Mathematical Models and Methods in Applied Sciences
Vol. 26, No. 5 (2016) 867–900
© World Scientific Publishing Company
DOI: 10.1142/S0218202516500202

A mixed-primal finite element approximation of a sedimentation–consolidation system

Mario Alvarez*
Sección de Matemática, Sede de Occidente,
Universidad de Costa Rica,
San Ramón de Alajuela, Costa Rica
mguadamuz@ci2ma.udec.cl

Gabriel N. Gatica
CI²MA and Departamento de Ingeniería Matemática,
Universidad de Concepción,
Casilla 160-C, Concepción, Chile
ggatica@ci2ma.udec.cl

Ricardo Ruiz-Baier†
Mathematical Institute, Oxford University,
Andrew Wiles Building, Woodstock Road,
OX2 6GG, Oxford, UK
ruizbaier@maths.ox.ac.uk

Received 20 March 2015
Revised 1 October 2015
Accepted 4 November 2015
Published 3 February 2016
Communicated by A. Quarteroni

This paper is devoted to the mathematical and numerical analysis of a strongly coupled flow and transport system typically encountered in continuum-based models of sedimentation–consolidation processes. The model focuses on the steady-state regime of a solid–liquid suspension immersed in a viscous fluid within a permeable medium, and the governing equations consist in the Brinkman problem with variable viscosity, written in terms of Cauchy pseudo-stresses and bulk velocity of the mixture; coupled with a nonlinear advection — nonlinear diffusion equation describing the transport of the solids volume fraction. The variational formulation is based on an augmented mixed approach for the Brinkman problem and the usual primal weak form for the transport equation. Solvability of the coupled formulation is established by combining fixed point arguments, certain regularity assumptions, and some classical results concerning variational problems and Sobolev spaces. In turn, the resulting augmented mixed-primal Galerkin scheme employs Raviart–Thomas approximations of order k for the stress and

*Present address: CI²MA and Departamento de Ingeniería Matemática, Universidad de Concepción, Casilla 160-C, Concepción, Chile.
†Corresponding author

piecewise continuous polynomials of order $k+1$ for velocity and volume fraction, and its solvability is deduced by applying a fixed-point strategy as well. Then, suitable Strang-type inequalities are utilized to rigorously derive optimal error estimates in the natural norms. Finally, a few numerical tests illustrate the accuracy of the augmented mixed-primal finite element method, and the properties of the model.

Keywords: Brinkman equations; nonlinear transport problem; augmented mixed-primal formulation; fixed point theory; sedimentation–consolidation process; finite element methods; *a priori* error analysis.

AMS Subject Classification: 65N30, 65N12, 76R05, 76D07, 65N15

1. Introduction

The interaction of solid–liquid suspensions is often encountered in a wide variety of natural and engineering applications, including fluidized beds, clot formation within the blood, solid–liquid separation and purification in wastewater treatment, macroscopic biofilm characterization, and many others. In sedimentation–consolidation processes, suspended solid particles settle down due to gravity acceleration and can be subsequently removed from the fluid. Here we are interested in a continuum-based framework where the viscous fluid is incompressible and the flow patterns are laminar so the mass and momentum balances are governed by the Brinkman equations with variable viscosity, and the mass balance of the solid phase (here allowed to sediment into the fluid phase) is described by a nonlinear advection–diffusion equation. A number of difficulties are associated to the understanding and prediction of the behavior of such a problem, including highly nonlinear (and typically degenerate) advection and diffusion terms, strong interaction of velocity and solids volume fraction via the Cauchy stress tensor and the forcing term, nonlinear structure of the overall coupled flow and transport problem, saddle-point structure of the flow problem, non-homogeneous and mixed boundary conditions, and so on. These complications are usually reflected, not only in the solvability analysis of the governing equations, but also in the construction of appropriate schemes for their numerical approximation, and in the derivation of stability results and error bounds.

The solvability of the sedimentation–consolidation problem has been previously discussed in Ref. 8 for the case of large fluid viscosity, using the technique of parabolic regularization. Moreover, a modified formulation based on Stokes flow has been recently studied in Ref. 2, where the solution of the transport equation required an explicit dependence of the effective diffusivity on the gradient of the concentration. With a similar restriction (the viscosity depending on the concentration and on the velocity gradient), the existence of solutions to a model of chemically reacting non-Newtonian fluid has been established in Ref. 6. In contrast, here these hypotheses have been modified, enlarging the applicability of the present results, in particular to classical models of sedimentation of suspensions. More specifically, we assume both the viscosity and the diffusivity to depend only on the scalar value of the concentration. Nevertheless, we still remain in the framework of non-degeneracy of the diffusion term. On the other hand, it is worth mentioning that models of sedimentation–consolidation share some structural similarities with Boussinesq- and Oldroyd-type models, for which several mixed formulations have been proposed.[13,14,16–18,26] In particular, the mixed finite element method for the Boussinesq problem developed in Ref. 17 is based on the introduction of the gradient of velocity as an auxiliary unknown, and the utilization of refined meshes near the singular corners and suitable finite element subspaces. In turn, the approach from Ref. 13 first introduces the same nonlinear pseudostress tensor linking the pseudostress and the convective term that has been employed before in Ref. 10 for the Navier–Stokes problem, and then augment the resulting mixed formulation of the stationary Boussinesq problem with Galerkin type terms arising from the constitutive and equilibrium equations, and the Dirichlet boundary condition. In this way, the Banach and Brouwer fixed point theorems, together with the Lax–Milgram lemma and the Babuška–Brezzi theory are applied to conclude the well-posedness of the continuous and discrete formulations. Nevertheless, up to our knowledge, mixed formulations specifically tailored for the study of sedimentation processes are not yet available from the literature.

According to the above bibliographic discussion, our present purpose is to examine mixed finite element approximations of the model problem, where also the Cauchy stress enters in the formulation as an additional unknown. Given the arrangement of the equations and the implicit smoothness requirements of the fluid velocity and its discrete approximation, we realize as in Ref. 2 that applying an augmentation strategy to the Brinkman problem simplifies the treatment of both the continuous and Galerkin schemes. More precisely, we propose an augmented variational formulation where stresses are sought in $\mathbb{H}(\mathbf{div};\Omega)$, the velocity is in $\mathbf{H}^1(\Omega)$, and the solids volume fraction has $\mathrm{H}^1(\Omega)$ regularity. Consequently, the rows of the Cauchy stress tensor will be approximated with Raviart–Thomas elements of order k, whereas the velocity and solids concentration will be discretized with continuous piecewise polynomials of degree $\leq k+1$. The solvability analysis of the continuous formulation is based on a strategy combining classical fixed-point arguments, suitable regularity assumptions on the decoupled problems, the Lax–Milgram lemma, and the Sobolev embedding and Rellich–Kondrachov compactness theorems. In addition, and provided that the data are sufficiently small, we also establish uniqueness of weak solution. On the other hand, well-posedness of the discrete problem relies on the Brouwer fixed point theorem and analogous arguments to those employed in the continuous analysis. Finally, applying a suitable Strang-type lemma valid for linear problems to the fluid flow equations, and explicitly deriving our own Strang-type estimate for the transport equations, we are also able to derive the corresponding Céa estimate, and to provide optimal *a priori* error bounds for the Galerkin solution.

The rest of the paper is organized as follows. Section 2 compiles some preliminary notation and outlines the boundary value problem of interest, which is rewritten by eliminating the pressure unknown from the system. In Sec. 3, we introduce the corresponding variational formulation following an augmented mixed approach for the Brinkman equations, coupled with a primal method for the transport problem. The associated Galerkin scheme is introduced in Sec. 4, followed by the development of its solvability analysis. In Sec. 5, we proceed with the study of accuracy of the augmented formulation, establishing optimal error bounds; and we close in Sec. 6 with some numerical examples illustrating the good performance of the mixed-primal method and confirming the predicted rates of convergence.

2. The Model Problem

2.1. *Preliminaries*

Let $\Omega \subseteq \mathrm{R}^n$, $n = 2, 3$ denote a bounded domain with polyhedral boundary $\Gamma = \bar{\Gamma}_{\mathrm{D}} \cup \bar{\Gamma}_{\mathrm{N}}$, with $\Gamma_{\mathrm{D}} \cap \Gamma_{\mathrm{N}} = \emptyset$, and denote by $\boldsymbol{\nu}$ the outward unit normal vector on Γ. We recall the standard notation for Lebesgue spaces $\mathrm{L}^p(\Omega)$ and Sobolev spaces $\mathrm{H}^s(\Omega)$ endowed with the norm $\|\cdot\|_{s,\Omega}$ and seminorm $|\cdot|_{s,\Omega}$. In particular, $\mathrm{H}^{1/2}(\Gamma)$ stands for the space of traces of functions of $\mathrm{H}^1(\Omega)$ and $\mathrm{H}^{-1/2}(\Gamma)$ denotes its dual. By $\mathbf{M}$ and $\mathbb{M}$ we will denote the corresponding vectorial and tensorial counterparts of the generic scalar functional space M, and $\|\cdot\|$, with no subscripts, will stand for the natural norm of either an element or an operator in any product functional space. In turn, for any vector field $\boldsymbol{v} = (v_i)_{i=1,n}$ we set the gradient and divergence operators as

$$\nabla \boldsymbol{v} := \left(\frac{\partial v_i}{\partial x_j}\right)_{i,j=1,n} \quad \text{and} \quad \operatorname{div} \boldsymbol{v} := \sum_{j=1}^{n} \frac{\partial v_j}{\partial x_j}.$$

In addition, for any tensor fields $\boldsymbol{\tau} = (\tau_{ij})_{i,j=1,n}$ and $\boldsymbol{\zeta} = (\zeta_{ij})_{i,j=1,n}$, we let $\mathbf{div}\,\boldsymbol{\tau}$ be the divergence operator div acting along the rows of $\boldsymbol{\tau}$, and define the transpose, the trace, the tensor inner product, and the deviatoric tensor, respectively, as

$$\boldsymbol{\tau}^{\mathtt{t}} := (\tau_{ji})_{i,j=1,n}, \quad \operatorname{tr}(\boldsymbol{\tau}) := \sum_{i=1}^{n} \tau_{ii},$$

$$\boldsymbol{\tau} : \boldsymbol{\zeta} := \sum_{i,j=1}^{n} \tau_{ij}\zeta_{ij} \quad \text{and} \quad \boldsymbol{\tau}^{\mathtt{d}} := \boldsymbol{\tau} - \frac{1}{n}\operatorname{tr}(\boldsymbol{\tau})\mathbb{I}.$$

Furthermore, we recall that

$$\mathbb{H}(\mathbf{div};\Omega) := \{\boldsymbol{\tau} \in \mathbb{L}^2(\Omega) : \mathbf{div}\,\boldsymbol{\tau} \in \mathbf{L}^2(\Omega)\},$$

equipped with the usual norm

$$\|\boldsymbol{\tau}\|^2_{\mathbf{div};\Omega} := \|\boldsymbol{\tau}\|^2_{0,\Omega} + \|\mathbf{div}\,\boldsymbol{\tau}\|^2_{0,\Omega}$$

is a standard Hilbert space in the realm of mixed problems. Finally, in what follows $\mathbb{I}$ stands for the identity tensor in $\mathbb{R} := \mathrm{R}^{n\times n}$, and $|\cdot|$ denotes the Euclidean norm in $\mathbf{R} := \mathrm{R}^n$.

2.2. *The sedimentation–consolidation system*

We consider the steady state of the sedimentation–consolidation process consisting on the transport and suspension of a solid phase into an immiscible fluid contained in a vessel Ω. The flow patterns are influenced by gravity and by the local fluctuations of the solids volume fraction. The process is governed by the following system of partial differential equations:

$$\begin{gathered}\boldsymbol{\sigma} = \mu(\phi)\nabla \boldsymbol{u} - p\mathbb{I}, \quad \boldsymbol{K}^{-1}\boldsymbol{u} - \mathbf{div}\,\boldsymbol{\sigma} = \boldsymbol{f}\phi, \quad \operatorname{div}\boldsymbol{u} = 0, \\ \tilde{\boldsymbol{\sigma}} = \vartheta(\phi)\nabla\phi - \phi\boldsymbol{u} - f_{\mathrm{bk}}(\phi)\boldsymbol{k}, \quad \beta\phi - \operatorname{div}\tilde{\boldsymbol{\sigma}} = g.\end{gathered} \tag{2.1}$$

The sought quantities are the Cauchy fluid pseudo-stress $\boldsymbol{\sigma}$, the average velocity of the mixture $\boldsymbol{u}$, the fluid pressure p and the volumetric fraction of the solids (in short, concentration) ϕ. In this model we also assume that the vessel initially contains an array of fixed-concentration particles (see the discussion in Ref. 3). In this context, the parameter β is a positive constant representing the porosity of the medium, and the permeability tensor $\boldsymbol{K} \in \mathbb{C}(\bar{\Omega}) := [C(\bar{\Omega})]^{n\times n}$ and its inverse are symmetric and uniformly positive definite, which means that there exists $\alpha_{\boldsymbol{K}} > 0$ such that

$$\boldsymbol{v}^{\mathrm{t}}\boldsymbol{K}^{-1}(\boldsymbol{x})\boldsymbol{v} \geq \alpha_{\boldsymbol{K}}|\boldsymbol{v}|^2, \quad \forall\,\boldsymbol{v} \in \mathrm{R}^n, \ \forall\,\boldsymbol{x} \in \Omega. \tag{2.2}$$

Here $\boldsymbol{k}$ is a constant vector pointing in the direction of gravity, and we assume that the kinematic effective viscosity, μ; the one-directional Kynch batch flux density function describing hindered settling, f_{bk}; and the diffusion or sediment compressibility, ϑ; are nonlinear scalar functions of the concentration ϕ. In particular, we can take

$$\mu(\phi) := \mu_\infty\left(1 - \frac{\phi}{\phi_m}\right)^{-\gamma_\mu}, \quad f_{\mathrm{bk}}(\phi) := f_\infty\left[1 + \phi\left(1 - \frac{\phi}{\phi_m}\right)^{\gamma_f}\right],$$
$$\vartheta(\phi) := \vartheta_\infty\left[\phi + \left(1 - \frac{\phi}{\phi_m}\right)^{-\gamma_\vartheta}\right],$$

where $\mu_\infty, \phi_m, f_\infty, \gamma_\mu, \gamma_f, \gamma_\vartheta, \vartheta_\infty$ are positive model parameters. Notice that f_{bk} and ϑ are regularized versions of the Kynch flux and compressibility functions typically employed in sedimentation models (see e.g. Refs. 7–9). Nevertheless, the subsequent analysis allows for arbitrary concentration-dependent functions, as long as the following properties are satisfied: there exist positive constants μ_1, μ_2, γ_1, γ_2, ϑ_1 and ϑ_2, such that

$$\mu_1 \leq \mu(s) \leq \mu_2, \quad \vartheta_1 \leq \vartheta(s) \leq \vartheta_2 \quad \text{and} \quad \gamma_1 \leq f_{\mathrm{bk}}(s) \leq \gamma_2 \quad \forall\, s \in \mathrm{R}. \tag{2.3}$$

Note that (2.3) guarantees, in particular, that the corresponding Nemytsky operators, say U for μ, defined by $U(\phi)(x) := \mu(\phi(x))\ \forall\,\phi \in \mathrm{L}^2(\Omega)$, $\forall\,x \in \Omega$ a.e., and analogously for ϑ, f_{bk}, μ^{-1}, ϑ^{-1}, and f_{bk}^{-1}, are all well defined and continuous from $\mathrm{L}^2(\Omega)$ into $\mathrm{L}^2(\Omega)$.

The driving force of the mixture also depends on the local fluctuations of the concentration, so the right-hand side of the second equation in (2.1) is linear with respect to ϕ, and $\boldsymbol{f} \in \mathbf{L}^\infty(\Omega)$ and $g \in \mathrm{L}^2(\Omega)$ are given functions. Finally, given $\boldsymbol{u}_\mathrm{D} \in \mathbf{H}^{1/2}(\Gamma_\mathrm{D})$, the following mixed boundary conditions complement (2.1):

$$\boldsymbol{u} = \boldsymbol{u}_\mathrm{D} \quad \text{on } \Gamma_\mathrm{D}, \qquad \boldsymbol{\sigma\nu} = \mathbf{0} \quad \text{on } \Gamma_\mathrm{N},$$

$$\phi = 0 \quad \text{on } \Gamma_\mathrm{D} \quad \text{and} \quad \tilde{\boldsymbol{\sigma}} \cdot \boldsymbol{\nu} = 0 \quad \text{on } \Gamma_\mathrm{N},$$

where we remark that the homogeneous datum for $\boldsymbol{\sigma}$ represents a pseudo-traction boundary condition, since we are employing $\nabla \boldsymbol{u}$ instead of the symmetrized gradient $\mathbf{e}(\boldsymbol{u}) := \frac{1}{2}(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^\mathrm{t})$ in the definition of the stress.

On the other hand, it is easy to see that the first and third equations in (2.1) are equivalent to

$$\boldsymbol{\sigma} = \mu(\phi)\nabla \boldsymbol{u} - p\mathbb{I} \quad \text{and} \quad p + \frac{1}{n}\,\mathrm{tr}(\boldsymbol{\sigma}) = 0 \quad \text{in } \Omega,$$

which permits us to eliminate the pressure p from the first equation. Consequently, we arrive at the following coupled system:

$$\begin{aligned}
&\frac{1}{\mu(\phi)}\boldsymbol{\sigma}^\mathrm{d} = \nabla \boldsymbol{u} \quad \text{in } \Omega, \qquad \boldsymbol{K}^{-1}\boldsymbol{u} - \mathbf{div}\,\boldsymbol{\sigma} = \boldsymbol{f}\phi \quad \text{in } \Omega,\\
&\tilde{\boldsymbol{\sigma}} = \vartheta(\phi)\nabla\phi - \phi\boldsymbol{u} - f_\mathrm{bk}(\phi)\boldsymbol{k} \quad \text{in } \Omega, \qquad \beta\phi - \mathrm{div}\,\tilde{\boldsymbol{\sigma}} = g \quad \text{in } \Omega,\\
&\boldsymbol{u} = \boldsymbol{u}_\mathrm{D} \quad \text{on } \Gamma_\mathrm{D}, \qquad \boldsymbol{\sigma\nu} = \mathbf{0} \quad \text{on } \Gamma_\mathrm{N},\\
&\phi = 0 \quad \text{on } \Gamma_\mathrm{D} \quad \text{and} \quad \tilde{\boldsymbol{\sigma}} \cdot \boldsymbol{\nu} = 0 \quad \text{on } \Gamma_\mathrm{N}.
\end{aligned} \tag{2.4}$$

We stress that the incompressibility constraint is implicitly present in the constitutive equation $(2.4)_1$ relating $\boldsymbol{\sigma}$ and $\boldsymbol{u}$. Systems like (2.1) are well established and have been extensively validated to describe sediment-flow patterns in permeable media (see Refs. 24, 25 and the references therein). Furthermore, if we wanted to deal with a traction boundary condition on Γ_N, then $(2.1)_1$ and $(2.4)_1$ would be replaced simply by $\boldsymbol{\sigma} = \mu(\phi)\mathbf{e}(\boldsymbol{u}) - p\mathbb{I}$ and $\frac{1}{\mu(\phi)}\boldsymbol{\sigma}^\mathrm{d} = \nabla \boldsymbol{u} - \gamma$, respectively, where $\gamma := \frac{1}{2}(\nabla \boldsymbol{u} - \nabla \boldsymbol{u}^\mathrm{t})$ is the additional unknown given by the vorticity. In this case, however, the rest of the corresponding analysis would be very close to the one to be developed in what follows.

3. The Variational Formulation

In this section, we proceed similarly as in Ref. 2 to derive a suitable variational formulation of (2.4) and analyze its corresponding solvability by using a fixed-point strategy.

3.1. *An augmented mixed-primal approach*

Notice that the homogeneous boundary condition for $\boldsymbol{\sigma}$ on Γ_N (cf. $(2.4)_3$) suggests the introduction of the functional space

$$\mathbb{H}_N(\mathbf{div};\Omega) := \{\boldsymbol{\tau} \in \mathbb{H}(\mathbf{div};\Omega) : \boldsymbol{\tau\nu} = \mathbf{0} \text{ on } \Gamma_\mathrm{N}\}.$$

Multiplying the first equation of (2.4) by $\boldsymbol{\tau} \in \mathbb{H}_N(\mathbf{div};\Omega)$, integrating by parts, and using the Dirichlet boundary condition for $\boldsymbol{u}$ (cf. third row of (2.4)), we obtain

$$\int_\Omega \frac{1}{\mu(\phi)}\boldsymbol{\sigma}^{\mathrm{d}} : \boldsymbol{\tau}^{\mathrm{d}} + \int_\Omega \boldsymbol{u}\cdot \mathbf{div}\,\boldsymbol{\tau} = \langle \boldsymbol{\tau\nu}, \boldsymbol{u}_{\mathrm{D}}\rangle_{\Gamma_{\mathrm{D}}}, \quad \forall\, \boldsymbol{\tau} \in \mathbb{H}_N(\mathbf{div};\Omega),$$

where $\langle\cdot,\cdot\rangle_{\Gamma_{\mathrm{D}}}$ is the duality pairing between $\mathbf{H}^{-1/2}(\Gamma_{\mathrm{D}})$ and $\mathbf{H}^{1/2}(\Gamma_{\mathrm{D}})$. Moreover, the momentum balance is then rewritten as

$$-\int_\Omega \boldsymbol{K}^{-1}\boldsymbol{u}\cdot\boldsymbol{v} + \int_\Omega \boldsymbol{v}\cdot\mathbf{div}\,\boldsymbol{\sigma} = -\int_\Omega \boldsymbol{f}\phi\cdot\boldsymbol{v}, \quad \forall\,\boldsymbol{v}\in \mathbf{L}^2(\Omega).$$

On the other hand, the Dirichlet boundary condition for ϕ (cf. fourth row of (2.4)) motivates the introduction of the space

$$\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) := \{\psi \in \mathrm{H}^1(\Omega) : \psi = 0 \text{ on } \Gamma_{\mathrm{D}}\},$$

for which, thanks to the generalized Poincaré inequality, there exists $c_p > 0$, depending only on Ω and Γ_{D}, such that

$$\|\psi\|_{1,\Omega} \le c_p|\psi|_{1,\Omega}, \quad \forall\,\psi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega). \tag{3.1}$$

Therefore, given $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$, we arrive at the following mixed formulation for the Brinkman flow: find $(\boldsymbol{\sigma},\boldsymbol{u}) \in \mathbb{H}_N(\mathbf{div};\Omega)\times\mathbf{L}^2(\Omega)$ such that

$$\begin{aligned} \mathbf{a}_\phi(\boldsymbol{\sigma},\boldsymbol{\tau}) + \mathbf{b}(\boldsymbol{\tau},\boldsymbol{u}) &= \langle \boldsymbol{\tau\nu}, \boldsymbol{u}_{\mathrm{D}}\rangle_{\Gamma_{\mathrm{D}}}, \quad \forall\,\boldsymbol{\tau}\in\mathbb{H}_N(\mathbf{div};\Omega),\\ \mathbf{b}(\boldsymbol{\sigma},\boldsymbol{v}) - \mathbf{c}(\boldsymbol{u},\boldsymbol{v}) &= -\int_\Omega \boldsymbol{f}\phi\cdot\boldsymbol{v}, \quad \forall\,\boldsymbol{v}\in\mathbf{L}^2(\Omega), \end{aligned} \tag{3.2}$$

where $\mathbf{a}_\phi : \mathbb{H}_N(\mathbf{div};\Omega)\times\mathbb{H}_N(\mathbf{div};\Omega)\to\mathbb{R}$, $\mathbf{b} : \mathbb{H}_N(\mathbf{div};\Omega)\times\mathbf{L}^2(\Omega)\to\mathbb{R}$ and $\mathbf{c} : \mathbf{L}^2(\Omega)\times\mathbf{L}^2(\Omega)\to\mathbb{R}$ are bounded bilinear forms defined as

$$\mathbf{a}_\phi(\boldsymbol{\zeta},\boldsymbol{\tau}) := \int_\Omega \frac{1}{\mu(\phi)}\boldsymbol{\zeta}^{\mathrm{d}} : \boldsymbol{\tau}^{\mathrm{d}}, \quad \mathbf{b}(\boldsymbol{\tau},\boldsymbol{v}) := \int_\Omega \boldsymbol{v}\cdot\mathbf{div}\,\boldsymbol{\tau}, \quad \mathbf{c}(\boldsymbol{u},\boldsymbol{v}) := \int_\Omega \boldsymbol{K}^{-1}\boldsymbol{u}\cdot\boldsymbol{v} \tag{3.3}$$

for $\boldsymbol{\zeta},\boldsymbol{\tau}\in\mathbb{H}_N(\mathbf{div};\Omega)$ and $\boldsymbol{u},\boldsymbol{v}\in\mathbf{L}^2(\Omega)$.

In turn, given $\boldsymbol{u}\in\mathbf{L}^2(\Omega)$, and using the homogeneous Neumann boundary condition for $\tilde{\boldsymbol{\sigma}}$ (cf. fourth row of (2.4)), we deduce that the primal formulation for the concentration equation becomes: find $\phi\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ such that

$$A_{\boldsymbol{u}}(\phi,\psi) = \int_\Omega f_{\mathrm{bk}}(\phi)\boldsymbol{k}\cdot\nabla\psi + \int_\Omega g\psi, \quad \forall\,\psi\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega), \tag{3.4}$$

where

$$A_{\boldsymbol{u}}(\phi,\psi) := \int_\Omega \vartheta(\phi)\nabla\phi\cdot\nabla\psi - \int_\Omega \phi\boldsymbol{u}\cdot\nabla\psi + \int_\Omega \beta\phi\psi, \quad \forall\,\phi,\psi\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega). \tag{3.5}$$

We remark at this point that the well-posedness of (3.2) is a straightforward consequence of the assumption on μ given in (2.3) and the well known Babuška–Brezzi theory (see, e.g. Theorem 2.1 in Ref. 22 and Proposition 4.3.1 in Ref. 4 for details). However, in order to deal with the analysis of (3.4)–(3.5), and particularly to estimate the second term defining $A_{\boldsymbol{u}}$, we would require $\boldsymbol{u} \in \mathbf{H}^1(\Omega)$. In fact, we know from the Rellich–Kondrachov compactness theorem (cf. Theorem 6.3 in Ref. 2, Theorem 1.3.5 in Ref. 28), that the injection $i_c : \mathrm{H}^1(\Omega) \to \mathrm{L}^4(\Omega)$ is compact, and hence continuous, which, after applying Hölder's inequality, yields the existence of a positive constant $c(\Omega) = \|i_c\|^2$, depending only on Ω, such that

$$\left|\int_\Omega \phi\boldsymbol{v}\cdot\nabla\psi\right| \le c(\Omega)\|\phi\|_{1,\Omega}\|\boldsymbol{v}\|_{1,\Omega}|\psi|_{1,\Omega}, \quad \forall\,\phi,\psi\in\mathrm{H}^1(\Omega),\ \ \forall\,\boldsymbol{v}\in\mathbf{H}^1(\Omega). \tag{3.6}$$

Furthermore, we now observe, as we did in Ref. 2, that while the exact solution of (3.2) actually satisfies $\nabla\boldsymbol{u} = \frac{1}{\mu(\phi)}\boldsymbol{\sigma}^{\mathrm{d}}$ in $\mathcal{D}'(\Omega)$, which implies that $\boldsymbol{u}$ does belong to $\mathbf{H}^1(\Omega)$, the foregoing distributional identity does not necessarily extend to the discrete setting of (3.2), and hence the aforementioned difficulty would appear again when trying to analyze the Galerkin scheme associated to (3.4). In order to overcome this inconvenience, we proceed similarly as in Sec. 3.1 of Ref. 2 (see also Sec. 3 in Ref. 19) and incorporate into (3.2) the following residual Galerkin-type terms:

$$\begin{aligned}
&\kappa_1\int_\Omega\left(\nabla\boldsymbol{u} - \frac{1}{\mu(\phi)}\boldsymbol{\sigma}^{\mathrm{d}}\right) : \nabla\boldsymbol{v} = 0, \quad \forall\,\boldsymbol{v}\in\mathbf{H}^1(\Omega),\\
&-\kappa_2\int_\Omega \boldsymbol{K}^{-1}\boldsymbol{u}\cdot\mathbf{div}\,\boldsymbol{\tau} + \kappa_2\int_\Omega \mathbf{div}\,\boldsymbol{\sigma}\cdot\mathbf{div}\,\boldsymbol{\tau}\\
&\qquad = -\kappa_2\int_\Omega \boldsymbol{f}\phi\cdot\mathbf{div}\,\boldsymbol{\tau}, \quad \forall\,\boldsymbol{\tau}\in\mathbb{H}_N(\mathbf{div};\Omega),
\end{aligned} \tag{3.7}$$

where (κ_1,κ_2) is a vector of positive parameters to be specified later. In this way, instead of (3.2), we consider from now on the following augmented mixed formulation: find $(\boldsymbol{\sigma},\boldsymbol{u})\in\mathbb{H}_N(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega)$ such that

$$B_\phi((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau},\boldsymbol{v})) = F_\phi(\boldsymbol{\tau},\boldsymbol{v}), \quad \forall\,(\boldsymbol{\tau},\boldsymbol{v})\in\mathbb{H}_N(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega), \tag{3.8}$$

where

$$\begin{aligned}
&B_\phi((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau},\boldsymbol{v}))\\
&\quad := \mathbf{a}_\phi(\boldsymbol{\sigma},\boldsymbol{\tau}) + \mathbf{b}(\boldsymbol{\tau},\boldsymbol{u}) - \mathbf{b}(\boldsymbol{\sigma},\boldsymbol{v}) + \mathbf{c}(\boldsymbol{u},\boldsymbol{v}) + \kappa_1\int_\Omega\left(\nabla\boldsymbol{u} - \frac{1}{\mu(\phi)}\boldsymbol{\sigma}^{\mathrm{d}}\right) : \nabla\boldsymbol{v}\\
&\qquad - \kappa_2\int_\Omega \boldsymbol{K}^{-1}\boldsymbol{u}\cdot\mathbf{div}\,\boldsymbol{\tau} + \kappa_2\int_\Omega\mathbf{div}\,\boldsymbol{\sigma}\cdot\mathbf{div}\,\boldsymbol{\tau},
\end{aligned} \tag{3.9}$$

and

$$F_\phi(\boldsymbol{\tau},\boldsymbol{v}) := \langle\boldsymbol{\tau\nu},\boldsymbol{u}_{\mathrm{D}}\rangle_{\Gamma_{\mathrm{D}}} + \int_\Omega\boldsymbol{f}\phi\cdot\boldsymbol{v} - \kappa_2\int_\Omega\boldsymbol{f}\phi\cdot\mathbf{div}\,\boldsymbol{\tau}. \tag{3.10}$$

We remark in advance that the well-posedness of (3.8) is proved below in Sec. 3.3. To this respect, it is important to highlight that, differently from Ref. 2, here we do not need to add any stabilization term on the Dirichlet boundary, as we did in Eq. (3.6) of Ref. 2, since the required $\mathbf{H}^1(\Omega)$-norm is obtained thanks to the first equation in (3.7) and the presence now of the positive definite bilinear form $\mathbf{c}$ (cf. (3.3)) in the definition of B_ϕ. Furthermore, since the unique solution of (3.2) is obviously a solution of (3.8) as well, we will conclude that both continuous problems share the same unique solution.

Summarizing the foregoing discussion, we find that the augmented mixed-primal formulation of the initial coupled problem (2.4) reduces to (3.8) and (3.4), that is: find $(\boldsymbol{\sigma}, \boldsymbol{u}, \phi) \in \mathbb{H}_N(\mathbf{div};\Omega) \times \mathbf{H}^1(\Omega) \times \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)$ such that

$$\begin{aligned} B_\phi((\boldsymbol{\sigma}, \boldsymbol{u}), (\boldsymbol{\tau}, \boldsymbol{v})) &= F_\phi(\boldsymbol{\tau}, \boldsymbol{v}), \quad \forall\, (\boldsymbol{\tau}, \boldsymbol{v}) \in \mathbb{H}_N(\mathbf{div};\Omega) \times \mathbf{H}^1(\Omega), \\ A_{\boldsymbol{u}}(\phi, \psi) &= \int_\Omega f_{\mathrm{bk}}(\phi)\boldsymbol{k} \cdot \nabla\psi + \int_\Omega g\psi, \quad \forall\, \psi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega). \end{aligned} \tag{3.11}$$

3.2. *Fixed point strategy*

We begin by noticing that the alternative formulation (3.8) will certainly require continuous and discrete solutions with second components living in $\mathbf{H}^1(\Omega)$. Now, given $\phi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)$ and the corresponding solution $(\boldsymbol{\sigma}, \boldsymbol{u}) \in \mathbb{H}_N(\mathbf{div};\Omega) \times \mathbf{H}^1(\Omega)$ of (3.8), we can set, instead of (3.4), the modified primal formulation: find $\widetilde{\phi} \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)$ such that

$$A_{\phi,\boldsymbol{u}}(\widetilde{\phi}, \widetilde{\psi}) = G_\phi(\widetilde{\psi}), \quad \forall\, \widetilde{\psi} \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega), \tag{3.12}$$

where

$$A_{\phi,\boldsymbol{u}}(\widetilde{\phi}, \widetilde{\psi}) := \int_\Omega \vartheta(\phi)\nabla\widetilde{\phi} \cdot \nabla\widetilde{\psi} - \int_\Omega \widetilde{\phi}\boldsymbol{u} \cdot \nabla\widetilde{\psi} + \int_\Omega \beta\widetilde{\phi}\widetilde{\psi}, \quad \forall\, \widetilde{\phi}, \widetilde{\psi} \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega), \tag{3.13}$$

and

$$G_\phi(\widetilde{\psi}) := \int_\Omega f_{\mathrm{bk}}(\phi)\boldsymbol{k} \cdot \nabla\widetilde{\psi} + \int_\Omega g\widetilde{\psi}, \quad \forall\, \widetilde{\psi} \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega). \tag{3.14}$$

The well-posedness of (3.12) will also be addressed in Sec. 3.3.

In turn, the description of problems (3.8) and (3.12) naturally suggests a fixed point strategy to analyze (3.11). Indeed, let $\mathbf{S} : \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega) \to \mathbb{H}_N(\mathbf{div};\Omega) \times \mathbf{H}^1(\Omega)$ be the operator defined by

$$\mathbf{S}(\phi) = (\mathbf{S}_1(\phi), \mathbf{S}_2(\phi)) := (\boldsymbol{\sigma}, \boldsymbol{u}) \in \mathbb{H}_N(\mathbf{div};\Omega) \times \mathbf{H}^1(\Omega), \quad \forall\, \phi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega),$$

where $(\boldsymbol{\sigma}, \boldsymbol{u})$ is the unique solution of (3.8) with the given ϕ. In turn, let $\widetilde{\mathbf{S}} : \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega) \times \mathbf{H}^1(\Omega) \to \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)$ be the operator defined by

$$\widetilde{\mathbf{S}}(\phi, \boldsymbol{u}) := \widetilde{\phi}, \quad \forall\, (\phi, \boldsymbol{u}) \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega) \times \mathbf{H}^1(\Omega),$$

where $\widetilde{\phi}$ is the unique solution of (3.12) with the given $(\phi, \boldsymbol{u})$. Then, we define the operator $\mathbf{T} : \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega) \to \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)$ by

$$\mathbf{T}(\phi) := \widetilde{\mathbf{S}}(\phi, \mathbf{S}_2(\phi)), \quad \forall\, \phi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega),$$

and realize that solving (3.11) is equivalent to seeking a fixed point of $\mathbf{T}$, that is: find $\phi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)$ such that

$$\mathbf{T}(\phi) = \phi. \tag{3.15}$$

We find it important to remark here that, due to the dependence on ϕ (instead of $|\nabla\phi|$ as in Ref. 2) of the diffusivity ϑ, our nonlinear operator $A_{\boldsymbol{u}}$ (cf. (3.5)) does not become strongly monotone (as it was the case for the corresponding nonlinear operator in Eq. (3.4) and Lemma 3.5 in Ref. 2), and hence we realize that for easing the present analysis we need to stay with the linear problem (3.12) instead of the nonlinear one suggested by (3.4) and (3.5).

3.3. *Well-posedness of the uncoupled problems*

In this section, we show that the uncoupled problems (3.8) and (3.12) are in fact well-posed. We begin by recalling (see, e.g. Ref. 5) that $\mathbb{H}(\mathbf{div};\Omega) = \mathbb{H}_0(\mathbf{div};\Omega)\oplus \mathrm{R}\mathbb{I}$, where

$$\mathbb{H}_0(\mathbf{div};\Omega) := \left\{\boldsymbol{\zeta} \in \mathbb{H}(\mathbf{div};\Omega) : \int_\Omega \mathrm{tr}(\boldsymbol{\zeta}) = 0\right\}.$$

More precisely, for each $\boldsymbol{\zeta} \in \mathbb{H}(\mathbf{div};\Omega)$ there exist unique $\boldsymbol{\zeta}_0 := \boldsymbol{\zeta} - \{\frac{1}{n|\Omega|}\int_\Omega \mathrm{tr}(\boldsymbol{\zeta})\}\mathbb{I} \in \mathbb{H}_0(\mathbf{div};\Omega)$ and $d := \frac{1}{n|\Omega|}\int_\Omega \mathrm{tr}(\boldsymbol{\zeta}) \in \mathrm{R}$, such that $\boldsymbol{\zeta} = \boldsymbol{\zeta}_0 + d\mathbb{I}$. As for the analysis in Ref. 2, the following two lemmas concerning the above decomposition will be instrumental in showing the well-posedness of (3.8) for a given ϕ.

Lemma 3.1. (Proposition 3.1 in Ref. 5) *There exists* $c_1 = c_1(\Omega) > 0$ *such that*

$$c_1\|\boldsymbol{\tau}_0\|^2_{0,\Omega} \le \|\boldsymbol{\tau}^\mathrm{d}\|^2_{0,\Omega} + \|\mathbf{div}\,\boldsymbol{\tau}\|^2_{0,\Omega}\,, \quad \forall\,\boldsymbol{\tau} = \boldsymbol{\tau}_0 + c\mathbb{I} \in \mathbb{H}(\mathbf{div};\Omega),$$

with $\boldsymbol{\tau}_0 \in \mathbb{H}_0(\mathbf{div};\Omega)$ *and* $c \in \mathrm{R}$.

Lemma 3.2. (Lemma 2.2 in Ref. 20) *There exists* $c_2 = c_2(\Omega,\Gamma_\mathrm{N}) > 0$ *such that*

$$c_2\|\boldsymbol{\tau}\|^2_{\mathbf{div};\Omega} \le \|\boldsymbol{\tau}_0\|^2_{\mathbf{div};\Omega} \quad \forall\,\boldsymbol{\tau} = \boldsymbol{\tau}_0 + c\mathbb{I} \in \mathbb{H}_N(\mathbf{div};\Omega),$$

with $\boldsymbol{\tau}_0 \in \mathbb{H}_0(\mathbf{div};\Omega)$ *and* $c \in \mathrm{R}$.

We now begin the solvability analysis of the uncoupled problems with the following result.

Lemma 3.3. *Assume that* $\kappa_1 \in (0, \frac{2\delta\mu_1}{\mu_2})$ *and* $\kappa_2 \in (0, \frac{2\tilde\delta\alpha_{\boldsymbol{K}}}{n\|\boldsymbol{K}^{-1}\|_\infty})$, *with* $\delta \in (0, 2\mu_1)$ *and* $\tilde\delta \in (0, \frac{2}{n\|\boldsymbol{K}^{-1}\|_\infty})$. *Then, for each* $\phi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)$, *problem* (3.8) *has a unique solution* $\mathbf{S}(\phi) := (\boldsymbol{\sigma},\boldsymbol{u}) \in H := \mathbb{H}_N(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega)$. *Moreover, there exists* $C_S > 0$, *independent of* ϕ, *such that*

$$\|\mathbf{S}(\phi)\|_H = \|(\boldsymbol{\sigma}, \boldsymbol{u})\|_H \le C_S\{\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + \|\phi\|_{0,\Omega}\|\boldsymbol{f}\|_{\infty,\Omega}\}, \quad \forall\, \phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega). \tag{3.16}$$

Proof. We first observe from (3.9) that, given $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$, B_ϕ is clearly a bilinear form. Next, applying Cauchy–Schwarz inequality and the lower bound for μ (cf. (2.3)), we find from (3.9) that there exists a positive constant $\|B\|$, depending on μ_1, κ_1, κ_2, n and $\|\boldsymbol{K}^{-1}\|_\infty$, such that

$$|B_\phi((\boldsymbol{\sigma}, \boldsymbol{u}), (\boldsymbol{\tau}, \boldsymbol{v}))| \le \|B\| \|(\boldsymbol{\sigma}, \boldsymbol{u})\|_H \|(\boldsymbol{\tau}, \boldsymbol{v})\|_H, \quad \forall\, (\boldsymbol{\sigma}, \boldsymbol{u}), (\boldsymbol{\tau}, \boldsymbol{v}) \in H, \tag{3.17}$$

which confirms the boundedness of B_ϕ independently of $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$. Next, we show that B_ϕ is H-elliptic. In fact, given $(\boldsymbol{\tau}, \boldsymbol{v}) \in H$, we have again from (3.9) that

$$\begin{aligned} B_\phi((\boldsymbol{\tau}, \boldsymbol{v}), (\boldsymbol{\tau}, \boldsymbol{v})) = \int_\Omega \frac{1}{\mu(\phi)} |\boldsymbol{\tau}^{\mathtt{d}}|^2 + \kappa_1 |\boldsymbol{v}|^2_{1,\Omega} - \kappa_1 \int_\Omega \frac{1}{\mu(\phi)} \boldsymbol{\tau}^{\mathtt{d}} : \nabla \boldsymbol{v} \\ + \kappa_2 \|\mathbf{div}\,\boldsymbol{\tau}\|^2_{0,\Omega} + \int_\Omega \boldsymbol{K}^{-1}\boldsymbol{v} \cdot \boldsymbol{v} - \kappa_2 \int_\Omega \boldsymbol{K}^{-1}\boldsymbol{v} \cdot \mathbf{div}\,\boldsymbol{\tau}, \end{aligned}$$

which, using the lower and upper bounds for μ (cf. (2.3)), the Cauchy–Schwarz and Young inequalities, and the estimate (2.2), yields for any δ, $\widetilde{\delta} > 0$,

$$\begin{aligned} B_\phi((\boldsymbol{\tau}, \boldsymbol{v}), (\boldsymbol{\tau}, \boldsymbol{v})) \ge \left(\frac{1}{\mu_2} - \frac{\kappa_1}{2\delta\mu_1}\right) \|\boldsymbol{\tau}^{\mathtt{d}}\|^2_{0,\Omega} + \kappa_2 \left(1 - \frac{n\widetilde{\delta}}{2}\|\boldsymbol{K}^{-1}\|_\infty\right) \|\mathbf{div}\,\boldsymbol{\tau}\|^2_{0,\Omega} \\ + \kappa_1 \left(1 - \frac{\delta}{2\mu_1}\right) |\boldsymbol{v}|^2_{1,\Omega} + \left(\alpha_{\boldsymbol{K}} - \frac{n\kappa_2}{2\widetilde{\delta}}\|\boldsymbol{K}^{-1}\|_\infty\right) \|\boldsymbol{v}\|^2_{0,\Omega}. \end{aligned} \tag{3.18}$$

Then, assuming the indicated hypotheses on $\delta, \kappa_1, \widetilde{\delta}$ and κ_2, we can introduce the positive constants:

$$\begin{aligned} \alpha_0(\Omega) &:= \min\left\{\left(\frac{1}{\mu_2} - \frac{\kappa_1}{2\delta\mu_1}\right), \frac{\kappa_2}{2}\left(1 - \frac{n\widetilde{\delta}}{2}\|\boldsymbol{K}^{-1}\|_\infty\right)\right\}, \\ \alpha_1(\Omega) &:= c_2 \min\left\{c_1\alpha_0(\Omega), \frac{\kappa_2}{2}\left(1 - \frac{n\widetilde{\delta}}{2}\|\boldsymbol{K}^{-1}\|_\infty\right)\right\}, \\ \alpha_2(\Omega) &:= \min\left\{\kappa_1\left(1 - \frac{\delta}{2\mu_1}\right), \left(\alpha_{\boldsymbol{K}} - \frac{n\kappa_2}{2\widetilde{\delta}}\|\boldsymbol{K}^{-1}\|_\infty\right)\right\}, \end{aligned}$$

which, according to Lemmas 3.1 and 3.2, and defining $\alpha(\Omega) := \min\{\alpha_1(\Omega), \alpha_2(\Omega)\}$, implies from (3.18) that

$$B_\phi((\boldsymbol{\tau}, \boldsymbol{v}), (\boldsymbol{\tau}, \boldsymbol{v})) \ge \alpha(\Omega)\|(\boldsymbol{\tau}, \boldsymbol{v})\|^2, \quad \forall\, (\boldsymbol{\tau}, \boldsymbol{v}) \in \mathbf{H}, \tag{3.19}$$

thus confirming the H-ellipticity of B_ϕ independently of $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ as well. Next, given $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$, it is easy to see from (3.10) that there exists a positive constant

$\|F\|$, depending only on κ_2, such that

$$\|F_\phi\| \le \|F\|\{\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + \|\phi\|_{0,\Omega}\|\boldsymbol{f}\|_{\infty,\Omega}\}. \tag{3.20}$$

Finally, a straightforward application of the Lax–Milgram lemma (see, e.g. Theorem 1.1 in Ref. 21), proves that, for each $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$, problem (3.8) has a unique solution $\mathbf{S}(\phi) := (\boldsymbol{\sigma}, \boldsymbol{u}) \in H$. Moreover, the corresponding continuous dependence result together with the estimates (3.19) and (3.20) yield (3.16) with $C_S := \frac{\|F\|}{\alpha(\Omega)}$, which completes the proof. □

We now establish the unique solvability of the linear problem (3.12).

Lemma 3.4. *Let* $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ *and* $\boldsymbol{u} \in \mathbf{H}^1(\Omega)$ *such that* $\|\boldsymbol{u}\|_{1,\Omega} < \frac{\vartheta_1}{2c_p c(\Omega)}$ *(cf.* (2.3), (3.1), (3.6)*). Then, there exists a unique* $\widetilde{\phi} := \widetilde{\mathbf{S}}(\phi, \boldsymbol{u}) \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ *solution of* (3.12)*. Moreover, there exists* $C_{\widetilde{S}} > 0$, *independent of* $(\phi, \boldsymbol{u})$, *such that*

$$\|\widetilde{\mathbf{S}}(\phi, \boldsymbol{u})\|_{1,\Omega} = \|\widetilde{\phi}\|_{1,\Omega} \le C_{\widetilde{S}}\{\gamma_2|\Omega|^{1/2}|\boldsymbol{k}| + \|g\|_{0,\Omega}\}. \tag{3.21}$$

Proof. First notice that $A_{\phi,\boldsymbol{u}}$ (cf. (3.13)) is clearly a bilinear form. In turn, according to (2.3) and (3.6), it readily follows from (3.13) that there exists a positive constant $\|A\|$, depending on ϑ_2, $c(\Omega)$, and the bound for $\|\boldsymbol{u}\|_{1,\Omega}$ assumed here, that

$$|A_{\phi,\boldsymbol{u}}(\widetilde{\phi}, \widetilde{\psi})| \le \|A\|\|\widetilde{\phi}\|_{1,\Omega}\|\widetilde{\psi}\|_{1,\Omega} \quad \forall\, \widetilde{\phi}, \widetilde{\psi} \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega),$$

which proves that $A_{\phi,\boldsymbol{u}}$ is bounded independently of ϕ and $\boldsymbol{u}$. Next, applying the estimate (3.6) and the inequality (3.1), we find that for each $\widetilde{\phi} \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ there holds

$$\begin{aligned} A_{\phi,\boldsymbol{u}}(\widetilde{\phi}, \widetilde{\phi}) &= \int_\Omega \vartheta(\phi)|\nabla\widetilde{\phi}|^2 - \int_\Omega \widetilde{\phi}\boldsymbol{u}\cdot\nabla\widetilde{\phi} + \beta\|\widetilde{\phi}\|^2_{0,\Omega} \\ &\ge (\vartheta_1 - c_p c(\Omega)\|\boldsymbol{u}\|_{1,\Omega})|\widetilde{\phi}|^2_{1,\Omega} \ge \frac{\vartheta_1}{2}|\widetilde{\phi}|^2_{1,\Omega} \ge \frac{\vartheta_1}{2c_p^2}\|\widetilde{\phi}\|^2_{1,\Omega}, \end{aligned} \tag{3.22}$$

which proves that $A_{\phi,\boldsymbol{u}}$ is $\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$-elliptic with constant $\widetilde{\alpha} := \frac{\vartheta_1}{2c_p^2}$, independently of ϕ and $\boldsymbol{u}$ as well. On the other hand, applying Cauchy–Schwarz inequality and the upper bound for f_{bk} given in (2.3), we easily deduce that

$$|G_\phi(\widetilde{\psi})| \le \{\gamma_2|\Omega|^{1/2}|\boldsymbol{k}| + \|g\|_{0,\Omega}\}\|\widetilde{\psi}\|_{1,\Omega}, \quad \forall\, \widetilde{\psi} \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega),$$

which says that $G_\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'$ and $\|G_\phi\| \le \{\gamma_2|\Omega|^{1/2}|\boldsymbol{k}| + \|g\|_{0,\Omega}\}$. Consequently, a direct application of the Lax–Milgram lemma implies the existence of a unique solution $\widetilde{\phi} = \widetilde{\mathbf{S}}(\phi, \boldsymbol{u}) \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ of (3.12), and the corresponding continuous dependence result becomes (3.21) with $C_{\widetilde{S}} = \frac{1}{\widetilde{\alpha}} = \frac{2c_p^2}{\vartheta_1}$. □

At this point we remark that the restriction on $\|\boldsymbol{u}\|_{1,\Omega}$ in Lemma 3.4 could also have been taken as $\|\boldsymbol{u}\|_{1,\Omega} < \varepsilon\frac{\vartheta_1}{c_p c(\Omega)}$ with any $\varepsilon \in (0, 1)$. However, we have chosen $\varepsilon = \frac{1}{2}$ for simplicity and because it yields a joint maximization of the ellipticity constant of $A_{\phi,\boldsymbol{u}}$ and the upper bound for $\|\boldsymbol{u}\|_{1,\Omega}$. In addition, when dropping the term $\beta\|\widetilde{\phi}\|_{0,\Omega}^2$ in (3.22) we have first assumed that β is small and then utilized the Poincaré inequality (3.1). In turn, when β is sufficiently large, say $\beta \geq \vartheta_1$, then the aforementioned expression is kept along the whole derivation of (3.22), so that in this case the Poincaré inequality (3.1) does not need to be applied.

We end this section by introducing suitable regularity hypotheses on the operators $\mathbf{S}$ and $\widetilde{\mathbf{S}}$, which will be employed later on. In fact, for the remainder of this paper we follow Eq. (3.22) in Ref. 2, and suppose that $\boldsymbol{u}_{\mathrm{D}} \in \mathrm{H}^{1/2+\delta}(\Gamma_{\mathrm{D}})$ and $g \in \mathrm{H}^{\delta}(\Omega)$, for some $\delta \in (0,1)$ (when $n = 2$) or $\delta \in (1/2, 1)$ (when $n = 3$). Then, we assume that for each $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ and $(\varphi, \boldsymbol{w}) \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) \times \mathbf{H}^1(\Omega)$, with $\|\phi\|_{1,\Omega} \leq r$ and $\|\varphi\|_{1,\Omega} + \|\boldsymbol{w}\|_{1,\Omega} \leq r, r > 0$ given, there holds, respectively, $\mathbf{S}(\phi) \in \mathbb{H}_N(\mathbf{div};\Omega) \cap \mathbb{H}^{\delta}(\Omega) \times \mathbf{H}^{1+\delta}(\Omega)$ and $\widetilde{\mathbf{S}}(\varphi, \boldsymbol{w}) \in \mathrm{H}^{1+\delta}_{\Gamma_{\mathrm{D}}}(\Omega)$, with

$$\|\mathbf{S}_1(\phi)\|_{\delta,\Omega} + \|\mathbf{S}_2(\phi)\|_{1+\delta,\Omega} \leq \widehat{C}_S(r)\{\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} + \|\phi\|_{0,\Omega}\|\boldsymbol{f}\|_{\infty,\Omega}\}, \tag{3.23}$$

and

$$\|\widetilde{\mathbf{S}}(\varphi, \boldsymbol{w})\|_{1+\delta,\Omega} \leq \widehat{C}_{\widetilde{S}}(r)\{\gamma_2|\Omega|^{1/2}|\boldsymbol{k}| + \|g\|_{\delta,\Omega}\}, \tag{3.24}$$

where $\widehat{C}_S(r)$ and $\widehat{C}_{\widetilde{S}}(r)$ are positive constants independent of ϕ and $(\varphi, \boldsymbol{w})$, respectively, but depending on the upper bound r of their norms. The reason of the stipulated ranges for δ will be clarified in the forthcoming analysis (see below proofs of Lemmas 3.6 and 3.7). More precisely, we remark in advance that the regularity estimate (3.23) is needed in the proof of Lemma 3.6 to bound an expression of the form $\|\mathbf{S}_1(\phi)\|_{\mathbb{L}^{2p}(\Omega)}$ in terms of $\|\mathbf{S}_1(\phi)\|_{\delta,\Omega}$, and hence of the data at the right-hand side of (3.23) (further details are available in the proof of Lemma 3.9 in Ref. 2). In turn, (3.24) is employed in the proof of Lemma 3.7 to bound an expression of the form $\|\nabla\widetilde{\mathbf{S}}(\varphi, \boldsymbol{w})\|_{\mathbb{L}^{2p}(\Omega)}$ in terms of $\|\widetilde{\mathbf{S}}(\varphi, \boldsymbol{w})\|_{1+\delta,\Omega}$, and hence of the data at the right-hand side of (3.24) (see (3.31) below for details).

Though the actual verification of (3.23) and (3.24) is beyond the goals of this paper, we remark that some insights confirming the feasibility of the assumed regularity for the nonlinear problem defining $\mathbf{S}$ were already provided in remarks below Eq. (3.22) in Ref. 2. In turn, the assumed regularity of the linear problem defining $\widetilde{\mathbf{S}}$ is quite standard in the realm of elliptic boundary value problems, and we just refer the interested reader to Ref. 15 or Ref. 23.

3.4. *Solvability of the fixed point equation*

We begin by emphasizing that the well-posedness of the uncoupled problems (3.8) and (3.12) confirms that the operators $\mathbf{S}, \widetilde{\mathbf{S}}$ and $\mathbf{T}$ (cf. Sec. 3.2) are well defined, and hence now we can address the solvability analysis of the fixed point Eq. (3.15). To this end, we will verify below the hypotheses of the Schauder fixed point theorem (see, e.g. Theorem 9.12-1(b) in Ref. 12), for which we require Lipschitz continuity of the nonlinear functions f_{bk}, ϑ and μ. More precisely, we assume that there exist positive constants L_μ, L_ϑ and L_f, such that for each $s, t \in \mathrm{R}$ there hold

$$\begin{gathered} |\mu(s) - \mu(t)| \leq L_\mu |s-t|, \quad |\vartheta(s) - \vartheta(t)| \leq L_\vartheta |s-t| \quad \text{and} \\ |f_{\mathrm{bk}}(s) - f_{\mathrm{bk}}(t)| \leq L_f |s-t|. \end{gathered} \tag{3.25}$$

We begin the analysis with the following straightforward consequence of Lemmas 3.3 and 3.4.

Lemma 3.5. *Given $r > 0$, we let $W := \{\phi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega) : \|\phi\|_{1,\Omega} \leq r\}$, and assume that*

$$C_S\{\|\boldsymbol{u}_\mathrm{D}\|_{1/2,\Gamma_\mathrm{D}} + r\|\boldsymbol{f}\|_{\infty,\Omega}\} < \frac{\vartheta_1}{2c_p c(\Omega)} \quad \text{and} \quad C_{\widetilde{S}}\{\gamma_2|\Omega|^{1/2}|\boldsymbol{k}| + \|g\|_{0,\Omega}\} \leq r, \tag{3.26}$$

where C_S and $C_{\widetilde{S}}$ are the constants specified in Lemmas 3.3 *and* 3.4, *respectively. Then $\mathbf{T}(W) \subseteq W$.*

Proof. It corresponds to a slight modification of the proof of Lemma 3.8 in Ref. 2. □

Next, similarly as in Ref. 2, the continuity and compactness of $\mathbf{T}$ will essentially be direct consequences of the following two lemmas providing the continuity of $\mathbf{S}$ and $\widetilde{\mathbf{S}}$, respectively.

Lemma 3.6. *There exists a positive constant C, depending on $\mu_1, \kappa_1, \kappa_2, L_\mu, \alpha(\Omega)$ and δ (cf.* (2.3), (3.7), (3.25), (3.19), (3.23)), *such that*

$$\|\mathbf{S}(\phi) - \mathbf{S}(\varphi)\|_H \leq C\{\|\boldsymbol{f}\|_{\infty,\Omega}\|\phi - \varphi\|_{0,\Omega} + \|\mathbf{S}_1(\varphi)\|_{\delta,\Omega}\|\phi - \varphi\|_{\mathrm{L}^{n/\delta}(\Omega)}\},$$
$$\forall\, \phi, \varphi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega).$$

Proof. Even though the present bilinear form B_ϕ (cf. (3.9)) and the corresponding one from Ref. 2 differ in a couple of linear terms, the present proof is almost verbatim as Lemma 3.9 in Ref. 2, particularly concerning the utilization of the Lipschitz-continuity of μ (cf. (3.25)), the regularity estimate (3.23), and the Sobolev embedding theorem (cf. Theorem 4.12 in Ref. 1, Theorem 1.3.4 in Ref. 28), and hence further details are omitted. □

On the contrary to the foregoing lemma, and due to the fact already mentioned that the diffusivity ϑ depends now on the scalar value of the concentration instead of the magnitude of its gradient (as it is in Ref. 2), the proof of the Lipschitz-continuity of the operator $\widetilde{\mathbf{S}}$, being more involved, differs substantially from the one given for the analogue result of Lemma 3.10 in Ref. 2. In particular, as a consequence of the aforementioned dependences, the regularity assumption (3.24), which was not required for the proof of Lemma 3.10 in Ref. 2, will definitely be employed next.

Lemma 3.7. *Let $C_{\widetilde{S}}$ be the constant provided by Lemma* 3.4. *Then, there exists a positive constant $\widetilde{C}$, depending on $C_{\widetilde{S}}, c(\Omega), L_f, L_\vartheta$ and δ (cf.* (3.6), (3.25), (3.24)), *such that for all* $(\phi, \boldsymbol{u}), (\varphi, \boldsymbol{w}) \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega) \times \mathbf{H}^1(\Omega)$, *with* $\|\boldsymbol{u}\|_{1,\Omega}, \|\boldsymbol{w}\|_{1,\Omega} < \frac{\vartheta_1}{2c_p c(\Omega)}$, *there holds*

$$\|\widetilde{\mathbf{S}}(\phi, \boldsymbol{u}) - \widetilde{\mathbf{S}}(\varphi, \boldsymbol{w})\|_{1,\Omega} \leq \widetilde{C}\{|\boldsymbol{k}|\|\phi - \varphi\|_{0,\Omega} + \|\widetilde{\mathbf{S}}(\varphi, \boldsymbol{w})\|_{1,\Omega}\|\boldsymbol{u} - \boldsymbol{w}\|_{1,\Omega} + \|\widetilde{\mathbf{S}}(\varphi, \boldsymbol{w})\|_{1+\delta,\Omega}\|\phi - \varphi\|_{\mathrm{L}^{n/\delta}(\Omega)}\}. \tag{3.27}$$

Proof. Given $(\phi, \boldsymbol{u}), (\varphi, \boldsymbol{w})$ as stated, we let $\widetilde{\phi} := \widetilde{\mathbf{S}}(\phi, \boldsymbol{u})$ and $\widetilde{\varphi} := \widetilde{\mathbf{S}}(\varphi, \boldsymbol{w})$, that is (cf. (3.12)):

$$A_{\phi,\boldsymbol{u}}(\widetilde{\phi}, \widetilde{\psi}) = G_\phi(\widetilde{\psi}) \quad \text{and} \quad A_{\varphi,\boldsymbol{w}}(\widetilde{\varphi}, \widetilde{\psi}) = G_\varphi(\widetilde{\psi}), \quad \forall\, \widetilde{\psi} \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega).$$

It follows, according to the ellipticity of $A_{\phi,\boldsymbol{u}}$ with constant $\widetilde{\alpha}$, and then subtracting and adding $G_\varphi(\widetilde{\phi} - \widetilde{\varphi}) = A_{\varphi,\boldsymbol{w}}(\widetilde{\varphi}, \widetilde{\phi} - \widetilde{\varphi})$, that

$$\begin{aligned}
\widetilde{\alpha}\|\widetilde{\phi} - \widetilde{\varphi}\|^2_{1,\Omega} &\leq A_{\phi,\boldsymbol{u}}(\widetilde{\phi}, \widetilde{\phi} - \widetilde{\varphi}) - A_{\phi,\boldsymbol{u}}(\widetilde{\varphi}, \widetilde{\phi} - \widetilde{\varphi}) \\
&= G_\phi(\widetilde{\phi} - \widetilde{\varphi}) - G_\varphi(\widetilde{\phi} - \widetilde{\varphi}) + A_{\varphi,\boldsymbol{w}}(\widetilde{\varphi}, \widetilde{\phi} - \widetilde{\varphi}) - A_{\phi,\boldsymbol{u}}(\widetilde{\varphi}, \widetilde{\phi} - \widetilde{\varphi}) \\
&= \int_\Omega (f_{\mathrm{bk}}(\phi) - f_{\mathrm{bk}}(\varphi))\boldsymbol{k} \cdot \nabla(\widetilde{\phi} - \widetilde{\varphi}) + \int_\Omega \widetilde{\varphi}(\boldsymbol{u} - \boldsymbol{w}) \cdot \nabla(\widetilde{\phi} - \widetilde{\varphi}) \\
&\quad + \int_\Omega (\vartheta(\varphi) - \vartheta(\phi))\nabla\widetilde{\varphi} \cdot \nabla(\widetilde{\phi} - \widetilde{\varphi}),
\end{aligned} \tag{3.28}$$

where the last equality has employed the definitions given by (3.13) and (3.14). Then applying Cauchy–Schwarz's inequality, the Lipschitz-continuity assumption (3.25) on the last term in (3.28), and then Hölder's inequality, we obtain

$$\begin{aligned}
\widetilde{\alpha}\|\widetilde{\phi} - \widetilde{\varphi}\|^2_{1,\Omega} &\leq \{L_f|\boldsymbol{k}|\|\phi - \varphi\|_{0,\Omega} + c(\Omega)\|\widetilde{\varphi}\|_{1,\Omega}\|\boldsymbol{u} - \boldsymbol{w}\|_{1,\Omega}\}|\widetilde{\phi} - \widetilde{\varphi}|_{1,\Omega} \\
&\quad + L_\vartheta\|\phi - \varphi\|_{\mathrm{L}^{2q}(\Omega)}\|\nabla\widetilde{\varphi}\|_{\mathrm{L}^{2p}(\Omega)}|\widetilde{\phi} - \widetilde{\varphi}|_{1,\Omega},
\end{aligned} \tag{3.29}$$

where $p, q \in [1, +\infty)$ are such that $1/p + 1/q = 1$. Next, given the further regularity δ assumed in (3.24), we recall that the Sobolev embedding theorem (cf. Theorem 4.12 in Ref. 1, Theorem 1.3.4 in Ref. 28) establishes the continuous injection $i_\delta : \mathrm{H}^\delta(\Omega) \to \mathrm{L}^{\delta^*}(\Omega)$ with boundedness constant C_δ, where

$$\delta^* := \begin{cases} \dfrac{2}{1-\delta} & \text{if } n = 2, \\[2ex] \dfrac{6}{3-2\delta} & \text{if } n = 3. \end{cases} \tag{3.30}$$

Thus, choosing p such that $2p = \delta^*$, we find that

$$\|\nabla\widetilde{\varphi}\|_{\mathrm{L}^{2p}(\Omega)} = \|\nabla\widetilde{\mathbf{S}}(\varphi, \boldsymbol{w})\|_{\mathrm{L}^{2p}(\Omega)} \leq C_\delta\|\nabla\widetilde{\mathbf{S}}(\varphi, \boldsymbol{w})\|_{\delta,\Omega} \leq C_\delta\|\widetilde{\mathbf{S}}(\varphi, \boldsymbol{w})\|_{1+\delta,\Omega}. \tag{3.31}$$

In turn, according to the above choice of p, that is $p = \delta^*/2$, it readily follows that

$$2q := \frac{2p}{p-1} = \begin{cases} \dfrac{2}{\delta} & \text{if } n = 2 \\ \dfrac{3}{\delta} & \text{if } n = 3 \end{cases} = \frac{n}{\delta}. \tag{3.32}$$

In this way, inequalities (3.28), (3.29) and (3.31) together with identity (3.32) imply (3.27), which finishes the proof. □

The following result, which is the analogue of Lemma 3.11 in Ref. 2, is a straightforward corollary of Lemmas 3.5–3.7.

Lemma 3.8. *Given* $r > 0$, *we let* $W := \{\phi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega) : \|\phi\|_{1,\Omega} \le r\}$, *and assume* (3.26) (*cf. Lemma* 3.5)*. Then, with the constants* C *and* $\widetilde{C}$ *from Lemmas* 3.6 *and* 3.7, *for all* $\phi, \varphi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)$ *there holds*

$$\begin{aligned} \|\mathbf{T}(\phi) - \mathbf{T}(\varphi)\|_{1,\Omega} &\le \widetilde{C}\{|\boldsymbol{k}| + C\|\mathbf{T}(\varphi)\|_{1,\Omega}\|\boldsymbol{f}\|_{\infty,\Omega}\}\|\phi - \varphi\|_{0,\Omega} \\ &\quad + \widetilde{C}\{C\|\mathbf{T}(\varphi)\|_{1,\Omega}\|\mathbf{S}_1(\varphi)\|_{\delta,\Omega} + \|\mathbf{T}(\varphi)\|_{1+\delta,\Omega}\}\|\phi - \varphi\|_{\mathrm{L}^{n/\delta}(\Omega)}. \end{aligned} \tag{3.33}$$

Proof. It suffices to recall from Sec. 3.2 that $\mathbf{T}(\phi) = \widetilde{\mathbf{S}}(\phi, \mathbf{S}_2(\phi))\ \forall\, \phi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)$, and then apply Lemmas 3.5, 3.6 and 3.7. □

The announced properties of $\mathbf{T}$ are proved now.

Lemma 3.9. *Given* $r > 0$, *we let* $W := \{\phi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega) : \|\phi\|_{1,\Omega} \le r\}$, *and assume* (3.26) (*cf. Lemma* 3.5)*. Then,* $\mathbf{T} : W \to W$ *is continuous and* $\overline{\mathbf{T}(W)}$ *is compact.*

Proof. It follows almost verbatim as the proof of Lemma 3.12 in Ref. 2. Indeed, it is basically a consequence of the Rellich–Kondrachov compactness theorem (cf. Theorem 6.3 in Ref. 2, Theorem 1.3.5 in Ref. 28), the specified range of the constant δ involved in the further regularity assumptions given by (3.23) and (3.24), and the well-known fact that every bounded sequence in a Hilbert space has a weakly convergent subsequence. We omit the rest of details. □

Finally, the main result of this section is given as follows, where the proof can be obtained very much as in Theorem 3.13 in Ref. 2.

Theorem 3.10. *Assume that the hypotheses of the Lemmas* 3.5–3.9 *are met. Then the augmented mixed-primal problem* (3.11) *has at least one solution* $(\boldsymbol{\sigma}, \boldsymbol{u}, \phi) \in \mathbb{H}_N(\mathbf{div}; \Omega) \times \mathbf{H}^1(\Omega) \times \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)$ *with* $\phi \in W$, *and there holds*

$$\|\phi\|_{1,\Omega} \le C_{\widetilde{S}}\{\gamma_2|\Omega|^{1/2}|\boldsymbol{k}| + \|g\|_{0,\Omega}\}, \tag{3.34}$$

and

$$\|(\boldsymbol{\sigma},\boldsymbol{u})\|_H \le C_S\{\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + r\|\boldsymbol{f}\|_{\infty,\Omega}\}, \tag{3.35}$$

where C_S *and* $C_{\widetilde{S}}$ *are the constants specified in Lemmas* 3.3 *and* 3.4, *respectively. Moreover, if the data* $\boldsymbol{k}, \boldsymbol{f}, g$ *and* $\boldsymbol{u}_{\mathrm{D}}$ *are sufficiently small so that, with the constants* $C, \widetilde{C}, \widehat{C}_S(r)$ *and* $\widehat{C}_{\widetilde{S}}(r)$ *from Lemmas* 3.6 *and* 3.7, *and estimates* (3.23) *and* (3.24), *and denoting by* $\widetilde{C}_\delta$ *the boundedness constant of the continuous injection of* $\mathrm{H}^1(\Omega)$ *into* $\mathrm{L}^{n/\delta}(\Omega)$, *there holds*

$$\begin{aligned}&\widetilde{C}(1+\widehat{C}_{\widetilde{S}}(r)\widetilde{C}_\delta C\gamma_2|\Omega|^{1/2})|\boldsymbol{k}| + \widehat{C}_{\widetilde{S}}(r)\widetilde{C}\widetilde{C}_\delta\|g\|_{\delta,\Omega} + \widetilde{C}\widetilde{C}_\delta C\widehat{C}_S(r)\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}}\\ &\quad + r\widetilde{C}C(1+r\widetilde{C}_\delta\widehat{C}_S(r))\|\boldsymbol{f}\|_{\infty,\Omega} < 1,\end{aligned} \tag{3.36}$$

then the solution ϕ *is unique in* W.

Proof. According to the equivalence between (3.11) and the fixed point equation (3.15), and thanks to the previous Lemmas 3.5 and 3.9, the existence of solution is just a straightforward application of the Schauder fixed point theorem (cf. Theorem 9.12-1(b) in Ref. 12). In turn, the estimates (3.34) and (3.35) follow from (3.16) (cf. Lemma 3.3) and (3.21) (cf. Lemma 3.4). Furthermore, given another solution $\varphi \in W$ of (3.15), the estimates $\|\mathbf{T}(\varphi)\|_{1,\Omega} = \|\varphi\|_{1,\Omega} \le r$,

$$\|\mathbf{S}_1(\varphi)\|_{\delta,\Omega} \le \widehat{C}_S(r)\{\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} + \|\varphi\|_{0,\Omega}\|\boldsymbol{f}\|_{\infty,\Omega}\} \quad (\text{cf. } (3.23)),$$

$$\|\widetilde{\varphi}\|_{1+\delta,\Omega} \le \widehat{C}_{\widetilde{S}}(r)\{\gamma_2|\Omega|^{1/2}|\boldsymbol{k}| + \|g\|_{\delta,\Omega}\} \quad (\text{cf. } (3.24)),$$

and

$$\|\psi\|_{\mathrm{L}^{n/\delta}(\Omega)} \le \widetilde{C}_\delta\|\psi\|_{1,\Omega}, \quad \forall\,\psi \in \mathrm{H}^1(\Omega), \tag{3.37}$$

confirm (3.36) as a sufficient condition for concluding, together with (3.33), that $\phi = \varphi$. In other words, (3.36) constitutes the condition arising from (3.33) — once (3.37), and the *a priori* and regularity estimates for $\|\mathbf{T}(\varphi)\|_{1,\Omega}$, $\|\mathbf{S}_1(\varphi)\|_{\delta,\Omega}$ and $\|\mathbf{T}(\varphi)\|_{1+\delta,\Omega}$, respectively, are employed — that makes the operator $\mathbf{T}$ to become a contraction, thus yielding the existence of a unique fixed point of $\mathbf{T}$ in W. □

4. The Galerkin Scheme

Let $\mathcal{T}_h$ be a regular triangulation of Ω by triangles K (respectively, tetrahedra K in R^3) of diameter h_K, and define the mesh size $h := \max\{h_K : K \in \mathcal{T}_h\}$. In addition, given an integer $k \ge 0$, for each $K \in \mathcal{T}_h$ we let $\mathrm{P}_k(K)$ be the space of polynomial functions on K of degree $\le k$, and define the corresponding local Raviart–Thomas space of order k as

$$\mathbf{RT}_k(K) := \mathbf{P}_k(K) \oplus \mathrm{P}_k(K)\boldsymbol{x},$$

where, according to the notations described in Sec. 1, $\mathbf{P}_k(K) = [\mathrm{P}_k(K)]^n$, and $\boldsymbol{x}$ is the generic vector in R^n. Then, we introduce the finite element subspaces approximating the unknowns $\boldsymbol{\sigma}$, $\boldsymbol{u}$, and ϕ, respectively, as the global Raviart–Thomas space of order k, and the corresponding Lagrange spaces given by the continuous piecewise polynomials of degree $\leq k+1$, that is:

$$\mathbb{H}_h^{\boldsymbol{\sigma}} := \{\boldsymbol{\tau}_h \in \mathbb{H}_N(\mathbf{div};\Omega) : \mathbf{c}^{\mathrm{t}}\boldsymbol{\tau}_h|_K \in \mathbf{RT}_k(K)\,\forall\,\mathbf{c} \in \mathrm{R}^n,\ \forall\,K \in \mathcal{T}_h\}, \tag{4.1}$$

$$\mathbf{H}_h^{\boldsymbol{u}} := \{\boldsymbol{v}_h \in \mathbf{C}(\Omega) : \boldsymbol{v}_h|_K \in \mathbf{P}_{k+1}(K)\,\forall\,K \in \mathcal{T}_h\}, \tag{4.2}$$

$$\mathrm{H}_h^{\phi} := \{\psi_h \in C(\Omega) \cap \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) : \psi_h|_K \in \mathrm{P}_{k+1}(K)\,\forall\,K \in \mathcal{T}_h\}. \tag{4.3}$$

In this way, the underlying Galerkin scheme, given by the discrete counterpart of (3.11), reads: find $(\boldsymbol{\sigma}_h, \boldsymbol{u}_h, \phi_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}} \times \mathrm{H}_h^{\phi}$ such that

$$\begin{aligned} B_{\phi_h}((\boldsymbol{\sigma}_h, \boldsymbol{u}_h), (\boldsymbol{\tau}_h, \boldsymbol{v}_h)) &= F_{\phi_h}(\boldsymbol{\tau}_h, \boldsymbol{v}_h), \quad \forall\,(\boldsymbol{\tau}_h, \boldsymbol{v}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}, \\ A_{\boldsymbol{u}_h}(\phi_h, \psi_h) &= \int_\Omega f_{\mathrm{bk}}(\phi_h)\boldsymbol{k} \cdot \nabla\psi_h + \int_\Omega g\psi_h, \quad \forall\,\psi_h \in \mathrm{H}_h^{\phi}. \end{aligned} \tag{4.4}$$

Throughout the rest of this section we adopt the discrete analogue of the fixed point strategy introduced in Sec. 3.2. Hence, we now let $\mathbf{S}_h : \mathrm{H}_h^{\phi} \to \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}$ be the operator defined by

$$\mathbf{S}_h(\phi_h) = (\mathbf{S}_{1,h}(\phi_h), \mathbf{S}_{2,h}(\phi_h)) := (\boldsymbol{\sigma}_h, \boldsymbol{u}_h), \quad \forall\,\phi_h \in \mathrm{H}_h^{\phi},$$

where $(\boldsymbol{\sigma}_h, \boldsymbol{u}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}$ is the unique solution of

$$B_{\phi_h}((\boldsymbol{\sigma}_h, \boldsymbol{u}_h), (\boldsymbol{\tau}_h, \boldsymbol{v}_h)) = F_{\phi_h}(\boldsymbol{\tau}_h, \boldsymbol{v}_h), \quad \forall\,(\boldsymbol{\tau}_h, \boldsymbol{v}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}, \tag{4.5}$$

with B_{ϕ_h} and F_{ϕ_h} being defined by (3.9) and (3.10), respectively, with $\phi = \phi_h$. In addition, we let $\widetilde{\mathbf{S}}_h : \mathrm{H}_h^{\phi} \times \mathbf{H}_h^{\boldsymbol{u}} \to \mathrm{H}_h^{\phi}$ be the operator defined by

$$\widetilde{\mathbf{S}}_h(\phi_h, \boldsymbol{u}_h) := \widetilde{\phi}_h, \quad \forall\,(\phi_h, \boldsymbol{u}_h) \in \mathrm{H}_h^{\phi} \times \mathbf{H}_h^{\boldsymbol{u}},$$

where $\widetilde{\phi}_h \in \mathrm{H}_h^{\phi}$ is the unique solution of

$$A_{\phi_h, \boldsymbol{u}_h}(\widetilde{\phi}_h, \widetilde{\psi}_h) = G_{\phi_h}(\widetilde{\psi}_h), \quad \forall\,\widetilde{\psi}_h \in \mathrm{H}_h^{\phi}, \tag{4.6}$$

with $A_{\phi_h, \boldsymbol{u}_h}$ and G_{ϕ_h} being defined by (3.13) and (3.14), respectively, with $\boldsymbol{u} = \boldsymbol{u}_h$ and $\phi = \phi_h$. Finally, we define the operator $\mathbf{T}_h : \mathrm{H}_h^{\phi} \to \mathrm{H}_h^{\phi}$ by

$$\mathbf{T}_h(\phi_h) := \widetilde{\mathbf{S}}_h(\phi_h, \mathbf{S}_{2,h}(\phi_h)), \quad \forall\,\phi_h \in \mathrm{H}_h^{\phi},$$

and realize that (4.4) can be rewritten, equivalently, as: find $\phi_h \in \mathrm{H}_h^{\phi}$ such that

$$\mathbf{T}_h(\phi_h) = \phi_h. \tag{4.7}$$

Certainly, all the above makes sense if we guarantee that the discrete problems (4.5) and (4.6) are well-posed. Indeed, it is easy to see that the respective proofs are almost verbatim of the continuous analogues provided in Sec. 3.3, and hence we simply state the corresponding results as follows.

Lemma 4.1. *Assume that* $\kappa_1 \in (0, \frac{2\delta\mu_1}{\mu_2})$ *and* $\kappa_2 \in (0, \frac{2\widetilde{\delta}\alpha_{\boldsymbol{K}}}{n\|\boldsymbol{K}^{-1}\|_\infty})$, *with* $\delta \in (0, 2\mu_1)$ *and* $\widetilde{\delta} \in (0, \frac{2}{n\|\boldsymbol{K}^{-1}\|_\infty})$. *Then, for each* $\phi_h \in \mathrm{H}_h^{\phi}$ *the problem* (4.5) *has a unique* *solution* $\mathbf{S}_h(\phi_h) := (\boldsymbol{\sigma}_h, \boldsymbol{u}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}$. *Moreover, with the same constant* $C_S > 0$ *from Lemma* 3.3, *there holds*

$$\|\mathbf{S}_h(\phi_h)\|_H = \|(\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\|_H \leq C_S\{\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + \|\phi_h\|_{0,\Omega}\|\boldsymbol{f}\|_{\infty,\Omega}\}, \quad \forall\, \phi_h \in \mathrm{H}_h^{\phi}.$$

Lemma 4.2. *Let* $\phi_h \in \mathrm{H}_h^{\phi}$ *and* $\boldsymbol{u}_h \in \mathbf{H}_h^{\boldsymbol{u}}$ *such that* $\|\boldsymbol{u}_h\|_{1,\Omega} < \frac{\vartheta_1}{2c_p c(\Omega)}$ (*cf.* (2.3), (3.1), (3.6))*. Then, there exists a unique* $\widetilde{\phi}_h := \widetilde{\mathbf{S}}_h(\phi_h, \boldsymbol{u}_h) \in \mathrm{H}_h^{\phi}$ *solution of* (4.6)*. Moreover, with the same constant* $C_{\widetilde{S}} > 0$ *from Lemma* 3.4, *there holds*

$$\|\widetilde{\mathbf{S}}_h(\phi_h, \boldsymbol{u}_h)\|_{1,\Omega} = \|\widetilde{\phi}_h\|_{1,\Omega} \leq C_{\widetilde{S}}\{\gamma_2|\Omega|^{1/2}|\boldsymbol{k}| + \|g\|_{0,\Omega}\}.$$

We now aim to show the solvability of (4.4) by analyzing the equivalent fixed point equation (4.7). To this end, in what follows we verify the hypotheses of the Brouwer fixed point theorem (cf. Theorem 9.9-2 in Ref. 12). We begin with the discrete version of Lemma 3.5.

Lemma 4.3. *Given* $r > 0$, *we let* $W_h := \{\phi_h \in \mathrm{H}_h^{\phi} : \|\phi_h\|_{1,\Omega} \leq r\}$, *and assume* (3.26) (*cf. Lemma* 3.5)*. Then* $\mathbf{T}_h(W_h) \subseteq W_h$*.*

Proof. It is a straightforward consequence of Lemmas 4.1 and 4.2. □

The discrete analogues of Lemmas 3.6 and 3.7 are provided next. We notice in advance that, instead of the regularity assumptions employed in the proof of those results, which actually are not needed nor could be applied in the present discrete case, we simply utilize a $\mathrm{L}^4 - \mathrm{L}^4 - \mathrm{L}^2$ argument.

Lemma 4.4. *There exist a positive constant* C, *depending on* $\mu_1, \kappa_1, \kappa_2, L_\mu$ *and* $\alpha(\Omega)$ (*cf.* (2.3), (3.7), (3.25), (3.19)), *such that*

$$\begin{aligned}&\|\mathbf{S}_h(\phi_h) - \mathbf{S}_h(\varphi_h)\|_H \\ &\quad \leq C\{\|\boldsymbol{f}\|_{\infty,\Omega}\|\phi_h - \varphi_h\|_{0,\Omega} + \|\mathbf{S}_{1,h}(\varphi_h)\|_{\mathbb{L}^4(\Omega)}\|\phi_h - \varphi_h\|_{\mathrm{L}^4(\Omega)}\},\end{aligned}$$

for all $\phi_h, \varphi_h \in \mathrm{H}_h^{\phi}$.

Proof. It proceeds exactly as in the proof of Lemma 3.6 (see Lemma 3.9 in Ref. 2), except for the derivation of the discrete analogue of Eq. (3.29), Lemma 3.9 in Ref. 2, where, instead of choosing the values of p and q determined by the regularity parameter δ, it suffices to take $p = q = 2$, thus obtaining

$$\begin{aligned}&|(B_{\varphi_h} - B_{\phi_h})((\boldsymbol{\zeta}_h, \boldsymbol{w}_h), (\sigma_h, \boldsymbol{u}_h) - (\boldsymbol{\zeta}_h, \boldsymbol{w}_h))| \\ &\quad \leq \frac{L_\mu(1+\kappa_1^2)^{1/2}}{\mu_1^2}\|\boldsymbol{\zeta}_h\|_{\mathbb{L}^4(\Omega)}\|\phi_h - \varphi_h\|_{\mathrm{L}^4(\Omega)}\|(\boldsymbol{\sigma}_h, \boldsymbol{u}_h) - (\boldsymbol{\zeta}_h, \boldsymbol{w}_h)\|_H,\end{aligned}$$

for all $\phi_h, \varphi_h \in \mathrm{H}_h^{\phi}$, with $(\sigma_h, \boldsymbol{u}_h) := \mathbf{S}_h(\phi_h)$ and $(\boldsymbol{\zeta}_h, \boldsymbol{w}_h) := \mathbf{S}_h(\varphi)$. Thus, the fact that the elements of $\mathbb{H}_h^{\boldsymbol{\sigma}}$ are piecewise polynomials insures that $\|\boldsymbol{\zeta}_h\|_{\mathbb{L}^4(\Omega)} < +\infty$ for each $\boldsymbol{\zeta}_h \in \mathbb{H}_h^{\boldsymbol{\sigma}}$. Further details are omitted. □

Lemma 4.5. *Let* $C_{\widetilde{S}}$ *be the constant provided by Lemma* 3.4. *Then, there exists a positive constant* $\widetilde{C}$, *depending on* $C_{\widetilde{S}}, c(\Omega), L_f$ *and* L_ϑ (*cf.* (3.6), (3.25)), *such that for all* $(\phi_h, \boldsymbol{u}_h), (\varphi_h, \boldsymbol{w}_h) \in \mathrm{H}_h^\phi \times \mathbf{H}_h^{\boldsymbol{u}}$, *with* $\|\boldsymbol{u}_h\|_{1,\Omega}, \|\boldsymbol{w}_h\|_{1,\Omega} < \frac{\vartheta_1}{2c_p c(\Omega)}$, *there holds*

$$\begin{aligned}
&\|\widetilde{\mathbf{S}}_h(\phi_h, \boldsymbol{u}_h) - \widetilde{\mathbf{S}}_h(\varphi_h, \boldsymbol{w}_h)\|_{1,\Omega} \\
&\quad \leq \widetilde{C}\{|\boldsymbol{k}|\|\phi_h - \varphi_h\|_{0,\Omega} + \|\widetilde{\mathbf{S}}_h(\varphi_h, \boldsymbol{w}_h)\|_{1,\Omega}\|\boldsymbol{u}_h - \boldsymbol{w}_h\|_{1,\Omega} \\
&\qquad + \|\nabla\widetilde{\mathbf{S}}_h(\varphi_h, \boldsymbol{w}_h)\|_{\mathrm{L}^4(\Omega)}\|\phi_h - \varphi_h\|_{\mathrm{L}^4(\Omega)}\}. \qquad (4.8)
\end{aligned}$$

Proof. Given $(\phi_h, \boldsymbol{u}_h)$ and $(\varphi_h, \boldsymbol{w}_h)$ as stated, we first let $\widetilde{\phi}_h := \widetilde{\mathbf{S}}_h(\phi_h, \boldsymbol{u}_h)$ and $\widetilde{\varphi}_h := \widetilde{\mathbf{S}}_h(\varphi_h, \boldsymbol{w}_h)$. Next, we proceed analogously as in the proof of Lemma 3.7, except for the derivation of the discrete analogue of the third term in (3.29), where, employing the same argument of the previous Lemma 4.4, it suffices to take $p = q = 2$, thus obtaining

$$\begin{aligned}
\widetilde{\alpha}\|\widetilde{\phi}_h - \widetilde{\varphi}_h\|_{1,\Omega}^2 &\leq \{L_f|\boldsymbol{k}|\|\phi_h - \varphi_h\|_{0,\Omega} + c(\Omega)\|\widetilde{\varphi}_h\|_{1,\Omega}\|\boldsymbol{u}_h - \boldsymbol{w}_h\|_{1,\Omega}\}|\widetilde{\phi}_h - \widetilde{\varphi}_h|_{1,\Omega} \\
&\quad + L_\vartheta\|\phi_h - \varphi_h\|_{\mathrm{L}^4(\Omega)}\|\nabla\widetilde{\varphi}_h\|_{\mathrm{L}^4(\Omega)}|\widetilde{\phi}_h - \widetilde{\varphi}_h|_{1,\Omega}.
\end{aligned}$$

Then, since the elements of H_h^ϕ are piecewise polynomials it follows that $\|\nabla\widetilde{\varphi}_h\|_{\mathrm{L}^4(\Omega)} < +\infty$, and hence the foregoing equation yields (4.8). Further details are omitted. □

Now, utilizing Lemmas 4.4 and 4.5, we can prove the discrete version of Lemma 3.8.

Lemma 4.6. *Suppose that the assumptions in Lemma* 4.3 *are satisfied. Then, with the constants* C *and* $\widetilde{C}$ *from Lemmas* 4.4 *and* 4.5, *for all* $\phi_h, \varphi_h \in \mathrm{H}_h^\phi$ *there holds*

$$\begin{aligned}
&\|\mathbf{T}_h(\phi_h) - \mathbf{T}_h(\varphi_h)\|_{1,\Omega} \\
&\quad \leq \widetilde{C}\{|\boldsymbol{k}| + C\|\mathbf{T}_h(\varphi_h)\|_{1,\Omega}\|\boldsymbol{f}\|_{\infty,\Omega}\}\|\phi_h - \varphi_h\|_{0,\Omega} \\
&\qquad + \widetilde{C}\{C\|\mathbf{T}_h(\varphi_h)\|_{1,\Omega}\|\mathbf{S}_{1,h}(\varphi_h)\|_{\mathbb{L}^4(\Omega)} + \|\nabla\mathbf{T}_h(\varphi_h)\|_{\mathrm{L}^4(\Omega)}\}\|\phi_h - \varphi_h\|_{\mathrm{L}^4(\Omega)}. \qquad (4.9)
\end{aligned}$$

Consequently, since the foregoing lemma and the continuous injection of $\mathrm{H}^1(\Omega)$ into $\mathrm{L}^4(\Omega)$ confirm the continuity of $\mathbf{T}_h$, we conclude, thanks to the Brouwer fixed point theorem (cf. Theorem 9.9-2 in Ref. 12) and Lemmas 4.3 and 4.6, the main result of this section.

Theorem 4.7. *Under the assumptions of Lemma* 4.3, *the Galerkin scheme* (4.4) *has at least one solution* $(\boldsymbol{\sigma}_h, \boldsymbol{u}_h, \phi_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}} \times \mathrm{H}_h^\phi$ *with* $\phi_h \in W_h$, *and there holds*

$$\|(\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\|_H \leq C_S\{\|\boldsymbol{u}_\mathrm{D}\|_{1/2,\Gamma_\mathrm{D}} + |\boldsymbol{k}|\|\phi_h\|_{1,\Omega}\},$$

and

$$\|\phi_h\|_{1,\Omega} \leq C_{\widetilde{S}}\{\gamma_2|\Omega|^{1/2}|\boldsymbol{k}| + \|g\|_{0,\Omega}\},$$

where C_S *and* $C_{\widetilde{S}}$ *are the constants provided by Lemmas* 3.3 *and* 3.4, *respectively.*

We end this section by remarking that the lack of suitable estimates for $\|\mathbf{S}_{1,h}(\varphi_h)\|_{\mathbb{L}^4(\Omega)}$ and $\|\nabla \mathbf{T}_h(\varphi_h)\|_{\mathrm{L}^4(\Omega)}$ stops us of trying to use (4.9) to derive a contraction estimate for $\mathbf{T}_h$. This is the reason why in the foregoing Theorem 4.7 we are able only to guarantee existence, but not uniqueness, of a discrete solution.

5. *A Priori* Error Analysis

Given $(\boldsymbol{\sigma}, \boldsymbol{u}, \phi) \in \mathbb{H}_N(\mathbf{div};\Omega) \times \mathbf{H}^1(\Omega) \times \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)$ with $\phi \in W$, and $(\boldsymbol{\sigma}_h, \boldsymbol{u}_h, \phi_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}} \times \mathrm{H}_h^{\phi}$ with $\phi_h \in W_h$, solutions of (3.11) and (4.4), respectively, we now aim to derive a corresponding *a priori* error estimate. For this purpose, we first observe from (3.11) and (4.4) that the above problems can be rewritten as two pairs of corresponding continuous and discrete formulations, namely

$$\begin{aligned} B_\phi((\boldsymbol{\sigma}, \boldsymbol{u}), (\boldsymbol{\tau}, \boldsymbol{v})) &= F_\phi(\boldsymbol{\tau}, \boldsymbol{v}) \quad \forall\, (\boldsymbol{\tau}, \boldsymbol{v}) \in \mathbb{H}_N(\mathbf{div};\Omega) \times \mathbf{H}^1(\Omega), \\ B_{\phi_h}((\boldsymbol{\sigma}_h, \boldsymbol{u}_h), (\boldsymbol{\tau}_h, \boldsymbol{v}_h)) &= F_{\phi_h}(\boldsymbol{\tau}_h, \boldsymbol{v}_h), \quad \forall\, (\boldsymbol{\tau}_h, \boldsymbol{v}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}, \end{aligned} \tag{5.1}$$

and

$$\begin{aligned} A_{\boldsymbol{u}}(\phi, \psi) &= G_\phi(\psi) \quad \forall\, \psi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega), \\ A_{\boldsymbol{u}_h}(\phi_h, \psi_h) &= G_{\phi_h}(\psi_h) \quad \forall\, \psi_h \in \mathrm{H}_h^{\phi}. \end{aligned} \tag{5.2}$$

Then, as suggested by the structure of the foregoing systems, in what follows we apply a suitable Strang-type lemma valid for linear problems to (5.1), and then derive our own Strang-type estimate for (5.2). The reason of the latter is that the present form $A_{\boldsymbol{u}}$ is not strongly monotone as it was in Ref. 2 where ϑ depended on $|\nabla\phi|$ instead of just ϕ, and hence it does not fit the corresponding Strang-type estimates for nonlinear problems (see, e.g. Lemma 5.1 in Ref. 2).

We begin our analysis by recalling from Ref. 11 the first Strang lemma for linear problems.

Lemma 5.1. *Let H be a Hilbert space, $F \in H'$, and $A : H \times H \to \mathrm{R}$ a bounded and elliptic bilinear form. In addition, let $\{H_n\}_{n\in\mathrm{N}}$ be a sequence of finite-dimensional subspaces of H, and for each $n \in \mathrm{N}$ consider a functional $F_n \in H_n'$ and a bounded bilinear form $A_n : H_n \times H_n \to \mathrm{R}$. Assume that the family $\{A\} \cup \{A_h\}_{n\in\mathrm{N}}$ is uniformly bounded and uniformly elliptic with constants Λ_B and Λ_E, respectively. In turn, let $u \in H$ and $u_n \in H_n$ such that*

$$A(u, v) = F(v), \quad \forall\, v \in H \quad \text{and} \quad A_n(u_n, v_n) = F_n(v_n), \quad \forall\, v_n \in H_n.$$

Then for each $n \in \mathrm{N}$ *there holds*

$$\|u - u_n\|_H \leq C_{\mathtt{ST}} \left\{ \sup_{\substack{w_n \in H_n \\ w_n \neq \mathbf{0}}} \frac{|F(w_n) - F_n(w_n)|}{\|w_n\|_H} \right.$$

$$\left. + \inf_{\substack{v_n \in H_n \\ v_n \neq \mathbf{0}}} \left(\|u - v_n\|_H + \sup_{\substack{w_n \in H_n \\ w_n \neq \mathbf{0}}} \frac{|A(v_n, w_n) - A_n(v_n, w_n)|}{\|w_n\|_H} \right) \right\},$$

with $C_{\mathtt{ST}} := \Lambda_E^{-1} \max\{1, \Lambda_E + \Lambda_B\}$.

Proof. See Lemma 4.1.1 in Ref. 11. □

We now denote as usual:

$$\mathrm{dist}(\phi, \mathrm{H}_h^\phi) := \inf_{\varphi_h \in \mathrm{H}_h^\phi} \|\phi - \varphi_h\|_{1,\Omega},$$

and

$$\mathrm{dist}((\boldsymbol{\sigma}, \boldsymbol{u}), \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}) := \inf_{(\boldsymbol{\tau}_h, \boldsymbol{v}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}} \|(\boldsymbol{\sigma}, \boldsymbol{u}) - (\boldsymbol{\tau}_h, \boldsymbol{v}_h)\|_H.$$

The following lemma provides a preliminary estimate for the error $\|(\boldsymbol{\sigma}, \boldsymbol{u}) - (\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\|_H$.

Lemma 5.2. *Let* $C_{\mathtt{ST}} := \alpha^{-1}(\Omega) \max\{1, \alpha(\Omega) + \|B\|\}$, *where* $\|B\|$ *and* $\alpha(\Omega)$ *are the boundedness and ellipticity constants*, *respectively*, *of the bilinear forms* B_ϕ (*cf.* (3.17), (3.19)). *Then there holds*

$$\begin{aligned} \|(\boldsymbol{\sigma}, \boldsymbol{u}) - (\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\|_H \leq C_{\mathtt{ST}} \Big\{ & (1 + 2\|B\|) \mathrm{dist}((\boldsymbol{\sigma}, \boldsymbol{u}), \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}) \\ & + (1 + \kappa_2^2)^{1/2} \|\boldsymbol{f}\|_{\infty,\Omega} \|\phi - \phi_h\|_{0,\Omega} \\ & + \frac{L_\mu (1 + \kappa_1^2)^{1/2}}{\mu_1^2} C_\delta \|\boldsymbol{\sigma}\|_{\delta,\Omega} \|\phi - \phi_h\|_{\mathrm{L}^{n/\delta}(\Omega)} \Big\}. \end{aligned} \tag{5.3}$$

Proof. By applying Lemma 5.1 to the context (5.1), we obtain

$$\|(\boldsymbol{\sigma}, \boldsymbol{u}) - (\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\|_H$$

$$\leq C_{\mathtt{ST}} \left\{ \sup_{\substack{(\boldsymbol{\tau}_h, \boldsymbol{v}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}} \\ (\boldsymbol{\tau}_h, \boldsymbol{v}_h) \neq \mathbf{0}}} \frac{|F_\phi(\boldsymbol{\tau}_h, \boldsymbol{v}_h) - F_{\phi_h}(\boldsymbol{\tau}_h, \boldsymbol{v}_h)|}{\|(\boldsymbol{\tau}_h, \boldsymbol{v}_h)\|} \right.$$

$$+ \inf_{\substack{(\boldsymbol{\zeta}_h,\boldsymbol{w}_h)\in\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}}\\(\boldsymbol{\zeta}_h,\boldsymbol{w}_h)\neq\mathbf{0}}} \left(\|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta}_h,\boldsymbol{w}_h)\|_H \right.$$

$$\left. + \sup_{\substack{(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\in\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}}\\(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\neq\mathbf{0}}} \frac{|B_\phi((\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h)) - B_{\phi_h}((\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h))|}{\|(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\|_H} \right) \Bigg\}. \tag{5.4}$$

Then, proceeding analogously as in the proof of Lemma 3.9 in Ref. 2, we easily deduce that

$$\sup_{\substack{(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\in\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}}\\(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\neq\mathbf{0}}} \frac{|F_\phi(\boldsymbol{\tau}_h,\boldsymbol{v}_h) - F_{\phi_h}(\boldsymbol{\tau}_h,\boldsymbol{v}_h)|_H}{\|(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\|_H} \leq (1+\kappa_2^2)^{1/2}\|\boldsymbol{f}\|_{\infty,\Omega}\|\phi-\phi_h\|_{0,\Omega}. \tag{5.5}$$

In turn, in order to estimate the supremum in (5.4), we add and subtract suitable terms to write

$$\begin{aligned} &B_\phi((\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h)) - B_{\phi_h}((\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h)) \\ &\quad = B_\phi((\boldsymbol{\zeta}_h,\boldsymbol{w}_h)-(\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau}_h,\boldsymbol{v}_h)) + (B_\phi - B_{\phi_h})((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau}_h,\boldsymbol{v}_h)) \\ &\quad\quad + B_{\phi_h}((\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h)), \end{aligned}$$

whence, applying the boundedness (3.17) to the first and third terms on the right-hand side of the foregoing equation, and proceeding analogously as for the derivation of Eqs. (3.29), (3.30) in Ref. 2 with the second one, we find that

$$\begin{aligned} &\sup_{\substack{(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\in\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}}\\(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\neq\mathbf{0}}} \frac{|B_\phi((\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h)) - B_{\phi_h}((\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h))|}{\|(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\|} \\ &\quad \leq 2\|B\|\|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta}_h,\boldsymbol{w}_h)\|_H + \frac{L_\mu(1+\kappa_1^2)^{1/2}}{\mu_1^2}C_\delta\|\boldsymbol{\sigma}\|_{\delta,\Omega}\|\phi-\phi_h\|_{\mathrm{L}^{n/\delta}(\Omega)}. \end{aligned} \tag{5.6}$$

Finally, by replacing the inequalities (5.5) and (5.6) into (5.4), we get (5.3), which ends the proof. □

Next, we have the following result concerning $\|\phi-\phi_h\|_{1,\Omega}$. To this end, and in order to simplify the subsequent writing, we introduce the following constants, independent of the data $\boldsymbol{k}$, g, $\boldsymbol{u}_D$ and $\boldsymbol{f}$,

$$K_1 := C_{\widetilde{S}}\{L_f + L_\vartheta C_\delta \widetilde{C}_\delta \widehat{C}_{\widetilde{S}}(r)\gamma_2|\Omega|^{1/2}\}, \quad K_2 := C_{\widetilde{S}}L_\vartheta C_\delta \widetilde{C}_\delta \widehat{C}_{\widetilde{S}}(r),$$

$$K_3 := 1 + C_{\widetilde{S}}(\vartheta_2+\beta), \quad K_4 := 3C_{\widetilde{S}}c(\Omega)C_S \quad \text{and} \quad K_5 := C_{\widetilde{S}}c(\Omega)r,$$

where C_δ is the boundedness constant of the continuous injection $i_\delta : \mathrm{H}^\delta(\Omega) \to \mathrm{L}^{\delta^*}(\Omega)$, with δ^* given by (3.30), and $\widetilde{C}_\delta$ is the boundedness constant of the compact injection $i : \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega) \to \mathrm{L}^{n/\delta}(\Omega)$.

Lemma 5.3. *Assume that the data* $\boldsymbol{k}$ *and* g *satisfy*

$$K_1|\boldsymbol{k}| + K_2\|g\|_{\delta,\Omega} \le \frac{1}{2}. \tag{5.7}$$

Then, there holds

$$\|\phi - \phi_h\|_{1,\Omega} \le \widetilde{K}_3(\boldsymbol{u}_D, \boldsymbol{f})\,\mathrm{dist}(\phi, \mathrm{H}_h^\phi) + \widetilde{K}_5\|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,\Omega}, \tag{5.8}$$

where

$$\widetilde{K}_3(\boldsymbol{u}_D, \boldsymbol{f}) := 2\{K_3 + K_4(\|\boldsymbol{u}_D\|_{1/2,\Gamma_\mathrm{D}} + r\|\boldsymbol{f}\|_{\infty,\Omega})\} \quad \textit{and} \quad \widetilde{K}_5 := 2K_5. \tag{5.9}$$

Proof. We first observe by triangle inequality that

$$\|\phi - \phi_h\|_{1,\Omega} \le \|\phi - \varphi_h\|_{1,\Omega} + \|\phi_h - \varphi_h\|_{1,\Omega}, \quad \forall\, \varphi_h \in \mathrm{H}_h^\phi. \tag{5.10}$$

Then employing the ellipticity of the bilinear form $A_{\phi_h,\boldsymbol{u}_h}$ with constant $\widetilde{\alpha}$, using that (cf. (5.2)) $A_{\phi_h,\boldsymbol{u}_h}(\phi_h, \phi_h - \varphi_h) = A_{\boldsymbol{u}_h}(\phi_h, \phi_h - \varphi_h) = G_{\phi_h}(\phi_h - \varphi_h)$, and then adding and subtracting the expression (cf. (5.2)) $G_\phi(\phi_h - \varphi_h) = A_{\boldsymbol{u}}(\phi, \phi_h - \varphi_h) = A_{\phi,\boldsymbol{u}}(\phi, \phi_h - \varphi_h)$, we deduce that

$$\begin{aligned} \widetilde{\alpha}\|\phi_h - \varphi_h\|_{1,\Omega}^2 &\le A_{\phi_h,\boldsymbol{u}_h}(\phi_h - \varphi_h, \phi_h - \varphi_h) \\ &\le |G_{\phi_h}(\phi_h - \varphi_h) - G_\phi(\phi_h - \varphi_h)| \\ &\quad + |A_{\phi,\boldsymbol{u}}(\phi, \phi_h - \varphi_h) - A_{\phi_h,\boldsymbol{u}_h}(\varphi_h, \phi_h - \varphi_h)|. \end{aligned} \tag{5.11}$$

Next, according to the definition of G_ϕ (cf. (3.14)), and applying Cauchy–Schwarz inequality, we get

$$|G_{\phi_h}(\phi_h - \varphi_h) - G_\phi(\phi_h - \varphi_h)| \le L_f|\boldsymbol{k}|\|\phi - \phi_h\|_{0,\Omega}|\phi_h - \varphi_h|_{1,\Omega}. \tag{5.12}$$

In turn, adding and subtracting $\vartheta(\phi_h)$ and $\boldsymbol{u}$ within appropriate expressions of $A_{\phi,\boldsymbol{u}}(\phi, \phi_h - \varphi_h)$ and $A_{\phi_h,\boldsymbol{u}_h}(\varphi_h, \phi_h - \varphi_h)$, respectively, and then applying Hölder's inequality, the upper bound of ϑ (cf. (2.3)), and (3.6), we find that

$$\begin{aligned} &|A_{\phi,\boldsymbol{u}}(\phi, \phi_h - \varphi_h) - A_{\phi_h,\boldsymbol{u}_h}(\varphi_h, \phi_h - \varphi_h)| \\ &\quad \le L_\vartheta\|\phi - \phi_h\|_{\mathrm{L}^{2q}(\Omega)}\|\nabla\phi\|_{\mathrm{L}^{2p}(\Omega)}|\phi_h - \varphi_h|_{1,\Omega} + \vartheta_2|\phi - \varphi_h|_{1,\Omega}|\phi_h - \varphi_h|_{1,\Omega} \\ &\qquad + c(\Omega)\|\varphi_h\|_{1,\Omega}\|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,\Omega}|\phi_h - \varphi_h|_{1,\Omega} \\ &\qquad + c(\Omega)\|\phi - \varphi_h\|_{1,\Omega}\|\boldsymbol{u}\|_{1,\Omega}|\phi_h - \varphi_h|_{1,\Omega} + \beta\|\phi - \varphi_h\|_{0,\Omega}\|\phi_h - \varphi_h\|_{0,\Omega}, \end{aligned} \tag{5.13}$$

where $p, q \in [1, +\infty)$ are such that $1/p + 1/q = 1$. In this way, using the Sobolev embedding theorem (cf. Theorem 4.12 in Ref. 1, Theorem 1.3.4 in Ref. 28), the regularity estimate (3.24), and applying the same arguments used for the derivation of (3.29) (cf. proof of Lemma 3.7), in particular the fact that $\mathrm{H}^1(\Omega)$ is compactly, and hence continuously embedded in $\mathrm{L}^{n/\delta}(\Omega)$ with boundedness constant $\widetilde{C}_\delta$, it follows from (5.13) that

$$\begin{aligned}
&|A_{\phi,\boldsymbol{u}}(\phi,\phi_h-\varphi_h) - A_{\phi_h,\boldsymbol{u}_h}(\varphi_h,\phi_h-\varphi_h)| \\
&\quad\le L_\vartheta C_\delta \widetilde{C}_\delta \widehat{C}_{\widetilde{S}}(r)\{\gamma_2|\Omega|^{1/2}|\boldsymbol{k}| + \|g\|_{\delta,\Omega}\}\|\phi-\phi_h\|_{1,\Omega}\|\phi_h-\varphi_h\|_{1,\Omega} \\
&\qquad + \vartheta_2\|\phi-\varphi_h\|_{1,\Omega}\|\phi_h-\varphi_h\|_{1,\Omega} + c(\Omega)\|\varphi_h\|_{1,\Omega}\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}\|\phi_h-\varphi_h\|_{1,\Omega} \\
&\qquad + c(\Omega)\|\phi-\varphi_h\|_{1,\Omega}\|\boldsymbol{u}\|_{1,\Omega}\|\phi_h-\varphi_h\|_{1,\Omega} + \beta\|\phi-\varphi_h\|_{1,\Omega}\|\phi_h-\varphi_h\|_{1,\Omega}.
\end{aligned} \tag{5.14}$$

Thus, by replacing (5.12) and (5.14) into (5.11), and then the resulting estimate into (5.10), employing the constants defined previously to the statement of the present lemma, using that both $\|\boldsymbol{u}\|_{1,\Omega}$ and $\|\boldsymbol{u}_h\|_{1,\Omega}$ are bounded by $C_S\{\|\boldsymbol{u}_D\|_{1/2,\Gamma_{\mathrm{D}}} + r\|\boldsymbol{f}\|_{\infty,\Omega}\}$ (cf. Lemmas 3.3 and 4.1), and recalling from the proof of Lemma 3.4 that $\widetilde{\alpha} = C_{\widetilde{S}}^{-1}$, we find, after several algebraic manipulations, that

$$\begin{aligned}
\|\phi-\phi_h\|_{1,\Omega} &\le \{K_1|\boldsymbol{k}| + K_2\|g\|_{\delta,\Omega}\}\|\phi-\phi_h\|_{1,\Omega} \\
&\quad + \{K_3 + K_4(\|\boldsymbol{u}_D\|_{1/2,\Gamma_{\mathrm{D}}} + r\|\boldsymbol{f}\|_{\infty,\Omega})\}\|\phi-\varphi_h\|_{1,\Omega} \\
&\quad + K_5\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}, \quad \forall\,\varphi_h\in\mathrm{H}_h^\phi,
\end{aligned}$$

which, according to the assumption (5.7) and the notation (5.9), and taking the infimum on $\varphi_h \in \mathrm{H}_h^\phi$, yields (5.8) and completes the proof. □

We now combine the inequalities provided by Lemmas 5.2 and 5.3 to derive the Céa estimate for the total error $\|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H + \|\phi-\phi_h\|_{1,\Omega}$. To this end, we now introduce the constants

$$K_6 := C_{\mathtt{ST}}\frac{L_\mu(1+\kappa_1^2)^{1/2}}{\mu_1^2}C_S\widetilde{C}_\delta\widehat{C}_S(r),$$

and

$$K_7 := C_{\mathtt{ST}}(1+\kappa_2^2)^{1/2} + \frac{L_\mu(1+\kappa_1^2)^{1/2}}{\mu_1^2}C_S\widetilde{C}_\delta\widehat{C}_S(r)r.$$

Then, employing from (3.23) that $\|\boldsymbol{\sigma}\|_{\delta,\Omega} \le \widehat{C}_S(r)\{\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} + \|\phi\|_{0,\Omega}\|\boldsymbol{f}\|_{\infty,\Omega}\}$, recalling that $\|\phi\|_{1,\Omega} \le r$, and using that $\widetilde{C}_\delta$ is the boundedness constant of the continuous injection of $\mathrm{H}^1(\Omega)$ into $\mathrm{L}^{n/\delta}(\Omega)$, we can assert from (5.3) that

$$\begin{aligned}
\|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H &\le C_{\mathtt{ST}}(1+2\|B\|)\mathrm{dist}((\boldsymbol{\sigma},\boldsymbol{u}),\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}}) \\
&\quad + \{K_6\|\boldsymbol{u}_D\|_{1/2+\delta,\Gamma_{\mathrm{D}}} + K_7\|\boldsymbol{f}\|_{\infty,\Omega}\}\|\phi-\phi_h\|_{1,\Omega},
\end{aligned}$$

which, employing the estimate for $\|\phi-\phi_h\|_{1,\Omega}$ given by (5.8), implies

$$\begin{aligned}
\|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H &\le C_{\mathtt{ST}}(1+2\|B\|)\,\mathrm{dist}((\boldsymbol{\sigma},\boldsymbol{u}),\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}}) + \widetilde{K}_6(\boldsymbol{u}_D,\boldsymbol{f})\,\mathrm{dist}(\phi,\mathrm{H}_h^\phi) \\
&\quad + \widetilde{K}_5\{K_6\|\boldsymbol{u}_D\|_{1/2+\delta,\Gamma_{\mathrm{D}}} + K_7\|\boldsymbol{f}\|_{\infty,\Omega}\}\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega},
\end{aligned}$$

where

$$\widetilde{K}_6(\boldsymbol{u}_D, \boldsymbol{f}) := \widetilde{K}_3(\boldsymbol{u}_D, \boldsymbol{f})\{K_6\|\boldsymbol{u}_D\|_{1/2+\delta,\Gamma_{\mathrm{D}}} + K_7\|\boldsymbol{f}\|_{\infty,\Omega}\}.$$

In this way, assuming now that $\boldsymbol{u}_D$ and $\boldsymbol{f}$ satisfy

$$\widetilde{K}_5\{K_6\|\boldsymbol{u}_D\|_{1/2+\delta,\Gamma_{\mathrm{D}}} + K_7\|\boldsymbol{f}\|_{\infty,\Omega}\} \leq \frac{1}{2},$$

we conclude from the foregoing equations that

$$\|(\boldsymbol{\sigma}, \boldsymbol{u}) - (\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\|_H \leq 2C_{\mathtt{ST}}(1 + 2\|B\|)\,\mathrm{dist}((\boldsymbol{\sigma}, \boldsymbol{u}), \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}) + 2\widetilde{K}_6(\boldsymbol{u}_D, \boldsymbol{f})\,\mathrm{dist}(\phi, \mathrm{H}_h^{\phi}). \quad (5.15)$$

Consequently, we can establish the following result providing the complete Céa estimate.

Theorem 5.4. *Assume that the data* $\boldsymbol{k}, g, \boldsymbol{u}_D$ *and* $\boldsymbol{f}$ *are sufficiently small so that*

$$K_1|\boldsymbol{k}| + K_2\|g\|_{\delta,\Omega} \leq \frac{1}{2} \quad \text{and} \quad \widetilde{K}_5\{K_6\|\boldsymbol{u}_D\|_{1/2+\delta,\Gamma_{\mathrm{D}}} + K_7\|\boldsymbol{f}\|_{\infty,\Omega}\} \leq \frac{1}{2}.$$

Then, there exists a positive constant C, *independent of* h, *but depending on data, parameters, and other constants, such that*

$$\|(\boldsymbol{\sigma}, \boldsymbol{u}) - (\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\|_H + \|\phi - \phi_h\|_{1,\Omega} \leq C\{\mathrm{dist}((\boldsymbol{\sigma}, \boldsymbol{u}), \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}) + \mathrm{dist}(\phi, \mathrm{H}_h^{\phi})\}. \quad (5.16)$$

Proof. It follows straightforwardly from (5.15) and (5.3). □

We end this section with the corresponding rates of convergence of our Galerkin scheme (4.4).

Theorem 5.5. *In addition to the hypotheses of Theorems* 3.10, 4.7 *and* 5.4, *assume that there exists* $s > 0$ *such that* $\boldsymbol{\sigma} \in \mathbb{H}^s(\Omega), \mathbf{div}\boldsymbol{\sigma} \in \mathbf{H}^s(\Omega), \boldsymbol{u} \in \mathbf{H}^{1+s}(\Omega)$ *and* $\phi \in \mathrm{H}^{1+s}(\Omega)$. *Then, there exists* $\widehat{C} > 0$, *independent of* h, *such that, with the finite element subspaces defined by* (4.1)–(4.3), *there holds*

$$\|(\boldsymbol{\sigma}, \boldsymbol{u}) - (\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\|_H + \|\phi - \phi_h\|_{1,\Omega} \leq \widehat{C}h^{\min\{s,k+1\}}\{\|\boldsymbol{\sigma}\|_{s,\Omega} + \|\mathbf{div}\,\boldsymbol{\sigma}\|_{s,\Omega} + \|\boldsymbol{u}\|_{1+s,\Omega} + \|\phi\|_{1+s,\Omega}\}.$$

Proof. It follows directly from the Céa estimate (5.16) and the approximation properties of $\mathbb{H}_h^{\boldsymbol{\sigma}}$, $\mathbf{H}_h^{\boldsymbol{u}}$ and H_h^{ϕ} (cf. Refs. 5, 11 and 21). □

6. Numerical Tests

Example 1. Our first example aims at testing the accuracy of our augmented finite element formulation. As usual, experimental errors and convergence rates are defined as

$$e(\boldsymbol{\sigma}) := \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathbf{div},\Omega}, \quad e(\boldsymbol{u}) := \|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,\Omega},$$
$$e(\phi) := \|\phi - \phi_h\|_{1,\Omega}, \quad r(\cdot) := \log(e(\cdot)/\widehat{e}(\cdot))[\log(h/\widehat{h})]^{-1},$$

where e and $\widehat{e}$ stand for errors computed on two consecutive meshes of sizes h and $\widehat{h}$, respectively. In all examples we consider $\boldsymbol{K} = K\mathbb{I}$, with K constant. The following exact solution to (2.1) defined on the unit disk is manufactured:

$$\phi(x_1, x_2) = c - c\exp(1 - x_1^2 - x_2^2),$$
$$\boldsymbol{u}(x_1, x_2) = \begin{pmatrix} \sin(\pi x_1)\cos(\pi x_2) \\ -\cos(\pi x_1)\sin(\pi x_2) \end{pmatrix},$$
$$\boldsymbol{\sigma}(x_1, x_2) = \mu(\phi)\boldsymbol{u} - (x_1^2 - x_2^2)\mathbb{I},$$

where $K^{-1} = 0.01$, $\beta = 10$, $\boldsymbol{k} = (0, -1)^{\mathrm{t}}$, $\mu(\phi) = (1 - b\phi)^{-2}$, $f_{\mathrm{bk}}(\phi) = b\phi(1 - b\phi)^2$, $\vartheta(\phi) = \phi + (1 - b\phi)^2$, and the source terms are

$$\boldsymbol{f}(x_1, x_2) = \phi^{-1}(K^{-1}\boldsymbol{u} - \mathbf{div}\,\boldsymbol{\sigma}),$$
$$g(x_1, x_2) = \beta\phi - \mathrm{div}(\vartheta(\phi)\nabla\phi) + \boldsymbol{u} \cdot \nabla\phi + f'_{\mathrm{bk}}(\phi)\boldsymbol{k} \cdot \nabla\phi,$$

for $(x_1, x_2) \in \overline{\Omega}$. We take $b = 1/2$, $c = 1/(1 - e)$ and set $\Gamma_{\mathrm{D}} = \partial\Omega$, where ϕ vanishes and the velocity is imposed accordingly to the exact solution. The mean value of $\mathrm{tr}(\boldsymbol{\sigma})_h$ over Ω is fixed via a penalization strategy. As defined above, the concentration is bounded in Ω and so are the concentration-dependent coefficients as well. In particular we have $\mu_1 = 1, \mu_2 = 4$ and as suggested by Lemma 3.3, the stabilization constants are chosen as $\kappa_1 = \frac{\delta\mu_1}{\mu_2}$ with $\delta = \mu_1$, and $\kappa_2 = 0.025$ for $\widetilde{\delta} = \frac{1}{4|K^{-1}|}$.

A Newton–Raphson algorithm with a fixed tolerance of 1e-6 has been used for the nonlinear problem (4.4). At each iteration the linear systems resulting from the linearization were solved by means of the multifrontal solver MUMPS. Independently of the refinement level, we observe that an average number of five steps was required to reach the desired tolerance. Values and plots of errors and corresponding rates associated to $\mathbf{RT}_k - \mathbf{P}_{k+1} - \mathrm{P}_{k+1}$ approximations with $k = 0$ and $k = 1$ are summarized in Table 1 and Fig. 1. The results show optimal asymptotic convergence rates for all fields (of order $k + 1$ for the pseudo-stress, the velocity and the concentration), which agree with the accuracy predicted in Theorem 5.5. We also remark that for both degrees of approximation, the concentration errors are always below the velocity errors, and both are dominated by the errors in the pseudo-stress approximation. The augmented mixed-primal approximations computed on a mesh of 204847 vertices and 409692 elements are depicted in Fig. 2, where stress, velocity, and concentration profiles are well resolved.

Table 1. Example 1: Convergence history and Newton iteration count for the mixed-primal $\mathbf{RT}_k - \mathbf{P}_{k+1} - \mathrm{P}_{k+1}$ approximations of the coupled problem, $k = 0, 1$. Here N_h stands for the number of degrees of freedom associated to each triangulation $\mathcal{T}_h$.

N_h	h	$e(\boldsymbol{\sigma})$	$r(\boldsymbol{\sigma})$	$e(\boldsymbol{u})$	$r(\boldsymbol{u})$	$e(\phi)$	$r(\phi)$	iter
			Augmented $\mathbf{RT}_0 - \mathbf{P}_1 - \mathrm{P}_1$ scheme					
45	1.000000	37.83630	—	5.078982	—	0.794267	—	7
150	0.752986	29.63322	0.861350	3.864984	0.962793	0.551649	1.284790	6
567	0.381608	14.53602	1.047988	1.893370	1.049950	0.261571	1.097927	6
1986	0.202981	7.685891	1.009446	0.917330	1.147899	0.142439	0.962789	5
7587	0.107277	3.855674	1.081757	0.449265	1.119414	0.071858	1.072957	5
29652	0.056293	1.929090	1.073920	0.222210	1.091739	0.036226	1.062147	5
116820	0.029796	0.967180	1.085224	0.111677	1.081444	0.018210	1.081158	5
465243	0.015539	0.480698	1.073919	0.056004	1.060163	0.009068	1.070848	5
1840545	0.008139	0.243228	1.053393	0.028080	1.067485	0.004619	1.043095	5
			Augmented $\mathbf{RT}_1 - \mathbf{P}_2 - \mathrm{P}_2$ scheme					
129	1.000000	32.06255	—	3.909169	—	0.549477	—	7
447	0.752986	16.36007	1.505829	1.686632	1.605829	0.163634	1.919349	6
2043	0.381608	6.518447	1.697863	0.374835	1.820361	0.040489	1.899365	5
6835	0.202981	2.511864	1.781622	0.097825	1.967394	0.011124	1.841454	5
26243	0.107277	0.695590	1.902700	0.026791	2.146465	0.002491	1.857032	5
104867	0.056293	0.209899	1.925679	0.006716	1.743774	0.000774	1.864551	5
412611	0.029796	0.056163	1.945132	0.001700	2.095484	0.000182	1.928363	5
1643907	0.015539	0.014448	1.960714	0.000427	2.132144	0.000039	1.987836	5

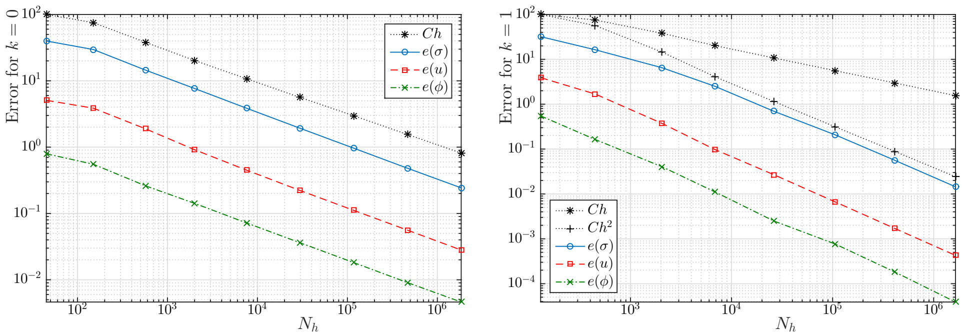

Fig. 1. Example 1: Computed errors associated to the mixed-primal approximation versus the number of degrees of freedom N_h for $\mathbf{RT}_0 - \mathbf{P}_1 - \mathrm{P}_1$ and $\mathbf{RT}_1 - \mathbf{P}_2 - \mathrm{P}_2$ finite elements (left and right, respectively). Values are detailed in Table 1.

For all remaining examples we stick to the case $k = 0$, i.e. row-wise lowest-order Raviart–Thomas finite element approximations for the Cauchy pseudo-stress, and piecewise linear approximations of velocity and concentration.

Example 2. Our next example corresponds to a test of batch sedimentation in a cylinder with a contraction (see e.g. Refs. 29 and 27). In this case the model parameters and concentration-dependent coefficients assume the values $K = 60$,

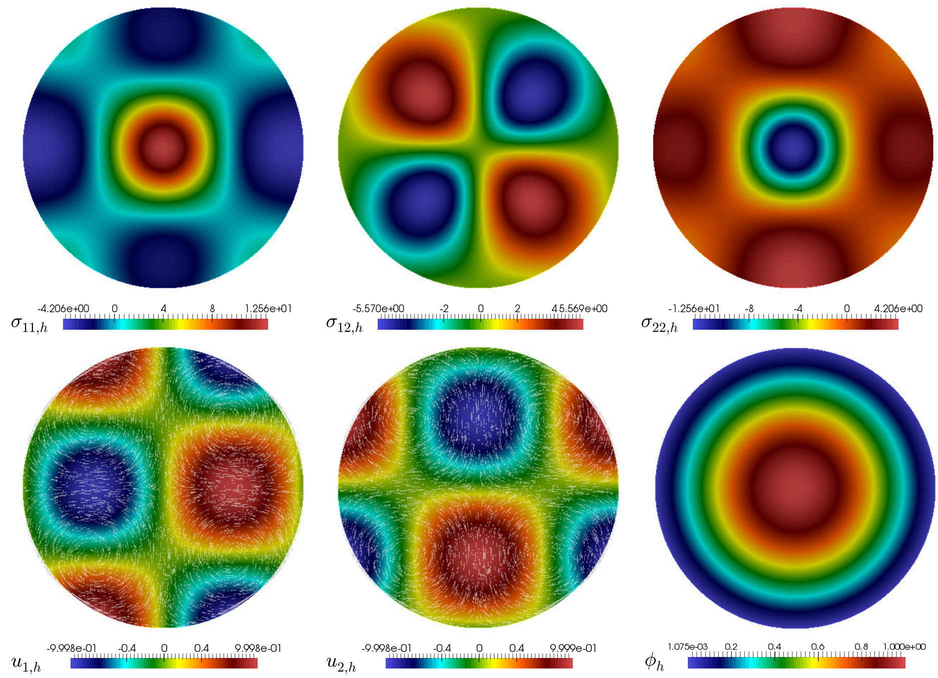

Fig. 2. Example 1: $\mathbf{RT}_0 - \mathbf{P}_1 - \mathrm{P}_1$ approximation of pseudo-stress components (top panels), velocity components with vector directions (bottom left and center, respectively), and concentration profile (bottom right), solutions to (3.11). The finest mesh has 204847 vertices and 409692 triangular elements.

$\beta = K^{-1}$, $\mu(\phi) = \mu_0(1 - \frac{\phi}{\phi_{\max}})^{-\eta}$, $f_{\mathrm{bk}}(\phi) = C_1 C_2 \frac{(1-\phi_0)}{\mu(\phi)}\phi$, $\vartheta(\phi) = C_1(C_3\gamma\phi^2 + C_2)$ and $\boldsymbol{f} = \Delta\rho G\boldsymbol{k}$, with $C_1 = (\rho_1 + \phi_{\max}\Delta\rho)(\rho_1\rho_2)^{-1}$, $C_2 = \frac{2}{9\mu_0}\Delta\rho G a^2$, $C_3 = 0.68355a^2$, $\Delta\rho = \rho_2 - \rho_1$, and $\gamma = 0.88$. Values for the remaining dimensional constants are collected in Table 2, and the model parameters yield the following stabilization constants $\mu_1 = \mu_0$, $\mu_2 = 6.5365$, $\kappa_1 = \delta\mu_1/\mu_2$ with $\delta = \mu_1$ and $\kappa_2 = 0.025$ for $\widetilde{\delta} = \frac{1}{4|K^{-1}|}$.

Table 2. Example 2: Model constants employed in the simulation of steady sedimentation of PMMA into glycerol/water within a contracted cylinder.

Quantity	Value
Density glycerol/water solution, ρ_1	$1.175\,\mathrm{g/cm^3}$
Density PMMA, ρ_2	$1.18\,\mathrm{g/cm^3}$
Viscosity glycerol/water solution, μ_0	$0.184\,\mathrm{g/cm^3}$
Initial volume fraction, ϕ_0	0.192
Maximum volume fraction, $\phi_{\max}$	0.64
Particle radius, a	$0.0397\,\mathrm{cm}$
Viscosity constant, η	1.82
Gravity, G	$980.665\,\mathrm{cm/s^2}$

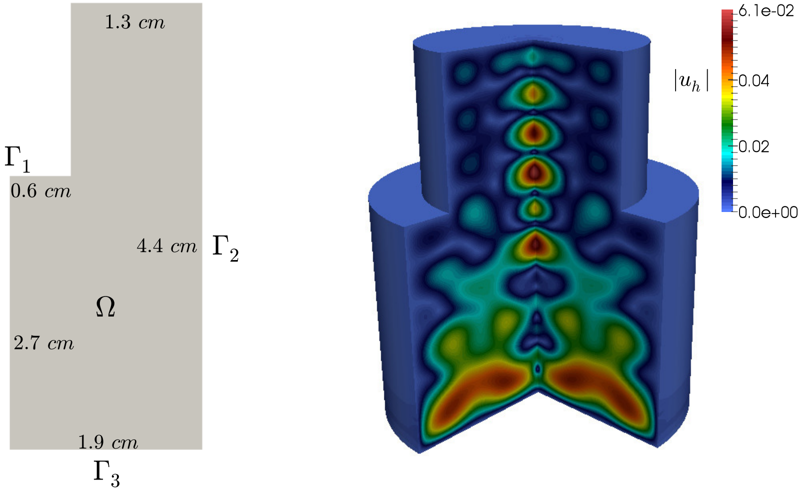

Fig. 3. Example 2: Typical dimensions and boundary setting for a two-dimensional computational domain representing the batch sedimentation within a cylinder with a contraction (left), and magnitude of the velocity field shown on the rotationally extruded domain (right).

On $\partial\Omega$ we impose zero-flux conditions for ϕ, that is $\tilde{\boldsymbol{\sigma}} \cdot \boldsymbol{\nu} = 0$. In addition, the following boundary conditions are imposed for the velocity (see sketch in Fig. 3): $\boldsymbol{u}|_{\Gamma_1} = \boldsymbol{u}|_{\Gamma_3} = \mathbf{0}$, and $u_2|_{\Gamma_2} = 0$ (representing a symmetry axis). The domain is discretized into 13131 vertices and 26260 triangles, and we represent the obtained field quantities of interest in Fig. 4. The maximum concentration has been packed at the bottom of the vessel, whereas throughout the rest of the domain is filled with low-concentration material. More interesting phenomena are observed from the velocity plots, where a main recirculation zone is observed at the center of the domain. Moreover, a countercurrent flow is observed along the symmetry axis (clearly identified in the horizontal velocity plot), and these flow patterns are further highlighted in the diagonal components of the pseudo-stress.

Example 3. Finally we turn to the simulation of the steady state of flow patterns on a box (see the domain, dimensions and boundary configuration illustrated in Fig. 5), using a modification to the single phase model described in Ref. 30 to reproduce the so-called Coandă effect, which corresponds to the tendency of a fluid jet to be attracted to a nearby surface.[31] In this case the nonlinear concentration-dependent coefficients are $\mu(\phi) = \mu_0(1 - \phi/\phi_{\max})^{\eta}$, $f_{\mathrm{bk}}(\phi) = u_{\infty}\phi(1 - \phi/\phi_{\max})^{\eta}$ and $\vartheta(\phi) = \vartheta_0(\phi^3 + 1)$ where $\eta = 1.82$, $\mu_0 = 0.02$, $\vartheta_0 = 0.0001$, $\beta = 0.01$, $K = 1000$, $g = 0$ and $\boldsymbol{f} = \Delta\rho G\boldsymbol{k}$, with $\Delta\rho = 0.0045$ and $G = 0.98$.

Concentration and velocities are fixed at the inlet surface Γ_{in} (a rectangle of width 0.5 cm and height 0.35 cm located on the top, at $x_1 = 0$) according to $\phi = \phi_{\mathrm{in}}$ and $\boldsymbol{u} = \boldsymbol{u}_{\mathrm{in}} = (u_{1,\mathrm{in}}, 0, 0)^{\mathtt{t}}$. At the outlet Γ_{out} (a rectangle with the same dimensions as the inlet, but located at $x_1 = 6$, on the bottom) we let the material exit

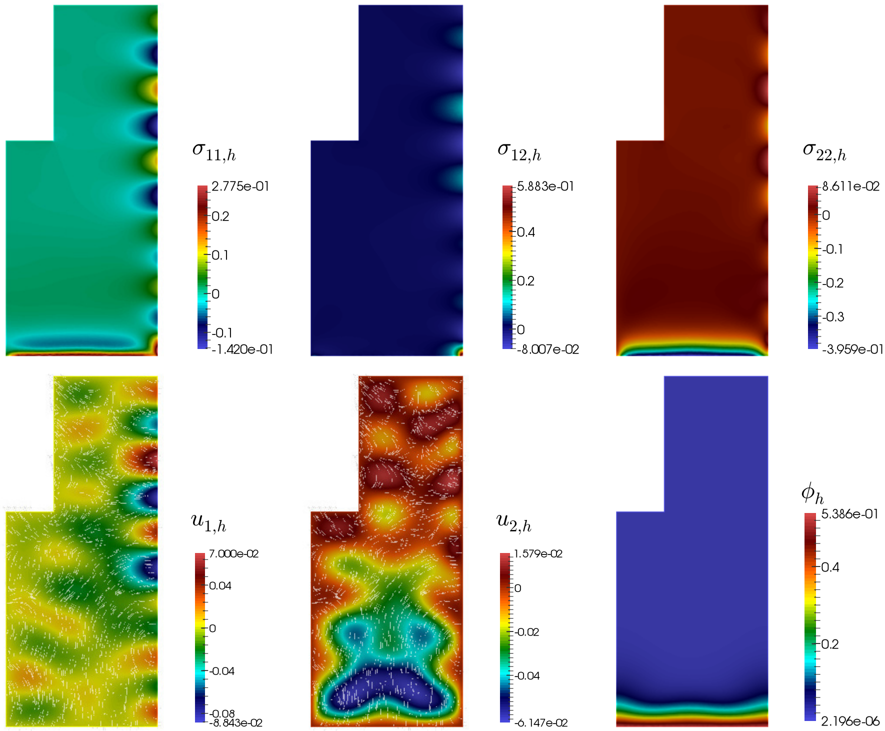

Fig. 4. Example 2: Principal components of the Cauchy pseudo-stress (top rows), velocity components $\boldsymbol{u}_h$ with vector directions (bottom left and center), and computed concentration ϕ_h (bottom right) for the test of batch sedimentation in a cylinder with a contraction.

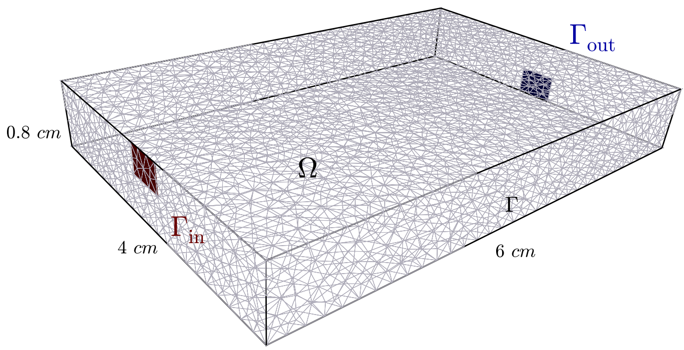

Fig. 5. Example 3: Sketch of the computational domain $\Omega = [0,6] \times [0,4] \times [0,0.8]$, a coarse mesh, and boundary setting, with $\partial\Omega = \Gamma \cup \Gamma_{\mathrm{in}} \cup \Gamma_{\mathrm{out}}$.

the domain with a velocity $\boldsymbol{u} = \boldsymbol{u}_{\mathrm{out}} = (u_{1,\mathrm{out}}, 0, 0)^{\mathtt{t}}$, but the concentration is not prescribed. On the remainder of $\partial\Omega$ we put no-slip boundary data for the velocity and zero-flux conditions for the concentration. Other model parameters are set as $u_{1,\mathrm{in}} = u_{1,\mathrm{out}} = 0.01$, $\phi_{\max} = 0.9$, $u_\infty = 0.0022$ and $\phi_{\mathrm{in}} = 0.3$.

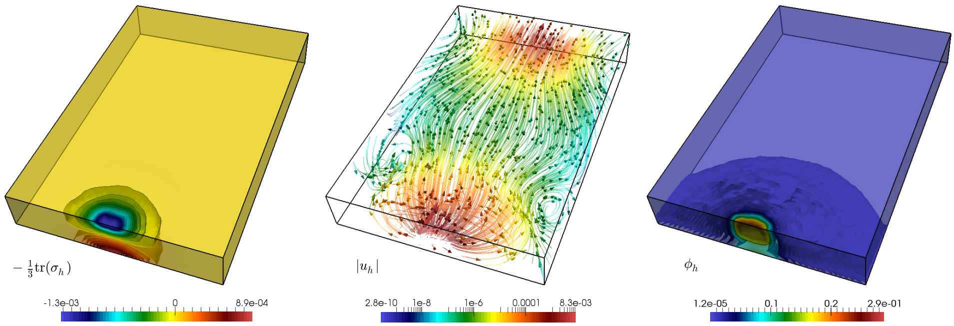

Fig. 6. Example 3: Approximate solutions to the so-called Coanda effect using an augmented mixed formulation. Trace of the Cauchy pseudo-stress tensor (left), velocity vectors and streamlines (middle), and concentration profile (right).

According to the bounds of the viscosity, the stabilization parameters were set as $\mu_1 = \mu_0$, $\mu_2 = \mu_0(1 - \phi_{\rm in}/\phi_{\max})^{-\eta}$, $\kappa_1 = \delta\frac{\mu_1}{\mu_2}$ with $\delta = \mu_1$ and $\kappa_2 = 0{,}055$ for $\widetilde{\delta} = \frac{1}{6|K^{-1}|}$. For this problem, seven Newton iterations were needed to achieve a tolerance of 1e-6 for the energy norm of the incremental approximations. The numerical results are depicted in Fig. 6 including concentration profiles, velocity vectors and streamlines, and trace of the Cauchy pseudo-stress tensor. As in Ref. 30, from the center plot of Fig. 6 we see a clear attachment of the fluid stream to the side walls, whereas the material with high concentration at the inlet dissolves almost completely at the outlet. This effect corresponds to a relatively high inlet velocity, and since higher concentration material is injected, it penetrates the clear fluid pushing it toward the outlet.

Acknowledgments

This work was partially supported by CONICYT-Chile through BASAL project CMM, Universidad de Chile, and project Anillo ACT1118 (ANANUM); by the Ministry of Education through the project REDOC.CTA of the Graduate School, Universidad de Concepción; by Centro de Investigación en Ingeniería Matemática ($\mathrm{CI^2MA}$), Universidad de Concepción; and by the Swiss National Science Foundation through the Research Grant SNSF PP00P2-144922.

References

1. R. A. Adams and J. J. F. Fournier, *Sobolev Spaces* (Academic Press, 2003).
2. M. Alvarez, G. N. Gatica and R. Ruiz-Baier, An augmented mixed-primal finite element method for a coupled flow–transport problem, *M2AN Math. Model. Numer. Anal.* **49** (2015) 1399–1427.
3. J.-L. Auriault, On the domain of validity of Brinkman's equation, *Transp. Porous Med.* **79** (2009) 215–223.

4. D. Boffi, F. Brezzi and M. Fortin, *Mixed Finite Element Methods and Applications* (Springer-Verlag, 2013).
5. F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods* (Springer-Verlag, 1991).
6. M. Bulíček and P. Pustějovská, Existence analysis for a model describing flow of an incompressible chemically reacting non-Newtonian fluid, *SIAM J. Math. Anal.* **46** (2014) 3223–3240.
7. R. Bürger, S. Kumar and R. Ruiz-Baier, Discontinuous finite volume element discretization for coupled flow-transport problems arising in models of sedimentation, *J. Comput. Phys.* **299** (2015) 446–471.
8. R. Bürger, C. Liu and W. L. Wendland, Existence and stability for mathematical models of sedimentation–consolidation processes in several space dimensions, *J. Math. Anal. Appl.* **264** (2001) 288–310.
9. R. Bürger, R. Ruiz-Baier and H. Torres, A stabilized finite volume element formulation for sedimentation–consolidation processes. *SIAM J. Sci. Comput.* **34** (2012) B265–B289.
10. J. Camaño, R. Oyarzúa and G. Tierra, Analysis of an augmented mixed-FEM for the Navier Stokes problem. Preprint 2014-33, Centro de Investigación en Ingeniería Matemática (CI^{2}MA), Universidad de Concepción, Chile (2014), http://www.ci2ma.udec.cl. *Math. Comput.*, to appear.
11. P. Ciarlet, *The Finite Element Method for Elliptic Problems* (North-Holland, 1978).
12. P. Ciarlet, *Linear and Nonlinear Functional Analysis with Applications* (Society for Industrial and Applied Mathematics, 2013).
13. E. Colmenares, G. N. Gatica and R. Oyarzúa, Analysis of an augmented mixed-primal formulation for the stationary Boussinesq problem, *Numer. Methods Partial Differential Equations* **32** (2016) 445–478.
14. C. Cox, H. Lee and D. Szurley, Finite element approximation of the non-isothermal Stokes–Oldroyd equations, *Int. J. Numer. Anal. Model.* **4** (2007) 425–440.
15. M. Dauge, *Elliptic Boundary Value Problems on Corner Domains*: *Smoothness and Asymptotics of Solutions*, Lecture Notes in Mathematics, Vol. 1341 (Springer-Verlag, 1988).
16. M. Farhloul, S. Nicaise and L. Paquet, A mixed formulation of Boussinesq equations: Analysis of nonsingular solutions, *Math. Comput.* **69** (2000) 965–986.
17. M. Farhloul, S. Nicaise and L. Paquet, A refined mixed finite element method for the Boussinesq equations in polygonal domains, *IMA J. Numer. Anal.* **21** (2001) 525–551.
18. M. Farhloul and A. Zine, A dual mixed formulation for non-isothermal Oldroyd–Stokes problem, *Math. Model. Nat. Phenom.* **6** (2011) 130–156.
19. L. E. Figueroa, G. N. Gatica and A. Márquez, Augmented mixed finite element methods for the stationary Stokes equations, *SIAM J. Sci. Comput.* **31** (2008/09) 1082–1119.
20. G. N. Gatica, Analysis of a new augmented mixed finite element method for linear elasticity allowing $\mathbb{RT}_0 - \mathbb{P}_1 - \mathbb{P}_0$ approximations, *M2AN Math. Model. Numer. Anal.* **40** (2006) 1–28.
21. G. N. Gatica, *A Simple Introduction to the Mixed Finite Element Method*: *Theory and Applications*, Springer Briefs in Mathematics (Springer, 2014).
22. G. N. Gatica, A. Márquez and M. A. Sánchez, *Analysis of a velocity-pressure-pseudo-stress formulation for the stationary Stokes equations*, Comput. Methods Appl. Mech. Engrg. **199** (2010) 1064–1079.

23. P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Monographs and Studies in Mathematics, Vol. 24 (Pitman Advanced Publishing Program, 1985).
24. A. Khalili, A. J. Basu, U. Pietrzyk and B. B. Jørgensen, Advective transport through permeable sediments: A new numerical and experimental approach, *Acta Mech.* **132** (1999) 221–227.
25. M. Ouriemi, P. Aussillous and E. Guazzelli, Sediment dynamics. Part 1. Bed-load transport by laminar shearing flows, *J. Fluid Mech.* **636** (2009) 295–319.
26. R. Oyarzúa, T. Qin and D. Schötzau, An exactly divergence-free finite element method for a generalized Boussinesq problem, *IMA J. Numer. Anal.* **34** (2014) 1104–1135.
27. K. V. Parchevsky, Numerical simulation of sedimentation in the presence of 2D compressible convection and reconstruction of the particle-radius distribution function, *J. Engrg. Math.* **41** (2001) 203–219.
28. A. Quarteroni and A. Valli, *Numerical Approximation of Partial Differential Equations*, Springer Series in Computational Mathematics, Vol. 23 (Springer-Verlag, 1994).
29. R. Rao, L. Mondy and S. Albotelli, Instabilities during batch sedimentation in geometries containing obstacles: A numerical and experimental study, *J. Numer. Methods Fluids* **55** (2007) 723–735.
30. R. Tarpagkou and A. Pantokratos, CFD methodology for sedimentation tanks: The effcects of secondary phase on fluid phase using DPM coupled calculations, *Appl. Math. Model.* **37** (2013) 3478–3494.
31. D. J. Tritton, *Physical Fluid Dynamics* (Van Nostrand, 1980).