

# An Augmented Stress-Based Mixed Finite Element Method for the Steady State Navier-Stokes Equations with Nonlinear Viscosity

Jessika Camaño,<sup>1,2</sup> Gabriel N. Gatica ,<sup>2,3</sup> Ricardo Oyarzúa,<sup>2,4</sup> Ricardo Ruiz-Baier<sup>5</sup>

<sup>1</sup>Departamento de Matemática y Física Aplicadas, Universidad Católica de la Santísima Concepción, Casilla 297, Concepción, Chile

<sup>2</sup>CI<sup>2</sup>MA, Universidad de Concepción, Casilla 160-C, Concepción, Chile

<sup>3</sup>Departamento de Ingeniería Matemática, Universidad de Concepción, Casilla 160-C, Concepción, Chile

<sup>4</sup>GIMNAP-Departamento de Matemática, Universidad del Bío-Bío, Casilla 5-C, Concepción, Chile

<sup>5</sup>Mathematical Institute, Oxford University, A. Wiles Building, Woodstock Road, OX2 6GG, Oxford, UK

Received 22 July 2016; accepted 26 April 2017

Published online 2 June 2017 in Wiley Online Library (wileyonlinelibrary.com).

DOI 10.1002/num.22166

A new stress-based mixed variational formulation for the stationary Navier-Stokes equations with constant density and variable viscosity depending on the magnitude of the strain tensor, is proposed and analyzed in this work. Our approach is a natural extension of a technique applied in a recent paper by some of the authors to the same boundary value problem but with a viscosity that depends nonlinearly on the gradient of velocity instead of the strain tensor. In this case, and besides remarking that the strain-dependence for the viscosity yields a more physically relevant model, we notice that to handle this nonlinearity we now need to incorporate not only the strain itself but also the vorticity as auxiliary unknowns. Furthermore, similarly as in that previous work, and aiming to deal with a suitable space for the velocity, the variational formulation is

*Correspondence to:* Gabriel N. Gatica, Departamento de Ingeniería Matemática and CI<sup>2</sup>MA, Universidad de Concepción, Casilla 160-C, Concepción, Chile (e-mail: ggatica@ci2ma.udec.cl)

Contract grant sponsor: CONICYT-Chile through BASAL project CMM

Contract grant sponsor: CONICYT-Chile through Anillo ACT1118 (ANANUM)

Contract grant sponsor: CONICYT-Chile through Inserción de Capital Humano Avanzado en la Academia; contract grant number: 79130048

Contract grant sponsor: Fondecyt; contract grant number: 1161325 and 11140691

Contract grant sponsor: Centro de Investigación en Ingeniería Matemática (CI<sup>2</sup>MA)

Contract grant sponsor: Universidad del Bío-Bío through DIUBB; contract grant number: 120808 GI/EF

Contract grant sponsor: Engineering and Physical Sciences Research Council EPSRC; contract grant number: EP/R00207X/1

augmented with Galerkin-type terms arising from the constitutive and equilibrium equations, the relations defining the two additional unknowns, and the Dirichlet boundary condition. In this way, and as the resulting augmented scheme can be rewritten as a fixed-point operator equation, the classical Schauder and Banach theorems together with monotone operators theory are applied to derive the well-posedness of the continuous and associated discrete schemes. In particular, we show that arbitrary finite element subspaces can be utilized for the latter, and then we derive optimal a priori error estimates along with the corresponding rates of convergence. Next, a reliable and efficient residual-based a posteriori error estimator on arbitrary polygonal and polyhedral regions is proposed. The main tools used include Raviart-Thomas and Clément interpolation operators, inverse and discrete inequalities, and the localization technique based on triangle-bubble and edge-bubble functions. Finally, several numerical essays illustrating the good performance of the method, confirming the reliability and efficiency of the a posteriori error estimator, and showing the desired behavior of the adaptive algorithm, are reported. © 2017 Wiley Periodicals, Inc. Numer Methods Partial Differential Eq 33: 1692–1725, 2017

*Keywords:* a priori error analysis; augmented mixed formulation; fixed point theory; mixed finite element methods; Navier-Stokes equations; nonlinear viscosity

## I. INTRODUCTION

The development of mixed finite element techniques for quasi-Newtonian fluids whose viscosity is a nonlinear function of the state variables, such as blood, polymers, and molten metals, among others, has gained considerable attention in the last few years. For instance, a mixed finite element method for the Navier-Stokes equations with a viscosity depending nonlinearly on the magnitude of the gradient of velocity, was introduced and analyzed recently in [1]. The approach there makes use of the same modified pseudostress tensor used in [2], which, similarly to the one from [3], involves the diffusive and convective terms, and the pressure. The latter unknown is then eliminated thanks to an equivalent statement implied by the incompressibility condition. In addition, to handle the nonlinear viscosity, and following [3] and [4], the gradient of velocity is incorporated as an auxiliary unknown. Furthermore, as the velocity actually lives in a smaller space than expected, the variational formulation is augmented with suitable Galerkin-type terms arising from the constitutive and equilibrium equations, the relation defining the aforementioned additional unknown, and the Dirichlet boundary condition. Moreover, the resulting augmented scheme can be rewritten as a fixed-point equation, and therefore the well-known Schauder and Banach theorems, combined with classical results on monotone operators, are applied to prove the well-posedness of the continuous and discrete systems. In particular, the unique solvability of the Galerkin schemes does not require any discrete inf-sup conditions, and hence arbitrary finite element subspaces of the respective continuous spaces can be used in [1]. For a complete bibliographic discussion on the wide variety of dual-mixed methods for Newtonian and Non-Newtonian incompressible flows, and particularly for the Navier-Stokes equations, including pseudostress-based, stress-based, least-squares, augmented, stabilized, and other related formulations, we refer to [1, Section I].

Conversely, it is well known that standard Galerkin procedures such as finite element and mixed finite element methods inevitably lose accuracy when they are applied to nonlinear problems on quasiuniform discretizations. This fact is usually due to the lack of previous knowledge on how to mesh the domains in these cases, and hence adaptive algorithms that are based on a posteriori error estimates are useful in overcoming such a difficulty. In this regard, a residual-based a posteriori error analysis for the model and method from [1] has been developed in the

recent work [5]. More precisely, the technique proposed in [6, 7] for a class of nonlinear problems in fluid mechanics is adapted in [5] to derive reliable and efficient residual-based a posteriori error estimators for the augmented mixed formulation introduced in [1] of the Navier-Stokes equations with viscosity depending on the gradient of velocity. In fact, the strategy in [5] begins with a global inf-sup condition for the linearization arising from the use of the Gâteaux derivatives of the nonlinear terms of the formulation. The rest of the analysis includes a suitable handling of the corresponding convective term of the Navier-Stokes equations, the introduction of continuous and discrete Helmholtz's decompositions, and the application of the local approximation properties of the Raviart-Thomas and Clément interpolation operators, inverse inequalities, and the localization technique based on triangle-bubble and edge-bubble functions. For an extensive list of references on a posteriori error analysis for linear and nonlinear problems, mainly in fluid mechanics, we refer to [5, Section I]. In particular, we remark that most of the main ideas and associated techniques can be found in the early works [8, 9] and the references therein.

In spite of the aforescribed contributions (cf. [1, 5]), we find it important to remark that a physically more meaningful model for the Navier-Stokes equations arises from a viscosity depending nonlinearly not on the full gradient of the velocity, but only on the symmetric part of it. According to it, the purpose of the present paper is to additionally contribute in the direction of mixed finite element methods for nonlinear problems in fluid mechanics, by extending the a priori and a posteriori error analyses developed in [1, 5] to the steady state Navier-Stokes equations with constant density and variable viscosity depending on the magnitude of the strain tensor. In this way, the physical relevance of the underlying model together with the fact that we now gather both the a priori and a posteriori error analyses in a single contribution, guarantee a greater visibility of our results. The rest of this work is organized as follows. Some preliminary notations, the nonlinear model of interest, and the definite unknowns to be considered in the variational formulation are discussed in Section II. In Section III, we first derive the augmented mixed variational formulation, which, differently from [1], and aiming to handle the new nonlinearity, includes now the strain and vorticity tensors as auxiliary unknowns. Next, we introduce and analyze the equivalent fixed point setting, and then we consider the particular case of homogeneous Dirichlet boundary conditions, for which one of the augmented equations is no longer needed. The section ends with the solvability analysis, mainly via the Schauder and Banach theorems and assuming sufficiently small data, of the corresponding fixed point operator equations. In turn, in Section IV, we study the associated Galerkin scheme by employing a discrete version of the fixed-point strategy developed in Section III. Similarly, as for [1], we remark that no discrete inf-sup conditions are required here for the discrete analysis, and hence arbitrary finite element subspaces can be used as well. In addition, the a priori error estimate and the corresponding rates of convergence for a particular choice of discrete subspaces are also deduced in Section IV under a similar assumption on the size of the data. Furthermore, in Section V, we derive a reliable and efficient residual-based a posteriori error estimator for our augmented mixed formulation on arbitrary polygonal and polyhedral regions of  $R^2$  and  $R^3$ , respectively. We provide most of the details for the 3D case, whereas the main aspects of the 2D case, being analogous, are simply summarized at the end of that Section. We remark that Raviart-Thomas and Clément interpolation operators, inverse and discrete inequalities, and the localization technique based on triangle-bubble and edge-bubble functions constitute the main tools used. Finally, in Section VI, we collect several numerical examples illustrating the good performance of the augmented mixed finite element method, confirming the theoretical rates of convergence, providing the expected bounded ranges for the effectivity indexes of the a posteriori error estimator in 2D and 3D, and showing the satisfactory behaviour of the corresponding adaptive refinement strategy.

II. THE MODEL PROBLEM

A. Preliminaries

Let us denote by  $\Omega \subseteq \mathbb{R}^n, n \in \{2, 3\}$ , a given bounded domain with polyhedral boundary  $\Gamma$ , and denote by  $\mathbf{v}$  the outward unit normal vector on  $\Gamma$ . Standard notation will be adopted for Lebesgue spaces  $L^p(\Omega)$  and Sobolev spaces  $H^s(\Omega)$  with norm  $\|\cdot\|_{s,\Omega}$  and seminorm  $|\cdot|_{s,\Omega}$ . In particular,  $H^{1/2}(\Gamma)$  is the space of traces of functions of  $H^1(\Omega)$  and  $H^{-1/2}(\Gamma)$  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 fields  $\mathbf{v} = (v_i)_{i=1,n}$  and  $\mathbf{w} = (w_i)_{i=1,n}$ , we set the gradient, divergence, and tensor product operators, as

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

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

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

where  $\mathbb{I}$  stands for the identity tensor in  $\mathbb{R} := \mathbb{R}^{n \times n}$ . Furthermore, we recall that

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

equipped with the usual norm

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

is a standard Hilbert space. Finally, in what follows,  $|\cdot|$  denotes the Euclidean norm in  $\mathbf{R} := \mathbb{R}^n$ .

We use  $C$ , with or without subscripts, bars, tildes or hats, to mean generic positive constants independent of the discretization parameters, which may take different values at different places.

B. The Steady-State Navier-Stokes Equations with Variable Viscosity

We consider the Navier-Stokes equations with constant density and variable viscosity, that is

$$\begin{aligned} -\mathbf{div}(\mu(|\mathbf{e}(\mathbf{u})|)\mathbf{e}(\mathbf{u})) + (\nabla \mathbf{u})\mathbf{u} + \nabla p &= \mathbf{f} && \text{in } \Omega, \\ \text{div } \mathbf{u} &= 0 && \text{in } \Omega, \\ \mathbf{u} &= \mathbf{g} && \text{on } \Gamma, \end{aligned} \tag{2.1}$$

where the unknowns are the velocity  $\mathbf{u}$  and the pressure  $p$  of a fluid occupying the region  $\Omega$ , and  $\mathbf{e}(\mathbf{u}) := \frac{1}{2} \{ \nabla \mathbf{u} + (\nabla \mathbf{u})^t \}$  stands for the strain rate tensor. In turn, the given data are the nonlinear fluid viscosity  $\mu : \mathbb{R}^+ \rightarrow \mathbb{R}$ , a volume force  $\mathbf{f} \in \mathbb{L}^2(\Omega)$ , and the boundary velocity  $\mathbf{g} \in \mathbf{H}^{1/2}(\Gamma)$ . Note that, according to the incompressibility of the fluid,  $\mathbf{g}$  must satisfy the compatibility condition

$$\int_{\Gamma} \mathbf{g} \cdot \mathbf{v} = 0, \tag{2.2}$$

and that uniqueness of a pressure solution of (2.1) is ensured in the space

$$L_0^2(\Omega) = \left\{ q \in L^2(\Omega) : \int_{\Omega} q = 0 \right\}.$$

We remark that the nonlinear function  $\mu$  depends now on the magnitude of  $\mathbf{e}(\mathbf{u})$  instead of that of  $\nabla \mathbf{u}$  as it was in [1]. Assumptions on the viscosity include  $\mu$  being of class  $C^1$ , and that there exist constants  $\mu_1, \mu_2 > 0$ , such that

$$\mu_1 \leq \mu(s) \leq \mu_2 \quad \text{and} \quad \mu_1 \leq \mu(s) + s\mu'(s) \leq \mu_2 \quad \forall s \geq 0 \tag{2.3}$$

which, according to [10, Theorem 3.8], imply Lipschitz continuity and strong monotonicity of the nonlinear operator induced by  $\mu$ . A classical example of viscosity functions is the well-known Carreau law

$$\mu(s) := \alpha_0 + \alpha_1(1 + s^2)^{(\beta-2)/2} \quad \forall s \geq 0 \tag{2.4}$$

where  $\alpha_0, \alpha_1 > 0$  and  $\beta \in [1, 2]$ . This law satisfies the assumptions (2.3) with  $(\mu_1, \mu_2) = (\alpha_0, \alpha_0 + \alpha_1)$ .

Next, proceeding similarly as in [1] (see also [2, 11]), that is defining now the tensor

$$\boldsymbol{\sigma} := \mu(|\mathbf{e}(\mathbf{u})|)\mathbf{e}(\mathbf{u}) - (\mathbf{u} \otimes \mathbf{u}) - p\mathbb{I} \quad \text{in } \Omega \tag{2.5}$$

using the incompressibility and the foregoing equation to eliminate the pressure, introducing the auxiliary unknowns

$$\mathbf{t} := \mathbf{e}(\mathbf{u}) \quad \text{and} \quad \boldsymbol{\rho} := \nabla \mathbf{u} - \mathbf{e}(\mathbf{u}),$$

which denote the strain and the vorticity, respectively, and observing from (2.5) that  $\boldsymbol{\sigma}$  is now required to be symmetric, which improves the approach from [1], we arrive at the following system of equations with unknowns  $\mathbf{t}$ ,  $\mathbf{u}$ ,  $\boldsymbol{\sigma}$ , and  $p$

$$\begin{aligned} \nabla \mathbf{u} &= \mathbf{t} + \boldsymbol{\rho} && \text{in } \Omega, \\ \mu(|\mathbf{t}|)\mathbf{t} - (\mathbf{u} \otimes \mathbf{u})^d &= \boldsymbol{\sigma}^d && \text{in } \Omega, \\ -\text{div } \boldsymbol{\sigma} &= \mathbf{f} && \text{in } \Omega, \\ \mathbf{u} &= \mathbf{g} && \text{on } \Gamma, \\ \boldsymbol{\sigma} &= \boldsymbol{\sigma}^t && \text{in } \Omega, \\ \int_{\Omega} \text{tr}(\boldsymbol{\sigma} + \mathbf{u} \otimes \mathbf{u}) &= 0 && . \end{aligned} \tag{2.6}$$

We notice here that the fluid incompressibility is implicitly incorporated in the new constitutive equation relating  $\boldsymbol{\sigma}$  and  $\mathbf{u}$  (second equation of (2.6)). In turn, the fact that the pressure must belong to  $L_0^2(\Omega)$  is guaranteed by the equivalent statement given by the last equation of (2.6). Indeed, it is easy to see (by taking the trace in [2.5]) that

$$p = -\frac{1}{n} \text{tr}(\boldsymbol{\sigma} + \mathbf{u} \otimes \mathbf{u}) \quad \text{in } \Omega. \tag{2.7}$$

### III. THE CONTINUOUS FORMULATION

#### A. The Augmented Mixed Formulation

We now proceed to derive a weak formulation of (2.6). First, we recall (cf. [12, 13]) that

$$\mathbb{H}(\mathbf{div}; \Omega) = \mathbb{H}_0(\mathbf{div}; \Omega) \oplus \mathbb{R}\mathbb{I} \tag{3.1}$$

where

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

In particular, decomposing  $\boldsymbol{\sigma}$  in (2.6) as  $\boldsymbol{\sigma} = \boldsymbol{\sigma}_0 + c\mathbb{I}$ , with  $\boldsymbol{\sigma}_0 \in \mathbb{H}_0(\mathbf{div}; \Omega)$ , we deduce from (3.1) and the last equation in (2.6) that  $c$  is given explicitly in terms of  $\mathbf{u}$  as

$$c = -\frac{1}{n|\Omega|} \int_{\Omega} \text{tr}(\mathbf{u} \otimes \mathbf{u}). \tag{3.2}$$

In this way, since  $\boldsymbol{\sigma}^d = \boldsymbol{\sigma}_0^d$  and  $\mathbf{div} \boldsymbol{\sigma} = \mathbf{div} \boldsymbol{\sigma}_0$  throughout the rest of the paper we rename  $\boldsymbol{\sigma}_0$  as  $\boldsymbol{\sigma} \in \mathbb{H}_0(\mathbf{div}; \Omega)$  and realize that the second, third, and fifth equations of (2.6) remain unchanged. In addition, thanks to the incompressibility condition and the first equation of (2.6), the unknown  $\mathbf{t}$  can be sought in the space

$$\mathbb{L}_{tr}^2(\Omega) := \{ \mathbf{s} \in \mathbb{L}^2(\Omega) : \text{trs} = 0 \}$$

whereas the vorticity  $\boldsymbol{\rho}$  lives in

$$\mathbb{L}_{skew}^2(\Omega) := \{ \boldsymbol{\eta} \in \mathbb{L}^2(\Omega) : \boldsymbol{\eta} = -\boldsymbol{\eta}^t \}.$$

Noticing first that it suffices to test the first equation of (2.6) against  $\boldsymbol{\tau} \in \mathbb{H}_0(\mathbf{div}; \Omega)$ , using the Dirichlet condition for  $\mathbf{u}$ , realizing that the constitutive equation given by the second equation of (2.6) needs to be tested only with  $\mathbf{s} \in \mathbb{L}_{tr}^2(\Omega)$ , and then imposing weakly the equilibrium equation and the symmetry of  $\boldsymbol{\sigma}$ , we arrive, at first instance, at the following weak formulation of (2.6): Find  $(\mathbf{t}, \boldsymbol{\sigma}, \boldsymbol{\rho}) \in \mathbb{L}_{tr}^2(\Omega) \times \mathbb{H}_0(\mathbf{div}; \Omega) \times \mathbb{L}_{skew}^2(\Omega)$ , and  $\mathbf{u}$  in a suitable space, such that

$$\begin{aligned} & \int_{\Omega} \mu(|\mathbf{t}|) \mathbf{t} : \mathbf{s} - \int_{\Omega} \boldsymbol{\sigma}^d : \mathbf{s} - \int_{\Omega} (\mathbf{u} \otimes \mathbf{u})^d : \mathbf{s} = 0 \quad \forall \mathbf{s} \in \mathbb{L}_{tr}^2(\Omega) \\ & \int_{\Omega} \boldsymbol{\tau}^d : \mathbf{t} + \int_{\Omega} \mathbf{u} \cdot \mathbf{div} \boldsymbol{\tau} + \int_{\Omega} \boldsymbol{\rho} : \boldsymbol{\tau} = \langle \boldsymbol{\tau} \mathbf{v}, \mathbf{g} \rangle \quad \forall \boldsymbol{\tau} \in \mathbb{H}_0(\mathbf{div}; \Omega) \\ & - \int_{\Omega} \mathbf{v} \cdot \mathbf{div} \boldsymbol{\sigma} - \int_{\Omega} \boldsymbol{\eta} : \boldsymbol{\sigma} = \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \quad \forall (\mathbf{v}, \boldsymbol{\eta}) \in \mathbf{L}^2(\Omega) \times \mathbb{L}_{skew}^2(\Omega) \end{aligned} \tag{3.3}$$

where  $\langle \cdot, \cdot \rangle$  denotes the duality pairing between  $\mathbf{H}^{-1/2}(\Gamma)$  and  $\mathbf{H}^{1/2}(\Gamma)$ .

We continue our analysis by observing, exactly as we did in [1], that by applying Cauchy-Schwarz and Hölder inequalities, and then using the compact (and hence continuous) injection  $i_c$  of  $\mathbf{H}^1(\Omega)$  into  $\mathbf{L}^4(\Omega)$  (see Rellich-Kondrachov compactness Theorem in [14, Theorem 6.3] or [15, Theorem 1.3.5]), that the third term in the first row of the foregoing system suggests to look for the unknown  $\mathbf{u}$  in  $\mathbf{H}^1(\Omega)$  and to restrict the set of corresponding test functions  $\mathbf{v}$  to the same

space. Consequently, we now augment (3.3) through the incorporation of the following redundant Galerkin terms:

$$\begin{aligned}
 \kappa_1 \int_{\Omega} \{ \sigma^d - \mu(|\mathbf{t}|)\mathbf{t} + (\mathbf{u} \otimes \mathbf{u})^d \} : \boldsymbol{\tau}^d &= 0 \quad \forall \boldsymbol{\tau} \in \mathbb{H}_0(\mathbf{div}; \Omega), \\
 \kappa_2 \int_{\Omega} \mathbf{div} \boldsymbol{\sigma} \cdot \mathbf{div} \boldsymbol{\tau} &= -\kappa_2 \int_{\Omega} \mathbf{f} \cdot \mathbf{div} \boldsymbol{\tau} \quad \forall \boldsymbol{\tau} \in \mathbb{H}_0(\mathbf{div}; \Omega), \\
 \kappa_3 \int_{\Omega} \{ \mathbf{e}(\mathbf{u}) - \mathbf{t} \} : \mathbf{e}(\mathbf{v}) &= 0 \quad \forall \mathbf{v} \in \mathbf{H}^1(\Omega), \\
 \kappa_4 \int_{\Omega} (\boldsymbol{\rho} - \{ \nabla \mathbf{u} - \mathbf{e}(\mathbf{u}) \}) : \boldsymbol{\eta} &= 0 \quad \forall \boldsymbol{\eta} \in \mathbb{L}_{skew}^2(\Omega), \\
 \kappa_5 \int_{\Gamma} \mathbf{u} \cdot \mathbf{v} &= \kappa_5 \int_{\Gamma} \mathbf{g} \cdot \mathbf{v} \quad \forall \mathbf{v} \in \mathbf{H}^1(\Omega),
 \end{aligned} \tag{3.4}$$

where  $\kappa_1, \kappa_2, \kappa_3, \kappa_4,$  and  $\kappa_5$  are positive parameters to be specified later. It is important to observe that, differently from the analysis in [1], here we have a third equation in (3.4) involving the strain tensor instead of the gradient of velocity, as well as a completely new fourth equation arising from the introduction of the vorticity  $\boldsymbol{\rho}$  as an auxiliary unknown. As we will see later on, these facts yield a number of modifications with respect to the solvability analysis carried out in [1].

The two foregoing systems of equations lead to the following augmented mixed formulation: Find  $\vec{\mathbf{t}} := (\mathbf{t}, \boldsymbol{\sigma}, \mathbf{u}, \boldsymbol{\rho}) \in \mathbf{H} := \mathbb{L}_{lr}^2(\Omega) \times \mathbb{H}_0(\mathbf{div}; \Omega) \times \mathbf{H}^1(\Omega) \times \mathbb{L}_{skew}^2(\Omega)$  such that

$$[(\mathbf{A} + \mathbf{B}_u)(\vec{\mathbf{t}}), \vec{\mathbf{s}}] = [\mathbf{F}, \vec{\mathbf{s}}] \quad \forall \vec{\mathbf{s}} := (s, \boldsymbol{\tau}, \mathbf{v}, \boldsymbol{\eta}) \in \mathbf{H}, \tag{3.5}$$

where  $[\cdot, \cdot]$  stands for the duality pairing between  $\mathbf{H}'$  and  $\mathbf{H}$ ,  $\mathbf{A} : \mathbf{H} \rightarrow \mathbf{H}'$  is the nonlinear operator

$$\begin{aligned}
 [\mathbf{A}(\vec{\mathbf{t}}), \vec{\mathbf{s}}] := & \int_{\Omega} \mu(|\mathbf{t}|)\mathbf{t} : s - \int_{\Omega} \sigma^d : s + \int_{\Omega} \boldsymbol{\tau}^d : \mathbf{t} + \int_{\Omega} \mathbf{u} \cdot \mathbf{div} \boldsymbol{\tau} - \int_{\Omega} \mathbf{v} \cdot \mathbf{div} \boldsymbol{\sigma} \\
 & + \int_{\Omega} \boldsymbol{\rho} : \boldsymbol{\tau} - \int_{\Omega} \boldsymbol{\eta} : \boldsymbol{\sigma} + \kappa_1 \int_{\Omega} \{ \sigma^d - \mu(|\mathbf{t}|)\mathbf{t} \} : \boldsymbol{\tau}^d + \kappa_2 \int_{\Omega} \mathbf{div} \boldsymbol{\sigma} \cdot \mathbf{div} \boldsymbol{\tau} \\
 & + \kappa_3 \int_{\Omega} \{ \mathbf{e}(\mathbf{u}) - \mathbf{t} \} : \mathbf{e}(\mathbf{v}) + \kappa_4 \int_{\Omega} (\boldsymbol{\rho} - \{ \nabla \mathbf{u} - \mathbf{e}(\mathbf{u}) \}) : \boldsymbol{\eta} + \kappa_5 \int_{\Gamma} \mathbf{u} \cdot \mathbf{v},
 \end{aligned} \tag{3.6}$$

$\mathbf{F} : \mathbf{H} \rightarrow \mathbf{R}$  is the bounded linear functional

$$[\mathbf{F}, \vec{\mathbf{s}}] := \langle \boldsymbol{\tau} \mathbf{v}, \mathbf{g} \rangle + \int_{\Omega} \mathbf{f} \cdot \{ \mathbf{v} - \kappa_2 \mathbf{div} \boldsymbol{\tau} \} + \kappa_5 \int_{\Gamma} \mathbf{g} \cdot \mathbf{v}, \tag{3.7}$$

and for each  $\mathbf{z} \in \mathbf{H}^1(\Omega)$ ,  $\mathbf{B}_z : \mathbf{H} \rightarrow \mathbf{H}'$  is the bounded linear operator

$$[\mathbf{B}_z(\vec{\mathbf{t}}), \vec{\mathbf{s}}] := \int_{\Omega} (\mathbf{z} \otimes \mathbf{u})^d : \{ \kappa_1 \boldsymbol{\tau}^d - s \}, \tag{3.8}$$

for all  $\vec{\mathbf{t}} := (\mathbf{t}, \boldsymbol{\sigma}, \mathbf{u}, \boldsymbol{\rho}), \vec{\mathbf{s}} := (s, \boldsymbol{\tau}, \mathbf{v}, \boldsymbol{\eta}) \in \mathbf{H}$ .

The boundedness of  $\mathbf{F}$  and  $\mathbf{B}_z$  will be confirmed in the following section, where we introduce our fixed-point approach to study the well-posedness of (3.5).

**B. A Fixed-Point Approach**

We begin by defining the operator  $\mathbf{T} : \mathbf{H}^1(\Omega) \rightarrow \mathbf{H}^1(\Omega)$  by

$$\mathbf{T}(z) := \mathbf{u} \quad \forall z \in \mathbf{H}^1(\Omega),$$

where  $\mathbf{u}$  is the third component of the unique solution (to be confirmed below) of the nonlinear problem: Find  $\vec{t} := (t, \sigma, \mathbf{u}, \rho) \in \mathbf{H}$  such that

$$[(\mathbf{A} + \mathbf{B}_z)(\vec{t}), \vec{s}] = [\mathbf{F}, \vec{s}] \quad \forall \vec{s} := (s, \boldsymbol{\tau}, \mathbf{v}, \eta) \in \mathbf{H}. \tag{3.9}$$

It follows that our augmented mixed formulation (3.5) can be rewritten, equivalently, as the fixed-point problem: Find  $\mathbf{u} \in \mathbf{H}^1(\Omega)$  such that

$$\mathbf{T}(\mathbf{u}) = \mathbf{u}.$$

The following useful inequalities will be employed below to analyze the well-posedness of (3.9) and a particular case of it to be considered in Section C below.

**Lemma 3.1.** *There exists  $c_1(\Omega) > 0$  such that*

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

**Proof.** See [12, Proposition 3.1, Chapter IV]. ■

**Lemma 3.2.** *There holds*

$$\|\mathbf{e}(\mathbf{v})\|_{0,\Omega}^2 \geq \frac{1}{2} \|\mathbf{v}\|_{1,\Omega}^2 \quad \forall \mathbf{v} \in \mathbf{H}_0^1(\Omega).$$

**Proof.** See [16, Theorem 10.1]. ■

**Lemma 3.3.** *There exists  $\kappa_0 > 0$  such that*

$$\|\mathbf{e}(\mathbf{v})\|_{0,\Omega}^2 + \|\mathbf{v}\|_{0,\Gamma}^2 \geq \kappa_0 \|\mathbf{v}\|_{1,\Omega}^2 \quad \forall \mathbf{v} \in \mathbf{H}^1(\Omega).$$

**Proof.** See [17, Lemma 3.1 and inequality (3.9)]. ■

Note that Lemmas 3.2 and 3.3 correspond to the Korn first inequality and a modified Korn inequality, respectively, which are more elaborated lower bounds than the usual Poincaré one used in [1]. As announced before, the application of these estimates is caused by the present dependence on the strain tensor in the third equation of (3.4). In turn, we also need to recall from [10] that, under the assumptions given by (2.3), the nonlinear operator induced by  $\mu$  is Lipschitz continuous and strongly monotone. More precisely, we have the following result.

**Lemma 3.4.** *Let  $L_\mu := \max\{\mu_2, 2\mu_2 - \mu_1\}$ , where  $\mu_1$  and  $\mu_2$  are the bounds of  $\mu$  given in (2.3). Then for each  $\mathbf{r}, \mathbf{s} \in \mathbb{L}^2(\Omega)$  there holds*

$$\|\mu(|\mathbf{r}|)\mathbf{r} - \mu(|\mathbf{s}|)\mathbf{s}\|_{0,\Omega} \leq L_\mu \|\mathbf{r} - \mathbf{s}\|_{0,\Omega}, \tag{3.10}$$

and

$$\int_{\Omega} \{\mu(|\mathbf{r}|)\mathbf{r} - \mu(|\mathbf{s}|)\mathbf{s}\} : (\mathbf{r} - \mathbf{s}) \geq \mu_1 \|\mathbf{r} - \mathbf{s}\|_{0,\Omega}^2. \tag{3.11}$$

**Proof.** See [10, Theorem 3.8] for details. ■

The following lemma provides sufficient conditions under which the operator  $\mathbf{T}$  is well-defined.

**Lemma 3.5.** Assume that  $\kappa_1 \in \left(0, \frac{2\delta\mu_1}{L_\mu}\right)$ ,  $\kappa_3 \in \left(0, 2\tilde{\delta}(\mu_1 - \frac{\kappa_1 L_\mu}{2\delta})\right)$ ,  $\kappa_4 \in \left(0, 2\hat{\delta}\kappa_0\beta(\Omega)\right)$ , and  $\kappa_2, \kappa_5 > 0$ , with  $\delta \in \left(0, \frac{2}{L_\mu}\right)$ ,  $\tilde{\delta}, \hat{\delta} \in (0, 2)$ , and  $\beta(\Omega) := \min\left\{\kappa_3\left(1 - \frac{\tilde{\delta}}{2}\right), \kappa_5\right\}$ . Then, there exists  $\varepsilon_0 > 0$  such that for each  $\varepsilon \in (0, \varepsilon_0)$ , problem (3.9) has a unique solution  $\vec{\mathbf{t}} := (\mathbf{t}, \boldsymbol{\sigma}, \mathbf{u}, \boldsymbol{\rho}) \in \mathbf{H}$  for each  $\mathbf{z} \in \mathbf{H}^1(\Omega)$  such that  $\|\mathbf{z}\|_{1,\Omega} \leq \varepsilon$ . Moreover, there exists  $c_T > 0$ , independent of  $\mathbf{z}$  and the data  $\mathbf{f}$  and  $\mathbf{g}$ , such that

$$\|\mathbf{T}(\mathbf{z})\|_{1,\Omega} = \|\mathbf{u}\|_{1,\Omega} \leq \|\vec{\mathbf{t}}\| \leq c_T \{\|\mathbf{f}\|_{0,\Omega} + \|\mathbf{g}\|_{0,\Gamma} + \|\mathbf{g}\|_{1/2,\Gamma}\}. \tag{3.12}$$

**Proof.** We proceed similarly as in the proof of [1, Lemma 3.4]. In fact, given  $\mathbf{z} \in \mathbf{H}^1(\Omega)$ , we first deduce from (3.6), using the Cauchy-Schwarz inequality, the Lipschitz continuity of the operator induced by  $\mu$  (cf. (3.10) in Lemma 3.4), and the trace operator  $\boldsymbol{\gamma}_0 : \mathbf{H}^1(\Omega) \rightarrow \mathbf{L}^2(\Gamma)$ , that there exists a positive constant  $L_A$ , depending on  $L_\mu$ , the parameters  $\kappa_i, i \in \{1, \dots, 5\}$ , and  $\|\boldsymbol{\gamma}_0\|$ , such that

$$[A(\vec{\mathbf{t}}) - A(\vec{\mathbf{r}}), \vec{\mathbf{s}}] \leq L_A \|\vec{\mathbf{t}} - \vec{\mathbf{r}}\| \|\vec{\mathbf{s}}\| \tag{3.13}$$

for all  $\vec{\mathbf{t}}, \vec{\mathbf{r}}, \vec{\mathbf{s}} \in \mathbf{H}$ . In turn, recalling that  $\mathbf{i}_c$  denotes the continuous injection of  $\mathbf{H}^1(\Omega)$  into  $\mathbf{L}^4(\Omega)$ , it readily follows (3.8), by applying Cauchy-Schwarz and Hölder inequalities, that

$$|[\mathbf{B}_z(\vec{\mathbf{t}}), \vec{\mathbf{s}}]| \leq \|\mathbf{i}_c\|^2 (\kappa_1^2 + 1)^{1/2} \|\mathbf{z}\|_{1,\Omega} \|\vec{\mathbf{t}}\| \|\vec{\mathbf{s}}\| \quad \forall \vec{\mathbf{t}}, \vec{\mathbf{s}} \in \mathbf{H}, \tag{3.14}$$

which, thanks to the linearity of  $\mathbf{B}_z$ , and together with (3.13), proves that the operator  $\mathbf{A} + \mathbf{B}_z$  is Lipschitz continuous with constant  $L_A + \|\mathbf{i}_c\|^2 (\kappa_1^2 + 1)^{1/2} \|\mathbf{z}\|_{1,\Omega}$ . Next, it is also clear from (3.6) that for each  $\vec{\mathbf{r}} := (\mathbf{r}, \boldsymbol{\zeta}, \mathbf{w}, \boldsymbol{\xi}), \vec{\mathbf{s}} := (\mathbf{s}, \boldsymbol{\tau}, \mathbf{v}, \boldsymbol{\eta}) \in \mathbf{H}$  there holds

$$\begin{aligned} & [\mathbf{A}(\vec{\mathbf{r}}) - \mathbf{A}(\vec{\mathbf{s}}), \vec{\mathbf{r}} - \vec{\mathbf{s}}] \\ &= \int_{\Omega} \{\mu(|\mathbf{r}|)\mathbf{r} - \mu(|\mathbf{s}|)\mathbf{s}\} : (\mathbf{r} - \mathbf{s}) + \kappa_1 \|(\boldsymbol{\zeta} - \boldsymbol{\tau})^d\|_{0,\Omega}^2 \\ &\quad - \kappa_1 \int_{\Omega} \{\mu(|\mathbf{r}|)\mathbf{r} - \mu(|\mathbf{s}|)\mathbf{s}\} : (\boldsymbol{\zeta} - \boldsymbol{\tau})^d + \kappa_2 \|\mathbf{div}(\boldsymbol{\zeta} - \boldsymbol{\tau})\|_{0,\Omega}^2 + \kappa_3 \|\mathbf{e}(\mathbf{w} - \mathbf{v})\|_{0,\Omega}^2 \\ &\quad - \kappa_3 \int_{\Omega} (\mathbf{r} - \mathbf{s}) : \mathbf{e}(\mathbf{w} - \mathbf{v}) + \kappa_4 \|\boldsymbol{\xi} - \boldsymbol{\eta}\|_{0,\Omega}^2 + \kappa_5 \|\mathbf{w} - \mathbf{v}\|_{0,\Gamma}^2 \\ &\quad - \kappa_4 \int_{\Omega} \{\nabla(\mathbf{w} - \mathbf{v}) - \mathbf{e}(\mathbf{w} - \mathbf{v})\} : (\boldsymbol{\xi} - \boldsymbol{\eta}), \end{aligned}$$

which, using the Cauchy-Schwarz and Young inequalities, the Lipschitz continuity and strong monotonicity properties of the operator induced by  $\mu$  (cf. [3.10] and [3.11]), and the fact that

$$\|\nabla(\mathbf{w} - \mathbf{v}) - \mathbf{e}(\mathbf{w} - \mathbf{v})\|_{0,\Omega}^2 = \|\mathbf{w} - \mathbf{v}\|_{1,\Omega}^2 - \|\mathbf{e}(\mathbf{w} - \mathbf{v})\|_{0,\Omega}^2,$$

yields for any  $\delta, \tilde{\delta}, \hat{\delta} > 0$ , the bound

$$\begin{aligned}
 & [\mathbf{A}(\vec{r}) - \mathbf{A}(\vec{s}), \vec{r} - \vec{s}] \\
 & \geq \left\{ \left( \mu_1 - \frac{\kappa_1 L_\mu}{2\delta} \right) - \frac{\kappa_3}{2\tilde{\delta}} \right\} \|\mathbf{r} - \mathbf{s}\|_{0,\Omega}^2 + \kappa_1 \left( 1 - \frac{L_\mu \delta}{2} \right) \|(\boldsymbol{\zeta} - \boldsymbol{\tau})^d\|_{0,\Omega}^2 \\
 & \quad + \kappa_2 \|\mathbf{div}(\boldsymbol{\zeta} - \boldsymbol{\tau})\|_{0,\Omega}^2 + \left\{ \kappa_3 \left( 1 - \frac{\tilde{\delta}}{2} \right) + \frac{\kappa_4}{2\tilde{\delta}} \right\} \|\mathbf{e}(\mathbf{w} - \mathbf{v})\|_{0,\Omega}^2 + \kappa_5 \|\mathbf{w} - \mathbf{v}\|_{0,\Gamma}^2 \\
 & \quad + \kappa_4 \left( 1 - \frac{\hat{\delta}}{2} \right) \|\boldsymbol{\xi} - \boldsymbol{\eta}\|_{0,\Omega}^2 - \frac{\kappa_4}{2\hat{\delta}} \|\mathbf{w} - \mathbf{v}\|_{1,\Omega}^2.
 \end{aligned} \tag{3.15}$$

Then, discarding the expression  $\frac{\kappa_4}{2\hat{\delta}}$  multiplying  $\|\mathbf{e}(\mathbf{w} - \mathbf{v})\|_{0,\Omega}^2$ , and according to the hypotheses on  $\delta, \kappa_1, \tilde{\delta}, \kappa_3, \hat{\delta}, \kappa_4, \kappa_2$ , and  $\kappa_5$ , and applying Lemmas 3.1 and 3.3, we can define the positive constants

$$\begin{aligned}
 \alpha_0(\Omega) & := \left( \mu_1 - \frac{\kappa_1 L_\mu}{2\delta} \right) - \frac{\kappa_3}{2\tilde{\delta}}, & \alpha_1(\Omega) & := \min \left\{ \kappa_1 \left( 1 - \frac{L_\mu \delta}{2} \right), \frac{\kappa_2}{2} \right\}, \\
 \alpha_2(\Omega) & := \min \left\{ \alpha_1(\Omega) c_1(\Omega), \frac{\kappa_2}{2} \right\}, & \alpha_3(\Omega) & := \kappa_0 \beta(\Omega) - \frac{\kappa_4}{2\tilde{\delta}}, & \text{and} & \quad \alpha_4(\Omega) := \kappa_4 \left( 1 - \frac{\hat{\delta}}{2} \right),
 \end{aligned}$$

which allow us to deduce from (3.15) that

$$[\mathbf{A}(\vec{r}) - \mathbf{A}(\vec{s}), \vec{r} - \vec{s}] \geq \alpha(\Omega) \|\vec{r} - \vec{s}\|^2 \quad \forall \vec{r}, \vec{s} \in \mathbf{H}, \tag{3.16}$$

where

$$\alpha(\Omega) := \min \{ \alpha_0(\Omega), \alpha_2(\Omega), \alpha_3(\Omega), \alpha_4(\Omega) \}$$

is the strong monotonicity constant of  $\mathbf{A}$ . We remark that the fourth equation in (3.4) motivates an additional application of the Young inequality, which implies the need of establishing suitable relationships between the constant  $\hat{\delta}$  and the remaining parameters appearing in the foregoing proof of strong monotonicity. Next, a combination of (3.14) and (3.16) implies that

$$[(\mathbf{A} + \mathbf{B}_z)(\vec{r}) - (\mathbf{A} + \mathbf{B}_z)(\vec{s}), \vec{r} - \vec{s}] \geq \frac{\alpha(\Omega)}{2} \|\vec{r} - \vec{s}\|^2 \quad \forall \vec{r}, \vec{s} \in \mathbf{H}, \tag{3.17}$$

provided  $\|\mathbf{z}\|_{1,\Omega} \leq \varepsilon_0$ , with

$$\varepsilon_0 := \frac{\alpha(\Omega)}{2 \|\mathbf{i}_c\|^2 (\kappa_1^2 + 1)^{1/2}}, \tag{3.18}$$

which confirms the strong monotonicity of the nonlinear operator  $\mathbf{A} + \mathbf{B}_z$ . Conversely, it follows from (3.7), using Cauchy-Schwarz's inequality and the trace theorems in  $\mathbb{H}(\mathbf{div}; \Omega)$  and  $\mathbf{H}^1(\Omega)$ , that  $\mathbf{F} \in \mathbf{H}'$  with

$$\|\mathbf{F}\| \leq M_{\mathbf{T}} \left\{ \|\mathbf{f}\|_{0,\Omega} + \|\mathbf{g}\|_{0,\Gamma} + \|\mathbf{g}\|_{1/2,\Gamma} \right\},$$

where  $M_T := \max \left\{ (1 + \kappa_2^2)^{1/2}, \kappa_5 \|\boldsymbol{y}_0\| \right\}$ . Consequently, a straightforward application of [18, Theorem 3.3.23] (which establishes the bijectivity of Lipschitz continuous and strongly monotone operators) implies that there exists a unique solution  $\vec{\boldsymbol{t}} \in \mathbf{H}$  of (3.9). Finally, applying (3.17) and performing simple algebraic manipulations, we derive (3.12) with the positive constant  $c_T := \frac{2M_T}{\alpha(\Omega)}$ . ■

We now observe that the constant  $\alpha(\Omega)$  yielding the strong monotonicity of  $\mathbf{A} + \mathbf{B}_z$  can be maximized by taking the parameters  $\delta, \kappa_1, \tilde{\delta}, \kappa_3, \hat{\delta}$ , and  $\kappa_4$  as the middle points of their feasible ranges, and by choosing  $\kappa_2$  and  $\kappa_5$  so that they maximize the minima defining  $\alpha_1(\Omega)$  and  $\beta(\Omega)$ , respectively. More precisely, we simply take

$$\begin{aligned} \delta &= \frac{1}{L_\mu}, & \kappa_1 &= \frac{\delta \mu_1}{L_\mu} = \frac{\mu_1}{L_\mu^2}, & \tilde{\delta} &= 1, & \kappa_3 &= \tilde{\delta} \left( \mu_1 - \frac{\kappa_1 L_\mu}{2\tilde{\delta}} \right) = \frac{\mu_1}{2}, \\ \kappa_2 &= 2\kappa_1 \left( 1 - \frac{L_\mu \delta}{2} \right) = \kappa_1 = \frac{\mu_1}{L_\mu^2}, & \kappa_5 &= \kappa_3 \left( 1 - \frac{\tilde{\delta}}{2} \right) = \frac{\kappa_3}{2} = \frac{\mu_1}{4}, \\ \hat{\delta} &= 1, & \text{and } \kappa_4 &= \hat{\delta} \kappa_0 \beta(\Omega) = \kappa_0 \kappa_5 = \frac{\kappa_0 \mu_1}{4}, \end{aligned} \tag{3.19}$$

which yields

$$\alpha_0(\Omega) = \frac{\mu_1}{4}, \quad \alpha_1(\Omega) = \frac{\mu_1}{2L_\mu^2}, \quad \alpha_2(\Omega) = \min \{c_1(\Omega), 1\} \frac{\mu_1}{2L_\mu^2}, \quad \alpha_3(\Omega) = \alpha_4(\Omega) = \frac{\kappa_0 \mu_1}{8},$$

and hence

$$\alpha(\Omega) = \min \left\{ \frac{\mu_1}{4}, \frac{\kappa_0 \mu_1}{8}, \min \{c_1(\Omega), 1\} \frac{\mu_1}{2L_\mu^2} \right\}.$$

Note that the values of the stabilization parameters  $\kappa_i, i \in \{1, \dots, 5\}$ , given in (3.19), are all explicitly computable in terms of the constants  $\mu_1$  and  $\mu_2$  (cf. (2.3)), except  $\kappa_4$ , which depends on the usually unknown constant  $\kappa_0$  appearing in the Korn-type inequality given by Lemma 3.3. According to this, the aforementioned explicit parameters in (3.19) together with an heuristic choice for  $\kappa_0$  (and hence for  $\kappa_4 = \frac{\kappa_0 \mu_1}{4}$ ) will be used below in Section VI for the corresponding numerical experiments.

### C. The Case of a Homogeneous Dirichlet Boundary Condition

We now address the case of a homogeneous Dirichlet condition for the velocity  $\boldsymbol{u}$  on the boundary  $\Gamma$ . In this way, the present section constitutes a significant complement not only of the foregoing discussion, but also of the results provided in [1], where the specific analysis for this kind of boundary conditions was not included. In particular, as  $\boldsymbol{u}$  lives now in  $\mathbf{H}_0^1(\Omega)$ , we first realize that the last equation of (3.4) is not needed anymore, which means that only four stabilization parameters are required. Consequently, our resulting augmented mixed formulation becomes: Find  $\vec{\boldsymbol{t}} := (\boldsymbol{t}, \boldsymbol{\sigma}, \boldsymbol{u}, \boldsymbol{\rho}) \in \mathbf{H}_0 := \mathbb{L}_{tr}^2(\Omega) \times \mathbb{H}_0(\mathbf{div}; \Omega) \times \mathbf{H}_0^1(\Omega) \times \mathbb{L}_{skew}^2(\Omega)$  such that

$$[(\mathbf{A} + \mathbf{B}_u)(\vec{\boldsymbol{t}}), \vec{\boldsymbol{s}}] = [\mathbf{F}, \vec{\boldsymbol{s}}] \quad \forall \vec{\boldsymbol{s}} := (s, \boldsymbol{\tau}, \boldsymbol{v}, \boldsymbol{\eta}) \in \mathbf{H}_0, \tag{3.20}$$

where the nonlinear operator  $\mathbf{A} : \mathbf{H}_0 \rightarrow \mathbf{H}'_0$  is defined by

$$\begin{aligned} [\mathbf{A}(\vec{t}), \vec{s}] &:= \int_{\Omega} \mu(|t|)t : s - \int_{\Omega} \sigma^d : s + \int_{\Omega} \tau^d : t + \int_{\Omega} \mathbf{u} \cdot \mathbf{div} \tau - \int_{\Omega} \mathbf{v} \cdot \mathbf{div} \sigma \\ &+ \int_{\Omega} \rho : \tau - \int_{\Omega} \eta : \sigma + \kappa_1 \int_{\Omega} \{\sigma^d - \mu(|t|)t\} : \tau^d + \kappa_2 \int_{\Omega} \mathbf{div} \sigma \cdot \mathbf{div} \tau \\ &+ \kappa_3 \int_{\Omega} \{\mathbf{e}(\mathbf{u}) - t\} : \mathbf{e}(\mathbf{v}) + \kappa_4 \int_{\Omega} (\rho - \{\nabla \mathbf{u} - \mathbf{e}(\mathbf{u})\}) : \eta, \end{aligned}$$

the bounded linear functional  $\mathbf{F} : \mathbf{H}_0 \rightarrow \mathbb{R}$  corresponds to

$$[\mathbf{F}, \vec{s}] := \int_{\Omega} \mathbf{f} \cdot \{\mathbf{v} - \kappa_2 \mathbf{div} \tau\}$$

and the operator  $\mathbf{B}_z$  is given, as before, by (3.8). In turn, the associated fixed point operator is given now by  $\mathbf{T}_0 : \mathbf{H}_0^1(\Omega) \rightarrow \mathbf{H}_0^1(\Omega)$ , where

$$\mathbf{T}_0(\mathbf{z}) := \mathbf{u} \quad \forall \mathbf{z} \in \mathbf{H}_0^1(\Omega),$$

and  $\mathbf{u}$  is the third component of the unique solution (to be confirmed below) of the nonlinear problem: Find  $\vec{t} := (t, \sigma, \mathbf{u}, \rho) \in \mathbf{H}_0$  such that

$$[(\mathbf{A} + \mathbf{B}_z)(\vec{t}), \vec{s}] = [\mathbf{F}, \vec{s}] \quad \forall \vec{s} := (s, \tau, \mathbf{v}, \eta) \in \mathbf{H}_0. \tag{3.21}$$

In this way, and similarly as in Section B, our augmented mixed formulation (3.20) can be rewritten, equivalently, as the fixed point problem: Find  $\vec{t} := (t, \sigma, \mathbf{u}, \rho) \in \mathbf{H}_0$  such that

$$\mathbf{T}_0(\mathbf{u}) = \mathbf{u}.$$

The analogue of Lemma 3.5, which will make use now of the first Korn inequality (cf. Lemma 3.2) instead of Lemma 3.3, is established as follows.

**Lemma 3.6.** *Assume that  $\kappa_1 \in (0, \frac{2\delta\mu_1}{L_\mu})$ ,  $\kappa_3 \in (0, 2\tilde{\delta}(\mu_1 - \frac{\kappa_1 L_\mu}{2\delta}))$ ,  $\kappa_4 \in (0, 2\hat{\delta}\kappa_3(1 - \frac{\tilde{\delta}}{2}))$ , and  $\kappa_2 > 0$ , with  $\delta \in (0, \frac{2}{L_\mu})$ , and  $\tilde{\delta}, \hat{\delta} \in (0, 2)$ . Then, there exists  $\varepsilon_0 > 0$  such that for each  $\varepsilon \in (0, \varepsilon_0)$ , problem (3.21) has a unique solution  $\vec{t} := (t, \sigma, \mathbf{u}, \rho) \in \mathbf{H}_0$  for each  $\mathbf{z} \in \mathbf{H}_0^1(\Omega)$  such that  $\|\mathbf{z}\|_{1,\Omega} \leq \varepsilon$ . Moreover, there holds*

$$\|\mathbf{T}_0(\mathbf{z})\|_{1,\Omega} = \|\mathbf{u}\|_{1,\Omega} \leq \|\vec{t}\| \leq \frac{2(1 + \kappa_2^2)^{1/2}}{\alpha(\Omega)} \|\mathbf{f}\|_{0,\Omega}.$$

**Proof.** As some details are either similar or almost verbatim to those provided in the proof of Lemma 3.5, we concentrate here on the main difference of the analysis, which has to do with the strong monotonicity of the nonlinear operator  $\mathbf{A}$ . In other words, our starting point here is inequality (3.15), which in the present case, and after applying Lemma 3.2, leads to

$$\begin{aligned} &[\mathbf{A}(\vec{r}) - \mathbf{A}(\vec{s}), \vec{r} - \vec{s}] \\ &\geq \left\{ \left( \mu_1 - \frac{\kappa_1 L_\mu}{2\delta} \right) - \frac{\kappa_3}{2\tilde{\delta}} \right\} \|\mathbf{r} - \mathbf{s}\|_{0,\Omega}^2 + \kappa_1 \left( 1 - \frac{L_\mu \delta}{2} \right) \|(\boldsymbol{\zeta} - \boldsymbol{\tau})^d\|_{0,\Omega}^2 \\ &\quad + \kappa_2 \|\mathbf{div}(\boldsymbol{\zeta} - \boldsymbol{\tau})\|_{0,\Omega}^2 + \left\{ \frac{\kappa_3}{2} \left( 1 - \frac{\tilde{\delta}}{2} \right) - \frac{\kappa_4}{4\hat{\delta}} \right\} \|\mathbf{w} - \mathbf{v}\|_{1,\Omega}^2 + \kappa_4 \left( 1 - \frac{\hat{\delta}}{2} \right) \|\boldsymbol{\xi} - \boldsymbol{\eta}\|_{0,\Omega}^2. \end{aligned}$$

Then, according to the hypotheses on  $\delta, \kappa_1, \tilde{\delta}, \kappa_3, \hat{\delta}, \kappa_4$ , and applying Lemma 3.1, we can define the positive constants

$$\begin{aligned} \alpha_0(\Omega) &:= \left( \mu_1 - \frac{\kappa_1 L_\mu}{2\delta} \right) - \frac{\kappa_3}{2\tilde{\delta}}, & \alpha_1(\Omega) &:= \min \left\{ \kappa_1 \left( 1 - \frac{L_\mu \delta}{2} \right), \frac{\kappa_2}{2} \right\}, \\ \alpha_2(\Omega) &:= \min \left\{ \alpha_1(\Omega) c_1(\Omega), \frac{\kappa_2}{2} \right\}, & \alpha_3(\Omega) &:= \frac{\kappa_3}{2} \left( 1 - \frac{\tilde{\delta}}{2} \right) - \frac{\kappa_4}{4\hat{\delta}}, \quad \text{and} \\ \alpha_4(\Omega) &:= \kappa_4 \left( 1 - \frac{\hat{\delta}}{2} \right), \end{aligned}$$

which, combined with the foregoing inequality, implies

$$[\mathbf{A}(\vec{r}) - \mathbf{A}(\vec{s}), \vec{r} - \vec{s}] \geq \alpha(\Omega) \|\vec{r} - \vec{s}\|^2 \quad \forall \vec{r}, \vec{s} \in \mathbf{H},$$

where

$$\alpha(\Omega) := \min \{ \alpha_0(\Omega), \alpha_2(\Omega), c_p \alpha_3(\Omega), \alpha_4(\Omega) \} \tag{3.22}$$

and  $c_p$  is the positive constant provided by Poincaré’s inequality. The rest proceeds exactly as in the proof of Lemma 3.5. In particular, the constant  $\varepsilon_0$  is given by (3.18) but with  $\alpha(\Omega)$  defined now by (3.22). We omit further details. ■

We end this section by remarking, as in Section B, that  $\alpha(\Omega)$  is maximized by taking the parameters  $\delta, \kappa_1, \tilde{\delta}, \kappa_3, \hat{\delta}$ , and  $\kappa_4$  as the midpoints of their feasible ranges, and by choosing  $\kappa_2$  so that it maximizes  $\alpha_1(\Omega)$ . The above means that we simply take

$$\begin{aligned} \delta &= \frac{1}{L_\mu}, & \kappa_1 &= \frac{\delta \mu_1}{L_\mu} = \frac{\mu_1}{L_\mu^2}, & \tilde{\delta} &= 1, & \kappa_3 &= \tilde{\delta} \left( \mu_1 - \frac{\kappa_1 L_\mu}{2\delta} \right) = \frac{\mu_1}{2}, \\ \kappa_2 &= 2\kappa_1 \left( 1 - \frac{L_\mu \delta}{2} \right) = \kappa_1 = \frac{\mu_1}{L_\mu^2}, & \hat{\delta} &= 1, & \text{and} & \kappa_4 &= \hat{\delta} \kappa_3 \left( 1 - \frac{\tilde{\delta}}{2} \right) = \frac{\kappa_3}{2} = \frac{\mu_1}{4}, \end{aligned} \tag{3.23}$$

which yields  $\alpha_0(\Omega) = \frac{\mu_1}{4}, \alpha_1(\Omega) = \frac{\mu_1}{2L_\mu^2}, \alpha_2(\Omega) = \min \{ c_1(\Omega), 1 \} \frac{\mu_1}{2L_\mu^2}, \alpha_3(\Omega) = \frac{\mu_1}{16}, \alpha_4(\Omega) = \frac{\mu_1}{8}$ , and hence

$$\alpha(\Omega) = \min \left\{ \min \{ c_1(\Omega), 1 \} \frac{\mu_1}{2L_\mu^2}, c_p \frac{\mu_1}{16}, \frac{\mu_1}{8} \right\}.$$

The explicit parameters defined in (3.23) will be used below in Section V for the corresponding numerical experiments with homogeneous Dirichlet boundary conditions for the velocity  $\mathbf{u}$ .

**D. Solvability Analysis of the Fixed-Point Equations**

We now aim to establish the existence of unique fixed points of the operators  $\mathbf{T}$  and  $\mathbf{T}_0$ . Actually, in what follows we just concentrate in the analysis of  $\mathbf{T}$  as the one of  $\mathbf{T}_0$  is completely analogous.

Moreover, as the approach follows very closely the ideas developed in [1], we simplify the presentation as much as possible and frequently refer to the results in that work. The same remarks apply for the subsequent sections.

To prove the existence of a unique fixed point of the operator  $\mathbf{T}$ , it suffices to verify the hypotheses of the classical Banach fixed point theorem, whose statement is recalled in what follows.

**Theorem 3.7.** *Let  $W$  be a closed subset of a Banach space  $X$ , and let  $T : W \rightarrow W$  be a contraction mapping. Then  $T$  has a unique fixed point.*

We begin the analysis with the following straightforward consequence of Lemma 3.5.

**Lemma 3.8.** *Let  $\varepsilon \in (0, \varepsilon_0)$ , with  $\varepsilon_0$  given by (3.18) (cf. proof of Lemma 3.5), let  $W_\varepsilon$  be the closed ball defined by  $W_\varepsilon := \{\mathbf{z} \in \mathbf{H}^1(\Omega) : \|\mathbf{z}\|_{1,\Omega} \leq \varepsilon\}$ , and assume that*

$$c_T \{ \|\mathbf{f}\|_{0,\Omega} + \|\mathbf{g}\|_{0,\Gamma} + \|\mathbf{g}\|_{1/2,\Gamma} \} \leq \varepsilon, \tag{3.24}$$

with  $c_T$  given at the end of the proof of Lemma 3.5. Then  $\mathbf{T}(W_\varepsilon) \subseteq W_\varepsilon$ .

In turn, the following lemma establishes a key estimate in deriving the continuity of  $\mathbf{T}$ .

**Lemma 3.9.** *Let  $\varepsilon \in (0, \varepsilon_0)$ , with  $\varepsilon_0$  given by (3.18), and let  $W_\varepsilon := \{\mathbf{z} \in \mathbf{H}^1(\Omega) : \|\mathbf{z}\|_{1,\Omega} \leq \varepsilon\}$ . Then there exists a positive constant  $C_T := \frac{2(\kappa_1^2+1)^{1/2}\|\mathbf{i}_c\|}{\alpha(\Omega)}$ , such that*

$$\|\mathbf{T}(\mathbf{z}) - \mathbf{T}(\tilde{\mathbf{z}})\|_{1,\Omega} \leq C_T \|\mathbf{T}(\tilde{\mathbf{z}})\|_{1,\Omega} \|\mathbf{z} - \tilde{\mathbf{z}}\|_{\mathbf{L}^4(\Omega)} \quad \forall \mathbf{z}, \tilde{\mathbf{z}} \in W_\varepsilon. \tag{3.25}$$

**Proof.** It follows as in the proof of [1, Lemma 3.7]. ■

The main result of this section is stated next.

**Theorem 3.10.** *Suppose that the parameters  $\kappa_i, i \in \{1, \dots, 5\}$ , satisfy the conditions required by Lemma 3.5. In addition, given  $\varepsilon \in (0, \varepsilon_0)$ , with  $\varepsilon_0$  defined by (3.18), we let  $W_\varepsilon := \{\mathbf{z} \in \mathbf{H}^1(\Omega) : \|\mathbf{z}\|_{1,\Omega} \leq \varepsilon\}$ , and assume that the data satisfy (3.24; cf. Lemma 3.8). Then, the augmented mixed formulation (3.5) has a unique solution  $\vec{\mathbf{t}} := (\mathbf{t}, \boldsymbol{\sigma}, \mathbf{u}, \boldsymbol{\rho}) \in \mathbf{H}$  with  $\mathbf{u} \in W_\varepsilon$ , and there holds*

$$\|\vec{\mathbf{t}}\| \leq c_T \{ \|\mathbf{f}\|_{0,\Omega} + \|\mathbf{g}\|_{0,\Gamma} + \|\mathbf{g}\|_{1/2,\Gamma} \}, \tag{3.26}$$

with  $c_T$  given at the end of the proof of Lemma 3.5.

**Proof.** We proceed similarly as in the proof of [1, Theorem 3.9] and make use of the classical Banach fixed point Theorem to prove that the mapping  $\mathbf{T}$  has a unique fixed point in  $W_\varepsilon$ . In fact, given  $\varepsilon \in (0, \varepsilon_0)$ , we first notice, using (3.25) and the continuity of  $\mathbf{i}_c : \mathbf{H}^1(\Omega) \rightarrow \mathbf{L}^4(\Omega)$ , that

$$\|\mathbf{T}(\mathbf{z}) - \mathbf{T}(\tilde{\mathbf{z}})\|_{1,\Omega} \leq C_T \|\mathbf{i}_c\| \|\mathbf{T}(\tilde{\mathbf{z}})\|_{1,\Omega} \|\mathbf{z} - \tilde{\mathbf{z}}\|_{1,\Omega} \quad \forall \mathbf{z}, \tilde{\mathbf{z}} \in W_\varepsilon.$$

Next, due to the definitions of the constants  $\varepsilon_0$  (cf. [3.18]) and  $C_T$  (cf. Lemma 3.9), we obtain

$$\|\mathbf{T}(\mathbf{z}) - \mathbf{T}(\tilde{\mathbf{z}})\|_{1,\Omega} \leq \frac{2(\kappa_1^2 + 1)^{1/2} \|\mathbf{i}_c\|^2}{\alpha(\Omega)} \|\mathbf{T}(\tilde{\mathbf{z}})\|_{1,\Omega} \|\mathbf{z} - \tilde{\mathbf{z}}\|_{1,\Omega} = \frac{1}{\varepsilon_0} \|\mathbf{T}(\tilde{\mathbf{z}})\|_{1,\Omega} \|\mathbf{z} - \tilde{\mathbf{z}}\|_{1,\Omega},$$

which, according to (3.12) and our assumption (3.24), yields

$$\|\mathbf{T}(z) - \mathbf{T}(\tilde{z})\|_{1,\Omega} \leq \frac{1}{\varepsilon_0} c_T \{ \|f\|_{0,\Omega} + \|g\|_{0,\Gamma} + \|g\|_{1/2,\Gamma} \} \|z - \tilde{z}\|_{1,\Omega} \leq \frac{\varepsilon}{\varepsilon_0} \|z - \tilde{z}\|_{1,\Omega}$$

for all  $z, \tilde{z} \in W_\varepsilon$ . This inequality proves that actually, under the hypothesis (3.24), the operator  $\mathbf{T} : W_\varepsilon \rightarrow W_\varepsilon$  becomes a contraction, and hence it has a unique fixed point. ■

#### IV. THE GALERKIN SCHEME

In this section, we introduce and study the Galerkin scheme of the augmented mixed formulation (3.5). We analyze its solvability by employing a discrete version of the fixed-point strategy developed in Section D. Finally, we derive the corresponding Céa estimate of our Galerkin scheme. We begin by introducing arbitrary finite dimensional subspaces  $\mathbb{H}'_h, \mathbb{H}^\sigma_h, \mathbf{H}^u_h$ , and  $\mathbb{H}^\rho_h$  of the continuous spaces  $\mathbb{L}^2_{ir}(\Omega), \mathbb{H}_0(\mathbf{div}; \Omega), \mathbf{H}^1(\Omega)$  and  $\mathbb{L}^2_{skew}(\Omega)$ , respectively. As usual,  $h$  denotes the size of a regular triangulation  $\mathcal{T}_h$  of  $\bar{\Omega}$  made up of triangles  $T$  (when  $n=2$ ) or tetrahedra  $T$  (when  $n=3$ ) of diameter  $h_T$ , that is  $h := \max \{h_T : T \in \mathcal{T}_h\}$ . Then, the Galerkin scheme associated with our problem (3.5) reads: Find  $\vec{t}_h := (t_h, \sigma_h, u_h, \rho_h) \in \mathbf{H}_h := \mathbb{H}'_h \times \mathbb{H}^\sigma_h \times \mathbf{H}^u_h \times \mathbb{H}^\rho_h$  such that

$$[(\mathbf{A} + \mathbf{B}_{u_h})(\vec{t}_h), \vec{s}_h] = [\mathbf{F}, \vec{s}_h] \quad \forall \vec{s}_h := (s_h, \tau_h, v_h, \eta_h) \in \mathbf{H}_h. \tag{4.1}$$

Next, analogously to the continuous case, we introduce the discrete version of  $\mathbf{T}$ :

$$\mathbf{T}_h : \mathbf{H}^u_h \rightarrow \mathbf{H}^u_h \quad \text{by} \quad \mathbf{T}_h(z_h) := u_h \quad \forall z_h \in \mathbf{H}^u_h,$$

where  $u_h$  is the third component of the unique solution (to be confirmed below) of the discrete nonlinear problem: Find  $\vec{t}_h := (t_h, \sigma_h, u_h, \rho_h) \in \mathbf{H}_h$  such that

$$[(\mathbf{A} + \mathbf{B}_{z_h})(\vec{t}_h), \vec{s}_h] = [\mathbf{F}, \vec{s}_h] \quad \forall \vec{s}_h := (s_h, \tau_h, v_h, \eta_h) \in \mathbf{H}_h. \tag{4.2}$$

Then, similarly as for the continuous case, we rewrite our Galerkin scheme (4.1) as the fixed-point equation: Find  $u_h \in \mathbf{H}^u_h$  such that

$$\mathbf{T}_h(u_h) = u_h.$$

We continue our analysis by observing, exactly as we did in [1], that the arguments used in the proof of Lemma 3.5 can also be applied to the present discrete setting. In particular, for each  $z_h \in \mathbf{H}^u_h$  the nonlinear operator  $\mathbf{A} + \mathbf{B}_{z_h} : \mathbf{H}_h \rightarrow \mathbf{H}'_h$  becomes Lipschitz continuous as well with constant  $L_A + \|\mathbf{i}_c\|^2(\kappa_1^2 + 1)^{1/2} \|z_h\|_{1,\Omega}$ . Moreover, under the same feasible ranges stipulated in Lemma 3.5 for the stabilization parameters and the given  $z_h \in \mathbf{H}^u_h$  (instead of  $z \in \mathbf{H}^1(\Omega)$ ), one finds that  $\mathbf{A} + \mathbf{B}_{z_h} : \mathbf{H}_h \rightarrow \mathbf{H}'_h$  becomes strongly monotone with the same constant  $\frac{\alpha(\Omega)}{2}$  provided in (3.17). Consequently, the classical result on the bijectivity of monotone operators given by [18, Theorem 3.3.23] implies now the following lemma.

**Lemma 4.1.** *Assume that  $\kappa_1 \in (0, \frac{2\delta\mu_1}{L\mu})$ ,  $\kappa_3 \in (0, 2\tilde{\delta}(\mu_1 - \frac{\kappa_1 L\mu}{2\delta}))$ ,  $\kappa_4 \in (0, 2\hat{\delta}\kappa_0\beta(\Omega))$ , and  $\kappa_2, \kappa_5 > 0$ , with  $\delta \in (0, \frac{2}{L\mu})$ ,  $\tilde{\delta}, \hat{\delta} \in (0, 2)$ , and  $\beta(\Omega) := \min \{ \kappa_3(1 - \frac{\tilde{\delta}}{2}), \kappa_5 \}$ . Then, for each  $\varepsilon \in (0, \varepsilon_0)$ , with  $\varepsilon_0$  given by (3.18), and for each  $z_h \in \mathbf{H}^u_h$  such that  $\|z_h\|_{1,\Omega} \leq \varepsilon$ , problem (4.2)*

has a unique solution  $\vec{\mathbf{t}}_h := (\mathbf{t}_h, \boldsymbol{\sigma}_h, \mathbf{u}_h, \boldsymbol{\rho}_h) \in \mathbf{H}_h$ . Moreover, with the same constant  $c_T > 0$  from Lemma 3.5, which is independent of  $\mathbf{z}_h$  and the data  $\mathbf{f}$  and  $\mathbf{g}$ , there holds

$$\|\mathbf{T}_h(\mathbf{z}_h)\|_{1,\Omega} = \|\mathbf{u}_h\|_{1,\Omega} \leq \|\vec{\mathbf{t}}_h\| \leq c_T \{ \|\mathbf{f}\|_{0,\Omega} + \|\mathbf{g}\|_{0,\Gamma} + \|\mathbf{g}\|_{1/2,\Gamma} \}. \tag{4.3}$$

Now, analogously to the continuous case, we are able to derive the following main result concerning the Galerkin scheme (4.1).

**Theorem 4.2.** *Suppose that the parameters  $\kappa_i, i \in \{1, \dots, 5\}$ , satisfy the conditions required by Lemma 4.1. In addition, given  $\varepsilon \in (0, \varepsilon_0)$ , with  $\varepsilon_0$  defined by (3.18), we let  $W_\varepsilon^h := \{\mathbf{z}_h \in \mathbf{H}_h^a : \|\mathbf{z}_h\|_{1,\Omega} \leq \varepsilon\}$ , and assume that the data satisfy (3.24) (cf. Lemma 3.8), that is*

$$c_T \{ \|\mathbf{f}\|_{0,\Omega} + \|\mathbf{g}\|_{0,\Gamma} + \|\mathbf{g}\|_{1/2,\Gamma} \} \leq \varepsilon, \tag{4.4}$$

with  $c_T$  given at the end of the proof of Lemma 3.5. Then, (4.1) has a unique solution  $\vec{\mathbf{t}}_h := (\mathbf{t}_h, \boldsymbol{\sigma}_h, \mathbf{u}_h, \boldsymbol{\rho}_h) \in \mathbf{H}_h$  with  $\mathbf{u}_h \in W_\varepsilon^h$ , and there holds

$$\|\vec{\mathbf{t}}_h\| \leq c_T \{ \|\mathbf{f}\|_{0,\Omega} + \|\mathbf{g}\|_{0,\Gamma} + \|\mathbf{g}\|_{1/2,\Gamma} \}. \tag{4.5}$$

**Proof.** We first observe, thanks to (4.3), that the assumption (4.4) guarantees that  $\mathbf{T}_h(W_\varepsilon^h) \subseteq W_\varepsilon^h$ . Then, using exactly the same arguments utilized in the proof of Theorem 3.10, we deduce that  $\mathbf{T}_h : W_\varepsilon^h \rightarrow W_\varepsilon^h$  is also a contraction. Hence, applying the Banach fixed point Theorem we obtain that there exists a unique fixed point for  $\mathbf{T}_h$ , or equivalently, there exists a unique solution to (4.1). In turn, the a priori estimate (4.5) follows directly from (4.3). ■

Next, we establish the corresponding Céa estimate for our Galerkin scheme (4.1). We remark in advance that, differently from [1, Lemma 4.3, Theorem 4.4], where a Strang-type lemma was used for its proof, we now follow a more straightforward approach, which is based on a suitable decomposition of the error and the Galerkin orthogonality condition. This alternative argumentation can be utilized for the error analysis of other nonlinear problems as well (see, e.g., [2]).

The announced result is stated as follows.

**Theorem 4.3.** *Let  $\vec{\mathbf{t}} \in \mathbf{H}$  and  $\vec{\mathbf{t}}_h \in \mathbf{H}_h$  be the unique solutions of the continuous and discrete problems (3.5) and (4.1), respectively, and let  $\text{dist}(\vec{\mathbf{t}}, \mathbf{H}_h)$  be the distance of  $\vec{\mathbf{t}}$  to  $\mathbf{H}_h$ , that is*

$$\text{dist}(\vec{\mathbf{t}}, \mathbf{H}_h) := \inf_{\vec{\mathbf{r}}_h \in \mathbf{H}_h} \|\vec{\mathbf{t}} - \vec{\mathbf{r}}_h\|_{\mathbf{H}}.$$

In addition, given  $\varepsilon \in (0, \varepsilon_0)$ , with  $\varepsilon_0$  defined by (3.18), we assume that the data  $\mathbf{f}$  and  $\mathbf{g}$  satisfy

$$c_T \{ \|\mathbf{f}\|_{0,\Omega} + \|\mathbf{g}\|_{0,\Gamma} + \|\mathbf{g}\|_{1/2,\Gamma} \} \leq \frac{\varepsilon}{2}, \tag{4.6}$$

with  $c_T$  given at the end of the proof of Lemma 3.5. Then, there exists a positive constant  $C$ , depending only on  $L_A$  and  $\alpha(\Omega)$ , such that

$$\|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| \leq C \text{dist}(\vec{\mathbf{t}}, \mathbf{H}_h). \tag{4.7}$$

**Proof.** To simplify the subsequent analysis, we define  $\mathbf{e}_\vec{\mathbf{t}} := \vec{\mathbf{t}} - \vec{\mathbf{t}}_h$ . As usual, given any  $\vec{\mathbf{r}}_h := (\mathbf{r}_h, \boldsymbol{\zeta}_h, \mathbf{w}_h, \boldsymbol{\xi}_h) \in \mathbf{H}_h$ , we decompose this error as

$$\mathbf{e}_\vec{\mathbf{t}} = \boldsymbol{\xi}_\vec{\mathbf{t}} + \boldsymbol{\chi}_\vec{\mathbf{t}} = (\vec{\mathbf{t}} - \vec{\mathbf{r}}_h) + (\vec{\mathbf{r}}_h - \vec{\mathbf{t}}_h). \tag{4.8}$$

First, from (3.5) and (4.1), we easily get the Galerkin orthogonality property

$$[\mathbf{A}(\vec{\mathbf{t}}) - \mathbf{A}(\vec{\mathbf{t}}_h), \vec{\mathbf{s}}_h] + [\mathbf{B}_u(\vec{\mathbf{t}}) - \mathbf{B}_{u_h}(\vec{\mathbf{t}}_h), \vec{\mathbf{s}}_h] = 0 \quad \forall \vec{\mathbf{s}}_h \in \mathbf{H}_h,$$

and adding and subtracting  $\mathbf{B}_u(\vec{\mathbf{t}}_h)$  and  $\mathbf{A}(\vec{\mathbf{r}}_h)$ , we obtain

$$[\mathbf{A}(\vec{\mathbf{r}}_h) - \mathbf{A}(\vec{\mathbf{t}}_h), \vec{\mathbf{s}}_h] + [\mathbf{B}_u(\vec{\mathbf{t}} - \vec{\mathbf{t}}_h), \vec{\mathbf{s}}_h] = -[\mathbf{A}(\vec{\mathbf{t}}) - \mathbf{A}(\vec{\mathbf{r}}_h), \vec{\mathbf{s}}_h] - [\mathbf{B}_{u-u_h}(\vec{\mathbf{t}}_h), \vec{\mathbf{s}}_h].$$

Hence, proceeding similarly with  $\mathbf{B}_{z_h}(\vec{\mathbf{t}}_h)$  on the right-hand side, and using the decomposition (4.8), we find that

$$\begin{aligned} [\mathbf{A}(\vec{\mathbf{r}}_h) - \mathbf{A}(\vec{\mathbf{t}}_h), \vec{\mathbf{s}}_h] + [\mathbf{B}_u(\chi_{\vec{t}}), \vec{\mathbf{s}}_h] &= -[\mathbf{A}(\vec{\mathbf{t}}) - \mathbf{A}(\vec{\mathbf{r}}_h), \vec{\mathbf{s}}_h] - [\mathbf{B}_u(\xi_{\vec{t}}), \vec{\mathbf{s}}_h] \\ &\quad - [\mathbf{B}_{u-z_h}(\vec{\mathbf{t}}_h), \vec{\mathbf{s}}_h] - [\mathbf{B}_{z_h-u_h}(\vec{\mathbf{t}}_h), \vec{\mathbf{s}}_h]. \end{aligned}$$

In particular, for  $\vec{\mathbf{s}}_h = \chi_{\vec{t}}$ , using that  $\mathbf{u} \in W_\varepsilon$ , and applying the strong monotonicity of the form on the left-hand side, and the Lipschitz continuity of  $\mathbf{A}$  and  $\mathbf{B}$  on the right-hand side, we deduce that

$$\begin{aligned} \frac{\alpha(\Omega)}{2} \|\chi_{\vec{t}}\|^2 &\leq L_A \|\xi_{\vec{t}}\| \|\chi_{\vec{t}}\| + \|\mathbf{i}_c\|^2 (\kappa_1^2 + 1)^{1/2} \|\mathbf{u}\|_{1,\Omega} \|\xi_{\vec{t}}\| \|\chi_{\vec{t}}\| \\ &\quad + \|\mathbf{i}_c\|^2 (\kappa_1^2 + 1)^{1/2} \|\mathbf{u} - \mathbf{z}_h\|_{1,\Omega} \|\vec{\mathbf{t}}_h\| \|\chi_{\vec{t}}\| \\ &\quad + \|\mathbf{i}_c\|^2 (\kappa_1^2 + 1)^{1/2} \|\mathbf{z}_h - \mathbf{u}_h\|_{1,\Omega} \|\vec{\mathbf{t}}_h\| \|\chi_{\vec{t}}\|, \end{aligned}$$

so that, using that  $\|\mathbf{z}_h - \mathbf{u}_h\|_{1,\Omega} \leq \|\chi_{\vec{t}}\|$  and  $\|\mathbf{u} - \mathbf{z}_h\|_{1,\Omega} \leq \|\xi_{\vec{t}}\|$ , we arrive at

$$\begin{aligned} \frac{\alpha(\Omega)}{2} \|\chi_{\vec{t}}\|^2 &\leq \left( L_A + \|\mathbf{i}_c\|^2 (\kappa_1^2 + 1)^{1/2} \|\mathbf{u}\|_{1,\Omega} + \|\mathbf{i}_c\|^2 (\kappa_1^2 + 1)^{1/2} \|\vec{\mathbf{t}}_h\| \right) \|\xi_{\vec{t}}\| \|\chi_{\vec{t}}\| \\ &\quad + \|\mathbf{i}_c\|^2 (\kappa_1^2 + 1)^{1/2} \|\vec{\mathbf{t}}_h\| \|\chi_{\vec{t}}\|^2. \end{aligned}$$

But, as  $\|\vec{\mathbf{t}}_h\| \leq c_T \{ \|\mathbf{f}\|_{0,\Omega} + \|\mathbf{g}\|_{0,\Gamma} + \|\mathbf{g}\|_{1/2,\Gamma} \} \leq \varepsilon/2$  and  $\|\mathbf{u}\|_{1,\Omega} \leq \varepsilon/2$ , we conclude that

$$\left( \frac{\alpha(\Omega)}{2} - \frac{\varepsilon}{2} \|\mathbf{i}_c\|^2 (\kappa_1^2 + 1)^{1/2} \right) \|\chi_{\vec{t}}\| \leq \left( L_A + \varepsilon \|\mathbf{i}_c\|^2 (\kappa_1^2 + 1)^{1/2} \right) \|\xi_{\vec{t}}\|,$$

which, together with the fact that  $\varepsilon \in (0, \varepsilon_0)$ , with  $\varepsilon_0$  defined in (3.18), and the triangle inequality, finish the proof. ■

Now, with the help of the previous theorem, we estimate the error for the postprocessed pressure. In fact, according to the Eqs. (2.7), and (3.2), we define our discrete approximation of the pressure as

$$p_h := -\frac{1}{n} \text{tr} \{ \boldsymbol{\sigma}_h + c_h \mathbb{I} + (\mathbf{u}_h \otimes \mathbf{u}_h) \} \quad \text{in } \Omega, \quad \text{with } c_h := -\frac{1}{n|\Omega|} \int_{\Omega} \text{tr}(\mathbf{u}_h \otimes \mathbf{u}_h),$$

which yields

$$p - p_h = \frac{1}{n} \text{tr} \{ (\boldsymbol{\sigma}_h - \boldsymbol{\sigma}) + (\mathbf{u}_h \otimes \mathbf{u}_h - \mathbf{u} \otimes \mathbf{u}) \} + (c_h - c),$$

and thus, applying the Cauchy-Schwarz inequality, we first deduce that

$$\|p - p_h\|_{0,\Omega} \leq \hat{C} \{ \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{0,\Omega} + \|\mathbf{u}_h \otimes \mathbf{u}_h - \mathbf{u} \otimes \mathbf{u}\|_{0,\Omega} + |c - c_h| \},$$

where  $\hat{C} > 0$  depends on  $n$  and  $|\Omega|$ . Next, bearing in mind the expression for  $c$  given by (3.2), decomposing

$$\mathbf{u}_h \otimes \mathbf{u}_h - \mathbf{u} \otimes \mathbf{u} = (\mathbf{u}_h - \mathbf{u}) \otimes \mathbf{u}_h + \mathbf{u} \otimes (\mathbf{u}_h - \mathbf{u}),$$

and using the triangle and Hölder inequalities, the compact embedding  $\mathbf{i}_c : \mathbf{H}^1(\Omega) \rightarrow \mathbf{L}^4(\Omega)$ , and the a priori bounds for  $\|\mathbf{u}\|_{1,\Omega}$  and  $\|\mathbf{u}_h\|_{1,\Omega}$  (cf. (3.26) in Theorem 3.10 and (4.5) in Theorem 4.2), we conclude from the foregoing equations that there exists a constant  $C > 0$ , independent of  $h$ , but depending on  $n$ ,  $|\Omega|$ ,  $\|\mathbf{i}_c\|$ , and the data  $\mathbf{f}$  and  $\mathbf{g}$ , such that

$$\|p - p_h\|_{0,\Omega} \leq C \{ \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\text{div};\Omega} + \|\mathbf{u} - \mathbf{u}_h\|_{1,\Omega} \}. \tag{4.9}$$

We now specify a concrete example of finite element subspaces for our Galerkin scheme (4.1). In what follows, given an integer  $k \geq 0$  and a set  $S \subseteq \mathbf{R} := \mathbf{R}^n$ ,  $\mathbf{P}_k(S)$  denotes the space of polynomial functions on  $S$  of degree  $\leq k$ . In addition, according to the notation described in Section 2.1, we set  $\mathbf{P}_k(S) := [\mathbf{P}_k(S)]^n$  and  $\mathbb{P}_k(S) := [\mathbf{P}_k(S)]^{n \times n}$ . Similarly,  $\mathbf{C}(\overline{\Omega}) = [\mathbf{C}(\overline{\Omega})]^n$  and  $\mathbb{C}(S) := [\mathbf{C}(S)]^{n \times n}$ . We start defining the corresponding local Raviart-Thomas spaces of order  $k$  as

$$\mathbf{RT}_k(T) := \mathbf{P}_k(T) \oplus \mathbf{P}_k(T)\mathbf{x} \quad \forall T \in \mathcal{T}_h,$$

where  $\mathbf{x}$  is a generic vector in  $\mathbf{R}^n$ . Then, we introduce examples of specific finite element subspaces  $\mathbb{H}_h^t, \mathbb{H}_h^\sigma, \mathbf{H}_h^\eta$ , and  $\mathbb{H}_h^\rho$  approximating the unknowns  $\mathbf{t}$ ,  $\boldsymbol{\sigma}$ ,  $\mathbf{u}$ , and  $\boldsymbol{\rho}$  as follows:

$$\mathbb{H}_h^t := \{ \mathbf{s}_h \in \mathbb{L}_{tr}^2(\Omega) : \mathbf{s}_h|_T \in \mathbb{P}_k(T) \quad \forall T \in \mathcal{T}_h \}, \tag{4.10}$$

$$\mathbb{H}_h^\sigma := \{ \boldsymbol{\tau}_h \in \mathbb{H}_0(\text{div}; \Omega) : \mathbf{c}^t \boldsymbol{\tau}|_T \in \mathbf{RT}_k(T), \quad \forall \mathbf{c} \in \mathbf{R}^n \quad \forall T \in \mathcal{T}_h \}, \tag{4.11}$$

$$\mathbf{H}_h^\eta := \{ \mathbf{v}_h \in \mathbf{C}(\overline{\Omega}) : \mathbf{v}_h|_T \in \mathbf{P}_{k+1}(T) \quad \forall T \in \mathcal{T}_h \}, \tag{4.12}$$

$$\mathbb{H}_h^\rho := \{ \boldsymbol{\eta}_h \in \mathbb{L}_{skew}^2(\Omega) : \boldsymbol{\eta}_h|_T \in \mathbb{P}_k(T) \quad \forall T \in \mathcal{T}_h \}. \tag{4.13}$$

The approximation properties of the above finite element subspaces are as follows (cf. [12, 13]):

( $\mathbf{AP}_h^t$ ) There exists  $C > 0$ , independent of  $h$ , such that for each  $s \in (0, k + 1]$ , and for each  $\mathbf{r} \in \mathbb{H}^s(\Omega) \cap \mathbb{L}_{tr}^2(\Omega)$ , there holds

$$\text{dist}(\mathbf{r}, \mathbb{H}_h^t) := \inf_{\mathbf{r}_h \in \mathbb{H}_h^t} \|\mathbf{r} - \mathbf{r}_h\|_{0,\Omega} \leq Ch^s \|\mathbf{r}\|_{s,\Omega}.$$

( $\mathbf{AP}_h^\sigma$ ) There exists  $C > 0$ , independent of  $h$ , such that for each  $s \in (0, k + 1]$ , and for each  $\boldsymbol{\zeta} \in \mathbb{H}^s(\Omega) \cap \mathbb{H}_0(\text{div}; \Omega)$  with  $\text{div } \boldsymbol{\zeta} \in \mathbf{H}^s(\Omega)$ , their holds

$$\text{dist}(\boldsymbol{\zeta}, \mathbb{H}_h^\sigma) := \inf_{\boldsymbol{\zeta}_h \in \mathbb{H}_h^\sigma} \|\boldsymbol{\zeta} - \boldsymbol{\zeta}_h\|_{\text{div};\Omega} \leq Ch^s \{ \|\boldsymbol{\zeta}\|_{s,\Omega} + \|\text{div } \boldsymbol{\zeta}\|_{s,\Omega} \}.$$

( $\mathbf{AP}_h^u$ ) There exists  $C > 0$ , independent of  $h$ , such that for each  $s \in (0, k + 1]$ , and for each  $\mathbf{w} \in \mathbf{H}^{s+1}(\Omega)$ , there holds

$$\text{dist}(\mathbf{w}, \mathbf{H}_h^u) := \inf_{\mathbf{w}_h \in \mathbf{H}_h^u} \|\mathbf{w} - \mathbf{w}_h\|_{1,\Omega} \leq Ch^s \|\mathbf{w}\|_{s+1,\Omega}.$$

( $\mathbf{AP}_h^\rho$ ) There exists  $C > 0$ , independent of  $h$ , such that for each  $s \in (0, k + 1]$ , and for each  $\boldsymbol{\eta} \in \mathbb{H}^s(\Omega) \cap \mathbb{L}_{skew}^2(\Omega)$ , there holds

$$\text{dist}(\boldsymbol{\eta}, \mathbb{H}_h^\rho) := \inf_{\boldsymbol{\eta}_h \in \mathbb{H}_h^\rho} \|\boldsymbol{\eta} - \boldsymbol{\eta}_h\|_{0,\Omega} \leq Ch^s \|\boldsymbol{\eta}\|_{s,\Omega}.$$

In consequence, we can establish the convergence of the Galerkin scheme (4.1) associated to the spaces specified in (4.10)–(4.13). We notice here that the main assumption (3.24) on the data guaranteeing the well-posedness of the continuous and discrete problems follows from (4.6), and hence it suffices to assume the latter only.

**Theorem 4.4.** *Besides the hypotheses of Lemma 4.1 (or Lemma 3.5) and Theorem 4.3, assume that there exists  $s > 0$  such that  $\mathbf{t} \in \mathbb{H}^s(\Omega)$ ,  $\boldsymbol{\sigma} \in \mathbb{H}^s(\Omega)$ ,  $\mathbf{div} \boldsymbol{\sigma} \in \mathbf{H}^s(\Omega)$ ,  $\mathbf{u} \in \mathbf{H}^{s+1}(\Omega)$ , and  $\boldsymbol{\rho} \in \mathbb{H}^s(\Omega)$ , and that the finite element subspaces are defined by (4.10)–(4.13). Then, there exists  $C > 0$ , independent of  $h$ , such that for each  $h > 0$  there holds*

$$\begin{aligned} & \|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| + \|p - p_h\|_{0,\Omega} \\ & \leq Ch^{\min\{s,k+1\}} \left\{ \|\mathbf{t}\|_{s,\Omega} + \|\boldsymbol{\sigma}\|_{s,\Omega} + \|\mathbf{div} \boldsymbol{\sigma}\|_{s,\Omega} + \|\mathbf{u}\|_{s+1,\Omega} + \|\boldsymbol{\rho}\|_{s,\Omega} \right\}. \end{aligned} \quad (4.14)$$

**Proof.** It follows from the Céa estimate (4.7), the upper bound given by (4.9), and the approximation properties ( $\mathbf{AP}_h^t$ ), ( $\mathbf{AP}_h^\sigma$ ), ( $\mathbf{AP}_h^u$ ) and ( $\mathbf{AP}_h^\rho$ ). ■

Before moving to the a posteriori error analysis of our Galerkin scheme, we highlight that, while the gradient of the velocity is not directly approximated here as it was in [1], we can still exploit the present auxiliary unknowns  $\mathbf{t}$  and  $\boldsymbol{\rho}$ , as they actually suggest a fairly simple way of obtaining a discrete approximation of  $\nabla \mathbf{u}$ . In fact, instead of performing a numerical differentiation of  $\mathbf{u}_h$ , which usually yields a loss of accuracy, we follow the first equation of (2.6) and define

$$(\nabla \mathbf{u})_h := \mathbf{t}_h + \boldsymbol{\rho}_h.$$

In this way, we readily obtain

$$\|\nabla \mathbf{u} - (\nabla \mathbf{u})_h\|_{0,\Omega} \leq \|\mathbf{t} - \mathbf{t}_h\|_{0,\Omega} + \|\boldsymbol{\rho} - \boldsymbol{\rho}_h\|_{0,\Omega} \leq \|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\|,$$

which, together with (4.14), certainly gives

$$\begin{aligned} & \|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| + \|p - p_h\|_{0,\Omega} + \|\nabla \mathbf{u} - (\nabla \mathbf{u})_h\|_{0,\Omega} \\ & \leq Ch^{\min\{s,k+1\}} \left\{ \|\mathbf{t}\|_{s,\Omega} + \|\boldsymbol{\sigma}\|_{s,\Omega} + \|\mathbf{div} \boldsymbol{\sigma}\|_{s,\Omega} + \|\mathbf{u}\|_{s+1,\Omega} + \|\boldsymbol{\rho}\|_{s,\Omega} \right\}. \end{aligned}$$

V. A POSTERIORI ERROR ANALYSIS

In this section, we derive a reliable and efficient residual-based a posteriori error estimate for (4.1), with the discrete spaces introduced in Section IV for  $n = 3$ . After that, we introduce our approach in the two-dimensional case and point out the differences between the estimator obtained for  $n = 3$  and  $n = 2$ . We first recall that the **curl** of a three-dimensional vector field  $\mathbf{v} := (v_1, v_2, v_3)$  is the vector

$$\mathbf{curl} \mathbf{v} = \nabla \times \mathbf{v} := \left( \frac{\partial v_3}{\partial x_2} - \frac{\partial v_2}{\partial x_3}, \frac{\partial v_1}{\partial x_3} - \frac{\partial v_3}{\partial x_1}, \frac{\partial v_2}{\partial x_1} - \frac{\partial v_1}{\partial x_2} \right).$$

Then, given a tensor function  $\boldsymbol{\tau} := (\tau_{ij})_{3 \times 3}$ , the operator **curl** denotes the operator **curl** acting along each row of  $\boldsymbol{\tau}$ , that is,  $\boldsymbol{\tau}$  is the  $3 \times 3$  tensor whose rows are given by

$$\mathbf{curl} \boldsymbol{\tau} := \begin{pmatrix} \mathbf{curl}(\tau_{11}, \tau_{12}, \tau_{13}) \\ \mathbf{curl}(\tau_{21}, \tau_{22}, \tau_{23}) \\ \mathbf{curl}(\tau_{31}, \tau_{32}, \tau_{33}) \end{pmatrix}.$$

Having defined **curl**, we now introduce the Sobolev space

$$\mathbb{H}(\mathbf{curl}; \Omega) := \{ \mathbf{w} \in \mathbb{L}^2(\Omega) : \mathbf{curl} \mathbf{w} \in \mathbb{L}^2(\Omega) \}.$$

In addition, we denote by  $\boldsymbol{\tau} \times \mathbf{v}$  the  $3 \times 3$  tensor whose rows are given by the tangential trace of each row of  $\boldsymbol{\tau}$ , that is,

$$\boldsymbol{\tau} \times \mathbf{v} := \begin{pmatrix} (\tau_{11}, \tau_{12}, \tau_{13}) \times \mathbf{v} \\ (\tau_{21}, \tau_{22}, \tau_{23}) \times \mathbf{v} \\ (\tau_{31}, \tau_{32}, \tau_{33}) \times \mathbf{v} \end{pmatrix}.$$

We denote by  $\mathcal{E}_h$  the set of faces  $e$  of  $\mathcal{T}_h$ , subdivided into interior and exterior faces  $\mathcal{E}_h = \mathcal{E}_h(\Omega) \cup \mathcal{E}_h(\Gamma)$ , with  $\mathcal{E}_h(\Omega) := \{e \in \mathcal{E}_h : e \subseteq \Omega\}$  and  $\mathcal{E}_h(\Gamma) := \{e \in \mathcal{E}_h : e \subseteq \Gamma\}$ . In turn, for each  $T \in \mathcal{T}_h$ , we let  $\mathcal{E}(T)$  denote the set of faces of  $T$ . Here,  $h_e$  stands for the diameter of a given  $e \in \mathcal{E}_h$ , and for each  $e \in \mathcal{E}_h$  we fix a unit normal  $\mathbf{v}_e$  to  $e$ . Then, given  $\boldsymbol{\tau} \in \mathbb{H}(\mathbf{curl}; \Omega)$  and  $e \in \mathcal{E}_h(\Omega)$ , we let  $[[\boldsymbol{\tau} \times \mathbf{v}_e]] := (\boldsymbol{\tau}|_{T'} - \boldsymbol{\tau}|_{T''})|_e \times \mathbf{v}_e$ , where  $T'$  and  $T''$  are elements of  $\mathcal{T}_h$  sharing the common face  $e$ . If no confusion arises, we will simply write  $\mathbf{v}$  instead of  $\mathbf{v}_e$ .

Now, let  $\vec{\mathbf{t}} := (\mathbf{t}, \boldsymbol{\sigma}, \mathbf{u}, \boldsymbol{\rho}) \in \mathbf{H}$  and  $\vec{\mathbf{t}}_h := (\mathbf{t}_h, \boldsymbol{\sigma}_h, \mathbf{u}_h, \boldsymbol{\rho}_h) \in \mathbf{H}_h$  be the unique solutions of the continuous and discrete problems (3.5) and (4.1), respectively. Then, we introduce the global a posteriori error estimator

$$\Theta := \left\{ \sum_{T \in \mathcal{T}_h} \Theta_T^2 \right\}^{1/2}, \tag{5.1}$$

where for each  $T \in \mathcal{T}_h$  we set:

$$\begin{aligned} \Theta_T^2 := & h_T^2 \|\mathbf{curl}(\mathbf{t}_h + \boldsymbol{\rho}_h)\|_{0,T}^2 + \|\boldsymbol{\sigma}_h^d - \mu(|\mathbf{t}_h|)\mathbf{t}_h + (\mathbf{u}_h \otimes \mathbf{u}_h)^d\|_{0,T}^2 \\ & + \|\boldsymbol{\sigma}_h - \boldsymbol{\sigma}_h^t\|_{0,T}^2 + \|\mathbf{f} - \mathcal{P}_h(\mathbf{f})\|_{0,T}^2 + \|\mathbf{f} + \mathbf{div}(\boldsymbol{\sigma}_h)\|_{0,T}^2 \\ & + \|\mathbf{e}(\mathbf{u}_h) - \mathbf{t}_h\|_{0,T}^2 + \|\boldsymbol{\rho}_h - \nabla \mathbf{u}_h + \mathbf{e}(\mathbf{u}_h)\|_{0,T}^2 \end{aligned}$$

$$\begin{aligned}
 &+ \sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h(\Omega)} h_e |[(\mathbf{t}_h + \boldsymbol{\rho}_h) \times \mathbf{v}]|_{0,e}^2 + \sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h(\Gamma)} \|\mathbf{g} - \mathbf{u}_h\|_{0,e}^2 \\
 &+ \sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h(\Gamma)} h_e |(\nabla \mathbf{g} - \mathbf{t}_h - \boldsymbol{\rho}_h) \times \mathbf{v}|_{0,e}^2,
 \end{aligned} \tag{5.2}$$

where  $\mathcal{P}_h$  is the  $L^2(\Omega)$ -orthogonal projector onto  $\mathbb{H}_h^t$ . Note that the above definition requires that  $\nabla \mathbf{g} \times \mathbf{v}|_e \in \mathbb{L}^2(e)$  for each  $e \in \mathcal{E}_h(\Gamma)$ , which is fixed below by assuming that  $\mathbf{g} \in \mathbf{H}^1(\Gamma)$ .

The main result of this section is stated as follows.

**Theorem 5.1.** *Let  $\vec{\mathbf{t}} \in \mathbf{H}$  and  $\vec{\mathbf{t}}_h \in \mathbf{H}_h$  be the unique solutions of the continuous and discrete problems (3.5) and (4.1), respectively, and assume that  $\mathbf{g} \in \mathbf{H}^1(\Gamma)$ . Then, there exist  $C_{rel} > 0$  and  $C_{eff} > 0$ , independent of  $h$ , such that*

$$C_{eff} \Theta \leq \|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| \leq C_{rel} \Theta. \tag{5.3}$$

The efficiency of  $\Theta$  (lower bound in [5.3]) is proved in Section B, whereas the corresponding reliability estimate (upper bound in [5.3]) is proved next in Section A.

### A. Reliability of the a posteriori Error Estimator

To prove the reliability of our a posteriori error estimator, we follow the strategy proposed originally in [7], and then used in [5], which is based on a linearization technique that involves the Gâteaux derivatives of the nonlinear terms of the formulation. More precisely, proceeding similarly as in [5], we begin with the result to be introduced next. However, we notice in advance that due to the new meaning of the unknown  $\mathbf{t}$  and the presence of additional terms in the present augmented formulation, two of the resulting bounded functionals, whose norms need to be estimated later on, do not coincide with those in [5] (see  $\mathbf{R}_3$  and  $\mathbf{R}_4$  below).

**Lemma 5.2.** *Let  $\vec{\mathbf{t}} \in \mathbf{H}$  and  $\vec{\mathbf{t}}_h \in \mathbf{H}_h$  be the unique solutions of the continuous and discrete problems (3.5) and (4.1), respectively. In addition, given  $\varepsilon \in (0, \varepsilon_0)$ , with  $\varepsilon_0$  defined by (3.18) (cf. proof of Lemma 3.5), we assume that the data  $\mathbf{f}$  and  $\mathbf{g}$  satisfy*

$$c_T \{ \|\mathbf{f}\|_{0,\Omega} + \|\mathbf{g}\|_{0,\Gamma} + \|\mathbf{g}\|_{1/2,\Gamma} \} \leq \frac{\varepsilon}{2}, \tag{5.4}$$

with  $c_T$  given at the end of the proof of Lemma 3.5. Then, there exists a constant  $C > 0$ , independent of  $h$ , such that

$$\|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| \leq C \|\mathbf{R}\|,$$

where

$$\mathbf{R}(\vec{\mathbf{s}}) := \mathbf{R}_1(\mathbf{s}) + \mathbf{R}_2(\boldsymbol{\tau}) + \mathbf{R}_3(\mathbf{v}) + \mathbf{R}_4(\boldsymbol{\eta}) \quad \forall \vec{\mathbf{s}} := (\mathbf{s}, \boldsymbol{\tau}, \mathbf{v}, \boldsymbol{\eta}) \in \mathbf{H},$$

and  $\mathbf{R}_1(\mathbf{s})$ ,  $\mathbf{R}_2(\boldsymbol{\tau})$ ,  $\mathbf{R}_3(\mathbf{v})$ ,  $\mathbf{R}_4(\boldsymbol{\eta})$  are defined by

$$\begin{aligned}
 \mathbf{R}_1(\mathbf{s}) &:= \int_{\Omega} \{ \sigma_h^d - \mu(|\mathbf{t}_h|) \mathbf{t}_h + (\mathbf{u}_h \otimes \mathbf{u}_h)^d \} : \mathbf{s}, \\
 \mathbf{R}_2(\boldsymbol{\tau}) &:= \langle \boldsymbol{\tau} \mathbf{v}, \mathbf{g} \rangle - \int_{\Omega} \mathbf{u}_h \cdot \mathbf{div} \boldsymbol{\tau} - \int_{\Omega} \mathbf{t}_h : \boldsymbol{\tau}^d - \kappa_2 \int_{\Omega} \{ \mathbf{f} + \mathbf{div} \sigma_h \} \cdot \mathbf{div} \boldsymbol{\tau}
 \end{aligned}$$

$$\begin{aligned}
 & - \int_{\Omega} \boldsymbol{\rho}_h : \boldsymbol{\tau} - \kappa_1 \int_{\Omega} \{ \boldsymbol{\sigma}_h^d - \mu(|\mathbf{t}_h|)\mathbf{t}_h + (\mathbf{u}_h \otimes \mathbf{u}_h)^d \} : \boldsymbol{\tau}^d, \\
 \mathbf{R}_3(\mathbf{v}) & := \int_{\Omega} \{ \mathbf{f} + \mathbf{div} \boldsymbol{\sigma}_h \} \cdot \mathbf{v} - \kappa_3 \int_{\Omega} \{ \mathbf{e}(\mathbf{u}_h) - \mathbf{t}_h \} : \mathbf{e}(\mathbf{v}) + \kappa_5 \int_{\Gamma} \{ \mathbf{g} - \mathbf{u}_h \} \cdot \mathbf{v}, \\
 \mathbf{R}_4(\boldsymbol{\eta}) & := \frac{1}{2} \int_{\Omega} \{ \boldsymbol{\sigma}_h - \boldsymbol{\sigma}_h^t \} : \boldsymbol{\eta} - \kappa_4 \int_{\Omega} (\boldsymbol{\rho}_h - \{ \nabla \mathbf{u}_h - \mathbf{e}(\mathbf{u}_h) \}) : \boldsymbol{\eta}.
 \end{aligned} \tag{5.5}$$

Furthermore, there holds

$$\mathbf{R}(\vec{\mathbf{s}}_h) = 0 \quad \forall \vec{\mathbf{s}}_h \in \mathbf{H}_h. \tag{5.6}$$

**Proof.** First, we note that the operator  $\mathbf{A}$  (cf. (3.6)) can be split into linear and nonlinear terms:

$$[\mathbf{A}(\vec{\mathbf{t}}), \vec{\mathbf{s}}] = [\mathbf{A}_1(\mathbf{t}), \mathbf{s}] - \kappa_1 [\mathbf{A}_1(\mathbf{t}), \boldsymbol{\tau}^d] + [\mathbf{A}_2(\vec{\mathbf{t}}), \vec{\mathbf{s}}],$$

where

$$\begin{aligned}
 [\mathbf{A}_1(\mathbf{t}), \mathbf{s}] & := \int_{\Omega} \mu(|\mathbf{t}|)\mathbf{t} : \mathbf{s}, \\
 [\mathbf{A}_2(\vec{\mathbf{t}}), \vec{\mathbf{s}}] & := - \int_{\Omega} \boldsymbol{\sigma}^d : \mathbf{s} + \int_{\Omega} \boldsymbol{\tau}^d : \mathbf{t} + \int_{\Omega} \mathbf{u} \cdot \mathbf{div} \boldsymbol{\tau} - \int_{\Omega} \mathbf{v} \cdot \mathbf{div} \boldsymbol{\sigma} \\
 & + \int_{\Omega} \boldsymbol{\rho} : \boldsymbol{\tau} - \int_{\Omega} \boldsymbol{\eta} : \boldsymbol{\sigma} + \kappa_1 \int_{\Omega} \boldsymbol{\sigma}^d : \boldsymbol{\tau}^d \\
 & + \kappa_2 \int_{\Omega} \mathbf{div} \boldsymbol{\sigma} \cdot \mathbf{div} \boldsymbol{\tau} + \kappa_3 \int_{\Omega} \{ \mathbf{e}(\mathbf{u}) - \mathbf{t} \} : \mathbf{e}(\mathbf{v}) \\
 & + \kappa_4 \int_{\Omega} (\boldsymbol{\rho} - \{ \nabla \mathbf{u} - \mathbf{e}(\mathbf{u}) \}) : \boldsymbol{\eta} + \kappa_5 \int_{\Gamma} \mathbf{u} \cdot \mathbf{v}.
 \end{aligned}$$

Next, since  $\mu$  is of class  $C^1$  and satisfies the assumptions (2.3), minor modifications of the proof of [7, Lemma 5.1] allow to show that the nonlinear operator  $\mathbf{A}_1$  is Gâteaux differentiable. This means that for each  $\mathbf{r} \in \mathbb{L}_{tr}^2(\Omega)$  there exists a bounded linear operator  $\mathcal{D}\mathbf{A}_1(\mathbf{r}) : \mathbb{L}_{tr}^2(\Omega) \rightarrow \mathbb{L}_{tr}^2(\Omega)'$  such that

$$\mathcal{D}\mathbf{A}_1(\mathbf{r})(\mathbf{t}) := \lim_{\epsilon \rightarrow 0} \frac{\mathbf{A}_1(\mathbf{r} + \epsilon \mathbf{t}) - \mathbf{A}_1(\mathbf{r})}{\epsilon} \quad \forall \mathbf{t} \in \mathbb{L}_{tr}^2(\Omega).$$

Notice that for each  $\mathbf{r} \in \mathbb{L}_{tr}^2(\Omega)$ ,  $\mathcal{D}\mathbf{A}_1(\mathbf{r})$  can be considered as a bilinear form satisfying

$$\mathcal{D}\mathbf{A}_1(\mathbf{r})(\mathbf{t}, \mathbf{s}) := \mathcal{D}\mathbf{A}_1(\mathbf{r})(\mathbf{t})(\mathbf{s}) \quad \forall \mathbf{t}, \mathbf{s} \in \mathbb{L}_{tr}^2(\Omega).$$

In addition, it is easy to prove, using (2.3), that  $\mathcal{D}\mathbf{A}_1(\mathbf{r})$  becomes both uniformly bounded and elliptic with constants  $L_{\mu}$  and  $\mu_1$ , respectively, that is

$$|\mathcal{D}\mathbf{A}_1(\mathbf{r})(\mathbf{t}, \mathbf{s})| \leq L_{\mu} \|\mathbf{t}\|_{0,\Omega} \|\mathbf{s}\|_{0,\Omega} \tag{5.7}$$

and

$$|\mathcal{D}\mathbf{A}_1(\mathbf{r})(\mathbf{s}, \mathbf{s})| \geq \mu_1 \|\mathbf{s}\|_{0,\Omega}^2, \tag{5.8}$$

for all  $\mathbf{r}, \mathbf{t}, \mathbf{s} \in \mathbb{L}_{rr}^2(\Omega)$  (see, for instance, [7, Lemma 5.1]). Then, given  $\mathbf{r} \in \mathbb{L}_{rr}^2(\Omega)$ , we introduce the linear operator

$$[\tilde{\mathbf{A}}(\vec{\mathbf{t}}), \vec{\mathbf{s}}] := \mathcal{DA}_1(\mathbf{r})(\mathbf{t}, \mathbf{s}) - \kappa_1 \mathcal{DA}_1(\mathbf{r})(\mathbf{t}, \boldsymbol{\tau}^d) + [\mathbf{A}_2(\vec{\mathbf{t}}), \vec{\mathbf{s}}] \quad \forall \vec{\mathbf{t}}, \vec{\mathbf{s}} \in \mathbf{H},$$

so that, using (5.7), (5.8), and the same arguments showing the strong monotonicity of  $[\mathbf{A}(\vec{\mathbf{t}}), (\vec{\mathbf{s}})]$ , we deduce the ellipticity of  $\tilde{\mathbf{A}}$

$$[\tilde{\mathbf{A}}(\vec{\mathbf{s}}), \vec{\mathbf{s}}] \geq \alpha(\Omega) \|\vec{\mathbf{s}}\|^2 \quad \forall \vec{\mathbf{s}} \in \mathbf{H}. \tag{5.9}$$

It readily follows that

$$[(\tilde{\mathbf{A}} + \mathbf{B}_w)(\vec{\mathbf{s}}), \vec{\mathbf{s}}] \geq \frac{\alpha(\Omega)}{2} \|\vec{\mathbf{s}}\|^2, \tag{5.10}$$

for all  $\vec{\mathbf{s}} \in \mathbf{H}$  and for all  $\mathbf{w} \in \mathbf{H}^1(\Omega)$  such that  $\|\mathbf{w}\|_{1,\Omega} \leq \varepsilon \in (0, \varepsilon_0)$ , with  $\varepsilon_0$  defined by (3.18), which yields the inf-sup condition

$$\frac{\alpha(\Omega)}{2} \|\vec{\mathbf{r}}\| \leq \sup_{\substack{\vec{\mathbf{s}} \in \mathbf{H} \\ \vec{\mathbf{s}} \neq \mathbf{0}}} \frac{[(\tilde{\mathbf{A}} + \mathbf{B}_w)(\vec{\mathbf{r}}), \vec{\mathbf{s}}]}{\|\vec{\mathbf{s}}\|} \quad \forall \vec{\mathbf{r}} \in \mathbf{H}.$$

Hence, taking in particular  $\mathbf{w} = \mathbf{u}$  and  $\vec{\mathbf{r}} = \vec{\mathbf{t}} - \vec{\mathbf{t}}_h$ , we deduce from the foregoing inequality that

$$\frac{\alpha(\Omega)}{2} \|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| \leq \sup_{\substack{\vec{\mathbf{s}} \in \mathbf{H} \\ \vec{\mathbf{s}} \neq \mathbf{0}}} \frac{[(\tilde{\mathbf{A}} + \mathbf{B}_u)(\vec{\mathbf{t}} - \vec{\mathbf{t}}_h), \vec{\mathbf{s}}]}{\|\vec{\mathbf{s}}\|}. \tag{5.11}$$

Conversely, using the Mean Value Theorem, we can assert that there exists a convex combination  $\mathbf{r}_h$  of  $\mathbf{t}$  and  $\mathbf{t}_h$  such that

$$\mathcal{DA}_1(\mathbf{r}_h)(\mathbf{t} - \mathbf{t}_h, \mathbf{s}) = [\mathbf{A}_1(\mathbf{t}), \mathbf{s}] - [\mathbf{A}_1(\mathbf{t}_h), \mathbf{s}] \quad \forall \mathbf{s} \in \mathbb{L}_{rr}^2(\Omega), \tag{5.12}$$

so that using now  $\mathbf{r}_h$  in the definition of  $\tilde{\mathbf{A}}$ , we can write

$$[\tilde{\mathbf{A}}(\vec{\mathbf{t}}), \vec{\mathbf{s}}] = \mathcal{DA}_1(\mathbf{r}_h)(\mathbf{t}, \mathbf{s}) - \kappa_1 \mathcal{DA}_1(\mathbf{r}_h)(\mathbf{t}, \boldsymbol{\tau}^d) + [\mathbf{A}_2(\vec{\mathbf{t}}), \vec{\mathbf{s}}] \quad \forall \vec{\mathbf{t}}, \vec{\mathbf{s}} \in \mathbf{H}.$$

The foregoing equality and (5.12) imply that

$$\begin{aligned} [(\tilde{\mathbf{A}} + \mathbf{B}_u)(\vec{\mathbf{t}} - \vec{\mathbf{t}}_h), \vec{\mathbf{s}}] &= [\tilde{\mathbf{A}}(\vec{\mathbf{t}}) - \tilde{\mathbf{A}}(\vec{\mathbf{t}}_h), \vec{\mathbf{s}}] + [\mathbf{B}_u(\vec{\mathbf{t}} - \vec{\mathbf{t}}_h), \vec{\mathbf{s}}] \\ &= [\mathbf{A}(\vec{\mathbf{t}}) - \mathbf{A}(\vec{\mathbf{t}}_h), \vec{\mathbf{s}}] + [\mathbf{B}_u(\vec{\mathbf{t}} - \vec{\mathbf{t}}_h), \vec{\mathbf{s}}] \\ &= [(\mathbf{A} + \mathbf{B}_u)(\vec{\mathbf{t}}), \vec{\mathbf{s}}] - [(\mathbf{A} + \mathbf{B}_u)(\vec{\mathbf{t}}_h), \vec{\mathbf{s}}] \\ &= [\mathbf{F}, \vec{\mathbf{s}}] - [(\mathbf{A} + \mathbf{B}_u)(\vec{\mathbf{t}}_h), \vec{\mathbf{s}}], \end{aligned}$$

and therefore, the inf-sup condition (5.11) becomes

$$\frac{\alpha(\Omega)}{2} \|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| \leq \sup_{\substack{\vec{\mathbf{s}} \in \mathbf{H} \\ \vec{\mathbf{s}} \neq \mathbf{0}}} \frac{[(\tilde{\mathbf{R}})(\vec{\mathbf{t}} - \vec{\mathbf{t}}_h), \vec{\mathbf{s}}]}{\|\vec{\mathbf{s}}\|},$$

where

$$[\tilde{\mathbf{R}}(\vec{\mathbf{t}} - \vec{\mathbf{t}}_h), \vec{\mathbf{s}}] := [\mathbf{F}, \vec{\mathbf{s}}] - [(\mathbf{A} + \mathbf{B}_{u_h})(\vec{\mathbf{t}}_h), \vec{\mathbf{s}}] + [\mathbf{B}_{u_h - u}(\vec{\mathbf{t}}_h), \vec{\mathbf{s}}].$$

Next, using that

$$|[\mathbf{B}_{u_h - u}(\vec{\mathbf{t}}_h), \vec{\mathbf{s}}]| \leq \|\mathbf{i}_c\|^2 (\kappa_1^2 + 1)^{1/2} \|\mathbf{u}_h - \mathbf{u}\|_{1,\Omega} \|\vec{\mathbf{t}}_h\| \|\vec{\mathbf{s}}\|,$$

and having in mind (4.5) and (5.4), we find that

$$\left( \frac{\alpha(\Omega)}{2} - \frac{\varepsilon \|\mathbf{i}_c\|^2 (\kappa_1^2 + 1)^{1/2}}{2} \right) \|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| \leq \sup_{\substack{\vec{\mathbf{s}} \in \mathbf{H} \\ \vec{\mathbf{s}} \neq \mathbf{0}}} \frac{[\mathbf{F}, \vec{\mathbf{s}}] - [(\mathbf{A} + \mathbf{B}_{u_h})(\vec{\mathbf{t}}_h), \vec{\mathbf{s}}]}{\|\vec{\mathbf{s}}\|}.$$

This estimate together with the fact that  $\varepsilon \in (0, \varepsilon_0)$ , with  $\varepsilon_0$  defined in (3.18), allow us to conclude that

$$\frac{\alpha(\Omega)}{4} \|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| \leq \sup_{\substack{\vec{\mathbf{s}} \in \mathbf{H} \\ \vec{\mathbf{s}} \neq \mathbf{0}}} \frac{[\mathbf{F}, \vec{\mathbf{s}}] - [(\mathbf{A} + \mathbf{B}_{u_h})(\vec{\mathbf{t}}_h), \vec{\mathbf{s}}]}{\|\vec{\mathbf{s}}\|} = \sup_{\substack{\vec{\mathbf{s}} \in \mathbf{H} \\ \vec{\mathbf{s}} \neq \mathbf{0}}} \frac{\mathbf{R}(\vec{\mathbf{s}})}{\|\vec{\mathbf{s}}\|}, \tag{5.13}$$

with

$$\mathbf{R}(\vec{\mathbf{s}}) := \mathbf{R}_1(\mathbf{s}) + \mathbf{R}_2(\boldsymbol{\tau}) + \mathbf{R}_3(\mathbf{v}) + \mathbf{R}_4(\boldsymbol{\eta}),$$

where  $\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3,$  and  $\mathbf{R}_4$  are defined in (5.5). Finally, the identity (5.6) is a straightforward consequence of (4.1). ■

We end this section by remarking that the supremum in (5.13) can be bounded in terms of  $\mathbf{R}_i, i = 1, \dots, 4$  as follows:

$$\frac{\alpha(\Omega)}{4} \|\vec{\mathbf{t}} - \vec{\mathbf{t}}_h\| \leq \{ \|\mathbf{R}_1\|_{\mathbb{L}^2_{Tr}(\Omega)'} + \|\mathbf{R}_2\|_{\mathbb{H}_0(\mathbf{div};\Omega)'} + \|\mathbf{R}_3\|_{\mathbf{H}^1(\Omega)'} + \|\mathbf{R}_4\|_{\mathbb{L}^2_{skew}(\Omega)'} \}, \tag{5.14}$$

and thus, the derivation of the upper bound in (5.3) is completed by providing suitable upper bounds for each one of the terms on the right-hand side of (5.14). To this respect, we first observe that direct applications of the Cauchy-Schwarz inequality give the corresponding estimates for the functionals  $\mathbf{R}_1, \mathbf{R}_3,$  and  $\mathbf{R}_4$ . Finally, the derivation of the upper bound for  $\|\mathbf{R}_2\|_{\mathbb{H}_0(\mathbf{div};\Omega)'}$  makes use of a stable Helmholtz decomposition for  $\mathbb{H}_0(\mathbf{div}; \Omega)$  which has been recently proved for  $n = 3$  in [19, Lemma 4.3] (see also [20, Theorem 3.1]), the Raviart–Thomas interpolation operator (see [12, 13]), the classical Clément interpolator ([21]), and the local approximation properties of them. This estimate follows basically from suitable modifications of the proofs of [5, Theorem 3.7] and [5, Lemmas 3.8 and 3.9]. In this regard, we just comment that within the process of bounding  $\|\mathbf{R}_2\|_{\mathbb{H}_0(\mathbf{div};\Omega)'}$  it also appears the local term  $h_T^2 \|\nabla \mathbf{u}_h - \mathbf{t}_h - \boldsymbol{\rho}_h\|_{0,T}^2$ , which being dominated by  $\|\mathbf{e}(\mathbf{u}_h) - \mathbf{t}_h\|_{0,T}^2 + \|\boldsymbol{\rho}_h - \nabla \mathbf{u}_h + \mathbf{e}(\mathbf{u}_h)\|_{0,T}^2$ , is then omitted from the final definition of  $\Theta_T^2$  (cf. (5.2)). Further details on all the reliability estimates can be found in the aforementioned bibliography.

**B. Efficiency of the a posteriori Error Estimator**

We now aim to prove the efficiency of  $\Theta$ , that is, the lower bound in (5.3). First, we deal with the zero order terms appearing in the definition of  $\Theta_T$ , for which we begin with the local estimates provided by the following lemma.

**Lemma 5.3.** *There hold*

$$\begin{aligned} \|\mathbf{f} + \mathbf{div}(\sigma_h)\|_{0,T} &\leq \|\sigma - \sigma_h\|_{\mathbf{div},T} \quad \forall T \in \mathcal{T}_h, \\ \|\sigma_h - \sigma_h^t\|_{0,T} &\leq 2\|\sigma - \sigma_h\|_{0,T} \quad \forall T \in \mathcal{T}_h, \end{aligned}$$

and there exist positive constants  $C_1, C_2$ , independent of  $h$ , such that

$$\|\mathbf{e}(\mathbf{u}_h) - \mathbf{t}_h\|_{0,T}^2 \leq C_1 \{ \|\mathbf{u} - \mathbf{u}_h\|_{1,T}^2 + \|\mathbf{t} - \mathbf{t}_h\|_{0,T}^2 \} \quad \forall T \in \mathcal{T}_h,$$

and

$$\|\rho_h - \nabla \mathbf{u}_h + \mathbf{e}(\mathbf{u}_h)\|_{0,T}^2 \leq C_2 \{ \|\rho - \rho_h\|_{0,T}^2 + \|\mathbf{u} - \mathbf{u}_h\|_{1,T}^2 \} \quad \forall T \in \mathcal{T}_h.$$

**Proof.** These inequalities follow using the relations  $\mathbf{f} = -\mathbf{div}(\sigma)$ ,  $\sigma = \sigma^t$ ,  $\mathbf{e}(\mathbf{u}) = \mathbf{t}$ , and  $\rho = \nabla \mathbf{u} - \mathbf{e}(\mathbf{u})$ , respectively. We omit further details. ■

We continue with the following nonlocal estimates.

**Lemma 5.4.** *There exists  $C_3, C_4 > 0$ , independent of  $h$ , such that there hold*

$$\sum_{e \in \mathcal{E}_h(\Gamma)} \|\mathbf{g} - \mathbf{u}_h\|_{0,e}^2 \leq C_3 \|\mathbf{u} - \mathbf{u}_h\|_{1,\Omega}^2,$$

and

$$\|\sigma_h^d - \mu(|\mathbf{t}_h|)\mathbf{t}_h + (\mathbf{u}_h \otimes \mathbf{u}_h)^d\|_{0,\Omega}^2 \leq C_4 \{ \|\sigma - \sigma_h\|_{0,\Omega}^2 + \|\mathbf{t} - \mathbf{t}_h\|_{0,\Omega}^2 + \|\mathbf{u} - \mathbf{u}_h\|_{1,\Omega}^2 \}.$$

**Proof.** The first estimate follows from the fact that  $\mathbf{g} = \mathbf{u}$  on  $\Gamma$  and the trace inequality, whereas the second one is a straightforward consequence of the constitutive relation  $\sigma^d = \mu(|\mathbf{t}|)\mathbf{t} - (\mathbf{u} \otimes \mathbf{u})^d$  in  $\Omega$ , the Lipschitz continuity of the nonlinear operator induced by  $\mu$  (cf. (3.10) in Lemma 3.4), a convenient decomposition of  $\mathbf{u} \otimes \mathbf{u} - \mathbf{u}_h \otimes \mathbf{u}_h$ , and the continuity of the injection  $i_c : \mathbf{H}^1(\Omega) \rightarrow \mathbf{L}^4(\Omega)$ . ■

The derivation of the upper bounds of the remaining terms defining the a posteriori error indicator  $\Theta_T^2$  proceeds similarly to [5], but adapting the results to the 3D case using some recent results from [19] and applying inverse inequalities and the localization technique based on element-bubble and edge-bubble functions. In this regard, we remark that the application of the aforementioned inequalities requires shape-regularity of the meshes. Nevertheless, the corresponding efficiency estimates should be easily extensible to more general triangulations by using, for instance, the results from [22] (where inverse inequalities for piecewise constant and continuous piecewise linear finite elements on locally refined shape-regular meshes were provided).

The announced estimates are summarized in the following three lemmas.

**Lemma 5.5.** *There exist positive constants  $C_5, C_6$ , independent of  $h$ , such that*

$$\begin{aligned} a) & h_T^2 \|\mathbf{curl}(\mathbf{t}_h + \boldsymbol{\rho}_h)\|_{0,T}^2 \leq C_5 \{ \|\mathbf{t} - \mathbf{t}_h\|_{0,T}^2 + \|\boldsymbol{\rho} - \boldsymbol{\rho}_h\|_{0,T}^2 \} \quad \forall T \in \mathcal{T}_h, \\ b) & h_e \|[(\mathbf{t}_h + \boldsymbol{\rho}_h) \times \mathbf{v}]\|_{0,e}^2 \leq C_6 \{ \|\mathbf{t} - \mathbf{t}_h\|_{0,\omega_e}^2 + \|\boldsymbol{\rho} - \boldsymbol{\rho}_h\|_{0,\omega_e}^2 \} \quad \forall e \in \mathcal{E}_h(\Omega), \end{aligned}$$

where  $\omega_e := \cup\{T' \in \mathcal{T}_h : e \in \mathcal{E}(T')\}$ .

**Proof.** We refer to [19, Lemmas 4.9 and 4.10] for the proofs of a) and b). ■

**Lemma 5.6.** *Assume that  $\mathbf{g}$  is piecewise polynomial. Then, there exists  $C_7 > 0$ , independent of  $h$ , such that*

$$h_e |(\nabla \mathbf{g} - \mathbf{t}_h - \boldsymbol{\rho}_h) \times \mathbf{v}|_{0,e}^2 \leq C_7 \{ \|\mathbf{t} - \mathbf{t}_h\|_{0,T_e}^2 + \|\boldsymbol{\rho} - \boldsymbol{\rho}_h\|_{0,T_e}^2 \} \quad \forall e \in \mathcal{E}_h(\Gamma),$$

where  $T_e$  is the tetrahedron of  $\mathcal{T}_h$  having  $e$  as a face.

**Proof.** It follows from a slight modification of the proof of [19, Lemma 4.13]. ■

**Lemma 5.7.** *There exists  $C_8 > 0$ , independent of  $h$ , such that*

$$\|\mathbf{f} - \mathcal{P}_h(\mathbf{f})\|_{0,T} \leq C_8 \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathbf{div},T} \quad \forall T \in \mathcal{T}_h.$$

**Proof.** It suffices to see that  $\|\mathbf{f} - \mathcal{P}_h(\mathbf{f})\|_{0,T}^2 = \|\mathcal{P}_h(\mathbf{div} \boldsymbol{\sigma}) - \mathbf{div} \boldsymbol{\sigma}\|_{0,T}^2$ , add and subtract  $\mathcal{P}_h(\mathbf{div} \boldsymbol{\sigma}_h)$ , and then apply continuity of the operator  $\mathcal{P}_h$ . ■

We close this section by remarking that the required efficiency of the a posteriori error estimator  $\Theta$  follows straightforwardly from Lemmas 5.3–5.7. In addition, we highlight that, on the contrary to [5], where the presentation concentrated in providing first the details of the 2D case and then a brief summary of the respective analysis for the 3D one, here we have privileged the discussion of the latter mainly because it involves the utilization of a recently established 3D Helmholtz decomposition. Moreover, the three-dimensional case is even more interesting from the applicability point of view, which is also taken into account below in the choice of our numerical results (cf. Section VI). In turn, some aspects of the 2D version of our a posteriori error analysis are provided next.

### C. Two-Dimensional Case

In what follows we briefly discuss the a posteriori error estimator in the two-dimensional case. We start by introducing some notations. For each  $T \in \mathcal{T}_h$  we let  $\mathcal{E}(T)$  be the set of edges of  $T$  and we denote by  $\mathcal{E}_h$  the set of all edges of  $\mathcal{T}_h$ , subdivided as in the three-dimensional case:

$$\mathcal{E}_h = \mathcal{E}_h(\Omega) \cup \mathcal{E}_h(\Gamma),$$

where  $\mathcal{E}_h(\Omega) := \{e \in \mathcal{E}_h : e \subseteq \Omega\}$  and  $\mathcal{E}_h(\Gamma) := \{e \in \mathcal{E}_h : e \subseteq \Gamma\}$ . In what follows,  $h_e$  stands for the length of a given edge  $e \in \mathcal{E}_h$ . Now, let  $v \in L^2(\Omega)$  such that  $v|_T \in C(T)$  for each  $T \in \mathcal{T}_h$ . Then, given  $e \in \mathcal{E}_h(\Omega)$ , we denote by  $[[v]]$  the jump of  $v$  across  $e$ , that is  $[[v]] := (v|_{T'})|_e - (v|_{T''})|_e$ , where  $T'$  and  $T''$  are the triangles of  $\mathcal{T}_h$  having  $e$  as an edge. Also, we fix a unit normal vector  $\mathbf{v}_e := (n_1, n_2)^t$  to the edge  $e$  (its particular orientation is not relevant) and let  $\mathbf{s}_e := (-n_2, n_1)^t$  be the corresponding fixed unit tangential vector along  $e$ . Hence, given  $\mathbf{v} \in \mathbf{L}^2(\Omega)$  and  $v \in \mathbf{L}^2(\Omega)$

such that  $\mathbf{v}|_T \in \mathbf{C}(T)$  and  $\boldsymbol{\tau}|_T \in \mathbf{C}(T)$ , respectively, for each  $T \in \mathcal{T}_h$ , we let  $[[\mathbf{v} \cdot \mathbf{s}_e]]$  and  $[[\boldsymbol{\tau} \mathbf{s}_e]]$  be the tangential jumps of  $\mathbf{v}$  and  $\boldsymbol{\tau}$ , across  $e$ , that is  $[[\mathbf{v} \cdot \mathbf{s}_e]] := \{(\mathbf{v}|_{T'})|_e - (\mathbf{v}|_{T''})|_e\} \cdot \mathbf{s}_e$  and  $[[\boldsymbol{\tau} \mathbf{s}_e]] := \{(\boldsymbol{\tau}|_{T'})|_e - (\boldsymbol{\tau}|_{T''})|_e\} \mathbf{s}_e$ , respectively. From now on, when no confusion arises, we will simply write  $\mathbf{s}$  and  $\mathbf{v}$  instead of  $\mathbf{s}_e$  and  $\mathbf{v}_e$ , respectively. Finally, for sufficiently smooth tensor fields  $\boldsymbol{\tau} := (\tau_{ij})_{2 \times 2}$ , we let

$$\mathbf{curl} \boldsymbol{\tau} := \begin{pmatrix} \frac{\partial \tau_{12}}{\partial x_1} - \frac{\partial \tau_{11}}{\partial x_2} \\ \frac{\partial \tau_{22}}{\partial x_1} - \frac{\partial \tau_{21}}{\partial x_2} \end{pmatrix}.$$

Now, let  $\vec{\mathbf{t}} := (\mathbf{t}, \boldsymbol{\sigma}, \mathbf{u}, \boldsymbol{\rho}) \in \mathbf{H}$  and  $\vec{\mathbf{t}}_h := (\mathbf{t}_h, \boldsymbol{\sigma}_h, \mathbf{u}_h, \boldsymbol{\rho}_h) \in \mathbf{H}_h$  be the unique solutions of the continuous and discrete problems (3.5) and (4.1), respectively. Then, we introduce the global a posteriori error estimator

$$\hat{\Theta} := \left\{ \sum_{T \in \mathcal{T}_h} \hat{\Theta}_T^2 \right\}^{1/2}, \tag{5.15}$$

where for each  $T \in \mathcal{T}_h$ :

$$\begin{aligned} \hat{\Theta}_T^2 &:= h_T^2 \|\mathbf{curl}(\mathbf{t}_h + \boldsymbol{\rho}_h)\|_{0,T}^2 + \|\boldsymbol{\sigma}_h^d - \mu(|\mathbf{t}_h|)\mathbf{t}_h + (\mathbf{u}_h \otimes \mathbf{u}_h)^d\|_{0,T}^2 \\ &+ \|\boldsymbol{\sigma}_h - \boldsymbol{\sigma}_h^t\|_{0,T}^2 + \|\mathbf{f} - \mathcal{P}_h(\mathbf{f})\|_{0,T}^2 + \|\mathbf{f} + \mathbf{div}(\boldsymbol{\sigma}_h)\|_{0,T}^2 \\ &+ \|\mathbf{e}(\mathbf{u}_h) - \mathbf{t}_h\|_{0,T}^2 + \|\boldsymbol{\rho}_h - \nabla \mathbf{u}_h + \mathbf{e}(\mathbf{u}_h)\|_{0,T}^2 \\ &+ \sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h(\Omega)} h_e |[[(\mathbf{t}_h + \boldsymbol{\rho}_h)\mathbf{s}]]|_{0,e}^2 + \sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h(\Gamma)} \|\mathbf{g} - \mathbf{u}_h\|_{0,e}^2 + h_e \left| \frac{d\mathbf{g}}{ds} - (\mathbf{t}_h + \boldsymbol{\rho}_h)\mathbf{s} \right|_{0,e}^2. \end{aligned}$$

The reliability of  $\hat{\Theta}$  can be proved similarly as in the three-dimensional case, that is, using a global inf-sup condition for a linearization of the problem, Helmholtz’s decomposition and local approximation properties of interpolation operators. Indeed, if we compare the definitions of  $\Theta$  (cf. (5.1)) and  $\hat{\Theta}$  we observe that most of the terms are exactly the same in both cases. In turn, to prove the efficiency of  $\hat{\Theta}$  it suffices to control the new terms, which is the purpose of the following two lemmas.

**Lemma 5.8.** *There exist positive constants  $C_1, C_2$ , independent of  $h$ , such that*

- a)  $h_T^2 \|\mathbf{curl}(\mathbf{t}_h + \boldsymbol{\rho}_h)\|_{0,T}^2 \leq C_1 \{ \|\mathbf{t} - \mathbf{t}_h\|_{0,T}^2 + \|\boldsymbol{\rho} - \boldsymbol{\rho}_h\|_{0,T}^2 \} \quad \forall T \in \mathcal{T}_h,$
- b)  $h_e |[[(\mathbf{t}_h + \boldsymbol{\rho}_h)\mathbf{s}]]|_{0,e}^2 \leq C_2 \{ \|\mathbf{t} - \mathbf{t}_h\|_{0,\omega_e}^2 + \|\boldsymbol{\rho} - \boldsymbol{\rho}_h\|_{0,\omega_e}^2 \} \quad \forall e \in \mathcal{E}_h(\Omega),$

where  $\omega_e := \cup\{T' \in \mathcal{T}_h : e \in \mathcal{E}(T')\}$ .

**Proof.** For a) we refer to [23, Lemma 4.3] (see also [5, Lemma 3.15] or [24, Lemma 4.9]). Similarly, for b) we refer to [23, Lemma 4.4] (see also [5, Lemma 3.15] or [24, Lemma 4.10]). ■

**Lemma 5.9.** *Assume that  $\mathbf{g}$  is piecewise polynomial. Then, there exists  $C_3 > 0$ , independent of  $h$ , such that*

$$h_e \left| \frac{d\mathbf{g}}{ds} - (\mathbf{t}_h + \boldsymbol{\rho}_h)\mathbf{s} \right|_{0,e}^2 \leq C_3 \{ \|\mathbf{t} - \mathbf{t}_h\|_{0,T_e}^2 + \|\boldsymbol{\rho} - \boldsymbol{\rho}_h\|_{0,T_e}^2 \} \quad \forall e \in \mathcal{E}_h(\Gamma),$$

where  $T_e$  is the triangle of  $\mathcal{T}_h$  having  $e$  as an edge.

**Proof.** The proof follows from a slight modification of the proof of [24, Lemma 4.15]. ■

Therefore, the main result in the 2D case is stated as follows.

**Theorem 5.10.** *Let  $\vec{t} \in \mathbf{H}$  and  $\vec{t}_h \in \mathbf{H}_h$  be the unique solutions of the continuous and discrete problems (3.5) and (4.1), respectively, and assume that  $\mathbf{g} \in \mathbf{H}^1(\Gamma)$ . Then, there exist  $\hat{C}_{rel} > 0$  and  $\hat{C}_{eff} > 0$ , independent of  $h$ , such that*

$$\hat{C}_{eff} \hat{\Theta} \leq \|\vec{t} - \vec{t}_h\| \leq \hat{C}_{rel} \hat{\Theta}.$$

## VI. NUMERICAL RESULTS

**Test 1.** Our first example serves to illustrate the accuracy of the mixed finite element method (as predicted by Theorem 4.4), and also to assess the practical performance of the method in a 3D computation. We construct the following analytical solution to (2.1) and (2.6)

$$\mathbf{u} = \begin{pmatrix} \sin(\pi x_1) \cos(\pi x_2) \cos(\pi x_3) \\ -2 \cos(\pi x_1) \sin(\pi x_2) \cos(\pi x_3) \\ \cos(\pi x_1) \cos(\pi x_2) \sin(\pi x_3) \end{pmatrix}, \quad \mathbf{t} = \mathbf{e}(\mathbf{u}), \quad \boldsymbol{\rho} = \nabla \mathbf{u} - \mathbf{e}(\mathbf{u}),$$

$$p = x_1^3 - x_2^3 - x_3^3, \quad \boldsymbol{\sigma} = \mu(|\mathbf{t}|)\mathbf{e}(\mathbf{u}) - (\mathbf{u} \otimes \mathbf{u}) - p\mathbb{I},$$

defined on the box  $\Omega = (0, 1) \times (0, 1) \times (-1, 1)$ , and where the viscosity is specified by (2.4) with  $\alpha_0 = 2$  and  $\alpha_1 = \beta = 1$ . The manufactured velocity is divergence free and it is used to apply the Dirichlet datum on  $\Gamma$ . It also satisfies the boundary compatibility condition  $\int_{\Gamma} \mathbf{u} \cdot \boldsymbol{\nu} = 0$ . Moreover, the proposed pressure has zero mean value in  $\Omega$ , implying that the last equation in (2.6) is fulfilled. The stabilization constants are chosen according to (3.19), leading to  $\kappa_1 = \kappa_2 = 0.0555$ ,  $\kappa_3 = 1$ ,  $\kappa_5 = 0.5$ , and as  $\kappa_4$  depends on the (unknown) Korn constant, we simply take  $\kappa_4 = \kappa_5$ . An experimental convergence analysis is performed, focusing on the lowest-order scheme (with  $k = 0$ ). Six steps of uniform mesh refinement were applied to an initial structured tetrahedral mesh, and on each nested mesh we denote computed errors and convergence rates as

$$e(\mathbf{t}) = \|\mathbf{t} - \mathbf{t}_h\|_{0,\Omega}, \quad e(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{0,\text{div}}, \quad e(\mathbf{u}) = \|\mathbf{u} - \mathbf{u}_h\|_{1,\Omega},$$

$$e(\boldsymbol{\rho}) = \|\boldsymbol{\rho} - \boldsymbol{\rho}_h\|_{0,\Omega}, \quad r(\cdot) = -2 \log(e(\cdot)/\hat{e}(\cdot))[\log(N/\hat{N})]^{-1},$$

where  $e, \hat{e}$  stand for errors generated by methods on meshes having  $N, \hat{N}$  degrees of freedom, respectively. These errors are tabulated by number of degrees of freedom and meshsize in Table I. Each individual error exhibits a clear  $O(h)$  rate of convergence, as expected from the a priori error estimates stated in Theorem 4.4. The values collected in the last column of the table confirm that a maximum of six Picard iterations are required to achieve a prescribed tolerance of 1E-6. All linear systems arising after fixed point linearization are solved with the unsymmetric multifrontal direct solver for sparse matrices UMFPAK. Computed solutions are shown in Fig. 1.

**Test 2.** Our next example assesses the accuracy of the proposed scheme along with the properties of the adaptive error estimator (5.15) (specialized for the 2D case). The domain is conformed by a rectangle  $(0, 1.5) \times (0, 1)$  with a pear-shaped hole inside. The viscosity is also given by (2.4)

TABLE I. Test 1. Experimental convergence and fixed point iteration count for the approximation of the Navier-Stokes equations with nonlinear viscosity.

D.o.f.	h	$e(t)$	$r(t)$	$e(\sigma)$	$r(\sigma)$	$e(u)$	$r(u)$	$e(\rho)$	$r(\rho)$	iter
270	1.4142	4.2007	–	42.5310	–	8.1726	–	2.3254	–	6
1887	0.7071	2.6383	0.6710	24.4819	0.7967	5.4160	0.5935	1.8920	0.2975	5
14211	0.3535	1.4840	0.8300	13.9391	0.8125	3.1370	0.7878	1.1008	0.7813	6
74322	0.2020	0.8804	0.9330	8.1546	0.9580	1.8845	0.9106	0.6505	0.9398	6
285090	0.1285	0.5683	0.9682	5.2564	0.9715	1.2177	0.9660	0.4143	0.9983	5
871875	0.0883	0.3930	0.9845	3.6869	0.9465	0.8416	0.9859	0.2843	1.0046	6
2741364	0.0524	0.2472	0.9937	2.2198	0.9687	0.5129	0.9803	0.1822	1.0175	6

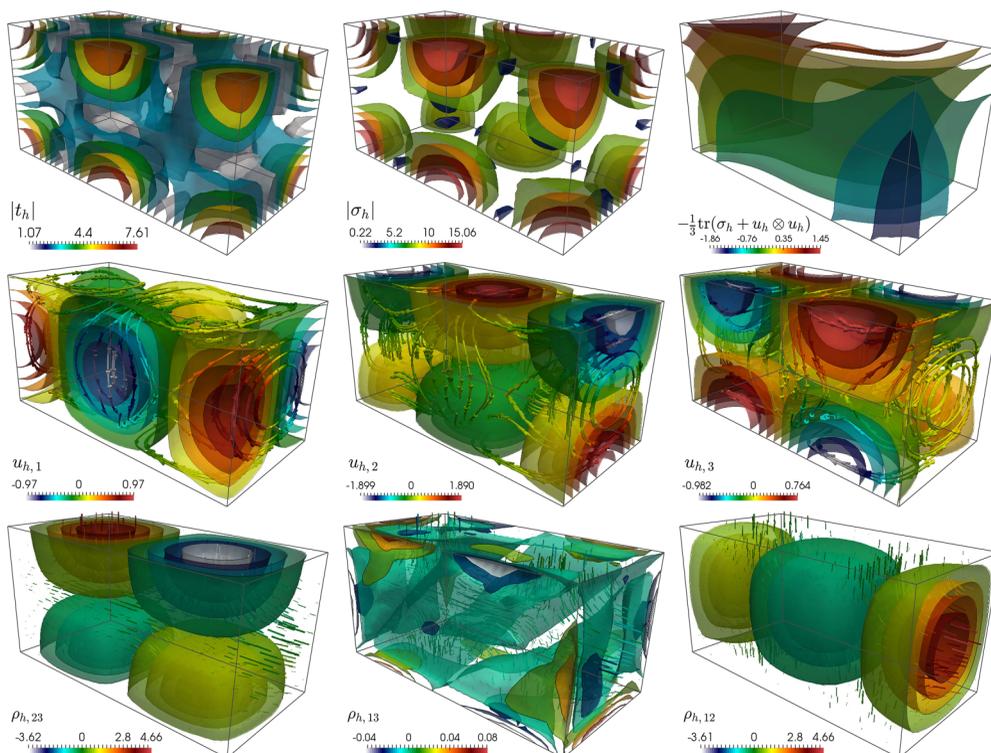


FIG. 1. Test 1. Iso-surfaces of Euclidean norm of the approximate strain tensor (top left), Euclidean norm of the Cauchy stress (top middle), postprocessed pressure (top right), approximate velocity components and computed streamlines (center row), and vorticity components (bottom row). [Color figure can be viewed at wileyonlinelibrary.com]

with  $\alpha_0 = 1$ ,  $\alpha_1 = 0.1$ , and  $\beta = 1$ , and the forcing and boundary terms  $f, g$  are chosen such that the exact solution to (2.6) is given by

$$\begin{aligned}
 \mathbf{u} &= \begin{pmatrix} 1 - \exp(\lambda x_1) \cos(2\pi x_2) \\ \frac{\lambda}{2\pi} \exp(\lambda x_1) \sin(2\pi x_2) \end{pmatrix}, \quad \mathbf{t} = \mathbf{e}(\mathbf{u}), \quad \rho = \nabla \mathbf{u} - \mathbf{e}(\mathbf{u}), \\
 p &= \frac{1 - \exp(2\lambda x_1)}{2(x_1 - a)^2 + 2(x_2 - b)^2}, \quad \sigma = \mu(|\mathbf{t}|)\mathbf{e}(\mathbf{u}) - (\mathbf{u} \otimes \mathbf{u}) - p\mathbb{I},
 \end{aligned}$$

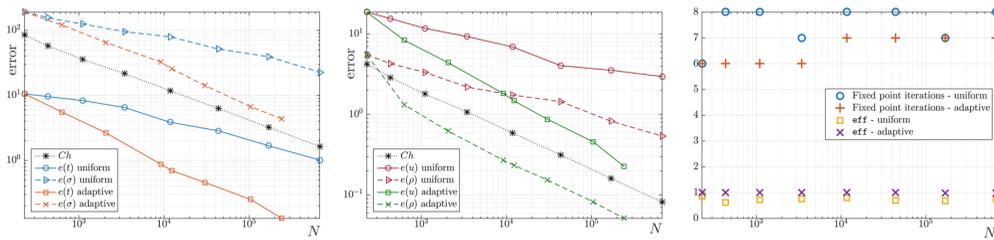


FIG. 2. Test 2. Individual error history vs. the number of degrees of freedom. Errors for strain and stress (left), velocity and vorticity (middle), and Picard iteration count with effectivity indexes (right) for two runs using uniform and adaptive mesh refinement according to the a posteriori error estimator (5.15). [Color figure can be viewed at wileyonlinelibrary.com]

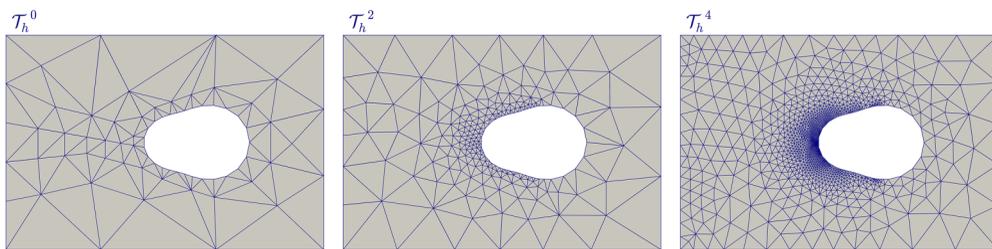


FIG. 3. Test 2. Initial and intermediate triangulations obtained after a few steps of mesh adaptation according to the a posteriori error estimator (5.15). [Color figure can be viewed at wileyonlinelibrary.com]

with the parameter  $\lambda = 0.5 - \sqrt{0.25 + 4\pi^2}$ , and  $(a, b) = (3/4, 1/2)$  is a point located within the hole in the domain, and close to the interior of  $\Omega$ . The stabilization parameters take the values  $\kappa_1 = \kappa_2 = 0.6944$ ,  $\kappa_3 = 0.5$ ,  $\kappa_4 = \kappa_5 = 0.25$ . Due to the singularity of the pressure (and therefore the Cauchy stress) at  $(a, b)$  we expect suboptimal convergence of the method under uniform mesh refinement, and so we apply an adaptive algorithm. The nonlinear system is linearized with a fixed-point strategy (stopped when the  $L^2$ - norm of the total residual attains the tolerance 1E-6), and in this case the subsequent linear systems were solved with the direct solver MUMPS. After computing locally, the estimator using (5.15), we proceed to tag elements for refinement using the Dörfler strategy, which consists in marking sufficiently many elements so that they represent a given fraction of the total estimated error (cf. [25]). A remeshing method is then applied, targeting the equidistribution of the local error indicators in the updated mesh.

In Fig. 2, we report on the convergence history of the method (in its lowest-order configuration) following both a uniform refinement, and successive mesh adaptation according to the algorithm described above. First, we notice a slightly larger Picard iteration count for this example (in comparison with the results in Table I). Second, we also observe a loss of optimality in the convergence rates under uniform mesh refinement, especially for the vorticity. In addition, the effectivity index (computed as  $eff(\hat{\theta}) := e/\hat{\theta}$ , where  $e$  denotes the total error) oscillates around 0.7 in the case of uniform refinement. This is remediated by the adaptive algorithm, which restores optimal convergence rates for all fields and a much more steady effectivity index. One can also notice that the method using adaptive mesh refinement outperforms the uniformly refined scheme by almost two orders of magnitude (in the sense of needed degrees of freedom to attain a given error). The initial coarse mesh, together with triangulations obtained after two and four adaptive

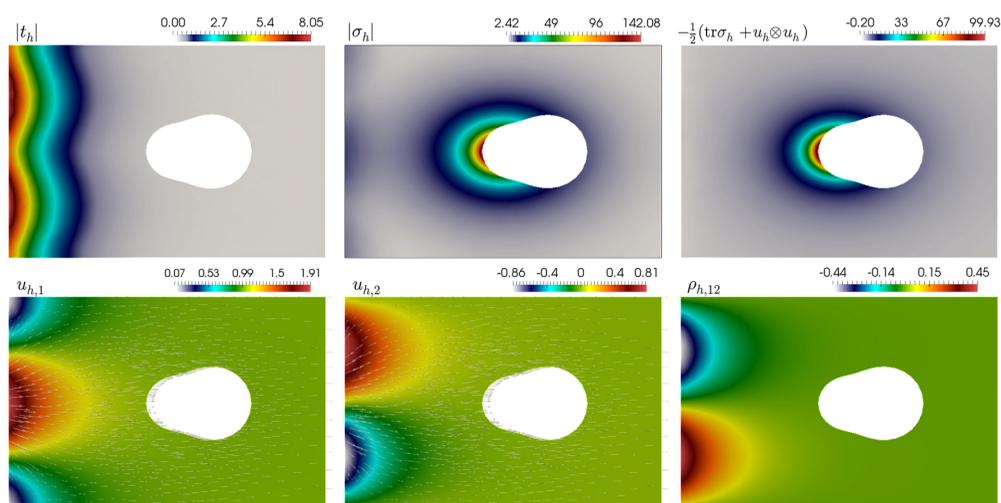


FIG. 4. Test 2. Numerical solutions computed on an adaptively refined mesh. Strain norm, Cauchy stress norm, postprocessed pressure, velocity components, and vorticity. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

steps are displayed in Fig. 3. We observe meshes heavily refined in the neighbourhood of  $(a, b)$  and near to the left wall. Finally, we present the computed numerical solutions in Fig. 4, exhibiting well-resolved profiles for all fields.

**Test 3.** The third example focuses on the driven cavity flow problem in the unit cube. The external body force is zero, and the three-dimensional flow patterns are determined by the boundary conditions only: an unidirectional Dirichlet velocity is set on the top lid  $\mathbf{u} = \mathbf{g} = (1, 0, 0)^T$ , and no-slip velocities  $\mathbf{u} = 0$  are imposed elsewhere on  $\Gamma$ . The viscosity is again taken as the Carreau law (2.4) with  $\alpha_0 = 0.005$ ,  $\alpha_1 = 0.01$  and  $\beta = 1$ . An initially coarse tetrahedral mesh of 1058 elements and 332 vertices was generated, and the proposed numerical scheme was used to solve the model problem, now via Newton linearization steps (with a fixed tolerance of  $1\text{E-}8$  on the residuals). After this first initial solve, we use an adaptive mesh refinement algorithm marking elements for refinement according to the locally computable error indicator (5.1), regenerating the mesh, and then solving again the discrete nonlinear problem until reaching the convergence of the Newton iterations. For this example, the linear systems were solved with a BiCGStab method with left Schur complement preconditioning. Five steps of adaptive mesh refinement were applied (with a maximum Newton iteration count of 6), and the approximate solutions computed on a mesh with 13685 elements are portrayed in Fig. 5. We observe smooth vorticity and strain profiles, and from the velocity streamlines we can evidence the formation of the typical asymmetric vortex parallel to the  $x_1 - x_3$  plane. In Fig. 6, we present examples of three adapted meshes resulting from adaptive refinement guided by the a posteriori error estimator (5.1), after one, three, and five adaptive steps. An agglomeration of tetrahedra is observed near the top corners of the domain, where the Dirichlet datum is discontinuous, and where the stress is concentrated.

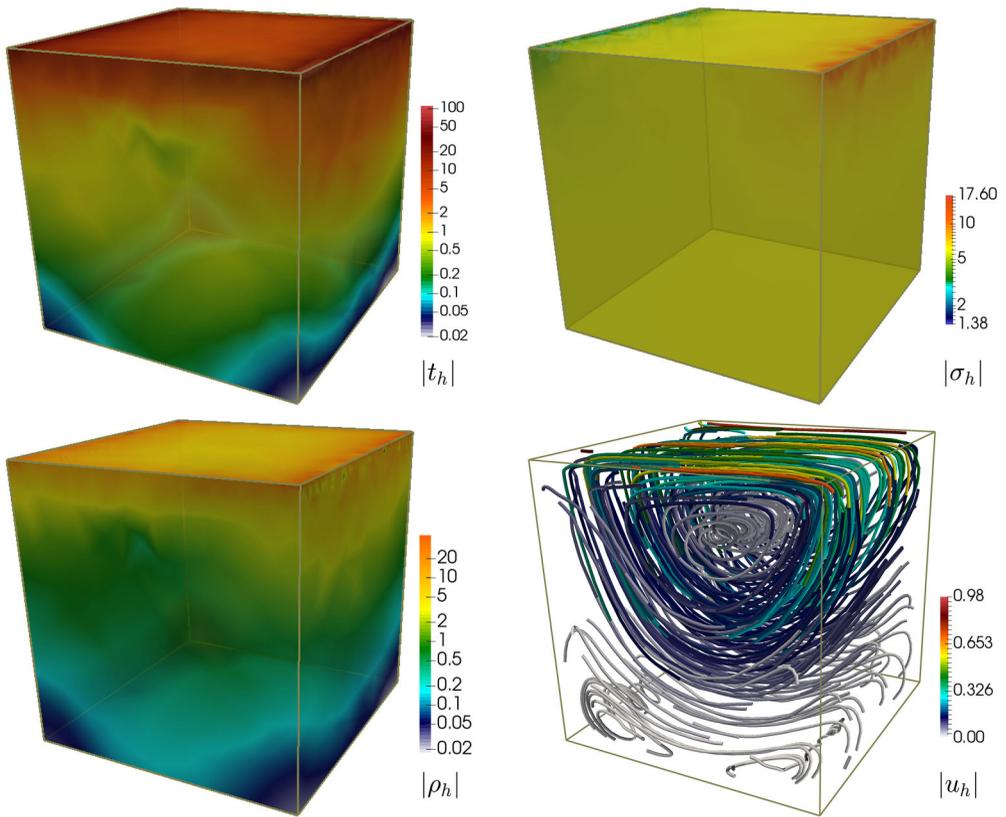


FIG. 5. Test 3. Numerically computed mixing patterns for the cavity flow benchmark. Results obtained with the lowest-order method. Strain norm, Cauchy stress norm, vorticity magnitude, and velocity streamlines. [Color figure can be viewed at wileyonlinelibrary.com]

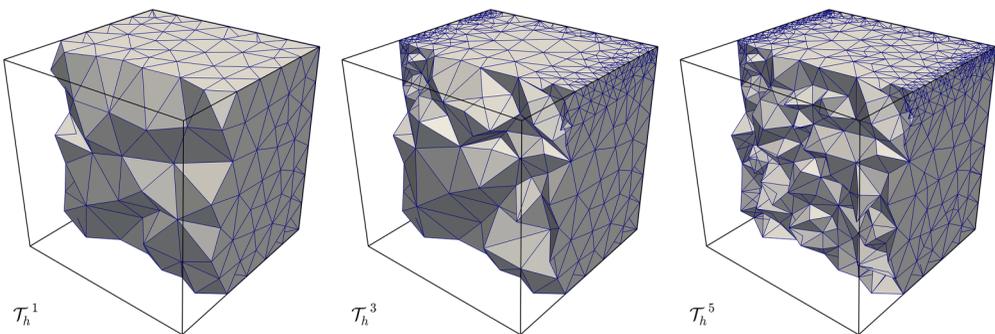


FIG. 6. Test 3. Clip of three intermediate tetrahedral meshes, refined adaptively according to the a posteriori error estimator (5.1). [Color figure can be viewed at wileyonlinelibrary.com]

## References

1. J. Camaño, G. N. Gatica, R. Oyarzúa, and G. Tierra, An augmented mixed finite element method for the Navier–Stokes equations with variable viscosity, *SIAM J Numer Anal* 54 (2016), 1069–1092.
2. J. Camaño, R. Oyarzúa, and G. Tierra, Analysis of an augmented mixed-FEM for the Navier Stokes problem, *Math Comp* 86 (2017), 589–615.
3. J. S. Howell and N. J. Walkington, Dual-mixed finite element methods for the Navier-Stokes equations, *ESAIM Math Model Numer Anal* 47 (2013), 789–805.
4. G. N. Gatica, A. Márquez, R. Oyarzúa, and R. Rebolledo, Analysis of an augmented fully-mixed approach for the coupling of quasi-Newtonian fluids and porous media, *Comput Methods Appl Mech Eng* 270 (2014), 76–112.
5. G. N. Gatica, R. Ruiz-Baier, and G. Tierra, A posteriori error analysis of an augmented mixed method for the Navier–Stokes equations with nonlinear viscosity, *Comput Math Appl* 72 (2016), 2289–2310.
6. G. N. Gatica, A. Márquez, and M. A. Sánchez, A priori and a posteriori error analyses of a velocity-pseudostress formulation for a class of quasi-Newtonian Stokes flows. *Comput Methods Appl Mech Eng* 200 (2011), 1619–1636.
7. A. I. Garralda-Guillem, M. Ruiz Galan, G. N. Gatica, and A. Márquez, A posteriori error analysis of twofold saddle point variational formulations for nonlinear boundary value problems, *IMA J Numer Anal* 34 (2014), 326–361.
8. A. Ainsworth and J. T. Oden, A posteriori error estimation in finite element analysis, *Comput Methods Appl Mech Eng* 142 (1997), 1–88.
9. R. Verfürth, A review of a-posteriori error estimation and adaptive mesh-refinement techniques, Wiley and Teubner Series, *Advances in Numerical Mathematics*, 1996.
10. G. N. Gatica and W. L. Wendland, Coupling of mixed finite elements and boundary elements for linear and nonlinear elliptic problems, *Appl Anal* 63 (1996), 39–75.
11. E. Colmenares, G. N. Gatica, and R. Oyarzúa, Analysis of an augmented mixed-primal formulation for the stationary Boussinesq problem, *Numer Methods Partial Differential Equations* 32 (2016), 445–478.
12. F. Brezzi, and M. Fortin, *Mixed and hybrid finite element methods*. Springer Series in Computational Mathematics, Vol. 15, Springer-Verlag, New York, 1991. pp. 350.
13. G. N. Gatica, *A simple introduction to the mixed finite element method: theory and applications*, Springer Briefs in Mathematics, Springer, Cham, 2014.
14. R. A. Adams, and J. J. F. Fournier, *Sobolev Spaces*, 2nd Ed., Pure and Applied Mathematics (Amsterdam), 140, Elsevier/Academic Press, Amsterdam, 2003, pp. xiv+305 pp.
15. A. Quarteroni and A. Valli, *Numerical approximation of partial differential equations*, Springer Series in Computational Mathematics, 23, Springer-Verlag, Berlin, 1994, pp. xvi+543
16. W. McLean, *Strongly elliptic systems and boundary integral equations*. Cambridge University Press, Cambridge, 2000.
17. G. N. Gatica, An augmented mixed finite element method for linear elasticity with non-homogeneous Dirichlet conditions, *Electron Trans Numer Anal* 26 (2007), 421–438.
18. J. Nečas, *Introduction to the theory of nonlinear elliptic equations*, Reprint of the 1983 Ed., A Wiley-Interscience Publication, Wiley, Chichester, 1986.
19. G. N. Gatica, L. F. Gatica, and F. Sequeira, A priori and a posteriori error analyses of a pseudostress-based mixed formulation for linear elasticity, *Comput Math Appl* 71 (2016), 585–614.
20. G. N. Gatica, A note on stable Helmholtz decompositions in 3D, Preprint 2016-03, Centro de Investigación en Ingeniería Matemática (CI<sup>2</sup>MA), Universidad de Concepción, Chile, Available at: <http://www.ci2ma.udec.cl/publicaciones/prepublicaciones>.

21. P. Clément, Approximation by finite element functions using local regularisation, *RAIRO Analyse Numérique* 9 (1975), 77–84.
22. W. Dahmen, B. Faermann, I.G. Graham, W. Hackbusch, and S.A. Sauter, Inverse inequalities on non-quasi-uniform meshes and application to the mortar element method, *Math Comp* 73 (2004), 1107–1138.
23. T. P. Barrios, G. N., Gatica, M. González, and N. Heuer, A residual based a posteriori error estimator for an augmented mixed finite element method in linear elasticity, *ESAIM Math Model Numer Anal* 40 (2006), 843–869.
24. G. N. Gatica, A. Márquez, and M. A. Sánchez, Analysis of a velocity-pressure-pseudostress formulation for the stationary Stokes equations, *Comput Methods Appl Mech Eng* 199 (2010), 1064–1079.
25. W. Dörfler, A convergent adaptive algorithm for Poisson’s equation, *SIAM J Numer Anal* 33 (1994), 1106–1124.