North Carolina Journal of
Mathematics and Statistics
Volume 12, Pages 1–27 (Accepted January 5, 2026, published January 12, 2026)
ISSN 2380-7539

Mixed finite element method for generalized Forchheimer in heterogeneous porous media

AYATI YADAV, TED PARK, TESSICA SELVAGANESAN, THINH KIEU, AND RAY YE

ABSTRACT. We analyzed a mixed finite element discretization of the generalized Darcy-Forchheimer model in a two- or three-dimensional porous domain. We established the existence, uniqueness, and stability of the solutions. Error estimates are presented based on the monotonicity possessed by the Forchheimer term. Numerical investigations were performed to confirm the theoretical accuracy of the discretization.

1. Introduction

The Darcy law is the most common equation to describe fluid flows in porous media

$$-\nabla p = \frac{\mu}{\kappa}\mathbf{v}, \tag{1.1}$$

where p, $\mathbf{v}$, μ, κ are, respectively (resp.), the pressure, velocity, absolute viscosity and permeability.

When the Reynolds number is large, Darcy's law becomes invalid, see Bear (1972); Muskat (1937). A nonlinear relationship between the velocity and gradient of pressure is introduced by adding the higher order terms of velocity to Darcy's law. Forchheimer established this in Forchheimer (1901) the following three nonlinear empirical models:

$$-\nabla p = a\mathbf{v}+b|\mathbf{v}|\mathbf{v}, \quad -\nabla p = a\mathbf{v}+b|\mathbf{v}|\mathbf{v}+c|\mathbf{v}|^2\mathbf{v}, \quad -\nabla p = a\mathbf{v}+d|\mathbf{v}|^{m-1}\mathbf{v}, m\in(1,2). \tag{1.2}$$

Above, the positive constants a, b, c, d are obtained from the experiments for each case.

All three Forchheimer equations in (1.2) can be written in a general form

$$-\nabla p = \sum_{i=0}^{N} a_i|\mathbf{v}|^{\alpha_i}\mathbf{v}. \tag{1.3}$$

This is called the generalized Forchheimer equation. When the media is heterogeneous, the coefficients a_i in equation (1.3) depend on the spatial variable $\mathbf{x}$. For compressible fluids, by the dimensional analysis in Muskat (1937), the equation (1.3) can be modified to become

$$-\nabla p = \sum_{i=0}^{N} a_i\rho^{\alpha_i}|\mathbf{v}|^{\alpha_i}\mathbf{v}, \tag{1.4}$$

Received by the editors August 27, 2025.

2020 *Mathematics Subject Classification.* 35Q35, 35D30, 35K55, 76S05.

Key words and phrases. Porous media, immersible flow, generalized Forchheimer equations, existence, error estimates, slightly compressible fluid, numerical analysis.

where $N \geq 1, \alpha_0 = 0 < \alpha_1 < \ldots < \alpha_N$ are fixed real numbers, the coefficients $a_0(\mathbf{x}), \ldots, a_N(\mathbf{x})$ are non-negative with

$$0 < a_* < a_0(\mathbf{x}), a_N(\mathbf{x}) < a^* < \infty, \quad 0 \leq a_i(\mathbf{x}) \leq a^* < \infty, \, i = 1, \ldots, N-1.$$

In particular, Ward Ward (1964) established from experimental data that

$$-\nabla p = \frac{\mu}{\kappa}\mathbf{v} + c_F \frac{\rho}{\sqrt{\kappa}}|\mathbf{v}|\mathbf{v}, \quad \text{where } c_F > 0. \tag{1.5}$$

The mathematical study of Darcy's law has been studied intensively, with a vast literature, see e.g. Vázquez (2007); Aronson (1986) and references there in. In contrast, the mathematical analysis of Forchheimer flows has received considerably less attention. For incompressible fluids, see Fabrie (1989); Franchi and Straughan (2003); Payne and Straughan (1996, 1999); Payne and Song (2000); Chadam and Qin (1997); Straughan (2008). Regarding compressible fluids, see Aulisa et al. (2009); Hoang and Ibragimov (2012); Hoang et al. (2015, 2014); Celik et al. (2017); Hoang and Kieu (2019); Celik et al. (2018, 2023) for single-phase flows and also Douglas et al. (1993); Park (2005); Kieu (2016) for numerical analysis. In particular, the papers Celik and Hoang (2016, 2017) deal with slightly compressible fluids in heterogeneous porous media.

Multiplying both sides of the equation (1.4) to ρ, we find that

$$\left(\sum_{i=0}^{N} a_i |\rho \mathbf{v}|^{\alpha_i}\right) \rho \mathbf{v} = -\rho \nabla p. \tag{1.6}$$

Denote the function $F : \Omega \times \mathbb{R}^+ \to \mathbb{R}^+$ as a generalized polynomial with non-negative coefficients by

$$F(\mathbf{x}, z) = a_0(\mathbf{x}) z^{\alpha_0} + a_1(\mathbf{x}) z^{\alpha_1} + \cdots + a_N(\mathbf{x}) z^{\alpha_N}, \quad z \geq 0. \tag{1.7}$$

The equation (1.6) can be rewritten as

$$F(\mathbf{x}, |\rho \mathbf{v}|)\rho \mathbf{v} = -\rho \nabla p. \tag{1.8}$$

Under isothermal conditions, the state equation only relates the density ρ with the pressure p, that is $\rho = \rho(p)$. Therefore, the equation of state for slightly compressible fluids is given by

$$\frac{d\rho}{dp} = \frac{\rho}{\bar{\omega}},$$

where $1/\bar{\omega} > 0$ represents the small compressibility.

Hence,

$$\nabla \rho = \frac{1}{\bar{\omega}} \rho \nabla p, \quad \text{or} \quad \rho \nabla p = \bar{\omega} \nabla \rho. \tag{1.9}$$

Combining (1.8) and (1.9) implies that

$$F(\mathbf{x}, |\rho \mathbf{v}|)\rho \mathbf{v} = -\bar{\omega} \nabla \rho.$$

By rescaling coefficients $a_i(\mathbf{x}) \to \bar{\omega}^{-1} a_i(\mathbf{x})$ of $F(.)$, we assume that $\bar{\omega} = 1$. Thus, the above equation is

$$F(\mathbf{x}, |\rho \mathbf{v}|)\rho \mathbf{v} = -\nabla \rho. \tag{1.10}$$

The equation (1.10) is coupled with the conservation law (or continuity equation)

$$\mathrm{div}\,(\rho\mathbf{v}) = f(\mathbf{x}), \tag{1.11}$$

where f is external mass flow rate.

In order to apply standard mixed formulation, it is more convenient to introduce the momentum variable $\mathbf{m} = \rho\mathbf{v}$, and to cast the governing equations in a density-momentum formulation

$$\begin{aligned} F(\mathbf{x}, |\mathbf{m}(\mathbf{x})|)\mathbf{m}(\mathbf{x}) &= -\nabla\rho(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \Omega, \\ \mathrm{div}\,\mathbf{m}(\mathbf{x}) &= f(\mathbf{x}) \qquad\quad \text{for all } \mathbf{x} \in \Omega. \end{aligned} \tag{1.12}$$

The Darcy- Forchheimer equation (1.12) leads to

$$\mathcal{F}(|\mathbf{m}|) = F(\mathbf{x}, |\mathbf{m}|)|\mathbf{m}| = |\nabla\rho|, \qquad \text{where } \mathcal{F}(s) = sF(s).$$

Since $\mathcal{F}$ is a one-to-one mapping from $[0,\infty)$ onto $[0,\infty)$, one can find a unique non-negative $|\mathbf{m}|$ as a function of $|\nabla\rho|$,

$$|\mathbf{m}| = \mathcal{F}^{-1}(|\nabla\rho|).$$

When solving for $\mathbf{m}$ from the first equation in (1.12), it gives

$$\mathbf{m} = \frac{-\nabla\rho}{F(\mathbf{x}, \mathcal{F}^{-1}(|\nabla\rho|))} = -K(\mathbf{x}, |\nabla\rho|)\nabla\rho, \tag{1.13}$$

where the function $K : \Omega \times \mathbb{R}^+ \to \mathbb{R}^+$ is defined for $\xi \geq 0$ by

$$K(\mathbf{x}, \xi) = \frac{1}{F(s(\mathbf{x}, \xi))}, \tag{1.14}$$

with $s = s(\mathbf{x}, \xi)$ being the unique non-negative solution of $sF(s) = \xi$.

Note that

$$\mathcal{F}^{-1}(0) = 0, \quad K(\mathbf{x}, 0) = \frac{1}{F(\mathbf{x}, 0)} = \frac{1}{a_0(\mathbf{x})} > 0.$$

When substituting (1.13) into the second equation of (1.12), we obtain a scalar partial differential equation (PDE) for the density:

$$-\mathrm{div}\,(K(\mathbf{x}, |\nabla\rho|)\nabla\rho) = f(\mathbf{x}), \quad \mathbf{x} \in \Omega. \tag{1.15}$$

This approach was widely exploited in Glowinski and Marroco (1975); Chow (1989); Baranger and Najib (1990); Sandri, D. (1993); Fabrie (1989) along with their numerical analysis.

In the present paper, the inhomogeneous continuity and the Forchheimer-Darcy's momentum equations are treated separately as a coupled system of first order PDE. This gives us the possibility to analyze the nonconstant coefficients. Using nonlinear monotone operator theory (e.g., Brézis (1973); Lions (1969); Showalter (1997); Zeidler (1990)), we can prove the existence and uniqueness of a weak solution of the corresponding elliptic problem of (1.15) for the Dirichlet boundary conditions with general coefficient functions, while imposing only minimal regularity assumptions. Moreover, we establish explicit estimates results which are not obtained in Fabrie (1989). This problem was not studied in the literature previously.

The mixed finite element method (MFEM) is valued for its ability to simultaneously compute scalar (e.g., pressure) and vector (e.g., velocity) functions with comparable accuracy. For second-order elliptic problems, mixed methods for semilinear and nonlinear cases are well-studied in Durán (1988); Milner and Suri (1992); Lee and Milner (1997). In this paper, we combine the techniques from Kieu (2016, 2020) with the mixed finite element framework in Arbogast et al. (1997) to utilize both the special structures of the equation as well as the advantages of the mixed finite element method in obtaining the optimal order error estimates for the solution in several norms of interest.

The paper is organized as follows. We introduce the notations and the relevant results in section 2. Section 3 is devoted to the analysis of the variational formulation. We prove the existence, uniqueness, and stability of weak solution. In section 4, we introduce the discrete problems, recall their main properties, and derive a priori error estimates. We end our paper with some numerical results validating the convergence analysis in section 5.

2. Preliminaries

2.1. Inequalities

The following are some commonly used consequences of Young's inequality.
If $x, y \geq 0$, $\gamma \geq \beta \geq \alpha > 0$, $p, q > 1$ with $1/p + 1/q = 1$, and $\varepsilon > 0$, then

$$x^\alpha \leq 1 + x^\beta, \quad x^\beta \leq x^\alpha + x^\gamma, \quad xy \leq \varepsilon x^p + \varepsilon^{-q/p} y^q. \tag{2.1}$$

For $z \in \mathbb{R}$, denote $z^+ = \max\{0, z\}$. For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $p > 0$, one has

$$(|\mathbf{x}| + |\mathbf{y}|)^p \leq 2^{(p-1)^+}(|\mathbf{x}|^p + |\mathbf{y}|^p) = \begin{cases} |\mathbf{x}|^p + |\mathbf{y}|^p, & \text{for } p \in (0,1], \\ 2^{p-1}(|\mathbf{x}|^p + |\mathbf{y}|^p), & \text{for } p > 1, \end{cases} \tag{2.2}$$

which consequently yields

$$(|\mathbf{x}| + |\mathbf{y}|)^p \leq 2^p(|\mathbf{x}|^p + |\mathbf{y}|^p) \qquad \text{for } p > 0, \tag{2.3}$$

$$||\mathbf{x}|^p - |\mathbf{y}|^p| \leq |\mathbf{x} - \mathbf{y}|^p \qquad \text{for } p \in (0,1]. \tag{2.4}$$

Lemma 2.1. *If $p > 0$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, then*

$$||\mathbf{x}|^p\mathbf{x} - |\mathbf{y}|^p\mathbf{y}| \leq 2^{(p-1)^+}(|\mathbf{x}|^p + |\mathbf{y}|^p)|\mathbf{x} - \mathbf{y}|, \tag{2.5}$$

$$(|\mathbf{x}|^p\mathbf{x} - |\mathbf{y}|^p\mathbf{y}) \cdot (\mathbf{x} - \mathbf{y}) \geq \frac{1}{2}(|\mathbf{x}|^p + |\mathbf{y}|^p)|\mathbf{x} - \mathbf{y}|^2, \tag{2.6}$$

$$(|\mathbf{x}|^p\mathbf{x} - |\mathbf{y}|^p\mathbf{y}) \cdot (\mathbf{x} - \mathbf{y}) \geq \frac{1}{2^{1+(p-1)^+}}|\mathbf{x} - \mathbf{y}|^{p+2}. \tag{2.7}$$

If $p \in (-1, 0)$, then

$$||x|^p x - |y|^p y| \leq 2^{-p}|x - y|^{1+p} \qquad \textit{for all } x, y \in \mathbb{R}, \tag{2.8}$$

$$(|\mathbf{x}|^p\mathbf{x} - |\mathbf{y}|^p\mathbf{y}) \cdot (\mathbf{x} - \mathbf{y}) \geq (1+p)(|\mathbf{x}| + |\mathbf{y}|)^p|\mathbf{x} - \mathbf{y}|^2 \qquad \textit{for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n. \tag{2.9}$$

It is meant, naturally in (2.8) and (2.9), that

$$|\mathbf{x}|^p\mathbf{x}, |\mathbf{y}|^p\mathbf{y}, (|\mathbf{x}|+|\mathbf{y}|)^p|\mathbf{x}-\mathbf{y}|^2 = 0 \text{ for } p\in(-1,0) \text{ and } \mathbf{x}=\mathbf{y}=0.$$

Proof. Proof of inequality (2.5). Consider Scenario 1 and define $h_2(t) = |\gamma(t)|^p\gamma(t)$ for $t\in[0,1]$. Then,

$$\begin{aligned}||\mathbf{x}|^p\mathbf{x}-|\mathbf{y}|^p\mathbf{y}| &= \left|\int_0^1 h_2'(t)dt\right| = \left|\int_0^1 |\gamma(t)|^p(\mathbf{x}-\mathbf{y}) + p|\gamma(t)|^{p-2}(\gamma(t)\cdot(\mathbf{x}-\mathbf{y}))\gamma(t)dt\right|\\ &\le (1+p)|\mathbf{x}-\mathbf{y}|\int_0^1|\gamma(t)|^p dt \le (1+p)|\mathbf{x}-\mathbf{y}|\int_0^1 2^{(p-1)^+}(t^p|\mathbf{x}|^p+(1-t)^p|\mathbf{y}|^p)dt\\ &= 2^{(p-1)^+}(|\mathbf{x}|^p+|\mathbf{y}|^p)|\mathbf{x}-\mathbf{y}|.\end{aligned}$$

This proves (2.5).

In Scenario 2, we can assume $\mathbf{y}=-k\mathbf{x}$ for some $k\ge 0$. We have

$$||\mathbf{x}|^p\mathbf{x}-|\mathbf{y}|^p\mathbf{y}| = |\mathbf{x}|^{p+1}(1+k^{p+1}) \le |\mathbf{x}|^{p+1}(1+k^p)(1+k) = (|\mathbf{x}|^p+|\mathbf{y}|^p)|\mathbf{x}-\mathbf{y}|.$$

Hence, we obtain (2.5).

Proofs of inequalities (2.6) *and* (2.7). We have

$$\begin{aligned}(|\mathbf{x}|^p\mathbf{x}-|\mathbf{y}|^p\mathbf{y})\cdot(\mathbf{x}-\mathbf{y}) &= \left(|\mathbf{x}|^p(\frac{\mathbf{x}+\mathbf{y}}{2}+\frac{\mathbf{x}-\mathbf{y}}{2}) - |\mathbf{y}|^p(\frac{\mathbf{x}+\mathbf{y}}{2}-\frac{\mathbf{x}-\mathbf{y}}{2})\right)\cdot(\mathbf{x}-\mathbf{y})\\ &= (|\mathbf{x}|^p-|\mathbf{y}|^p)\frac{\mathbf{x}+\mathbf{y}}{2}\cdot(\mathbf{x}-\mathbf{y}) + \frac{1}{2}(|\mathbf{x}|^p+|\mathbf{y}|^p)|\mathbf{x}-\mathbf{y}|^2\\ &= \frac{1}{2}(|\mathbf{x}|^p-|\mathbf{y}|^p)(|\mathbf{x}|^2-|\mathbf{y}|^2) + \frac{1}{2}(|\mathbf{x}|^p+|\mathbf{y}|^p)|\mathbf{x}-\mathbf{y}|^2.\end{aligned}$$

Since $(|\mathbf{x}|^p-|\mathbf{y}|^p)(|\mathbf{x}|^2-|\mathbf{y}|^2)\ge 0$, we obtain (2.6). Using $(|\mathbf{x}|+|\mathbf{y}|)^p \ge 2^{-(p-1)^+}|\mathbf{x}-\mathbf{y}|^p$, we then deduce (2.7) from (2.6).

Now, consider $p\in(-1,0)$.

Proof of inequality (2.8). Let $x,y\in\mathbb{R}$. The inequality obviously holds true when $x=0$ or $y=0$. Also, by switching the roles of x and y, we can assume $x>0$ and $y\neq 0$.

If $y>0$, then $||x|^px-|y|^py| = |x^{1+p}-y^{1+p}|$. Noting that $1+p\in(0,1)$, we apply inequality (2.4) to have

$$||x|^px-|y|^py| \le |x-y|^{1+p} \le 2^{-p}|x-y|^{1+p}.$$

If $y<0$, then $||x|^px-|y|^py| = |x|^{1+p}+|y|^{1+p}$. Applying Hölder's inequality to the dot product of two vectors $(|x|^{1+p},|y|^{1+p})$ and $(1,1)$ with powers $1/(1+p)$ and $-1/p$, we obtain

$$||x|^px-|y|^py| \le (|x|+|y|)^{1+p}\cdot 2^{-p} = 2^{-p}|x-y|^{1+p},$$

which yields (2.8).

Proof of inequality (2.9). Let $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$. Consider Scenario 1 and define the function

$$h_3(t) = |\gamma(t)|^p\gamma(t)\cdot(\mathbf{x}-\mathbf{y}) \text{ for } t\in[0,1].$$

Then,

$$(|\mathbf{x}|^p\mathbf{x}-|\mathbf{y}|^p\mathbf{y})\cdot(\mathbf{x}-\mathbf{y})=\int_0^1 h_3'(t)dt=\int_0^1|\gamma(t)|^p|\mathbf{x}-\mathbf{y}|^2+p|\gamma(t)|^{p-2}|\gamma(t)\cdot(\mathbf{x}-\mathbf{y})|^2dt$$
$$\geq(1+p)|\mathbf{x}-\mathbf{y}|^2\int_0^1|\gamma(t)|^p dt.$$

Note that $-p\in(0,1)$, hence $|\gamma(t)|^{-p}\leq(|\mathbf{x}|+|\mathbf{y}|)^{-p}$. Therefore, we obtain (2.9).

In Scenario 2, we can assume $\mathbf{y}=-k\mathbf{x}$ for some $k\geq 0$. We have

$$(|\mathbf{x}|^p\mathbf{x}-|\mathbf{y}|^p\mathbf{y})\cdot(\mathbf{x}-\mathbf{y})=|\mathbf{x}|^{2+p}(1+k^{1+p})(1+k).$$

Since $0<1+p<1$, from (2.3), we have that $1+k^{1+p}\geq(1+k)^{1+p}$. Hence,

$$(|\mathbf{x}|^p\mathbf{x}-|\mathbf{y}|^p\mathbf{y})\cdot(\mathbf{x}-\mathbf{y})\geq|\mathbf{x}|^{2+p}(1+k)^{2+p}=|\mathbf{x}-\mathbf{y}|^2(|\mathbf{x}|+|\mathbf{y}|)^p,$$

which proves (2.9) again. □

Lemma 2.2. *The following inequalities hold for all* $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$.

$$|F(|\mathbf{x}|)\mathbf{x}-F(|\mathbf{y}|)\mathbf{y}|\leq c_1\left(1+|\mathbf{x}|^{\alpha_N}+|\mathbf{y}|^{\alpha_N}\right)|\mathbf{x}-\mathbf{y}|. \tag{2.10}$$

$$(F(|\mathbf{x}|)\mathbf{x}-F(|\mathbf{y}|)\mathbf{y})\cdot(\mathbf{x}-\mathbf{y})\geq c_2\left(|\mathbf{x}-\mathbf{y}|^2+|\mathbf{x}-\mathbf{y}|^{\alpha_N+2}\right), \tag{2.11}$$

where the constants $c_1=a^*2^{1+(\alpha_N-1)^+}(N+1)>0$*, and* $c_2=a_*2^{-1-(\alpha_N-1)^+}>0$.

Proof. Proof of inequality (2.10). We have

$$|F(|\mathbf{x}|)\mathbf{x}-F(|\mathbf{y}|)\mathbf{y}|=|a_0(\mathbf{x}-\mathbf{y})+a_1(|\mathbf{x}|^{\alpha_1}\mathbf{x}-|\mathbf{y}|^{\alpha_1}\mathbf{y})+\cdots+a_N(|\mathbf{x}|^{\alpha_N}\mathbf{x}-|\mathbf{y}|^{\alpha_N}\mathbf{y})|$$
$$\leq a^*\left(|\mathbf{x}-\mathbf{y}|+||\mathbf{x}|^{\alpha_1}\mathbf{x}-|\mathbf{y}|^{\alpha_1}\mathbf{y}|+\cdots+||\mathbf{x}|^{\alpha_N}\mathbf{x}-|\mathbf{y}|^{\alpha_N}\mathbf{y}|\right).$$

Applying (2.5) to the terms $||\mathbf{x}|^{\alpha_i}\mathbf{x}-|\mathbf{y}|^{\alpha_i}\mathbf{y}|$, we find that

$$|F(|\mathbf{x}|)\mathbf{x}-F(|\mathbf{y}|)\mathbf{y}|\leq a^*\Big(1+2^{(\alpha_1-1)^+}(|\mathbf{x}|^{\alpha_1}+|\mathbf{y}|^{\alpha_1})+\cdots+2^{(\alpha_N-1)^+}(|\mathbf{x}|^{\alpha_N}+|\mathbf{y}|^{\alpha_N})\Big)|\mathbf{x}-\mathbf{y}|.$$

By Young's inequality (2.1), $|\mathbf{w}|^{\alpha_i}\leq 1+|\mathbf{w}|^{\alpha_N}$, we find that

$$|F(|\mathbf{x}|)\mathbf{x}-F(|\mathbf{y}|)\mathbf{y}|\leq a^*2^{1+(\alpha_N-1)^+}(N+1)\left(1+|\mathbf{x}|^{\alpha_N}+|\mathbf{y}|^{\alpha_N}\right)|\mathbf{x}-\mathbf{y}|,$$

which proves (2.10).

Proof of the inequality (2.11). We have

$$(F(|\mathbf{x}|)\mathbf{x}-F(|\mathbf{y}|)\mathbf{y})\cdot(\mathbf{x}-\mathbf{y})=a_0|\mathbf{x}-\mathbf{y}|^2+\ldots+a_N(|\mathbf{x}|^{\alpha_N}\mathbf{x}-|\mathbf{y}|^{\alpha_N}\mathbf{y})\cdot(\mathbf{x}-\mathbf{y})$$
$$\geq a_*\left(|\mathbf{x}-\mathbf{y}|^2+\ldots+(|\mathbf{x}|^{\alpha_N}\mathbf{x}-|\mathbf{y}|^{\alpha_N}\mathbf{y})\cdot(\mathbf{x}-\mathbf{y})\right).$$

Applying (2.7) to the terms $(|\mathbf{x}|^{\alpha_i}\mathbf{x}-|\mathbf{y}|^{\alpha_i}\mathbf{y})\cdot(\mathbf{x}-\mathbf{y})$, we obtain

$$(F(|\mathbf{x}|)\mathbf{x}-F(|\mathbf{y}|)\mathbf{y})\cdot(\mathbf{x}-\mathbf{y})\geq a_*\left(|\mathbf{x}-\mathbf{y}|^2+\ldots+\frac{1}{2^{1+(\alpha_N-1)^+}}|\mathbf{x}-\mathbf{y}|^{\alpha_N+2}\right)$$
$$\geq\frac{a_*}{2^{1+(\alpha_N-1)^+}}\left(|\mathbf{x}-\mathbf{y}|^2+|\mathbf{x}-\mathbf{y}|^{\alpha_N+2}\right).$$

Therefore, we obtain (2.11). □

2.2. Functional spaces

Next, we review the Sobolev spaces and trace theorems. Hereafter, the spatial dimension $n \geq 2$ is fixed. Let Ω be an open, bounded subset of $\mathbb{R}^n$ with the boundary $\partial\Omega$ of class C^1.

For $1 \leq s < \infty$, let $L^s(\Omega)$ be the standard Lebesgue space of scalar functions and denote $\mathbf{L}^s(\Omega) = (L^s(\Omega))^n$. The notation $\|\cdot\|_{0,s}$ is used to denote both norms $\|\cdot\|_{L^s(\Omega)}$ and $\|\cdot\|_{\mathbf{L}^s(\Omega)}$.

For a nonnegative integer m, let $W^{m,p}(\Omega)$ be the standard Sobolev space with the norm

$$\|u\|_{m,s} = \Big(\sum_{|\alpha|\leq m} \|D^\alpha u\|^s_{L^s(\Omega)} \Big)^{\frac{1}{s}}.$$

For any normed space X, its dual space is denoted by X', and the product between X' and X is denoted by $\langle \cdot,\cdot\rangle_{X',X}$, i.e., $\langle y, x\rangle_{X',X} = y(x)$ for $y \in X'$ and $x \in X$.

Consider $1 < s < \infty$ now. The function $\gamma_0 : f \in C^\infty(\overline{\Omega}) \mapsto f\big|_{\partial\Omega}$ can be extended to a bounded linear mapping $\gamma_{0,s} : W^{1,s}(\Omega) \to L^s(\partial\Omega)$. The function $\gamma_{0,s}(f)$ is called the trace of f on $\partial\Omega$.

Define $X_s = W^{1-1/s,s}(\partial\Omega)$ to be the range of $\gamma_{0,s}$ equipped with the norm

$$\|f\|_{X_s} = \inf\{\|\varphi\|_{1,s} : \varphi \in W^{1,s}(\Omega), \gamma_{0,s}(\varphi) = f\}.$$

Define the space

$$\mathbf{W}_s(\mathrm{div},\Omega) = \{\mathbf{v} \in \mathbf{L}^s(\Omega) : \mathrm{div}\,\mathbf{v} \in \mathrm{L}^s(\Omega)\} \tag{2.12}$$

equipped with the norm

$$\|\mathbf{v}\|_{\mathbf{W}_s(\mathrm{div},\Omega)} = \Big(\|\mathbf{v}\|^s_{0,s} + \|\mathrm{div}\,\mathbf{v}\|^s_{0,s}\Big)^{1/s}. \tag{2.13}$$

Then, $\mathbf{W}_s(\mathrm{div},\Omega)$ is a reflexive Banach space, see Lemma A.1 below.

Let $r > 1$ be the Hölder conjugate of s, i.e., $1/s + 1/r = 1$. Thanks to (2.2),

$$\|\mathbf{v}\|_{\mathbf{W}_s(\mathrm{div},\Omega)} \leq \|\mathbf{v}\|_{0,s} + \|\mathrm{div}\,\mathbf{v}\|_{0,s} \leq 2^{1/r}\|\mathbf{v}\|_{\mathbf{W}_s(\mathrm{div},\Omega)}. \tag{2.14}$$

Let $\vec{\nu}$ denote the outward normal vector to the boundary $\partial\Omega$. Then one can extend the normal trace $\gamma_{\mathrm{n}}(\mathbf{v}) = \mathbf{v}\cdot\vec{\nu}$ for $\mathbf{v} \in (C^\infty(\overline{\Omega}))^n$ to a bounded, linear mapping $\gamma_{\mathrm{n},s}$ from $\mathbf{W}_s(\mathrm{div},\Omega)$ into X_r'. In particular, there is $\bar{c}_1 > 0$ such that

$$\|\gamma_{\mathrm{n},s}(\mathbf{v})\|_{X_r'} \leq \bar{c}_1 \|\mathbf{v}\|_{\mathbf{W}_s(\mathrm{div},\Omega)} \text{ for all } \mathbf{v} \in \mathbf{W}_s(\mathrm{div},\Omega), \tag{2.15}$$

and Green's formula

$$\int_\Omega \mathbf{v}\cdot\nabla q\,d\mathbf{x} + \int_\Omega \mathrm{div}\,\mathbf{v}\,q\,d\mathbf{x} = \int_{\partial\Omega} (\mathbf{v}\cdot\vec{\nu})q\,d\sigma \tag{2.16}$$

holds for every $\mathbf{v} \in \mathbf{W}_s(\mathrm{div},\Omega)$ and $q \in W^{1,r}(\Omega)$.

Finally, we recall an important norm estimate, see (Baranger and Najib, 1990, Inequality (4.2)) or (Sandri, D., 1993, Lemma A.1) or (Knabner and Summ, 2016, Lemma A.3).

Lemma 2.3. *Let $r, s \in (1, \infty)$ be Hölder conjugates of each other and let $V = \mathbf{W}_s(\mathrm{div}, \Omega)$. Then there exists a constant $C_* > 0$ such that, for all $q \in L^r(\Omega)$, it holds*

$$\|q\|_{0,r} \leq C_* \sup_{\mathbf{v}\in V\setminus\{0\}} \frac{\int_\Omega \mathrm{div}\,\mathbf{v}\, q\, d\mathbf{x}}{\|\mathbf{v}\|_V}. \tag{2.17}$$

Proof. Let $q \in L^r(\Omega)$. If $q = 0$, then (2.17) holds true. Consider $q \neq 0$. Denote by $W_0^{1,r}(\Omega)$ the space of functions in $W_0^{1,r}(\Omega)$ having zero trace on the boundary. Note that $|q|^{r-2}q \in L^s(\Omega)$. By the Browder–Minty Theorem, there exists a unique solution $w \in W_0^{1,s}(\Omega)$ of the problem

$$\int_\Omega |\nabla w|^{r-2}\nabla w \cdot \nabla v\, d\mathbf{x} = \int_\Omega |q|^{r-2} q v\, d\mathbf{x} \quad \text{for all } v \in W_0^{1,r}(\Omega). \tag{2.18}$$

Choosing $v = w$ in (2.18) and applying the Hölder and Poincaré inequalities give

$$\|\nabla w\|_{0,r}^r = \int_\Omega |\nabla w|^r\, d\mathbf{x} = \int_\Omega |q|^{r-2} q w\, d\mathbf{x} \leq \|q\|_{0,r}^{r-1} \|w\|_{0,r} \leq C \|q\|_{0,r}^{r-1} \|\nabla w\|_{0,r},$$

where C is a positive constant. Hereafter, C denotes a generic positive constant. It follows that

$$\|\nabla w\|_{0,r} \leq C \|q\|_{0,r}.$$

Set $\mathbf{u} = -|\nabla w|^{r-2}\nabla w$. Then $\mathbf{u} \in \mathbf{L}^s(\Omega)$ and, by (2.18), $\mathrm{div}\,\mathbf{u} = |q|^{r-2}q \in L^s(\Omega)$. Therefore, $\mathbf{u} \in V \setminus \{0\}$. Observe that

$$\|\mathbf{u}\|_V^s = \|\mathbf{u}\|_{0,s}^s + \|\mathrm{div}\,\mathbf{u}\|_{0,s}^s = \|\nabla w\|_{0,r}^r + \|q\|_{0,r}^r \leq C \|q\|_{0,r}^r.$$

Thus, $\|\mathbf{u}\|_V \leq C \|q\|_{0,r}^{r-1}$. We then have

$$\int_\Omega (\mathrm{div}\,\mathbf{u}) q\, d\mathbf{x} = \int_\Omega |q|^r\, d\mathbf{x} = \|q\|_{0,r}^r = \|q\|_{0,r} \|q\|_{0,r}^{r-1} \geq C \|q\|_{0,r} \|\mathbf{u}\|_V.$$

Consequently, we obtain inequality (2.17). □

Our calculations frequently use the following exponents

$$s = \alpha_N + 2 \in (2, \infty), \quad r = \frac{s}{s-1} \in (1, 2). \tag{2.19}$$

The arguments $C, C_1, \ldots$ denote generic positive constants whose values may vary from place to place. These constants depend on the exponents, the coefficients of the polynomial F, the spatial dimension n, and the domain Ω, but are independent of the boundary data and the spatial discretization step.

3. The mixed formulation

We consider the problem governed by the Darcy-Forchheimer equation and the continuity equation together with Dirichlet boundary condition

$$\begin{cases} F(|\mathbf{m}(\mathbf{x})|)\mathbf{m}(\mathbf{x}) = -\nabla\rho(\mathbf{x}) & \text{for all } \mathbf{x} \in \Omega, \\ \mathrm{div}\,\mathbf{m}(\mathbf{x}) = f(\mathbf{x}) & \text{for all } \mathbf{x} \in \Omega, \\ \rho(\mathbf{x}) = \psi(\mathbf{x}) & \text{for all } \mathbf{x} \in \partial\Omega. \end{cases} \tag{3.1}$$

The mixed formulation of (3.1) reads as follows: Find $(\mathbf{m}, \rho) \in \mathbf{W}_s(\mathrm{div}, \Omega) \times \mathrm{L^r}(\Omega) \equiv \mathrm{V} \times \mathrm{Q}$ such that

$$\begin{cases} \displaystyle\int_\Omega F(|\mathbf{m}|)\mathbf{m} \cdot \mathbf{v}\, d\mathbf{x} - \int_\Omega \rho \,\mathrm{div}\, \mathbf{v}\, d\mathbf{x} = -\int_{\partial\Omega} \psi\,(\mathbf{v} \cdot \vec{\nu})\, d\sigma & \text{for all } \mathbf{v} \in V, \\ \displaystyle\int_\Omega (\mathrm{div}\, \mathbf{m}) q\, d\mathbf{x} = \int_\Omega f\, q\, d\mathbf{x} & \text{for all } q \in Q. \end{cases} \tag{3.2}$$

Forcing function

For the fifth integral in (3.2), we assume $f \in L^s(\Omega)$ and define $\Lambda_f : Q \to \mathbb{R}$ by

$$\Lambda_f(q) = \int_\Omega f q \, d\mathbf{x} \quad \text{for all } q \in Q.$$

Then $\Lambda_f \in Q'$ and

$$\|\Lambda_f\|_{Q'} = \|f\|_{0,s} \,. \tag{3.3}$$

Boundary data

For the third integral in (3.2), we assume $\psi \in X_r$ and define $\Upsilon_\psi : V \to \mathbb{R}$ by

$$\Upsilon_\psi(\mathbf{v}) = \int_{\partial\Omega} \psi(\mathbf{v} \cdot \vec{\nu})\, d\sigma \quad \text{for all } \mathbf{v} \in V.$$

Thanks to the Green's formula (2.16), $\Upsilon_\psi(\mathbf{v})$ is the rigorous formulation for the boundary integral in (3.2). By (2.15), one has $\Upsilon_\psi \in V'$ and

$$\|\Upsilon_\psi\|_{V'} \le \bar{c}_1 \|\psi\|_{X_r} \,. \tag{3.4}$$

For the second and fourth integrals in (3.2), we define a bilinear form $b : V \times Q \to \mathbb{R}$ by

$$b(\mathbf{v}, q) = \int_\Omega (\mathrm{div}\, \mathbf{v}) q \, d\mathbf{x} \quad \text{for all } \mathbf{v} \in V, q \in Q.$$

Then for any $\mathbf{v} \in V$ and $q \in Q$, applying Hölder's inequality gives

$$|b(\mathbf{v}, q)| \le \|\mathrm{div}\, \mathbf{v}\|_{0,s} \|p\|_{0,r} \,. \tag{3.5}$$

For the first integral in (3.2), we define $a : \mathbf{L}^s(\Omega) \times \mathbf{L}^s(\Omega) \to \mathbb{R}$ by

$$a(\mathbf{u}, \mathbf{v}) = \int_\Omega F(|\mathbf{u}(\mathbf{x})|)\mathbf{u}(\mathbf{x}) \cdot \mathbf{v}(\mathbf{x})\, d\mathbf{x} \quad \text{for all } \mathbf{u}, \mathbf{v} \in \mathbf{L}^s(\Omega).$$

The following are the basic properties of $a(\cdot, \cdot)$.

Lemma 3.1. *For any* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbf{L}^s(\Omega)$*, one has*

$$|a(\mathbf{u}, \mathbf{w}) - a(\mathbf{v}, \mathbf{w})| \le c_4 \left(1 + \|\mathbf{u}\|_{0,s}^{s-2} + \|\mathbf{v}\|_{0,s}^{s-2}\right) \|\mathbf{u} - \mathbf{v}\|_{0,s} \|\mathbf{w}\|_{0,s} \,, \tag{3.6}$$

$$a(\mathbf{u}, \mathbf{u} - \mathbf{v}) - a(\mathbf{v}, \mathbf{u} - \mathbf{v}) \ge c_2 \|\mathbf{u} - \mathbf{v}\|_{0,s}^s \,, \tag{3.7}$$

where c_2*,* $c_4 = c_1(1 + |\Omega|^{1/s})$ *are positive constants.*

Consequently,

$$|a(\mathbf{u},\mathbf{w})| \le c_4 \left(\|\mathbf{u}\|_{0,s} + \|\mathbf{u}\|_{0,s}^{s-1}\right) \|\mathbf{w}\|_{0,s}, \tag{3.8}$$

$$|a(\mathbf{u},\mathbf{u})| \ge c_2 \|\mathbf{u}\|_{0,s}^{s}. \tag{3.9}$$

Proof. Using property (2.10), we have

$$\begin{aligned} |a(\mathbf{u},\mathbf{w}) - a(\mathbf{v},\mathbf{w})| &\le \int_\Omega |F(|\mathbf{u}|)\,\mathbf{u} - F(|\mathbf{v}|)\,\mathbf{v}|\,|\mathbf{w}|\,d\mathbf{x} \\ &\le c_1 \int_\Omega (1 + |\mathbf{u}|^{s-2} + |\mathbf{v}|^{s-2})\,|\mathbf{u}-\mathbf{v}|\,|\mathbf{w}|\,d\mathbf{x}. \end{aligned}$$

Applying Hölder's inequality for three powers $s/(s-2), s, s$ gives

$$\int_\Omega (1 + |\mathbf{u}|^{s-2} + |\mathbf{v}|^{s-2})\,|\mathbf{u}-\mathbf{v}|\,|\mathbf{w}|\,d\mathbf{x} \le \left(\|1\|_{0,s} + \|\mathbf{u}\|_{0,s}^{s-2} + \|\mathbf{v}\|_{0,s}^{s-2}\right) \|\mathbf{u}-\mathbf{v}\|_{0,s}\,\|\mathbf{w}\|_{0,s}\,.$$

Thus, we obtain (3.6).

Finally, by (2.11), we have

$$a(\mathbf{u},\mathbf{u}-\mathbf{v}) - a(\mathbf{v},\mathbf{u}-\mathbf{v}) = \int_\Omega (F(|\mathbf{u}|)\mathbf{u} - F(|\mathbf{v}|)\mathbf{v}) \cdot (\mathbf{u}-\mathbf{v})\,d\mathbf{x} \ge c_2 \int_\Omega |\mathbf{u}-\mathbf{v}|^s\,d\mathbf{x},$$

which proves (3.7).

Taking $\mathbf{v} = 0$ in (3.6) and (3.7), we obtain (3.8) and (3.9). □

Definition 3.2. *Given $f \in L^s(\Omega)$ and $\psi \in X_r$, a weak solution of Problem* (3.2) *is a pair* $(\mathbf{m},\rho) \in V \times Q$ *that satisfies*

$$\begin{cases} a(\mathbf{m},\mathbf{v}) - b(\mathbf{v},\rho) = -\Upsilon_\psi(\mathbf{v}) & \text{for all } \mathbf{v} \in V, \\ \qquad\qquad\; b(\mathbf{m},q) = \Lambda_f(q) & \text{for all } q \in Q. \end{cases} \tag{3.10}$$

We will establish the existence and uniqueness of a weak solution of the problem (3.2).

Theorem 3.3. *The following statements hold true.*

(1) For any $f \in L^s(\Omega)$ and $\psi \in X_r$, there exists a unique weak solution $(\mathbf{m},\rho) \in V \times Q$ of Problem (3.2)*.*

(2) There is $c_3 > 0$ such that if $f, \psi, (\mathbf{m},\rho)$ are as in part 1, then

$$\|\mathbf{m}\|_{0,s} + \|\operatorname{div}\mathbf{m}\|_{0,s} + \|\rho\|_{0,r} \le c_3\big(\|f\|_{0,s}^{r-1} + \|f\|_{0,s}^{s-1} + \|\psi\|_{X_r}^{r-1} + \|\psi\|_{X_r}\big). \tag{3.11}$$

The proof of Theorem 3.3 will be presented in subsection 3.4 below.

We use regularization to show the existence of a weak solution $(\mathbf{m},\rho) \in V \times Q$ to the problem (3.2).

3.1. The regularized problem

For $\mathbf{u}, \mathbf{v} \in V$ and $p, q \in Q$, define

$$I(\mathbf{u},\mathbf{v}) = \int_\Omega |\mathrm{div}\,\mathbf{u}|^{s-2}\mathrm{div}\,\mathbf{u}\cdot \mathrm{div}\,\mathbf{v}\,d\mathbf{x} \quad \text{and} \quad J(p,q) = \int_\Omega |p|^{r-2}p\cdot q\,d\mathbf{x}. \tag{3.12}$$

For the fixed $\varepsilon > 0$, we consider the following regularized problem: Find $(\mathbf{m}_\varepsilon, \rho_\varepsilon) \in V \times Q$ such that

$$\begin{cases} a(\mathbf{m}_\varepsilon,\mathbf{v}) + \varepsilon I(\mathbf{m}_\varepsilon,\mathbf{v}) - b(\mathbf{v},\rho_\varepsilon) = -\Upsilon_\psi(\mathbf{v}) & \text{for all } \mathbf{v}\in V\\ \varepsilon J(\rho_\varepsilon, q) + b(\mathbf{m}_\varepsilon, q) = \Lambda_f(q) & \text{for all } q \in Q. \end{cases} \tag{3.13}$$

Lemma 3.4. *For every $\varepsilon > 0$, there is a unique solution $(\mathbf{m}_\varepsilon, \rho_\varepsilon) \in V \times Q$ of the regularized problem* (3.13).

Proof. Adding the left hand side of (3.13), we obtain the nonlinear form defined on $V \times Q$,

$$a_\varepsilon((\mathbf{m}_\varepsilon,\rho_\varepsilon),(\mathbf{v},q)) := a(\mathbf{m}_\varepsilon,\mathbf{v}) + \varepsilon I(\mathbf{m}_\varepsilon,\mathbf{v}) - b(\mathbf{v},\rho_\varepsilon) + \varepsilon J(\rho_\varepsilon,q) + b(\mathbf{m}_\varepsilon,q), \text{ for } (\mathbf{v},q)\in V\times Q. \tag{3.14}$$

A nonlinear operator $\mathcal{A}_\varepsilon : (V\times Q) \to (V \times Q)'$ defined by

$$\langle \mathcal{A}_\varepsilon((\mathbf{u},p)),(\mathbf{v},q)\rangle_{(V\times Q)'\times(V\times Q)} = a_\varepsilon((\mathbf{u},p),(\mathbf{v},q)).$$

Then, $\mathcal{A}_\varepsilon$ is continuous, coercive and strictly monotone.

Applying the theorem of Browder and Minty (see in Zeidler and Boron (1989), Thm. 26.A) for every $\tilde{f} \in (V\times Q)'$, there exists unique a solution $(\mathbf{m}_\varepsilon,\rho_\varepsilon) \in V\times Q$ of the operator equation $\mathcal{A}_\varepsilon(\mathbf{m}_\varepsilon,\rho_\varepsilon) = \tilde{f}$. In particular, we choose the linear form $\tilde{f}$ defined by $\tilde{f}(\mathbf{v},q) := -\Upsilon_\psi(\mathbf{v}) + \Lambda_f(q)$, which arises by adding the right hand sides of (3.13). Therefore, (3.13) has a unique solution.

Below, we establish that $\mathcal{A}_\varepsilon$ is continuous, coercive and strictly monotone.

Proof of the fact $\mathcal{A}_\varepsilon$ is continuous. For any $(\mathbf{u}_1,p_1),(\mathbf{u}_2,p_2),(\mathbf{v},q)\in V\times Q$, we have

$$\begin{aligned} \langle \mathcal{A}_\varepsilon((\mathbf{u}_1,p_1) - \mathcal{A}_\varepsilon((\mathbf{u}_2,p_2)),(\mathbf{v},q)\rangle_{(V\times Q)'\times(V\times Q)} &= a(\mathbf{u}_1,\mathbf{v}) - a(\mathbf{u}_2,\mathbf{v})\\ + \varepsilon(I(\mathbf{u}_1,\mathbf{v}) - I(\mathbf{u}_2,\mathbf{v})) - b(\mathbf{v},p_1-p_2) &+ \varepsilon(J(p_1,q) - J(p_2,q)) + b(\mathbf{u}_1-\mathbf{u}_2,q). \end{aligned} \tag{3.15}$$

Using the (3.6), we have

$$\begin{aligned} |a(\mathbf{u}_1,\mathbf{v}) - a(\mathbf{u}_2,\mathbf{v})| &\le c_4\left(1 + \|\mathbf{u}_1\|_{0,s}^{s-2} + \|\mathbf{u}_2\|_{0,s}^{s-2}\right)\|\mathbf{u}_1-\mathbf{u}_2\|_{0,s}\|\mathbf{v}\|_{0,s}\\ &\le c_4\left(1 + \|\mathbf{u}_1\|_V^{s-2} + \|\mathbf{u}_2\|_V^{s-2}\right)\|\mathbf{u}_1-\mathbf{u}_2\|_V\|\mathbf{v}\|_V. \end{aligned} \tag{3.16}$$

From (2.5) and Hölder's inequality, it follows that

$$\begin{aligned} |I(\mathbf{u}_1,\mathbf{v}) - I(\mathbf{u}_2,\mathbf{v})| &\le 2^{s-2}\int_\Omega (|\mathrm{div}\,\mathbf{u}_1|^{s-2} + |\mathrm{div}\,\mathbf{u}_2|^{s-2})\cdot|\mathrm{div}\,(\mathbf{u}_1-\mathbf{u}_2)|\cdot|\mathrm{div}\,\mathbf{v}|\,d\mathbf{x}\\ &\le 2^{s-2}\big(\|\mathrm{div}\,\mathbf{u}_1\|_{0,s}^{s-2} + \|\mathrm{div}\,\mathbf{u}_2\|_{0,s}^{s-2}\big)\|\mathrm{div}\,(\mathbf{u}_1-\mathbf{u}_2)\|_{0,s}\|\mathrm{div}\,\mathbf{v}\|_{0,s}\\ &\le 2^{s-2}\big(\|\mathbf{u}_1\|_V^{s-2} + \|\mathbf{u}_2\|_V^{s-2}\big)\|\mathbf{u}_1-\mathbf{u}_2\|_V\|\mathbf{v}\|_V. \end{aligned}$$

Using (2.8), together with Hölder's inequality and the fact that $r-2 \in (-1,0)$, we have

$$
\begin{aligned}
|J(p_1,q)-J(p_2,q)| &\le \int_\Omega \left||p_1|^{r-2}p_1-|p_2|^{r-2}p_2\right|\cdot|q|\,d\mathbf{x}\\
&\le 2^{2-r}\int_\Omega |p_1-p_2|^{r-1}|q|d\mathbf{x}\\
&\le 2^{2-r}\left\|p_1-p_2\right\|_Q^{r-1}\left\|q\right\|_Q.
\end{aligned}
$$

By Hölder's inequality,

$$
\begin{aligned}
|b(\mathbf{v},p_1-p_2)|+|b(\mathbf{u}_1-\mathbf{u}_2,q)| &\le \left\|\operatorname{div}\mathbf{v}\right\|_{0,s}\left\|p_1-p_2\right\|_{0,r}+\left\|\operatorname{div}(\mathbf{u}_1-\mathbf{u}_2)\right\|_{0,s}\left\|q\right\|_{0,r}\\
&\le \left\|p_1-p_2\right\|_Q\left\|\mathbf{v}\right\|_V+\left\|\mathbf{u}_1-\mathbf{u}_2\right\|_V\left\|q\right\|_Q.
\end{aligned}
$$

From (3.15) and the above estimates, it follows that

$$
\begin{aligned}
|\left\langle \mathcal{A}_\varepsilon((\mathbf{u}_1,p_1)-\mathcal{A}_\varepsilon((\mathbf{u}_2,p_2)),(\mathbf{v},q)\right\rangle_{(V\times Q)'\times(V\times Q)}| &\le c_5(1+\varepsilon)\big(1+\left\|\mathbf{u}_1\right\|_V^{s-2}+\left\|\mathbf{u}_2\right\|_V^{s-2}\big)\\
&\quad\cdot\big(\left\|\mathbf{u}_1-\mathbf{u}_2\right\|_V+\left\|p_1-p_2\right\|_Q+\left\|p_1-p_2\right\|_Q^{r-1}\big)\left\|(\mathbf{v},q)\right\|_{V\times Q},
\end{aligned}
$$

where $c_5=\max\{c_4,2^{s-2},2^{2-r}\}$. This yields

$$
\begin{aligned}
\left\|\mathcal{A}_\varepsilon((\mathbf{u}_1,p_1)-\mathcal{A}_\varepsilon((\mathbf{u}_2,p_2)\right\|_{(V\times Q)'} &\le c_5(1+\varepsilon)\big(1+\left\|\mathbf{u}_1\right\|_V^{s-2}+\left\|\mathbf{u}_2\right\|_V^{s-2}\big)\\
&\quad\cdot\big(\left\|\mathbf{u}_1-\mathbf{u}_2\right\|_V+\left\|p_1-p_2\right\|_Q+\left\|p_1-p_2\right\|_Q^{r-1}\big).
\end{aligned}
$$

Thus, $\mathcal{A}_\varepsilon$ is continuous.

Proof of the fact $\mathcal{A}_\varepsilon$ is coercive. For any $(\mathbf{u},p)\in V\times Q$, we have from (3.9) and (3.14), that

$$
\begin{aligned}
\left\langle \mathcal{A}_\varepsilon(\mathbf{u},p),(\mathbf{u},p)\right\rangle_{(V\times Q)'\times(V\times Q)} &= a(\mathbf{u},\mathbf{u})+\varepsilon I(\mathbf{u},\mathbf{u})+\varepsilon J(p,p)\\
&\ge c_2\left\|\mathbf{u}\right\|_{0,s}^s+\varepsilon\left\|\operatorname{div}\mathbf{u}\right\|_{0,s}^s+\varepsilon\left\|p\right\|_{0,r}^r\\
&\ge \min\{c_2,\varepsilon\}\big(\left\|\mathbf{u}\right\|_V^s+\left\|p\right\|_Q^r\big).
\end{aligned}
$$

Note that $s>2>r>1$. We consider $\left\|\mathbf{u}\right\|_V+\left\|p\right\|_Q\ge 2$.
If $\left\|\mathbf{u}\right\|_V\ge 1$, then

$$
\left\|\mathbf{u}\right\|_V^s+\left\|p\right\|_Q^r\ge\left\|\mathbf{u}\right\|_V^r+\left\|p\right\|_Q^r\ge 2^{1-r}(\left\|\mathbf{u}\right\|_V+\left\|p\right\|_Q)^r.
$$

If $\left\|\mathbf{u}\right\|_V<1$, then $\left\|p\right\|_Q>1>\left\|\mathbf{u}\right\|_V$, and

$$
\left\|\mathbf{u}\right\|_V^s+\left\|p\right\|_Q^r\ge\left\|p\right\|_Q^r\ge\left(\frac{1}{2}\left\|p\right\|_Q+\frac{1}{2}\left\|\mathbf{u}\right\|_V\right)^r=2^{-r}(\left\|\mathbf{u}\right\|_V+\left\|p\right\|_Q)^r.
$$

In both cases, we find that

$$
\frac{\left\langle \mathcal{A}_\varepsilon(\mathbf{u},p),(\mathbf{u},p)\right\rangle_{(V\times Q)'\times(V\times Q)}}{\left\|(\mathbf{v},q)\right\|_{V\times Q}}\ge 2^{-r}\min\{c_2,\varepsilon\}\left\|(\mathbf{v},q)\right\|_{V\times Q}^{r-1}\to\infty \text{ as } \left\|(\mathbf{v},q)\right\|_{V\times Q}\to\infty.
$$

Therefore, $\mathcal{A}_\varepsilon$ is coercive.

Proof of the fact $\mathcal{A}_\varepsilon$ is strictly monotone. By "strictly monotonotone", we mean that

$$\langle \mathcal{A}_\varepsilon(\mathbf{u}, p) - \mathcal{A}_\varepsilon(\mathbf{v}, q), (\mathbf{u} - \mathbf{v}, p - q)\rangle_{(V\times Q)'\times(V\times Q)} > 0$$

for all $(\mathbf{u}, p), (\mathbf{v}, q) \in V \times Q$ with $(\mathbf{u}, p) \neq (\mathbf{v}, q)$.

Let $(\mathbf{u}, p), (\mathbf{v}, q) \in V \times Q$. We have

$$\begin{aligned}\langle \mathcal{A}_\varepsilon(\mathbf{u}, p) - \mathcal{A}_\varepsilon(\mathbf{v}, q), (\mathbf{u} - \mathbf{v}, p - q)\rangle_{(V\times Q)'\times(V\times Q)} &= a(\mathbf{u}, \mathbf{u} - \mathbf{v}) - a(\mathbf{v}, \mathbf{u} - \mathbf{v}) \\ &+ \varepsilon(I(\mathbf{u}, \mathbf{u} - \mathbf{v}) - I(\mathbf{v}, \mathbf{u} - \mathbf{v})) + \varepsilon(J(p, p - q) - J(q, p - q)).\end{aligned}$$

Applying inequality (3.7), we obtain

$$a(\mathbf{u}, \mathbf{u} - \mathbf{v}) - a(\mathbf{u}, \mathbf{u} - \mathbf{v}) \geq c_2 \|\mathbf{u} - \mathbf{v}\|_{0,s}^s .$$

In addition, inequality (2.7) yields

$$I(\mathbf{u}, \mathbf{u} - \mathbf{v}) - I(\mathbf{v}, \mathbf{u} - \mathbf{v}) \geq 2^{1-s} \|\mathrm{div}\,(\mathbf{u} - \mathbf{v})\|_{0,s}^s .$$

Utilizing inequality (2.9), we have

$$J(p, p - q) - J(q, p - q) \geq (r - 1) \int_\Omega (|p| + |q|)^{r-2} |p - q|^2 \, d\mathbf{x}.$$

Consequently, by putting these estimates together, we arrive at

$$\begin{aligned}&\langle \mathcal{A}_\varepsilon(\mathbf{u}, p) - \mathcal{A}_\varepsilon(\mathbf{v}, q), (\mathbf{u} - \mathbf{v}, p - q)\rangle_{(V\times Q)'\times(V\times Q)} \\ &\qquad \geq C_2 \left(\|\mathbf{u} - \mathbf{v}\|_{0,s}^s + \varepsilon \|\mathrm{div}\,(\mathbf{u} - \mathbf{v})\|_{0,s}^s + \varepsilon \int_\Omega (|p| + |q|)^{r-2} |p - q|^2 \, d\mathbf{x} \right),\end{aligned}$$

where $C_2 = \min\{c_2, 2^{1-s}, r-1\}$. This implies that $\langle \mathcal{A}_\varepsilon(\mathbf{u}, p) - \mathcal{A}_\varepsilon(\mathbf{v}, q), (\mathbf{u} - \mathbf{v}, p - q)\rangle_{(V\times Q)'\times(V\times Q)}$ is positive whenever $(\mathbf{u}, p) \neq (\mathbf{v}, q)$.

Therefore, $\mathcal{A}_\varepsilon$ is strictly monotone. □

Next, we show that the solution $(\mathbf{m}_\varepsilon, \rho_\varepsilon)$ is bounded independently of ε.

Lemma 3.5. *There exist constants $\mathcal{C} > 0$, independent of ε, such that for sufficiently small $\varepsilon > 0$ the solution $(\mathbf{m}_\varepsilon, \rho_\varepsilon)$ of (3.13) satisfies the following estimates*

$$\|\mathbf{m}_\varepsilon\|_V + \|\rho_\varepsilon\|_Q \leq \mathcal{C}. \tag{3.17}$$

Proof. We begin with a bound for the norm of $\mathrm{div}\,\mathbf{m}_\varepsilon$. Choosing $q = |\mathrm{div}\,\mathbf{m}_\varepsilon|^{s-2}\mathrm{div}\,\mathbf{m}_\varepsilon \in Q$ in the second equation of (3.13) and using Hölder's inequality, we find that

$$\|\mathrm{div}\,\mathbf{m}_\varepsilon\|_{0,s}^s \leq \|\Lambda_f\|_{Q'} \|\mathrm{div}\,\mathbf{m}_\varepsilon\|_{0,s}^{s-1} + \varepsilon \|\rho_\varepsilon\|_{0,r}^{r-1} \|\mathrm{div}\,\mathbf{m}_\varepsilon\|_{0,s}^{s-1} .$$

It implies that

$$\|\mathrm{div}\,\mathbf{m}_\varepsilon\|_{0,s} \leq \|\Lambda_f\|_{Q'} + \varepsilon \|\rho_\varepsilon\|_{0,r}^{r-1} . \tag{3.18}$$

Choosing $(\mathbf{v}, q) = (\mathbf{m}_\varepsilon, \rho_\varepsilon)$ in (3.13) and utilizing (2.14) results in

$$\begin{aligned} a(\mathbf{m}_\varepsilon, \mathbf{m}_\varepsilon) + \varepsilon\|\operatorname{div}\mathbf{m}_\varepsilon\|_{0,s}^s + \varepsilon\|\rho_\varepsilon\|_{0,r}^r &= -\Upsilon_\psi(\mathbf{m}_\varepsilon) + \Lambda_f(\rho_\varepsilon) \\ &\le \|\Upsilon_\psi\|_{V'}\|\mathbf{m}_\varepsilon\|_V + \|\Lambda_f\|_{Q'}\|\rho_\varepsilon\|_Q \\ &\le \|\Upsilon_\psi\|_{V'}\big(\|\mathbf{m}_\varepsilon\|_{0,s} + \|\operatorname{div}\mathbf{m}_\varepsilon\|_{0,s}\big) + \|\Lambda_f\|_{Q'}\|\rho_\varepsilon\|_{0,r}\,. \end{aligned} \tag{3.19}$$

Applying inequality (3.8) to the first term of (3.19), neglecting the next two terms, and utilizing the estimate (3.18) for the last divergence term, we obtain

$$c_2\|\mathbf{m}_\varepsilon\|_{0,s}^s \le \|\Upsilon_\psi\|_{V'}\big(\|\mathbf{m}_\varepsilon\|_{0,s} + \|\Lambda_f\|_{Q'} + \varepsilon\|\rho_\varepsilon\|_{0,r}^{r-1}\big) + \|\Lambda_f\|_{Q'}\|\rho_\varepsilon\|_{0,r}\,.$$

By Young's inequality, specifically, the last one in (2.1), we have

$$\|\Upsilon_\psi\|_{V'}\|\mathbf{m}_\varepsilon\|_{0,s} \le (c_2/2)\|\mathbf{m}_\varepsilon\|_{0,s}^s + (2/c_2)^{r/s}\|\Upsilon_\psi\|_{V'}^r\,.$$

It follows that

$$\|\mathbf{m}_\varepsilon\|_{0,s}^s \le C\left(\|\Upsilon_\psi\|_{V'}^r + \|\Upsilon_\psi\|_{V'}\|\Lambda_f\|_{Q'} + \varepsilon\|\Upsilon_\psi\|_{V'}\|\rho_\varepsilon\|_{0,r}^{r-1} + \|\Lambda_f\|_{Q'}\|\rho_\varepsilon\|_{0,r}\right). \tag{3.20}$$

To bound ρ_ε, we employ the inf-sup condition (2.17). The first equation in (3.13) and the above estimate for $\|\operatorname{div}\mathbf{m}_\varepsilon\|_{0,s}$, we have

$$\begin{aligned} b(\mathbf{v}, \rho_\varepsilon) &= a(\mathbf{m}_\varepsilon, \mathbf{v}) + \varepsilon I(\mathbf{m}_\varepsilon, \mathbf{v}) + \Upsilon_\psi(\mathbf{v}) \\ &\le C(\|\mathbf{m}_\varepsilon\|_{0,s} + \|\mathbf{m}_\varepsilon\|_{0,s}^{s-1})\|\mathbf{v}\|_{0,s} + \varepsilon\|\operatorname{div}\mathbf{m}_\varepsilon\|_{0,s}^{s-1}\|\operatorname{div}\mathbf{v}\|_{0,s} + \|\Upsilon_\psi\|_{V'}(\|\mathbf{v}\|_{0,s} + \|\operatorname{div}\mathbf{v}\|_{0,s}) \\ &\le \left[C\left(\|\mathbf{m}_\varepsilon\|_{0,s} + \|\mathbf{m}_\varepsilon\|_{0,s}^{s-1}\right) + \varepsilon\|\operatorname{div}\mathbf{m}_\varepsilon\|^{s-1} + \|\Upsilon_\psi\|_{V'}\right]\|\mathbf{v}\|_V\,. \end{aligned}$$

Consequently,

$$\begin{aligned} \|\rho_\varepsilon\|_{0,r} &\le C_* \sup_{\mathbf{v}\in V\setminus\{0\}} \frac{b(\mathbf{v}, \rho_\varepsilon)}{\|\mathbf{v}\|_V} \le C_*\left[C\left(\|\mathbf{m}_\varepsilon\|_{0,s} + \|\mathbf{m}_\varepsilon\|_{0,s}^{s-1}\right) + \varepsilon\|\operatorname{div}\mathbf{m}_\varepsilon\|^{s-1} + \|\Upsilon_\psi\|_{V'}\right] \\ &\le C_0\left[\|\mathbf{m}_\varepsilon\|_{0,s} + \|\mathbf{m}_\varepsilon\|_{0,s}^{s-1} + \varepsilon\big(\|\Lambda_f\|_{Q'} + \varepsilon\|\rho_\varepsilon\|_{0,r}^{r-1}\big)^{s-1} + \|\Upsilon_\psi\|_{V'}\right] \\ &\le 2^{s-1}C_0\left[\|\mathbf{m}_\varepsilon\|_{0,s} + \|\mathbf{m}_\varepsilon\|_{0,s}^{s-1} + \varepsilon\|\Lambda_f\|_{Q'}^{s-1} + \varepsilon^s\|\rho_\varepsilon\|_{0,r} + \|\Upsilon_\psi\|_{V'}\right]. \end{aligned}$$

Here, we used the fact that $(r-1)(s-1) = 1$ in the last inequality.

By setting $\varepsilon_0 = \min\{1, (2^{s-2}C_0)^{-1/s}\}$ and considering $\varepsilon \in (0, \varepsilon_0)$, we deduce

$$\|\rho_\varepsilon\|_{0,r} \le 2^s C_0\left(\|\mathbf{m}_\varepsilon\|_{0,s} + \|\mathbf{m}_\varepsilon\|_{0,s}^{s-1} + \varepsilon\|\Lambda_f\|_{Q'}^{s-1} + \|\Upsilon_\psi\|_{V'}\right). \tag{3.21}$$

Consequently, one has

$$\|\rho_\varepsilon\|_{0,r}^{r-1} \le C_1\left(\|\mathbf{m}_\varepsilon\|_{0,s}^{r-1} + \|\mathbf{m}_\varepsilon\|_{0,s} + \varepsilon^{r-1}\|\Lambda_f\|_{Q'} + \|\Upsilon_\psi\|_{V'}^{r-1}\right). \tag{3.22}$$

Substituting (3.21) and (3.22) into (3.20) leads to

$$\begin{aligned}\|\mathbf{m}_\varepsilon\|_{0,s}^s \le & C\Big[\, \|\Upsilon_\psi\|_{V'}^r + \|\Upsilon_\psi\|_{V'} \|\Lambda_f\|_{Q'} \\ & + \varepsilon \|\Upsilon_\psi\|_{V'} \left(\|\mathbf{m}_\varepsilon\|_{0,s}^{r-1} + \|\mathbf{m}_\varepsilon\|_{0,s} + \varepsilon^{r-1} \|\Lambda_f\|_{Q'} + \|\Upsilon_\psi\|_{V'}^{r-1} \right) \\ & + \|\Lambda_f\|_{Q'} \left(\|\mathbf{m}_\varepsilon\|_{0,s} + \|\mathbf{m}_\varepsilon\|_{0,s}^{s-1} + \varepsilon \|\Lambda_f\|_{Q'}^{s-1} + \|\Upsilon_\psi\|_{V'} \right) \Big] \\ \le & C\Big[(1+\varepsilon) \|\Upsilon_\psi\|_{V'}^r + (1+\varepsilon^r) \|\Upsilon_\psi\|_{V'} \|\Lambda_f\|_{Q'} + \varepsilon \|\Lambda_f\|_{Q'}^s \\ & + \varepsilon \|\Upsilon_\psi\|_{V'} \left(\|\mathbf{m}_\varepsilon\|_{0,s}^{r-1} + \|\mathbf{m}_\varepsilon\|_{0,s} \right) + \|\Lambda_f\|_{Q'} \left(\|\mathbf{m}_\varepsilon\|_{0,s} + \|\mathbf{m}_\varepsilon\|_{0,s}^{s-1} \right) \Big].\end{aligned}$$

Let $s_* = s/(s-r+1) = s(s-1)/(s(s-1)-1) \in (1, r)$. Then using Young's inequality, we obtain

$$\begin{aligned}\|\mathbf{m}_\varepsilon\|_{0,s}^s \le C\Big[(1+\varepsilon) \|\Upsilon_\psi\|_{V'}^r + (1+\varepsilon^r) \|\Upsilon_\psi\|_{V'} \|\Lambda_f\|_{Q'} + \varepsilon \|\Lambda_f\|_{Q'}^s \Big] \\ + \frac{1}{2} \|\mathbf{m}_\varepsilon\|_{0,s}^s + C\big(\varepsilon^{s_*} \|\Upsilon_\psi\|_{V'}^{s_*} + \varepsilon^r \|\Upsilon_\psi\|_{V'}^r \big) + C(\|\Lambda_f\|_{Q'}^r + \|\Lambda_f\|_{Q'}^s).\end{aligned}$$

This implies

$$\|\mathbf{m}_\varepsilon\|_{0,s} \le C_1 d_1(\varepsilon), \tag{3.23}$$

where

$$d_1(\varepsilon) = \Big[(1+\varepsilon^r)(\|\Upsilon_\psi\|_{V'}^r + \|\Upsilon_\psi\|_{V'} \|\Lambda_f\|_{Q'}) + (1+\varepsilon) \|\Lambda_f\|_{Q'}^s + \varepsilon^{s_*} \|\Upsilon_\psi\|_{V'}^{s_*} + \|\Lambda_f\|_{Q'}^r \Big]^{1/s}.$$

Inserting (3.23) into (3.21) yields

$$\|\rho_\varepsilon\|_{0,r} \le C_2 d_2(\varepsilon), \tag{3.24}$$

where $d_2(\varepsilon) = d_1(\varepsilon) + d_1(\varepsilon)^{s-1} + \varepsilon \|\Lambda_f\|_{Q'}^{s-1} + \|\Upsilon_\psi\|_{V'}$.
Using this estimate in (3.18) yields

$$\|\mathrm{div}\, \mathbf{m}_\varepsilon\|_{0,s} \le C_3 d_3(\varepsilon), \tag{3.25}$$

where $d_3(\varepsilon) = \|\Lambda_f\|_{Q'} + \varepsilon d_2(\varepsilon)^{r-1}$.
Observe that $d_i(\varepsilon)$, for $i = 1, 2, 3$, are increasing functions with respect to ε. Summing up the estimates (3.23), (3.24) and (3.25) gives

$$\|\mathbf{m}_\varepsilon\|_V + \|\rho_\varepsilon\|_Q \le \mathcal{C} \stackrel{\text{def}}{=\!=} \sum_{i=1}^{3} C_i d_i(1).$$

Thus, we obtain the desired estimate (3.17).

The proof is complete. □

3.2. Existence and uniqueness

This subsection is dedicated to establish the existence and uniqueness of a weak solution to problem (3.2).

Theorem 3.6. *Suppose* $f \in L^s(\Omega)$ *and* $\psi \in X_r$*. The mixed formulation* (3.2) *of the problem* (3.1) *has a unique solution* $(\mathbf{m}, \rho) \in \mathbf{W}_s(\mathrm{div}, \Omega) \times L^r(\Omega)$.

Proof. Adding the left hand side of (3.2), we obtain the following nonlinear form defined on $V \times Q$ by

$$a((\mathbf{m}, \rho), (\mathbf{v}, q)) := a(\mathbf{m}, \mathbf{v}) - b(\mathbf{v}, \rho) + b(\mathbf{m}, q).$$

Consider the nonlinear operator $\mathcal{A} : V \times Q \to (V \times Q)'$ defined by

$$\langle \mathcal{A}(\mathbf{u}, p), (\mathbf{v}, q)\rangle_{(V\times Q)'\times(V\times Q)} := a((\mathbf{u}, p), (\mathbf{v}, q)).$$

Set $\varepsilon = 1/n$ and let $(\mathbf{m}_n, \rho_n)$ be the unique solution of the regularized problem (3.13). Since $(\mathbf{m}_n, \rho_n)$ is a bounded sequence in $V \times Q$, there exists a weakly convergent subsequence, again denoted by $(\mathbf{m}_n, \rho_n)$, with weak limit $(\mathbf{m}, \rho) \in V \times Q$. For $\tilde{f}(\mathbf{v}, q) := -\Upsilon_\psi(\mathbf{v}) + \Lambda_f(q) \in (V \times Q)'$,

$$\begin{aligned} \left\|\mathcal{A}(\mathbf{m}_n, \rho_n) - \tilde{f}\right\|_{(V\times Q)'} &= \sup_{(\mathbf{v},q)\neq\mathbf{0}} \frac{|a((\mathbf{m}_n, \rho_n), (\mathbf{v}, q)) - \tilde{f}(\mathbf{v}, q)|}{\|(\mathbf{v}, q)\|_{V\times Q}} \\ &= \sup_{(\mathbf{v},q)\neq\mathbf{0}} \frac{|a(\mathbf{m}_n, \mathbf{v}) - b(\mathbf{v}, \rho_n) + b(\mathbf{m}_n, q) - \tilde{f}(\mathbf{v}, q)|}{\|(\mathbf{v}, q)\|_{V\times Q}}. \end{aligned} \tag{3.26}$$

Noting from (3.13) that

$$\begin{aligned} \left|a(\mathbf{m}_n, \mathbf{v}) - b(\mathbf{v}, \rho_n) + b(\mathbf{m}_n, q) - \tilde{f}(\mathbf{v}, q)\right| &= \frac{1}{n}\left|I(\mathbf{m}_n, \mathbf{v}) + J(\rho_n, q)\right| \\ &\leq \frac{1}{n}\left[\|\mathrm{div}\,\mathbf{m}_n\|_{0,s}^{s-1} \|\mathrm{div}\,\mathbf{v}\|_{0,s} + \|\rho_n\|_{0,r}^{r-1} \;\; \|q\|_{0,r}\right] \\ &\leq \frac{1}{n}\left[\|\mathrm{div}\,\mathbf{m}_n\|_{0,s}^{s-1} + \|\rho_n\|_{0,r}^{r-1}\right] \|(\mathbf{v}, q)\|_{V\times Q}. \end{aligned}$$

Hence,

$$\left\|\mathcal{A}(\mathbf{m}_n, \rho_n) - \tilde{f}\right\|_{(V\times Q)'} \leq \frac{C}{n}\left[\|\mathrm{div}\,\mathbf{m}_n\|_{0,s}^{s-1} + \|\rho_n\|_{0,r}^{r-1}\right] \overset{n\to\infty}{\longrightarrow} 0. \tag{3.27}$$

The sequence $\mathcal{A}(\mathbf{m}_n, \rho_n)$ converges strongly in $(V \times Q)'$ to $\tilde{f}$. Thus, we can conclude that $\mathcal{A}(\mathbf{m}, \rho) = \tilde{f}$ in $(V \times Q)'$ (see e.g. Zeidler (1990), p. 474), i.e., $(\mathbf{m}, \rho)$ is a solution of problem (3.10).

To show the uniqueness, we consider two solutions $(\mathbf{m}_1, \rho_1)$ and $(\mathbf{m}_2, \rho_2)$ of (3.10). Using the test function $(\mathbf{v}, q) = (\mathbf{m}_1 - \mathbf{m}_2, \rho_1 - \rho_2)$, we obtain

$$\begin{aligned} a(\mathbf{m}_1, \mathbf{m}_1 - \mathbf{m}_2) - a(\mathbf{m}_2, \mathbf{m}_1 - \mathbf{m}_2) - (b(\mathbf{m}_1 - \mathbf{m}_2, \rho_1) - b(\mathbf{m}_1 - \mathbf{m}_2, \rho_2)) &= 0, \\ b(\mathbf{m}_1, \rho_1 - \rho_2) - b(\mathbf{m}_2, \rho_1 - \rho_2) &= 0. \end{aligned} \tag{3.28}$$

Adding these equations yields

$$0 = a(\mathbf{m}_1, \mathbf{m}_1 - \mathbf{m}_2) - a(\mathbf{m}_2, \mathbf{m}_1 - \mathbf{m}_2) \geq c_2 \|\mathbf{m}_1 - \mathbf{m}_2\|_{0,s}^s .$$

It follows that $\mathbf{m}_1 = \mathbf{m}_2$.

For $i = 1, 2$, we have the variational equation $a(\mathbf{m}_i, \mathbf{v}) - b(\mathbf{v}, \rho_i) = -\Upsilon_\psi(\mathbf{v})$ for all $\mathbf{v} \in V$. Subtracting these two equations implies $b(\mathbf{v}, p_1 - p_2) = a(\mathbf{m}_1, \mathbf{v}) - a(\mathbf{m}_2, \mathbf{v}) = 0$ for all $\mathbf{v} \in V$. By applying inequality (2.17) in Lemma 2.3 to $q = \rho_1 - \rho_2$, we obtain $\rho_1 = \rho_2$. □

3.3. Estimates

Regarding the unique solutions of equation (3.10), we have the following estimates.

Theorem 3.7. *Let* $(\mathbf{m}, \rho) \in V \times Q$ *be the solution of* (3.10). *Then*

$$\begin{aligned} &\|\mathbf{m}\|_{0,s} \leq C(\|\psi\|_{X^r}^{r-1} + \|f\|_{0,s}^{r-1} + \|f\|_{0,s}), \\ &\|\operatorname{div} \mathbf{m}\|_{0,s} \leq \|f\|_{0,s} , \\ &\|\rho\|_{0,r} \leq C(\|f\|_{0,s}^{r-1} + \|f\|_{0,s}^{s-1} + \|\psi\|_{X^r}^{r-1} + \|\psi\|_{X^r}). \end{aligned} \tag{3.29}$$

Proof. We repeat the calculations in Lemma 3.5 with $\varepsilon = 0$. It follows (3.18), (3.20) and (3.21) that

$$\|\operatorname{div} \mathbf{m}\|_{0,s} \leq \|\Lambda_f\|_{Q'} , \tag{3.30}$$

$$\|\mathbf{m}\|_{0,s}^s \leq C \left(\|\Upsilon_\psi\|_{V'}^r + \|\Upsilon_\psi\|_{V'} \|\Lambda_f\|_{Q'} + \|\Lambda_f\|_{Q'} \|\rho\|_{0,r} \right) , \tag{3.31}$$

$$\|\rho\|_{0,r} \leq C \left(\|\mathbf{m}\|_{0,s} + \|\mathbf{m}\|_{0,s}^{s-1} + \|\Upsilon_\psi\|_{V'} \right) . \tag{3.32}$$

Substituting (3.32) into (3.31) leads to

$$\|\mathbf{m}_\varepsilon\|_{0,s}^s \leq C\Big(\|\Upsilon_\psi\|_{V'}^r + \|\Upsilon_\psi\|_{V'} \|\Lambda_f\|_{Q'} + \|\Lambda_f\|_{Q'} (\|\mathbf{m}_\varepsilon\|_{0,s} + \|\mathbf{m}_\varepsilon\|_{0,s}^{s-1})\Big).$$

Then by using Young's inequality, we obtain

$$\|\mathbf{m}\|_{0,s} \leq C\mathcal{Z}, \quad \|\rho\|_{0,r} \leq C\mathcal{W}, \quad \|\operatorname{div} \mathbf{m}\|_{0,s} \leq \|\Lambda_f\|_{Q'} , \tag{3.33}$$

where $\mathcal{Z} = (\|\Upsilon_\psi\|_{V'}^r + \|\Upsilon_\psi\|_{V'} \|\Lambda_f\|_{Q'} + \|\Lambda_f\|_{Q'}^r + \|\Lambda_f\|_{Q'}^s)^{1/s}$, $\mathcal{W} = \mathcal{Z} + \mathcal{Z}^{s-1} + \|\Upsilon_\psi\|_{V'}$.

Let C denote a generic positive constant as in the proof of Lemma 3.5. Using inequalities (2.3) and (2.1) yields

$$\begin{aligned} \mathcal{Z} &\leq C\big(\|\Upsilon_\psi\|_{V'}^r + \|\Lambda_f\|_{Q'}^r + \|\Lambda_f\|_{Q'}^s \big)^{1/s} \leq C(\|\Upsilon_\psi\|_{V'}^{r-1} + \|\Lambda_f\|_{Q'}^{r-1} + \|\Lambda_f\|_{Q'}), \\ \mathcal{W} &\leq C\big(\|\Upsilon_\psi\|_{V'}^{r-1} + \|\Lambda_f\|_{Q'}^{r-1} + \|\Lambda_f\|_{Q'} \big) + C\big(\|\Upsilon_\psi\|_{V'}^{r-1} + \|\Lambda_f\|_{Q'}^{r-1} + \|\Lambda_f\|_{Q'} \big)^{s-1} + \|\Upsilon_\psi\|_{V'} \\ &\leq C(\|\Upsilon_\psi\|_{V'}^{r-1} + \|\Upsilon_\psi\|_{V'} + \|\Lambda_f\|_{Q'}^{r-1} + \|\Lambda_f\|_{Q'}^{s-1}). \end{aligned}$$

Note from (3.3) and (3.4) that

$$\|\Upsilon_\psi\|_{V'} \leq \bar{c}_1 \|\psi\|_{X^r} \text{ and } \|\Lambda_f\|_{Q'} = \|f\|_{0,s} .$$

Then, we obtain the estimates (3.29). □

3.4. Proof of Theorem 3.3

Proof. Part 1. The statement follows Theorem 3.6.
Part 2. We apply Theorem 3.7, then estimate (3.11) follows (3.29). □

4. A mixed finite element approximation

We assume that the boundary $\partial\Omega$ of Ω is polygonal or polyhedral. Let $\{\mathcal{T}_h\}_h$ be a regular triangulation of $\overline{\Omega}$ with $\max_{\tau\in\mathcal{T}_h} \operatorname{diam} \tau \leq \mathrm{h}$. The discrete subspaces $V_h \times Q_h \subset V \times Q$ are defined as

$$Q_h = \{\rho_h \in L^2(\Omega), \forall\tau \in \mathcal{T}_h, \rho_h|_\tau \in P_k(\tau)\},$$
$$V_h = \{\mathbf{m}_h \in V, \forall\tau \in \mathcal{T}_h, \mathbf{m}_h|_\tau \in RT_k(\tau)\},$$

with $P_k(\tau)$ being the space of polynomial of degree at most k on the element τ and

$$RT_k(\tau) = (P_k(\tau))^n + \mathbf{x}P_k(\tau).$$

For momentum, let $\Pi : V \to V_h$ be the Raviart-Thomas projection Raviart and Thomas (1977), which satisfies

$$\int_\Omega \operatorname{div}(\Pi\mathbf{m} - \mathbf{m})\, q\, d\mathbf{x} = 0, \quad \text{for all } \mathbf{m} \in V, q \in Q_h.$$

For density, we use the standard L^2-projection operator, see in Ciarlet (1978), $\pi : Q \to Q_h$, satisfying

$$\int_\Omega (\pi\rho - \rho)\, q\, d\mathbf{x} = 0, \quad \text{for all } \rho \in Q, q \in Q_h,$$
$$\int_\Omega (\pi\rho - \rho)\operatorname{div} \mathbf{m}_h\, d\mathbf{x} = 0, \quad \text{for all } \mathbf{m}_h \in V_h, \rho \in Q.$$

This projection has well-known approximation properties, e.g. Brezzi and Fortin (1991); Johnson and Thomée (1981); Bramble et al. (2002).

$$\|\Pi\mathbf{m} - \mathbf{m}\|_{0,q} \leq Ch^p \|\mathbf{m}\|_{p,q}, \quad 1/q < p \leq k+1, \forall\mathbf{m} \in V \cap (W^{p,q}(\Omega))^d. \tag{4.1}$$
$$\|\pi\rho - \rho\|_{0,q} \leq Ch^p \|\rho\|_{p,q}, \quad 0 \leq p \leq k+1, q \in [1,\infty], \forall\rho \in p,q. \tag{4.2}$$

The two projections π and Π preserve the commuting property $\operatorname{div} \circ \Pi = \pi \circ \operatorname{div} : V \to Q_h$.

The discrete formulation of (3.2) can read as follows: Find $\mathbf{m}_h \in V_h$, $\rho_h \in Q_h$ such that

$$\begin{cases} a(\mathbf{m}_h, \mathbf{v}) - b(\mathbf{v}, \rho_h) = -\Upsilon_\psi(\mathbf{v}) & \text{for all } \mathbf{v} \in V_h, \\ b(\mathbf{m}_h, q) = \Lambda_f(q) & \text{for all } q \in Q_h. \end{cases} \tag{4.3}$$

In a similar manner to problem (3.2), we obtain the following:

Theorem 4.1. *For any $f \in L^s(\Omega)$ and $\psi \in X_r$, the problem* (4.3) *has a unique solution $(\mathbf{m}_h, \rho_h) \in V_h \times Q_h$, and there exists a constant $C > 0$ such that*

$$\|\mathbf{m}_h\|_{0,s} + \|\operatorname{div} \mathbf{m}_h\|_{0,s} + \|\rho_h\|_{0,r} \le C\big(\|f\|_{0,s}^{r-1} + \|f\|_{0,s}^{s-1} + \|\psi\|_{X_r}^{r-1} + \|\psi\|_{X_r}\big). \tag{4.4}$$

4.1. Error estimates

In this subsection, we will give the error estimate between the analytical solution and the approximate solution.

Lemma 4.2. *For any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbf{L}^s(\Omega)$, one has*

$$|a(\mathbf{u},\mathbf{w}) - a(\mathbf{v},\mathbf{w})| \le c_3 \Big[\|\mathbf{u}-\mathbf{v}\|_{0,2} \|\mathbf{w}\|_{0,2} + \Phi(\mathbf{u},\mathbf{v})\, \Psi(\mathbf{u},\mathbf{v}) \|\mathbf{w}\|_{0,s} \Big], \tag{4.5}$$

$$a(\mathbf{u},\mathbf{u}-\mathbf{v}) - a(\mathbf{v},\mathbf{u}-\mathbf{v}) \ge c_0 \Big[\|\mathbf{u}-\mathbf{v}\|_{0,2}^2 + \|\mathbf{u}-\mathbf{v}\|_{0,s}^s + \Phi^2(\mathbf{u},\mathbf{v}) \Big], \tag{4.6}$$

where $c_3 = \sqrt{2}c_1$, $c_0 = \min\{1, a_*\} 2^{-2(s-2)-1}$ are positive constants and

$$\Phi(\mathbf{u},\mathbf{v}) = \left(\int_\Omega (|\mathbf{u}|^{s-2} + |\mathbf{v}|^{s-2}) |\mathbf{u}-\mathbf{v}|^2 \, d\mathbf{x} \right)^{1/2}, \quad \Psi(\mathbf{u},\mathbf{v}) = \left(\|\mathbf{u}\|_{0,s}^{s-2} + \|\mathbf{v}\|_{0,s}^{s-2} \right)^{1/2}. \tag{4.7}$$

Proof. Using property (2.10) and Hölder's inequality, we have

$$\begin{aligned}
|a(\mathbf{u},\mathbf{w}) - a(\mathbf{v},\mathbf{w})| &\le c_1 \int_\Omega (1 + |\mathbf{u}|^{s-2} + |\mathbf{v}|^{s-2}) |\mathbf{u}-\mathbf{v}||\mathbf{w}| \, d\mathbf{x} \\
&= c_1 \int_\Omega |\mathbf{u}-\mathbf{v}||\mathbf{w}| + (|\mathbf{u}|^{s-2} + |\mathbf{v}|^{s-2})^{1/2} |\mathbf{u}-\mathbf{v}| (|\mathbf{u}|^{s-2} + |\mathbf{v}|^{s-2})^{1/2} |\mathbf{w}| \, d\mathbf{x} \\
&\le c_1 \|\mathbf{u}-\mathbf{v}\|_{0,2} \|\mathbf{w}\|_{0,2} + c_1 \Phi(\mathbf{u},\mathbf{v}) \Big(\int_\Omega (|\mathbf{u}|^{s-2} + |\mathbf{v}|^{s-2}) |\mathbf{w}|^2 \, d\mathbf{x} \Big)^{1/2}.
\end{aligned}$$

Applying Hölder's inequality gives

$$\begin{aligned}
\int_\Omega (|\mathbf{u}|^{s-2} + |\mathbf{v}|^{s-2}) |\mathbf{w}|^2 \, d\mathbf{x} &\le \left(\int_\Omega (|\mathbf{u}|^{s-2} + |\mathbf{v}|^{s-2})^{s/(s-2)} \, d\mathbf{x} \right)^{(s-2)/s} \left(\int_\Omega |\mathbf{w}|^s \, d\mathbf{x} \right)^{2/s} \\
&\le 2^{2/s} \left(\int_\Omega (|\mathbf{u}|^s + |\mathbf{v}|^s) \, d\mathbf{x} \right)^{(s-2)/s} \left(\int_\Omega |\mathbf{w}|^s \, d\mathbf{x} \right)^{2/s} \\
&\le 2 (\|\mathbf{u}\|_{0,s}^{s-2} + \|\mathbf{v}\|_{0,s}^{s-2}) \|\mathbf{w}\|_{0,s}^2 .
\end{aligned}$$

Thus, we obtain (4.5).

Finally, by (2.6), we have

$$\begin{aligned}
a(\mathbf{u},\mathbf{u}-\mathbf{v}) - a(\mathbf{v},\mathbf{u}-\mathbf{v}) &= \int_\Omega (F(|\mathbf{u}|)\mathbf{u} - F(|\mathbf{v}|)\mathbf{v}) \cdot (\mathbf{u}-\mathbf{v}) \, d\mathbf{x} \\
&\ge \int_\Omega \Big(a_0 |\mathbf{u}-\mathbf{v}|^2 + \frac{1}{2} \sum_{i=1}^N a_i (|\mathbf{u}|^{\alpha_i} + |\mathbf{v}|^{\alpha_i}) |\mathbf{u}-\mathbf{v}|^2 \Big) \, d\mathbf{x} \\
&\ge 2^{-1} a_* \left[\int_\Omega |\mathbf{u}-\mathbf{v}|^2 \, d\mathbf{x} + \int_\Omega (|\mathbf{u}|^{s-2} + |\mathbf{v}|^{s-2}) |\mathbf{u}-\mathbf{v}|^2 \, d\mathbf{x} \right].
\end{aligned}$$

Using (2.3) $|\mathbf{x}|^p + |\mathbf{y}^p| \ge 2^{-p}(|\mathbf{x}| + |\mathbf{y}|)^p$ and $(|\mathbf{x}| + |\mathbf{y}|)^p \ge 2^{-(p-1)^+}|\mathbf{x} - \mathbf{y}|^p, p > 0$, it follows that

$$\begin{aligned}\int_\Omega (|\mathbf{u}|^{s-2} + |\mathbf{v}|^{s-2})|\mathbf{u} - \mathbf{v}|^2\, d\mathbf{x} &= \int_\Omega (\frac{1}{2} + \frac{1}{2})(|\mathbf{u}|^{s-2} + |\mathbf{v}|^{s-2})|\mathbf{u} - \mathbf{v}|^2\, d\mathbf{x} \\ &\ge 2^{-2(s-2)} \|\mathbf{u} - \mathbf{v}\|_{0,s}^s + 2^{-1}\Phi^2(\mathbf{u}, \mathbf{v}),\end{aligned} \tag{4.8}$$

which proves (4.6). □

Theorem 4.3. *Let* $(\mathbf{m}, \rho) \in V \times Q$ *be the solution of* (3.2) *and* $(\mathbf{m}_h, \rho_h) \in V_h \times Q_h$ *be the solution of* (4.3)*. Then, there exist positive constants* C *independent of* h *such that*

$$\|\mathbf{m} - \mathbf{m}_h\|_{0,2}^2 + \|\mathbf{m} - \mathbf{m}_h\|_{0,s}^s \le C(\|\mathbf{m} - \Pi\mathbf{m}\|_{0,2}^2 + \|\mathbf{m} - \Pi\mathbf{m}\|_{0,s}^2). \tag{4.9}$$

$$\|\rho - \rho_h\|_{0,r} \le C(\|\mathbf{m} - \Pi\mathbf{m}\|_{0,2} + \|\mathbf{m} - \Pi\mathbf{m}\|_{0,s} + \|\rho - \pi\rho\|_{0,r}). \tag{4.10}$$

Proof. By (3.10) and (4.3), we have the error equations

$$\begin{aligned} a(\mathbf{m}, \mathbf{v}) - a(\mathbf{m}_h, \mathbf{v}) - b(\mathbf{v}, \rho - \rho_h) &= 0 \quad \text{for all } \mathbf{v} \in V_h, \\ b(\mathbf{m} - \mathbf{m}_h, q) &= 0 \quad \text{for all } q \in Q_h. \end{aligned} \tag{4.11}$$

Using L^2-project and Raviart-Thomas projection, we rewrite (4.11) in the form

$$\begin{aligned} a(\mathbf{m}, \mathbf{v}) - a(\mathbf{m}_h, \mathbf{v}) - b(\mathbf{v}, \pi\rho - \rho_h) &= 0, \\ b(\Pi\mathbf{m} - \mathbf{m}_h, q) &= 0. \end{aligned}$$

Choosing $q = \pi\rho - \rho_h \in Q_h$ and $\mathbf{v} = \Pi\mathbf{m} - \mathbf{m}_h \in V_h$ and adding the two resulting equations, we obtain

$$a(\mathbf{m}, \Pi\mathbf{m} - \mathbf{m}_h) - a(\mathbf{m}_h, \Pi\mathbf{m} - \mathbf{m}_h) = 0,$$

that is

$$a(\mathbf{m}, \mathbf{m} - \mathbf{m}_h) - a(\mathbf{m}_h, \mathbf{m} - \mathbf{m}_h) = a(\mathbf{m}, \mathbf{m} - \Pi\mathbf{m}) - a(\mathbf{m}_h, \mathbf{m} - \Pi\mathbf{m}). \tag{4.12}$$

By (4.6),

$$a(\mathbf{m}, \mathbf{m} - \mathbf{m}_h) - a(\mathbf{m}_h, \mathbf{m} - \mathbf{m}_h) \ge c_0 \left(\|\mathbf{m} - \mathbf{m}_h\|_{0,2}^2 + \|\mathbf{m} - \mathbf{m}_h\|_{0,s}^s + \Phi^2(\mathbf{m}, \mathbf{m}_h)\right). \tag{4.13}$$

Using (4.5), it follows that

$$\begin{aligned} &a(\mathbf{m}, \mathbf{m} - \Pi\mathbf{m}) - a(\mathbf{m}_h, \mathbf{m} - \Pi\mathbf{m}) \\ &\quad \le c_3 \left[\|\mathbf{m} - \mathbf{m}_h\|_{0,2} \|\mathbf{m} - \Pi\mathbf{m}\|_{0,2} + \Phi(\mathbf{m}, \mathbf{m}_h)\Psi(\mathbf{m}, \mathbf{m}_h) \|\mathbf{m} - \Pi\mathbf{m}\|_{0,s}\right] \\ &\quad \le \varepsilon \left[\|\mathbf{m} - \mathbf{m}_h\|_{0,2}^2 + \Phi^2(\mathbf{m}, \mathbf{m}_h)\right] + c_3^2\varepsilon^{-1} \left[\|\mathbf{m} - \Pi\mathbf{m}\|_{0,2}^2 + \Psi^2(\mathbf{m}, \mathbf{m}_h) \|\mathbf{m} - \Pi\mathbf{m}\|_{0,s}^2\right]. \end{aligned}$$

Choosing $\varepsilon = 2^{-1}c_0$ and applying the above inequalities, we deserve

$$\begin{aligned} \|\mathbf{m} - \mathbf{m}_h\|_{0,2}^2 + \|\mathbf{m} - \mathbf{m}_h\|_{0,s}^s + \Phi^2(\mathbf{m}, \mathbf{m}_h) &\le c_4 \left(\|\mathbf{m} - \Pi\mathbf{m}\|_{0,2}^2 + \Psi^2(\mathbf{m}, \mathbf{m}_h) \|\mathbf{m} - \Pi\mathbf{m}\|_{0,s}^2\right) \\ &\le c_4(1 + \Psi^2(\mathbf{m}, \mathbf{m}_h)) \left[\|\mathbf{m} - \Pi\mathbf{m}\|_{0,2}^2 + \|\mathbf{m} - \Pi\mathbf{m}\|_{0,s}^2\right], \end{aligned} \tag{4.14}$$

where $c_4 = 2c_3^2 c_0^{-1}$.

If we omit the non-negative term $\Phi^2(\mathbf{m}, \mathbf{m}_h)$ and apply (3.11) along with (4.4), then (4.9) immediately follows.

By means of (4.5) and the Minkowski inequality, we find that

$$\begin{aligned} b(\mathbf{v}, \pi\rho - \rho_h) &= a(\mathbf{m}, \mathbf{v}) - a(\mathbf{m}_h, \mathbf{v}) \\ &\le \|\mathbf{m} - \mathbf{m}_h\|_{0,2} \|\mathbf{v}\|_{0,2} + \Phi(\mathbf{m}, \mathbf{m}_h)\, \Psi(\mathbf{m}, \mathbf{m}_h) \|\mathbf{v}\|_{0,s} \\ &\le (\|\mathbf{m} - \mathbf{m}_h\|_{0,2} + \Phi(\mathbf{m}, \mathbf{m}_h)\, \Psi(\mathbf{m}, \mathbf{m}_h))\ \|\mathbf{v}\|_{V_h} \\ &\le \left(1 + \Psi^2(\mathbf{m}, \mathbf{m}_h)\right)^{1/2} \left[\|\mathbf{m} - \mathbf{m}_h\|_{0,2}^2 + \Phi^2(\mathbf{m}, \mathbf{m}_h)\right]^{1/2} \|\mathbf{v}\|_{V_h}. \end{aligned}$$

Applying (4.14) to the term $\|\mathbf{m} - \mathbf{m}_h\|_{0,2}^2 + \Phi^2(\mathbf{m}, \mathbf{m}_h)$ results in

$$\begin{aligned} b(\mathbf{v}, \pi\rho - \rho_h) &\le c_4^{1/2}(1 + \Psi^2(\mathbf{m}, \mathbf{m}_h)) \left[\|\mathbf{m} - \Pi\mathbf{m}\|_{0,2}^2 + \|\mathbf{m} - \Pi\mathbf{m}\|_{0,s}^2\right]^{1/2} \|\mathbf{v}\|_{V_h} \\ &\le c_4^{1/2}(1 + \Psi^2(\mathbf{m}, \mathbf{m}_h)) \left[\|\mathbf{m} - \Pi\mathbf{m}\|_{0,2} + \|\mathbf{m} - \Pi\mathbf{m}\|_{0,s}\right] \|\mathbf{v}\|_{V_h}. \end{aligned} \tag{4.15}$$

We have

$$\|\pi\rho - \rho_h\|_{0,r} \le C_* \sup_{\mathbf{v}\in V_h} \frac{b(\mathbf{v}, \pi\rho - \rho_h)}{\|\mathbf{v}\|_{V_h}} \le C(1 + \Psi^2(\mathbf{m}, \mathbf{m}_h)) \left[\|\mathbf{m} - \Pi\mathbf{m}\|_{0,2} + \|\mathbf{m} - \Pi\mathbf{m}\|_{0,s}\right].$$

By invoking the triangle inequality, $\|\rho - \rho_h\|_{0,r} \le \|\rho - \pi\rho\|_{0,r} + \|\pi\rho - \rho_h\|_{0,r}$, together with (3.11) and (4.4), the inequality (4.10) readily follows.

The proof is concluded. □

Theorem 4.4. *Let* $(\mathbf{m}, \rho) \in V \times Q$ *be the solution of* (3.2) *and* $(\mathbf{m}_h, \rho_h) \in V_h \times Q_h$ *be the solution of* (4.3). *If* $(\mathbf{m}, \rho) \in V \cap (W^{p,s}(\Omega))^n \times W^{p,r}(\Omega)$, *then there exist positive constants* C *independent of* h *such that*

$$\|\mathbf{m} - \mathbf{m}_h\|_{0,2}^2 + \|\mathbf{m} - \mathbf{m}_h\|_{0,s}^s \le Ch^{2p} \quad 1 \le p \le k+1. \tag{4.16}$$

$$\|\rho - \rho_h\|_{0,r} \le Ch^p \quad 1 \le p \le k+1. \tag{4.17}$$

Proof. The estimates (4.16) and (4.17) result from substituting the interpolation error inequalities (4.1) and (4.2) into the inequalities (4.9) and (4.10) of Theorem 4.3. □

5. Numerical results

In this section, we conduct the numerical experiments using the lowest order Raviart-Thomas finite element to solve problem (4.3) in a two-dimensional region. We test several examples using triangular elements in a two-dimension domain to verify the rates of convergence. For simplicity, we use the unit square $\Omega = [0, 1]^2$ as the example domain. The simulations are performed using FEniCS Logg et al. (2012). The unit square is divided into a $\mathcal{N} \times \mathcal{N}$ mesh of squares, each split into two right triangles via the UnitSquareMesh class in FEniCS, ensuring uniform triangulation in each dimension.

To test the convergence of our method, we consider the Forchheimer two-term law given by $F(|\mathbf{m}|)\mathbf{m} = \mathbf{m} + |\mathbf{m}|\mathbf{m}$, corresponding to the parameters $s = 3$ and $r = 3/2$. The generalized Forchheimer equation is solved numerically on each mesh using Newton's method, with a nonlinear solver tolerance of 10^{-6}. We evaluate the L^r-error for the density along with both the L^2-error and L^s-errors for the momentum based on the regularity of the analytical solutions. The convergence rates are computed using the formula $r = \frac{\ln(e_i)-\ln(e_{i-1})}{\ln(h_i)-\ln(h_{i-1})}$ across eight levels of mesh refinement with the discretization parameters $h = 1/4, 1/8, 1/16, 1/32, 1/64, 1/128, 1/256, 1/512$.

Example 1. (Forchheimer without source) We choose the analytical solution

$$\rho(\mathbf{x}) = x_1 - \sqrt{3}x_2 \quad \text{and} \quad \mathbf{m}(\mathbf{x}) = \frac{-2\nabla\rho}{1+\sqrt{1+4|\nabla\rho|}} = \left(-\frac{1}{2}, \frac{\sqrt{3}}{2}\right)^T \quad \forall \mathbf{x} \in [0,1]^2.$$

The forcing term f is determined from equation $\operatorname{div}\mathbf{m} = f$. Explicitly, $f(\mathbf{x}) = 0$. The boundary condition is determined according to the analytical solution as follows:

$$\psi(\mathbf{x}) = \begin{cases} x_1 & \text{if } x_2 = 0 \\ x_1 - \sqrt{3} & \text{if } x_2 = 1 \\ -\sqrt{3}x_2 & \text{if } x_1 = 0 \\ 1 - \sqrt{3}x_2 & \text{if } x_1 = 1 \end{cases}.$$

The numerical results are listed in Table 1.

$\mathcal{N}$	$\|\rho - \rho_h\|_{0,r}$	Rates	$\|\mathbf{m} - \mathbf{m}_h\|_{0,2}$	Rates	$\|\mathbf{m} - \mathbf{m}_h\|_{0,s}$	Rates
4	$6.390e-02$	–	$8.200e-02$	–	$9.434e-02$	
8	$4.272e-02$	0.453	$5.336e-02$	0.620	$7.150e-02$	0.400
16	$2.817e-02$	0.580	$3.262e-02$	0.710	$4.948e-02$	0.531
32	$1.612e-02$	0.601	$1.771e-02$	0.881	$3.363e-02$	0.557
64	$8.627e-03$	0.805	$9.296e-03$	0.930	$2.129e-02$	0.660
128	$4.463e-03$	0.902	$4.795e-03$	0.955	$1.356e-02$	0.650
256	$2.243e-03$	0.951	$2.463e-03$	0.961	$8.561e-03$	0.664
512	$2.243e-03$	0.992	$1.261e-03$	0.966	$5.414e-03$	0.661

Table 1. *The convergence study for Forchheimer flows using the mixed finite elements in 2D.*

Example 2. (Forchheimer with the source, zero boundary data) In this example, the exact (analytical) solution is given by

$$\rho(\mathbf{x}) = \sin(\pi x_1)\sin(\pi x_2) \quad \text{and} \quad \mathbf{m}(\mathbf{x}) = \frac{-2\nabla\rho}{1+\sqrt{1+4|\nabla\rho|}} \quad \forall \mathbf{x} \in [0,1]^2,$$

where $\nabla\rho = (\pi\cos(\pi x_1)\sin(\pi x_2), \pi\sin(\pi x_1)\cos(\pi x_2))^T$, and

$$|\nabla\rho| = \pi\sqrt{(\cos(\pi x_1)\sin(\pi x_2))^2 + (\sin(\pi x_1)\cos(\pi x_2))^2}.$$

The forcing term $f(\mathbf{x})$ and the boundary condition $\psi(\mathbf{x})$ are as follows

$$f(\mathbf{x}) = \frac{4\pi^2 \sin(\pi x_1)\sin(\pi x_2)}{1+\sqrt{1+4|\nabla\rho|}} + \frac{2\pi^2 \sin 2\pi(x_1+x_2)}{(1+\sqrt{1+4|\nabla\rho|})^2\sqrt{(1+4|\nabla\rho|}\,|\nabla\rho|}, \quad \psi(\mathbf{x}) = 0.$$

The numerical results are listed in Table 2.

$\mathcal{N}$	$\|\rho-\rho_h\|_{0,r}$	Rates	$\|\mathbf{m}-\mathbf{m}_h\|_{0,2}$	Rates	$\|\mathbf{m}-\mathbf{m}_h\|_{0,s}$	Rates
4	$1.252e-02$	–	$4.420e-02$	–	$6.400e-02$	
8	$8.416e-03$	0.562	$3.102e-02$	0.511	$4.682e-02$	0.451
16	$4.932e-03$	0.573	$2.160e-02$	0.522	$3.189e-02$	0.554
32	$2.831e-03$	0.771	$1.389e-02$	0.637	$2.168e-02$	0.557
64	$1.533e-03$	0.801	$8.340e-03$	0.736	$1.474e-02$	0.556
128	$8.073e-04$	0.885	$4.685e-03$	0.832	$9.918e-03$	0.572
256	$4.009e-04$	0.925	$2.509e-03$	0.901	$6.575e-03$	0.593
512	$4.009e-04$	1.010	$1.284e-03$	0.966	$4.332e-03$	0.602

Table 2. *The convergence study for Forchheimer flows using the mixed finite elements in 2D.*

Example 3. (Forchheimer with the source, nonzero boundary data) The analytical solution in this example is

$$\rho(\mathbf{x}) = w^2(\mathbf{x}) \quad \text{and} \quad \mathbf{m}(\mathbf{x}) = -\frac{4(x_1,x_2)^T}{1+\sqrt{1+8w(\mathbf{x})}} \quad \forall \mathbf{x}\in[0,1]^2,$$

where $w(\mathbf{x}) = \sqrt{x_1^2+x_2^2}$. The forcing term f and the boundary condition $\psi(\mathbf{x})$ are as follows

$$f(\mathbf{x}) = -\frac{8(1+6w(\mathbf{x})+\sqrt{1+8w(\mathbf{x})})}{\sqrt{1+8w(\mathbf{x})}(1+\sqrt{1+8w(\mathbf{x})})^2}, \quad \psi(\mathbf{x}) = \begin{cases} x_2^2 & \text{on } x_1=0,\\ 1+x_2^2 & \text{on } x_1=1,\\ x_1^2 & \text{on } x_2=0,\\ x_1^2+1 & \text{on } x_2=1 \end{cases}.$$

The numerical results are listed in Table 3.

$\mathcal{N}$	$\|\rho-\rho_h\|_{0,r}$	Rates	$\|\mathbf{m}-\mathbf{m}_h\|_{0,2}$	Rates	$\|\mathbf{m}-\mathbf{m}_h\|_{0,s}$	Rates
4	$1.520e-02$	–	$4.210e-01$	–	$6.140e-01$	
8	$9.520e-03$	0.621	$2.854e-01$	0.561	$4.309e-01$	0.511
16	$5.732e-03$	0.675	$1.920e-01$	0.572	$2.976e-01$	0.534
32	$3.195e-03$	0.732	$1.234e-01$	0.637	$2.028e-01$	0.553
64	$1.754e-03$	0.843	$7.417e-02$	0.735	$1.367e-01$	0.569
128	$9.266e-04$	0.865	$4.257e-02$	0.801	$8.946e-02$	0.612
256	$4.820e-04$	0.921	$2.376e-02$	0.841	$5.756e-02$	0.636
512	$4.820e-04$	0.943	$1.251e-02$	0.926	$3.663e-02$	0.652

Table 3. *The convergence study for Forchheimer flows using the mixed finite elements in 2D.*

Acknowledgments. The author sincerely thanks the references for their careful reading of this article and for their helpful comments and valuable suggestions which improve the manuscript.

Appendix A.

Lemma A.1. *For any* $s \in (1, \infty)$, $\mathbf{W}_s(\mathrm{div}, \Omega)$ *is a reflexive Banach space.*

Proof. Firstly, one can verify that $\mathbf{W}_s(\mathrm{div}, \Omega)$ is a Banach space. Similarly in Knabner and Summ (2016), we use the mapping $E\mathbf{v} = (\mathbf{v}, \mathrm{div}\,\mathbf{v})$ for $\mathbf{v} \in \mathbf{W}_s(\mathrm{div}, \Omega)$ to embed $\mathbf{W}_s(\mathrm{div}, \Omega)$ into $(L^s(\Omega))^{n+1}$. Also, denote $\widetilde{\mathbf{W}}_s = E(\mathbf{W}_s(\mathrm{div}, \Omega)) \subset (\mathrm{L^s}(\Omega))^{\mathrm{n+1}}$. Then, the norm $\|E\mathbf{v}\|_{(L^s(\Omega))^{n+1}}$ in $\widetilde{\mathbf{W}}_s$ is equivalent to the norm $\|\mathbf{v}\|_V$. This way, we identify $\mathbf{W}_s(\mathrm{div}, \Omega)$ as $\widetilde{\mathbf{W}}_s$ and vice versa. As a consequence, $\widetilde{\mathbf{W}}_s$ is a closed subspace of $(L^s(\Omega))^{n+1}$, hence it is a reflexive Banach space.

For dual and double dual spaces, we identify $F \in \mathbf{W}_s(\mathrm{div}, \Omega)'$ as $\widetilde{F} = F \circ E^{-1} \in \widetilde{\mathbf{W}}'_s$, and we identify $G \in \mathbf{W}_s(\mathrm{div}, \Omega)''$ as $\widetilde{G} \in \widetilde{\mathbf{W}}''_s$ defined by $\widetilde{G}(\widetilde{F}) = G(\widetilde{F} \circ E)$ for any $\widetilde{F} \in \widetilde{\mathbf{W}}'_s$. Then, $\widetilde{\mathbf{W}}_s$ being reflexive implies that $\mathbf{W}_s(\mathrm{div}, \Omega)$ is a reflexive Banach space. □

References

Arbogast, T., Wheeler, M. F., and Yotov, I. (1997). Mixed finite elements for elliptic problems with tensor coefficients as cell-centered finite differences. *SIAM Journal on Numerical Analysis*, 34(2):828–852.

Aronson, D. G. (1986). The porous medium equation. In *Nonlinear diffusion problems (Montecatini Terme, 1985)*, volume 1224 of *Lecture Notes in Math.*, pages 1–46. Springer, Berlin.

Aulisa, E., Bloshanskaya, L., Hoang, L., and Ibragimov, A. (2009). Analysis of generalized Forchheimer flows of compressible fluids in porous media. *J. Math. Phys.*, 50(10):103102, 44.

Baranger, J. and Najib, K. (1990). Numerical analysis of quasi-Newtonian flow obeying the power low or the carreau flow. *Numerische Mathematik*, 58(1):35–49.

Bear, J. (1972). *Dynamics of Fluids in Porous Media*. Dover, New York.

Bramble, J. H., Pasciak, J. E., and Steinbach, O. (2002). On the stability of the l2 projection in h1(ω). *Mathematics of Computation*, 71(237):pp. 147–156.

Brézis, H. (1973). *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. North-Holland Publishing Co., Amsterdam. North-Holland Mathematics Studies, No. 5. Notas de Matemática (50).

Brezzi, F. and Fortin, M. (1991). *Mixed and hybrid finite element methods*, volume 15 of *Springer Series in Computational Mathematics*. Springer-Verlag, New York.

Celik, E. and Hoang, L. (2016). Generalized Forchheimer flows in heterogeneous porous media. *Nonlinearity*, 29(3):1124–1155.

Celik, E. and Hoang, L. (2017). Maximum estimates for generalized Forchheimer flows in heterogeneous porous media. *J. Differential Equations*, 262(3):2158–2195.

Celik, E., Hoang, L., Ibragimov, A., and Kieu, T. (2017). Fluid flows of mixed regimes in porous media. *Journal of Mathematical Physics*, 58(2):023102.

Celik, E., Hoang, L., and Kieu, T. (2018). Generalized Forchheimer flows of isentropic gases. *Journal of Mathematical Fluid Mechanics*, 20(1):83–115.

Celik, E., Hoang, L., and Kieu, T. (2023). Studying a doubly nonlinear model of slightly compressible Forchheimer flows in rotating porous media. *Turkish J. Math.*, 47(3):949–987.

Chadam, J. and Qin, Y. (1997). Spatial decay estimates for flow in a porous medium. *SIAM J. Math. Anal.*, 28(4):808–830.

Chow, S. S. (1989). Finite element error estimates for non-linear elliptic equations of monotone type. *Numerische Mathematik*, 54(4):373–393.

Ciarlet, P. G. (1978). *The finite element method for elliptic problems*. North-Holland Publishing Co., Amsterdam. Studies in Mathematics and its Applications, Vol. 4.

Douglas, J. J., Paes-Leme, P. J., and Giorgi, T. (1993). Generalized Forchheimer flow in porous media. In *Boundary value problems for partial differential equations and applications*, volume 29 of *RMA Res. Notes Appl. Math.*, pages 99–111. Masson, Paris.

Durán, R. G. (1988). Error analysis in $l^p \leqslant p \leqslant \infty$, for mixed finite element methods for linear and quasi-linear elliptic problems. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique*, 22(3):371–387.

Fabrie, P. (1989). Regularity of the solution of Darcy-Forchheimer's equation. *Nonlinear Anal.*, 13(9):1025–1049.

Forchheimer, P. (1901). *Wasserbewegung durch Boden Zeit*, volume 45. Ver. Deut. Ing.

Franchi, F. and Straughan, B. (2003). Continuous dependence and decay for the Forchheimer equations. *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.*, 459(2040):3195–3202.

Glowinski, R. and Marroco, A. (1975). Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. *Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique*, 9(R2):41–76.

Hoang, L. and Ibragimov, A. (2012). Qualitative study of generalized Forchheimer flows with the flux boundary condition. *Adv. Differential Equations*, 17(5–6):511–556.

Hoang, L., Ibragimov, A., Kieu, T., and Sobol, Z. (2015). Stability of solutions to generalized Forchheimer equations of any degree. *J. Math. Sci. (N.Y.)*, 210(4, Problems in mathematical analysis. No. 81 (Russian)):476–544.

Hoang, L. and Kieu, T. (2019). Global estimates for generalized Forchheimer flows of slightly compressible fluids. *J. Anal. Math.*, 137(1):1–55.

Hoang, L. T., Kieu, T. T., and Phan, T. V. (2014). Properties of generalized Forchheimer flows in porous media. *J. Math. Sci.*, 202(2):259–332.

Johnson, C. and Thomée, V. (1981). Error estimates for some mixed finite element methods for parabolic type problems. *RAIRO Anal. Numér.*, 15(1):41–78.

Kieu, T. (2016). Analysis of expanded mixed finite element methods for the generalized Forchheimer flows of slightly compressible fluids. *Numer. Methods Partial Differential Equations*, 32(1):60–85.

Kieu, T. (2020). Existence of a solution for generalized Forchheimer flow in porous media with minimal regularity conditions. *Journal of Mathematical Physics*, 61(1):013507.

Knabner, P. and Summ, G. (2016). Solvability of the mixed formulation for Darcy-Forchheimer flow in porous media. pages 1–26. preprint. DOI:10.48550/arXiv.1608.08829.

Lee, M. and Milner, F. A. (1997). Mixed finite element methods for nonlinear elliptic problems: the h-p version. *Journal of Computational and Applied Mathematics*, 85(2):239–261.

Lions, J.-L. (1969). *Quelques méthodes de résolution des problèmes aux limites non linéaires*. Dunod.

Logg, A., Mardal, K.-A., and Wells, G. N., editors (2012). *Automated Solution of Differential Equations by the Finite Element Method*, volume 84 of *Lecture Notes in Computational Science and Engineering*. Springer.

Milner, F. A. and Suri, M. (1992). Mixed finite element methods for quasilinear second order elliptic problems : the p-version. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique*, 26(7):913–931.

Muskat, M. (1937). *The flow of homogeneous fluids through porous media*. McGraw-Hill Book Company, inc.

Park, E.-J. (2005). Mixed finite element methods for generalized Forchheimer flow in porous media. *Numer. Methods Partial Differential Equations*, 21(2):213–228.

Payne, L. E. and Song, J. C. (2000). Spatial decay for a model of double diffusive convection in Darcy and Brinkman flows. *Z. Angew. Math. Phys.*, 51(6):867–889.

Payne, L. E. and Straughan, B. (1996). Stability in the initial-time geometry problem for the Brinkman and Darcy equations of flow in porous media. *J. Math. Pures Appl. (9)*, 75(3):225–271.

Payne, L. E. and Straughan, B. (1999). Convergence and continuous dependence for the Brinkman-Forchheimer equations. *Stud. Appl. Math.*, 102(4):419–439.

Raviart, P. A. and Thomas, J. M. (1977). A mixed finite element method for 2-nd order elliptic problems. In Galligani, I. and Magenes, E., editors, *Mathematical Aspects of Finite Element Methods*, pages 292–315, Berlin, Heidelberg. Springer Berlin Heidelberg.

Sandri, D. (1993). Sur lápproximation numérique des écoulements quasi-newtoniens dont la viscosité suit la loi puissance ou la loi de carreau. *ESAIM: M2AN*, 27(2):131–155.

Showalter, R. E. (1997). *Monotone operators in Banach space and nonlinear partial differential equations*, volume 49 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI.

Straughan, B. (2008). *Stability and wave motion in porous media*, volume 165 of *Applied Mathematical Sciences*. Springer, New York.

Vázquez, J. L. (2007). *The porous medium equation*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, Oxford. Mathematical theory.

Ward, J. C. (1964). Turbulent flow in porous media. *Journal of the Hydraulics Division, Proc. Am. Soc. Civ. Eng.*, 90(HY5):1–12.

Zeidler, E. (1990). *Nonlinear functional analysis and its applications. II/B*. Springer-Verlag, New York. Nonlinear monotone operators, Translated from the German by the author and Leo F. Boron.

Zeidler, E. and Boron, L. (1989). *Nonlinear Functional Analysis and Its Applications: Part 2 B: Nonlinear Monotone Operators*. Nonlinear Functional Analysis and Its Applications. Springer New York.

(A. Yadav) SOUTH FORSYTH HIGH SCHOOL, 585 PEACHTREE PARKWAYCUMMING, GA 30041, U.S.A.
Email address: ayatistar17@gmail.com

(T. Park) NORTH GWINNETT HIGH SCHOOL, 20 LEVEL CREEK RD, SUWANEE, GA 30024
Email address: teddybearbs101@gmail.com

(T. Selvaganesan) LAMBERT HIGH SCHOOL, 805 NICHOLS RD, SUWANEE, GA 30024
Email address: selvaganesan.tessica@gmail.com

(T. Kieu) DEPARTMENT OF MATHEMATICS, UNIVERSITY OF NORTH GEORGIA, GAINESVILLE CAMPUS, 3820 MUNDY MILL RD., OAKWOOD, GA 30566, U.S.A.
Email address, Corresponding author: thinh.kieu@ung.edu

(R. Ye) SOUTH FORSYTH HIGH SCHOOL, 585 PEACHTREE PARKWAYCUMMING, GA 30041, U.S.A.
Email address: 11556ye@gmail.com