Journal of Scientific Computing (2019) 80:141–174
https://doi.org/10.1007/s10915-019-00931-4

New Mixed Finite Element Methods for Natural Convection with Phase-Change in Porous Media

Mario Alvarez[1] · **Gabriel N. Gatica**[2] · **Bryan Gomez-Vargas**[1,2] · **Ricardo Ruiz-Baier**[3,4]

Received: 9 November 2018 / Revised: 8 February 2019 / Accepted: 20 February 2019 /
Published online: 6 March 2019

Abstract
This article is concerned with the mathematical and numerical analysis of a steady phase change problem for non-isothermal incompressible viscous flow. The system is formulated in terms of pseudostress, strain rate and velocity for the Navier–Stokes–Brinkman equation, whereas temperature, normal heat flux on the boundary, and an auxiliary unknown are introduced for the energy conservation equation. In addition, and as one of the novelties of our approach, the symmetry of the pseudostress is imposed in an ultra-weak sense, thanks to which the usual introduction of the vorticity as an additional unknown is no longer needed. Then, for the mathematical analysis two variational formulations are proposed, namely mixed-primal and fully-mixed approaches, and the solvability of the resulting coupled formulations is established by combining fixed-point arguments, Sobolev embedding theorems and certain regularity assumptions. We then construct corresponding Galerkin discretizations based on adequate finite element spaces, and derive optimal a priori error estimates. Finally, numerical experiments in 2D and 3D illustrate the interest of this scheme and validate the theory.

Keywords Natural convection · Viscous flow in porous media · Change of phase · Mixed-primal formulation · Fully-mixed formulation · Fixed-point theory · Finite element methods

Mathematics Subject Classification 65N30 · 65N12 · 65N15 · 76R05 · 76D07

1 Introduction

We are interested in the mathematical and numerical investigation of phase change models for natural convection in porous media. Natural convection is a largely studied phenomenon due

This work was partially supported by CONICYT-Chile through the project AFB170001 of the PIA Program: Concurso Apoyo a Centros Cientificos y Tecnológicos de Excelencia con Financiamiento Basal; by Becas-Chile Programme for foreign students; by Centro de Investigación en Ingeniería Matemática (CI^{2}MA), Universidad de Concepción; and by Vicerrectoría de Investigación, project 540-B7-233, Sede de Occidente, Universidad de Costa Rica.

✉ Gabriel N. Gatica
ggatica@ci2ma.udec.cl

Extended author information available on the last page of the article

to its presence in different applications: melting and solidification processes [18,30,38,39], design of latent heat based energy storage devices [21], ocean and atmosphere dynamics [20,27], crystalization in magma chambers [9,35], etc. Differently from other works where the phase change is incorporated into the Boussinesq approximation by means of enthalpy-porosity methods [33] or enthalpy-viscosity models [18], in this article the problem is modeled either as a viscous Newtonian fluid where the change of phase is encoded in the viscosity itself, or using a Brinkman–Boussinesq approximation where the solidification process influences the drag directly.

A variety of numerical methods dealing with phase change Boussinesq models have been proposed in recent years, including bioconvective flows [13,28], porosity-based models [33, 39], and viscosity based-models [7,18,42]. Mathematical analysis of other related models as natural convection [29,41], and Boussinesq-type, such as time-dependent problems under different contexts [2,3,19], primal and mixed formulations [14,16,22,32], with viscosity of the fluid depending on the temperature [4,5], and exactly divergence free [32] are available in the literature. However, up to our knowledge, a rigorous mixed analysis for phase change models for natural convection is something that has not had great attention until now. Therefore, in the present work, we focus on the mathematical and numerical analysis of that problem, which has been proposed in [7, Section 4.2], and where the authors studied a fully-primal formulation for a non-stationary phase-change model. Here, and similarly to [16], we propose mixed-primal and fully-mixed approaches.

The rest of this work is organized as follows. In the remainder of this section, we recall some preliminary notations. The nonlinear model of interest, and the definitive unknowns to be considered in the variational formulation are presented in Sect. 2. For the Navier–Stokes–Brinkman equations, the main unknowns are the velocity, a pseudostress tensor relating the strain tensor with the convective term and the strain rate tensor. The pressure is eliminated using the fluid incompressibility and can be recovered as a post-process of the pseudostress. Moreover, because of the convective term, the velocity is sought in $\mathbf{H}^1(\Omega)$, which requires augmentation via Galerkin terms arising from the constitutive and equilibrium equations, and therefore, imposing in an ultra-weak sense the symmetry of the pseudostress, we do not need to introduce the vorticity as unknown in our variational formulation. In turn, for the energy equation, and in addition to the temperature, we introduce the normal heat flux through the boundary as a Lagrange multiplier for the primal formulation and a further unknown for the mixed approach. We remark that including these Galerkin terms allows us to circumvent the necessity of proving inf-sup conditions for both problems, and as a result, to relax the hypotheses on the corresponding discrete spaces. In this way, the classical Banach fixed-point theorem, the Lax–Milgram lemma, the Babuška–Brezzi theory, suitable regularity and smallness-of-data assumptions, can be applied to prove well-posedness of the continuous problem. In Sect. 3, we also define the Galerkin scheme considering arbitrary finite dimensional subspaces and provide its unique solvability (this time, by means of Brouwer fixed-point theorem), together with the corresponding Céa estimate. Then, we make precise the definition of the involved discrete spaces. In Sect. 4 we establish the corresponding fully-mixed variational formulation and its associated Galerkin scheme, and show that both systems are well-posed. Then, considering specific finite element spaces for the unknowns together with its approximation properties, we deduce the corresponding rates of convergence. We close in Sect. 5 with several numerical examples illustrating the performance of the augmented mixed-primal and fully-mixed finite element methods, as well as confirming the theoretical rates of convergence.

1.1 Preliminaries

Let us denote by $\Omega \subseteq \mathrm{R}^n$, $n = 2, 3$, a given bounded domain with polyhedral boundary $\Gamma = \partial\Omega$, and 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, divergence and tensor product operators as

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

In addition, given 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

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

where $\mathbb{I}$ is the identity matrix of $\mathrm{R}^{n\times n}$. Furthermore, we recall the following Hilbert space equipped with its usual norm

$$\mathbb{H}(\mathbf{div};\Omega) := \left\{\boldsymbol{\tau} \in \mathbb{L}^2(\Omega) : \quad \mathbf{div}\,\boldsymbol{\tau} \in \mathbf{L}^2(\Omega)\right\}, \qquad \|\boldsymbol{\tau}\|^2_{\mathbf{div};\Omega} := \|\boldsymbol{\tau}\|^2_{0,\Omega} + \|\mathbf{div}\,\boldsymbol{\tau}\|^2_{0,\Omega}.$$

In addition, by $|\cdot|$ we will denote both the Euclidean norm in R^n and the Frobenius norm in $\mathrm{R}^{n\times n}$.

2 The Model Problem

Let us consider the following PDE system, describing phase change mechanisms involving viscous fluids within porous media. The governing equations in this case correspond to the Navier–Stokes–Brinkman equations coupled with a generalized energy equation (related to the well-known Stefan problem)

$$(\nabla \boldsymbol{u})\boldsymbol{u} - \alpha\,\mathbf{div}\big[\mu(\theta)\mathbf{e}(\boldsymbol{u})\big] + \nabla p + \eta(\theta)\boldsymbol{u} = f(\theta)\boldsymbol{k} \qquad \text{in } \Omega, \tag{2.1a}$$

$$\operatorname{div} \boldsymbol{u} = 0 \qquad \text{in } \Omega, \tag{2.1b}$$

$$-\rho\,\operatorname{div}(\kappa\nabla\theta) + \boldsymbol{u}\cdot\nabla\theta + \boldsymbol{u}\cdot\nabla s(\theta) = 0 \qquad \text{in } \Omega, \tag{2.1c}$$

$$\boldsymbol{u} = \boldsymbol{u}_{\mathrm{D}} \quad \text{and} \quad \theta = \theta_{\mathrm{D}} \qquad \text{on } \Gamma, \tag{2.1d}$$

with $\alpha := \frac{1}{\mathrm{Re}}$, $\rho := \frac{1}{C\mathrm{Pr}}$, where Re and Pr are the Reynolds and Prandtl numbers, respectively, κ and C are the non-dimensional heat conductivity tensor (here assumed isotropic) and specific heat, respectively, $\boldsymbol{k}$ stands for the unit vector pointing oppositely to gravity, $\mathbf{e}(\boldsymbol{u}) = \frac{1}{2}(\nabla\boldsymbol{u} + \nabla\boldsymbol{u}^{\mathrm{t}})$ is the strain rate tensor, and $\boldsymbol{u} : \Omega \to \mathrm{R}^n$, $p : \Omega \to \mathrm{R}$ and $\theta : \Omega \to \mathrm{R}$, correspond to the velocity, pressure, and the temperature of the fluid flow, respectively. Finally, μ, η, s and f are the nonlinear viscosity, porosity, enthalpy and buoyancy terms, respectively, which depend on the temperature. Notice that here $s(\theta)$ denotes the regularized enthalpy function and it accounts for the latent heat of fusion, i.e. the energy needed to change the phase of a material [7,36,37].

For the subsequent analysis we assume that the functions $\mu,\ \eta,\ s$ are uniformly bounded and Lipschitz continuous: there exist positive constants $\mu_1, \mu_2, \eta_1, \eta_2, s_1, s_2, L_\mu, L_\eta, L_s$ such that

$$\begin{array}{lll} \mu_1 \leq \mu(\psi) \leq \mu_2, & |\mu(\psi) - \mu(\phi)| \leq L_\mu |\psi - \phi| & \forall\, \psi, \phi \in \mathrm{R}, \\ \eta_1 \leq \eta(\psi) \leq \eta_2, & |\eta(\psi) - \eta(\phi)| \leq L_\eta |\psi - \phi| & \forall\, \psi, \phi \in \mathrm{R}, \\ s_1 \leq s(\psi) \leq s_2, & |s(\psi) - s(\phi)| \leq L_s |\psi - \phi| & \forall\, \psi, \phi \in \mathrm{R}. \end{array} \tag{2.2}$$

Similar assumptions are placed on the buoyancy function f: there exist positive constants C_f and L_f such that

$$|f(\psi)| \leq C_f |\psi|, \quad |f(\psi) - f(\phi)| \leq L_f |\psi - \phi| \quad \forall\, \psi, \phi \in \mathrm{R}. \tag{2.3}$$

On the other hand, we will suppose that for every $\psi \in \mathrm{H}^1(\Omega)$, we have $s(\psi) \in \mathrm{H}^1(\Omega)$, and that there exist positive constants s_3 and $L_{\tilde{s}}$ such that

$$|\nabla s(\psi)| \leq s_3 |\nabla \psi|, \quad |\nabla s(\psi) - \nabla s(\phi)| \leq L_{\tilde{s}} |\psi - \phi|, \quad \forall\, \psi, \phi \in \mathrm{H}^1(\Omega)\,. \tag{2.4}$$

Finally, we suppose that κ and κ^{-1} are uniform bounded and uniformly positive definite tensors, meaning that there exist positive constants K_0, K_1, $\widetilde{K}_0$ and $\widetilde{K}_1$ such that

$$|\kappa| \leq K_1, \quad \kappa \boldsymbol{v} \cdot \boldsymbol{v} \geq K_0 |\boldsymbol{v}|^2, \quad |\kappa^{-1}| \leq \widetilde{K}_1, \quad \kappa^{-1} \boldsymbol{v} \cdot \boldsymbol{v} \geq \widetilde{K}_0 |\boldsymbol{v}|^2 \quad \forall\, \boldsymbol{v} \in \mathrm{R}^n. \tag{2.5}$$

With respect to the boundary conditions in (2.1d), we assume that $\boldsymbol{u}_{\mathrm{D}} \in \mathbf{H}^{1/2}(\Gamma)$, $\theta_{\mathrm{D}} \in \mathrm{H}^{1/2}(\Gamma)$, and that $\boldsymbol{u}_{\mathrm{D}}$ verifies the compatibility condition

$$\int_\Gamma \boldsymbol{u}_{\mathrm{D}} \cdot \boldsymbol{\nu} \,=\, 0. \tag{2.6}$$

In addition, it is well-known (see, e.g [31]) that uniqueness of pressure is ensured in the space

$$\mathrm{L}_0^2(\Omega) \,:=\, \left\{ q \in \mathrm{L}^2(\Omega) : \int_\Omega q \,=\, 0 \right\}.$$

We end this section by remarking that, due to the laminar regime of the fluid in each one of the numerical tests reported in Sect. 5, the module of the velocity field is small, and hence it might not be necessary to compute the Reynolds number, besides the fact that there seems to be no formula available in the literature for the case (as the present) of a non-constant viscosity. Nevertheless, if an estimation of this number is in fact needed, we would suggest to take a characteristic viscosity μ_c defined as the mean value of it, that is $\mu_c := \dfrac{1}{|\Omega|} \displaystyle\int_\Omega \mu(\theta)$, which can be controlled by μ_1 and μ_2 (cf. (2.2)), and then compute the model parameters (Reynolds and Prandtl numbers) and rewrite the coupled-system based on this choice.

3 The Mixed-Primal Approach

In this section we proceed similarly as in [4,11,14] to propose a mixed-primal approach for (2.1). Then, we establish the corresponding continuous and discrete formulations, analyze their solvability by using a fixed-point approach, and derive the corresponding a priori error estimates.

3.1 The Continuous Formulation

We first proceed as in [4] and set the strain rate tensor as an auxiliary unknown:

$$\mathbf{t} := \mathbf{e}(\boldsymbol{u}) = \nabla \boldsymbol{u} - \boldsymbol{\gamma}(\boldsymbol{u}) \in \mathbb{L}^2_{\mathtt{tr}}(\Omega),$$

where, for each $\boldsymbol{v} \in \mathbf{H}^1(\Omega)$, $\boldsymbol{\gamma}(\boldsymbol{v}) = \frac{1}{2}(\nabla \boldsymbol{v} - (\nabla \boldsymbol{v})^{\mathrm{t}})$ is the skew-symmetric part of the velocity gradient tensor $\nabla \boldsymbol{v}$, and

$$\mathbb{L}^2_{\mathtt{tr}}(\Omega) := \left\{ \mathbf{s} \in \mathbb{L}^2(\Omega) : \quad \mathbf{s} = \mathbf{s}^{\mathrm{t}} \quad \text{and} \quad \mathrm{tr}(\mathbf{s}) = 0 \right\}.$$

Then, introducing also the pseudostress tensor as a new unknown:

$$\boldsymbol{\sigma} := \alpha\, \mu(\theta)\mathbf{t} - (\boldsymbol{u} \otimes \boldsymbol{u}) - p\, \mathbb{I}, \tag{3.1}$$

we deduce that (2.1b) together with (3.1) are equivalent to the pair of equations

$$\begin{aligned} \alpha\, \mu(\theta)\, \mathbf{t} - (\boldsymbol{u} \otimes \boldsymbol{u})^{\mathrm{d}} &= \boldsymbol{\sigma}^{\mathrm{d}} \quad \text{in} \quad \Omega, \\ p &= -\frac{1}{n}\mathrm{tr}(\boldsymbol{\sigma} + \boldsymbol{u} \otimes \boldsymbol{u}) \quad \text{in} \quad \Omega. \end{aligned}$$

Consequently, we arrive at the following coupled system without pressure:

$$\begin{aligned} \mathbf{t} + \boldsymbol{\gamma}(\boldsymbol{u}) &= \nabla \boldsymbol{u} && \text{in} \quad \Omega, && (3.2\mathrm{a}) \\ \alpha\, \mu(\theta)\, \mathbf{t} - (\boldsymbol{u} \otimes \boldsymbol{u})^{\mathrm{d}} &= \boldsymbol{\sigma}^{\mathrm{d}} && \text{in} \quad \Omega, && (3.2\mathrm{b}) \\ \eta(\theta)\, \boldsymbol{u} - \mathbf{div}\, \boldsymbol{\sigma} &= f(\theta)\boldsymbol{k} && \text{in} \quad \Omega, && (3.2\mathrm{c}) \\ -\rho\, \mathrm{div}(\kappa \nabla \theta) + \boldsymbol{u} \cdot \nabla \theta + \boldsymbol{u} \cdot \nabla s(\theta) &= 0 && \text{in} \quad \Omega, && (3.2\mathrm{d}) \\ \boldsymbol{u} &= \boldsymbol{u}_{\mathrm{D}} && \text{on} \quad \Gamma, && (3.2\mathrm{e}) \\ \theta &= \theta_{\mathrm{D}} && \text{on} \quad \Gamma, && (3.2\mathrm{f}) \\ \int_\Omega \mathrm{tr}(\boldsymbol{\sigma} + \boldsymbol{u} \otimes \boldsymbol{u}) &= 0. && && (3.2\mathrm{g}) \end{aligned}$$

Note that the incompressibility constraint is implicitly present in (3.2b), relating $\boldsymbol{\sigma}$, $\mathbf{t}$ and $\boldsymbol{u}$. In turn, the fact that the pressure p must belong to $\mathrm{L}^2_0(\Omega)$ (for uniqueness reasons) is guaranteed by the equivalent statement given by (3.2g).

Thus, in order to derive a primal formulation for the energy equation, we proceed to multiply (3.2d) by $\psi \in \mathrm{H}^1(\Omega)$, integrate by parts, and introduce, as a new unknown, the normal heat flux on Γ, $\lambda := -\rho\, \kappa \nabla \theta \cdot \boldsymbol{\nu} \in \mathrm{H}^{-1/2}(\Gamma)$, so that we arrive at

$$\rho \int \kappa \nabla \theta \cdot \nabla \psi + \langle \lambda, \psi \rangle_\Gamma + \int_\Omega \psi \boldsymbol{u} \cdot \nabla(\theta + s(\theta)) = 0 \quad \forall\, \psi \in \mathrm{H}^1(\Omega)\,, \tag{3.3}$$

where $\langle \cdot, \cdot \rangle_\Gamma$ denotes from now on the duality pairing between $\mathrm{H}^{-1/2}(\Gamma)$ and $\mathrm{H}^{1/2}(\Gamma)$. In turn, the Dirichlet condition $\theta = \theta_{\mathrm{D}}$ on Γ is imposed weakly as

$$\langle \xi, \theta \rangle_\Gamma = \langle \xi, \theta_{\mathrm{D}} \rangle_\Gamma \quad \forall\, \xi \in \mathrm{H}^{-1/2}(\Gamma).$$

On the other hand, multiplying (3.2b) by a suitable test function, we obtain

$$\alpha \int_\Omega \mu(\theta)\mathbf{t} : \mathbf{s} - \int_\Omega \boldsymbol{\sigma}^{\mathrm{d}} : \mathbf{s} - \int_\Omega (\boldsymbol{u} \otimes \boldsymbol{u})^{\mathrm{d}} : \mathbf{s} = 0 \quad \forall \mathbf{s} \in \mathbb{L}^2_{\mathtt{tr}}(\Omega). \tag{3.4}$$

Here we readily note that in order to bound the third terms on the LHS of (3.3) and (3.4), and thanks to the continuous injection of $\mathrm{H}^1(\Omega)$ into $\mathrm{L}^4(\Omega)$, we require the unknown $\boldsymbol{u}$ to live in $\mathbf{H}^1(\Omega)$ (see e.g. [4–6]). Such regularity can be exploited to cast the Navier–Stokes–Brinkman equations uniquely in terms of the pseudostress and the velocity. Indeed, testing (3.2a) against $\boldsymbol{\tau} \in \mathbb{H}(\mathbf{div}; \Omega)$ and employing (3.2e), we readily obtain

$$\int_\Omega \mathbf{t} : \boldsymbol{\tau}^{\mathrm{d}} + \int_\Omega \gamma(\boldsymbol{u}) : \boldsymbol{\tau} + \int_\Omega \boldsymbol{u} \cdot \mathbf{div}\, \boldsymbol{\tau} = \langle \boldsymbol{\tau}\boldsymbol{\nu}, \boldsymbol{u}_{\mathrm{D}} \rangle_\Gamma \quad \forall\, \boldsymbol{\tau} \in \mathbb{H}(\mathbf{div}; \Omega)\,.$$

Afterwards, testing (3.2c) against $\boldsymbol{v} \in \mathbf{H}^1(\Omega)$, we deduce that

$$-\int_\Omega \boldsymbol{v} \cdot \mathbf{div}\, \boldsymbol{\sigma} + \int_\Omega \eta(\theta) \boldsymbol{u} \cdot \boldsymbol{v} = \int_\Omega f(\theta) \boldsymbol{k} \cdot \boldsymbol{v} \quad \forall\, \boldsymbol{v} \in \mathbf{H}^1(\Omega),$$

and finally, defining $\mathcal{A} := \Big\{ \boldsymbol{\gamma}(\boldsymbol{v}) : \ \boldsymbol{v} \in \mathbf{H}^1(\Omega) \Big\}$, we impose the symmetry of $\boldsymbol{\sigma}$ in an ultra-weak sense, as follows:

$$\int_\Omega \boldsymbol{\sigma} : \boldsymbol{\gamma}(\boldsymbol{v}) = 0 \quad \forall\, \boldsymbol{v} \in \mathbf{H}^1(\Omega). \tag{3.5}$$

We stress here that the usual way of imposing this property of $\boldsymbol{\sigma}$ is in the form: $\int_\Omega \boldsymbol{\sigma} : \boldsymbol{\eta} = 0$ $\forall\, \boldsymbol{\eta} \in \mathbb{L}^2_{\mathrm{skew}}(\Omega) := \Big\{ \boldsymbol{\omega} \in \mathbb{L}^2(\Omega) : \ \boldsymbol{\omega} + \boldsymbol{\omega}^{\mathrm{t}} = \mathbf{0} \Big\}$, which is known as the weak sense. However, in the present approach we propose to take advantage of the further regularity of $\boldsymbol{u}$ and its corresponding test functions, which are all now in $\mathbf{H}^1(\Omega)$, and simply test $\boldsymbol{\sigma}$ against tensors in $\mathcal{A}$. In this way, the fact that $\mathcal{A}$ is a proper subspace of $\mathbb{L}^2_{\mathrm{skew}}(\Omega)$ constitutes the reason why this alternative imposition of the symmetry of the pseudostress is called *ultra-weak*.

Hence, a preliminary weak formulation for the coupled problem (2.1) reads: Find $(\mathbf{t}, \boldsymbol{\sigma}, \boldsymbol{u}, \theta, \lambda) \in \mathbb{L}^2_{\mathtt{tr}}(\Omega) \times \mathbb{H}(\mathbf{div}; \Omega) \times \mathbf{H}^1(\Omega) \times \mathrm{H}^1(\Omega) \times \mathrm{H}^{-1/2}(\Gamma)$ such that

$$\begin{aligned}
\alpha \int_\Omega \mu(\theta) \mathbf{t} : \mathbf{s} - \int_\Omega (\boldsymbol{u} \otimes \boldsymbol{u})^{\mathrm{d}} : \mathbf{s} - \int_\Omega \boldsymbol{\sigma}^{\mathrm{d}} : \mathbf{s} = 0 \quad \forall\, \mathbf{s} \in \mathbb{L}^2_{\mathtt{tr}}(\Omega),\\
\int_\Omega \mathbf{t} : \boldsymbol{\tau}^{\mathrm{d}} + \int_\Omega \gamma(\boldsymbol{u}) : \boldsymbol{\tau} + \int_\Omega \boldsymbol{u} \cdot \mathbf{div}\, \boldsymbol{\tau} = \langle \boldsymbol{\tau}\boldsymbol{\nu}, \boldsymbol{u}_{\mathrm{D}} \rangle_\Gamma \quad \forall\, \boldsymbol{\tau} \in \mathbb{H}(\mathbf{div}; \Omega)\,,\\
-\int_\Omega \boldsymbol{v} \cdot \mathbf{div}\, \boldsymbol{\sigma} - \int_\Omega \boldsymbol{\sigma} : \gamma(\boldsymbol{v}) + \int_\Omega \eta(\theta) \boldsymbol{u} \cdot \boldsymbol{v} = \int_\Omega f(\theta) \boldsymbol{k} \cdot \boldsymbol{v} \quad \forall\, \boldsymbol{v} \in \mathbf{H}^1(\Omega),\\
\rho \int \kappa \nabla\theta \cdot \nabla\psi + \langle \lambda, \psi \rangle_\Gamma = -\int_\Omega \psi \boldsymbol{u} \cdot \nabla(\theta + s(\theta)) \quad \forall\, \psi \in \mathrm{H}^1(\Omega),\\
\langle \xi, \theta \rangle_\Gamma = \langle \xi, \theta_{\mathrm{D}} \rangle_\Gamma \quad \forall\, \xi \in \mathrm{H}^{-1/2}(\Gamma).
\end{aligned} \tag{3.6}$$

On the other hand, by virtue of the orthogonal decomposition $\mathbb{H}(\mathbf{div}; \Omega) = \mathbb{H}_0(\mathbf{div}; \Omega) \oplus \mathrm{R}\mathbb{I}$, where

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

and (3.2g), we can write $\boldsymbol{\sigma} = \boldsymbol{\sigma}_0 + c\mathrm{I}$, with $\boldsymbol{\sigma}_0$ in $\mathbb{H}_0(\mathbf{div}; \Omega)$, and c given explicitly in terms of $\boldsymbol{u}$ by

$$c = -\frac{1}{n|\Omega|} \int_\Omega \mathrm{tr}(\boldsymbol{u} \otimes \boldsymbol{u}).$$

Then, denoting from now on the unknown $\boldsymbol{\sigma}_0$ simply by $\boldsymbol{\sigma}$, the variational formulation (3.6) can be reformulated in terms of the $\mathbb{H}_0(\mathbf{div}; \Omega)$-component of the pseudostress (see [5, Lemma 3.1]). Accordingly, in order to analyze (3.6) we augment using residual Galerkin-type terms arising from (3.2), but all them tested differently from (3.6), namely:

$$\begin{aligned}
\kappa_1 \int_\Omega \left\{\boldsymbol{\sigma}^{\mathrm{d}} + (\boldsymbol{u}\otimes\boldsymbol{u})^{\mathrm{d}} - \alpha\mu(\theta)\mathbf{t}\right\} : \boldsymbol{\tau}^{\mathrm{d}} &= 0 && \forall\, \boldsymbol{\tau} \in \mathbb{H}_0(\mathbf{div};\Omega)\,,\\
\kappa_2 \int_\Omega \{\mathbf{div}\,\boldsymbol{\sigma} - \eta(\theta)\boldsymbol{u}\}\cdot\mathbf{div}\boldsymbol{\tau} &= -\kappa_2\int_\Omega f(\theta)\boldsymbol{k}\cdot\mathbf{div}\boldsymbol{\tau} && \forall\, \boldsymbol{\tau} \in \mathbb{H}_0(\mathbf{div};\Omega)\,,\\
\kappa_3 \int_\Omega \{\mathbf{e}(\boldsymbol{u}) - \mathbf{t}\}\cdot\mathbf{e}(\boldsymbol{v}) &= 0 && \forall\, \boldsymbol{v} \in \mathbf{H}^1(\Omega),
\end{aligned}$$

where κ_1, κ_2 and κ_3 are positive parameters to be specified later on. In this way, denoting $H := \mathbb{L}^2_{\mathtt{tr}}(\Omega)\times\mathbb{H}_0(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega)$, $\vec{\mathbf{t}} := (\mathbf{t},\boldsymbol{\sigma},\boldsymbol{u})$, and $\vec{\mathbf{s}} := (\mathbf{s},\boldsymbol{\tau},\boldsymbol{v})$, we arrive at the following augmented mixed-primal formulation for (2.1): Find $(\vec{\mathbf{t}},(\theta,\lambda)) \in H\times \mathrm{H}^1(\Omega)\times\mathrm{H}^{-1/2}(\Gamma)$ such that

$$\mathbf{A}_\theta(\vec{\mathbf{t}},\vec{\mathbf{s}}) + \mathbf{B}_{\boldsymbol{u}}(\vec{\mathbf{t}},\vec{\mathbf{s}}) = F_\theta(\vec{\mathbf{s}}) + F_{\mathrm{D}}(\vec{\mathbf{s}}) \qquad \forall\, \vec{\mathbf{s}}\in H, \tag{3.7a}$$

$$\mathbf{a}(\theta,\psi) + \mathbf{b}(\psi,\lambda) = H_{\boldsymbol{u},\theta}(\psi) \qquad \forall\,\psi\in\mathrm{H}^1(\Omega)\,, \tag{3.7b}$$

$$\mathbf{b}(\theta,\xi) = G(\xi) \qquad \forall\,\xi\in\mathrm{H}^{-1/2}(\Gamma), \tag{3.7c}$$

where, given an arbitrary $(\boldsymbol{w},\phi)\in\mathbf{H}^1(\Omega)\times\mathrm{H}^1(\Omega)$, the forms $\mathbf{A}_\phi$, $\mathbf{B}_{\boldsymbol{w}}$, $\mathbf{a}$, $\mathbf{b}$, and the functionals F_ϕ, F_{D}, $H_{\boldsymbol{w},\phi}$, and G are defined as

$$\begin{aligned}
\mathbf{A}_\phi(\vec{\mathbf{t}},\vec{\mathbf{s}}) := {}& \alpha\int \mu(\phi)\mathbf{t} : \left\{\mathbf{s}-\kappa_1\boldsymbol{\tau}^{\mathrm{d}}\right\} + \int_\Omega \mathbf{t} : \left\{\boldsymbol{\tau}^{\mathrm{d}} - \kappa_3\,\mathbf{e}(\boldsymbol{v})\right\} - \int_\Omega \boldsymbol{\sigma}^{\mathrm{d}} : \left\{\mathbf{s}-\kappa_1\boldsymbol{\tau}^{\mathrm{d}}\right\}\\
&+ \int_\Omega \boldsymbol{u}\cdot\mathbf{div}\,\boldsymbol{\tau} - \int_\Omega \boldsymbol{v}\cdot\mathbf{div}\,\boldsymbol{\sigma} + \int_\Omega \gamma(\boldsymbol{u})\colon\boldsymbol{\tau} - \int_\Omega \boldsymbol{\sigma}:\gamma(\boldsymbol{v}) \qquad (3.8a)\\
&+ \int_\Omega \eta(\phi)\boldsymbol{u}\cdot\left\{\boldsymbol{v} - \kappa_2\,\mathbf{div}\boldsymbol{\tau}\right\} + \kappa_2\int_\Omega \mathbf{div}\,\boldsymbol{\sigma}\cdot\mathbf{div}\boldsymbol{\tau} + \kappa_3\int_\Omega \mathbf{e}(\boldsymbol{u}):\mathbf{e}(\boldsymbol{v}),
\end{aligned}$$

$$\mathbf{B}_{\boldsymbol{w}}(\vec{\mathbf{t}},\vec{\mathbf{s}}) := \int_\Omega (\boldsymbol{u}\otimes\boldsymbol{w})^{\mathrm{d}} : \left\{\kappa_1\boldsymbol{\tau}^{\mathrm{d}} - \mathbf{s}\right\}, \tag{3.8b}$$

for all $\vec{\mathbf{t}},\vec{\mathbf{s}}\in H$, and

$$\mathbf{a}(\theta,\psi) := \rho\int_\Omega \kappa\nabla\theta\cdot\nabla\psi \qquad \forall\,\theta,\psi\in\mathrm{H}^1(\Omega), \tag{3.9a}$$

$$\mathbf{b}(\psi,\xi) := \langle\xi,\psi\rangle_\Gamma \qquad \forall\,(\psi,\xi)\in\mathrm{H}^1(\Omega)\times\mathrm{H}^{-1/2}(\Gamma), \tag{3.9b}$$

$$F_\phi(\vec{\mathbf{s}}) := \int_\Omega f(\phi)\boldsymbol{k}\cdot\left\{\boldsymbol{v}-\kappa_2\,\mathbf{div}\,\boldsymbol{\tau}\right\} \qquad \forall\vec{\mathbf{s}}\in H, \tag{3.9c}$$

$$F_{\mathrm{D}}(\vec{\mathbf{s}}) := \langle\boldsymbol{\tau}\boldsymbol{\nu},\boldsymbol{u}_{\mathrm{D}}\rangle_\Gamma \qquad \forall\vec{\mathbf{s}}\in H, \tag{3.9d}$$

$$H_{\boldsymbol{w},\phi}(\psi) := -\int_\Omega \psi\boldsymbol{w}\cdot\nabla(\phi+s(\phi)) \qquad \forall\,\psi\in\mathrm{H}^1(\Omega), \tag{3.9e}$$

$$G(\xi) := \langle\xi,\theta_{\mathrm{D}}\rangle_\Gamma \qquad \forall\,\xi\in\mathrm{H}^{-1/2}(\Gamma). \tag{3.9f}$$

We notice in advance that the forms $\mathbf{B}_{\boldsymbol{w}}$, $\boldsymbol{a}$ and $\mathbf{b}$ are exactly defined as in [15, Section 3.1] and therefore we omit parts of the proofs whenever necessary. Finally, we remark that, in contrast with other recent strain-based formulations [4,12,23,25], here we do not introduce vorticity as additional unknown. Also, the presence of the drag term in the momentum equation allows us to complete the $\mathbf{H}^1(\Omega)-$norm of the velocity without the need of a fourth residual term in the augmentation procedure.

3.2 Solvability Analysis

We proceed similarly as in [4,25] and utilize a fixed-point scheme to prove the well-posedness of the continuous formulation (3.7). Let us write $\mathbf{H} := \mathbf{H}^1(\Omega) \times \mathrm{H}^1(\Omega)$ and define $\mathbf{S} : \mathbf{H} \to H$ as

$$\mathbf{S}(\boldsymbol{w}, \phi) = (\mathbf{S}_1(\boldsymbol{w}, \phi), \mathbf{S}_2(\boldsymbol{w}, \phi), \mathbf{S}_3(\boldsymbol{w}, \phi)) := \vec{\mathbf{t}} \qquad \forall\, (\boldsymbol{w}, \phi) \in \mathbf{H}\,, \tag{3.10}$$

where $\vec{\mathbf{t}} \in H$ is the unique solution of the problem defined by (3.7a) with $(\boldsymbol{w}, \phi)$ instead of $(\boldsymbol{u}, \theta)$, that is

$$\mathbf{A}_\phi(\vec{\mathbf{t}}, \vec{\mathbf{s}}) \,+\, \mathbf{B}_{\boldsymbol{w}}(\vec{\mathbf{t}}, \vec{\mathbf{s}}) = F_\phi(\vec{\mathbf{s}}) \,+\, F_{\mathrm{D}}(\vec{\mathbf{s}}) \quad \forall \vec{\mathbf{s}} \in H. \tag{3.11}$$

In turn, let $\widetilde{\mathbf{S}} : \mathbf{H} \to \mathrm{H}^1(\Omega)$ be the operator defined by

$$\widetilde{\mathbf{S}}(\boldsymbol{w}, \phi) := \theta \qquad \forall\, (\boldsymbol{w}, \phi) \in \mathbf{H}\,, \tag{3.12}$$

where θ is the first component of the unique solution $(\theta, \lambda) \in \mathrm{H}^1(\Omega) \times \mathrm{H}^{-1/2}(\Gamma)$ of the problem defined by (3.7b)–(3.7c) with $(\boldsymbol{w}, \phi)$ instead of $(\boldsymbol{u}, \theta)$, that is

$$\begin{aligned} \mathbf{a}(\theta, \psi) \,+\, \mathbf{b}(\psi, \lambda) &= H_{\boldsymbol{w},\phi}(\psi) \quad \forall\, \psi \in \mathrm{H}^1(\Omega)\,, \\ \mathbf{b}(\theta, \xi) &= G(\xi) \qquad \forall\, \xi \in \mathrm{H}^{-1/2}(\Gamma). \end{aligned} \tag{3.13}$$

Then, we define the operator $\mathbf{T} : \mathbf{H} \to \mathbf{H}$ by

$$\mathbf{T}(\boldsymbol{w}, \phi) = \big(\mathbf{S}_3(\boldsymbol{w}, \phi), \widetilde{\mathbf{S}}(\mathbf{S}_3(\boldsymbol{w}, \phi), \phi)\big) \quad \forall\, (\boldsymbol{w}, \phi) \in \mathbf{H}, \tag{3.14}$$

and one readily realizes that solving (3.7) is equivalent to seeking a fixed point of $\mathbf{T}$, that is: Find $(\boldsymbol{u}, \theta) \in \mathbf{H}$ such that

$$\mathbf{T}(\boldsymbol{u}, \theta) = (\boldsymbol{u}, \theta).$$

We now provide sufficient conditions under which the uncoupled problems (3.11) and (3.13) are indeed uniquely solvable. In what follows, for each $\vec{\mathbf{s}} \in H$, $\|\vec{\mathbf{s}}\|$ denotes the corresponding product norm.

Lemma 3.1 *Assume that* $\kappa_1 \in \left(0, \frac{2\mu_1\delta_1}{\mu_2}\right)$, $\kappa_2 \in \left(0, \frac{2\eta_1\delta_3}{\eta_2}\right)$ *and* $\kappa_3 \in \left(0, 2\alpha\delta_2\left(\mu_1 - \frac{\kappa_1\mu_2}{2\delta_1}\right)\right)$ *with* $\delta_1 \in \left(0, \frac{2}{\alpha\mu_2}\right)$, $\delta_2 \in (0, 2)$, $\delta_3 \in \left(0, \frac{2}{\eta_2}\right)$. *Then, there exists* $r_0 > 0$ *such that for each* $r \in (0, r_0)$, *problem* (3.11) *has a unique solution* $\mathbf{S}(\boldsymbol{w}, \phi) := \vec{\mathbf{t}} \in H$, *for each* $(\boldsymbol{w}, \phi) \in \mathbf{H}$ *with* $\|\boldsymbol{w}\|_{1,\Omega} \le r$. *Moreover, there exists* $c_{\mathbf{S}} > 0$, *independent of* $(\boldsymbol{w}, \phi)$, *such that*

$$\|\mathbf{S}(\boldsymbol{w}, \phi)\| = \|\vec{\mathbf{t}}\| \le c_{\mathbf{S}}\left\{C_f\,\|\phi\|_{0,\Omega} + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma}\right\} \quad \forall\, (\boldsymbol{w}, \phi) \in \mathbf{H}. \tag{3.15}$$

Proof Let us start the discussion by deriving the continuity of the forms involved. First, employing the assumptions (2.2), we deduce that

$$|\mathbf{A}_\phi(\vec{\mathbf{t}}, \vec{\mathbf{s}})| \le C_{\mathbf{A}}\|\vec{\mathbf{t}}\|\|\vec{\mathbf{s}}\| \quad \forall \vec{\mathbf{t}}, \vec{\mathbf{s}} \in H, \tag{3.16}$$

where $C_{\mathbf{A}}$ is a constant depending on $\alpha, \kappa_1, \kappa_2, \kappa_3, \mu_2$, and η_2. In turn, by applying the continuous injection $\boldsymbol{i}_c : \mathbf{H}^1(\Omega) \to \mathbf{L}^4(\Omega)$, we obtain that

$$|\mathbf{B}_{\boldsymbol{w}}(\vec{\mathbf{t}}, \vec{\mathbf{s}})| \le \|\boldsymbol{i}_c\|^2\,(1 + \kappa_1^2)^{1/2}\|\boldsymbol{u}\|\|\boldsymbol{w}\|\|\vec{\mathbf{s}}\| \quad \forall \vec{\mathbf{t}}, \vec{\mathbf{s}} \in H. \tag{3.17}$$

Hence, from (3.16) and (3.17), there exists a positive constant denoted by $\|\mathbf{A}_\phi + \mathbf{B}_{\boldsymbol{w}}\|$, such that

$$|(\mathbf{A}_\phi + \mathbf{B}_{\boldsymbol{w}})(\vec{\mathbf{t}}, \vec{\mathbf{s}})| \le \|\mathbf{A}_\phi + \mathbf{B}_{\boldsymbol{w}}\|\,\|\vec{\mathbf{t}}\|\,\|\vec{\mathbf{s}}\| \quad \forall \vec{\mathbf{t}}, \vec{\mathbf{s}} \in H.$$

On the other hand, in order to show that $(\mathbf{A}_\phi + \mathbf{B}_{\boldsymbol{w}})$ is elliptic, we first prove that $\mathbf{A}_\phi$ satisfies this property. In fact, by using Cauchy–Schwarz and Young inequalities, and the results provided in [10, Prop. 3.1] and [17, Thm. 6.15-1] with constants $c_3(\Omega)$ and $\kappa_0(\Omega)$, respectively, it is possible to find a constant $\widetilde{\alpha}(\Omega) := \min\{\alpha_1, \alpha_3\kappa_0(\Omega), \alpha_4\}$, independent of $(\boldsymbol{w}, \phi)$,such that

$$\mathbf{A}_\phi(\vec{\mathbf{s}}, \vec{\mathbf{s}}) \geq \widetilde{\alpha}(\Omega)\|\vec{\mathbf{s}}\|^2 \quad \forall \vec{\mathbf{s}} \in H, \tag{3.18}$$

where

$$\alpha_1 := \alpha\mu_1 - \frac{\kappa_1\alpha\mu_2}{2\delta_1} - \frac{\kappa_3}{2\delta_2}, \quad \alpha_2 := \min\left\{\kappa_1\left(1 - \frac{\alpha\mu_2\delta_1}{2}\right), \frac{\kappa_2}{2}\left(1 - \frac{\eta_2\delta_3}{2}\right)\right\},$$
$$\alpha_3 := \min\left\{\kappa_3\left(1 - \frac{\delta_2}{2}\right), \eta_1 - \frac{\kappa_2\eta_2}{2\delta_3}\right\}, \quad \alpha_4 := \min\left\{\alpha_2 c_3(\Omega), \frac{\kappa_2}{2}\left(1 - \frac{\eta_2\delta_3}{2}\right)\right\},$$

where $\kappa_1, \kappa_2, \kappa_3, \delta_1, \delta_2$ and δ_3 are defined as in the statement of the present lemma. Moreover, by combining (3.17) and (3.18), we obtain that

$$(\mathbf{A}_\phi + \mathbf{B}_{\boldsymbol{w}})(\vec{\mathbf{t}}, \vec{\mathbf{s}}) \geq \frac{\widetilde{\alpha}(\Omega)}{2}\|\vec{\mathbf{s}}\|^2 \quad \forall \vec{\mathbf{s}} \in H, \tag{3.19}$$

provided $\|\boldsymbol{w}\|_{1,\Omega} \leq r_0$, with

$$r_0 := \frac{\widetilde{\alpha}(\Omega)}{2\,\|\boldsymbol{i}_c\|^2\,(1+\kappa_1^2)^{1/2}}, \tag{3.20}$$

which confirms the ellipticity of the nonlinear operator $\mathbf{A}_\phi + \mathbf{B}_{\boldsymbol{w}}$. On the other hand, by applying Cauchy–Schwarz inequality and the trace theorem in $\mathbb{H}(\mathbf{div};\Omega)$, we deduce that $F_\phi, F_{\mathrm{D}} \in H'$ with

$$\|F_\phi\| \leq C_f(1+\kappa_2^2)^{1/2}\,\|\phi\|_{0,\Omega} \quad \text{and} \quad \|F_{\mathrm{D}}\| \leq \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma}. \tag{3.21}$$

Consequently, a straightforward application of the Lax–Milgram lemma implies that there exists a unique solution $\vec{\mathbf{t}} \in H$ of (3.11). Finally, using (3.19) and (3.21), and performing simple algebraic manipulations, we derive (3.15) with $c_{\mathrm{S}} := \frac{2(1+\kappa_2^2)^{1/2}}{\widetilde{\alpha}(\Omega)} > 0$, independent of $(\boldsymbol{w}, \phi)$. □

Lemma 3.2 *For each* $(\boldsymbol{w}, \phi) \in \mathbf{H}$, *problem* (3.13) *has a unique solution* $(\theta, \lambda) = \big(\widetilde{\mathbf{S}}(\boldsymbol{w}, \phi), \lambda\big) \in \mathrm{H}^1(\Omega) \times \mathrm{H}^{-1/2}(\Gamma)$. *Moreover, there exists a constant* $\widetilde{c}_{\mathrm{S}} > 0$ *independent of* $(\boldsymbol{w}, \phi)$, *such that*

$$\|\widetilde{\mathbf{S}}(\boldsymbol{w}, \phi)\| \leq \|(\theta, \lambda)\| \leq \widetilde{c}_{\mathrm{S}}\Big\{\|\boldsymbol{w}\|_{1,\Omega}|\phi|_{1,\Omega} + \|\theta_{\mathrm{D}}\|_{1/2,\Gamma}\Big\}. \tag{3.22}$$

Proof From [14, Lemma 3.4] we know that $\boldsymbol{a}$ and $\mathbf{b}$ are bounded independently of $(\boldsymbol{w}, \phi)$, and that the bilinear form $\mathbf{b}$ satisfies the inf-sup condition. Furthermore, recalling that ϑ (cf. (2.5)) is a uniformly positive definite tensor, and using the Friedrichs–Poincaré inequality, we also deduce that $\boldsymbol{a}$ is V-elliptic with constant $\alpha_{\boldsymbol{a}}(\Omega)$, where V is the kernel of the operator induced by $\mathbf{b}$. Now, using (3.9e), (3.9f) and applying the continuous injection $i_c : \mathrm{H}^1(\Omega) \to \mathrm{L}^4(\Omega)$, we find that

$$\left\|H_{\boldsymbol{w},\phi}\right\| \leq \|i_c\|^2\,\left\{1 + s_3\right\}\|\boldsymbol{w}\|_{1,\Omega}|\phi|_{1,\Omega} \quad \text{and} \quad \|G\| \leq \|\theta_{\mathrm{D}}\|_{1/2,\Gamma},$$

which implies that $H_{\boldsymbol{w},\phi}$ and G are bounded functionals. Thus, a straightforward application of the Babŭska–Brezzi theory (see, e.g. [24, Thm. 2.3]) proves that for each $(\boldsymbol{w}, \phi) \in \mathbf{H}$, problem (3.13) has a unique solution $(\theta, \lambda) \in \mathrm{H}^1(\Omega) \times \mathrm{H}^{-1/2}(\Gamma)$. Moreover, there exists a positive constant $\widetilde{c}_{\mathrm{S}}$ depending on ρ, K_1, $\alpha_{\boldsymbol{a}}(\Omega)$, $\|i_c\|$, s_3 and the inf-sup constant of $\mathbf{b}$, such that the estimate (3.22) holds. □

At this point, we remark that for computational purposes, the constants α_1, α_2 and α_3 defining $\widetilde{\alpha}(\Omega)$ in Lemma 3.1, can be maximized by choosing the parameters $\delta_1, \delta_2, \delta_3, \kappa_1\, \kappa_2$, and κ_3 as the middle points of their feasible ranges. Adequate choices for these parameters are then

$$\delta_1 = \frac{1}{\alpha\mu_2}, \quad \kappa_1 = \frac{\mu_1}{\alpha\mu_2^2}, \quad \delta_2 = 1, \quad \kappa_3 = \frac{\alpha\mu_1}{2}, \quad \delta_3 = \frac{1}{\eta_2}, \quad \kappa_2 = \frac{\eta_1}{\eta_2^2}. \tag{3.23}$$

Continuing with the analysis, we assume further regularity on the problem defining $\mathbf{S}$. More precisely, we assume that $\boldsymbol{u}_{\mathrm{D}} \in \mathbf{H}^{1/2+\varepsilon}(\Gamma)$ for some $\varepsilon \in (0, 1)$ (when $n = 2$) or $\varepsilon \in \left[\frac{1}{2}, 1\right)$ (when $n = 3$), and that for each $(\boldsymbol{w}, \phi) \in \mathbf{H}$ with $\|\boldsymbol{w}\|_{1,\Omega} + \|\phi\|_{1,\Omega} \le r$, $r > 0$ given, there holds $(\mathbf{r}, \boldsymbol{\zeta}, z) := \mathbf{S}(\boldsymbol{w}, \phi) \in \mathbb{L}^2_{\mathtt{tr}}(\Omega) \cap \mathbb{H}^{\varepsilon}(\Omega) \times \mathbb{H}_0(\mathbf{div};\Omega) \cap \mathbb{H}^{\varepsilon}(\Omega) \times \mathbf{H}^{1+\varepsilon}(\Omega)$, with

$$\|\mathbf{r}\|_{\varepsilon,\Omega} + \|\boldsymbol{\zeta}\|_{\varepsilon,\Omega} + \|z\|_{1+\varepsilon,\Omega} \le \widehat{C}(r)\left\{C_f\, \|\phi\|_{0,\Omega} + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma}\right\}, \tag{3.24}$$

where C_f is given by (2.3) and $\widehat{C}(r)$ is a positive constant independent of $(\boldsymbol{w}, \phi)$ but depending on the upper bound r of its norm. The reason of the stipulated ranges for ε will be clarified in the forthcoming analysis (specifically in the proofs of Lemmas 3.4 and 3.6, below). Also, we pay attention to the fact that while the estimate (3.24) will be employed only to bound $\|\mathbf{r}\|_{\varepsilon,\Omega}$, we have stated it including the terms $\|\boldsymbol{\zeta}\|_{\varepsilon,\Omega}$ and $\|z\|_{1+\varepsilon,\Omega}$ as well, since due to the first and second equations of (3.2), the regularities of $\mathbf{r}$, $\boldsymbol{\zeta}$ and z will most likely be connected.

On the other hand, we emphasize that the well-posedness of the uncoupled problems (3.11) and (3.13) ensure that the operators $\mathbf{S}$, $\widetilde{\mathbf{S}}$ and $\mathbf{T}$ are well-defined. Hence, the existence of a unique fixed-point of $\mathbf{T}$ follows after verifying the hypotheses of the Banach fixed-point theorem.

Lemma 3.3 *Given $r \in (0, r_0)$, with r_0 given by (3.20), we let $W := \Big\{(\boldsymbol{w}, \phi) \in \mathbf{H} : \|(\boldsymbol{w}, \phi)\| \le r\Big\}$, and assume that*

$$c(r)\left\{C_f + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma}\right\} + \widetilde{c}_{\mathbf{S}}\, \|\theta_{\mathrm{D}}\|_{1/2,\Gamma} \le r, \tag{3.25}$$

where $c(r) := (1 + \widetilde{c}_{\mathbf{S}})\, c_{\mathbf{S}} \max\{1, r\}$, and C_f, $c_{\mathbf{S}}$ and $\widetilde{c}_{\mathbf{S}}$ are the constants specified in (2.3), and Lemmas 3.1 and 3.2, respectively. Then $\mathbf{T}(W) \subseteq W$.

Proof It follows exactly as in [14, Lemma 3.5]. □

Next, the Lipschitz continuity of $\mathbf{T}$ will essentially be a direct consequence of the following two lemmas providing the same property for $\mathbf{S}$ and $\widetilde{\mathbf{S}}$, respectively.

Lemma 3.4 *Let $r \in (0, r_0)$ with r_0 given by (3.20). Then, there exists a constant $\widetilde{C}_{\mathbf{S}} > 0$, independent of r, such that for all $(\boldsymbol{w}_1, \phi_1)$, $(\boldsymbol{w}_2, \phi_2) \in \mathbf{H}$, with $\|\boldsymbol{w}_1\|_{1,\Omega}$, $\|\boldsymbol{w}_2\|_{1,\Omega} \le r$, there holds*

$$\begin{aligned}\|\mathbf{S}(\boldsymbol{w}_1, \phi_1) - \mathbf{S}(\boldsymbol{w}_2, \phi_2)\| \le \widetilde{C}_{\mathbf{S}} \Big\{ &\|\mathbf{S}_3(\boldsymbol{w}_2, \phi_2)\|_{1,\Omega} \Big(\|\boldsymbol{w}_1 - \boldsymbol{w}_2\|_{1,\Omega} + \|\phi_1 - \phi_2\|_{1,\Omega} \Big) \\ &+ \|\phi_1 - \phi_2\|_{\mathrm{L}^{n/\varepsilon}(\Omega)}\, \|\mathbf{S}_1(\boldsymbol{w}_2, \phi_2)\|_{\varepsilon,\Omega} + L_f\, \|\phi_1 - \phi_2\|_{0,\Omega} \Big\}. \end{aligned} \tag{3.26}$$

Proof Given $(\boldsymbol{w}_1, \phi_1)$, $(\boldsymbol{w}_2, \phi_2)$ as stated, we let $\vec{\mathbf{t}}_j := (\mathbf{t}_j, \boldsymbol{\sigma}_j, \boldsymbol{u}_j) = \mathbf{S}(\boldsymbol{w}_j, \phi_j) \in H$, $j \in \{1, 2\}$, which, according to (3.11), means that for all $\vec{\mathbf{s}} \in H$ there hold:

$$\mathbf{A}_{\phi_1}(\vec{\mathbf{t}}_1, \vec{\mathbf{s}}) + \mathbf{B}_{\boldsymbol{w}_1}(\vec{\mathbf{t}}_1, \vec{\mathbf{s}}) = F_{\phi_1}(\vec{\mathbf{s}}) + F_{\mathrm{D}}(\vec{\mathbf{s}}) \quad \text{and} \quad \mathbf{A}_{\phi_2}(\vec{\mathbf{t}}_2, \vec{\mathbf{s}}) + \mathbf{B}_{\boldsymbol{w}_2}(\vec{\mathbf{t}}_2, \vec{\mathbf{s}}) = F_{\phi_2}(\vec{\mathbf{s}}) + F_{\mathrm{D}}(\vec{\mathbf{s}}).$$

Now, applying the ellipticity of $\mathbf{A}_{\phi_1} + \mathbf{B}_{\boldsymbol{w}_1}$ (cf. 3.19), and then adding and subtracting the equality $\mathbf{A}_{\phi_2}(\vec{\mathbf{t}}_2, \vec{\mathbf{s}}) + \mathbf{B}_{\boldsymbol{w}_2}(\vec{\mathbf{t}}_2, \vec{\mathbf{s}}) = F_{\phi_2}(\vec{\mathbf{s}}) + F_{\mathrm{D}}(\vec{\mathbf{s}})$, we find that

$$\begin{aligned}\frac{\widetilde{\alpha}(\Omega)}{2}\|\vec{\mathbf{t}}_1 - \vec{\mathbf{t}}_2\|^2 &\leq (\mathbf{A}_{\phi_1} + \mathbf{B}_{\boldsymbol{w}_1})(\vec{\mathbf{t}}_1 - \vec{\mathbf{t}}_2, \vec{\mathbf{t}}_1 - \vec{\mathbf{t}}_2)\\ &= (F_{\phi_1} - F_{\phi_2})(\vec{\mathbf{t}}_1 - \vec{\mathbf{t}}_2) + (\mathbf{A}_{\phi_2} - \mathbf{A}_{\phi_1})(\vec{\mathbf{t}}_2, \vec{\mathbf{t}}_1 - \vec{\mathbf{t}}_2) + (\mathbf{B}_{\boldsymbol{w}_2 - \boldsymbol{w}_1})(\vec{\mathbf{t}}_2, \vec{\mathbf{t}}_1 - \vec{\mathbf{t}}_2).\end{aligned} \tag{3.27}$$

Next, for the first and third terms on the right hand side of (3.27), we exploit the assumption (2.3) and the estimate given in [14, Lemma 3.6], respectively, to obtain

$$\begin{aligned}&\left|\int_\Omega \Big(f(\phi_1) - f(\phi_2)\Big)\boldsymbol{k}\cdot\Big\{(\boldsymbol{u}_1 - \boldsymbol{u}_2) - \kappa_2\,\mathbf{div}\,(\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2)\Big\}\right|\\ &\quad\leq L_f(1+\kappa_2^2)^{1/2}\,\|\phi_1 - \phi_2\|_{0,\Omega}\,\|\vec{\mathbf{t}}_1 - \vec{\mathbf{t}}_2\|,\end{aligned} \tag{3.28}$$

and

$$\begin{aligned}&\left|\int_\Omega \Big(\boldsymbol{u}_2\otimes(\boldsymbol{w}_2 - \boldsymbol{w}_1)\Big)^{\mathrm{d}} : \Big\{\kappa_1(\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2)^{\mathrm{d}} - (\mathbf{t}_1 - \mathbf{t}_2)\Big\}\right|\\ &\quad\leq \|\boldsymbol{i}_c\|^2\,(1+\kappa_1^2)^{1/2}\,\|\boldsymbol{u}_2\|_{1,\Omega}\,\|\boldsymbol{w}_2 - \boldsymbol{w}_1\|_{1,\Omega}\,\|\vec{\mathbf{t}}_1 - \vec{\mathbf{t}}_2\|.\end{aligned} \tag{3.29}$$

On the other hand, for the second term of (3.27), we apply the assumptions (2.2), and the Cauchy–Schwarz and Hölder inequalities, to deduce that

$$\begin{aligned}\Big|(\mathbf{A}_{\phi_2} - \mathbf{A}_{\phi_1})(\vec{\mathbf{t}}_2, \vec{\mathbf{t}}_1 - \vec{\mathbf{t}}_2)\Big| &= \Big|\alpha\int_\Omega \Big(\mu(\phi_2) - \mu(\phi_1)\Big)\mathbf{t}_2 : \Big\{(\mathbf{t}_1 - \mathbf{t}_2) - \kappa_1(\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2)^{\mathrm{d}}\Big\}\\ &+ \int_\Omega\Big(\eta(\phi_2) - \eta(\phi_1)\Big)\boldsymbol{u}_2\cdot\Big\{(\boldsymbol{u}_1 - \boldsymbol{u}_2) - \kappa_2\,\mathbf{div}\,(\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2)\Big\}\Big|\\ &\leq \Big(\alpha L_\mu(1+\kappa_1^2)^{1/2}\,\|\phi_2 - \phi_1\|_{\mathrm{L}^{2q}(\Omega)}\,\|\mathbf{t}_2\|_{\mathbb{L}^{2p}(\Omega)}\\ &+ L_\eta\,\|i_c\|^2\,(1+\kappa_2^2)^{1/2}\,\|\phi_2 - \phi_1\|_{1,\Omega}\,\|\boldsymbol{u}_2\|_{1,\Omega}\Big)\|\vec{\mathbf{t}}_1 - \vec{\mathbf{t}}_2\|,\end{aligned} \tag{3.30}$$

with $p, q \in (1, +\infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$. At this point, we proceed as in [6, Lemma 3.9]. In fact, given the further regularity ε assumed in (3.24), we recall that the Sobolev embedding theorem (see, e.g [1, Thm. 4.12]) establishes the continuous injection $i_\varepsilon : \mathrm{H}^\varepsilon(\Omega) \to \mathrm{L}^{2p}(\Omega)$ with boundedness constant C_ε, where

$$2p = \begin{cases} \dfrac{2}{1-\varepsilon} & \text{if } n = 2,\\[2mm] \dfrac{6}{3-2\varepsilon} & \text{if } n = 3,\end{cases}$$

and $2q = \frac{n}{\varepsilon}$, and therefore, there holds

$$\|\mathbf{t}_2\|_{\mathbb{L}^{2p}(\Omega)} \leq C_\varepsilon\,\|\mathbf{t}_2\|_{\varepsilon,\Omega} \quad \forall\, \mathbf{t}_2 \in \mathbb{H}^\varepsilon(\Omega). \tag{3.31}$$

Then, (3.31) could be bounded by (3.24), yielding for each $(\boldsymbol{w}_2, \phi_2) \in \mathbf{H}$ with $\|\boldsymbol{w}_2\|_{1,\Omega} + \|\phi_2\|_{1,\Omega} \leq r$, the estimate

$$\|\mathbf{t}_2\|_{\mathbb{L}^{2p}(\Omega)} \leq C_\varepsilon\widehat{C}(r)\left\{C_f\,\|\phi_2\|_{0,\Omega} + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma}\right\}.$$

Finally, denoting

$$\widetilde{C}_{\mathbf{S}} := \frac{2}{\widetilde{\alpha}(\Omega)}\max\Big\{(1+\kappa_2^2)^{1/2}, \|\boldsymbol{i}_c\|^2\,(1+\kappa_1^2)^{1/2}, \alpha C_\varepsilon L_\mu(1+\kappa_1^2)^{1/2}, L_\eta\,\|i_c\|^2\,(1+\kappa_2^2)^{1/2}\Big\},$$

inequalities (3.27), (3.28), (3.29), (3.30) and (3.31), imply (3.26) and complete the proof. □

Lemma 3.5 *There exists $\widetilde{C}_{\widetilde{\mathbf{S}}} > 0$, such that for all $(\boldsymbol{w}_1, \phi_1), (\boldsymbol{w}_2, \phi_2) \in \mathbf{H}$ there holds*

$$\begin{aligned} &\|\widetilde{\mathbf{S}}(\boldsymbol{w}_1, \phi_1) - \widetilde{\mathbf{S}}(\boldsymbol{w}_2, \phi_2)\| \\ &\quad \leq \widetilde{C}_{\widetilde{\mathbf{S}}} \Big\{ \|\boldsymbol{w}_1 - \boldsymbol{w}_2\|_{1,\Omega} |\phi_1|_{1,\Omega} + \|\boldsymbol{w}_2\|_{1,\Omega} |\phi_1 - \phi_2|_{1,\Omega} + \|\boldsymbol{w}_2\|_{1,\Omega} \|\phi_1 - \phi_2\|_{0,\Omega} \Big\}. \end{aligned} \tag{3.32}$$

Proof Given $(\boldsymbol{w}_1, \phi_1), (\boldsymbol{w}_2, \phi_2) \in \mathbf{H}$, we let $(\theta_1, \lambda_1), (\theta_2, \lambda_2) \in \mathrm{H}^1(\Omega) \times \mathrm{H}^{-1/2}(\Gamma)$ be solutions to (3.13) corresponding to $(\boldsymbol{w}_1, \phi_1)$ and $(\boldsymbol{w}_2, \phi_2)$, respectively, that is $\theta_j = \widetilde{\mathbf{S}}(\boldsymbol{w}_j, \phi_j)$, $\quad j \in \{1, 2\}$. Then invoking the linearity of the forms $\boldsymbol{a}$ and $\mathbf{b}$, and performing algebraic manipulations, we deduce (using both formulations arising from (3.13)) that

$$\begin{aligned} \boldsymbol{a}(\theta_1 - \theta_2, \psi) + \mathbf{b}(\psi, \lambda_1 - \lambda_2) &= H_{\boldsymbol{w}_1 - \boldsymbol{w}_2, \phi_1}(\psi) + H_{\boldsymbol{w}_2, \phi_1}(\psi) - H_{\boldsymbol{w}_2, \phi_2}(\psi) \quad \forall\, \psi \in \mathrm{H}^1(\Omega), \\ \mathbf{b}(\theta_1 - \theta_2, \xi) &= 0 \quad \forall\, \xi \in \mathrm{H}^{-1/2}(\Gamma). \end{aligned} \tag{3.33}$$

Next, noting from the second equation of (3.33) that $\theta_1 - \theta_2$ belongs to the kernel V of $\mathbf{b}$, taking $\psi = \theta_1 - \theta_2$ and $\xi = \lambda_1 - \lambda_2$ in (3.33), applying the ellipticity of $\boldsymbol{a}$ in V, and using the assumption (2.4), we readily deduce from the first equation of (3.33) that

$$\begin{aligned} &\alpha_{\boldsymbol{a}}(\Omega) \|\theta_1 - \theta_2\|_{1,\Omega}^2 \leq \boldsymbol{a}(\theta_1 - \theta_2, \theta_1 - \theta_2) \\ &\quad = H_{\boldsymbol{w}_1 - \boldsymbol{w}_2, \phi_1}(\theta_1 - \theta_2) + H_{\boldsymbol{w}_2, \phi_1}(\theta_1 - \theta_2) - H_{\boldsymbol{w}_2, \phi_2}(\theta_1 - \theta_2) \\ &\quad \leq \|i_c\|^2 \Big\{ (1 + s_3) \|\boldsymbol{w}_1 - \boldsymbol{w}_2\|_{1,\Omega} |\phi_1|_{1,\Omega} + \|\boldsymbol{w}_2\|_{1,\Omega} |\phi_1 - \phi_2|_{1,\Omega} \\ &\quad + L_{\tilde{s}} \|\boldsymbol{w}_2\|_{1,\Omega} \|\phi_1 - \phi_2\|_{0,\Omega} \Big\} \|\theta_1 - \theta_2\|_{1,\Omega}\,, \end{aligned}$$

which gives (3.32) with $\widetilde{C}_{\widetilde{\mathbf{S}}} := \frac{\|i_c\|^2}{\alpha_{\boldsymbol{a}}} \max\{1 + s_3, L_{\tilde{s}}\}$. □

The announced property of $\mathbf{T}$ is proved now.

Lemma 3.6 *Let r and W as in Lemma 3.3. Then, there exists a positive constant $C_{\mathbf{T}}$ such that for all $(\boldsymbol{w}_1, \phi_1), (\boldsymbol{w}_2, \phi_2) \in W$ there holds*

$$\begin{aligned} &\|\mathbf{T}(\boldsymbol{w}_1, \phi_1) - \mathbf{T}(\boldsymbol{w}_2, \phi_2)\| \\ &\quad \leq C_{\mathbf{T}} \Big\{ C_f + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma} + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma} + L_f \Big\} \|(\boldsymbol{w}_1, \phi_1) - (\boldsymbol{w}_2, \phi_2)\|\,. \end{aligned}$$

Proof It follows directly from the definition of $\mathbf{T}$ (cf. (3.14)) and the estimates (3.26) and (3.32). We remit to [14, Lemma 3.8] for similar further details. □

Finally, the main result of this section is given as follows.

Theorem 3.7 *Suppose that the parameters κ_1, κ_2 and κ_3 satisfy the conditions required by Lemma 3.1. Let r and W as in Lemma 3.3, and assume that the data satisfy (3.25) and*

$$C_{\mathbf{T}} \Big\{ C_f + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma} + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma} + L_f \Big\} < 1. \tag{3.34}$$

Then, problem (3.7) has a unique solution $(\vec{\mathbf{t}}, (\theta, \lambda)) \in H \times \mathrm{H}^1(\Omega) \times \mathrm{H}^{-1/2}(\Gamma)$, with $(\boldsymbol{u}, \theta) \in W$, and there holds

$$\|\vec{\mathbf{t}}\| \leq c_{\mathbf{S}} \left\{ C_f\, r + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma} \right\},$$

and

$$\|(\theta, \lambda)\| \leq \widetilde{c}_{\mathrm{S}} \{ r \|\boldsymbol{u}\|_{1,\Omega} + \|\theta_{\mathrm{D}}\|_{1/2,\Gamma} \}.$$

Proof It follows as a combination of Lemmas 3.3 and 3.6, the assumption (3.34), the Banach fixed-point theorem, and the a priori estimates (3.15) and (3.22). We omit further details. □

3.3 The Galerkin Scheme

In this section we analyze a Galerkin scheme associated with (3.7). We remark in advance that most of the details are omitted since they follow straightforwardly by adapting the fixed-point strategy from Sect. 3.2. We start by considering generic finite dimensional subspaces

$$\begin{gathered}\mathbb{H}_h^{\mathbf{t}} \subseteq \mathbb{L}^2_{\mathtt{tr}}(\Omega), \quad \mathbb{H}_h^{\boldsymbol{\sigma}} \subseteq \mathbb{H}_0(\mathbf{div};\Omega), \quad \mathbf{H}_h^{\boldsymbol{u}} \subseteq \mathbf{H}^1(\Omega),\\ \mathrm{H}_h^{\theta} \subseteq \mathrm{H}^1(\Omega), \quad \text{and} \quad \mathrm{H}_h^{\lambda} \subseteq \mathrm{H}^{-1/2}(\Gamma),\end{gathered}$$

which will be specified later on. Hereafter, h denotes the size of a regular triangulation $\mathcal{T}_h$ of $\overline{\Omega}$ made up of triangles K (in R^2) or tetrahedra K (in R^3) of diameter h_K, i.e. $h := \max\{h_K : K \in \mathcal{T}_h\}$. Defining $H_h := \mathbb{H}_h^{\mathbf{t}} \times \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}$, and denoting $\vec{\mathbf{t}}_h := (\mathbf{t}_h, \boldsymbol{\sigma}_h, \boldsymbol{u}_h)$ and $\vec{\mathbf{s}}_h := (\mathbf{s}_h, \boldsymbol{\tau}_h, \boldsymbol{v}_h)$, the Galerkin scheme for (3.7) reads: Find $(\vec{\mathbf{t}}_h, (\theta_h, \lambda_h)) \in H_h \times \mathrm{H}_h^{\theta} \times \mathrm{H}_h^{\lambda}$ such that

$$\begin{aligned}\mathbf{A}_{\theta_h}(\vec{\mathbf{t}}_h, \vec{\mathbf{s}}_h) + \mathbf{B}_{\boldsymbol{u}_h}(\vec{\mathbf{t}}_h, \vec{\mathbf{s}}_h) &= F_{\theta_h}(\vec{\mathbf{s}}_h) + F_{\mathrm{D}}(\vec{\mathbf{s}}_h) && \forall\, \vec{\mathbf{s}}_h \in H_h,\\ \mathbf{a}(\theta_h, \psi_h) + \mathbf{b}(\psi_h, \lambda_h) &= H_{\boldsymbol{u}_h,\theta_h}(\psi_h) && \forall\, \psi_h \in \mathrm{H}_h^{\theta},\\ \mathbf{b}(\theta_h, \xi_h) &= G(\xi_h) && \forall\, \xi_h \in \mathrm{H}_h^{\lambda}.\end{aligned} \tag{3.35}$$

Next, we set $\mathbf{H}_h := \mathbf{H}_h^{\boldsymbol{u}} \times \mathrm{H}_h^{\theta}$ and let $\mathbf{S}_h : \mathbf{H}_h \to H_h$ be the operator defined as

$$\mathbf{S}_h(\boldsymbol{w}_h, \phi_h) = (\mathbf{S}_{1,h}(\boldsymbol{w}_h, \phi_h), \mathbf{S}_{2,h}(\boldsymbol{w}_h, \phi_h), \mathbf{S}_{3,h}(\boldsymbol{w}_h, \phi_h)) := \vec{\mathbf{t}}_h \qquad \forall\, (\boldsymbol{w}_h, \phi_h) \in \mathbf{H}_h\,, \tag{3.36}$$

where $\vec{\mathbf{t}}_h \in H_h$ is the unique solution of the problem given by the first equation of (3.35) with $(\boldsymbol{w}_h, \phi_h)$ instead of $(\boldsymbol{u}_h, \theta_h)$, that is

$$\mathbf{A}_{\phi_h}(\vec{\mathbf{t}}_h, \vec{\mathbf{s}}_h) + \mathbf{B}_{\boldsymbol{w}_h}(\vec{\mathbf{t}}_h, \vec{\mathbf{s}}_h) = F_{\phi_h}(\vec{\mathbf{s}}_h) + F_{\mathrm{D}}(\vec{\mathbf{s}}_h) \quad \forall\, \vec{\mathbf{s}}_h \in H_h. \tag{3.37}$$

Just for sake of completeness, we recall here that the functional F_{D} is defined in (3.9d). In turn, for a given pair $(\boldsymbol{w}_h, \phi_h)$, the bilinear forms $\mathbf{A}_{\phi_h}$ and $\mathbf{B}_{\boldsymbol{w}_h}$, and the functional F_{ϕ_h} are those corresponding to (3.8a), (3.8b) and (3.9c), respectively, with $\boldsymbol{w} = \boldsymbol{w}_h$ and $\phi = \phi_h$.

Furthermore, we define $\widetilde{\mathbf{S}}_h : \mathbf{H}_h \to \mathrm{H}_h^{\theta}$ as

$$\widetilde{\mathbf{S}}_h(\boldsymbol{w}_h, \phi_h) := \theta_h \qquad \forall\, (\boldsymbol{w}_h, \phi_h) \in \mathbf{H}_h\,, \tag{3.38}$$

where θ_h is the first component of the unique solution $(\theta_h, \lambda_h) \in \mathrm{H}_h^{\theta} \times \mathrm{H}_h^{\lambda}$ of the problem given by the second and third equations of (3.35) with $(\boldsymbol{w}_h, \phi_h)$ instead of $(\boldsymbol{u}_h, \theta_h)$, that is

$$\begin{aligned}\mathbf{a}(\theta_h, \psi_h) + \mathbf{b}(\psi_h, \lambda_h) &= H_{\boldsymbol{w}_h,\phi_h}(\psi_h) \quad \forall\, \psi_h \in \mathrm{H}_h^{\theta}(\Omega)\,,\\ \mathbf{b}(\theta_h, \xi_h) &= G(\xi_h) \qquad\quad \forall\, \xi_h \in \mathrm{H}_h^{\lambda}.\end{aligned} \tag{3.39}$$

The forms $\boldsymbol{a}$ and $\mathbf{b}$ and the functional G are defined in (3.9a), (3.9b) and (3.9f), respectively, whereas $H_{\boldsymbol{w}_h,\phi_h}$ is defined as in (3.9e) with $\boldsymbol{w} = \boldsymbol{w}_h$ and $\phi = \phi_h$.

Finally, by introducing the operator $\mathbf{T}_h : \mathbf{H}_h \to \mathbf{H}_h$ as

$$\mathbf{T}_h(\boldsymbol{w}_h, \phi_h) = \big(\mathbf{S}_{3,h}(\boldsymbol{w}_h, \phi_h), \widetilde{\mathbf{S}}_h(\mathbf{S}_{3,h}(\boldsymbol{w}_h, \phi_h), \phi_h)\big) \qquad \forall\, (\boldsymbol{w}_h, \phi_h) \in \mathbf{H}_h,$$

we see that solving (3.35) is equivalent to seeking $(\boldsymbol{u}_h, \theta_h) \in \mathbf{H}_h$ such that

$$\mathbf{T}_h(\boldsymbol{u}_h, \theta_h) = (\boldsymbol{u}_h, \theta_h). \tag{3.40}$$

Certainly, all the above makes sense if we guarantee that the uncoupled discrete problems (3.37) and (3.39) are well-posed, which is addressed in what follows. We begin with the corresponding result for $\mathbf{S}_h$, which actually follows almost verbatim to that of its continuous counterpart $\mathbf{S}$, and proof can be omitted.

Lemma 3.8 *Assume that* $\kappa_1 \in \left(0, \frac{2\mu_1\delta_1}{\mu_2}\right)$, $\kappa_2 \in \left(0, \frac{2\eta_1\delta_3}{\eta_2}\right)$ *and* $\kappa_3 \in \left(0, 2\alpha\delta_2\left(\mu_1 - \frac{\kappa_1\mu_2}{2\delta_1}\right)\right)$, *with* $\delta_1 \in \left(0, \frac{2}{\alpha\mu_2}\right)$, $\delta_2 \in (0,2)$ *and* $\delta_3 \in \left(0, \frac{2}{\eta_2}\right)$. *Then, there exists* $r_0 > 0$ *such that for each* $r \in (0, r_0)$, *problem* (3.37) *has a unique solution* $\mathbf{S}_h(\boldsymbol{w}_h, \phi_h) := \vec{\mathbf{t}}_h \in H_h$ *for each* $(\boldsymbol{w}_h, \phi_h) \in \mathbf{H}_h$ *with* $\|\boldsymbol{w}_h\|_{1,\Omega} \leq r$. *Moreover, there exists* $c_{\mathbf{S}} > 0$, *independent of* $(\boldsymbol{w}_h, \phi_h)$, *such that*

$$\|\mathbf{S}_h(\boldsymbol{w}_h, \phi_h)\| = \|\vec{\mathbf{t}}_h\| \leq c_{\mathbf{S}} \left\{C_f \|\phi_h\|_{0,\Omega} + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma}\right\} \quad \forall\, (\boldsymbol{w}_h, \phi_h) \in \mathbf{H}_h. \tag{3.41}$$

In turn, in order to analyze the problem (3.39), we need to incorporate further hypotheses on the discrete spaces H_h^θ and H_h^λ. For this purpose, we now let

$$V_h := \left\{\psi_h \in \mathrm{H}_h^\theta : \quad \mathbf{b}(\psi_h, \xi_h) = 0 \quad \forall\, \xi_h \in \mathrm{H}_h^\lambda\right\},$$

be the discrete kernel of $\mathbf{b}$. Then, assuming the following discrete inf-sup conditions (which do hold for some finite element spaces, as those listed at the end of this section):

(**H.0**) There exists a constant $\alpha_1 > 0$, independent of h, such that

$$\sup_{\substack{\psi_h \in V_h \\ \psi_h \neq 0}} \frac{\boldsymbol{a}(\psi_h, \varphi_h)}{\|\psi_h\|_{1,\Omega}} \geq \alpha_1 \|\varphi_h\|_{1,\Omega} \quad \forall\, \varphi_h \in V_h. \tag{3.42}$$

(**H.1**) There exists a constant $\alpha_2 > 0$, independent of h, such that

$$\sup_{\substack{\psi_h \in \mathrm{H}_h^\theta \\ \psi_h \neq 0}} \frac{\mathbf{b}(\psi_h, \xi_h)}{\|\psi_h\|_{1,\Omega}} \geq \alpha_2 \|\xi_h\|_{-1/2,\Gamma} \quad \forall\, \xi_h \in \mathrm{H}_h^\lambda,$$

we can prove that the operator $\widetilde{\mathbf{S}}_h$ is well-posed, which is abridged in the following lemma. We refer to [14, Lemma 4.2] for further details.

Lemma 3.9 *For each* $(\boldsymbol{w}_h, \phi_h) \in \mathbf{H}_h$, *problem* (3.39) *has a unique solution* $(\theta_h, \lambda_h) = (\widetilde{\mathbf{S}}_h(\boldsymbol{w}_h, \phi_h), \lambda_h) \in \mathrm{H}_h^\theta \times \mathrm{H}_h^\lambda$. *Moreover, there exists a constant* $\widetilde{C} > 0$ *independent of* $(\boldsymbol{w}_h, \phi_h)$, *such that*

$$\|\widetilde{\mathbf{S}}_h(\boldsymbol{w}_h, \phi_h)\| \leq \|(\theta_h, \lambda_h)\| \leq \widetilde{C} \left\{\|\boldsymbol{w}_h\|_{1,\Omega} |\phi_h|_{1,\Omega} + \|\theta_{\mathrm{D}}\|_{1/2,\Gamma}\right\}.$$

The solvability of the fixed-point problem (3.40) is now proved by means of the Brouwer fixed-point theorem (see, e.g. [17, Thm. 9.9-2]). We begin with the discrete version of Lemma 3.3.

Lemma 3.10 *Given* $r \in (0, r_0)$, *with* r_0 *as in* (3.20), *we let* $W_h := \Big\{(\boldsymbol{w}_h, \phi_h) \in \mathbf{H}_h : \|(\boldsymbol{w}_h, \phi_h)\| \leq r\Big\}$, *and assume that*

$$\widetilde{c}(r) \left\{C_f + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma}\right\} + \widetilde{C}\, \|\theta_{\mathrm{D}}\|_{1/2,\Gamma} \leq r, \tag{3.43}$$

where $\widetilde{c}(r) := (1 + \widetilde{C})\, c_{\mathbf{S}} \max\{1, r\}$, *and* $c_{\mathbf{S}}$ *and* $\widetilde{C}$ *are the constants specified in Lemmas* 3.1 *and* 3.9, *respectively. Then* $\mathbf{T}_h(W_h) \subseteq W_h$.

The discrete analogue of Lemma 3.4 is provided next. We notice in advance that, instead of the regularity assumptions employed in the continuous case (not applicable in the present discrete case), we simply utilize an $\mathrm{L}^4 - \mathrm{L}^4 - \mathrm{L}^2$ argument.

Lemma 3.11 *Let $r \in (0, r_0)$ with r_0 given by (3.20). Then, there exists a constant $\widetilde{C}_{\mathbf{S}} > 0$, independent of r, such that for all $(\boldsymbol{w}_h, \phi_h)$, $(\widetilde{\boldsymbol{w}}_h, \widetilde{\phi}_h) \in \mathbf{H}_h$, with $\|\boldsymbol{w}_h\|_{1,\Omega}$, $\|\widetilde{\boldsymbol{w}}_h\|_{1,\Omega} \le r$, there holds*

$$\|\mathbf{S}_h(\boldsymbol{w}_h, \phi_h) - \mathbf{S}_h(\widetilde{\boldsymbol{w}}_h, \widetilde{\phi}_h)\| \le \widetilde{C}_{\mathbf{S}} \left\{ \|\mathbf{S}_{3,h}(\widetilde{\boldsymbol{w}}_h, \widetilde{\phi}_h)\|_{1,\Omega} \left(\|\boldsymbol{w}_h - \widetilde{\boldsymbol{w}}_h\|_{1,\Omega} + \|\phi_h - \widetilde{\phi}_h\|_{1,\Omega} \right) \right.$$
$$\left. + \|\phi_h - \widetilde{\phi}_h\|_{4,\Omega} \|\mathbf{S}_{1,h}(\widetilde{\boldsymbol{w}}_h, \widetilde{\phi}_h)\|_{4,\Omega} + L_f \|\phi_h - \widetilde{\phi}_h\|_{0,\Omega} \right\}.$$

Proof It proceeds exactly as in the proof of Lemma 3.4, except for the derivation of the discrete analogue of (3.30), where, instead of choosing the values of p, q determined by the regularity parameter ε, it suffices to take $p = q = 2$, thus obtaining

$$|(\mathbf{A}_{\widetilde{\phi}_h} - \mathbf{A}_{\phi_h})(\vec{\boldsymbol{r}}_h, \vec{\mathbf{t}}_h - \vec{\boldsymbol{r}}_h)| \le \left(\alpha L_\mu (1 + \kappa_1^2)^{1/2} \|\widetilde{\phi}_h - \phi_h\|_{4,\Omega} \|\boldsymbol{r}_h\|_{4,\Omega} \right.$$
$$\left. + L_\eta c_4(\Omega)(1 + \kappa_2^2)^{1/2} \|\widetilde{\phi}_h - \phi_h\|_{1,\Omega} \|z_h\|_{1,\Omega} \right) \|\vec{\mathbf{t}}_h - \vec{\boldsymbol{r}}_h\|,$$

for all $(\boldsymbol{w}_h, \phi_h)$, $(\widetilde{\boldsymbol{w}}_h, \widetilde{\phi}_h)$, with $\vec{\mathbf{t}}_h = (\mathbf{t}_h, \boldsymbol{\sigma}_h, \boldsymbol{u}_h) := \mathbf{S}_h(\boldsymbol{w}_h, \phi_h) \in H_h$ and $\vec{\boldsymbol{r}}_h := (\boldsymbol{r}_h, \boldsymbol{\zeta}_h, z_h) = \mathbf{S}_h(\widetilde{\boldsymbol{w}}_h, \widetilde{\phi}_h) \in H_h$. Thus, since the elements of $\mathbb{H}_h^{\mathbf{t}}$ are piecewise polynomials, we know that $\|\boldsymbol{r}_h\|_{4,\Omega} < +\infty$ for each $\boldsymbol{r}_h \in \mathbb{H}_h^{\mathbf{t}}$. □

The discrete version of Lemma 3.5 is given as follows.

Lemma 3.12 *There exists a constant $\widehat{C}_{\widetilde{\mathbf{S}}} > 0$, such that for all $(\boldsymbol{w}_h, \phi_h)$, $(\widetilde{\boldsymbol{w}}_h, \widetilde{\phi}_h) \in \mathbf{H}_h$, there holds*

$$\|\widetilde{\mathbf{S}}_h(\boldsymbol{w}_h, \phi_h) - \widetilde{\mathbf{S}}_h(\widetilde{\boldsymbol{w}}_h, \widetilde{\phi}_h)\|$$
$$\le \widehat{C}_{\widetilde{\mathbf{S}}} \left\{ \|\boldsymbol{w}_h - \widetilde{\boldsymbol{w}}_h\|_{1,\Omega} \|\phi_h\|_{1,\Omega} + \|\widetilde{\boldsymbol{w}}_h\|_{1,\Omega} |\phi_h - \widetilde{\phi}_h|_{1,\Omega} + \|\widetilde{\boldsymbol{w}}_h\|_{1,\Omega} \|\phi_h - \widetilde{\phi}_h\|_{0,\Omega} \right\}.$$

Proof It follows the same arguments from Lemma 3.5, but now using the inf-sup condition (3.42) rather than the V-ellipticity of $\boldsymbol{a}$. □

Next, utilizing Lemmas 3.11 and 3.12, we can prove the discrete version of Lemma 3.6.

Lemma 3.13 *Let r and W_h as in Lemma 3.10. Then, there exists a constant $\widetilde{C}_{\mathbf{T}} > 0$, such that for all $(\boldsymbol{w}_h, \phi_h)$, $(\widetilde{\boldsymbol{w}}_h, \widetilde{\phi}_h) \in \mathbf{H}_h$, there holds*

$$\|\mathbf{T}_h(\boldsymbol{w}_h, \phi_h) - \mathbf{T}_h(\widetilde{\boldsymbol{w}}_h, \widetilde{\phi}_h)\|$$
$$\le \widetilde{C}_{\mathbf{T}} \left\{ \|\mathbf{S}_{3,h}(\widetilde{\boldsymbol{w}}_h, \widetilde{\phi}_h)\|_{1,\Omega} + \|\mathbf{S}_{1,h}(\widetilde{\boldsymbol{w}}_h, \widetilde{\phi}_h)\|_{4,\Omega} + L_f \right\} \|(\boldsymbol{w}_h, \phi_h) - (\widetilde{\boldsymbol{w}}_h, \widetilde{\phi}_h)\|.$$

Notice that the previous lemma provides the continuity required by the Brouwer fixed-point theorem, in the convex and compact set $W_h \subseteq \mathbf{H}_h$. Therefore, we have the following result.

Theorem 3.14 *Suppose that the parameters κ_1, κ_2 and κ_3 satisfy the conditions required by Lemma 3.8. Let r and W_h as in Lemma 3.10, and assume that the data satisfy (3.43). Then, the problem (3.35) has at least one solution $(\vec{\mathbf{t}}_h, (\theta_h, \lambda_h)) \in H_h \times \mathrm{H}_h^\theta \times \mathrm{H}_h^\lambda$, with $(\boldsymbol{u}_h, \theta_h) \in W_h$, and there holds*

$$\|\vec{\mathbf{t}}_h\| \le c_{\mathbf{S}} \left\{ C_f\, r + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma} \right\},$$

and

$$\|(\theta_h, \lambda_h)\| \le \widetilde{C} \left\{ r \|\boldsymbol{u}_h\|_{1,\Omega} + \|\theta_{\mathrm{D}}\|_{1/2,\Gamma} \right\}.$$

3.4 A Priori Error Analysis

Our next goal is to derive an a priori error estimate for our Galerkin scheme (3.35). More precisely, given $(\mathbf{t}, \boldsymbol{\sigma}, \boldsymbol{u}, (\theta, \lambda)) := (\vec{\mathbf{t}}, (\theta, \lambda)) \in H \times \mathrm{H}^1(\Omega) \times \mathrm{H}^{-1/2}(\Gamma)$, with $(\boldsymbol{u}, \theta) \in W$, and $(\mathbf{t}_h, \boldsymbol{\sigma}_h, \boldsymbol{u}_h, (\theta_h, \lambda_h)) := (\vec{\mathbf{t}}_h, (\theta_h, \lambda_h)) \in H_h \times \mathrm{H}_h^\theta \times \mathrm{H}_h^\lambda$, with $(\boldsymbol{u}_h, \theta_h) \in W_h$, solutions of the problems (3.7) and (3.35), respectively, we are interested in obtaining an upper bound for

$$\|(\vec{\mathbf{t}}, (\theta, \lambda)) - (\vec{\mathbf{t}}_h, (\theta_h, \lambda_h))\|.$$

To this end, we apply two instrumental results from [34, Thm. 11.1 and 11.2] concerning Strang-type estimates for elliptic and saddle point problems, respectively, where continuous and discrete formulations differ only in the functionals involved. We begin with the following preliminary estimate.

Lemma 3.15 *There exists a constant $C_{\mathbf{ST}} > 0$, independent of h, such that*

$$\begin{aligned}\|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| \;\leq\; C_{\mathbf{ST}}\Big\{&\mathrm{dist}(\vec{\mathbf{t}}, H_h) + L_f\,\|\theta - \theta_h\|_{1,\Omega} + \|\theta - \theta_h\|\,\|\mathbf{t}\|_{\varepsilon,\Omega}\\ &+\; \|\boldsymbol{u}\|_{1,\Omega}\,\|\theta - \theta_h\|_{1,\Omega} + \|\boldsymbol{u}\|_{1,\Omega}\,\|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,\Omega}\Big\}. \end{aligned} \tag{3.44}$$

Proof From Lemma (3.1) we observe that $\mathbf{A}_\theta + \mathbf{B}_{\boldsymbol{u}}$ and $\mathbf{A}_{\theta_h} + \mathbf{B}_{\boldsymbol{u}_h}$ are bounded and uniformly elliptic bilinear forms with ellipticity constant $\frac{\widetilde{\alpha}(\Omega)}{2}$. Also, $F_\theta + F_{\mathrm{D}}$ and $F_{\theta_h} + F_{\mathrm{D}}$ are linear bounded functionals in H and H_h, respectively. Thus, a straightforward application of [34, Thm. 11.1] to the context given by the first equations of (3.7) and (3.35), yields

$$\begin{aligned}\|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| \;\leq\; \overline{C}_1 \Bigg\{ &\sup_{\substack{\vec{\mathbf{s}}_h \in H_h\\ \vec{\mathbf{s}}_h \neq 0}} \frac{|F_\theta(\vec{\mathbf{s}}_h) - F_{\theta_h}(\vec{\mathbf{s}}_h)|}{\|\vec{\mathbf{s}}_h\|}\\ &+ \inf_{\substack{\vec{\mathbf{q}}_h \in H_h\\ \vec{\mathbf{q}}_h \neq 0}} \left(\|\vec{\mathbf{t}} - \vec{\mathbf{q}}_h\| + \sup_{\substack{\vec{\mathbf{s}}_h \in H_h\\ \vec{\mathbf{s}}_h \neq 0}} \frac{|(\mathbf{A}_\theta + \mathbf{B}_{\boldsymbol{u}})(\vec{\mathbf{q}}_h, \vec{\mathbf{s}}_h) - (\mathbf{A}_{\theta_h} + \mathbf{B}_{\boldsymbol{u}_h})(\vec{\mathbf{q}}_h, \vec{\mathbf{s}}_h)|}{\|\vec{\mathbf{s}}_h\|} \right) \Bigg\}, \end{aligned} \tag{3.45}$$

where $\overline{C}_1 := \frac{2}{\widetilde{\alpha}(\Omega)} \max\{1, \|\mathbf{A}_\theta + \mathbf{B}_{\boldsymbol{u}}\|\}$. Hence, in order to estimate the last supremum in (3.45), we add and subtract suitable terms to obtain

$$\begin{aligned}(\mathbf{A}_\theta + \mathbf{B}_{\boldsymbol{u}})(\vec{\mathbf{q}}_h, \vec{\mathbf{s}}_h) - (\mathbf{A}_{\theta_h} + \mathbf{B}_{\boldsymbol{u}_h})(\vec{\mathbf{q}}_h, \vec{\mathbf{s}}_h) &= (\mathbf{A}_\theta - \mathbf{A}_{\theta_h})(\vec{\mathbf{t}}, \vec{\mathbf{s}}_h) + (\mathbf{B}_{\boldsymbol{u}} - \mathbf{B}_{\boldsymbol{u}_h})(\vec{\mathbf{t}}, \vec{\mathbf{s}}_h)\\ &+ (\mathbf{A}_{\theta_h} + \mathbf{B}_{\boldsymbol{u}_h})(\vec{\mathbf{q}}_h - \vec{\mathbf{t}}, \vec{\mathbf{s}}_h) + (\mathbf{A}_\theta + \mathbf{B}_{\boldsymbol{u}})(\vec{\mathbf{q}}_h - \vec{\mathbf{t}}, \vec{\mathbf{s}}_h),\end{aligned}$$

and then, using the boundedness of the bilinear forms $\mathbf{A}_\theta + \mathbf{B}_{\boldsymbol{u}}$ and $\mathbf{A}_{\theta_h} + \mathbf{B}_{\boldsymbol{u}_h}$, the estimate (3.31), and the continuous embedding $\mathrm{H}^1(\Omega) \to \mathrm{L}^{n/\varepsilon}(\Omega)$ with constant $\widetilde{C}_\varepsilon$, we obtain

$$\begin{aligned}&|(\mathbf{A}_\theta + \mathbf{B}_{\boldsymbol{u}})(\vec{\mathbf{q}}_h, \vec{\mathbf{s}}_h) - (\mathbf{A}_{\theta_h} + \mathbf{B}_{\boldsymbol{u}_h})(\vec{\mathbf{q}}_h, \vec{\mathbf{s}}_h)|\\ &\leq \Big\{\alpha L_\mu C_\varepsilon \widetilde{C}_\varepsilon (1 + \kappa_1^2)^{1/2}\,\|\mathbf{t}\|_{\varepsilon,\Omega}\,\|\theta - \theta_h\|_{1,\Omega} + L_\eta (1 + \kappa_2^2)^{1/2}\,\|\theta - \theta_h\|_{1,\Omega}\,\|\boldsymbol{u}\|_{1,\Omega}\\ &+\; \|\boldsymbol{i}_c\|^2\,(1 + \kappa_1^2)^{1/2}\,\|\boldsymbol{u}\|_{1,\Omega}\,\|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,\Omega} + 2\|\mathbf{A}_\theta + \mathbf{B}_{\boldsymbol{u}}\|\|\vec{\mathbf{q}}_h - \vec{\mathbf{t}}\|\Big\}\|\vec{\mathbf{s}}_h\|. \end{aligned} \tag{3.46}$$

In turn, similarly as in (3.28), we note that

$$|(F_{\theta_h} - F_\theta)(\vec{\mathbf{s}}_h)| \leq L_f (1 + \kappa_2^2)^{1/2}\,\|\theta - \theta_h\|_{0,\Omega}\,\|\vec{\mathbf{s}}_h\|. \tag{3.47}$$

Finally, by replacing (3.46) and (3.47) back into (3.45), one obtains (3.44) with constant $C_{\mathbf{ST}}$ depending on $\widetilde{\alpha}(\Omega)$, L_μ, C_ε, $\widetilde{C}_\varepsilon$, L_η, $\|\boldsymbol{i}_c\|$ and $\|\mathbf{A}_\theta + \mathbf{B}_{\boldsymbol{u}}\|$. □

Next, we have the following complementary result.

Lemma 3.16 *There exists a constant $\widetilde{C}_{\mathbf{ST}} > 0$ independent of h, such that*

$$\|(\theta,\lambda) - (\theta_h,\lambda_h)\| \;\leq\; \widetilde{C}_{\mathbf{ST}} \Big\{ \mathrm{dist}\left((\theta,\lambda), \mathrm{H}_h^\theta \times \mathrm{H}_h^\lambda\right) + \|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,\Omega}\, |\theta|_{1,\Omega} \\ +\; \|\boldsymbol{u}_h\|_{1,\Omega}\, |\theta - \theta_h|_{1,\Omega} + \|\boldsymbol{u}_h\|_{1,\Omega}\, \|\theta - \theta_h\|_{0,\Omega} \Big\}. \quad (3.48)$$

Proof We first observe that (**H.0**) and (**H.1**) guarantee the main hypothesis in [34, Thm. 11.2]. Hence, by applying this lemma to the context given by the second and third equations of (3.7) and (3.35), we arrive at

$$\|(\theta,\lambda) - (\theta_h,\lambda_h)\| \leq \overline{C}_2 \left\{ \|(H_{\boldsymbol{u},\theta} - H_{\boldsymbol{u}_h,\theta_h})|_{\mathrm{H}_h^\theta}\| + \mathrm{dist}\left((\theta,\lambda), \mathrm{H}_h^\theta \times \mathrm{H}_h^\lambda\right) \right\}, \quad (3.49)$$

where $\overline{C}_2$ is a constant depending on α_1, α_2, $\|\boldsymbol{a}\|$, $\|\mathbf{b}\|$. Next, analogously to the proof of Lemma 3.5, we can assert that

$$\|(H_{\boldsymbol{u},\theta} - H_{\boldsymbol{u}_h,\theta_h})|_{\mathrm{H}_h^\theta}\| = \|(H_{\boldsymbol{u}-\boldsymbol{u}_h,\theta} + H_{\boldsymbol{u}_h,\theta} - H_{\boldsymbol{u}_h,\theta_h})|_{\mathrm{H}_h^\theta}\| \\ \leq \|i_c\|^2 \left\{ (1+s_3)\, \|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,\Omega}\, |\theta|_{1,\Omega} + \|\boldsymbol{u}_h\|_{1,\Omega}\, |\theta - \theta_h|_{1,\Omega} + L_{\tilde{s}}\, \|\boldsymbol{u}_h\|_{1,\Omega}\, \|\theta - \theta_h\|_{0,\Omega} \right\}. \quad (3.50)$$

Finally, the required estimate (3.48) follows by replacing (3.50) back into (3.49), with constant $\widetilde{C}_{\mathbf{ST}}$ depending on α_1, α_2, $\|\boldsymbol{a}\|$, $\|\mathbf{b}\|$, $\|i_c\|$, s_3 and $L_{\tilde{s}}$. □

We remark that an alternative way to prove the previous results follows similarly as in [25, Lemma 3.11] and [24, Thm. 2.6], respectively.

Having established bounds for $\|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\|$ and $\|(\theta,\lambda) - (\theta_h,\lambda_h)\|$, we are now able to derive the Céa estimate for the global error. In fact, by adding the estimates (3.44) and (3.48), and applying the continuous injection $\mathrm{H}^1(\Omega) \to \mathrm{L}^2(\Omega)$, we obtain

$$\|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| + \|(\theta,\lambda) - (\theta_h,\lambda_h)\| \leq C_{\mathbf{ST}}\, \mathrm{dist}(\vec{\mathbf{t}}, H_h) + \widetilde{C}_{\mathbf{ST}}\, \mathrm{dist}\left((\theta,\lambda), \mathrm{H}_h^\theta \times \mathrm{H}_h^\lambda\right) \\ + \left\{ C_{\mathbf{ST}}\Big(L_f + \|\mathbf{t}\|_{\varepsilon,\Omega} + \|\boldsymbol{u}\|_{1,\Omega}\Big) + 2\widetilde{C}_{\mathbf{ST}}\, \|\boldsymbol{u}_h\|_{1,\Omega} \right\} \|\theta - \theta_h\|_{1,\Omega} \\ + \left\{ C_{\mathbf{ST}}\, \|\boldsymbol{u}\|_{1,\Omega} + \widetilde{C}_{\mathbf{ST}} |\theta|_{1,\Omega} \right\} \|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,\Omega}\,.$$

Now, we note that the terms $\|\boldsymbol{u}\|_{1,\Omega}$, $|\theta|_{1,\Omega}$, $\|\boldsymbol{u}_h\|_{1,\Omega}$ and $\|\mathbf{t}\|_{\varepsilon,\Omega}$ can be bounded by data using the estimates (3.15), (3.22), (3.41) and (3.24), respectively. Therefore, performing some algebraic manipulations, and introducing the constants:

$$C_5 := C_{\mathbf{ST}} C_\varepsilon \widehat{C}(r), \quad C_6 := 2C_{\mathbf{ST}} c_{\mathbf{S}} + \widetilde{C}_{\mathbf{ST}} \widetilde{c}_{\mathbf{S}} c_{\mathbf{S}} r + 2\widetilde{C}_{\mathbf{ST}} c_{\mathbf{S}}, \\ C_7 := \max\{C_{\mathbf{ST}}, C_5, (C_5 + C_6) r, C_6, \widetilde{C}_{\mathbf{ST}} \widetilde{c}_{\mathbf{S}}\}, \quad (3.51)$$

it can be show that

$$\|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| + \|(\theta,\lambda) - (\theta_h,\lambda_h)\| \leq C_{\mathbf{ST}}\, \mathrm{dist}(\vec{\mathbf{t}}, H_h) + \widetilde{C}_{\mathbf{ST}}\, \mathrm{dist}\left((\theta,\lambda), \mathrm{H}_h^\theta \times \mathrm{H}_h^\lambda\right) \\ + C_7 \Big(L_f + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Omega} + C_f + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Omega} + \|\theta_{\mathrm{D}}\|_{1/2,\Gamma} \Big) \\ \times \left\{ \|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| + \|(\theta,\lambda) - (\theta_h,\lambda_h)\| \right\}. \quad (3.52)$$

Consequently, we can establish the following main result.

Theorem 3.17 *Assume that the data satisfy*

$$C_7\Big\{L_f + \|\boldsymbol{u}_{\rm D}\|_{1/2+\varepsilon,\Omega} + C_f + \|\boldsymbol{u}_{\rm D}\|_{1/2,\Omega} + \|\theta_{\rm D}\|_{1/2,\Gamma}\Big\} < \frac{1}{2}. \tag{3.53}$$

Then, there exists a positive constant C_8 independent of h, such that

$$\|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| + \|(\theta,\lambda) - (\theta_h,\lambda_h)\| \le C_8\Big\{\mathrm{dist}(\vec{\mathbf{t}}, H_h) + \mathrm{dist}\big((\theta,\lambda), \mathrm{H}_h^\theta \times \mathrm{H}_h^\lambda\big)\Big\}. \tag{3.54}$$

Proof It follows directly from (3.52) and (3.53). □

As a first remark of the previous theorem, we stress that the ultra-weak sense in which the symmetry of $\boldsymbol{\sigma}$ was imposed (cf. 3.5) does not affect the expected asymptotic symmetry of the discrete tensor $\boldsymbol{\sigma}_h$. In fact, adding and subtracting the symmetric unknown $\boldsymbol{\sigma}$ in the below estimate, we obtain

$$\|\boldsymbol{\sigma}_h - \boldsymbol{\sigma}_h^{\rm t}\| = \|\boldsymbol{\sigma}_h - \boldsymbol{\sigma} + \boldsymbol{\sigma}^{\rm t} - \boldsymbol{\sigma}_h^{\rm t}\| \le C_8\Big\{\mathrm{dist}(\vec{\mathbf{t}}, H_h) + \mathrm{dist}\big((\theta,\lambda), \mathrm{H}_h^\theta \times \mathrm{H}_h^\lambda\big)\Big\}, \tag{3.55}$$

which yields $\lim\limits_{h\to 0}\|\boldsymbol{\sigma}_h - \boldsymbol{\sigma}_h^{\rm t}\| = 0$, and then, we have actually proved that $\boldsymbol{\sigma}_h$ tends to a symmetric tensor. In second place, exactly as in [12, Section 4] we obtain the error for the postprocessed pressure: there exists a positive constant $\widehat{C}$, independent of h, such that

$$\|p - p_h\|_{0,\Omega} \le \widehat{C}\Big\{\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathbf{div};\Omega} + \|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,\Omega}\Big\}.$$

3.5 Specific Finite Element Subspaces

In this section we specify concrete discrete subspaces and make precise the convergence rate for (3.35). 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 local Raviart–Thomas space of order k as

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

where $\mathbf{P}_k(K) = [\mathrm{P}_k(K)]^n$, and $\boldsymbol{x}$ is the generic vector in R^n. Then, we consider piecewise polynomials of degree $\le k$ for approximating entries of the strain rate $\mathbf{t}$, the global Raviart–Thomas space of order k to approximate rows of the pseudostress $\boldsymbol{\sigma}$, and the Lagrange space given by the continuous piecewise polynomial vectors of degree $\le k+1$ for the velocity $\boldsymbol{u}$, respectively, that is

$$\begin{aligned}
\mathbb{H}_h^{\mathbf{t}} &:= \big\{\mathbf{s}_h \in \mathbb{L}^2_{\rm tr}(\Omega) : \quad \mathbf{s}_h|_K \in \mathbb{P}_k(K) \quad \forall\, K \in \mathcal{T}_h\big\},\\
\mathbb{H}_h^{\boldsymbol{\sigma}} &:= \big\{\boldsymbol{\tau}_h \in \mathbb{H}_0(\mathbf{div};\Omega) : \quad \boldsymbol{c}^{\rm t}\boldsymbol{\tau}_h|_K \in \mathbf{RT}_k(K), \quad \forall\, \boldsymbol{c} \in \mathrm{R}^n \quad \forall\, K \in \mathcal{T}_h\big\},\\
\mathbf{H}_h^{\boldsymbol{u}} &:= \big\{\boldsymbol{v}_h \in \mathbf{C}(\overline{\Omega}) : \quad \boldsymbol{v}_h|_K \in \mathbf{P}_{k+1}(K) \quad \forall\, K \in \mathcal{T}_h\big\}.
\end{aligned} \tag{3.56}$$

The approximating space for temperature will consist of continuous piecewise polynomials of degree $\le k+1$

$$\mathrm{H}_h^\theta := \big\{\psi_h \in \mathrm{C}(\overline{\Omega}) : \quad \psi_h|_K \in \mathrm{P}_{k+1}(K) \quad \forall\, K \in \mathcal{T}_h\big\}. \tag{3.57}$$

For the normal heat flux, we let $\{\widetilde{\Gamma}_1, \widetilde{\Gamma}_2, \ldots, \widetilde{\Gamma}_m\}$ be an independent triangulation of Γ (made of straight segments in R^2, or triangles in R^3), and define $\widetilde{h} := \max_{j\in\{1,\ldots,m\}} |\widetilde{\Gamma}_j|$. Then, with the same integer $k \ge 0$ used in definitions (3.56) and (3.57), we approximate λ by piecewise polynomials of degree $\le k$ over this new mesh, that is

$$\mathrm{H}_{\widetilde{h}}^\lambda := \Big\{\xi_{\widetilde{h}} \in \mathrm{L}^2(\Gamma) : \quad \xi_{\widetilde{h}}|_{\widetilde{\Gamma}_j} \in \mathrm{P}_k(\widetilde{\Gamma}_j) \quad \forall\, j \in \{1,\ldots,m\}\Big\}. \tag{3.58}$$

We remark that the spaces H_h^θ and $\mathrm{H}_{\widetilde{h}}^\lambda$ satisfy the inf-sup conditions $\mathbf{H.0}$ and $\mathbf{H.1}$. We remit to (cf. [14, Lemma 4.10], [24, Lemma 4.7]) for further details.

Finally, approximation properties of the spaces in (3.56), (3.57) and (3.58) can be found in e.g [4,10,24], which combined with the Céa estimate (3.54) produce the theoretical rate of convergence of (3.35), summarized in what follows.

Theorem 3.18 *In addition to the hypotheses of Theorems 3.7, 3.14 and 3.17, assume that there exists $s > 0$ such that $\mathbf{t} \in \mathbb{H}^s(\Omega)$, $\boldsymbol{\sigma} \in \mathbb{H}^s(\Omega)$, $\mathbf{div}\,\boldsymbol{\sigma} \in \mathbf{H}^s(\Omega)$, $\boldsymbol{u} \in \mathbf{H}^{1+s}(\Omega)$, $\theta \in \mathrm{H}^{1+s}(\Omega)$ and $\lambda \in \mathrm{H}^{-1/2+s}(\Gamma)$. Then, there exist positive constants C_0, $C > 0$, independent of h and $\widetilde{h}$, such that for all $h \le C_0\widetilde{h}$, with the finite element subspaces defined by (3.56), (3.57) and (3.58), there holds*

$$\begin{aligned}\|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| + \big\|(\theta,\lambda) - (\theta_h, \lambda_{\widetilde{h}})\big\| \;&\le\; C\widetilde{h}^{\min\{s,k+1\}}\,\|\lambda\|_{-1/2+s,\Gamma}\\ &+ Ch^{\min\{s,k+1\}}\Big\{\|\mathbf{t}\|_{s,\Omega} + \|\boldsymbol{\sigma}\|_{s,\Omega} + \|\mathbf{div}\,\boldsymbol{\sigma}\|_{s,\Omega} + \|\boldsymbol{u}\|_{1+s,\Omega} + \|\theta\|_{1+s,\Omega}\Big\}.\end{aligned}$$

Finally, we point out that (3.55) and the previous theorem imply that, under the same foregoing regularity assumptions, the approximating unknown $\boldsymbol{\sigma}_h$ converges to a symmetric tensor with the same rate of convergence of all the unknowns involved.

4 The Fully-Mixed Approach

In this section we proceed similarly as in [15] to put forward a fully-mixed approach for (2.1). Then, we establish the corresponding continuous and discrete formulations, analyze their solvability by using fixed-point strategies, and derive the corresponding a priori error estimates.

4.1 The Continuous Formulation

Having established in Sect. 3 the mixed formulation for the Navier–Stokes–Brinkman problem, it only remains to define a mixed formulation for the energy equation. Let us introduce the unknown

$$\boldsymbol{\Theta} := \rho\kappa\nabla\theta - \theta\boldsymbol{u} - s(\theta)\boldsymbol{u} \quad \text{in } \Omega,$$

and then, denoting from now on the tensor $\rho^{-1}\kappa^{-1}$ simply as κ^{-1}, applying (2.1b) and performing some algebraic computations, we obtain

$$\kappa^{-1}\boldsymbol{\Theta} + \kappa^{-1}\theta\boldsymbol{u} + \kappa^{-1}s(\theta)\boldsymbol{u} = \nabla\theta \quad \text{in } \Omega, \quad \operatorname{div}\boldsymbol{\Theta} = 0 \quad \text{in } \Omega, \quad \theta = \theta_{\mathrm{D}} \quad \text{on } \Gamma. \tag{4.1}$$

In this way, testing the first equation in (4.1) against functions $\boldsymbol{\Phi} \in \mathbf{H}(\mathrm{div};\Omega)$, integrating by parts, and using the Dirichlet boundary condition for θ, we obtain

$$\int_\Omega \kappa^{-1}\boldsymbol{\Theta}\cdot\Phi + \int_\Omega \theta\,\operatorname{div}\Phi + \int_\Omega \kappa^{-1}\theta\boldsymbol{u}\cdot\boldsymbol{\Phi} \;=\; -\int_\Omega \kappa^{-1}s(\theta)\boldsymbol{u}\cdot\boldsymbol{\Phi} + \langle\boldsymbol{\Phi}\cdot\boldsymbol{\nu}, \theta_{\mathrm{D}}\rangle_\Gamma\,. \tag{4.2}$$

In turn, testing the equilibrium equation in (4.1) against a suitable function ψ, we get

$$-\int_\Omega \psi\,\operatorname{div}\boldsymbol{\Theta} \;=\; 0.$$

Similarly as in Sect. 3, we note from the last term on the left-hand side of (4.2), that we require to seek the temperature θ in $\mathrm{H}^1(\Omega)$. Thus we are left with the preliminary weak formulation: Find $(\boldsymbol{\Theta},\theta) \in \mathbf{H}(\mathrm{div};\Omega) \times \mathrm{H}^1(\Omega)$, such that

$$
\begin{aligned}
\int_\Omega \kappa^{-1}\boldsymbol{\Theta}\cdot\Phi + \int_\Omega \theta\,\mathrm{div}\,\Phi + \int_\Omega \kappa^{-1}\theta\boldsymbol{u}\cdot\boldsymbol{\Phi} &= -\int_\Omega \kappa^{-1}s(\theta)\boldsymbol{u}\cdot\boldsymbol{\Phi} + \langle\boldsymbol{\Phi}\cdot\boldsymbol{\nu},\theta_{\mathrm{D}}\rangle_\Gamma\,,\\
-\int_\Omega \psi\,\mathrm{div}\,\boldsymbol{\Theta} &= 0\,,
\end{aligned}
\tag{4.3}
$$

for all $(\boldsymbol{\Phi},\psi) \in \mathbf{H}(\mathrm{div};\Omega)\times \mathrm{H}^1(\Omega)$. Again, the analysis will require to incorporate the following redundant terms:

$$
\begin{aligned}
\kappa_4\int_\Omega \left(\nabla\theta - \kappa^{-1}\theta\boldsymbol{u} - \kappa^{-1}s(\theta)\boldsymbol{u} - \kappa^{-1}\boldsymbol{\Theta}\right)\cdot\nabla\psi &= 0 \quad \forall\,\psi\in\mathrm{H}^1(\Omega),\\
\kappa_5\int_\Omega \mathrm{div}\,\boldsymbol{\Theta}\,\mathrm{div}\,\boldsymbol{\Phi} &= 0 \quad \forall\,\boldsymbol{\Phi}\in\mathbf{H}(\mathrm{div};\Omega),\\
\kappa_6\int_\Gamma \theta\,\psi &= \kappa_6\int_\Gamma \theta_{\mathrm{D}}\,\psi \quad \forall\,\psi\in\mathrm{H}^1(\Omega),
\end{aligned}
$$

where κ_4,κ_5 and κ_6 are positive parameters to be specified later on. Then, now we may consider the following mixed formulation for the energy equation: Find $(\boldsymbol{\Theta},\theta)\in\mathbf{H}(\mathrm{div};\Omega)\times \mathrm{H}^1(\Omega)$, such that

$$
\widetilde{\boldsymbol{a}}((\boldsymbol{\Theta},\theta),(\boldsymbol{\Phi},\psi)) + \widetilde{\mathbf{b}}_{\boldsymbol{u}}((\boldsymbol{\Theta},\theta),(\boldsymbol{\Phi},\psi)) = \widetilde{F}_{\boldsymbol{u},\theta}(\boldsymbol{\Phi},\psi) + \widetilde{F}_{\mathrm{D}}(\boldsymbol{\Phi},\psi),
\tag{4.4}
$$

for all $(\boldsymbol{\Phi},\psi)\in\mathbf{H}(\mathrm{div};\Omega)\times\mathrm{H}^1(\Omega)$, where, given an arbitrary $(\boldsymbol{w},\phi)\in\mathbf{H}$, the forms $\widetilde{\boldsymbol{a}}$, $\widetilde{\mathbf{b}}_{\boldsymbol{w}}$ and the functionals $\widetilde{F}_{\boldsymbol{w},\phi}$ and $\widetilde{F}_{\mathrm{D}}$ are defined, respectively, as

$$
\begin{aligned}
\widetilde{\boldsymbol{a}}((\boldsymbol{\Theta},\theta),(\boldsymbol{\Phi},\psi)) :=& \int_\Omega \kappa^{-1}\boldsymbol{\Theta}\cdot(\Phi-\kappa_4\nabla\psi) + \int_\Omega \theta\,\mathrm{div}\,\Phi - \int_\Omega \psi\,\mathrm{div}\,\boldsymbol{\Theta}\\
&+\kappa_4\int_\Omega \nabla\theta\cdot\nabla\psi + \kappa_5\int_\Omega \mathrm{div}\,\boldsymbol{\Theta}\,\mathrm{div}\,\boldsymbol{\Phi} + \kappa_6\int_\Gamma \theta\,\psi\,,
\end{aligned}
\tag{4.5a}
$$

$$
\widetilde{\mathbf{b}}_{\boldsymbol{w}}((\boldsymbol{\Theta},\theta),(\boldsymbol{\Phi},\psi)) := \int_\Omega \kappa^{-1}\theta\boldsymbol{w}\cdot(\boldsymbol{\Phi}-\kappa_4\nabla\psi)\,,
\tag{4.5b}
$$

for all $(\boldsymbol{\Theta},\theta),(\boldsymbol{\Phi},\psi)\in\mathbf{H}(\mathrm{div};\Omega)\times\mathrm{H}^1(\Omega)$, and

$$
\widetilde{F}_{\boldsymbol{w},\phi}(\boldsymbol{\Phi},\psi) := \int_\Omega \kappa^{-1}s(\phi)\boldsymbol{w}\cdot(\kappa_4\nabla\psi-\boldsymbol{\Phi}),
\tag{4.6a}
$$

$$
\widetilde{F}_{\mathrm{D}}(\boldsymbol{\Phi},\psi) := \langle\boldsymbol{\Phi}\cdot\boldsymbol{\nu},\theta_{\mathrm{D}}\rangle_\Gamma + \kappa_6\int_\Gamma \theta_{\mathrm{D}}\,\psi,
\tag{4.6b}
$$

for all $(\boldsymbol{\Phi},\psi)\in\mathbf{H}(\mathrm{div};\Omega)\times\mathrm{H}^1(\Omega)$. The fully-mixed variational formulation for (2.1) reduces therefore to the first equation of (3.7) and (4.4), i.e.: Find $(\vec{\mathbf{t}},(\boldsymbol{\Theta},\theta))\in H\times\mathbf{H}(\mathrm{div};\Omega)\times\mathrm{H}^1(\Omega)$ such that

$$
\begin{aligned}
\mathbf{A}_\theta(\vec{\mathbf{t}},\vec{\mathbf{s}}) + \mathbf{B}_{\boldsymbol{u}}(\vec{\mathbf{t}},\vec{\mathbf{s}}) &= F_\theta(\vec{\mathbf{s}}) + F_{\mathrm{D}}(\vec{\mathbf{s}}),\\
\widetilde{\boldsymbol{a}}((\boldsymbol{\Theta},\theta),(\boldsymbol{\Phi},\psi)) + \widetilde{\mathbf{b}}_{\boldsymbol{u}}((\boldsymbol{\Theta},\theta),(\boldsymbol{\Phi},\psi)) &= \widetilde{F}_{\boldsymbol{u},\theta}(\boldsymbol{\Phi},\psi) + \widetilde{F}_{\mathrm{D}}(\boldsymbol{\Phi},\psi),
\end{aligned}
\tag{4.7}
$$

for all $(\vec{\mathbf{s}},(\boldsymbol{\Phi},\psi))\in H\times\mathbf{H}(\mathrm{div};\Omega)\times\mathrm{H}^1(\Omega)$.

We end this section by noticing that the present use of a mixed approach for the heat equation avoids the introduction of the unknown given by the normal boundary heat flux λ, as it was required in the primal formulation from Sect. 3.

4.2 Solvability Analysis

The forms $\widetilde{\boldsymbol{a}}$ and $\widetilde{\mathbf{b}}_{\boldsymbol{u}}$ are defined exactly as in [15, Section 3.1] and therefore we omit parts of the proofs whenever necessary. On the other hand, for the solvability of (4.7), we propose a fixed-point approach as in Sect. 3.2. More precisely, in addition to using the operator $\mathbf{S}$ (cf. (3.10)–(3.11)), and instead of (3.12) and (3.14), we define the operators $\widehat{\mathbf{S}} : \mathbf{H} \to \mathrm{H}^1(\Omega)$ and $\widehat{\mathbf{T}} := \mathbf{H} \to \mathbf{H}$ as $\widehat{\mathbf{S}}(\boldsymbol{w}, \phi) := \theta \quad \forall\, (\boldsymbol{w}, \phi) \in \mathbf{H}$, where θ is the second component of the unique solution $(\boldsymbol{\Theta}, \theta) \in \mathbf{H}(\mathrm{div}; \Omega) \times \mathrm{H}^1(\Omega)$ of the problem given by the second equation of (4.7) with $(\boldsymbol{w}, \phi)$ instead of $(\boldsymbol{u}, \theta)$, that is

$$\widetilde{\boldsymbol{a}}((\boldsymbol{\Theta}, \theta), (\boldsymbol{\Phi}, \psi)) + \widetilde{\mathbf{b}}_{\boldsymbol{w}}((\boldsymbol{\Theta}, \theta), (\boldsymbol{\Phi}, \psi)) = \widetilde{F}_{\boldsymbol{w},\phi}(\boldsymbol{\Phi}, \psi) + \widetilde{F}_{\mathrm{D}}(\boldsymbol{\Phi}, \psi), \tag{4.8}$$

for all $(\boldsymbol{\Phi}, \psi) \in \mathbf{H}(\mathrm{div}; \Omega) \times \mathrm{H}^1(\Omega)$, and

$$\widehat{\mathbf{T}}(\boldsymbol{w}, \phi) = \big(\mathbf{S}_3(\boldsymbol{w}, \phi), \widehat{\mathbf{S}}(\mathbf{S}_3(\boldsymbol{w}, \phi), \phi)\big) \quad \forall\, (\boldsymbol{w}, \phi) \in \mathbf{H},$$

respectively. A first result concerning the solvability of the mixed formulation (4.8) is provided next.

Lemma 4.1 *Assume that* $\kappa_4 \in \left(0, \frac{2\widetilde{K}_0\delta_4}{\widetilde{K}_1}\right)$, *with* $\delta_4 \in \left(0, \frac{2}{\widetilde{K}_1}\right)$, *and* $\kappa_5, \kappa_6 > 0$. *Then, there exists* $\widetilde{r}_0 > 0$ *such that for each* $\widetilde{r} \in (0, \widetilde{r}_0)$, *problem* (4.8) *has a unique solution* $(\boldsymbol{\Theta}, \widehat{\mathbf{S}}(\boldsymbol{w}, \phi)) := (\boldsymbol{\Theta}, \theta) \in \mathbf{H}(\mathrm{div}; \Omega) \times \mathrm{H}^1(\Omega)$ *for each* $(\boldsymbol{w}, \phi) \in \mathbf{H}$ *with* $\|\boldsymbol{w}\|_{1,\Omega} \le \widetilde{r}$. *Moreover, there exists* $k_{\mathbf{S}} > 0$, *independent of* $(\boldsymbol{w}, \phi)$, *such that*

$$\|\widehat{\mathbf{S}}(\boldsymbol{w}, \phi)\| \;=\; \|\theta\|_{1,\Omega} \;\le\; \|(\boldsymbol{\Theta}, \theta)\| \le k_{\mathbf{S}} \left\{\|\boldsymbol{w}\|_{0,\Omega} + \|\theta_{\mathrm{D}}\|_{0,\Gamma} + \|\theta_{\mathrm{D}}\|_{1/2,\Gamma}\right\} \quad \forall\, (\boldsymbol{w}, \phi) \in \mathbf{H}. \tag{4.9}$$

Proof From [15, Lemma 3.3] we recall that the bilinear form $\widetilde{\boldsymbol{a}} + \widetilde{\mathbf{b}}_{\boldsymbol{w}}$ [cf. (4.5a), (4.5b)] is elliptic with constant $\frac{\widetilde{\alpha}_1(\Omega)}{2}$, provided $\|\boldsymbol{w}\|_{1,\Omega} \le \widetilde{r}_0$, with

$$\widetilde{r}_0 := \frac{\widetilde{\alpha}_1(\Omega)}{2\,\|i_c\|^2\,(\Omega)(1+\kappa_4^2)^{1/2}\,\widetilde{K}_1}. \tag{4.10}$$

Now, from (4.6a) and (4.6b) we note that the functionals $\widetilde{F}_{\boldsymbol{w},\phi}$ and $\widetilde{F}_{\mathrm{D}}$ are bounded with

$$\|\widetilde{F}_{\boldsymbol{w},\phi}\| \le \widetilde{K}_1 s_2 (1+\kappa_4^2)^{1/2}\,\|\boldsymbol{w}\|_{0,\Omega} \quad \text{and} \quad \|\widetilde{F}_{\mathrm{D}}\| \le \kappa_6 c_0(\Omega)\,\|\theta_{\mathrm{D}}\|_{0,\Gamma} + \|\theta_{\mathrm{D}}\|_{1/2,\Gamma}\,,$$

where $c_0(\Omega)$ is the norm of the trace operator in $\mathrm{H}^1(\Omega)$. Finally, a direct application of the Lax–Milgram lemma proves that for each $(\boldsymbol{w}, \phi) \in \mathbf{H}$, problem (4.8) has a unique solution $(\boldsymbol{\Theta}, \theta) \in \mathbf{H}(\mathrm{div}; \Omega) \times \mathrm{H}^1(\Omega)$. Moreover, the continuous dependence result establishes that

$$\|\widehat{\mathbf{S}}(\boldsymbol{w}, \phi)\| \le \|(\boldsymbol{\Theta}, \theta)\| \le \frac{2}{\widetilde{\alpha}_1}\|\widetilde{F}_{\boldsymbol{w},\phi} + \widetilde{F}_{\mathrm{D}}\| \le k_{\mathbf{S}} \left\{\|\boldsymbol{w}\|_{0,\Omega} + \|\theta_{\mathrm{D}}\|_{0,\Gamma} + \|\theta_{\mathrm{D}}\|_{1/2,\Gamma}\right\},$$

where $k_{\mathbf{S}} := \frac{2}{\widetilde{\alpha}_1}\max\{\widetilde{K}_1 s_2(1+\kappa_4^2)^{1/2}, \kappa_6 c_0(\Omega), 1\}$, which ends the proof. □

The analogue of Lemma 3.3 is stated next.

Lemma 4.2 *Given* $r \in (0, \min\{r_0, \widetilde{r}_0\})$, *with* r_0 *and* $\widetilde{r}_0$ *given by* (3.20) *and* (4.10), *respectively, we let* $\widehat{W} := \{(\boldsymbol{w}, \phi) \in \mathbf{H} : \|(\boldsymbol{w}, \phi)\| \le r\}$, *and assume that*

$$c(r)\left\{C_f + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma}\right\} + k_{\mathbf{S}}\left\{\|\theta_{\mathrm{D}}\|_{0,\Gamma} + \|\theta_{\mathrm{D}}\|_{1/2,\Gamma}\right\} \le r, \tag{4.11}$$

where $c(r) := (1+k_{\mathbf{S}})\,c_{\mathbf{S}}\max\{1, r\}$, *and* $c_{\mathbf{S}}$ *and* $k_{\mathbf{S}}$ *are the constants specified in Lemmas* 3.1 *and* 4.1, *respectively. Then* $\widehat{\mathbf{T}}(\widehat{W}) \subseteq \widehat{W}$.

Proof It follows exactly as in [15, Lemma 3.5]. □

Next, we aim to prove the continuity of $\widehat{\mathbf{T}}$, which basically will be direct consequence of Lemma 3.4 and the following result providing the continuity of $\mathbf{S}$ and $\widehat{\mathbf{S}}$, respectively.

Lemma 4.3 *There exists $\widetilde{K}_{\widetilde{\mathbf{S}}} > 0$, such that for all $(\boldsymbol{w}_1, \phi_1), (\boldsymbol{w}_2, \phi_2) \in \mathbf{H}$, there holds*

$$\begin{aligned}&\|\widehat{\mathbf{S}}(\boldsymbol{w}_1, \phi_1) - \widehat{\mathbf{S}}(\boldsymbol{w}_2, \phi_2)\|\\&\quad \leq \widetilde{K}_{\widetilde{\mathbf{S}}}\Big\{\|\widehat{\mathbf{S}}(\boldsymbol{w}_2, \phi_2)\|_{1,\Omega}\,\|\boldsymbol{w}_1 - \boldsymbol{w}_2\|_{1,\Omega} + \|\boldsymbol{w}_2\|_{1,\Omega}\,\|\phi_1 - \phi_2\|_{0,\Omega} + \|\boldsymbol{w}_1 - \boldsymbol{w}_2\|_{1,\Omega}\Big\}.\end{aligned} \tag{4.12}$$

Proof Given $r \in (0, \widetilde{r}_0)$, and $(\boldsymbol{w}_1, \phi_1), (\boldsymbol{w}_2, \phi_2) \in \mathbf{H}$ with $\|\boldsymbol{w}_1\|_{1,\Omega}, \|\boldsymbol{w}_2\|_{1,\Omega} \leq r$, we let $(\boldsymbol{\Theta}_1, \theta_1), (\boldsymbol{\Theta}_2, \theta_2) \in \mathbf{H}(\mathrm{div};\Omega) \times \mathrm{H}^1(\Omega)$ be solutions to (4.8) corresponding to $(\boldsymbol{w}_1, \phi_1)$ and $(\boldsymbol{w}_2, \phi_2)$, respectively, that is

$$\widetilde{\boldsymbol{a}}((\boldsymbol{\Theta}_1, \theta_1), (\boldsymbol{\Phi}, \psi)) + \widetilde{\mathbf{b}}_{\boldsymbol{w}_1}((\boldsymbol{\Theta}_1, \theta_1), (\boldsymbol{\Phi}, \psi)) = \widetilde{F}_{\boldsymbol{w}_1,\phi_1}(\boldsymbol{\Phi}, \psi) + \widetilde{F}_{\mathrm{D}}(\boldsymbol{\Phi}, \psi),$$

and

$$\widetilde{\boldsymbol{a}}((\boldsymbol{\Theta}_2, \theta_2), (\boldsymbol{\Phi}, \psi)) + \widetilde{\mathbf{b}}_{\boldsymbol{w}_2}((\boldsymbol{\Theta}_2, \theta_2), (\boldsymbol{\Phi}, \psi)) = \widetilde{F}_{\boldsymbol{w}_2,\phi_2}(\boldsymbol{\Phi}, \psi) + \widetilde{F}_{\mathrm{D}}(\boldsymbol{\Phi}, \psi),$$

for all $(\boldsymbol{\Phi}, \psi) \in \mathbf{H}(\mathrm{div};\Omega) \times \mathrm{H}^1(\Omega)$. Then, similarly to Lemma 3.4, we add and subtract suitable terms to get

$$\begin{aligned}(\widetilde{\boldsymbol{a}} + \widetilde{\mathbf{b}}_{\boldsymbol{w}_2})((\boldsymbol{\Theta}_1, \theta_1) - (\boldsymbol{\Theta}_2, \theta_2), (\boldsymbol{\Theta}_1, \theta_1) - (\boldsymbol{\Theta}_2, \theta_2))&\\ = -\widetilde{\mathbf{b}}_{\boldsymbol{w}_1 - \boldsymbol{w}_2}((\boldsymbol{\Theta}_1, \theta_1), (\boldsymbol{\Theta}_1, \theta_1) - (\boldsymbol{\Theta}_2, \theta_2))&\\ + (\widetilde{F}_{\boldsymbol{w}_1,\phi_1} - \widetilde{F}_{\boldsymbol{w}_2,\phi_2})((\boldsymbol{\Theta}_1, \theta_1) - (\boldsymbol{\Theta}_2, \theta_2))\,,&\end{aligned}$$

from which, applying the ellipticity of $\widetilde{\boldsymbol{a}} + \widetilde{\mathbf{b}}_{\boldsymbol{w}_2}$, we deduce that

$$\begin{aligned}&\frac{\widetilde{\alpha}_1}{2}\|(\boldsymbol{\Theta}_1, \theta_1) - (\boldsymbol{\Theta}_2, \theta_2)\|^2\\&\quad \leq -\widetilde{\mathbf{b}}_{\boldsymbol{w}_1 - \boldsymbol{w}_2}((\boldsymbol{\Theta}_1, \theta_1), (\boldsymbol{\Theta}_1, \theta_1) - (\boldsymbol{\Theta}_2, \theta_2)) + (\widetilde{F}_{\boldsymbol{w}_1,\phi_1} - \widetilde{F}_{\boldsymbol{w}_2,\phi_2})((\boldsymbol{\Theta}_1, \theta_1) - (\boldsymbol{\Theta}_2, \theta_2))\\&\quad \leq \widetilde{K}_1\left\{(1 + \kappa_4^2)^{1/2}\,\|i_c\|^2\,\|\theta_1\|_{1,\Omega}\,\|\boldsymbol{w}_1 - \boldsymbol{w}_2\| + L_s(1 + \kappa_4^2)^{1/2}\,\|\boldsymbol{w}_2\|_{1,\Omega}\,\|\phi_1 - \phi_2\|_{1,\Omega}\right.\\&\quad \left. + s_2\,\|\boldsymbol{w}_1 - \boldsymbol{w}_2\|_{0,\Omega}\right\}\|(\boldsymbol{\Theta}_1, \theta_1) - (\boldsymbol{\Theta}_2, \theta_2)\|\,.\end{aligned}$$

The foregoing inequality yields (4.12) with $\widetilde{K}_{\widetilde{\mathbf{S}}} := \frac{2\widetilde{K}_1}{\widetilde{\alpha}_1}\max\{(1 + \kappa_4^2)^{1/2}\,\|i_c\|^2\,, L_s(1 + \kappa_4^2)^{1/2}, s_2\}$, which finishes the proof. □

We are now in a position to establish the announced property of the operator $\widehat{\mathbf{T}}$. We omit the corresponding proof and refer to [15, Lemma 3.8] for details.

Lemma 4.4 *Given $r \in (0, \min\{r_0, \widetilde{r}_0\})$, with r_0 and $\widetilde{r}_0$ given by (3.20) and (4.10), respectively, we let $\widehat{W}$ as in Lemma 4.2. Then, there exists a constant $K_{\mathbf{T}} > 0$ such that for all $(\boldsymbol{w}_1, \phi_1), (\boldsymbol{w}_2, \phi_2) \in \widehat{W}$, there holds*

$$\begin{aligned}&\|\widehat{\mathbf{T}}(\boldsymbol{w}_1, \phi_1) - \widehat{\mathbf{T}}(\boldsymbol{w}_2, \phi_2)\|\\&\quad \leq K_{\mathbf{T}}\Big\{C_f + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma} + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma} + L_f\Big\}\,\|(\boldsymbol{w}_1, \phi_1) - (\boldsymbol{w}_2, \phi_2)\|\,.\end{aligned} \tag{4.13}$$

The existence and uniqueness of a fixed point of $\widehat{\mathbf{T}}$ (and therefore well-posedness of 4.7), is stated as follows.

Theorem 4.5 *Suppose that the parameters κ_4, κ_5 and κ_6 satisfy the conditions required by Lemma 4.1. In addition, let r and $\widehat{W}$ as in Lemma 4.2, and assume that the data verify (4.11) and*

$$K_{\mathbf{T}}\Big\{C_f + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma} + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma} + L_f\Big\} < 1. \tag{4.14}$$

Then (4.7) has a unique solution $(\vec{\mathbf{t}}, (\boldsymbol{\Theta}, \theta)) \in H \times \mathbf{H}(\mathrm{div};\Omega) \times \mathrm{H}^1(\Omega)$ with $(\boldsymbol{u}, \theta) \in \widehat{W}$. Moreover

$$\|\vec{\mathbf{t}}\| \le c_{\mathbf{S}}\left\{C_f\, r + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma}\right\},$$

and

$$\|(\boldsymbol{\Theta}, \theta)\| \le k_{\mathbf{S}}\{\|\boldsymbol{u}\|_{1,\Omega} + \|\theta_{\mathrm{D}}\|_{0,\Gamma} + \|\theta_{\mathrm{D}}\|_{1/2,\Gamma}\}.$$

Proof It suffices to apply the Banach fixed-point Theorem (bearing in mind (4.13) - (4.14)), and then employ the a priori estimates (3.15) and (4.9). We omit further details. □

4.3 The Galerkin Scheme

Similarly to Sect. 3.3, we begin by considering the arbitrary finite dimensional subspaces

$$\begin{gathered}\mathbb{H}_h^{\mathbf{t}} \subseteq \mathbb{L}^2_{\mathtt{tr}}(\Omega), \quad \mathbb{H}_h^{\boldsymbol{\sigma}} \subseteq \mathbb{H}_0(\mathbf{div};\Omega), \quad \mathbf{H}_h^{\boldsymbol{u}} \subseteq \mathbf{H}^1(\Omega),\\ \mathbf{H}_h^{\boldsymbol{\Theta}} \subseteq \mathbf{H}(\mathrm{div};\Omega), \quad \text{and} \quad \mathrm{H}_h^{\theta} \subseteq \mathrm{H}^1(\Omega).\end{gathered} \tag{4.15}$$

A Galerkin scheme for (4.7) then reads: Find $(\vec{\mathbf{t}}_h, (\boldsymbol{\Theta}_h, \theta_h)) \in H_h \times \mathrm{H}_h^{\boldsymbol{\Theta}} \times \mathrm{H}_h^{\theta}$ such that

$$\begin{aligned}\mathbf{A}_{\theta_h}(\vec{\mathbf{t}}_h, \vec{\mathbf{s}}_h) + \mathbf{B}_{\boldsymbol{u}_h}(\vec{\mathbf{t}}_h, \vec{\mathbf{s}}_h) &= F_{\theta_h}(\vec{\mathbf{s}}_h) + F_{\mathrm{D}}(\vec{\mathbf{s}}_h),\\ \widetilde{\boldsymbol{a}}((\boldsymbol{\Theta}_h, \theta_h), (\boldsymbol{\Phi}_h, \psi_h)) + \widetilde{\mathbf{b}}_{\boldsymbol{u}_h}((\boldsymbol{\Theta}_h, \theta_h), (\boldsymbol{\Phi}_h, \psi_h)) &= \widetilde{F}_{\boldsymbol{u}_h,\theta_h}(\boldsymbol{\Phi}_h, \psi_h) + \widetilde{F}_{\mathrm{D}}(\boldsymbol{\Phi}_h, \psi_h),\end{aligned} \tag{4.16}$$

for all $(\vec{\mathbf{s}}_h, (\boldsymbol{\Phi}_h, \psi_h)) \in H_h \times \mathbf{H}_h^{\boldsymbol{\Theta}} \times \mathrm{H}_h^{\theta}$. We emphasize that the analysis of (4.16) uses the discrete version of the fixed-point strategy from Sect. 4.2. Results and the used arguments are almost verbatim to those in that section, and we omit them here simply stating the main result.

Theorem 4.6 *Suppose that the parameters κ_4, κ_5 and κ_6 satisfy the conditions required by Lemma 4.1. In addition, let $\widehat{W}_h := \big\{(\boldsymbol{w}_h, \phi_h) \in \mathbf{H}_h^{\boldsymbol{u}} \times \mathrm{H}_h^{\theta} : \|(\boldsymbol{w}_h, \phi_h)\| \le r\big\}$, with r defined as in Lemma 4.2, and assume that the data satisfy (4.11). Then, the problem (4.16) has at least one solution $(\vec{\mathbf{t}}_h, (\boldsymbol{\Theta}_h, \theta_h)) \in H_h \times \mathbf{H}_h^{\boldsymbol{\Theta}} \times \mathrm{H}_h^{\theta}$, with $(\boldsymbol{u}_h, \theta_h) \in \widehat{W}_h$, and there holds*

$$\|\vec{\mathbf{t}}_h\| \le c_{\mathbf{S}}\left\{C_f\, r + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma}\right\},$$

and

$$\|(\boldsymbol{\Theta}_h, \theta_h)\| \le k_{\mathbf{S}}\{\|\boldsymbol{u}_h\|_{1,\Omega} + \|\theta_{\mathrm{D}}\|_{0,\Gamma} + \|\theta_{\mathrm{D}}\|_{1/2,\Gamma}\}.$$

4.4 A Priori Error Analysis

Let $(\boldsymbol{\Theta}, \theta)$ and $(\boldsymbol{\Theta}_h, \theta_h)$ be solutions to the problems

$$\begin{aligned}\widetilde{\boldsymbol{a}}((\boldsymbol{\Theta}, \theta), (\boldsymbol{\Phi}, \psi)) + \widetilde{\mathbf{b}}_{\boldsymbol{u}}((\boldsymbol{\Theta}, \theta), (\boldsymbol{\Phi}, \psi)) &= \widetilde{F}_{\boldsymbol{u},\theta}(\boldsymbol{\Phi}, \psi) + \widetilde{F}_{\mathrm{D}}(\boldsymbol{\Phi}, \psi) \quad \text{and}\\ \widetilde{\boldsymbol{a}}((\boldsymbol{\Theta}_h, \theta_h), (\boldsymbol{\Phi}_h, \psi_h)) + \widetilde{\mathbf{b}}_{\boldsymbol{u}_h}((\boldsymbol{\Theta}_h, \theta_h), (\boldsymbol{\Phi}_h, \psi_h)) &= \widetilde{F}_{\boldsymbol{u}_h,\theta_h}(\boldsymbol{\Phi}_h, \psi_h) + \widetilde{F}_{\mathrm{D}}(\boldsymbol{\Phi}_h, \psi_h),\end{aligned} \tag{4.17}$$

for all $(\boldsymbol{\Phi}, \psi) \in \mathbf{H}(\mathrm{div};\Omega) \times \mathrm{H}^1(\Omega)$, and for all $(\boldsymbol{\Phi}_h, \psi_h) \in \mathbf{H}_h^{\boldsymbol{\Theta}} \times \mathrm{H}_h^{\theta}$, respectively. A preliminary error estimate is provided by the following lemma.

Lemma 4.7 *There exists a positive constant $K_{\mathbf{ST}}$, independent of h, such that*

$$\|(\boldsymbol{\Theta},\theta)-(\boldsymbol{\Theta}_h,\theta_h)\| \le K_{\mathbf{ST}}\Big\{\left(1+\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}\right)\mathrm{dist}\left((\boldsymbol{\Theta},\theta),\mathbf{H}_h^{\boldsymbol{\Theta}}\times\mathrm{H}_h^{\theta}\right)$$
$$+\ \|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}\,\|\theta\|_{1,\Omega} + \|\boldsymbol{u}_h\|_{1,\Omega}\,\|\theta-\theta_h\|_{1,\Omega} + s_2\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{0,\Omega}\Big\}. \tag{4.18}$$

Proof Proceeding as in the proof of Lemma 3.15, a straightforward application of the Strang lemma provided in [34, Thm. 11.1] to the context (4.17), yields

$$\|(\boldsymbol{\Theta},\theta)-(\boldsymbol{\Theta}_h,\theta_h)_h\| \le \overline{K}_1\left\{\sup_{\substack{(\boldsymbol{\Phi}_h,\psi_h)\in\mathbf{H}_h^{\boldsymbol{\Theta}}\times\mathrm{H}_h^{\theta}\\(\boldsymbol{\Phi}_h,\psi_h)\neq 0}}\frac{|\widetilde{F}_{\boldsymbol{u},\theta}(\boldsymbol{\Phi}_h,\psi_h)-\widetilde{F}_{\boldsymbol{u}_h,\theta_h}(\boldsymbol{\Phi}_h,\psi_h)|}{\|(\boldsymbol{\Phi}_h,\psi_h)\|}\right.$$
$$\left.+\inf_{\substack{(\Psi_h,\phi_h)\in\mathbf{H}_h^{\boldsymbol{\Theta}}\times\mathrm{H}_h^{\theta}\\(\Psi_h,\phi_h)\neq 0}}\left(\|(\boldsymbol{\Theta},\theta)-(\boldsymbol{\Psi}_h,\phi_h)\|+\sup_{\substack{(\boldsymbol{\Phi}_h,\psi_h)\in\mathbf{H}_h^{\boldsymbol{\Theta}}\times\mathrm{H}_h^{\theta}\\(\boldsymbol{\Phi}_h,\psi_h)\neq 0}}\frac{|\widetilde{\mathbf{b}}_{\boldsymbol{u}-\boldsymbol{u}_h}((\boldsymbol{\Psi}_h,\phi_h),(\boldsymbol{\Phi}_h,\psi_h))|}{\|(\boldsymbol{\Phi}_h,\psi_h)\|}\right)\right\}, \tag{4.19}$$

where $\overline{K}_1 := \frac{2}{\widetilde{\alpha}_1(\Omega)}\max\{1,\|\widetilde{\boldsymbol{a}}+\widetilde{\mathbf{b}}_{\boldsymbol{u}}\|\}$. Thus, employing [15, Lemma 5.3], we have

$$\sup_{\substack{(\boldsymbol{\Phi}_h,\psi_h)\in\mathbf{H}_h^{\boldsymbol{\Theta}}\times\mathrm{H}_h^{\theta}\\(\boldsymbol{\Phi}_h,\psi_h)\neq 0}}\frac{|\widetilde{\mathbf{b}}_{\boldsymbol{u}-\boldsymbol{u}_h}((\boldsymbol{\Psi}_h,\phi_h),(\boldsymbol{\Phi}_h,\psi_h))|}{\|(\boldsymbol{\Phi}_h,\psi_h)\|} \le \|i_c\|^2\,(1+\kappa_4^2)^{1/2}\widetilde{K}_1\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}\,\|\theta\|_{1,\Omega}$$
$$+\ \|i_c\|^2\,(1+\kappa_4^2)^{1/2}\widetilde{K}_1\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}\,\|(\boldsymbol{\Theta},\theta)-(\boldsymbol{\Psi}_h,\phi_h)\|\,, \tag{4.20}$$

and similarly as in Lemma 4.3 we get

$$\sup_{\substack{(\boldsymbol{\Phi}_h,\psi_h)\in\mathbf{H}_h^{\boldsymbol{\Theta}}\times\mathrm{H}_h^{\theta}\\(\boldsymbol{\Phi}_h,\psi_h)\neq 0}}\frac{|\widetilde{F}_{\boldsymbol{u},\theta}(\boldsymbol{\Phi}_h,\psi_h)-\widetilde{F}_{\boldsymbol{u}_h,\theta_h}(\boldsymbol{\Phi}_h,\psi_h)|}{\|(\boldsymbol{\Phi}_h,\psi_h)\|}$$
$$\le \widetilde{K}_1(1+\kappa_4^2)^{1/2}L_s\,\|\boldsymbol{u}_h\|_{1,\Omega}\,\|\theta_h-\theta\|_{1,\Omega}+\widetilde{K}_1 s_2\,\|\boldsymbol{u}_h-\boldsymbol{u}\|_{0,\Omega}\,. \tag{4.21}$$

Therefore, (4.18) follows by replacing (4.20) and (4.21) back into (4.19), with a constant $K_{\mathbf{ST}}$ depending on $\widetilde{\alpha}_1$, $\|\widetilde{\boldsymbol{a}}+\widetilde{\mathbf{b}}_{\boldsymbol{u}}\|$, $\widetilde{K}_1$, $\|i_c\|$, κ_4, and L_s. □

In much the same way as in Sect. 3.3, denoting

$$C_9 = K_{\mathbf{ST}}k_{\mathbf{S}}c_{\mathbf{S}}, \quad C_{10} := C_{\mathbf{ST}}c_{\mathbf{S}}+K_{\mathbf{ST}}c_{\mathbf{S}}+C_{\mathbf{ST}}c_{\mathbf{S}},$$
$$C_{11} := \max\{C_{\mathbf{ST}}, C_5, (C_5+C_{10}+C_9 r)r, C_9+C_{10}, K_{\mathbf{ST}}k_{\mathbf{S}}, K_{\mathbf{ST}}\},$$

where C_5 is the constant defined in (3.51), and applying the estimates given in Lemmas 3.15 and 4.7, we can prove that

$$\|\vec{\mathbf{t}}-\vec{\mathbf{t}}_h\|+\|(\boldsymbol{\Theta},\theta)-(\boldsymbol{\Theta}_h,\theta_h)\| \le C_{\mathbf{ST}}\,\mathrm{dist}(\vec{\mathbf{t}},H_h)$$
$$+\ K_{\mathbf{ST}}\left(1+\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}\right)\,\mathrm{dist}\left((\boldsymbol{\Theta},\theta),\mathrm{H}_h^{\boldsymbol{\Theta}}\times\mathrm{H}_h^{\theta}\right)$$

$$+ C_{11}\Big(L_f + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Omega} + C_f + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Omega} + \|\theta_{\mathrm{D}}\|_{1/2,\Gamma} + \|\theta_{\mathrm{D}}\|_{0,\Gamma} + s_2\Big) \times \Big\{\|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| + \|(\boldsymbol{\Theta},\theta) - (\boldsymbol{\Theta}_h,\theta_h)\|\Big\}.$$

We stress here that the constants multiplying $\mathrm{dist}(\vec{\mathbf{t}}, H_h)$ and $\mathrm{dist}\big((\boldsymbol{\Theta},\theta), \mathrm{H}_h^{\boldsymbol{\Theta}} \times \mathrm{H}_h^{\theta}\big)$ are both controlled by constants, parameters, and data only since $\|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,\Omega}$ can be controlled by (3.15) and (3.41). Consequently, we can establish the following main result.

Theorem 4.8 *Assume that the data satisfy*

$$C_7\Big\{L_f + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Omega} + C_f + \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Omega} + \|\theta_{\mathrm{D}}\|_{1/2,\Gamma} + \|\theta_{\mathrm{D}}\|_{0,\Gamma} + s_2\Big\} < \frac{1}{2}.$$

Then, there exists a positive constant C_{12}, independent of h, such that

$$\|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| + \|(\boldsymbol{\Theta},\theta) - (\boldsymbol{\Theta}_h,\theta_h)\| \le C_{12}\Big\{\mathrm{dist}(\vec{\mathbf{t}}, H_h) + \mathrm{dist}\Big((\boldsymbol{\Theta},\theta), \mathrm{H}_h^{\boldsymbol{\Theta}} \times \mathrm{H}_h^{\theta}\Big)\Big\}. \quad (4.22)$$

4.5 Specific Finite Element Subspaces

Here we consider the global Raviart–Thomas space of order k to approximate $\boldsymbol{\Theta}$, and the Lagrange space of degree $\le k+1$ for the temperature θ, that is

$$\begin{aligned}\mathbf{H}_h^{\boldsymbol{\Theta}} &:= \Big\{\boldsymbol{\Phi}_h \in \mathbf{H}(\mathrm{div};\Omega) : \quad \boldsymbol{c}^{\mathrm{t}}\boldsymbol{\Phi}_h|_K \in \mathbf{RT}_k(K), \quad \forall\, \boldsymbol{c} \in \mathrm{R}^n \quad \forall\, K \in \mathcal{T}_h\Big\},\\ \mathrm{H}_h^{\theta} &:= \Big\{\psi_h \in \mathrm{C}(\overline{\Omega}) : \quad \psi_h|_K \in \mathrm{P}_{k+1}(K) \quad \forall\, K \in \mathcal{T}_h\Big\}.\end{aligned} \quad (4.23)$$

The approximation properties of the spaces in (3.56) and (4.23) (that can be found in e.g [4,10,24]) are then combined with the Céa estimate (4.22) to produce the theoretical rate of convergence of (4.16), summarized as follows.

Theorem 4.9 *Apart from the hypotheses of Theorems 4.5, 4.6 and 4.8, assume that there exists $s > 0$ such that $\mathbf{t} \in \mathbb{H}^s(\Omega)$, $\boldsymbol{\sigma} \in \mathbb{H}^s(\Omega)$, $\mathbf{div}\,\boldsymbol{\sigma} \in \mathbf{H}^s(\Omega)$, $\boldsymbol{u} \in \mathbf{H}^{1+s}(\Omega)$, $\boldsymbol{\Theta} \in \mathbf{H}^s(\Omega)$, $\mathrm{div}\,\boldsymbol{\Theta} \in \mathrm{H}^s(\Omega)$, and $\theta \in \mathrm{H}^{1+s}(\Omega)$. Then there exists $C > 0$ independent of h, such that with (3.56) and (4.23), one has*

$$\|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| + \|(\boldsymbol{\Theta},\theta) - (\boldsymbol{\Theta}_h,\theta_h)\| \le C h^{\min\{s,k+1\}}\Big\{\|\mathbf{t}\|_{s,\Omega} + \|\boldsymbol{\sigma}\|_{s,\Omega} + \|\mathbf{div}\,\boldsymbol{\sigma}\|_{s,\Omega} + \|\boldsymbol{u}\|_{1+s,\Omega} + \|\boldsymbol{\Theta}\|_{s,\Omega} + \|\mathrm{div}\,\boldsymbol{\Theta}\|_{s,\Omega} + \|\theta\|_{1+s,\Omega}\Big\}. \quad (4.24)$$

5 Numerical Tests

We now present a set of computational tests. For the mixed-primal scheme (3.35) we consider an example that shows the convergence rates anticipated by Theorem 3.18, and a second test that addresses the application of our method to the three-dimensional modeling of gallium melting in a cuboid cavity. We will also present two examples that illustrate the performance of the fully-mixed scheme (4.16), and that will serve as confirmation for the rates of convergence provided by Theorem 4.9.

5.1 Preliminary Notations

A Picard algorithm with tolerance of $1E-6$ on the ℓ^2-norm of the residual has been employed for our fixed-point problems. The convergence of the approximate solutions is assessed by computing errors in the respective norms and experimental rates, that we define as usual

$$
\begin{aligned}
&\mathrm{e}(\mathbf{t}) = \|\mathbf{t}-\mathbf{t}_h\|_{0,\Omega}\,, \quad \mathrm{e}(\boldsymbol{u}) = \|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}\,, \quad \mathrm{e}(p) = \|p-p_h\|_{0,\Omega}\,, \quad \mathrm{e}(\theta) = \|\theta-\theta_h\|_{1,\Omega}\,,\\
&\mathrm{e}(\lambda) = \left\|\lambda-\lambda_{\widetilde{h}}\right\|_{0,\Gamma}\,, \quad \mathrm{e}(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathbf{div};\Omega}\,,\\
&\quad \mathrm{e}(\boldsymbol{\Theta}) = \|\boldsymbol{\Theta}-\boldsymbol{\Theta}_h\|_{\mathbf{div};\Omega} \quad \widehat{\mathrm{e}}(\boldsymbol{\sigma}) = \left\|\boldsymbol{\sigma}_h-\boldsymbol{\sigma}_h^{\mathrm{t}}\right\|_{0,\Omega}\,,\\
&r(\lambda) = \frac{\log(\mathrm{e}(\lambda)/\mathrm{e}'(\lambda))}{\log(\widetilde{h}/\widetilde{h}')}, \quad r(\%) = \frac{\log(\mathrm{e}(\%)/\mathrm{e}'(\%))}{\log(h/h')},
\end{aligned}
$$

with $\% \in \{\mathbf{t}, \boldsymbol{\sigma}, \boldsymbol{u}, p, \boldsymbol{\Theta}, \theta\}$, and where e, e$'$ denote errors computed on two consecutive meshes of sizes h, h' ($\widetilde{h}$ and $\widetilde{h}'$ for λ), respectively. The trace condition on the stress is enforced through a penalization strategy. Furthermore, for the Examples 5.2.1, 5.3.1 and 5.3.2 described below, we remark that the Navier–Stokes–Brinkman and heat equations are considered non-homogeneous and the extra source terms are chosen according to the given exact solutions. This treatment does not compromise the continuous and discrete analysis, as the regularity of the exact solution provides sufficiently smooth right-hand sides, thus only requiring a slight modification of the functionals in the variational formulation.

5.2 Tests for the Mixed-Primal Scheme

Example 5.2.1 In our first numerical test, we consider problem (2.1) defined in the unit square $\Omega = (0,1)^2$ and choose the following manufactured exact solutions, viscosity, porosity, enthalpy, buoyancy and thermal conductivity:

$$
\begin{aligned}
&\boldsymbol{u} = \begin{pmatrix} \sin(\pi x)\cos(\pi y)\\ -\sin(\pi y)\cos(\pi x)\end{pmatrix}, \quad \theta = 1+\sin(\pi x)\cos(\pi y), \quad p = x^2-y^2, \quad \mathbf{t} = \mathbf{e}(\boldsymbol{u}),\\
&\quad \boldsymbol{\sigma} = \alpha\,\mu(\theta)\mathbf{t} - (\boldsymbol{u}\otimes\boldsymbol{u}) - p\,\mathbb{I}, \quad \lambda = -\rho\kappa\nabla\theta\cdot\boldsymbol{\nu}, \quad \mu(\theta) = \exp(-0.25\,\theta),\\
&\quad \eta(\theta) = 2-\tanh(0.5-\theta), \quad s(\theta) = 1+\tanh(1-\theta), \quad f(\theta) = 0.01\frac{\mathrm{Ra}}{\mathrm{PrRe}^2}\theta, \quad \kappa = \mathbb{I}.
\end{aligned} \tag{5.1}
$$

These closed-form solutions feature a divergence-free velocity that satisfies the compatibility condition (2.6) and it is used as a non-homogeneous Dirichlet datum on Γ. In turn, the exact temperature is uniformly bounded and it is also exploited as Dirichlet datum. Moreover, the nonlinear functions satisfy (2.2)–(2.4). We consider $\boldsymbol{k} = (0,1)^{\mathrm{t}}$ and the parameters given by: $\mathrm{Re} = 1, \mathrm{Pr} = 0.71, C = 1$ and $\mathrm{Ra} = 100$, where Ra is the Rayleigh number. The stabilization parameters κ_1, κ_2 and κ_3 are taken as in (3.23), where the viscosity and porosity bounds are estimated as $\mu_1 = 0.6$, $\mu_2 = 1$ and $\eta_1 = 1$, $\eta_2 = 3$, respectively, thus resulting in $\kappa_1 = 0.6$, $\kappa_2 = 0.33$ and $\kappa_3 = 0.3$. An average of six Picard steps were required to reach the desired tolerance. Errors and corresponding rates associated with first and second order approximations are summarized in Table 1. The results show optimal asymptotic convergence rates for all fields, which are the expected ones according to Theorem 3.18. Also, Fig. 1 shows that the rates of convergence for $\widehat{e}(\boldsymbol{\sigma})$ are the expected ones. Finally, samples of augmented mixed-primal approximations obtained with 1M DoFs are depicted in Fig. 2.

Table 1 Example 5.2.1 convergence history for $k = 0, 1$

Mixed-primal $\mathbb{P}_0 - \mathbf{RT}_0 - \mathbf{P}_1 - \mathrm{P}_1 - \mathrm{P}_0$ scheme							
DoFs	h	e($\mathbf{t}$)	e($\boldsymbol{\sigma}$)	e($\boldsymbol{u}$)	e(p)	e(θ)	e(λ)
1114	0.1900	0.27795	0.81133	0.46689	0.08997	0.32250	1.04403
4138	0.0950	0.14164	0.39563	0.23877	0.04233	0.16783	0.50832
16088	0.0490	0.07030	0.19703	0.11721	0.02047	0.08227	0.25242
63531	0.0244	0.03513	0.09902	0.05920	0.01045	0.04157	0.12559
255319	0.0139	0.01751	0.04928	0.02931	0.00514	0.02057	0.06272
1010150	0.0077	0.00878	0.02435	0.01450	0.00249	0.01020	0.03135
$\widehat{\mathrm{e}}(\boldsymbol{\sigma})$	$\widetilde{h}$	$r(\mathbf{t})$	$r(\boldsymbol{\sigma})$	$r(\boldsymbol{u})$	$r(p)$	$r(\theta)$	$r(\lambda)$
0.28394	0.5000	–	–	–	–	–	–
0.14441	0.2500	0.97260	1.03613	0.96744	1.08761	0.94224	1.03826
0.07225	0.1250	1.05788	1.05290	1.07467	1.09690	1.07679	1.00987
0.03542	0.0625	0.99623	0.98802	0.98078	0.96534	0.98014	1.00705
0.01767	0.0312	1.24455	1.24779	1.25722	1.26803	1.25808	1.00167
0.00901	0.0156	1.17723	1.20254	1.19973	1.23193	1.19550	1.00041
Mixed-primal $\mathbb{P}_1 - \mathbf{RT}_1 - \mathbf{P}_2 - \mathrm{P}_2 - \mathrm{P}_1$ scheme							
DoFs	h	e($\mathbf{t}$)	e($\boldsymbol{\sigma}$)	e($\boldsymbol{u}$)	e(p)	e(θ)	e(λ)
3610	0.1900	0.02055	0.06020	0.03517	0.01120	0.02617	0.08327
13690	0.1025	0.00494	0.01494	0.00824	0.00324	0.00607	0.01984
53826	0.0492	0.00120	0.00365	0.00200	0.00078	0.00145	0.00476
213782	0.0256	0.00030	0.00092	0.00051	0.00020	0.00036	0.00116
861670	0.0139	0.00008	0.00023	0.00013	0.00006	0.00008	0.00028
$\widehat{\mathrm{e}}(\boldsymbol{\sigma})$	$\widetilde{h}$	$r(\mathbf{t})$	$r(\boldsymbol{\sigma})$	$r(\boldsymbol{u})$	$r(p)$	$r(\theta)$	$r(\lambda)$
0.01935	0.5000	–	–	–	–	–	–
0.00427	0.0250	2.30696	2.25787	2.35075	2.01078	2.36693	2.06950
0.00108	0.1250	1.91309	1.90661	1.91476	1.91739	1.93990	2.05693
0.00027	0.0625	2.10199	2.12297	2.10057	2.09550	2.14255	2.03644
0.00006	0.0312	2.15117	2.26398	2.23191	1.95851	2.29949	2.01195

Example 5.2.2 We continue with a simulation involving phase change in a cuboid cavity. The problem corresponds to the steady thermal convective flow occurring in the melting of gallium. Numerical results for the transient version of this problem, as well as detailed experimental considerations, can be found in e.g. [8,40,42]. We have adapted the model to comply with (2.1), using a porosity-enthalpy framework (i.e., setting a constant viscosity), but employing mixed boundary conditions as prescribed below. The physical properties of the problem are defined by the model constants $\mathrm{Ra} = 2E5$, $\mathrm{Re} = 10$, $\mathrm{Pr} = 0.71$, $\mu = 1$, $\eta_1 = 1E-3$, $\eta_2 = 1E5$, $\theta_r = 0.01$, $r = 0.05$, $\boldsymbol{g} = (0, 0, 1)^{\mathrm{t}}$. The temperature-dependent enthalpy and porosity functions adopt the forms

$$s(\theta) = \frac{1}{2}\left\{1 + \tanh(\frac{\theta_r - \theta}{r})\right\}, \quad \eta(\theta) = \eta_1 + \eta_2\left\{1 + \tanh(\frac{\theta_r - \theta}{r})\right\}.$$

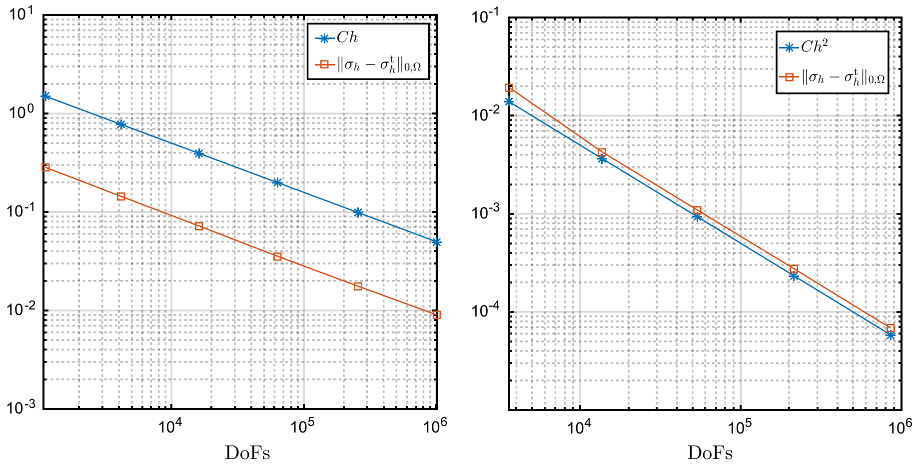

Fig. 1 Example 5.2.1 errors associated with the mixed-primal approximation *versus* DoFs for $\mathbb{P}_0 - \mathbf{RT}_0 - \mathbf{P}_1 - \mathrm{P}_1 - \mathrm{P}_0$ and $\mathbb{P}_1 - \mathbf{RT}_1 - \mathbf{P}_2 - \mathrm{P}_2 - \mathrm{P}_1$ finite elements (left and right, respectively)

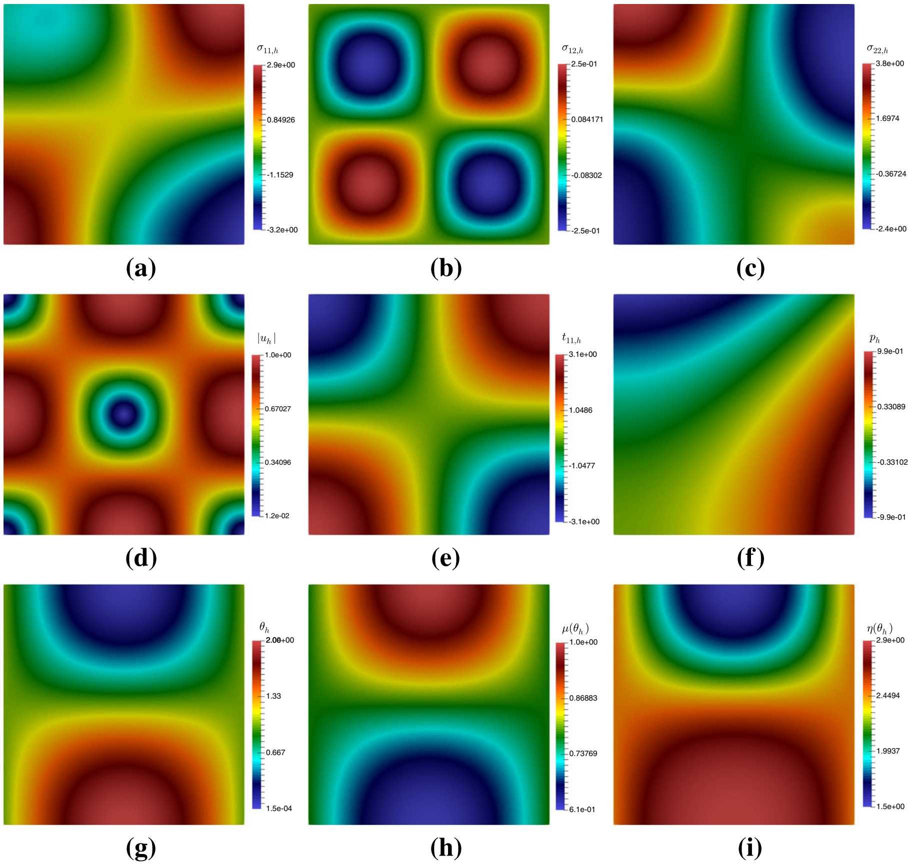

Fig. 2 Example 5.2.1 lowest-order approximate solutions: **a**–**c** pseudostress entries, **d** displacement magnitude, **e** strain rate, **f** postprocessed pressure, **g** temperature, **h** effective viscosity, and **i** effective porosity fields

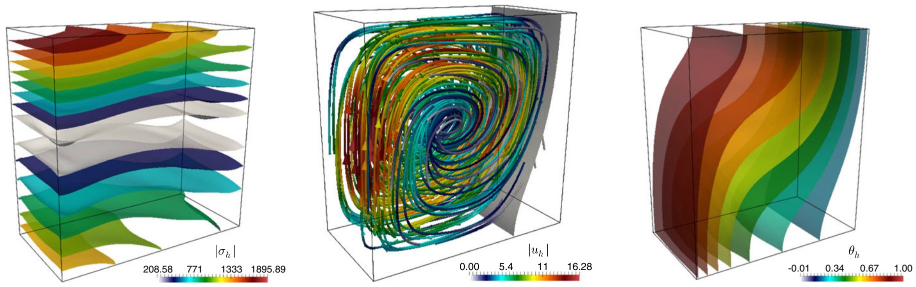

Fig. 3 Example 5.2.2 Computed solutions with the lowest-order mixed-primal scheme. **a** pseudo-stress magnitude, **b** velocity magnitude, **c** temperature

Table 2 Example 5.3.1 convergence history and Picard iteration count for $k = 0, 1$

$\mathbb{P}_0 - \mathbf{RT}_0 - \mathbf{P}_1 - \mathbf{RT}_0 - \mathrm{P}_1$ scheme							
DoFs	h	e($\mathbf{t}$)	e($\boldsymbol{\sigma}$)	e($\boldsymbol{u}$)	e(p)	e($\boldsymbol{\Theta}$)	e(θ)
1332	0.1900	0.27796	0.81134	0.46690	0.08977	1.75126	0.32186
5012	0.0950	0.14164	0.39564	0.23877	0.04228	0.86291	0.16783
19636	0.0490	0.07030	0.19703	0.11721	0.02047	0.43528	0.08226
77860	0.0244	0.03513	0.09902	0.05920	0.01045	0.21694	0.04158
313572	0.0139	0.01751	0.04928	0.02931	0.0051	0.10825	0.02057
1241924	0.0077	0.00878	0.02435	0.01450	0.00249	0.05320	0.01020

iter	$r(\mathbf{t})$	$r(\boldsymbol{\sigma})$	$r(\boldsymbol{u})$	$r(p)$	$r(\boldsymbol{\Theta})$	$r(\theta)$
6	–	–	–	–	–	–
6	0.97261	1.03610	0.96744	1.08623	1.02110	0.93942
6	1.05793	1.05293	1.07468	1.09557	1.03359	1.07691
6	0.99624	0.98803	0.98078	0.96496	1.00003	0.97988
6	1.24456	1.24779	1.25722	1.26790	1.24230	1.25816
6	1.17723	1.20254	1.19973	1.23190	1.21162	1.19554

Fully-mixed $\mathbb{P}_1 - \mathbf{RT}_1 - \mathbf{P}_2 - \mathbf{RT}_1 - \mathrm{P}_2$ scheme							
DoFs	h	e($\mathbf{t}$)	e($\boldsymbol{\sigma}$)	e($\boldsymbol{u}$)	e(p)	e($\boldsymbol{\Theta}$)	e(θ)
4354	0.1900	0.02055	0.06020	0.03517	0.01120	0.10995	0.02402
16642	0.1025	0.00494	0.01494	0.00824	0.00324	0.02824	0.00571
65734	0.0492	0.00120	0.00365	0.00200	0.00078	0.00694	0.00139
261712	0.0256	0.00030	0.00092	0.00051	0.00020	0.00174	0.00035
1056184	0.0139	0.00008	0.00023	0.00013	0.00006	0.00042	0.00008

iter	$r(\mathbf{t})$	$r(\boldsymbol{\sigma})$	$r(\boldsymbol{u})$	$r(p)$	$r(\boldsymbol{\Theta})$	$r(\theta)$
6	–	–	–	–	–	–
6	2.30685	2.25790	2.35075	2.01040	2.20238	2.32777
6	1.91308	1.90661	1.91475	1.91735	1.90140	1.90685
6	2.10194	2.12295	2.10055	2.09544	2.12817	2.11582
6	2.15690	2.26665	2.23441	1.97252	2.31585	2.28408

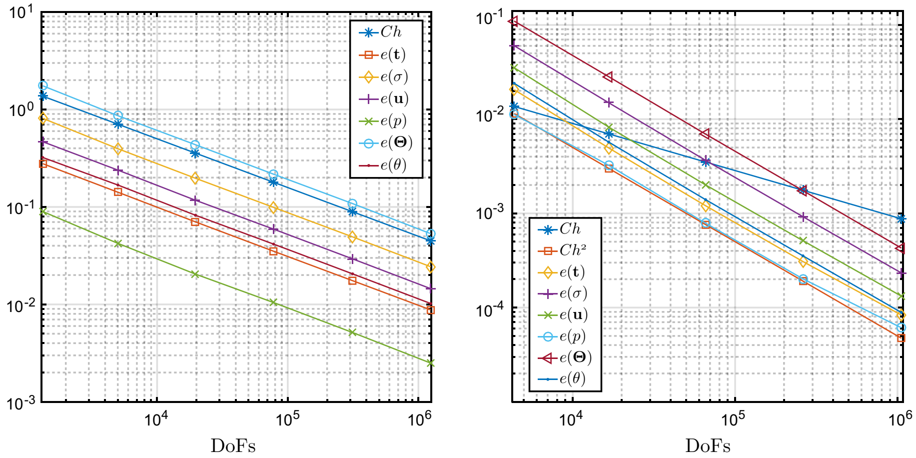

Fig. 4 Example 5.3.1 errors associated with the fully-mixed approximation *versus* DoFs for $\mathbb{P}_0 - \mathbf{RT}_0 - \mathbf{P}_1 - \mathbf{RT}_0 - \mathrm{P}_1$ and $\mathbb{P}_1 - \mathbf{RT}_1 - \mathbf{P}_2 - \mathbf{RT}_1 - \mathrm{P}_2$ finite elements (left and right, respectively)

The computational domain is the box $\Omega = (0, 2) \times (0, 2) \times (0, 1)$ and we generate a structured mesh composed by 255K tetrahedral elements and about 46K vertices. Considering the lowest-order mixed-primal finite element method (3.35), the assembled linear systems appearing at each Picard iteration consist of about 3M DoFs for the Navier–Stokes–Brinkman block and near 46K DoFs for the energy conservation equation. No-slip conditions were imposed for the velocity on the whole boundary. Moreover, the walls defined by $x = 0$ and $x = 2$ are maintained at fixed temperatures of $\theta = 1$ and $\theta = -0.01$, respectively; whereas on the remaining walls we impose zero-flux boundary conditions for the temperature. Such a setting implies in particular, that the Lagrange multiplier λ is not required in the formulation. Primary features of the flow can be observed in Fig. 3. We do not expect to produce the flow separation vortices as those seen in [40] because our test focuses on the steady regime and we employ a enthalpy-porosity model. Nevertheless, we do see streamlines avoiding the solid region (on the right hand side of the gray wall), as well as a qualitative match with the temperature profiles observed in [8,40], where thermal convection occurs mainly on the xy plane. Under the considered flow regime, 15 fixed-point iterations were needed to reach the desired residual tolerance of $1E-6$.

5.3 Tests for the Fully-Mixed Scheme

Example 5.3.1 In this example we consider the domain, exact solution, nonlinear functions, parameters and stabilization parameters for the Navier–Stokes–Brinkman equation exactly as in Example 5.2.1 of Sect. 5.2 (cf. (5.1)). We recall that $\boldsymbol{\Theta} := \rho\kappa\nabla\theta - \theta\boldsymbol{u} - s(\theta)$ and for the values κ_4, κ_5 and κ_6, we follow [15, Section 6] to obtain $\kappa_4 = 0.99$, $\kappa_5 = 0.5$ and κ_6=0.49. Values and plots of errors and corresponding rates associated with first and second order approximations are summarized in Table 2 and Fig. 4. The results show optimal asymptotic convergence rates for all fields, which are the expected ones according to Theorem 4.9. We remark here that the errors reported in Tables 1 and 2 for the unknowns $\mathbf{t}$, $\boldsymbol{\sigma}$, and $\boldsymbol{u}$, are basically the same for the two methods considered in the paper, which is due to the fact that both formulations consider a mixed approach for the Navier–Stokes–Brinkman equation.

However, since for the heat equation primal and mixed approaches are employed, which yields the two different coupled schemes that are proposed and analyzed in the paper, some very slight changes (even only after two or three decimals) can be observed in those tables for the rates of convergences of $\mathbf{t}$, $\boldsymbol{\sigma}$, $\boldsymbol{u}$, and p.

Example 5.3.2 In our second example, we produce the error and rate history associated with the finite element approximation for the three-dimensional case. Let us consider the following closed-form solutions to the model problem, defined on the unit cube domain $\Omega = (0, 1)^3$:

$$\boldsymbol{u} = \begin{pmatrix} \cos(x)\sin(y)\sin(z) \\ \sin(x)\cos(y)\sin(z) \\ -2\sin(x)\sin(y)\cos(z) \end{pmatrix}, \quad \theta = 1 + \sin(\pi x)\cos(\pi y)\sin(\pi z), \quad p = x^2 - 2y^2 - z^2.$$

These functions are smooth and they are used to generate non-homogeneous forcing and source terms. Also, the manufactured velocity and temperature are used as Dirichlet datum on Γ. The porosity, enthalpy and thermal conductivity are taken as in Example 5.2.1, and the remaining nonlinear functions are defined as: $\mu(\theta) = \exp(-\theta)$, $\quad f(\theta) = \theta$. All model constants assume the adimensional value 1. The stabilization parameters are taken again as in Example 5.3.1. Part of the solution is shown in Fig. 5, and a convergence history for a set of quasi-uniform refinements is shown in Table 3, confirming that this fully-mixed finite element method converges optimally with order $\mathcal{O}(h^{k+1})$.

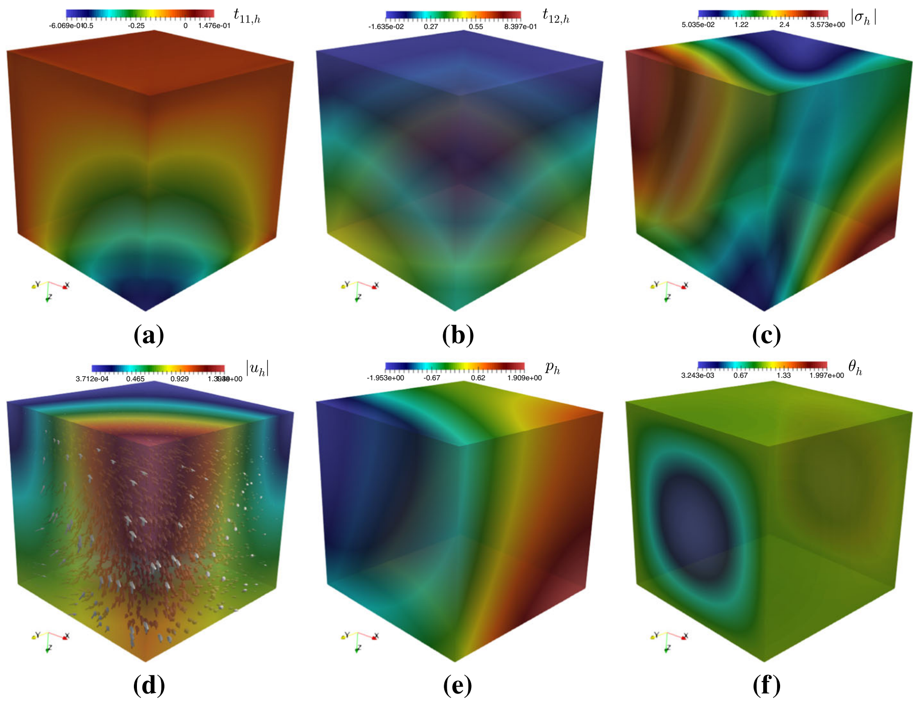

Fig. 5 Example 5.3.2 lowest-order approximate solutions: **a**–**b** relevant components of the strain rate, **c** pseudostress magnitude, **d** displacement magnitude, **e** postprocessed pressure, and **f** temperature

Table 3 Example 5.3.2 convergence history and Picard iteration count for $k = 0, 1$

Fully-mixed $\mathbb{P}_0 - \mathbf{RT}_0 - \mathbf{P}_1 - \mathbf{RT}_0 - \mathrm{P}_1$ scheme							
DoFs	h	e($\mathbf{t}$)	e($\boldsymbol{\sigma}$)	e($\boldsymbol{u}$)	e(p)	e($\boldsymbol{\Theta}$)	e(θ)
828	0.7071	0.38526	0.79824	0.54203	0.23637	5.37964	1.85821
5876	0.3535	0.24931	0.39517	0.27068	0.13072	2.88208	0.95736
44388	0.1767	0.13775	0.19102	0.12628	0.06088	1.46682	0.48293
345284	0.0883	0.07088	0.09400	0.06013	0.02935	0.73668	0.24289
2724228	0.0441	0.03572	0.04676	0.02947	0.01449	0.36875	0.12175
iter	$r(\mathbf{t})$	$r(\boldsymbol{\sigma})$	$r(\boldsymbol{u})$	$r(p)$	$r(\boldsymbol{\Theta})$	$r(\theta)$	
7	–	–	–	–	–	–	
6	0.62787	1.01432	1.00177	1.02698	0.90040	0.95678	
6	0.85581	1.04871	1.09991	1.10229	0.97441	0.98723	
6	0.95859	1.02296	1.07049	1.05256	0.99357	0.99151	
6	0.98869	1.00727	1.02845	1.01791	0.99839	0.99630	
Fully-mixed$\mathbb{P}_1 - \mathbf{RT}_1 - \mathbf{P}_2 - \mathbf{RT}_1 - \mathrm{P}_2$ scheme							
DoFs	h	e($\mathbf{t}$)	e($\boldsymbol{\sigma}$)	e($\boldsymbol{u}$)	e(p)	e($\boldsymbol{\Theta}$)	e(θ)
3476	0.7071	0.05677	0.11816	0.07065	0.04174	1.90822	0.59405
25572	0.3535	0.01657	0.03067	0.01769	0.01044	0.51742	0.15859
196292	0.1767	0.00441	0.00784	0.00437	0.00260	0.13213	0.04294
1538436	0.0883	0.00113	0.00199	0.00108	0.00065	0.03354	0.01128
iter	$r(\mathbf{t})$	$r(\boldsymbol{\sigma})$	$r(\boldsymbol{u})$	$r(p)$	$r(\boldsymbol{\Theta})$	$r(\theta)$	
6	–	–	–	–	–	–	
6	1.77670	1.94583	1.99761	1.99851	1.88281	1.90524	
6	1.90946	1.96778	2.01746	2.00408	1.96929	1.88463	
6	1.96175	1.97533	2.00942	1.99138	1.97773	1.92771	

Acknowledgements The authors are sincerely thankful to Miroslav Kutcha and Kent-André Mardal (University of Oslo), for extending the `(FEniCS)_ii` library [26] (a set of dedicated routines for coupling forms associated with meshes of embedded manifolds) to the case of non-conforming traces, and for kindly assisting with some technical issues in the implementation of the numerical tests in this paper.

References

1. Adams, R.A., Fournier, J.J.F.: Sobolev Spaces. Academic Press, Elsevier Ltd, Cambridge (2003)
2. Agroum, R., Bernardi, C., Satouri, J.: Spectral discretization of the time-dependent Navier-Stokes problem coupled with the heat equation. Appl. Math. Comp. **268**, 59–82 (2015)
3. Aldbaissy, R., Hecht, F., Mansour, G., Sayah, T.: A full discretisation of the time-dependent Boussinesq (buoyancy) model with nonlinear viscosity. Calcolo **55**(44), 1–29 (2018)
4. Almonacid, J., Gatica, G.N., Oyarzúa, R., and Ruiz-Baier, R.: A new mixed finite element method for the n-dimensional Boussinesq problem with temperature-dependent viscosity. Preprint 2018-18, Centro de Investigación en Ingeniería Matemática (CI^{2}MA), Universidad de Concepción, Chile, (2018). https://www.ci2ma.udec.cl
5. Almonacid, J., Gatica, G.N., Oyarzúa, R.: A mixed-primal finite element method for the Boussinesq problem with temperature-dependent viscosity. Calcolo **55**(3), 42 (2018)

6. Alvarez, M., Gatica, G.N., Ruiz-Baier, R.: An augmented mixed–primal finite element method for a coupled flow-transport problem. ESAIM: Math. Model. Numer. Anal. **49**(5), 1399–1427 (2015)
7. Alvarez, M., Gomez-Vargas, B., Ruiz-Baier, R., and Woodfield, J.: *Stability and finite element approximation of phase change models for natural convection in porous media*. J. Comput. Appl. Math., in press (2019).
8. Ben-David, O., Levy, A., Mikhailovich, B., Azulay, A.: 3D numerical and experimental study of gallium melting in a rectangular container. Int. J. Heat Mass Transf. **67**, 260–271 (2013)
9. Brandeis, G., Marsh, B.D.: The convective liquidus in solidifying magma chamber: a fluid dynamic investigation. Nature **339**, 613–616 (1989)
10. Brezzi, F., Fortin, M.: Mixed and Hybrid Finite Element Methods. Springer, Berlin (1991)
11. Camaño, J., Oyarzúa, R., Ruiz-Baier, R., Tierra, G.: Error analysis of an augmented mixed method for the Navier-Stokes problem with mixed boundary conditions. IMA J. Numer. Anal. **38**(3), 1452–1484 (2018)
12. Camaño, J., Gatica, G.N., Oyarzúa, R., Ruiz-Baier, R.: An augmented stress-based mixed fi- nite element method for the Navier–Stokes equations with nonlinear viscosity. Numer. Methods Partial Differ. Equ. **33**(5), 1692–1725 (2017)
13. Cao, Y., Chen, S.: Analysis and finite element approximation of bioconvection flows with concentration dependent viscosity. Int. J. Numer. Anal. Mod. **11**(1), 86–101 (2013)
14. Colmenares, E., Gatica, G.N., Oyarzúa, R.: Analysis of an augmented mixed-primal formulation for the stationary Boussinesq problem. Numer. Methods Partial Differ. Equ. **32**(2), 445–478 (2016)
15. Colmenares, E., Gatica, G.N., Oyarzúa, R.: An augmented fully-mixed finite element method for the stationary Boussinesq problem. Calcolo **54**(1), 167–205 (2017)
16. Colmenares, E., Neilan, M.: Dual-mixed finite element methods for the stationary Boussinesq problem. Comp. Math. Appl. **72**, 1828–1850 (2016)
17. Ciarlet, P.: Linear and Nonlinear Functional Analysis with Applications. Society for Industrial and Applied Mathematics, Philadelphia, PA (2013)
18. Danaila, I., Moglan, R., Hecht, F., Le Masson, S.: A Newton method with adaptive finite elements for solving phase-change problems with natural convection. J. Comput. Phys. **274**, 826–840 (2014)
19. Deteix, J., Jendoubi, A., Yakoubi, D.: A coupled prediction scheme for solving the Navier-Stokes and convection-diffusion equations. SIAM J. Numer. Anal. **52**(5), 2415–2439 (2014)
20. Dinniman, M.S., Asay-Davis, X.S., Galton-Fenzi, B.K., Holland, P.R., Jenkins, A., Timmermann, R.: Modeling ice shelf/ocean interaction in antarctica: a review. Oceanography **29**(4), 144–153 (2016)
21. Dutil, Y., Rousse, D.R., Salah, N.B., Lassue, S., Zalewski, L.: A review on phase-change materials: mathematical modeling and simulations. Renew. Sustain. Energy Rev. **15**(1), 112–130 (2011)
22. Farhloul, M., Nicaise, S., Paquet, L.: A mixed formulation of Boussinesq equations: analysis of nonsingular solutions. Math. Comp. **69**(231), 965–986 (2000)
23. Gatica, G.N.: 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**(1), 1–28 (2006)
24. Gatica, G.N.: A Simple Introduction to the Mixed Finite Element Method: Theory and Applications. Springer Briefs in Mathematics. Springer, Cham (2014)
25. Gatica, G.N., Gomez-Vargas, B., Ruiz-Baier, R.: Analysis and mixed-primal finite element discretisations for stress-assisted diffusion problems. Comput. Methods Appl. Mech. Eng. **337**, 411–438 (2018)
26. Holter, K.E., Kuchta, M., and Mardal, K.-A.: Trace constrained problems in FEniCS. In: Hale, J.S. (ed.) Proceedings of the FEniCS Conference 2017 (2017). https://doi.org/10.6084/m9.figshare.5086369
27. Hsu, H.W., Postberg, F., Sekine, Y., et al.: Ongoing hydrothermal activities within Enceladus. Nature **7542**(519), 207–210 (2015)
28. Kan-on, Y., Narukawa, K., Teramoto, Y.: On the equations of bioconvective flow. J. Math. Kyoto Univ. **32**(1), 135–153 (1992)
29. Ma, X., Tao, Z., and Zhang, T.: A variational multiscale method for steady natural convection problem based on two-grid discretization. Adv. Diff. Equ. 2–20 (2016)
30. Morgan, K.: A numerical analysis of freezing and melting with convection. Comput. Methods Appl. Mech. Eng. **28**(3), 275–284 (1981)
31. Oyarzúa, R., Qin, T., Schötzau, D.: An exactly divergence-free finite element method for a generalized Boussinesq problem. IMA J. Numer. Anal. **34**(3), 1104–1135 (2014)
32. Oyarzúa, R., Zúñiga, P.: Analysis of a conforming finite element method for the Boussinesq problem with temperature-dependent parameters. J. Comput. Appl. Math. **323**, 71–94 (2017)
33. Schroeder, P.W., Lube, G.: Stabilised dG-FEM for incompressible natural convection flows with boundary and moving interior layers on non-adapted meshes. J. Comput. Phys. **335**, 760–779 (2017)
34. Roberts, J.E., Thomas, J.M.: Mixed and hybrid methods. In: Ciarlet, P.G., Lions, J.L. (eds.) Hand-book of Numerical Analysis. Finite Element Methods (Part 1), vol. II. North-Holland, Amsterdam (1991)

35. Ulvrová, M., Labrosse, S., Coltice, N., Røaback, P., Tackley, P.J.: Numerical modelling of convection interacting with a melting and solidification front: application to the thermal evolution of the basal magma ocean. Phys. Earth Planet. Inter. **206–207**, 51–66 (2012)
36. Voller, V.R., Cross, M., Markatos, N.C.: An enthalpy method for convection/diffusion phase change. Int. J. Numer. Methods Eng. **24**(1), 271–284 (1987)
37. Voller, V.R., Prakash, C.: A fixed grid numerical modelling methodology for convection and phase transition efficiently. J. Comput. Phys. **30**(8), 1709–1719 (1987)
38. Vidalain, G., Gosselin, L., Lacroix, M.: An enhanced thermal conduction model for the prediction of convection dominated solidliquid phase change. Int. J. Heat Mass Transf. **52**, 1753–1760 (2009)
39. Wang, S., Faghri, A., Bergman, T.L.: A comprehensive numerical model for melting with natural convection. Int. J. Heat Mass Transfer. **53**(9–10), 1986–2000 (2010)
40. Wittig, K., Nikrityuk, P.A.: Three-dimensionality of fluid flow in the benchmark experiment for a pure metal melting on a vertical wall. IOP Conf. Ser.: Mater. Sci. Eng. **27**, 012054 (2012)
41. Zhang, Y., Hou, Y., Jia, H.: Subgrid stabilized defect-correction method for a steady-state natural convection problem. Comput. Math. Appl. **67**, 497–514 (2014)
42. Zimmerman, A.G., and Kowalski, J.:Monolithic simulation of convection-coupled phase-change- verification and reproducibility. Preprint (2018). arXiv:1801.03429v1

Affiliations

Mario Alvarez[1] · Gabriel N. Gatica[2] · Bryan Gomez-Vargas[1,2] · Ricardo Ruiz-Baier[3,4]

Mario Alvarez
mario.alvarezguadamuz@ucr.ac.cr

Bryan Gomez-Vargas
bgomez@ci2ma.udec.cl

Ricardo Ruiz-Baier
ruizbaier@maths.ox.ac.uk

1 Sección de Matemática, Sede de Occidente, Universidad de Costa Rica, San Ramón, Costa Rica

2 CI^{2}MA and Departamento de Ingeniería Matemática, Universidad de Concepción, Casilla 160-C, Concepción, Chile

3 Mathematical Institute, University of Oxford, Andrew Wiles Building, Oxford OX2 6GG, UK

4 Laboratory of Mathematical Modeling, Institute of Personalized Medicine, Sechenov University, Moscow, Russia 119991