J. Appl. Numer. Optim. 4 (2022), No. 3, pp. 299-314
Available online at http://jano.biemdas.com
https://doi.org/10.23952/jano.4.2022.3.01

THE PENALTY METHOD FOR THE BOUNDARY CONDITION OF THE DARCY SYSTEM

GUANYU ZHOU

Institute of Fundamental and Frontier Sciences,
University of Electronic Science and Technology of China, Chengdu 610054, China
School of Mathematical Sciences,
University of Electronic Science and Technology of China, Chengdu 610031, China

Abstract. We propose a penalty method to approximate the boundary condition of the Darcy system. For the penalty variational problem, we establish the well-posedness theorem and prove the optimal error estimates of the penalty in the continuous sense. Moreover, we apply the finite element method using RT0/P0 element to discretize the penalty variational problem. The convergence rate depending on both the penalty parameter and mesh size, as well as the applicability of the discrete scheme, are investigated through several numerical experiments on the cases with smooth/non-smooth and convex/non-convex domains.

Keywords. Error analysis; Finite element method; Linearizing-decoupling; Peterlin viscoelastic model.

1. INTRODUCTION

The Darcy system is a widely used PDE model for the flow in the porous medium. Coupled with the free flow system (e.g., the (Navier-)Stokes equations), it has been applied to various real-world simulations, for instance, industrial filtering [1, 2, 3], groundwater system [4, 5], vuggy porous medium [6], blood flow in tumors [7], and so on.

The velocity $\boldsymbol{u}$ and pressure p of the Darcy flow in a bounded smooth domain $\Omega \subset \mathbb{R}^d$ ($d = 2,3$) are described by the following PDE system:

$$-\boldsymbol{u} = K\nabla p \quad \text{in } \Omega, \tag{1.1a}$$

$$\nabla \cdot \boldsymbol{u} = f \quad \text{in } \Omega, \tag{1.1b}$$

$$\boldsymbol{u} \cdot \boldsymbol{n} = g \quad \text{on } \Gamma, \tag{1.1c}$$

where $\Gamma := \partial\Omega$, K is a positive-definite symmetry matrix representing the permeability, and $f : \Omega \to \mathbb{R}$ and $g : \Gamma \to \mathbb{R}$ are given functions satisfying the compatibility condition:

$$\int_\Omega f \, dx = \int_\Gamma g \, ds. \tag{1.2}$$

The Governing equation (1.1a) is called the Darcy law, saying the velocity of the porous medium flow is proportional to the pressure. The flow is incompressible if $f = 0$. Note that (1.1) is

E-mail address: zhoug@uestc.edu.cn.
Received September 9, 2022; Accepted October 10, 2022.

equivalent to the elliptic problem of the pressure p:

$$-\nabla\cdot(K\nabla p) = f \qquad \text{in } \Omega, \tag{1.3a}$$

$$K\nabla p\cdot \boldsymbol{n} = g \qquad \text{on } \Gamma. \tag{1.3b}$$

The numerical approximation for the Darcy system (equivalently, the mixed form of the elliptic problem) has been extensively studied. It is worth mentioning that solving the elliptic problem (1.3) directly is not equivalent to solving the mixed from (1.1) in the discretization sense. One can apply the finite element method (FEM) with conforming P_k-element to approximate p of (1.3), and then obtain the approximation of $\boldsymbol{u}$ by taking the derivative of the numerical solution p_h at each element. This approach is a standard FEM to the elliptic problem and is simple to carry out, and the numerical analysis of which has been fully developed [8]. For the Stokes-Darcy system, the elliptic problem (1.3) of the Darcy flow was widely used in numerical computation, and there exist many studies on the numerical analysis (see, e.g., [4, 9, 10, 11, 12], just to name a few). However, such approach yields a discontinuous numerical velocity $\boldsymbol{u}_h := -K\nabla p_h$, and even worse, the local conservation law breaks down, i.e., $\int_T \nabla\cdot\boldsymbol{u}_h\, dx \neq \int_T f\, dx$.

For Darcy flow computation, elliptic problem (1.3) is utilized if one wants to obtain a continuous, well-approximated pressure. On the other hand, to achieve a discrete velocity field satisfying the local conservation law, the discretization of the mixed form (1.1) was recommended [13]. The mixed form is also helpful for the case with non-smooth and non-convex domains, which may cause singularity. For the Stokes-Darcy coupled flow, we can use various types of elements to treat (1.1) (e.g., RT (Raviart-Thomas), CR (Crouzeix-Raviart), BDM, BDFM, and DG method), the numerical analysis of which was extensively studied [14, 15, 16, 17, 18, 19].

We have to point out that there are quite a few existing works dealing with the curved boundary/interface because implementing the boundary condition (1.1c) (or the interface version $[\boldsymbol{u}\cdot\boldsymbol{n}] = g$ with $[\cdot]$ denoting the jump) on a curved boundary is not trivial. Several approaches have been proposed to treat this issue. The straightforward idea is to approximate Γ by a polygonal surface Γ_h and adjust the matrix and the right-hand side of the linear system from the discrete scheme, such that $\boldsymbol{u}_h\cdot\boldsymbol{n}_h = g$ is directly enforced at some appropriate points of the edges on Γ_h [20]. This approach looks simple; however, it is not popular in practical use because it requires very technical coding skills and may cause ill-posedness. An alternative way is to transform the curved boundary into a straight line or a plane surface. But such transformation could be unavailable or nontrivial to construct. The Lagrange multiplier method [21, 22] is a successful method, which implements $\boldsymbol{u}_h\cdot\boldsymbol{n}_h = g$ in the weak sense. But it introduces an additional new unknown (called the Lagrange multiplier) in a finite element space defined on the mesh of the boundary (also an additional equation as well), which increases the computational cost. The analysis requires a coupled inf-sup condition involving the multiplier, which is nontrivial to obtain.

In the present work, we continue the previous works for the Stokes and Navier-Stokes' cases [23, 24, 25], and study the penalty approach for the Darcy system. The idea is to add the penalty term $\frac{1}{\varepsilon}\int_\Gamma(\boldsymbol{u}_\varepsilon - g)\cdot\boldsymbol{n}\boldsymbol{v}\cdot\boldsymbol{n}\, ds$ to the variational equation with a tiny positive penalty parameter ε. As ε goes to 0, we expect that $\boldsymbol{u}_\varepsilon\cdot\boldsymbol{n} \to g$ to approximate (1.1c). Since the penalty term is just an integration, the penalty method can facilitate the implementation. But the error estimates now depend on the penalty parameter ε and the mesh size h. We present the formulations of the penalty approaches and study the well-posedness and error estimate in the continuous sense. For discretization, we apply the FEM using RT0/P0 element and investigate the convergence rates on both h and ε through several numerical experiments, in smooth/non-smooth convex/non-convex domains.

The rest of this paper is organized as follows. In Section 2, we describe the penalty approach and obtain the error of the penalty. Section 3 is devoted to the penalty finite element scheme. We carry out numerical experiments in Section 4 to investigate the convergence behavior.

Notations. Throughout this paper, for any domain G in $\mathbb{R}^d$ or $\mathbb{R}^{d-1}$, we denote by $\|\cdot\|_{H^k(G)}$ the norm of the Hilbert spaces $H^k(G)$ and $H^k(G)^d$, by $(\cdot,\cdot)_G$ the inner-product of $L^2(G)$ or $L^2(G)^d$, and by $\langle\cdot,\cdot\rangle_G$ the dual-product between $H^{-\frac{1}{2}}(G)$ or $H_{00}^{\frac{1}{2}}(G)$. Note that $H_{00}^{\frac{1}{2}}(G)=H^{\frac{1}{2}}(G)$ when G is a closed surface. We use C and C_i to express a generic constant independent of h and ε.

2. The Weak Formulations and Penalty Method

In this section, we present the weak formulations of the Darcy system and introduce the Lagrange multiplier problem. Then we propose the penalty method to treat the boundary condition and evaluate the penalty error.

2.1. The weak formulation.

Assume that $f\in L^2(\Omega)$ and $g\in H^{-\frac{1}{2}}(\Gamma)$. We introduce the function spaces:

$$V:=H(\mathrm{div};\Omega)=\{\boldsymbol{v}\in L^2(\Omega)^d \mid \nabla\cdot\boldsymbol{v}\in L^2(\Omega)\},$$
$$V_g:=\{\boldsymbol{v}\in V \mid \langle\boldsymbol{v}\cdot\boldsymbol{n}-g,\mu\rangle_\Gamma=0 \ \ \forall\mu\in H^{\frac{1}{2}}(\Gamma)\},$$
$$Q:=L^2(\Omega),\quad \mathring{Q}:=L_0^2(\Omega)=\{q\in Q \mid (1,q)_\Omega=0\}.$$

For any $\boldsymbol{v}\in V$, we have the trace inequality [13]:

$$\|\boldsymbol{v}\cdot\boldsymbol{n}\|_{H^{-\frac{1}{2}}(\Gamma)}\le C\|\boldsymbol{v}\|_V. \tag{2.1}$$

Reversely, for any $g\in H^{-\frac{1}{2}}(\Gamma)$, there exists a $\boldsymbol{v}_g\in V$ such that (see [26]):

$$\boldsymbol{v}_g\cdot\boldsymbol{n}=g \text{ on } \Gamma,\quad \|\boldsymbol{v}_g\|_V:=\sqrt{\|\boldsymbol{v}\|_{L^2(\Omega)}^2+\|\nabla\cdot\boldsymbol{v}\|_{L^2(\Omega)}^2}\le C\|g\|_{H^{-\frac{1}{2}}(\Gamma)}. \tag{2.2}$$

Note that the above $\boldsymbol{v}_g\cdot\boldsymbol{n}|_\Gamma=g$ should be understood in the weak sense, i.e.,

$$\langle\boldsymbol{v}\cdot\boldsymbol{n},\mu\rangle_\Gamma=\langle g,\mu\rangle_\Gamma\quad\forall\mu\in H^{\frac{1}{2}}(\Gamma).$$

There exists a unique solution $\phi\in H^1(\Omega)/\mathbb{R}$ to the Poisson equation:

$$\begin{aligned}\Delta\phi=f_g:=\langle g,1\rangle_\Gamma/|\Omega| &\qquad \text{in } \Omega,\\ \nabla\phi\cdot\boldsymbol{n}=g &\qquad \text{on } \Gamma.\end{aligned}$$

It is easy to check that $\boldsymbol{v}_g:=\nabla\phi$ satisfies (2.2).

Testing (1.1a)(1.1b) by $(\boldsymbol{v},q)\in V_0\times\mathring{Q}$, and using the integration by parts, we obtain the weak formulation of (1.1):

Find $\boldsymbol{u}\in V_0+\{\boldsymbol{v}_g\}$ and $p\in\mathring{Q}$ such that

$$(K^{-1}\boldsymbol{u},\boldsymbol{v})_\Omega-(\nabla\cdot\boldsymbol{v},p)_\Omega=0 \qquad \forall\boldsymbol{v}\in V_0, \tag{2.3a}$$
$$(\nabla\cdot\boldsymbol{u},q)_\Omega-(f,q)=0 \qquad \forall q\in\mathring{Q}. \tag{2.3b}$$

It is apparent that the classical solution of (1.1) satisfies (2.3), and conversely, if the weak solution of (2.3) is smooth enough, then it becomes a classical solution to (1.1). Note that (2.3) is also regarded as the mixed variational form of elliptic problem (1.3).

Setting $\boldsymbol{u}_0 := \boldsymbol{u} - \boldsymbol{v}_g$, we reformulate (2.3) into the following variational problem.

Find $(\boldsymbol{u}_0, p) \in V_0 \times \mathring{Q}$ such that

$$(K^{-1}\boldsymbol{u}_0, \boldsymbol{v})_\Omega - (\nabla\cdot\boldsymbol{v}, p)_\Omega = -(K^{-1}\boldsymbol{v}_g, \boldsymbol{v})_\Omega \qquad \forall \boldsymbol{v} \in V_0, \tag{2.4a}$$

$$(\nabla\cdot\boldsymbol{u}_0, q)_\Omega = (f - \nabla\cdot\boldsymbol{v}_g, q)_\Omega \qquad \forall q \in \mathring{Q}, \tag{2.4b}$$

Since $f - \nabla\cdot\boldsymbol{v}_g \in \mathring{Q}$ (due to (1.2)), there exists a $\boldsymbol{u}_1 \in H_0^1(\Omega)^d$ (or $\boldsymbol{u}_1 \in H^1(\Omega)^d$ with $\boldsymbol{u}_1\cdot\boldsymbol{n}|_\Gamma = 0$) such that (see [26]):

$$\begin{aligned}\nabla\cdot\boldsymbol{u}_1 &= f - \nabla\cdot\boldsymbol{v}_g,\\ \|\boldsymbol{u}_1\|_{H^1(\Omega)} &\le C\|f - \nabla\cdot\boldsymbol{v}_g\|_{L^2(\Omega)}\\ &\le C\big(\|f\|_{L^2(\Omega)} + \|g\|_{H^{-\frac{1}{2}}(\Gamma)}\big) \quad \text{(by (2.2))}.\end{aligned}$$

We set $\boldsymbol{f} := \boldsymbol{v}_g + \boldsymbol{u}_1 \in Q$ and $\boldsymbol{u} := \boldsymbol{u}_0 - \boldsymbol{u}_1 \in V_0$. To avoid the abundance of notations, here we have used the same notation $\boldsymbol{u}$ with a different meaning from those in the above equations.

Now we present the weak form equivalent to (2.4):

(V) Find $(\boldsymbol{u}, p) \in V_0 \times \mathring{Q}$ such that

$$(K^{-1}\boldsymbol{u}, \boldsymbol{v})_\Omega - (\nabla\cdot\boldsymbol{v}, p)_\Omega = -(K^{-1}\boldsymbol{f}, \boldsymbol{v})_\Omega \qquad \forall \boldsymbol{v} \in V_0, \tag{2.5a}$$

$$(\nabla\cdot\boldsymbol{u}, q)_\Omega = 0 \qquad \forall q \in \mathring{Q}, \tag{2.5b}$$

which corresponds to the PDE model

$$-(\boldsymbol{u} + \boldsymbol{f}) = K\nabla p \qquad \text{in } \Omega, \tag{2.6a}$$

$$\nabla\cdot\boldsymbol{u} = 0 \qquad \text{in } \Omega, \tag{2.6b}$$

$$\boldsymbol{u}\cdot\boldsymbol{n} = 0 \qquad \text{on } \Gamma. \tag{2.6c}$$

Instead of (2.3) (resp. (1.1)), here and hereafter, we consider (2.5) (resp. (2.6)) with incompressibility (2.6b) and homogeneous boundary condition (2.6c). According to the above argument, these two problems are equivalently ($\boldsymbol{u}$ of (1.1) equals to $\boldsymbol{u} + \boldsymbol{v}_g + \boldsymbol{u}_1$ of (2.6)).

Moreover, setting

$$\mathring{V}_0 := \{\boldsymbol{v} \in V_0 \mid \nabla\cdot\boldsymbol{v} = 0\},$$

we see that $\boldsymbol{u}$ of (2.5) is also a solution to the following problem.

(V°) Find $\boldsymbol{u} \in \mathring{V}_0$ such that

$$(K^{-1}\boldsymbol{u}, \boldsymbol{v})_\Omega = -(K^{-1}\boldsymbol{f}, \boldsymbol{v})_\Omega \qquad \forall \boldsymbol{v} \in \mathring{V}_0. \tag{2.7}$$

Proposition 2.1. *Let $k_{\max}$ and $k_{\min}$ be the largest and smallest eigenvalue of the symmetric positive definite matrix K. There exists a unique solution $\boldsymbol{u} \in \mathring{V}_0$ of (V°) satisfying*

$$\|\boldsymbol{u}\|_{L^2(\Omega)} \le k_{\max}k_{\min}^{-1}\|\boldsymbol{f}\|_{L^2(\Omega)}. \tag{2.8}$$

And there exists a unique $p \in \mathring{Q}$ such that $(\boldsymbol{u}, p)$ solves (V) with

$$\|p\|_{\mathring{Q}} \le Ck_{\min}^{-1}(\|f\|_{L^2(\Omega)} + \|\boldsymbol{u}\|_{L^2(\Omega)}). \tag{2.9}$$

Proposition 2.1 can be regarded as a well-posedness result of the mixed method for the Poisson equation [13]. For the sake of the convenience of the readers unfamiliar with this topic, let us sketch a brief proof.

Proof. Note that $k_{\max}^{-1}$ and k_{mim}^{-1} are the minimum and maximum eigenvalues of K^{-1}. We have the coercivity and continuity (by $\nabla\cdot\boldsymbol{v}=0$ for all $\boldsymbol{v}\in\mathring{V}_0$):

$$(K^{-1}\boldsymbol{v},\boldsymbol{v})_\Omega \geq k_{\max}^{-1}\|\boldsymbol{v}\|_{L^2(\Omega)}^2 = k_{\max}^{-1}\|\boldsymbol{v}\|_V^2 \quad \forall \boldsymbol{v}\in\mathring{V}_0, \tag{2.10a}$$

$$(K^{-1}\boldsymbol{u},\boldsymbol{v})_\Omega \leq k_{\text{mim}}^{-1}\|\boldsymbol{u}\|_{L^2(\Omega)}\|\boldsymbol{v}\|_{L^2(\Omega)} = k_{\text{mim}}^{-1}\|\boldsymbol{u}\|_V\|\boldsymbol{v}\|_V. \tag{2.10b}$$

By the Lax-Milgram theorem, there exists a unique $\boldsymbol{u}\in\mathring{V}_0$ of (V°). Substituting $\boldsymbol{v}=\boldsymbol{u}$ into (2.7) yields (2.8). The unique existence of p and (2.9) follows from the inf-sup condition: there exists a positive constant C such that

$$\inf_{p\in\mathring{Q}}\sup_{\boldsymbol{v}\in H_0^1(\Omega)^d}\frac{(\nabla\cdot\boldsymbol{v},p)}{\|\boldsymbol{v}\|_V\|p\|_{\mathring{Q}}}\geq C. \tag{2.11}$$

The proof is complete. □

The function space V_0 incorporating the boundary condition $\boldsymbol{v}\cdot\boldsymbol{n}|_\Gamma=0$ is not convenient to deal with in numerical implementation. To this end, we introduce the Lagrange multiplier formulation. For briefness, set the notations:

$$\Lambda := H^{\frac{1}{2}}(\Gamma), \quad \Lambda^* := \text{the dual of }\Lambda.$$

The Lagrange multiplier problem is stated as follows.

(VL) Find $(\boldsymbol{u},p,\lambda)\in V\times\mathring{Q}\times\Lambda$ such that: for all $(\boldsymbol{v},q,\mu)\in V\times\mathring{Q}\times\Lambda$,

$$(K^{-1}\boldsymbol{u},\boldsymbol{v})_\Omega-(\nabla\cdot\boldsymbol{v},p)_\Omega+\langle\boldsymbol{v}\cdot\boldsymbol{n},\lambda\rangle_\Gamma=-(K^{-1}\boldsymbol{f},\boldsymbol{v})_\Omega, \tag{2.12a}$$

$$(\nabla\cdot\boldsymbol{u},q)_\Omega=0, \tag{2.12b}$$

$$\langle\boldsymbol{u}\cdot\boldsymbol{n},\mu\rangle_\Gamma=0. \tag{2.12c}$$

(VL) ⇒ (V) is trivial. And (V) ⇒ (VL) follows from (2.2) and the duality

$$\|\mu\|_{H^{\frac{1}{2}}(\Gamma)}=\sup_{\xi\in H^{-\frac{1}{2}}(\Gamma)}\frac{\langle\xi,\mu\rangle_\Gamma}{\|\xi\|_{H^{-\frac{1}{2}}(\Gamma)}},\quad \|\xi\|_{H^{-\frac{1}{2}}(\Gamma)}=\sup_{\mu\in H^{\frac{1}{2}}(\Gamma)}\frac{\langle\xi,\mu\rangle_\Gamma}{\|\mu\|_{H^{\frac{1}{2}}(\Gamma)}}. \tag{2.13}$$

Remark 2.1. Here we restrict the solution and testing function $\boldsymbol{u},\boldsymbol{v}$ in V (not V_0). The boundary condition (2.6c) is enforced in the weak sense by (2.12c). If p is sufficiently smooth (e.g. $p\in H_0^1(\Omega)$), one can easily verify that $\lambda=p|_\Gamma$.

Using the Lagrange multiplier method in computation avoids the direct implementation of $\boldsymbol{u}\cdot\boldsymbol{n}|_\Gamma=0$. However, it adds the new unknown λ and increases the computational cost. Next, we propose the penalty method to approximate the boundary condition (2.6c), which can facilitate the implementation without enlarging the size of the linear system of the discrete scheme.

2.2. **The penalty method.** We denote by ε $(0 < \varepsilon \ll 1)$ the penalty parameter. The idea is to add the bilinear form $\frac{1}{\varepsilon}\int_\Gamma(\boldsymbol{u}_\varepsilon\cdot\boldsymbol{n})(\boldsymbol{v}\cdot\boldsymbol{n})\,ds$ to the weak formulation such that, $\boldsymbol{u}_\varepsilon\cdot\boldsymbol{n}\to 0$ as $\varepsilon\to 0$. We define the function space $\tilde{V} := \{\boldsymbol{v}\in V \mid \boldsymbol{v}\cdot\boldsymbol{n}\in L^2(\Gamma)\}$ endowed with the norm $\|\boldsymbol{v}\|_{\tilde{V}}^2 := \|\boldsymbol{v}\|_V^2 + \|\boldsymbol{v}\cdot\boldsymbol{n}\|_{L^2(\Gamma)}^2$. The penalty method is stated as follows.

(V$_\varepsilon$) Find $(\boldsymbol{u}_\varepsilon, p_\varepsilon)\in\tilde{V}\times Q$ such that: for all $(\boldsymbol{v},q)\in\tilde{V}\times Q$,

$$(K^{-1}\boldsymbol{u}_\varepsilon,\boldsymbol{v})_\Omega - (\nabla\cdot\boldsymbol{v},p_\varepsilon)_\Omega + \varepsilon^{-1}(\boldsymbol{u}_\varepsilon\cdot\boldsymbol{n},\boldsymbol{v}\cdot\boldsymbol{n})_\Gamma = -(K^{-1}\boldsymbol{f},\boldsymbol{v})_\Omega, \tag{2.14a}$$

$$(\nabla\cdot\boldsymbol{u}_\varepsilon,q)_\Omega = 0. \tag{2.14b}$$

Notice that in general $\boldsymbol{u}_\varepsilon\cdot\boldsymbol{n}\neq 0$, which means that there are inflow and outflow on the boundary, so the pressure constant, saying $\int_\Omega p_\varepsilon\,dx$, is uniquely determined. Thus we search p_ε in Q, not in $\mathring{Q}$. Set $\mathring{\tilde{V}} := \{\boldsymbol{v}\in\tilde{V} \mid \nabla\cdot\boldsymbol{v}=0\}$. We consider the variational problem:

(V$_\varepsilon^\circ$) Find $\boldsymbol{u}_\varepsilon\in\mathring{\tilde{V}}$ such that

$$(K^{-1}\boldsymbol{u}_\varepsilon,\boldsymbol{v})_\Omega + \varepsilon^{-1}(\boldsymbol{u}_\varepsilon\cdot\boldsymbol{n},\boldsymbol{v}\cdot\boldsymbol{n})_\Gamma = -(K^{-1}\boldsymbol{f},\boldsymbol{v})_\Omega \qquad \forall\boldsymbol{v}\in\mathring{\tilde{V}}. \tag{2.15}$$

As with Proposition 2.1, we have the well-posedness for (2.15) and (2.14).

Proposition 2.2. *There exists a unique solution $\boldsymbol{u}_\varepsilon\in\mathring{\tilde{V}}$ of (V$_\varepsilon^\circ$) satisfying*

$$\|\boldsymbol{u}_\varepsilon\|_{L^2(\Omega)} + k_{\max}^{\frac{1}{2}}\varepsilon^{-\frac{1}{2}}\|\boldsymbol{u}_\varepsilon\cdot\boldsymbol{n}\|_{L^2(\Gamma)} \le k_{\max}k_{\min}^{-1}\|\boldsymbol{f}\|_{L^2(\Omega)}. \tag{2.16}$$

And there exists a unique $p_\varepsilon\in Q$ such that $(\boldsymbol{u}_\varepsilon,p_\varepsilon)$ solves (V) with

$$\|p_\varepsilon\|_{\mathring{Q}} \le Ck_{\min}^{-1}(\|\boldsymbol{f}\|_{L^2(\Omega)} + \|\boldsymbol{u}_\varepsilon\|_{L^2(\Omega)}), \tag{2.17}$$

$$\frac{1}{|\Omega|}\left|\int_\Omega p_\varepsilon\,dx\right| \le C(k_{\max}^{\frac{1}{2}}k_{\mathrm{mim}}^{-2} + k_{\min}^{-1})\|\boldsymbol{f}\|_{L^2(\Omega)}. \tag{2.18}$$

Proof. Set the bilinear form

$$a_\varepsilon(\boldsymbol{w},\boldsymbol{v}) := (K^{-1}\boldsymbol{w},\boldsymbol{v})_\Omega + \varepsilon^{-1}(\boldsymbol{w}\cdot\boldsymbol{n},\boldsymbol{v}\cdot\boldsymbol{n})_\Gamma \qquad \forall\boldsymbol{w},\boldsymbol{v}\in\mathring{\tilde{V}}.$$

Since $a_\varepsilon(\cdot,\cdot)$ satisfies the coercivity and continuity, i.e.,

$$\begin{aligned}
a_\varepsilon(\boldsymbol{v},\boldsymbol{v}) &\ge k_{\max}^{-1}\|\boldsymbol{v}\|_{L^2(\Omega)}^2 + \varepsilon^{-1}\|\boldsymbol{v}\cdot\boldsymbol{n}\|_{L^2(\Gamma)}^2\\
&\ge \min(k_{\max}^{-1},\varepsilon^{-1})\|\boldsymbol{v}\|_{\tilde{V}}^2 \quad \forall\boldsymbol{v}\in\mathring{\tilde{V}},\\
a_\varepsilon(\boldsymbol{w},\boldsymbol{v}) &\le k_{\min}^{-1}\|\boldsymbol{w}\|_{L^2(\Omega)}\|\boldsymbol{v}\|_{L^2(\Omega)} + \varepsilon^{-1}\|\boldsymbol{w}\cdot\boldsymbol{n}\|_{L^2(\Gamma)}\|\boldsymbol{v}\cdot\boldsymbol{n}\|_{L^2(\Gamma)}\\
&\le \max(k_{\min}^{-1},\varepsilon^{-1})\|\boldsymbol{w}\|_{\tilde{V}}\|\boldsymbol{v}\|_{\tilde{V}} \quad \forall\boldsymbol{w},\boldsymbol{v}\in\mathring{\tilde{V}},
\end{aligned}$$

by Lax-Milgram's theorem, (V$_\varepsilon^\circ$) admits a unique solution $\boldsymbol{u}_\varepsilon\in\mathring{\tilde{V}}$. Substituting $\boldsymbol{v}=\boldsymbol{u}_\varepsilon$ into (2.15), we calculate as

$$k_{\max}^{-1}\|\boldsymbol{u}_\varepsilon\|_{L^2(\Omega)}^2 + \varepsilon^{-1}\|\boldsymbol{u}_\varepsilon\cdot\boldsymbol{n}\|_{L^2(\Gamma)}^2 \le k_{\mathrm{mim}}^{-1}\|\boldsymbol{f}\|_{L^2(\Omega)}\|\boldsymbol{u}_\varepsilon\|_{L^2(\Omega)},$$

which yields

$$\|\boldsymbol{u}_\varepsilon\|_{L^2(\Omega)} \le k_{\max}k_{\mathrm{mim}}^{-1}\|\boldsymbol{f}\|_{L^2(\Omega)}, \quad \varepsilon^{-\frac{1}{2}}\|\boldsymbol{u}_\varepsilon\cdot\boldsymbol{n}\|_{L^2(\Gamma)} \le k_{\max}^{\frac{1}{2}}k_{\mathrm{mim}}^{-1}\|\boldsymbol{f}\|_{L^2(\Omega)}. \tag{2.19}$$

Here we introduce an inf-sup condition different from (2.11): there exists a positive constant C such that

$$\inf_{p\in Q}\sup_{\boldsymbol{v}\in H^1(\Omega)^d}\frac{(\nabla\cdot\boldsymbol{v},p)}{\|\boldsymbol{v}\|_V\|p\|_Q}\geq C. \tag{2.20}$$

It follows from (2.20) that there exists a unique $p_\varepsilon\in Q$ such that

$$(\nabla\cdot\boldsymbol{v},p_\varepsilon)_\Omega=(K^{-1}\boldsymbol{u}_\varepsilon,\boldsymbol{v})_\Omega+\varepsilon^{-1}(\boldsymbol{u}_\varepsilon\cdot\boldsymbol{n},\boldsymbol{v}\cdot\boldsymbol{n})_\Gamma+(K^{-1}\boldsymbol{f},\boldsymbol{v})_\Omega,\quad\forall\boldsymbol{v}\in H^1(\Omega)^d. \tag{2.21}$$

Since $H^1(\Omega)^d$ is dense in $\tilde{V}$, (2.21) also holds for any $\boldsymbol{v}\in\tilde{V}$. Noting that $\boldsymbol{v}\cdot\boldsymbol{n}|_\Gamma=0$ for all $\boldsymbol{v}\in H_0^1(\Omega)^d$, it follows from (2.11) that

$$\begin{aligned}C\|p_\varepsilon\|_{\mathring{Q}}\leq\sup_{\boldsymbol{v}\in H_0^1(\Omega)^d}\frac{(\nabla\cdot\boldsymbol{v},p_\varepsilon)_\Omega}{\|\boldsymbol{v}\|_{H^1(\Omega)}}&=\sup_{\boldsymbol{v}\in H_0^1(\Omega)^d}\frac{(K^{-1}\boldsymbol{u}_\varepsilon,\boldsymbol{v})_\Omega+(K^{-1}\boldsymbol{f},\boldsymbol{v})_\Omega}{\|\boldsymbol{v}\|_{H^1(\Omega)}}\\&\leq k_{\min}^{-1}(\|\boldsymbol{f}\|_{L^2(\Omega)}+\|\boldsymbol{u}_\varepsilon\|_{L^2(\Omega)}).\end{aligned} \tag{2.22}$$

We make the decomposition

$$p_\varepsilon=\mathring{p}_\varepsilon+k_\varepsilon\quad\text{where}\quad\mathring{p}_\varepsilon\in\mathring{Q},\quad k_\varepsilon:=|\Omega|^{-1}\int_\Omega p_\varepsilon\,dx. \tag{2.23}$$

Note that k_ε is the mean value of p_ε and $\|p_\varepsilon\|_{\mathring{Q}}=\|\mathring{p}_\varepsilon\|_Q$. It remains to bound k_ε. There exists a $\boldsymbol{v}\in H^1(\Omega)^d$ such that $\boldsymbol{v}|_\Gamma=\boldsymbol{n}$ and $\|\boldsymbol{v}\|_{H^1(\Omega)}\leq C\|\boldsymbol{n}\|_{H^{\frac{1}{2}}(\Gamma)}$. Inserting such $\boldsymbol{v}$ into (2.21) and in use of (2.23), we find that:

$$\begin{aligned}k_\varepsilon|\Gamma|&=k_\varepsilon\int_\Gamma\boldsymbol{v}\cdot\boldsymbol{n}\,ds=(\nabla\cdot\boldsymbol{v},k_\varepsilon)_\Omega\quad(\text{by }\boldsymbol{v}\cdot\boldsymbol{n}|_\Gamma=1)\\&=-(\nabla\cdot\boldsymbol{v},\mathring{p}_\varepsilon)_\Omega+(K^{-1}\boldsymbol{u}_\varepsilon,\boldsymbol{v})_\Omega+\varepsilon^{-1}(\boldsymbol{u}_\varepsilon\cdot\boldsymbol{n},1)_\Gamma+(K^{-1}\boldsymbol{f},\boldsymbol{v})_\Omega\\&=-(\nabla\cdot\boldsymbol{v},\mathring{p}_\varepsilon)_\Omega+(K^{-1}\boldsymbol{u}_\varepsilon,\boldsymbol{v})_\Omega+(K^{-1}\boldsymbol{f},\boldsymbol{v})_\Omega,\end{aligned} \tag{2.24}$$

where we have applied $(\boldsymbol{u}_\varepsilon\cdot\boldsymbol{n},1)_\Gamma=(\nabla\cdot\boldsymbol{u}_\varepsilon,1)_\Omega=0$ in the last equation. From (2.24), (2.19), and (2.22), we have

$$\begin{aligned}|k_\varepsilon|&\leq C(\|\mathring{p}_\varepsilon\|_Q+k_{\min}^{-1}\|\boldsymbol{u}_\varepsilon\|_{L^2(\Omega)}+k_{\min}^{-1}\|\boldsymbol{f}\|_{L^2(\Omega)})\|\boldsymbol{v}\|_V\\&\leq C(k_{\max}^{\frac{1}{2}}k_{\mathrm{mim}}^{-2}+k_{\min}^{-1})\|\boldsymbol{f}\|_{L^2(\Omega)}\|\boldsymbol{n}\|_{H^{\frac{1}{2}}(\Gamma)}.\end{aligned} \tag{2.25}$$

Hence we complete the proof. □

Remark 2.2. The *a-priori* estimate (2.16) implies that

$$\varepsilon^{-\frac{1}{2}}\|\boldsymbol{u}_\varepsilon\cdot\boldsymbol{n}\|_{L^2(\Gamma)}\leq\varepsilon^{\frac{1}{2}}k_{\max}^{\frac{1}{2}}k_{\min}^{-1}\|\boldsymbol{f}\|_{L^2(\Omega)}\to0\quad\text{as }\varepsilon\to0,$$

which approximates to the boundary condition $\boldsymbol{u}\cdot\boldsymbol{n}|_\Gamma=0$ of (2.6).

2.3. **The convergence analysis of penalty approach.** We turn to $\|\boldsymbol{u}-\boldsymbol{u}_\varepsilon\|_V$. For briefness, we set the notations:

$$\boldsymbol{e}_{\boldsymbol{u}}:=\boldsymbol{u}-\boldsymbol{u}_\varepsilon,\quad e_p=p-p_\varepsilon,\quad e_\lambda:=\lambda-\lambda_\varepsilon\text{ with }\lambda_\varepsilon:=\varepsilon^{-1}\boldsymbol{u}_\varepsilon\cdot\boldsymbol{n}.$$

Theorem 2.1. *Assume that* $\lambda\in L^2(\Gamma)$ *and* $\boldsymbol{u}\cdot\boldsymbol{n}|_\Gamma\in L^2(\Gamma)$*. Then we have the error estimate*

$$\|\boldsymbol{e}_{\boldsymbol{u}}\|_V+Ck_{\min}\|\mathring{e}_p\|_Q\leq k_{\max}^{\frac{1}{2}}\varepsilon^{\frac{1}{2}}\|\lambda\|_{L^2(\Gamma)}. \tag{2.26}$$

If $\lambda \in H^{\frac{1}{2}}(\Gamma)$, *then*

$$\|K^{-1}\boldsymbol{e_u}\|_{L^2(\Omega)} + \|\mathring{e}_p\|_Q \le C\varepsilon\big(\|\lambda\|_{H^{\frac{1}{2}}(\Gamma)} + (k_{\max}^{\frac{1}{2}}k_{\min}^{-2} + k_{\min}^{-1})\|\boldsymbol{f}\|_{L^2(\Omega)}\big). \tag{2.27}$$

Proof. Since we have assumed $\lambda \in L^2(\Gamma)$ and $\boldsymbol{u} \in \tilde{V} \subset V$ (note that the dual product $\langle\cdot,\cdot\rangle_\Gamma$ of (2.12) now equals to $(\cdot,\cdot)_\Gamma$), subtracting (2.12) from (2.14), we find that $(\boldsymbol{e_u}, e_p, e_\lambda)$ satisfies

$$(K^{-1}\boldsymbol{e_u},\boldsymbol{v})_\Omega - (\nabla\cdot\boldsymbol{v}, e_p)_\Omega + (\boldsymbol{v}\cdot\boldsymbol{n}, e_\lambda)_\Gamma = 0 \qquad \forall \boldsymbol{v} \in \tilde{V}, \tag{2.28a}$$
$$(\nabla\cdot\boldsymbol{e_u}, q)_\Omega = 0 \qquad \forall q \in \mathring{Q}, \tag{2.28b}$$
$$(\boldsymbol{e_u}\cdot\boldsymbol{n}, \mu)_\Gamma + \varepsilon(\lambda_\varepsilon, \mu)_\Gamma = 0 \qquad \forall \mu \in \Lambda. \tag{2.28c}$$

Substituting $\boldsymbol{v} = \boldsymbol{e_u}$ into (2.28a), together with (2.28b) and (2.28c), we deduce

$$\begin{aligned}0 &= \|K^{-1}\boldsymbol{e_u}\|_{L^2(\Omega)}^2 + (e_\lambda, \boldsymbol{e_u}\cdot\boldsymbol{n})_\Gamma \\ &= \|K^{-1}\boldsymbol{e_u}\|_{L^2(\Omega)}^2 + (e_\lambda, -\boldsymbol{u}_\varepsilon\cdot\boldsymbol{n})_\Gamma && (\text{by } \boldsymbol{u}\cdot\boldsymbol{n} = 0) \\ &= \|K^{-1}\boldsymbol{e_u}\|_{L^2(\Omega)}^2 + \varepsilon\|e_\lambda\|_{L^2(\Gamma)}^2 - \varepsilon(e_\lambda, \lambda)_\Gamma && (\text{by } \boldsymbol{u}_\varepsilon\cdot\boldsymbol{n} = \varepsilon\lambda_\varepsilon).\end{aligned}$$

Therefore, we obtain

$$\|K^{-1}\boldsymbol{e_u}\|_{L^2(\Omega)}^2 + \varepsilon\|e_\lambda\|_{L^2(\Gamma)}^2 \le \varepsilon\|e_\lambda\|_{L^2(\Gamma)}\|\lambda\|_{L^2(\Gamma)},$$

which yields

$$\|e_\lambda\|_{L^2(\Gamma)} \le \|\lambda\|_{L^2(\Gamma)}, \quad \|\boldsymbol{e_u}\|_{L^2(\Omega)} \le k_{\max}^{\frac{1}{2}}\varepsilon^{\frac{1}{2}}\|\lambda\|_{L^2(\Gamma)}. \tag{2.29}$$

By (2.23) (the decomposition of p_ε), we set

$$\mathring{e}_p := p - \mathring{p}_\varepsilon \in \mathring{Q}, \quad \hat{e}_\lambda := \lambda - (\lambda_\varepsilon - k_\varepsilon).$$

It follows from (2.28a) that

$$(K^{-1}\boldsymbol{e_u},\boldsymbol{v})_\Omega - (\nabla\cdot\boldsymbol{v}, \mathring{e}_p)_\Omega + (\boldsymbol{v}\cdot\boldsymbol{n}, \hat{e}_\lambda)_\Gamma = 0 \quad \forall \boldsymbol{v} \in \tilde{V}. \tag{2.30}$$

By restricting $\boldsymbol{v} \in H_0^1(\Omega)^d \subset \tilde{V}$, and applying the inf-sup condition (2.11), we have

$$\begin{aligned}C\|\mathring{e}_p\|_Q &\le \sup_{\boldsymbol{v}\in H_0^1(\Omega)^d} \frac{(\nabla\cdot\boldsymbol{v}, \mathring{e}_p)_\Omega}{\|\boldsymbol{v}\|_{H^1(\Omega)}} = \sup_{\boldsymbol{v}\in H_0^1(\Omega)^d} \frac{(K^{-1}\boldsymbol{e_u},\boldsymbol{v})_\Omega}{\|\boldsymbol{v}\|_{H^1(\Omega)}} \\ &\le \|K^{-1}\boldsymbol{e_u}\|_{L^2(\Omega)} \le k_{\min}^{-1}\|\boldsymbol{e_u}\|_{L^2(\Omega)}.\end{aligned} \tag{2.31}$$

A combination of (2.29), (2.31), and $\nabla\cdot\boldsymbol{e_u} = 0$ results (2.26).

Now we proceed to evaluate $(\boldsymbol{e_u}, e_p, e_\lambda)$ under the assumption that $\lambda \in H^{\frac{1}{2}}(\Gamma)$. In this case, we have

$$(K^{-1}\boldsymbol{e_u},\boldsymbol{v})_\Omega - (\nabla\cdot\boldsymbol{v}, \mathring{e}_p)_\Omega + \langle\boldsymbol{v}\cdot\boldsymbol{n}, \hat{e}_\lambda\rangle_\Gamma = 0 \quad \forall \boldsymbol{v} \in V. \tag{2.32}$$

It follows from (2.13) and (2.32) that

$$\begin{aligned}
\|\hat{e}_\lambda\|_{H^{-\frac{1}{2}}(\Gamma)} &= \sup_{\xi\in H^{\frac{1}{2}}(\Gamma)} \frac{\langle \xi, \hat{e}_\lambda\rangle_\Gamma}{\|\xi\|_{H^{\frac{1}{2}}(\Gamma)}} = \sup_{\boldsymbol{v}\in H^1(\Omega)^d} \frac{\langle \boldsymbol{v}\cdot\boldsymbol{n}, \hat{e}_\lambda\rangle_\Gamma}{\|\boldsymbol{v}\cdot\boldsymbol{n}\|_{H^{\frac{1}{2}}(\Gamma)}} \\
&= \sup_{\boldsymbol{v}\in H^1(\Omega)^d} \frac{(\boldsymbol{v}\cdot\boldsymbol{n}, \hat{e}_\lambda)_\Gamma}{\|\boldsymbol{v}\cdot\boldsymbol{n}\|_{H^{\frac{1}{2}}(\Gamma)}} \quad (\text{by } \hat{e}_\lambda \in L^2(\Gamma)), \\
&\le \sup_{\boldsymbol{v}\in V} \frac{-(K^{-1}\boldsymbol{e}_{\boldsymbol{u}}, \boldsymbol{v})_\Omega + (\nabla\cdot\boldsymbol{v}, \mathring{e}_p)_\Omega}{\|\boldsymbol{v}\|_V} \\
&\le C(\|K^{-1}\boldsymbol{e}_{\boldsymbol{u}}\|_{L^2(\Omega)} + \|\mathring{e}_p\|_{L^2(\Omega)}).
\end{aligned} \tag{2.33}$$

Substituting $\boldsymbol{v} = \boldsymbol{e}_{\boldsymbol{u}}$ into (2.32), we obtain

$$\begin{aligned}
0 &= \|K^{-1}\boldsymbol{e}_{\boldsymbol{u}}\|^2_{L^2(\Omega)} + \langle \hat{e}_\lambda, \boldsymbol{e}_{\boldsymbol{u}}\cdot\boldsymbol{n}\rangle_\Gamma \qquad (\text{by } \nabla\cdot\boldsymbol{e}_{\boldsymbol{u}} = 0) \\
&= \|K^{-1}\boldsymbol{e}_{\boldsymbol{u}}\|^2_{L^2(\Omega)} + \varepsilon\|\hat{e}_\lambda\|^2_{L^2(\Gamma)} \\
&\quad - \varepsilon\langle \hat{e}_\lambda, \lambda - k_\varepsilon\rangle_\Gamma \qquad (\text{by } \boldsymbol{u}\cdot\boldsymbol{n} = 0,\ \boldsymbol{u}_\varepsilon\cdot\boldsymbol{n} = \varepsilon\lambda_\varepsilon),
\end{aligned} \tag{2.34}$$

which implies

$$\begin{aligned}
&\|K^{-1}\boldsymbol{e}_{\boldsymbol{u}}\|^2_{L^2(\Omega)} + \varepsilon\|e_\lambda\|^2_{L^2(\Gamma)} \le \|e_\lambda\|_{H^{-\frac{1}{2}}(\Gamma)}\|\lambda - k_\varepsilon\|_{H^{\frac{1}{2}}(\Gamma)} \\
&\le C\varepsilon\|e_\lambda\|_{H^{-\frac{1}{2}}(\Gamma)}(\|\lambda\|_{H^{\frac{1}{2}}(\Gamma)} + |k_\varepsilon|) \\
&\le C\varepsilon(\|K^{-1}\boldsymbol{e}_{\boldsymbol{u}}\|_{L^2(\Omega)} + \|\mathring{e}_p\|_{L^2(\Omega)})\big(\|\lambda\|_{H^{\frac{1}{2}}(\Gamma)} + (k_{\max}^{\frac{1}{2}}k_{\mathrm{mim}}^{-2} + k_{\min}^{-1})\|\boldsymbol{f}\|_{L^2(\Omega)}\big) \\
&\le C\varepsilon\|K^{-1}\boldsymbol{e}_{\boldsymbol{u}}\|_{L^2(\Omega)}\big(\|\lambda\|_{H^{\frac{1}{2}}(\Gamma)} + (k_{\max}^{\frac{1}{2}}k_{\mathrm{mim}}^{-2} + k_{\min}^{-1})\|\boldsymbol{f}\|_{L^2(\Omega)}\big) \quad (\text{by } (2.31)),
\end{aligned}$$

where we have applied (2.18) and (2.33) to derive the third inequality in above. As a result, we have

$$\|K^{-1}\boldsymbol{e}_{\boldsymbol{u}}\|_{L^2(\Omega)} \le C\varepsilon\big(\|\lambda\|_{H^{\frac{1}{2}}(\Gamma)} + (k_{\max}^{\frac{1}{2}}k_{\mathrm{mim}}^{-2} + k_{\min}^{-1})\|\boldsymbol{f}\|_{L^2(\Omega)}\big),$$

which together with (2.29) concludes (2.27). □

3. The Finite Element Method with Penalty

We consider the 2D case and apply the finite element method to discretize the penalty problem (V_ε). The first step is to approximate the smooth domain Ω by a polygonal domain Ω_h and introduce a regular triangulation $\mathscr{T}_h$ to Ω_h, where $h := \max_{T\in\mathscr{T}_h}\operatorname{diam}(T)$ denotes the mesh size (the maximum diameter of the triangles of $\mathscr{T}_h$). We denote by Γ_h the boundary of Ω_h, and by $\boldsymbol{n}_h$ the unit outer normal vector to Γ_h. We adopt the RT0/P0 element (see [13]) and introduce the finite element spaces:

$$\begin{aligned}
V_h &:= \mathscr{R}\mathscr{T}_0(\Omega_h, \mathscr{T}_h) \\
&= \{\boldsymbol{v}_h \in L^2(\Omega_h) \mid \boldsymbol{v}_h|_T = (a,b)^\top + c(x_1,x_2)^\top\ \forall T\in\mathscr{T}_h, a,b,c\in\mathbb{R}\}, \\
V_{h0} &:= \{\boldsymbol{v}_h \in V_h \mid \boldsymbol{v}_h\cdot\boldsymbol{n}_h = 0 \text{ on } \Gamma_h\}, \\
Q_h &:= \{q_h \in L^2(\Omega_h) \mid q_h|_T \in P_0(T)\},
\end{aligned}$$

where $P_0(T)$ represents the set of constant functions on T and $(x_1,x_2)=x$. We have

$$\mathrm{div}V_h := \{\nabla\cdot \boldsymbol{v}_h \mid \boldsymbol{v}_h \in V_h\} = Q_h. \tag{3.1}$$

The discrete penalty problem reads as:

($\mathrm{V}_{\varepsilon,h}$) Find $(\boldsymbol{u}_{\varepsilon,h}, p_{\varepsilon,h}) \in V_h \times Q_h$ such that: for all $(\boldsymbol{v}_h, q_h) \in V_h \times Q_h$,

$$\begin{aligned}(K^{-1}\boldsymbol{u}_{\varepsilon,h},\boldsymbol{v}_h)_{\Omega_h} - (\nabla\cdot\boldsymbol{v}_h, p_{\varepsilon,h})_{\Omega_h} + \varepsilon^{-1}(\boldsymbol{u}_{\varepsilon,h}\cdot\boldsymbol{n}_h, \boldsymbol{v}_h\cdot\boldsymbol{n}_h)_{\Gamma_h}& \\ = -(K^{-1}\boldsymbol{f},\boldsymbol{v}_h)_{\Omega_h},& \end{aligned} \tag{3.2a}$$

$$(\nabla\cdot\boldsymbol{u}_{\varepsilon,h}, q_h)_{\Omega_h} = 0, \tag{3.2b}$$

where we have continuously extended $\boldsymbol{f}$ to $\Omega\cup\Omega_h$ and added a suitable constant such that $(\boldsymbol{f},1)_{\Omega_h}=0$. It follows from (3.1) and (3.2b) that

$$\nabla\cdot\boldsymbol{u}_{\varepsilon,h} = 0 \quad \text{in } \Omega_h, \tag{3.3}$$

which will be confirmed by our numerical examples in Section 4.

Set $\mathring{V}_h := \{\boldsymbol{v}_h \in V_h \mid (\nabla\cdot\boldsymbol{v}_h, q_h)_{\Omega_h} = 0\ \forall q_h \in Q_h\}$. We find that $\boldsymbol{u}_{\varepsilon,h}$ of ($\mathrm{V}_{\varepsilon,h}$) is a solution to the following problem.

($\mathrm{V}^{\circ}_{\varepsilon,h}$) Find $\boldsymbol{u}_{\varepsilon,h} \in \mathring{\tilde{V}}$ such that, for all $\boldsymbol{v}_h \in \mathring{V}_h$,

$$(K^{-1}\boldsymbol{u}_{\varepsilon,h},\boldsymbol{v}_h)_{\Omega_h} + \varepsilon^{-1}(\boldsymbol{u}_{\varepsilon,h}\cdot\boldsymbol{n}_h, \boldsymbol{v}_h\cdot\boldsymbol{n}_h)_{\Gamma_h} = -(K^{-1}\boldsymbol{f},\boldsymbol{v}_h)_{\Omega_h}. \tag{3.4}$$

The well-posedness of (3.4) follows from the standard argument using Lax-Milgram's theorem. The unique existence of $p_{\varepsilon,h}$ can be proved by a discrete inf-sup condition (or equivalently, by (3.1)). The error estimate of $\|\boldsymbol{u}_{\varepsilon,h}-\boldsymbol{u}\|$ is very technical because we have to consider both the finite element approximation error and the domain/boundary perturbation error (i.e., the error caused by $\Omega_h \approx \Omega$). In the present work, we do not establish the error analysis for the finite element approximation, which is left for future work. However, in the next section, we will comprehensively investigate the convergence rate on both h and ε via numerical experiments.

4. NUMERICAL EXPERIMENTS

We carry out numerical experiments to observe the convergence behavior of the penalty finite element method. In particular, we investigate the experimental convergence rate on both h and ε for the cases with the smooth domain (see Example 1), and non-convex non-smooth domain (see Example 4), respectively. Note that the non-convex non-smooth case may cause a singularity in the solution. We will compute the following errors (noting that $\nabla\cdot\boldsymbol{u}=0$):

$$\begin{aligned}&\mathrm{uL2} := \frac{\|\boldsymbol{u}-\boldsymbol{u}_{\varepsilon,h}\|_{L^2(\Omega_h)}}{\|\boldsymbol{u}\|_{L^2(\Omega_h)}}, \quad \mathrm{pL2} := \frac{\|p-p_{\varepsilon,h}\|_{L^2(\Omega_h)}}{\|p\|_{L^2(\Omega_h)}},\\ &\mathrm{uHdiv} := \|\nabla\cdot\boldsymbol{u}_{\varepsilon,h}\|_{L^2(\Omega_h)}.\end{aligned}$$

We will calculate their convergence rate for the L^2-norm errors uL2 and pL2. And we compute uHdiv to check (3.3).

In Examples 2 and 3, we compare the numerical solutions with different values of ε for the cases with convex and non-convex smooth domains, respectively.

Table 1 Example 1: the errors with fixed $\varepsilon = 10^{-7}$

h	uL2	rate	pL2	rate	uHdiv
1.29e-1	7.48e-2	-	7.62e-2	-	2.36e-14
6.16e-2	3.34e-2	1.08	3.73e-2	1.02	4.75e-14
3.38e-2	1.81e-2	0.97	1.88e-2	0.99	9.90e-14
1.81e-2	9.07e-3	0.99	9.33e-3	1.01	2.00e-13
1.12e-2	4.57e-3	0.99	4.76e-3	0.97	4.00e-13

Table 2 Example 1: the errors with fixed $h = 0.0112$

ε	uL2	rate	pL2	rate	uHdiv
0.1	5.35e-2	-	9.38e-2	-	4.02e-13
5e-2	2.81e-2	0.93	4.92e-2	0.93	4.03e-13
2.5e-2	1.49e-2	0.92	2.56e-2	0.95	4.01e-13
1.25e-2	8.48e-3	0.81	1.36e-2	0.91	4.01e-13
6.25e-3	5.81e-3	0.54	7.96e-3	0.77	4.02e-13

4.1. **Example 1: the convergence rate of the case with smooth domain.** We consider (2.6) in the unit disk domain $\Omega = \{(x_1,x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 < 1\}$ with $K = I$ (the indentity matrix) and the exact solution:

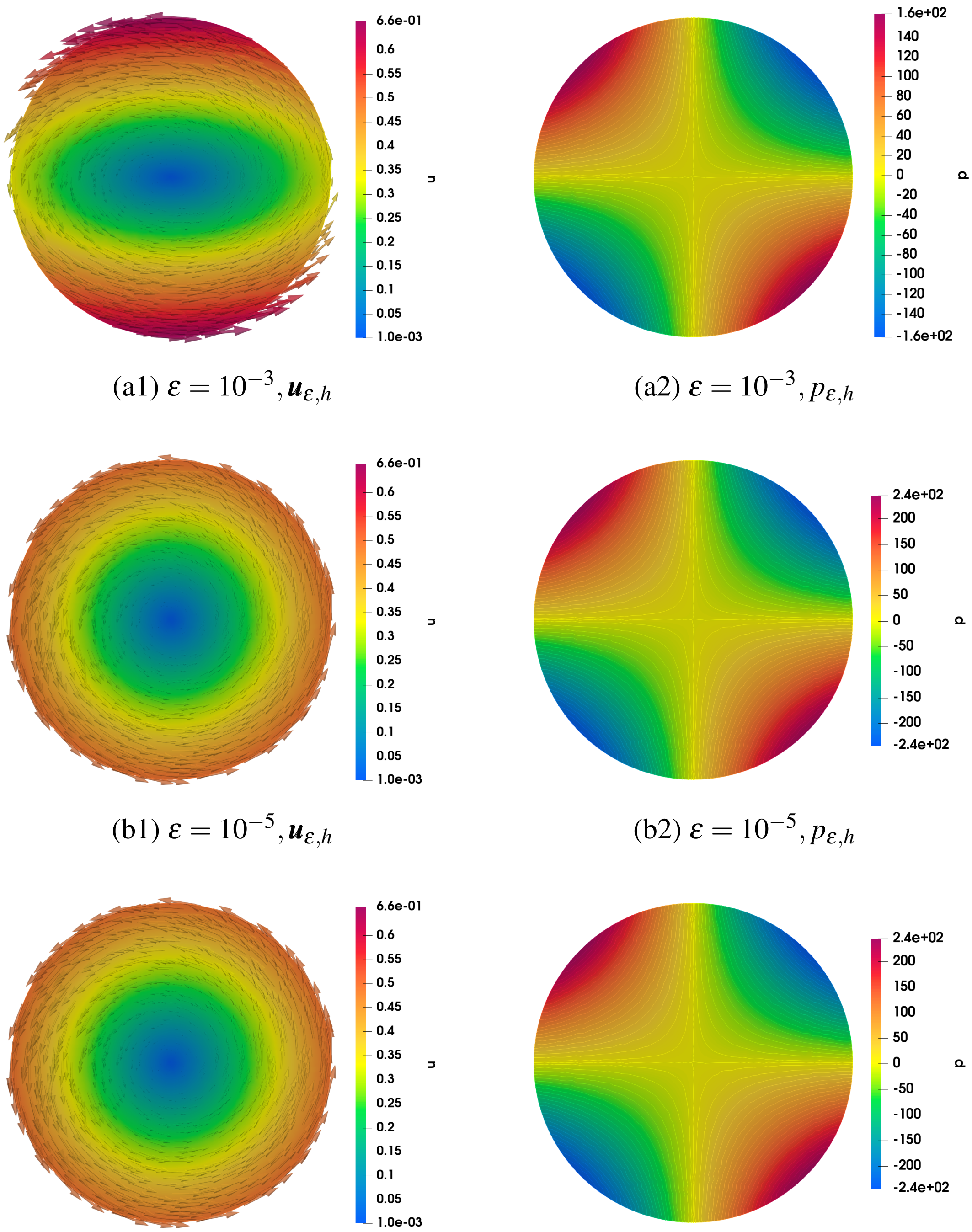
$$\boldsymbol{u} = (10x_2(x_1^2+x_2^2), -10x_1(x_1^2+x_2^2))^\top, \quad p = 10x_1x_2^2.$$

In view of $\boldsymbol{n}|_\Gamma = (x_1,x_2)^\top$, it is easy to check that such $\boldsymbol{u}$ satisfies (2.6b)(2.6a).

To illustrate the experimental convergence rate on h, we fix $\varepsilon = 10^{-7}$ and solve (3.2) for difference mesh size h. The error data are presented in Table 1, where we observe the first-order convergence, i.e., $\|\boldsymbol{u} - \boldsymbol{u}_{\varepsilon,h}\|_{L^2(\Omega)} + \|p - p_{\varepsilon,h}\|_{L^2(\Omega)} \leq Ch$. We see that $\|\nabla \cdot \boldsymbol{u}_{\varepsilon,h}\| \approx 10^{-13}$, which actually confirms (3.3) (note that because of machine error we cannot obtain $\nabla \cdot \boldsymbol{u}_{\varepsilon,h} = 0$ exactly).

Next, we fix the mesh $\mathscr{T}_h$ with $h = 0.0112$ and solve (3.2) for $\varepsilon = 10^{-1}2^{-i}$ $(i = 0,1,\cdots,4)$. The experimental errors and convergence rates are listed in Table 2. It demonstrates the first-order convergence $O(\varepsilon)$ when ε is not too small (here $\varepsilon \geq 0.025$), which supports our theoretical error estimates (2.27) in Theorem 2.1. Note that the errors are saturated for smaller ε and the experimental rate decreases because the errors are also bounded below by Ch. In other words, no matter how tiny ε is chosen, the error always has a lower bound depending on the fixed mesh size h.

4.2. **Example 2: the solutions of different ε for the smooth convex domain.** We take $\Omega = \{(x_1,x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 < 1\}$, $K = 0.001I$, and $\boldsymbol{f} = (x_2, 0)^\top$. We perform the simulation in a fine mesh with $\varepsilon = 10^{-3}, 10^{-5}, 10^{-7}$, and plot the numerical solution $(\boldsymbol{u}_{\varepsilon,h}, p_{\varepsilon,h})$ in Figure 1. For $\varepsilon = 10^{-3}$, the boundary condition $\boldsymbol{u} \cdot \boldsymbol{n} = 0$ has not been well approximated (see Figure 1 (a1)). Although the variational problem $(\mathrm{V}_{\varepsilon,h})$ involves the coefficient ε^{-1}, it seems we can take extreme tiny ε (e.g., $\varepsilon = 10^{-7}$) without causing instability (see Figure 1 (c1)).



(a1) $\varepsilon = 10^{-3}, \boldsymbol{u}_{\varepsilon,h}$ (a2) $\varepsilon = 10^{-3}, p_{\varepsilon,h}$

(b1) $\varepsilon = 10^{-5}, \boldsymbol{u}_{\varepsilon,h}$ (b2) $\varepsilon = 10^{-5}, p_{\varepsilon,h}$

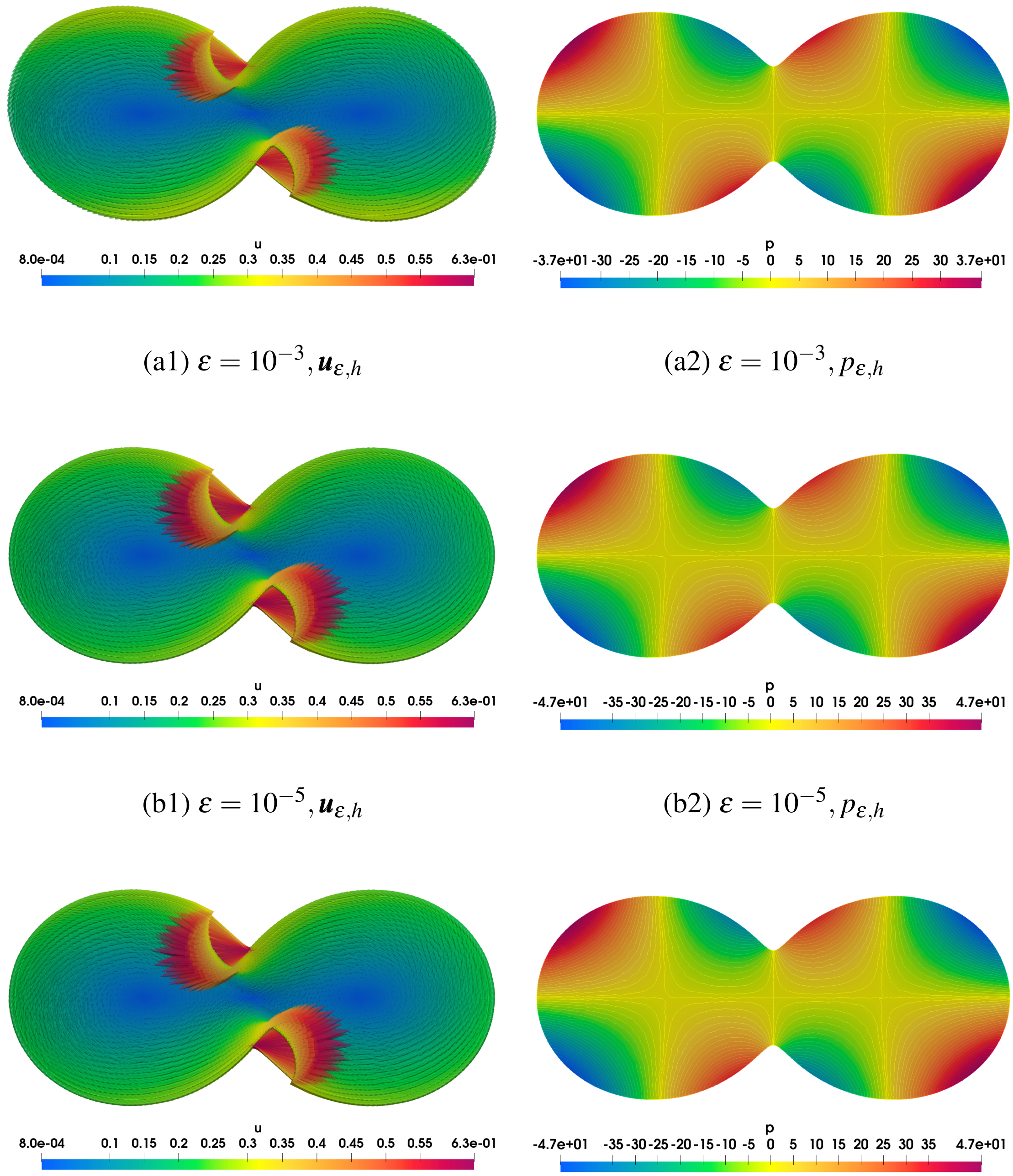
(c1) $\varepsilon = 10^{-7}, \boldsymbol{u}_{\varepsilon,h}$ (c2) $\varepsilon = 10^{-7}, p_{\varepsilon,h}$

Figure 1 Example 2: $(\boldsymbol{u}_{\varepsilon,h}, p_{\varepsilon,h})$ for different ε

4.3. **Example 3: the solutions of different ε for the smooth non-convex domain.** Now we take a non-convex smooth domain Ω surrounded by the boundary

$$\Gamma := \{\big((1-0.8\sin^2(t))\cos(t), 1-0.8\sin^2(t))\cos(t)\big) \mid t \in [0,2\pi)\},$$

and set $K = 0.001I$, $\boldsymbol{f} = (x_2, 0)^\top$. As with the previous example, we apply the penalty finite element method with $\varepsilon = 10^{-3}, 10^{-5}, 10^{-7}$ to solve (2.6). The numerical results are plotted in



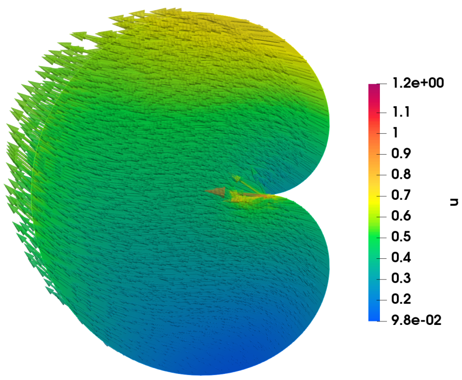
(a1) $\varepsilon = 10^{-3}, \boldsymbol{u}_{\varepsilon,h}$ (a2) $\varepsilon = 10^{-3}, p_{\varepsilon,h}$

(b1) $\varepsilon = 10^{-5}, \boldsymbol{u}_{\varepsilon,h}$ (b2) $\varepsilon = 10^{-5}, p_{\varepsilon,h}$

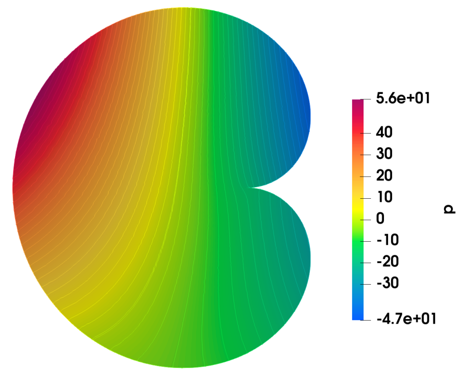
(c1) $\varepsilon = 10^{-7}, \boldsymbol{u}_{\varepsilon,h}$ (c2) $\varepsilon = 10^{-7}, p_{\varepsilon,h}$

Figure 2 Example 3: $(\boldsymbol{u}_{\varepsilon,h}, p_{\varepsilon,h})$ for different ε

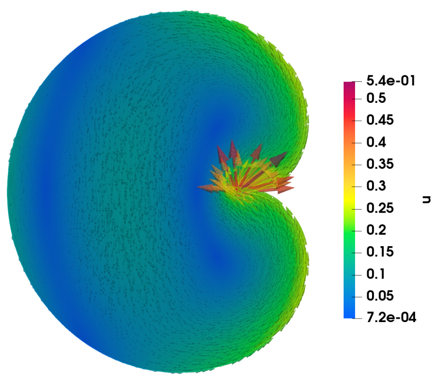
Figure 2. When $\varepsilon = 10^{-3}$, we fail to approximate the boundary condition (see Figure 2 (a1)). For smaller penalty parameters (e.g. $\varepsilon = 10^{-5}, 10^{-7}$), $\boldsymbol{u} \cdot \boldsymbol{n} = 0$ is approximated very well, and the "singular behavior" of the flow near the corners are captured by the simulation (see Figure 2 (b1)(c1)). It reveals the good applicability of the penalty method for the case with the non-convex domain.



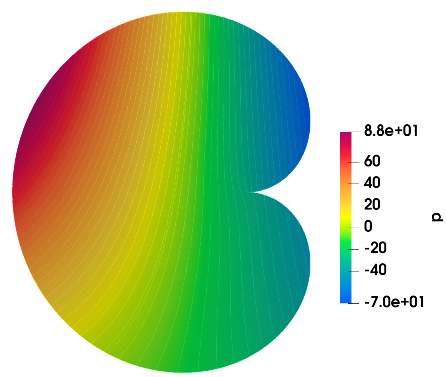
(a1) $\varepsilon = 10^{-3}, \boldsymbol{u}_{\varepsilon,h}$



(a2) $\varepsilon = 10^{-3}, p_{\varepsilon,h}$



(b1) $\varepsilon = 10^{-5}, \boldsymbol{u}_{\varepsilon,h}$



(b2) $\varepsilon = 10^{-5}, p_{\varepsilon,h}$

Figure 3 Example 4: $(\boldsymbol{u}_{\varepsilon,h}, p_{\varepsilon,h})$ for different ε

4.4. **Example 4: the convergence rate of the case with non-smooth non-convex domain.** We turn to the convergence behavior for the case with non-convex non-smooth domain Ω, which is surrounded by the boundary

$$\Gamma := \left\{ \left(\sin(t)\cos(2t), y = \sin(t)\sin(2t) \right) \mid t \in [0, \pi) \right\},$$

and set $K = 0.01I$, $\boldsymbol{f} = (x_2 + 1, -x_1^2)^\top$. It is well-known that the non-convex non-smooth domain may cause the singularity of the elliptic problem [27] (so does the Darcy system as a mixed form of the elliptic problem), which brings difficulty to the numerical approximation. Moreover, because of the loss of regularity, we do not expect the first-order convergence, namely $O(h)$.

Before exploring the convergence rate, we first compute the numerical solutions with $\varepsilon = 10^{-2}, 10^{-5}$. The profiles of the numerical solution are plotted in Figure 3. The case of $\varepsilon = 10^{-2}$ totally fails to recover $\boldsymbol{u} \cdot \boldsymbol{n}|_\Gamma = 0$ (see Figure 3 (a1)), while it is well approximated by taking $\varepsilon = 10^{-5}$ (see Figure 3 (a2)). And we see that singularity occurs at the cusp.

Let us pay attention to the convergence behavior. Since we do not have the exact solution for this example, we take a very fine mesh (with $h = 6.20e-3$) and a very tiny penalty parameter ($\varepsilon = 1e-8$) and compute the numerical solution, which we adopt as the reference solution. In the following, we compare the numerical solution on coarse meshes using a not-so-small penalty parameter with the reference solution and compute the error. First, we fix $\varepsilon = 10^{-8}$, and

Table 3 Example 4: the errors with fixed $\varepsilon = 10^{-8}$

h	uL2	rate	pL2	rate	uHdiv
1.65e-1	1.99e-1	-	1.13e-1	-	3.56e-16
8.41e-2	1.24e-1	0.70	3.33e-2	1.82	7.37e-16
4.50e-2	7.91e-2	0.72	1.73e-2	1.05	1.44e-15
2.31e-2	5.09e-2	0.66	8.59e-3	1.05	2.91e-15

Table 4 Example 1: the errors with fixed $h = 0.0112$

ε	uL2	rate	pL2	rate	uHdiv
$2^{-8}/5$	5.05e-1	0.93	4.82e-2	0.93	1.25e-14
$2^{-9}/5$	2.59e-1	0.96	2.47e-2	0.96	1.17e-14
$2^{-10}/5$	1.31e-1	0.98	1.25e-2	0.98	1.15e-14
$2^{-11}/5$	6.59e-2	0.99	6.29e-2	0.99	1.14e-14

solve (3.2) on different meshes. The experimental errors and convergence rates are presented in Table 3. We notice that the convergence rate of pressure is still first-order, whereas the error of velocity is near $O(h^{0.7})$. The reduced convergence rate may be owing to the singularity of $\boldsymbol{u}$. Next, we fix the mesh size $h = 6.20e-3$ and compute the errors of the numerical solutions with different ε (see Table 4), which shows the first-order convergence, i.e., $O(\varepsilon)$. This example indicates that the singularity does not affect the convergence rate on ε.

REFERENCES

[1] P. Chidyagwai, B. Rivière, Numerical modelling of coupled surface and subsurface flow systems, Adv. Water Resour. 33 (2010), 92-105.

[2] N.S. Hanspal, A.N. Waghode, V. Nassehi, R.J. Wakeman, Numerical analysis of coupled Stokes/Darcy flows in industrial filtrations, Transp. Porous Media 64 (2006), 73-101.

[3] V. Nassehi, Modelling of combined Navier–Stokes and Darcy flows in crossflow membrane filtration, Chem. Eng. Sci. 53 (1998), 1253-1265.

[4] Y. Cao, M. Gunzburger, F. Hua, X. Wang, Coupled Stokes-Darcy model with Beavers–Joseph interface boundary condition, Comput. Math. Sci. 8 (2010), 1-25.

[5] Y. Cao, M. Gunzburger, X. Hu, F. Hua, X. Wang, W. Zhao, Finite element approximations for Stokes-Darcy flow with Beavers-Joseph interface conditions, SIAM J. Numer. Anal. 47 (2010), 4239-4256.

[6] T. Arbogast, D.S. Brunson, A computational method for approximating a Darcy–Stokes system governing a vuggy porous medium, Comput. Geosci. 11 (2007), 207-218.

[7] C. Pozrikidis, D.A. Farrow, A model of fluid flow in solid tumors, Ann. Biomed. Eng. 31 (2003), 181-194.

[8] S.C. Brenner, L.R. Scott, The Mathematical Theory of Finite Element Methods, Springer New York, 2007.

[9] N. Chaabane, V. Girault, C. Poelz, B. Rivière, Convergence of IPDG for coupled time-dependent Navier-Stokes and Darcy equations, J. Comput. Appl. Math. 324 (2017), 25-48.

[10] M. Discacciati, A. Quarteroni, Navier-Stokes/Darcy coupling: modeling, analysis, and numerical approximation, Rev. Mat. Complut. 20 (2009), 315-426.

[11] V. Girault, B. Rivière, DG approximation of coupled Navier-Stokes and Darcy equations by Beaver-Joseph-Saffman interface condition, SIAM J. Numer. Anal. 47 (2009), 2052-2089.

[12] C. Qiu, X. He, J. Li, Y. Lin, A domain decomposition method for the time-dependent Navier-Stokes-Darcy model with Beavers-Joseph interface condition and defective boundary condition, J. Comput. Phys. 411 (2020), 109400.

[13] D. Boffi, F. Brezzi, M. Fortin, Mixed Finite Element Methods and Applications, Springer Series in Computational Mathematics vol. 44, Springer, Heidelberg, 2013.

[14] V.J. Ervin, E.W. Jenkins, S. Sun, Coupled generalized nonlinear Stokes flow with flow through a porus medium, SIAM J. Numer. Anal. 47 (2009), 929-952.

[15] G.N. Gatica, S. Meddahi, R. Oyarzúa, A conforming mixed finite-element method for the coupling of fluid flow with porous media flow, IMA J. Numer. Anal. 29 (2009), 86-108.

[16] G.N. Gatica, S. Meddahi, F.-J. Sayas, Convergence of a family of Galerkin discretizations for the Stokes–Darcy coupled problem, Numer. Methods Partial Differential Equations 27 (2011), 712-748.

[17] W.J. Layton. F. Schieweck, I. Yotov, Coupling fluid flow with porous media flow, SIAM J. Numer. Anal. 40 (2003), 2195-2218.

[18] P. Song, C. Wang, I. Yotov, Domain decomposition for Stokes–Darcy flows with curved interfaces, Procedia Computer Sci. 18 (2013), 1077-1086.

[19] D. Vassilev, I. Yotov, Coupling Stokes-Darcy flow with transport, SIAM J. Sci. Comput. 31 (2009), 3661-3684.

[20] M. Tabata, Finite element approximation to infinite Prandtl number Boussinesq equations with temperature-dependent coefficients-Thermal convection problems in a spherical shell, Future Gener. Comput. Syst. 22 (2006), 521-531.

[21] R. Verfürth, Finite element approximation of incompressible Navier–Stokes equations with slip boundary condition, Numer. Math. 50 (1987), 697-721.

[22] G. Zhou, I. Oikawa, T. Kashiwabara, The Crouzeix-Raviart element for the Stokes equations with the slip boundary condition on a curved boundary, J. Comput. Appl. Math. 383 (2021), 113123.

[23] T. Kashiwabara, I. Oikawa, G. Zhou, Penalty method with P1/P1 finite element approximation for the Stokes equations under the slip boundary condition, Numer. Math. 134 (2016), 705-740.

[24] G. Zhou, T. Kashiwabara, I. Oikawa, Penalty method for the stationary Navier–Stokes problems under the slip boundary condition, J. Sci. Comput. 68 (2016), 339-374.

[25] G. Zhou, T. Kashiwabara, I. Oikawa, A penalty method for the time-dependent Stokes problem with the slip boundary condition and its finite element approximation, Appl. Math. 62 (2017), 377-403.

[26] V. Girault, P.V. Raviart, Finite Element Methods for Navier-Stokes Equations: Theory and Algorithms, Springer, Berlin Heidelberg 1986.

[27] P. Grisvard, Elliptic Problems in Nonsmooth Domains, Pitman, Boston, 1985.