, Comput. Methods Appl. Math., 12 (2012), pp. 168–183, https://doi.org/10.2478/cmam-2012-0011.

[40] Y. Liu, J. Wang, and Q. Zou, *A conservative flux optimization finite element method for convection-diffusion equations*, SIAM J. Numer. Anal., 57 (2019), pp. 1238–1262.

[41] A. Logg, K.-A. Mardal, and G. Wells, *Automated Solution of Differential Equations by the Finite Element Method: The FEniCS Book*, Springer Science & Business Media, New York, 2012, https://doi.org/10.1007/978-3-642-23099-8.

[42] A. Masud and J. Kwack, *A stabilized mixed finite element method for the first-order form of advection-diffusion equation*, Internat. J. Numer. Methods Fluids, 57 (2008), pp. 1321–1348, https://doi.org/10.1002/fld.1842.

[43] W. F. Mitchell, *A collection of 2d elliptic problems for testing adaptive grid refinement algorithms*, Appl. Math. Comput., 220 (2013), pp. 350–364, https://doi.org/10.1016/j.amc.2013.05.068.

[44] P. Monk and E. Süli, *The adaptive computation of far-field patterns by a posteriori error estimation of linear functionals*, SIAM J. Numer. Anal., 36 (1998), pp. 251–274, https://doi.org/10.1137/S0036142997315172.

[45] P.-A. Raviart and J.-M. Thomas, *A mixed finite element method for 2-nd order elliptic problems*, in Mathematical Aspects of Finite Element Methods, I. Galligani and E. Magenes, eds. Springer, Berlin, 1977, pp. 292–315.

[46] J.-M. Thomas, *Mixed finite elements methods for convection-diffusion problems*, in Numerical Approximation of Partial Differential Equations, E. L. Ortiz, ed., North-Holland, Math. Stud. 133, North-Holland, Amsterdam, 1987, pp. 241–250, https://doi.org/10.1016/S0304-0208(08)71735-3.

[47] C. Wang and J. Wang, *A Primal-Dual Weak Galerkin Finite Element Method for Fokker-Planck Type Equations*, e-print, arXiv:1704.05606 [math.NA], 2017, https://arxiv.org/abs/1704.05606.

[48] C. Wang and J. Wang, *A primal-dual weak Galerkin finite element method for second order elliptic equations in non-divergence form*, Math. Comp., 87 (2018), pp. 515–545, https://doi.org/10.1090/mcom/3220.

[49] C. Wang and J. Wang, *Primal-dual weak Galerkin finite element methods for elliptic Cauchy problems*, Comput. Math. Appl., in press.