AIMS Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 8(4): 8034–8059.
DOI: 10.3934/math.2023407
Received: 07 November 2022
Revised: 09 January 2023
Accepted: 17 January 2023
Published: 31 January 2023

Research article

Immersed finite element methods for convection diffusion equations

Gwanghyun Jo[1] **and Do Y. Kwak**[2,*]

[1] Department of Mathematics, Kunsan National University, Republic of Korea
[2] Department of Mathematical Sciences, KAIST, Daejeon, Republic of Korea

* **Correspondence:** Email: kdy@kaist.ac.kr; Tel: +82423502720.

Abstract: In this work, we develop two IFEMs for convection-diffusion equations with interfaces. We first define bilinear forms by adding judiciously defined convection-related line integrals. By establishing Gårding's inequality, we prove the optimal error estimates both in L^2 and H^1-norms. The second method is devoted to the convection-dominated case, where test functions are piecewise constant functions on vertex-associated control volumes. We accompany the so-called upwinding concepts to make the control-volume based IFEM robust to the magnitude of convection terms. The H^1 optimal error estimate is proven for control-volume based IFEM. We document numerical experiments which confirm the analysis.

Keywords: immersed finite element method; convection-diffusion problem; interface problem; control volume; upwinding scheme
Mathematics Subject Classification: 65M08, 65M15, 65M60

1. Introduction

There are many physical phenomena where the parameters of the model problem change abruptly across some interfaces. For example, the parameter of multi-phase flows in porous media could be discontinuous on material interfaces across which materials have different permeabilities [1, 2]. Also, dielectric coefficients in biomolecular cells are discontinuous along solute-solvent interfaces [3, 4]. To numerically solve interface problems using the finite element method (FEM), people usually use fitted grids, since unfitted grids yield suboptimal convergence rates [5].

Recently, there appeared some structured grid-based methods for interface problems in the FEM community since data structures are simple. Extended finite element methods (XFEMs) are one of the popular structured grid-based methods for interface problems whose local spaces are enriched by some basis functions [6–9]. By truncating the shape function along the interfaces, additional basis functions are obtained near the interfaces.

On the other hand, immersed finite element methods (IFEMs) are another type of structured grid-based methods for interface problems. In IFEMs, the basis functions are modified along the interfaces to satisfy certain interface conditions. In this way, no extra degrees of freedom are necessary. Owing to the simple data structures of algebraic systems, multigrid algorithms for IFEM were proposed and analyzed in [10]. Error estimates for IFEM on elliptic and elasticity interface problems can be found in [11–16]. IFEM has been proven to be effective in various problems including multiphase flows in porous media [17], elasticity problem [15, 16], Poisson-Boltzmann-Nerst-Plank model [18] and Hele-Shaw flows [19]. Meanwhile, finite volume-based IFEMs for elliptic interface problems were developed in [20–23]. In these works, test functions are piecewise constant functions on so-called *control-volume*s or *dual-volume*s, and the resulting bilinear forms are simpler than those of conventional IFEM.

So far, most works are concentrated on symmetric problems while there are many examples of convection-diffusion problems involving interfaces, especially in porous media problems, see [2, 17]. For the convection-dominated flows problem, it is required that bilinear forms are modified properly to avoid non-physical oscillations. In this work, we develop two IFEMs for convection-diffusion elliptic interface problems. We start by defining a bilinear form similar to the original IFEM, modifying the line integrals of convection terms. We show that the proposed IFEM has optimal convergence rates. In order to handle convection-dominated case, we modify the first version by considering vertex-associated control volumes. The test function space is defined by the set of piecewise constant functions on each control volume. Finally, we apply the upwinding concepts so that control volume-based IFEM is robust to the magnitude of convection parameters. Optimal H^1-error estimate is carried out for the second scheme.

The rest of the paper is organized as follows. The model problem is described in Section 2. First type of IFEM is proposed in Section 3 and control volume based IFEM for convection dominating case is developed in Section 4. Numerical experiments which support the theory are reported in Section 5. Finally, some conclusive remarks are given in Section 6.

2. Governing equations

Assume that Ω is a convex polygonal domain in $\mathbb{R}^2$ divided by the interface Γ resulting in two subdomains, i.e., $\Omega = \Omega^- \cup \Omega^+$. We consider the second order elliptic interface model problem whose diffusion and convection parameters are not necessarily continuous across the interface.

$$-\nabla \cdot \beta \nabla u + \mathbf{b} \cdot \nabla u + Ru = f, \quad \text{in } \Omega, \tag{2.1}$$

$$[u]_\Gamma = 0, \quad \text{on } \Gamma, \tag{2.2}$$

$$[\beta \nabla u \cdot \mathbf{n}_\Gamma]_\Gamma = 0, \quad \text{on } \Gamma, \tag{2.3}$$

$$u = 0, \quad \text{in } \partial\Omega, \tag{2.4}$$

where $\mathbf{n}_\Gamma$ is a unit vector normal to the interface Γ^- and $[\cdot]_\Gamma$ implies the jumps across Γ, i.e., $[u]_\Gamma = u|_{\Omega^-} - u|_{\Omega^+}$. Here, the diffusion parameter β is allowed to be discontinuous across the interface Ω, with $\beta = \beta^+ \in C^1(\Omega^+)$ and $\beta = \beta^- \in C^1(\Omega^-)$. As usual, β is positive and uniformly bounded, i.e.,

$$0 < \underline{\beta} \leq \beta < \overline{\beta}. \tag{2.5}$$

We need similar assumptions for the convection parameter. While $\mathbf{b}$ is allowed to be discontinuous along Γ with $\mathbf{b} = \mathbf{b}^- \in C^1(\Omega^-)$ on Ω^- and $\mathbf{b} = \mathbf{b}^+ \in C^1(\Omega^+)$ on Ω^+, it is assumed that $\mathbf{b}$ is uniformly bounded, i.e.,

$$|\mathbf{b}| \le |\bar{\mathbf{b}}|,$$

and

$$[\mathbf{b} \cdot \mathbf{n}_\Gamma]_\Gamma = 0. \tag{2.6}$$

Finally, the reaction parameter R is assumed to be in $L^2(\Omega)$.

We introduce some notations. For a given subdomain $D \subset \Omega$ and $m = 1, 2$, $H^m(D)$, $H^1_0(D)$, $H^m(\partial D)$ are Sobolev spaces of order m with the norm $\|\cdot\|_{m,D}$ and the semi-norm $|\cdot|_{m,D}$. For any real number between m and $m + 1$, we define fractional Sobolev space $H^s(D)$ as the interpolation between $H^m(D)$ and $H^{m+1}(D)$. In particular, the norm of fractional space $H^{m+\sigma}(D)$, $0 < \sigma < 1$ (in two dimensional case) is defined as

$$\|u\|^2_{H^{m+\sigma}(D)} := \|u\|^2_{H^m(D)} + \sum_{|\alpha|=m} |D^\alpha u|^2_{H^\sigma(D)}$$

where semi-norm $|\cdot|_{H^\sigma(D)}$ is defined as

$$|u|^2_{H^\sigma(D)} := \int_D \int_D \frac{(u(\mathbf{x}) - u(\mathbf{y}))^2}{|x - y|^{2+2s}} \mathrm{d}x\mathrm{d}y.$$

In case $D = \Omega$, we simply denote the norm $\|\cdot\|_{m,\Omega}$ and the inner product $(\cdot,\cdot)_{m,\Omega}$ by $\|\cdot\|_m$ and $(\cdot,\cdot)_m$, respectively. For $0 < \alpha \le 1$, we also introduce the broken Sobolev space $\widetilde{H}^{1+\alpha}(\Omega)$ defined as

$$\widetilde{H}^{1+\alpha}(\Omega) := \{u \in H^1(\Omega) \;:\; u|_{\Omega^-} \in H^{1+\alpha}(\Omega^-) \text{ and } u|_{\Omega^+} \in H^{1+\alpha}(\Omega^+)\},$$

equipped with the norm:

$$\|u\|^2_{\widetilde{H}^{1+\alpha}(\Omega)} := \|u\|^2_1 + \|u\|^2_{H^{1+\alpha}(\Omega^-)} + \|u\|^2_{H^{1+\alpha}(\Omega^+)}.$$

Also, we introduce some subspaces.

$$H^1_0(\Omega) := \{u \in H^1(\Omega) \,|\, u = 0 \text{ on } \partial\Omega\},$$
$$\widetilde{H}^{1+\alpha}_\Gamma(\Omega) := \{u \in \widetilde{H}^{1+\alpha} \,|[u]_\Gamma = [\beta\nabla u \cdot \mathbf{n}_\Gamma]_\Gamma = 0\}.$$

By integration by parts, we can derive the weak formulation of the model problem as follows: find $u \in H^1_0(\Omega)$ such that

$$t(u, v) = (f, v)_0, \qquad \forall v \in H^1_0(\Omega), \tag{2.7}$$

where

$$t(u, v) = \int_{\Omega^-} \beta\nabla u \cdot \nabla v \,\mathrm{d}\mathbf{x} + \int_{\Omega^+} \beta\nabla u \cdot \nabla v \,\mathrm{d}\mathbf{x} + \int_\Omega (\mathbf{b} \cdot \nabla u + Ru)v \,\mathrm{d}\mathbf{x}.$$

The following assumption is commonly used to prove the existence of convection diffusion problems (see Part III. Section 1.1 in [24]).

$$-\frac{1}{2}\mathrm{div}\mathbf{b} + R > 0. \tag{2.8}$$

Proposition 2.1. *Under the condition (2.8), there exists a unique solution for (2.7).*

Proof. By Lax-Milgram theorem [25], it suffices to show the coerciveness of $t(\cdot,\cdot)$ on $H_0^1(\Omega)$. By the definition of the bilinear form, we have

$$t(v,v) = \int_{\Omega^-} \beta \nabla v \cdot \nabla v \, \mathrm{d}\mathbf{x} + \int_{\Omega^+} \beta \nabla v \cdot \nabla v \, \mathrm{d}\mathbf{x} + \int_{\Omega} \left(\nabla \frac{v^2}{2} \right) \cdot \mathbf{b} \mathrm{d}\mathbf{x} + \int_{\Omega} R v^2 \, \mathrm{d}\mathbf{x}.$$

By applying the integration by parts on third term together with (2.6), we have

$$t(v,v) = \int_{\Omega^-} \beta \nabla v \cdot \nabla v \, \mathrm{d}\mathbf{x} + \int_{\Omega^+} \beta \nabla v \cdot \nabla v \, \mathrm{d}\mathbf{x} \int_{\Omega} \left(-\frac{1}{2} \mathrm{div} \mathbf{b} + R \right) v^2 \mathrm{d}\mathbf{x}.$$

Finally, the coerciveness of $t(\cdot,\cdot)$ on $H_0^1(\Omega)$ is obtained by the conditions (2.5) and (2.8)

$$\begin{aligned} t(v,v) &\geq \underline{\beta} |v|_1^2 + \left(-\frac{1}{2} \mathrm{div} \mathbf{b} + R \right) \|v\|_0^2 \\ &\geq \min \left\{ \underline{\beta}, -\frac{1}{2} \mathrm{div} \mathbf{b} + R \right\} \|v\|_1^2. \end{aligned}$$

□

By Proposition 2.1, we have the regularity theorem for the model problem (2.1)–(2.4), see [26, 27].

Proposition 2.2. *There exists an* $0.5 < \alpha \leq 1$ *and a unique solution* $u \in H_0^1(\Omega) \cap \widetilde{H}^{1+\alpha}(\Omega)$ *of problem (2.1)–(2.4). Furthermore, u satisfies*

$$\|u\|_{\widetilde{H}^{1+\alpha}(\Omega)} \leq C \|f\|_{H^{-1+\alpha}(\Omega)}, \tag{2.9}$$

where C is some constant depending on β *and* $\mathbf{b}$.

Throughout the paper, we use generic constants $C, C_1, C_2, ...$ independent of the mesh size, not necessarily the same for each appearance.

3. Immersed finite element method for nonsymmetric elliptic interface problem

In this section, we develop a new version of IFEM for a model problem. For convenience's sake, we assume that $R = 0$ in (2.1) from now on. For the case $R \neq 0$, the proof proceeds in exactly the same way if one adds some obvious terms in the bilinear forms. We start by introducing some definitions of spaces and norms. For the sake of convenience, we assume that Ω is a rectangular domain throughout the rest of the manuscript. Let $\mathcal{T}_h$ be a triangulation of Ω by right triangles. Since nodes are not aligned with the interface Γ there arise triangles in $\mathcal{T}_h$ which is cut by Γ. In such a case, we say that T are the *interface* elements and denote $\mathcal{T}_h^I$ to be the set of *interface* element. Slight modifications are required for the inner products on $T \in \mathcal{T}_h^I$, i.e.,

$$(u,v)_{m,T} = (u,v)_{m,T^+} + (u,v)_{m,T^-}, \quad \|\cdot\|_{m,T}^2 = \|\cdot\|_{m,T^+}^2 + \|\cdot\|_{m,T^-}^2, \quad m = 0, 1,$$

where $T^+ = T \cap \Omega^+$ and $T^- = T \cap \Omega^-$. We define a broken Sobolev space

$$H^1(\mathcal{T}_h) = \left\{ \phi \in L^2(\Omega) : \phi|_T \in H^1(T), \quad \forall T \in \mathcal{T}_h \right\},$$

with a broken norm

$$\|\phi\|_{1,h}^2 = \sum_{\phi\in\mathcal{T}_h} \|\phi\|_{1,T}^2.$$

We define $\mathcal{E}_h$ to be the set of edges of $\mathcal{T}_h$. Here, we let $\mathcal{E}_h^o$ be the set of the interior edges and $\mathcal{E}_h^\partial$ be the boundary edges. For each $e \in \mathcal{E}_h$, we associate a unit vector $\mathbf{n}_e$ at e. We define the jump $[\phi]_e$ and average $\{\phi\}_e$ for $\phi \in H^1(\mathcal{T}_h)$ as follows:

$$[\phi]_e(\mathbf{x}) := \lim_{\delta\to 0+}(\phi(\mathbf{x}-\delta\mathbf{n}_e) - \phi(x+\delta\mathbf{n}_e)),$$
$$\{\phi\}_e(\mathbf{x}) := \frac{1}{2}\lim_{\delta\to 0+}(\phi(\mathbf{x}-\delta\mathbf{n}_e) + \phi(x+\delta\mathbf{n}_e)),$$

if e does not belong to $\partial\Omega$ and

$$[\phi]_e(\mathbf{x}) := \lim_{\delta\to 0+}(\phi(\mathbf{x}-\delta\mathbf{n}_{\partial\Omega})), \quad \{\phi\}_e(\mathbf{x}) := \lim_{\delta\to 0+}(\phi(\mathbf{x}-\delta\mathbf{n}_{\partial\Omega}))$$

if e belongs to $\partial\Omega$.

3.1. Immersed finite element method space

In this subsection, we describe P_1-conforming based immersed finite element (IFE) spaces. For any $T \in \mathcal{T}_h$, let $S_h(T)$ be a P_1-conforming space. If T is an interface element, we modify the space $S_h(T)$. For example, suppose T having node A_i's $(i = 1, 2, 3)$ is cut through the Γ at edge points E_1 and E_3 as in Figure 1.

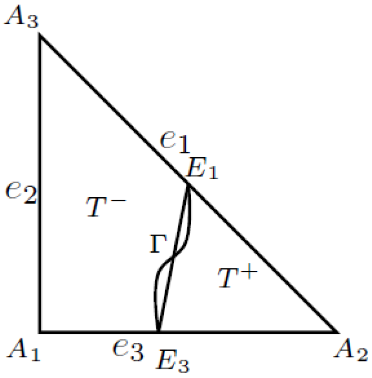


Figure 1. An interface element T cut by interface Γ.

For a function $\phi \in \mathcal{T}_h$, we modify it to be a piecewise linear function $\widehat{\phi}$ satisfying flux-continuity conditions, i.e., the new function $\widehat{\phi}$ has a form that

$$\widehat{\phi} = \begin{cases} \phi^+ = a^+ + b^+x + c^+y, & (x,y)\in T^+, \\ \phi^- = a^- + b^-x + c^-y, & (x,y)\in T^-, \end{cases} \tag{3.1}$$

and it satisfies

$$\widehat{\phi}(A_i) = \phi(A_i), \quad i = 1, 2, 3 \tag{3.2}$$

$$\phi^+(E_i) = \phi^-(E_i), \quad i = 1, 2, \tag{3.3}$$

$$\int_{\overline{E_1E_2}} \beta^+ \nabla\phi^+ \cdot \mathbf{n}_{\overline{E_1E_2}} = \int_{\overline{E_1E_2}} \beta^- \nabla\phi^- \cdot \mathbf{n}_{\overline{E_1E_2}}. \tag{3.4}$$

The uniqueness and existence of such modified functions are proven in [13]. Such modified piecewise linear IFE space on T is denoted by $\widehat{S}_h(T)$. Finally, we define global IFE space:

$$\widehat{S}_h(\Omega) := \left\{ \phi \in L^2(\Omega) : \begin{array}{l} \phi|_T \in S_h(T) \text{ if } T \in \mathcal{T}_h / \mathcal{T}_h^I \quad \text{and } \phi|_T \in \widehat{S}_h(T) \text{ if } T \in \mathcal{T}_h^I, \\ \phi \text{ is continuous on vertexes of } \mathcal{T}_h, \\ \phi(X) = 0 \text{ if } X \text{ is a nodes on } \partial\Omega \end{array} \right\}.$$

We intent to state the approximation property of the space $\widehat{S}_h(\Omega)$. For this purpose, we introduce nodal interpolation operator $I_h : \widetilde{H}_\Gamma^{1+\alpha} \to \widehat{S}_h(\Omega)$ defined by for $v \in \widetilde{H}_\Gamma^{1+\alpha}$,

$$(I_h u)(X) = u(X), \quad \text{for all nodes } X \text{ of } \mathcal{T}_h.$$

The following lemma was proven in [11, 12, 14].

Lemma 3.1. *There exists a constant $C > 0$ such that*

$$\sum_{T\in\mathcal{T}_h} (\|\phi - I_h\phi\|_{0,T} + h\|\phi - I_h\phi\|_{1,T}) \le Ch^{1+\alpha}\|\phi\|_{\widetilde{H}^{1+\alpha}(\Omega)}, \quad \forall\phi \in \widetilde{H}_\Gamma^{1+\alpha}(\Omega).$$

3.2. Associated bilinear forms

In this subsection, we develop IFEM for nonsymmetric elliptic interface problem. First, we multiply $v_h \in \widehat{S}_h(\Omega)$ to (2.1) and apply integration by parts to obtain that

$$\begin{aligned} \int_\Omega f v_h \mathrm{d}\mathbf{x} &= \sum_{T\in\mathcal{T}_h} \left(\int_T \beta\nabla u \cdot \nabla v_h \mathrm{d}\mathbf{x} - \int_e \beta\nabla u \cdot \mathbf{n} v_h \mathrm{d}s \right) + \sum_{T\in\mathcal{T}_h} \int_T v_h \mathbf{b} \cdot \nabla u \mathrm{d}\mathbf{x} \\ &= \sum_{T\in\mathcal{T}_h} \int_T \beta\nabla u \cdot \nabla v_h \mathrm{d}\mathbf{x} - \sum_{e\in\mathcal{E}_h} \int_e \{\beta\nabla u \cdot \mathbf{n}_e\}_e [v_h]_e \mathrm{d}s + \sum_{T\in\mathcal{T}_h} \int_T (\mathbf{b} \cdot \nabla u) v_h \mathrm{d}\mathbf{x}. \end{aligned} \tag{3.5}$$

We define the bilinear form $t_h(\cdot,\cdot) : H_h(\Omega) \times H_h(\Omega) \mapsto \mathbb{R}^2$ by for all $v, w \in H_h(\Omega) := \widehat{S}_h(\Omega) + H_0^1(\Omega)$,

$$t_h(v, w) = a_h(v, w) + b_h(v, w), \tag{3.6}$$

$$\begin{aligned} a_h(v, w) = &\sum_{T\in\mathcal{T}_h} \int_T \beta\nabla v \cdot \nabla w \mathrm{d}\mathbf{x} - \sum_{e\in\mathcal{E}_h} \int_e \{\beta\nabla v \cdot \mathbf{n}_e\}_e [w]_e \mathrm{d}s \\ &+ \theta \sum_{e\in\mathcal{E}_h} \int_e \{\beta\nabla w\}_e [v]_e \mathrm{d}s + \sum_{e\in\mathcal{E}_h} \frac{\sigma}{|e|} \int_e [v]_e [w]_e \,\mathrm{d}s, \end{aligned} \tag{3.7}$$

$$b_h(v, w) = \sum_{T\in\mathcal{T}_h} \int_T (\mathbf{b} \cdot \nabla v) w \,\mathrm{d}\mathbf{x} + \eta \sum_{e\in\mathcal{E}_h} \int_e \{\mathbf{b} \cdot \mathbf{n} w\}_e [v]_e \,\mathrm{d}s, \tag{3.8}$$

where $|e|$ is the measure of e. Finally, we propose IFEM scheme for convection diffusion type elliptic interface problem: find $u_h \in \widehat{S}_h(\Omega)$ such that

$$t_h(u_h, v_h) = (f, v_h)_\Omega, \quad \forall v_h \in \widehat{S}_h(\Omega). \tag{3.9}$$

Let us explain the parameters appearing in $t_h(\cdot,\cdot)$ in (3.6). First, the parameter θ is motivated by the interior penalty discontinuous Galerkin (DG) method [28, 29] whose values are among $\{-1, 0, 1\}$ and $\sigma > 0$ is a stabilization parameter. The second term of $b_h(\cdot,\cdot)$ is motivated by [28]. Comparing with the equation (3.5), the parameter $\eta \neq 0$ seems non-natural. However, it will be shown that by choosing $\eta = -1$ and $\theta = -1$, we can show the optimal convergence error estimates by defining the dual problem (see proof of Lemma 3.8).

Before closing this subsection, we state a lemma regarding the consistency of the proposed scheme.

Lemma 3.2. *Suppose u is the solution of (2.1)–(2.4) and u_h is the solution of (3.9), then*

$$t_h(u - u_h, v_h) = 0, \quad \forall v_h \in \widehat{S}_h(\Omega). \tag{3.10}$$

Proof. It suffices to show that

$$t_h(u, v_h) = (f, v_h)_\Omega, \quad \forall v_h \in \widehat{S}_h(\Omega).$$

This is simply obtained by the Eq (3.6) and the fact that $[u]_e = 0$ for all $e \in \mathcal{E}_h$ since $u \in H^1(\Omega)$. □

3.3. Optimal error estimates

This subsection is devoted to prove the optimal convergence of the proposed scheme. For this purpose, we introduce an energy like norm on $H_h(\Omega)$:

$$|||\phi|||_h^2 = \sum_{T\in\mathcal{T}_h} \|\beta^{\frac{1}{2}}\nabla\phi\|_{0,T}^2 + \sum_{e\in\mathcal{E}_h} \frac{1}{|e|}\|[\phi]_e\|_{0,e}^2.$$

We list some lemmas which will play essential roles in the forthcoming analysis.

Lemma 3.3. *([29]) Let e be an edge of $T \in \mathcal{T}_h$. Then following holds,*

$$\|\phi\|_{0,e} \leq C_0 h^{-1/2}\left(\|\phi\|_{0,T} + h|\phi|_{1,T}\right), \quad \phi \in H^1(T).$$

Also, the following trace-like inequality can be established for the discrete space.

Lemma 3.4. *There exists some $C_t > 0$ such that the following holds for all $\phi \in \widehat{S}_h(\Omega)$ and $T \in \mathcal{T}_h$ and edges e of T.*

$$\|\beta^{\frac{1}{2}}\nabla\phi\cdot\mathbf{n}_e\|_{0,e} \leq C_t h^{-\frac{1}{2}}\|\beta^{\frac{1}{2}}\nabla\phi\|_{0,T}.$$

Here, C_t is independent of location of interface but C_t depends on β.

Proof. We start by decompose $\nabla\phi$ as

$$\nabla\phi = (\nabla\phi\cdot\mathbf{n}_\Gamma)\mathbf{n}_\Gamma + (\nabla\phi\cdot\mathbf{t}_\Gamma)\mathbf{t}_\Gamma := \mathbf{w} + \mathbf{z},$$

where $\mathbf{n}_\Gamma$ and $\mathbf{t}_\Gamma$ are the unit normal and tangent vector to the interface Γ. Since the functions in $\widehat{S}_h(T)$ satisfies the flux continuity condition, $\beta\mathbf{w} \in H^1(T)$. Also, $\mathbf{z}$ belongs to $H^1(T)$ since $\nabla\phi$ has well defined trace on Γ. Thus, we have that

$$\|\beta\mathbf{w}\cdot\mathbf{n}_e\|_{0,e} \leq C_0 h^{-1/2}\|\beta\mathbf{w}\|_{0,T} \tag{3.11}$$

$$\|\mathbf{z}\cdot\mathbf{n}_e\|_{0,e} \le C_0 h^{-1/2}\|\mathbf{z}\|_{0,T}. \tag{3.12}$$

By the triangular inequality and inequalities (3.11) and (3.12), we have

$$\begin{aligned}
\|\beta^{\frac{1}{2}}\nabla\phi\cdot\mathbf{n}_e\|_{0,e} &\le \|\beta^{\frac{1}{2}}\mathbf{w}\cdot\mathbf{n}_e\|_{0,e} + \|\beta^{\frac{1}{2}}\mathbf{z}\cdot\mathbf{n}_e\|_{0,e}\\
&\le \frac{1}{\underline{\beta}^{\frac{1}{2}}}\|\beta\mathbf{w}\cdot\mathbf{n}_e\|_{0,e} + \overline{\beta}^{\frac{1}{2}}\|\mathbf{z}\cdot\mathbf{n}_e\|_{0,e}\\
&\le C_0 h^{-\frac{1}{2}}\left(\frac{1}{\underline{\beta}^{\frac{1}{2}}}\|\beta\mathbf{w}\|_{0,T} + \overline{\beta}^{\frac{1}{2}}\|\mathbf{z}\|_{0,T}\right)\\
&\le 2C_0\left(\frac{\overline{\beta}}{\underline{\beta}}\right)^{\frac{1}{2}} h^{-\frac{1}{2}}\|\beta^{\frac{1}{2}}\nabla\phi\|_{0,T}.
\end{aligned}$$

□

Also, we remark that interpolation property can be written in $|||\cdot|||_h$-norm.

Lemma 3.5. *There exists a constant $C > 0$ such that*

$$|||\phi - I_h\phi|||_h \le Ch^{\alpha}\|\phi\|_{\widetilde{H}^{1+\alpha}(\Omega)}, \quad \forall\phi\in\widetilde{H}^{1+\alpha}_{\Gamma}(\Omega).$$

Proof. This inequality follows from Lemma 3.1 and trace inequality. □

We are in a position to establish the Gårding inequality of the bilinear form $t_h(\cdot,\cdot)$.

Lemma 3.6. *There exist some constants $C_1 > 0$ and $C_2 > 0$ independent of h and the location of interface such that the following holds whenever $\sigma > \sigma_0$,*

$$C_1|||\phi_h|||_h^2 - C_2\|\phi_h\|^2_{L^2(\Omega)} \le t_h(\phi_h,\phi_h), \quad \forall\phi_h\in\widehat{S}_h(\Omega), \tag{3.13}$$

for some $\sigma_0 > 0$.

Proof. From the definition of the bilinear form, we have

$$\begin{aligned}
t_h(\phi_h,\phi_h) = &\sum_{T\in\mathcal{T}_h}\int_T \beta\nabla\phi_h\cdot\nabla\phi_h\,dx - (1-\theta)\sum_{e\in\mathcal{E}_h}\int_e\{\beta\nabla\phi_h\cdot\mathbf{n}_e\}_e[\phi_h]_e\,ds\\
&+\sum_{T\in\mathcal{T}_h}\int_T(\mathbf{b}\cdot\nabla\phi_h)\phi_h\,dx + \eta\sum_{e\in\mathcal{E}_h}\int_e\{\mathbf{b}\cdot\mathbf{n}\phi_h\}_e[\phi_h]_e\,ds + \sum_{e\in\mathcal{E}_h}\frac{\sigma}{|e|}\int_e[\phi_h]_e^2\,ds.
\end{aligned}$$

Using Cauchy's inequality, we have that

$$\begin{aligned}
&\sum_{e\in\mathcal{E}_h}\int_e|\{\beta\nabla\phi_h\cdot\mathbf{n}_e\}_e[\phi_h]_e\,ds\\
&\quad\le\overline{\beta}^{\frac{1}{2}}\left(h\sum_{e\in\mathcal{E}_h}\|\{\beta^{\frac{1}{2}}\nabla\phi_h\cdot\mathbf{n}_e\}_e\|^2_{0,e}\right)^{\frac{1}{2}}\left(h^{-1}\sum_{e\in\mathcal{E}_h}\|[\phi_h]_e\|^2_{0,e}\right)^{\frac{1}{2}}.
\end{aligned}\tag{3.14}$$

Let T_1^e and T_2^e be neighboring elements sharing the edge e. By applying Lemma 3.4, we have that

$$\begin{aligned}
h\sum_{e\in\mathcal{E}_h}\|\{\beta^{\frac{1}{2}}\nabla\phi_h\cdot\mathbf{n}_e\}_e\|_{0,e}^2 &\le \frac{h}{2}\sum_{e\in\mathcal{E}_h}\left(\|(\beta^{\frac{1}{2}}\nabla\phi_h)_{|_{T_1^e}}\cdot\mathbf{n}_e\|_{0,e}^2+\|(\beta^{\frac{1}{2}}\nabla\phi_h)_{|_{T_2^e}}\cdot\mathbf{n}_e\|_{0,e}^2\right)\\
&\le \frac{C_t^2}{2}\sum_{e\in\mathcal{E}_h}(\|\beta^{\frac{1}{2}}\nabla\phi_h\|_{0,T_1^e}^2+\|\beta^{\frac{1}{2}}\nabla\phi_h\|_{0,T_2^e}^2)\\
&\le C_t^2\sum_{T\in\mathcal{T}_h}\|\beta^{\frac{1}{2}}\nabla\phi_h\|_{0,T}^2. \qquad (3.15)
\end{aligned}$$

Using Eqs (3.14), (3.15) and invoking Young's inequality, for $\delta_1 > 0$, we have that

$$\begin{aligned}
&(1-\theta)\sum_{e\in\mathcal{E}_h}\int_e|\{\beta\nabla\phi_h\cdot\mathbf{n}_e\}_e[\phi_h]_e|\,ds\\
&\qquad\le\frac{\delta_1}{2}\sum_{T\in\mathcal{T}_h}\|\beta^{\frac{1}{2}}\nabla\phi_h\|_{0,T}^2+\frac{(1-\theta)^2\overline{\beta}C_t^2}{2\delta_1}\sum_{e\in\mathcal{E}_h}\frac{1}{|e|}\|[\phi_h]\|_{0,e}^2.
\end{aligned}$$

Similarly, invoking Cauchy Schwarz, Young's inequality and trace inequality, it holds that

$$\begin{aligned}
&\left|\eta\sum_{e\in\mathcal{E}_h}\int_e\{\mathbf{b}\cdot\mathbf{n}\phi_h\}_e[\phi_h]_e\,ds\right|\le|\bar{\mathbf{b}}||\eta|\left(\frac{h}{2}\sum_{e\in\mathcal{E}_h}(\|\phi_h|_{T_1}\|_{0,e}^2+\|\phi_h|_{T_2}\|_{0,e}^2)\right)^{\frac{1}{2}}\left(h^{-1}\sum_{e\in\mathcal{E}_h}\|[\phi_h]_e\|_{0,e}^2\right)^{\frac{1}{2}}\\
&\le|\bar{\mathbf{b}}||\eta|\left(\frac{C_0^2}{2}\sum_{e\in\mathcal{E}_h}(\|\phi_h\|_{0,T_1}^2+h\|\nabla\phi_h\|_{0,T_1}^2+\|\phi_h\|_{0,T_2}^2+h\|\nabla\phi_h\|_{0,T_2}^2)\right)^{\frac{1}{2}}\left(h^{-1}\sum_{e\in\mathcal{E}_h}\|[\phi_h]_e\|_{0,e}^2\right)^{\frac{1}{2}}\\
&\le\frac{\delta_2}{2}\sum_{T\in\mathcal{T}_h}(\|\phi_h\|_{0,T}^2+h\|\nabla\phi_h\|_{0,T}^2)+\frac{|\bar{\mathbf{b}}|^2|\eta|^2C_0^2}{2\delta_2}\sum_{e\in\mathcal{E}_h}\frac{1}{|e|}\|[\phi_h]_e\|_{0,e}^2.
\end{aligned}$$

for any $\delta_2 > 0$. Therefore, we have

$$\begin{aligned}
t_h(\phi_h,\phi_h)\ge&\left(1-\frac{\delta_1}{2}-\frac{h\delta_2}{2\underline{\beta}}\right)\sum_{T\in\mathcal{T}_h}\|\beta^{\frac{1}{2}}\nabla\phi_h\|_{0,T}^2-\frac{\delta_2}{2}\|\phi_h\|_0^2-|\bar{\mathbf{b}}|\sum_{T\in\mathcal{T}_h}\|\phi_h\|_{0,T}\|\nabla\phi_h\|_{0,T}\\
&+\left(\sigma-\frac{(1-\theta)^2\overline{\beta}C_t^2}{2\delta_1}-\frac{|\bar{\mathbf{b}}|^2|\eta|^2C_0^2}{2\delta_2}\right)\sum_{e\in\mathcal{E}_h}\frac{1}{|e|}\|[\phi_h]\|_{0,e}^2.
\end{aligned}$$

Here, invoking Young's inequality

$$\begin{aligned}
|\bar{\mathbf{b}}|\sum_{T\in\mathcal{T}_h}\|\phi_h\|_{0,T}\|\nabla\phi_h\|_{0,T}&\le\frac{\delta_3}{2}\sum_{T\in\mathcal{T}_h}\|\nabla\phi_h\|_{0,T}^2+\frac{|\bar{\mathbf{b}}|^2}{2\delta_3}\sum_{T\in\mathcal{T}_h}\|\phi_h\|_{0,T}^2\\
&\le\frac{\delta_3}{2}\sum_{T\in\mathcal{T}_h}\frac{1}{\underline{\beta}}\|\beta^{\frac{1}{2}}\nabla\phi_h\|_{0,T}^2+\frac{|\bar{\mathbf{b}}|^2}{2\delta_3}\sum_{T\in\mathcal{T}_h}\|\phi_h\|_{0,T}^2
\end{aligned}$$

for $\delta_3 > 0$, Combining above, there holds

$$t_h(\phi_h,\phi_h) \geq \left(1 - \frac{\delta_1}{2} - \frac{h\delta_2}{2\underline{\beta}} - \frac{\delta_3}{2\underline{\beta}}\right)\sum_{T\in\mathcal{T}_h} \|\beta^{\frac{1}{2}}\nabla\phi_h\|_{0,T}^2 - \left(\frac{\delta_2}{2} + \frac{|\overline{\mathbf{b}}|^2}{2\delta_3}\right)\|\phi_h\|_0^2$$
$$+ \left(\sigma - \frac{(1-\theta)^2\overline{\beta}C_t^2}{2\delta_1} - \frac{|\overline{\mathbf{b}}|^2|\eta|^2C_0^2}{2\delta_2}\right)\sum_{e\in\mathcal{E}_h}\frac{1}{|e|}\|[\phi_h]\|_{0,e}^2.$$

Finally, we choose the parameters.

$$\delta_1 = 1/2, \quad \delta_2 = \delta_3 = \frac{\underline{\beta}}{2}.$$

Then, given a sufficiently large $\sigma > 0$, desired inequality holds with

$$C_1 = \frac{1}{4}, \qquad C_2 = \frac{\underline{\beta}}{4} + \frac{|\overline{\mathbf{b}}|^2}{\underline{\beta}}.$$

□

The continuity of the bilinear form $t_h(\cdot,\cdot)$ can be proven by the same techniques used in the proof of Lemma 3.6.

Lemma 3.7. *There exists some C_b such that the following holds when $\sigma > 0$,*

$$t_h(\phi_h,\psi_h) \leq C_b|||\phi_h|||_h \cdot |||\psi_h|||_h, \quad \forall\phi_h,\psi_h \in \widehat{S}_h(\Omega).$$

The following Lemma will play an important role for the proof of the optimal error estimates.

Lemma 3.8. *Let u be the solution of (2.1)–(2.4) and u_h be the solution of (3.9). Suppose that Proposition 2.1. holds with $0.5 < \alpha \leq 1$ and that $\theta = -1$ and $\eta = -1$ in (3.6). Then there exists some $C_3 > 0$ such that following holds*

$$\|u - u_h\|_{L^2(\Omega)} \leq C_3h^\alpha|||u - u_h|||_h. \tag{3.16}$$

Proof. Let $e_h = u - u_h$. We consider dual problem: find $\phi \in \widetilde{H}^{1+\alpha}(\Omega)$ such that

$$\begin{aligned} -\mathrm{div}\beta\nabla\phi - \mathrm{div}(\mathbf{b}\phi) &= e_h, && \text{in } \Omega,\\ [\phi]_\Gamma &= 0, && \text{on } \Gamma,\\ [\beta\nabla\phi\cdot\mathbf{n}]_\Gamma &= 0, && \text{on } \Gamma,\\ \phi &= 0, && \text{in } \partial\Omega. \end{aligned}$$

By the integration by parts, we have

$$\begin{aligned} \|e_h\|_{L^2(\Omega)}^2 &= \sum_{T\in\mathcal{T}_h}\left(\int_T \beta\nabla\phi\cdot\nabla e_h dx + \int_T(\mathbf{b}\phi)\cdot\nabla e_h dx - \sum_{e\in\partial T}(\int_e \beta\nabla\phi\cdot\mathbf{n}e_h ds + \int_e(\mathbf{b}\phi)\cdot\mathbf{n}e_h ds)\right)\\ &= \sum_{T\in\mathcal{T}_h}\left(\int_T \beta\nabla\phi\cdot\nabla e_h dx + \int_T(\mathbf{b}\phi)\cdot\nabla e_h dx\right) - \sum_{e\in\partial T}\left(\int_e\{\beta\nabla\phi\cdot\mathbf{n}\}[e_h]_e ds + \int_e\{\mathbf{b}\cdot\mathbf{n}\phi\}_e[e_h]_e ds\right) \end{aligned}$$

$$= t_h(e_h, \phi).$$

By consistency of the scheme (see (3.10)), we have

$$\|e_h\|^2_{L^2(\Omega)} = t_h(e_h, \phi - I_h\phi).$$

Finally, by Lemma 3.7, Lemma 3.5 and Proposition 2.2, we obtain the following inequality,

$$\begin{aligned}\|e_h\|^2_{L^2(\Omega)} &\le |||e_h|||_h \, |||\phi - I_h\phi|||_h \\ &\le Ch^\alpha |||e_h|||_h \, \|\phi\|_{\widetilde{H}^{1+\alpha}(\Omega)} \\ &\le Ch^\alpha |||e_h|||_h \, \|e_h\|_{L^2(\Omega)}.\end{aligned}$$

□

Finally, we prove the main theorem.

Theorem 3.9. *Under the same assumptions in Lemma 3.8, there exists some* h_0 *and* $C = C(\beta, \mathbf{b}) > 0$ *such that when* $0 < h < h_0$ *the following holds.*

$$\|u - u_h\|_{L^2(\Omega)} + h^\alpha |||u - u_h|||_h \le Ch^{2\alpha}\|f\|_{H^{-1+\alpha}(\Omega)}. \tag{3.17}$$

Proof. It suffices to show that

$$|||u - u_h|||_h \le Ch^\alpha \|f\|_{H^{-1+\alpha}(\Omega)}.$$

From the Gårding's inequality and Lemma 3.8, we have

$$\begin{aligned}C_1|||u - u_h|||^2_h &\le t_h(u - u_h, u - u_h) + C_2\|u - u_h\|^2_{L^2(\Omega)} \\ &\le t_h(u - u_h, u - u_h) + C_2C_3^2h^{2\alpha}|||u - u_h|||^2_h.\end{aligned}$$

We choose h_0 as

$$h_0 := \left(\frac{C_1}{2C_2C_3^2}\right)^{1/2\alpha}.$$

Then, for $0 < h < h_0$, we have

$$\frac{C_1}{2}|||u - u_h|||^2_h \le t_h(u - u_h, u - u_h).$$

Finally, by (3.10), Lemma 3.5, and Proposition 2.2, we have

$$\begin{aligned}\frac{C_1}{2}|||u - u_h|||^2_h &\le t_h(u - u_h, u - u_h) \\ &= t_h(u - u_h, u - I_hu) \\ &= C_bC_Ih^\alpha |||u - u_h|||_h \, \|u\|_{\widetilde{H}^{1+\alpha}(\Omega)} \\ &\le Ch^\alpha |||u - u_h|||_h \, \|f\|_{\widetilde{H}^{-1+\alpha}(\Omega)}.\end{aligned}$$

Dividing $|||u - u_h|||_h$, the desired inequality is obtained. □

4. Control volume based IFEM for convection dominated case

In this section, we introduce a control volume-based IFEM for the convection-dominated case. By multiplying test functions which are piecewise constants on vertex-associated control volumes, a new formulation for the governing equation (2.1)–(2.4) is derived. We accompany the so-called *upwinding* concept, which makes the proposed scheme robust to the magnitude of convection terms. Such upwinding scheme was implemented by following the standard techniques in [30, 31] where vertex values are judiciously chosen in bilinear for the convection term to avoid spurious oscillations. For the convenience of analysis, we assume that β is piecewise constants on each subdomain Ω^- and Ω^+ and that $\alpha = 1$.

4.1. Derivation of control volume-based IFEM

We introduce some notations and spaces. Let V_h be the set of all nodes in $\mathcal{T}_h$. Given an arbitrary vertex $P_i \in V_h$, we denote $\{T_k\}_{k=1}^{N_{P_i}}$ to be the set of triangles sharing P_i as a common vertex. Connecting the barycenters and edge midpoints in $\{T_k\}_{k=1}^{N_{P_i}}$, vertex P_i-associated control volume is obtained, which we denote T_i^* (see Figure 2).

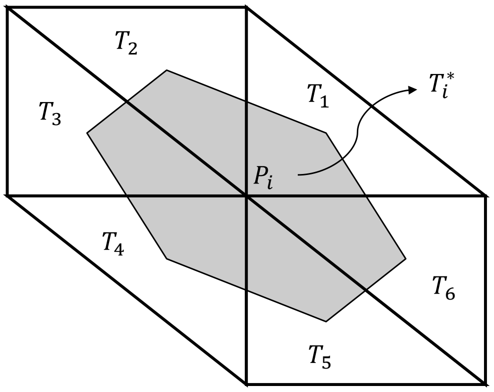


Figure 2. An illustration of a control volume associated with a vertex P_i (gray region). Triangles T_1,T_2,...,T_6 shares P_i as a common node. By connecting barycenters and edge midpoints of triangles, we obtain T_i^*.

Let $\mathcal{D}_h$ be a collection of such control volumes. The test function space is defined as

$$W_h(\Omega) = \{\psi \in L^2(\Omega) \,|\, \psi \text{ is piecewise constant on each control volume in } \mathcal{D}_h\}.$$

In particular, we let $w_i \in W_h(\Omega)$ be the function with $w_i = 1$ on T_i^* and $w_i = 0$ otherwise. Let L_h be the lumping operator from $\widehat{S}_h(\Omega)$ onto $W_h(\Omega)$, i.e., for $u_h \in \widehat{S}_h(\Omega)$, a piecewise constant function $L_h u_h$ is determined by

$$(L_h u_h|_{T_i^*})(P_i) = u_h(P_i), \quad \forall T_i^* \in \mathcal{D}_h.$$

The following Lemma holds by an interpolation property for piecewise-H^1 functions [22].

Lemma 4.1. *There exists some $C > 0$ such that satisfying*

$$\|v - L_h v\|_{0,\Omega} \le Ch|v|_{H^1(\Omega)}, \quad \forall v \in H_h(\Omega). \tag{4.1}$$

Now, let us derive a control volume method. By the relation that $\mathbf{b} \cdot \nabla u = \text{div}(u\mathbf{b}) - u\text{div}\mathbf{b}$, the governing equation (2.1) can be written as

$$-\nabla \cdot \beta \nabla u + \text{div}(u\mathbf{b}) - u\text{div}\mathbf{b} = f, \quad \text{on } \Omega. \tag{4.2}$$

Integrating over each control volume T_i^*, we have

$$-\int_{\partial T_i^*} \beta \nabla u \cdot \mathbf{n} ds + \int_{\partial T_i^*} u\mathbf{b} \cdot \mathbf{n} ds - \int_{T_i^*} u\text{div}\mathbf{b} d\mathbf{x} = \int_{T_i^*} f d\mathbf{x}. \tag{4.3}$$

Motivated by the above equation, the (naive) control-volume method is defined as: find $u_h \in \widehat{S}_h(\Omega)$ such that

$$\bar{a}_h(u_h, v_h) + \bar{b}(u_h, v_h) = (f, L_h v_h), \quad \forall v_h \in \widehat{S}_h, \tag{4.4}$$

where

$$\begin{aligned}
\bar{a}_h(u_h, v_h) &= -\sum_{T_i^* \in \mathcal{D}_h} v_h(P_i) \int_{\partial T_i^*} \beta \nabla u_h \cdot \mathbf{n}\, ds, \quad \bar{b}_h(u_h, v_h) = \bar{b}_h^1(u_h, v_h) + \bar{b}_h^2(u_h, v_h), \\
\bar{b}_h^1(u_h, v_h) &= \sum_{T_i^* \in \mathcal{D}_h} v_h(P_i) \int_{\partial T_i^*} u_h(\mathbf{b} \cdot \mathbf{n}) ds, \\
\bar{b}_h^2(u_h, v_h) &= -\sum_{T_i^* \in \mathcal{D}_h} v_h(P_i) \int_{T_i^*} u_h \text{div}\mathbf{b} dx.
\end{aligned}$$

Here, we introduce some notations. Consider a triangle T^k having P_i as a node and C_k as a center. We define Λ_i^k be the set of adjacent node of P_i in element T_k. For $\ell \in \Lambda_i^k$, let $M_{i\ell}$ be the midpoints of edge $\overline{P_i P_\ell}$. Let us denote $\gamma_{i\ell}^k$ to be the segment $\overline{C_k M_{i\ell}}$, whose outward normal vector is $\mathbf{n}_{i\ell}^k$ (see Figure 3).

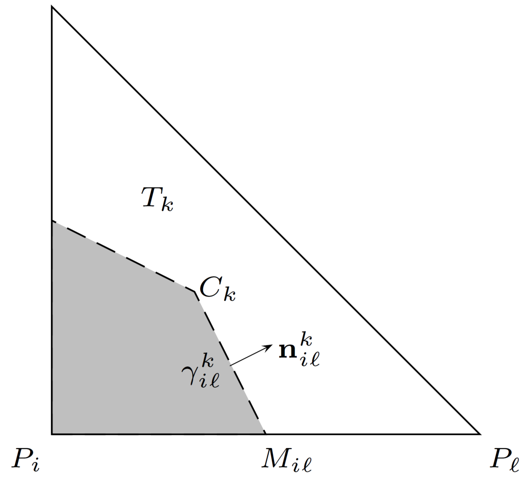


Figure 3. Illustration of $\gamma_{i\ell}^k$ in T_k.

Based on the notations introduced above, we note that

$$\overline{b}_h^1(u_h, v_h) = \sum_{T_i^* \in \mathcal{D}_h} v_h(P_i) \sum_{k=1}^{N_{P_i}} \sum_{\ell \in \Lambda_i^k} \int_{\gamma_{i\ell}^k} u_h \mathbf{b} \cdot \mathbf{n}_{i\ell}^k ds.$$

Here, we use simplified notation $\sum_{\ell \in \Lambda_i}$ (and $\mathbf{n}_i$ resp.) for $\sum_{k=1}^{N_{P_i}} \sum_{\ell \in \Lambda_i^k}$ (and $\mathbf{n}_{i\ell}^k$ resp.) when there is no worry of confusion, i.e.,

$$\overline{b}_h^1(u_h, v_h) = \sum_{T_i^* \in \mathcal{D}_h} v_h(P_i) \sum_{\ell \in \Lambda_i} \int_{\gamma_{i\ell}^k} u_h \mathbf{b} \cdot \mathbf{n}_i ds.$$

Scheme (4.4) can be improved in two directions. First, following Wang [22], we define a new bilinear form:

$$\widetilde{a}_h(u_h, v_h) = \bar{a}_h(u_h, v_h) + \sum_{e \in \mathcal{E}_h} \frac{\sigma}{|e|} \int_e [u_h]_e [v_h]_e \, ds.$$

Also, we modify $\overline{b}_h^1$ using the *upwinding* concepts [30]:

$$\widetilde{b}_h^1(u_h, v_h) = \sum_{T_i^* \in \mathcal{D}_h} v_h(P_i) \sum_{\ell \in \lambda_i} \int_{\gamma_{i\ell}^k} u_h^{i\ell} \mathbf{b} \cdot \mathbf{n}_i ds,$$

where

$$\begin{aligned} u_h^{i\ell} &= \lambda_{i\ell} u_h(P_i) + (1 - \lambda_{i\ell}) u_h(P_\ell), \\ \lambda_{i\ell} &= \begin{cases} 1, & \text{if } \mathbf{b} \cdot \mathbf{n}_i > 0 \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Next, we modify $\overline{b}_h^2$ as

$$\begin{aligned} \widetilde{b}_h^2(u_h, v_h) &= - \sum_{T_i^* \in \mathcal{D}_h} v_h(P_i) u_h(P_i) \int_{T_i^*} \text{div} \mathbf{b} dx \\ &= - \sum_{T_i^* \in \mathcal{D}_h} v_h(P_i) u_h(P_i) \int_{\partial T_i^*} \mathbf{b} \cdot \mathbf{n} ds. \end{aligned}$$

Finally, by defining modified bilinear forms as

$$\begin{aligned} \widetilde{t}_h(u_h, v_h) &= \widetilde{a}_h(u_h, v_h) + \widetilde{b}_h(u_h, v_h), \\ \widetilde{b}_h(u_h, v_h) &= \widetilde{b}_h^1(u_h, v_h) + \widetilde{b}_h^2(u_h, v_h), \end{aligned}$$

we propose (upwinding) control volume method: find $\tilde{u}_h \in \widehat{S}_h(\Omega)$ satisfying

$$\widetilde{t}_h(\tilde{u}_h, v_h) = (f, L_h v_h)_\Omega, \quad \forall v_h \in \widehat{S}_h(\Omega). \tag{4.5}$$

4.2. A consistency error estimate and coerciveness

In this subsection, we estimate the difference between $\widetilde{t_h}(u, v_h)$ and $\widetilde{t_h}(u_h, v_h)$ and show the coerciveness of the control volume based IFEM.

Lemma 4.2. *Suppose u is the solution of (2.1)–(2.4) and $\tilde{u}_h$ is the solution of (4.5). Then, for any $v_h \in \widehat{S}_h(\Omega)$, the following relation holds.*

$$\widetilde{t_h}(u, v_h) - \widetilde{t_h}(u_h, v_h) = \widetilde{b}_h(u, v_h) - \bar{b}_h(u, v_h). \tag{4.6}$$

Moreover, we have that

$$|\bar{b}_h(u, v_h) - \widetilde{b}_h(u, v_h)| \le Ch\|u\|_{\widetilde{H}^2(\Omega)} \|v_h\|_{1,h}. \tag{4.7}$$

Proof. From (4.3), we have

$$\bar{a}_h(u, v_h) + \bar{b}_h(u, v_h) = (f, L_h v_h).$$

However, by the fact that $[u]_e = 0$,

$$\bar{a}_h(u, v_h) = \widetilde{a}_h(u, v_h).$$

Therefore,

$$\widetilde{t_h}(u, v_h) - \widetilde{b}_h(u, v_h) + \bar{b}_h(u, v_h) = (f, L_h v_h),$$

which is equal to $\widetilde{t}(\tilde{u}_h, v_h)$. This proves (4.6). For the proof of (4.7), it suffices to estimate

$$\begin{aligned} A_1 &:= \bar{b}_h^1(u, v_h) - \widetilde{b}_h^1(u, v_h), \\ A_2 &:= \bar{b}_h^2(u, v_h) - \widetilde{b}_h^2(u, v_h). \end{aligned}$$

First, from the definitions of bilinear forms, we have

$$A_1 = \sum_{T_i^* \in \mathcal{D}_h} \sum_{\ell \in \Lambda_i} v_h(P_i) \int_{\gamma_{i\ell}^k} \mathbf{b} \cdot \mathbf{n}_i (u(\mathbf{x}) - \lambda_{i\ell} u(P_i) - (1 - \lambda_{i\ell}) u(P_\ell)) ds.$$

Here, by the fact that $\gamma_{i\ell}^k = \gamma_{\ell i}^k$ and $\mathbf{n}_i = -\mathbf{n}_\ell$, we have

$$A_1 = \frac{1}{2} \sum_{T_i^* \in \mathcal{D}_h} \sum_{\ell \in \Lambda_i} (v_h(P_i) - v_h(P_\ell)) \int_{\gamma_{i\ell}^k} \mathbf{b} \cdot \mathbf{n}_i (u(\mathbf{x}) - \lambda_{i\ell} u(P_i) - (1 - \lambda_{i\ell}) u(P_\ell)) ds. \tag{4.8}$$

By Cauchy's inequality, we have

$$|v_h(P_i) - v_h(P_\ell)| = \left| \int_{\overline{P_i P_\ell}} \frac{\partial v_h}{\partial \mathbf{s}} ds \right| \le h^{\frac{1}{2}} \|\nabla v_h \cdot \mathbf{s}\|_{0, \overline{P_i P_\ell}}, \tag{4.9}$$

where $\mathbf{s}$ is a unit vector in the direction of $\overline{P_i P_\ell}$. When T^k is not an interface element, we have

$$\|\nabla v_h \cdot \mathbf{s}\|_{0, \overline{P_i P_\ell}} \le Ch^{-\frac{1}{2}} |\nabla v_h|_{0, T^k}.$$

When T^k is cut by the interface, we decompose ∇v_h as

$$\nabla v_h = (\nabla v_h \cdot \mathbf{n}_\Gamma)\mathbf{n}_\Gamma + (\nabla v_h \cdot \mathbf{t}_\Gamma) \cdot \mathbf{t}_\Gamma := \mathbf{w} + \mathbf{z},$$

where $\mathbf{n}_\Gamma$ and $\mathbf{t}_\Gamma$ are unit normal and tangential vectors to the interface. We note that $\beta\mathbf{w} \in H^1(T^k)$ and $\mathbf{z} \in H^1(T^k)$ by the construction of the space $\widehat{S}_h(\Omega)$. By the triangle inequality and trace inequality, we have

$$\begin{aligned}
\|\nabla v_h \cdot \mathbf{s}\|_{0,\overline{P_iP_\ell}} &\le \frac{1}{\underline{\beta}}\|\beta\mathbf{w} \cdot \mathbf{s}\|_{0,\overline{P_iP_\ell}} + \|\mathbf{z} \cdot \mathbf{s}\|_{0,\overline{P_iP_\ell}} \\
&\le \left(\frac{1}{\underline{\beta}} C h^{-\frac{1}{2}} |\beta\nabla v_h|_{0,T^k} + C h^{-\frac{1}{2}} |\nabla v_h|_{0,T^k}\right) \\
&\le C h^{-\frac{1}{2}} |\nabla v_h|_{0,T^k}.
\end{aligned} \tag{4.10}$$

Hence, by (4.9) and (4.10), we have that

$$|v_h(P_i) - v_h(P_\ell)| \le C|\nabla v_h|_{0,T^k}. \tag{4.11}$$

For convenience' sake, we use notation u^z for $\lambda_{i\ell}u(P_i) + (1-\lambda_{i\ell})u(P_\ell)$, where u^z can be either $u(P_i)$ or $u(P_\ell)$. Using the similar technique, we can show that

$$|u(\mathbf{x}) - u^z| \le C\|u\|_{\widetilde{H}^2(T^k)}, \quad \mathbf{x} \in \gamma_{i\ell}^k, \quad u^z = u(P_i),\ u(P_\ell). \tag{4.12}$$

Finally, by (4.11) and (4.12) and the fact that $|\gamma_{i\ell}^k| \le h$, we have

$$\begin{aligned}
&\left|(v_h(P_i) - v_h(P_\ell)) \int_{\gamma_{i\ell}^k} \mathbf{b} \cdot \mathbf{n}_i (u(\mathbf{x}) - u^z) ds\right| \\
&\le C\|\nabla v_h\|_{0,T^k} \int_{\gamma_{i\ell}^k} |\mathbf{b} \cdot \mathbf{n}_i| C \|u\|_{\widetilde{H}^2(T^k)} \\
&\le C h \|\nabla v_h\|_{0,T^k} \|u\|_{\widetilde{H}^2(T^k)}.
\end{aligned}$$

Summing over control volumes in (4.8), we have

$$|A_1| \le Ch\|u\|_{\widetilde{H}^2(\Omega)}\|v_h\|_{1,h}. \tag{4.13}$$

To bound A_2, we first note that

$$\begin{aligned}
A_2 &= -\sum_{T_i^* \in \mathcal{D}_h} v_h(P_i) \int_{T_i^*} u \operatorname{div}\mathbf{b} d\mathbf{x} + \sum_{T_i^* \in \mathcal{D}_h} v_h(P_i) \int_{T_i^*} L_h(u) \operatorname{div}\mathbf{b} d\mathbf{x} \\
&= \sum_{T_i^* \in \mathcal{D}_h} v_h(P_i) \int_{T_i^*} (L_h(u) - u) \operatorname{div}\mathbf{b} d\mathbf{x}.
\end{aligned}$$

By applying (4.1), we have

$$|A_2| \le Ch\|u\|_{1,h}\|v_h\|_{1,h}. \tag{4.14}$$

Hence, by (4.13), (4.14), we obtain the desired inequality. □

To show that control-volume based IFEM is coercive, we need some lemmas.

Lemma 4.3. *[20, 22] The following holds for $u \in H_h(\Omega)$, $v_h \in \widehat{S}_h(\Omega)$.*

$$\left|\sum_{T\in\mathcal{T}_h}\int_T \beta\nabla u\cdot\nabla v\,dx + \sum_{T_i^*\in\mathcal{D}_h} v(P_i)\int_{\partial T_i^*}\beta\nabla u\cdot\mathbf{n}\,ds\right| = \left|\sum_{T\in T_h}\int_{\partial T}(\beta\nabla u_h\cdot\mathbf{n}_e)(v_h - L_h v_h)ds\right| \tag{4.15}$$

$$\leq Ch|||u|||_h|||v|||_h. \tag{4.16}$$

The following lemma also plays an important role in proving the coerciveness.

Lemma 4.4. *For $v_h \in \widehat{S}_h(\Omega)$, we have*

$$\widetilde{b}_h^1(v_h, v_h) + \frac{1}{2}\widetilde{b}_h^2(v_h, v_h) = \frac{1}{2}\sum_{T_i^*\in\mathcal{D}_h}\sum_{\ell\in\Lambda_i}\int_{\partial T_i^*}(\mathbf{b}\cdot\mathbf{n}_i)(v_h(P_i) - v_h(P_\ell))^2\left(\lambda_{i\ell} - \frac{1}{2}\right)ds.$$

Proof. From the definitions of bilinear forms, we have

$$\widetilde{b}_h^1(v_h, v_h) + \frac{1}{2}\widetilde{b}_h^2(v_h, v_h) = \sum_{T_i^*\in\mathcal{D}_h}\sum_{\ell\in\Lambda_i} v_h(P_i)\int_{\partial T_i^*}(\mathbf{b}\cdot\mathbf{n}_i)(v_h^{i\ell} - \frac{1}{2}v_h(P_i))ds.$$

Here, from the fact that $\gamma_{i\ell}^k = \gamma_{\ell i}^k$ and $\mathbf{n}_\ell = -\mathbf{n}_i$, we have that

$$\widetilde{b}_h^1(v_h, v_h) + \frac{1}{2}\widetilde{b}_h^2(v_h, v_h) = \frac{1}{2}\sum_{T_i^*\in\mathcal{D}_h}\sum_{\ell\in\Lambda_i}\int_{\partial T_i^*}(\mathbf{b}\cdot\mathbf{n}_i)\left[v_h(P_i)\left(v_h^{i\ell} - \frac{1}{2}v_h(P_i)\right) - v_h(P_\ell)\left(v_h^{\ell i} - \frac{1}{2}v_h(P_\ell)\right)\right]ds.$$

Using the relation that $v_h^{i\ell} = \lambda_{i\ell}v_h(P_i) + (1 - \lambda_{i\ell})v_h(P_i)$ and that $\lambda_{i\ell} = -\lambda_{\ell i}$, we have

$$v_h(P_i)\left(v_h^{i\ell} - \frac{1}{2}v_h(P_i)\right) - v_h(P_\ell)\left(v_h^{\ell i} - \frac{1}{2}v_h(P_\ell)\right) = (v_h(P_j) - v_h(P_\ell))^2\left(\lambda_{i\ell} - \frac{1}{2}\right),$$

from which we have the desired equation. □

We now state theorem regarding coerciveness of $\widetilde{t}_h$ on $\widehat{S}_h(\Omega)$.

Theorem 4.5. *Suppose the condition (2.8) holds. There exists a constant $C > 0$ and h_0 such that whenever $0 < h < h_0$, the following holds*

$$\widetilde{t}_h(v_h, v_h) \geq C|||v_h|||_h^2, \tag{4.17}$$

for all $v_h \in \widehat{S}_h(\Omega)$.

Proof. By the definition of the bilinear form and by the Eq (4.15), we have

$$\begin{aligned}\widetilde{a}_h(v_h, v_h) &= -\sum_{T_i^*\in\mathcal{D}_h} v_h(P_i)\int_{\partial T_i^*}\beta\nabla u_h\cdot\mathbf{n}\,ds + \sum_{e\in\mathcal{E}_h}\frac{\sigma}{|e|}\int_e [v_h]_e^2 ds\\ &= \sum_{T\in\mathcal{T}_h}\int_T \beta\nabla v_h\cdot\nabla v_h d\mathbf{x} + \sum_{e\in\mathcal{E}_h}\frac{\sigma}{|e|}\int_e [v_h]_e^2 ds + \sum_{T\in T_h}\int_{\partial T}(\beta\nabla u_h\cdot\mathbf{n}_e)(L_h v_h - v_h)ds.\end{aligned}$$

Here, by the Eq (4.16) and the definition of $|||\cdot|||_h$, we have that

$$\widetilde{a}_h(v_h, v_h) \geq |||v_h|||_h^2 - Ch|||v_h|||_h^2. \tag{4.18}$$

Next, we bound $\widetilde{b}_h(v_h, v_h)$. By Lemma 4.4, we obtain

$$\begin{aligned}\widetilde{b}_h(v_h, v_h) &= \widetilde{b}_h^1(v_h, v_h) + \frac{1}{2}\widetilde{b}_h^2(v_h, v_h) + \frac{1}{2}\widetilde{b}_h^2(v_h, v_h)\\ &= \frac{1}{2}\sum_{T_i^*\in\mathcal{D}_h}\sum_{\ell\in\Lambda_i}\int_{\partial T_i^*}(\mathbf{b}\cdot\mathbf{n}_i)(v_h(P_i)-v_h(P_\ell))^2\left(\lambda_{i\ell}-\frac{1}{2}\right)ds\\ &\quad -\frac{1}{2}\sum_{T_i^*\in\mathcal{D}_h}v_h(P_i)\sum_{\ell\in\Lambda_i}\int_{\Gamma_{i\ell}}\mathbf{b}\cdot\mathbf{n}_i v_h(P_i)ds\end{aligned}$$

Here, since $(\mathbf{b}\cdot\mathbf{n}_i)(\lambda_{i\ell}-1/2) > 0$, we have that

$$\widetilde{b}_h(v_h, v_h) \geq -\frac{1}{2}\sum_{T_i^*\in\mathcal{D}_h}v_h(P_i)\sum_{\ell\in\Lambda_i}\int_{\Gamma_{i\ell}}\mathbf{b}\cdot\mathbf{n}_i v_h(P_\ell)ds. \tag{4.19}$$

Finally, combining (4.18) and (4.19), we have

$$\begin{aligned}\widetilde{t}_h(v_h, v_h) &= \widetilde{a}_h(v_h, v_h) + \widetilde{b}_h(v_h, v_h)\\ &\geq |||v_h|||_h^2 - Ch|||v_h|||_h^2 - \frac{1}{2}\sum_{T_i^*\in\mathcal{D}_h}v_h(P_i)\sum_{k\in\Lambda_i}\int_{\Gamma_{i\ell}}\mathbf{b}\cdot\mathbf{n}_i v_h(P_\ell)ds\\ &= |||v_h|||_h^2 - Ch|||v_h|||_h^2 - \frac{1}{2}\sum_{T_i^*\in\mathcal{D}_h}\int_{T_i^*}(\mathrm{div}\mathbf{b})|L_h v_h|^2dx.\end{aligned}$$

Here, under the condition (2.8), the desired inequality holds by taking h_0 small enough. □

Corollary 4.1. *By Lax-Milgram theorem [25], Theorem 4.5 ensures the existence and uniqueness of the proposed control volume based IFEM when h is sufficiently small.*

4.3. An optimal error estimate in $|||\cdot|||_h$-norm

In this subsection, we prove optimal error estimate for control volume based IFEM in energy-like norm. We start by listing some lemmas.

Lemma 4.6. *For $u \in \widetilde{H}_\Gamma^2(\Omega)$, the following inequality holds.*

$$h\sum_{e\in\mathcal{E}_h}\|\nabla(u-I_hu)\|_{0,e}^2 + h^{-1}\sum_{e\in\mathcal{E}_h}\|[u-I_hu]_e\|_{0,e}^2 \leq Ch^2\|u\|_{\widetilde{H}^2(\Omega)}^2. \tag{4.20}$$

Proof. Let $\phi = u - I_hu \in \widetilde{H}^1(\Omega)$. Suppose e is the common edge of elements T_1 and T_2 in $\mathcal{T}_h$. Then,

$$h^{-1/2}\|[\phi]_e\|_{0,e} \leq \frac{h^{-1/2}}{2}\left(\|\phi|_{T_1}\|_{0,e} + \|\phi|_{T_2}\|_{0,e}\right).$$

Here, we apply the trance inequality (Lemma 3.3) and interpolation property (Lemma 3.1) to obtain

$$\begin{aligned} h^{-1/2}\|[\phi]_e\|_{0,e} &\le \frac{1}{2}\left[h^{-1}\left(\|\phi\|_{0,T_1} + \|\phi\|_{0,T_2}\right) + \left(|\phi|_{1,T_1} + |\phi|_{1,T_2}\right)\right] \\ &\le Ch\|u\|_{\widetilde{H}^2(T)}. \end{aligned} \tag{4.21}$$

The inequality

$$h^{1/2}\|\nabla\phi\|_{0,e} \le Ch\|u\|_{\widetilde{H}^2(T)}. \tag{4.22}$$

can be proved by the similar techniques used in the proof of Lemma 3.4. By summing inequalities (4.21) and (4.22) over all edges, we have the desired inequality. □

We introduce some notations.

$$b_h^1(u,v) = \sum_{T\in\mathcal{T}_h}\int_T \mathrm{div}(u\mathbf{b})vd\mathbf{x}, \quad b_h^2(u,v) = -\sum_{T\in\mathcal{T}_h}\int_T u\mathrm{div}(\mathbf{b})vd\mathbf{x}.$$

Clearly, for the bilinear form $b_h(\cdot,\cdot)$ defined in Section 3 (with parameter $\eta = -1$), it holds that

$$b_h(u,v) = b_h^1(u,v) + b_h^2(u,v) - \sum_{e\in\mathcal{E}_h}\int_e \{\mathbf{b}\cdot\mathbf{n}v\}_e[u]_e\,\mathrm{d}s.$$

Lemma 4.7. *Suppose $u \in \widetilde{H}^2_\Gamma(\Omega)$ and $v_h \in \widehat{S}_h(\Omega)$. Let $w = u - I_h u \in H_h(\Omega)$. Then, the following relation holds.*

$$b_h^1(w,v_h) + b_h^2(w,v_h) - \widetilde{b}_h^1(w,v_h) - \widetilde{b}_h^2(w,v_h) \le Ch\|u\|_{\widetilde{H}^2(\Omega)}\|v_h\|_{1,h}. \tag{4.23}$$

Proof. Let E be the left hand side of (4.23). Then,

$$\begin{aligned} E = &(b_h^1(w,v_h) - b_h^1(w,L_hv_h)) + (b_h^1(w,L_hv_h) - \widetilde{b}_h^1(w,v_h)) + (b_h^2(w,v_h) - b_h^2(w,L_hv_h)) \\ &+ (b_h^2(w,L_hv_h) - \widetilde{b}_h^2(w,v_h)) := E_1 + E_2 + E_3 + E_4. \end{aligned}$$

By Lemma 4.1,

$$|E_1| \le Ch\|w\|_{1,h}\|v_h\|_{1,h} \tag{4.24}$$
$$|E_3| \le Ch\|w\|_{1,h}\|v_h\|_{1,h}. \tag{4.25}$$

By the definitions of bilinear forms, we have

$$|E_4| = \left|\sum_{T_i^*\in\mathcal{D}_h} v_h(P_i)\int_{T_i^*}\mathrm{div}\mathbf{b}(L_u(w) - w)d\mathbf{x}\right|$$

Then, by Lemma 4.1

$$|E_4| = Ch\|w\|_{1,h}\|v_h\|_{1,h}. \tag{4.26}$$

Finally, it remains to bound E_2. Here, we notice that $b_h^1(w, L_h v_h) = \overline{b}_h(w, v_h)$. By the definitions of the bilinear forms and by the fact that $\gamma_{i\ell}^k = \gamma_{\ell i}^k$ and $\mathbf{n}_i = -\mathbf{n}_\ell$, we have

$$\begin{aligned}
E_2 &= \overline{b}_h^1(w, v_h) - \widetilde{b}_h^1(w, v_h) \\
&= \sum_{T_i^* \in \mathcal{D}_h} \sum_{\ell \in \Lambda_i} v_h(P_i) \int_{\gamma_{i\ell}^k} \mathbf{b} \cdot \mathbf{n}_i (w(\mathbf{x}) - \lambda_{i\ell} w(P_i) - (1 - \lambda_{i\ell}) w(P_\ell)) ds \\
&= \frac{1}{2} \sum_{T_i^* \in \mathcal{D}_h} \sum_{\ell \in \Lambda_i} (v_h(P_i) - v_h(P_\ell)) \int_{\gamma_{i\ell}^k} \mathbf{b} \cdot \mathbf{n}_i (w(\mathbf{x}) - w^z) ds,
\end{aligned}$$

where $w^z = \lambda_{i\ell} w(P_i) + (1 - \lambda_{i\ell}) w(P_\ell)$. Here, we can establish the below inequalities following the same techniques to prove (4.11), i.e.,

$$\begin{aligned}
&|v_h(P_i) - v_h(P_\ell)| \leq C \|\nabla v_h\|_{0,T^k} \\
&|w(\mathbf{x}) - w^z| \leq \|w\|_{\widetilde{H}^2(T^k)}, \quad \mathbf{x} \in \gamma_{i\ell}^k, \quad w^z = w(P_i),\ w(P_z).
\end{aligned}$$

Then, using the fact that $|\gamma_{i\ell}^k| \leq h$, we have that

$$\left| (v_h(P_i) - v_h(P_\ell)) \int_{\gamma_{i\ell}^k} \mathbf{b} \cdot \mathbf{n}_i (w(\mathbf{x}) - w^z) ds \right| \leq Ch \|w\|_{\widetilde{H}^2(T^k)} \|\nabla v_h\|_{0,h}.$$

Summing over the control volumes, we have that

$$|E_2| = \left| \overline{b}_h^1(w, v_h) - \widetilde{b}_h^1(w, v_h) \right| \leq Ch \|w\|_{\widetilde{H}^2(\Omega)} \|v_h\|_{1,h}. \tag{4.27}$$

Then, by noting that

$$\|w\|_{1,h} \leq Ch \|u\|_{\widetilde{H}^2(\Omega)},$$

in (4.24), (4.25), (4.26) and noting that

$$\|w\|_{\widetilde{H}^2(\Omega)} \leq C \|u\|_{\widetilde{H}^2(\Omega)},$$

in (4.27), we have

$$|E| \leq Ch \|u\|_{\widetilde{H}^2(\Omega)} \|v_h\|_{1,h}.$$

□

Before stating the main theorem, we define some notations.

$$\begin{aligned}
&E_h(u, v_h) = E_h^a(u, v_h) + E_h^b(u, v_h), \\
&E_h^a(u, v_h) = a_h(u, v_h) - \widetilde{a}_h(u, v_h), \quad E_h^b(u, v_h) = b_h(u, v_h) - \widetilde{b}_h(u, v_h).
\end{aligned}$$

Finally, we are in a position to prove the main theorem.

Theorem 4.8. *Suppose (2.8) holds. Let $\tilde{u}_h$ be the solution of (4.5) and let u be the solution of (2.1)–(2.4). Then,*

$$|||u - \tilde{u}_h|||_h \le Ch\|u\|_{\widetilde{H}^2(\Omega)}.$$

Here, the constant C depends on β and $\mathbf{b}$.

Proof. Let $v_h = \tilde{u}_h - I_h u$. By the coerciveness property (4.17), we have

$$\begin{aligned} C|||\tilde{u}_h - I_h u|||^2 \le& \widetilde{t_h}(\tilde{u}_h - I_h u, v_h) \\ =& \widetilde{t_h}(u - I_h u, v_h) + (\widetilde{t_h}(\tilde{u}, v_h) - \widetilde{t_h}(u, v_h)) \\ =& t_h(u - I_h u, v_h) - E_h(u - I_h u, v_h) + (\widetilde{t_h}(\tilde{u}, v_h) - \widetilde{t_h}(u, v_h)) \\ :=& B_1 + B_2 + B_3 \end{aligned} \tag{4.28}$$

By the continuity of t_h and interpolation property, we have

$$\begin{aligned} |B_1| \le& C|||u - I_h u|||_h|||v_h|||_h \\ \le& Ch\|u\|_{\widetilde{H}^2(\Omega)}|||v_h|||_h. \end{aligned}$$

Next, B_3 is bounded by Lemma 4.2, i.e.,

$$|B_3| \le Ch\|u\|_{\widetilde{H}^2(\Omega)}|||v_h|||_h.$$

Finally, we bound B_2. For the convenience, we denote $u - I_h u$ by w. By the definitions of bilinear forms and Lemma 4.3, we have

$$\begin{aligned} E_h^a(w, v_h) =& \sum_{T\in\mathcal{T}_h} \int_T \beta\nabla w \cdot \nabla v_h \, d\mathbf{x} + \sum_{T_i^*\in\mathcal{D}_h} v_h(P_i) \int_{\partial T_i^*} \beta\nabla w \cdot \mathbf{n} \, ds \\ & - \sum_{e\in\mathcal{E}_h} \int_e \{\beta\nabla w \cdot \mathbf{n}_e\}_e [v_h]_e ds - \sum_{e\in\mathcal{E}_h} \int_e \{\beta\nabla v_h \cdot \mathbf{n}_e\}_e [w]_e ds \\ =& \sum_{T\in\mathcal{T}_h} \int_{\partial T} (\beta\nabla w \cdot \mathbf{n}_e)(v_h - L_h v_h) ds - \sum_{e\in\mathcal{E}_h} \int_e \{\beta\nabla w \cdot \mathbf{n}_e\}_e [v_h]_e ds - \sum_{e\in\mathcal{E}_h} \int_e \{\beta\nabla v_h \cdot \mathbf{n}_e\}_e [w]_e ds. \end{aligned}$$

By applying (4.16) and (4.20), we have that

$$E_h^a(w, v_h) \le Ch\|u\|_{\widetilde{H}^2(\Omega)}|||v_h|||_h. \tag{4.29}$$

Next, we bound $E_h^b(w, v_h)$. From the definitions of bilinear forms,

$$E_h^b(w, v_h) = b_h^1(w, v_h) + b_h^2(w, v_h) - \widetilde{b}_h^1(w, v_h) - \widetilde{b}_h^2(w, v_h) - \sum_{e\in\mathcal{E}_h} \int_e \{\mathbf{b}v_h \cdot \mathbf{n}_e\}_e [w]_e ds.$$

By (4.23), we have

$$\left| b_h^1(w, v_h) + b_h^2(w, v_h) - \widetilde{b}_h^1(w, v_h) - \widetilde{b}_h^2(w, v_h) \right| \le Ch\|u\|_{\widetilde{H}^2(\Omega)}|||v_h|||_h.$$

Also, by (4.20) and trace inequality, we have

$$\left|\sum_{e\in\mathcal{E}_h}\int_e \{\mathbf{b}v_h\cdot\mathbf{n}_e\}_e[w]_e ds\right| \le Ch\|u\|_{\widetilde{H}^2(\Omega)}|||v_h|||_h.$$

Hence, we have that

$$E_h^b(u - I_h u, v_h) \le Ch\|u\|_{\widetilde{H}^2(\Omega)}|||v_h|||_h. \tag{4.30}$$

From the (4.29) and (4.30), we have

$$|B_2| \le Ch\|u\|_{\widetilde{H}^2(\Omega)}|||v_h|||_h.$$

Now, combining estimates for B_1, B_2 and B_3 in (4.28), we obtain

$$|||\tilde{u}_h - I_h u|||_h \le Ch\|u\|_{\widetilde{H}^2(\Omega)}.$$

Finally, by the triangle inequality and interpolation property, and Proposition 2.2, the desired inequality is obtained.

$$|||u - \tilde{u}_h|||_h \le |||u - I_h u|||_h + |||\tilde{u}_h - I_h u|||_h \le Ch\|u\|_{\widetilde{H}^2(\Omega)}.$$

□

5. Numerical results

In this section, we document two examples which support the theory in the previous sections. For the first example, the L^2 and H^1-errors of IFEM (proposed in Section 3) are reported. In the second example, we consider convection-dominated problem where both boundary and internal layers appear. We see that control volume-based IFEM (proposed in Section 4) yields numerical solutions without non-physical oscillations. In both examples, the domain is $\Omega = [-1, 1]^2$ and the interface is given by the level set of $L(x, y) := (x/r_0)^2 + (y/r_0)^2 - 1$ where r_0 will be chosen later.

Example 1

We consider the exact solution

$$p = \begin{cases} L(x,y)/\beta^- & \text{in } \Omega^-, \\ L(x,y)/\beta^+ & \text{in } \Omega^+, \end{cases}$$

and parameters $\beta^- = 1000$, $\beta^+ = 1$, $r_0 = 0.8$ and $\mathbf{b} = (\sin y + x, \cos x + y)$. We choose $\theta = -1$ in the definition of $t_h(\cdot,\cdot)$. To see the effect of the line integral for the convection terms (second term of $b_h(\cdot,\cdot)$), we provide the results with $\eta = -1$ and $\eta = 0$. The L^2 and piecewise H^1-errors by IFEM with $\eta = -1$ are reported in Table 1 and that by IFEM with $\eta = 0$ are reported in Table 2. We observe the optimal error convergent rates in Table 1 which confirms the error estimates in Section 3. However, similar results are obtained by IFEM with $\eta = 0$ in Table 2. This suggests that although optimal error estimates are carried out with $\eta = -1$, one may simply set $\eta = 0$ in practice. Other choices of β and $\mathbf{b}$ parameters yield similar results.

Table 1. L^2 and H^1-errors of IFEM with $\eta = -1$ for Example 1.

h	$\|\|u - u_h\|\|_{L^2(\Omega)}$	order	$\|\|u - u_h\|\|_{1,h}$	order
$1/2^3$	9.184E-03		2.702E-01	
$1/2^4$	2.946E-03	1.640	1.275E-01	1.083
$1/2^5$	5.742E-04	2.359	5.905E-02	1.111
$1/2^6$	1.022E-04	2.491	2.845E-02	1.054
$1/2^7$	2.672E-05	1.935	1.416E-02	1.007
$1/2^8$	6.578E-06	2.022	7.057E-03	1.004
$1/2^9$	1.609E-06	2.032	3.521E-03	1.003

Table 2. L^2 and H^1-errors of IFEM $\eta = 0$ for Example 1.

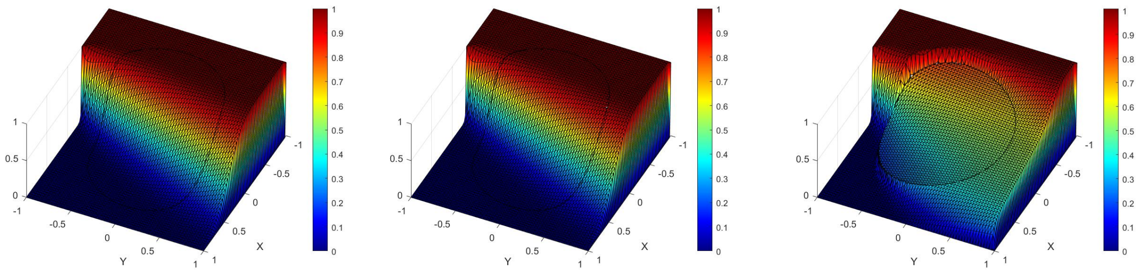
h	$\|\|u - u_h\|\|_{L^2(\Omega)}$	order	$\|\|u - u_h\|\|_{1,h}$	order
$1/2^3$	9.199E-03		2.701E-01	0.747
$1/2^4$	2.971E-03	1.631	1.275E-01	1.083
$1/2^5$	5.752E-04	2.369	5.905E-02	1.111
$1/2^6$	1.018E-04	2.498	2.845E-02	1.054
$1/2^7$	2.659E-05	1.937	1.416E-02	1.007
$1/2^8$	6.542E-06	2.023	7.057E-03	1.004
$1/2^9$	1.599E-06	2.033	3.521E-03	1.003

Example 2

Here, we consider an example with layers whose exact solution is unknown. We impose the boundary conditions as

$$\begin{cases} u(x,y) = 1, & x = -1, \\ u(x,y) = 1, & y = -1 \text{ and } -1 \leq x \leq -2/3, \\ u(x,y) = 0, & \text{otherwise,} \end{cases}$$

and a homogeneous outer-source, i.e., $f = 0$. The convection parameter is $\mathbf{b} = (t, 3t)$ where $t = 10^3$ or 10^9. Graphs of numerical solutions obtained by control volume-based IFEM with different choices of parameters are reported in Figure 4. We see that control volume-based IFEM generate solutions without any non-physical oscillations. We believe similar effect can be obtained with other stabilizing schemes such as streamline upwind/Petrov-Galerkin [32] or local projection stabilization [33].



$\beta^{-} = 1, \beta^{+} = 1$
$\boldsymbol{b} = (10^9, 3 \cdot 10^9)$

$\beta^{-} = 1000, \beta^{+} = 1$
$\boldsymbol{b} = (10^9, 3 \cdot 10^9)$

$\beta^{-} = 1000, \beta^{+} = 1$
$\boldsymbol{b} = (10^3, 3 \cdot 10^3)$

Figure 4. Graphs of numerical solutions obtained by control volume-based IFEM with respect to different parameters. The parameters are $\beta^{-} = 1, \beta^{+} = 1, t = 10^9$ (left), $\beta^{-} = 1000, \beta^{+} = 1, t = 10^9$, (middle), $\beta^{-} = 1000, \beta^{+} = 1, t = 10^3$, (right).

6. Conclusions and discussions

In this work, we develop two structured grid-based methods for nonsymmetric elliptic interface problems. We first develop immersed finite element method whose bilinear form contains judiciously defined line integrals for convection terms. By establishing Gårding's inequality on immersed finite element space, we prove the optimal error estimates. For the convection-dominated case, we design control volume-based immersed finite element methods. Using the upwinding schemes, the proposed scheme is robust to the magnitude of the convection terms. Optimal error estimates in the H^1-norm are carried out for the second scheme. Numerical experiments support our analysis. For the convection-dominated case, we see that the results obtained by control-volume based IFEM show no non-physical oscillations.

Acknowledgements

First author was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1C1C1A01005396).

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. J. Bear, *Dynamics of fluids in porous media*, Elsevier, New York, 1972.
2. P. Bastian, A fully-coupled discontinuous Galerkin method for two-phase flow in porous media with discontinuous capillary pressure, *Computat. Geosci.*, **18** (2014), 779–796. https://doi.org/10.1007/s10596-014-9426-y
3. I. L. Chern, J. G. Liu, W. C. Wang, Accurate evaluation of electrostatics for macromolecules in solution, *Meth. Appl. Anal.*, **10** (2003), 309–328. https://dx.doi.org/10.4310/MAA.2003.v10.n2.a9

4. L. Chen, M. J. Holst, J. Xu, The finite element approximation of the nonlinear Poisson–Boltzmann equation, *SIAM J. Numer. Anal.*, **45** (2007), 2298–2320. https://doi.org/10.1137/060675514

5. I. Babuška, The finite element method for elliptic equations with discontinuous coefficients *Computing*, **5** (1970), 207–213. https://doi.org/10.1007/BF02248021

6. N. Moës, T. Belytschko, Extended finite element method for cohesive crack growth, *Eng. Fract. Mech.*, **69** (2002), 813–833. https://doi.org/10.1016/S0013-7944(01)00128-X

7. J. Chessa, T. Belytschko, An extended finite element method for two-phase fluids, *J. Appl. Mech.*, **70** (2003), 10–17. https://doi.org/10.1115/1.1526599

8. G. Legrain, N. Moës, E. Verron, Stress analysis around crack tips in finite strain problems using the extended finite element method, *Int. J. Numer. Meth. Eng.*, **63** (2005), 290–314. https://doi.org/10.1002/nme.1291

9. M. Cervera, G. B. Barbat, M. Chiumenti, J. Y. Wu, A comparative review of xfem, mixed fem and phase-field models for quasi-brittle cracking, *Arch. Comput. Method. E.*, **29**, (2022), 1009–1083. https://doi.org/10.1007/s11831-021-09604-8

10. G. Jo, D. Y. Kwak, Geometric multigrid algorithms for elliptic interface problems using structured grids, *Numer. Algorithms*, **81** (2019), 211–235. https://doi.org/10.1007/s11075-018-0544-9

11. Z. Li, T. Lin, Y. Lin, R. C. Rogers, An immersed finite element space and its approximation capability, *Numer. Meth. Part. D. E.*, **20** (2004), 338–367. https://doi.org/10.1002/num.10092

12. X. He, T. Lin, Y. Lin, Approximation capability of a bilinear immersed finite element space, *Numer. Meth. Part. D. E.*, **24** (2008), 1265–1300. https://doi.org/10.1002/num.20318

13. S. H. Chou, D. Y. Kwak, K. T. Wee, Optimal convergence analysis of an immersed interface finite element method, *Adv. Comput. Math.*, **33** (2010), 149–168. https://doi.org/10.1007/s10444-009-9122-y

14. D. Y. Kwak, K. T. Wee, K. S. Chang, An analysis of a broken P_1-nonconforming finite element method for interface problems, *SIAM J. Numer. Anal.*, **48** (2010), 2117–2134. https://doi.org/10.1137/080728056

15. D. Y. Kwak, S. Jin, D. Kyeong, A stabilized P_1-nonconforming immersed finite element method for the interface elasticity problems, *ESAIM: Math. Model. Num.*, **51** (2017), 187–207. https://doi.org/10.1051/m2an/2016011

16. G. Jo, D. Y. Kwak, A reduced Crouzeix-Raviart immersed finite element method for elasticity problems with interfaces, *Comput. Meth. Appl. Math.*, **20** (2020), 501–516 https://doi.org/10.1515/cmam-2019-0046

17. G. Jo, D. Y. Kwak, An IMPES scheme for a two-phase flow in heterogeneous porous media using a structured grid, *Comput. Method. Appl. M.*, **317** (2017), 684–701. https://doi.org/10.1016/j.cma.2017.01.005

18. I. Kwon, D. Y. Kwak, G. Jo, Discontinuous bubble immersed finite element method for Poisson-Boltzmann-Nernst-Planck model, *J. Comput. Phys.*, **438** (2021), 110370. https://doi.org/10.1016/j.jcp.2021.110370

19. Y. Choi, G. Jo, D. Y. Kwak, Y. J. Lee, Locally conservative discontinuous bubble scheme for Darcy flow and its application to Hele-Shaw equation based on structured grids, *Numer. Algorithms*, (2022), https://doi.org/10.1007/s11075-022-01333-8

20. R. E. Ewing, Z. Li, T. Lin, Y. Lin, The immersed finite volume element methods for the elliptic interface problems, *Math. Comput. Simulat.*, **50** (1999), 63–76. https://doi.org/10.1016/S0378-4754(99)00061-0
21. X. M. He, T. Lin, Y. Lin, A bilinear immersed finite volume element method for the diffusion equation with discontinuous coefficient, *Commun. Comput. Phys.*, **6** (2009), 185–202. 10.4208/cicp.2009.v6.p185
22. Q. Wang, Z. Zhang, A stabilized immersed finite volume element method for elliptic interface problems, *Appl. Numer. Math.*, **143** (2019), 75–87. https://doi.org/10.1016/j.apnum.2019.03.010
23. Q. Wang, Z. Zhang, L. Wang, New immersed finite volume element method for elliptic interface problems with non-homogeneous jump conditions, *J. Comput. Phys.*, **427** (2021), 110075. https://doi.org/10.1016/j.jcp.2020.110075
24. H. G. Roos, M. Stynes, L. Tobiska, *Robust numerical methods for singularly perturbed differential equations: Convection-diffusion-reaction and flow problems*, Springer Science and Business Media, 2008. https://doi.org/10.1007/978-3-540-34467-4
25. S. C. Brenner, L. R. Scott, *The mathematical theory of finite element methods*, New York: Springer, 2008. https://doi.org/10.1007/978-1-4757-4338-8
26. Ja A. Roĭtberg, Z. G. Šeftel, A theorem on homeomorphisms for elliptic systems and its applications, *Math. USSR-Sbornik*, **7** (1969), 439–465. https://doi.org/10.1070/SM1969v007n03ABEH001099
27. J. H. Bramble, J. T. King, A finite element method for interface problems in domains with smooth boundaries and interfaces, *Adv. Comput. Math.*, **6** (1996), 109–138. https://doi.org/10.1007/BF02127700
28. M. F. Wheeler, An elliptic collocation-finite element method with interior penalties, *SIAM J. Numer. Anal.,* **15** (1978), 152–161. https://doi.org/10.1137/0715010
29. D. N. Arnold, An interior penalty finite element method with discontinuous elements, *SIAM J. Numer. Anal.,* **19** (1982), 742–760. https://doi.org/10.1137/0719052
30. K. Ohmori, T. Ushijima, A technique of upstream type applied to a linear nonconforming finite element approximation of convective diffusion equations, *RAIRO Anal. Numérique*, **18** (1984), 309–322. https://doi.org/10.1051/m2an/1984180303091
31. R. E. Bank, J. F. Burgler, W. Fichner, R. K. Smith, Some upwinding techniques for finite element approximations of convection-diffusion equations, *Numer. Math.*, **58** (1990), 185–202. https://doi.org/10.1007/BF01385618
32. A. N. Brooks, T. J. R. Hughes, Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations, *Comput. Method. Appl. M.*, **32** (1982), 199–259. https://doi.org/10.1016/0045-7825(82)90071-8
33. N. Ahmed, V. John, G. Matthies, J. Novo, A local projection stabilization/continuous Galerkin–Petrov method for incompressible flow problems, *Appl. Math. Comput.*, **333** (2018), 304–324. https://doi.org/10.1016/j.amc.2018.03.088