Analysis of a conservative mixed-FEM for the stationary Navier–Stokes problem *

Jessika Camaño[†], Carlos García[‡] and Ricardo Oyarzúa[§]

Dedicated to Professor Gabriel N. Gatica on the occasion of his 60th birthday

Abstract

In this paper we propose and analyze a new conservative mixed finite element method for the Navier-Stokes problem posed in non-standard Banach spaces. Our approach is based on the introduction of a pseudostress tensor relating the velocity gradient with the convective term, leading to a mixed formulation where the aforementioned pseudostress tensor and the velocity are the main unknowns of the system. Then the associated Galerkin scheme can be defined by employing Raviart–Thomas elements of degree k for the pseudostress tensor and discontinuous piecewise polynomial elements of degree k for the velocity. With this choice of spaces, the equilibrium equation is exactly satisfied if the external force belongs to the velocity discrete space, thus the method is conservative, which constitutes one of the main feature of our approach. For both, the continuous and discrete problems, the Banach–Nečas–Babuška and Banach's fixed point theorems are employed to prove unique solvability. We also provide the convergence analysis and particularly prove that the error decay with optimal rate of convergence. Further variables of interest, such as the fluid pressure, the fluid vorticity and the fluid velocity gradient, can be easily approximated as a simple postprocess of the finite element solutions with the same rate of convergence. Finally, several numerical results illustrating the performance of the method are provided.

Key words: Navier–Stokes, conservativity, mixed finite element method, Banach spaces, Raviart-Thomas elements

Mathematics Subject Classifications (1991): 65N15, 65N30, 76D05, 76M10

---

*This research was partially supported by CONICYT-Chile through project Fondecyt 1161325, project Fondecyt 1180859 and project AFB170001 of the PIA Program: Concurso Apoyo a Centros Cientificos y Tecnologicos de Excelencia con Financiamiento Basal.

[†]Departamento de Matemática y Física Aplicadas, Universidad Católica de la Santísima Concepción, Casilla 297, Concepción, Chile, and CI^{2}MA, Universidad de Concepción, Casilla 160-C, Concepción, Chile, email: `jecamano@ucsc.cl`

[‡]Departamento de Matemática, Universidad del Bío-Bío, Casilla 5-C, Concepción, Chile, and CI^{2}MA, Universidad de Concepción, Casilla 160-C, Concepción, Chile, email: `carlos.garcia.vera.ingmath@gmail.com`

[§]Departamento de Matemática, Universidad del Bío-Bío, Casilla 5-C, Concepción, Chile, and CI^{2}MA, Universidad de Concepción, Casilla 160-C, Concepción, Chile, email: `royarzua@ubiobio.cl`

1 Introduction

The Navier–Stokes (NS) problem is one of the most challenging problems in different areas of mathematics and particularly the numerical analysis community has been working for decades in the devising of accurate and efficient numerical methods to approximate the solution of NS, all of them with pros and cons. Probably, the most commonly used methods by engineers to approximate the solution of this important problem are those based on conforming discretizations of the classical velocity-pressure formulation, mainly because they are relatively cheap and easy of implementing. Actually, in most of the softwares designed to solve partial differential equations, such as *Freefem++* and *Fenics*, the classical families of finite elements for the Stokes problem are already available (see [22] for a detailed study of these classical families). However, it is well-known that, in general for flow problems, conforming H^1-discretizations lead to non-conservative methods, as it is the case of conforming velocity-pressure discretizations of NS. In order to circumvent this lack of conservativity, many researchers have turned to other type of discretizations, such as Finite Volumes and Discontinuous Galerkin methods, among others (see for instance [8, 9, 26, 27, 29], and the references therein).

One of the classical approaches to obtain conservative methods for flow problems is the discretization by means of mixed finite element methods. In fact, since the equilibrium equation is discretized at the same time with the constitutive equation, by construction, they are naturally conservative. This is the case, for instance, of pseudostress-based mixed methods for the Stokes equations (see e.g. [2], [15], [19], [20], and the references therein). In fact, by introducing a presudsotress tensor $\boldsymbol{\sigma}$, relating the gradient of the velocity with the pressure, the Stokes equations can be rewritten as a first–order system of equations, where the equilibrium equation has the form $-\mathbf{div}\,\boldsymbol{\sigma} = \mathbf{f}$, with $\mathbf{f}$ being the datum, thus leading to a conservative scheme. There are other two advantages to utilize mixed methods, particularly for fluid-flow problems. On the one hand, they have a natural applicability to non-Newtonian flows. Indeed, since in this case the constitutive equation is nonlinear, the stress cannot be eliminated, and hence it becomes an unavoidable unknown in the corresponding solvability analysis. Actually, they allow for a unified analysis for linear and nonlinear flows (see e.g. [12, 18, 21, 24]). On the other hand, further variables of interest can be approximated without loosing any accuracy (see e.g. [6]).

In recent years, the study of dual-mixed formulations for fluid-flow problems has been also extended to the Navier–Stokes equations. In particular, Farhloul et al. in [13] and [14] extended the analysis of dual-mixed formulations for the Stokes equations to the Navier–Stokes problem. They introduce the strain tensor (in [13]) and the velocity gradient tensor (in [14]) as the main unknowns of the systems and derive quasi-optimal numerical methods for the fluid flow problem. In turn, in [5] (see also [3] and [4]), Cai et al. extended the analysis of pseudostress-based mixed methods for the Stokes problem to the Navier–Stokes equations. They introduce and analyse a conforming $Hdiv$ method for a pseudostress-based mixed formulation of accuracy $O(h^{k+1-n/6})$ in the L^3 norm, with $k \geq 0$ being the polynomial degree and $n = 2, 3$. Later on, in [25], Howell and Walkington, introduced a new dual-mixed finite element method for the Navier–Stokes problem, considering the velocity gradient (in L^2), the velocity (in L^2) and a modified pseudostress (or stress) tensor (in $Hdiv$) linking the gradient of the velocity and the pressure with the convective term, as the main unknowns of the system. Since the analysis hinges on non-standard inf-sup conditions, in [25] the authors propose new families of finite elements obtained by enriching well-known families of finite elements designed for the elasticity problem, such as the Arnold–Falk–Winther and PEERS elements. With these spaces it can be proved optimal convergence with a computational cost relatively high. Finally, in [6] it has been introduced a new augmented-mixed finite element method for NS where, similarly to [25], it is introduced a non-standard pseudostress tensor relating the gradient of the velocity with the convective term in such a way this pseudostress (in $Hdiv$) together to the velocity in H^1 are the only unknowns of the system and the pressure, as well as other variables of interest, can be recovered by employing a simple post-processing procedure. In [6], well-posedness of the continuous and discrete problems, as well as optimal convergence, are obtained owing to the incorporation of Galerkin least-squares type terms in the formulation, which prevents, as it is also the case of the aforementioned works [5] and [13], the obtaining of a conservative method.

According to the above discussion, and with the idea of contributing with new methods to approximate the solution of NS, in this paper we introduce and analyze a new $Hdiv$-conforming finite element method for the stationary Navier–Stokes problem with constant viscosity, which up to the author's knowledge, is the first conservative conforming method for NS. Our approach consists in rewriting the corresponding system of equations in terms of the pseudostress tensor previously utilized in [6], say $\boldsymbol{\sigma}$, in such a way after eliminating the fluid pressure from the system, a first-order set of equations can be derived. One of the advantages of employing this procedure is that the equilibrium equation can be written in the form $-\mathbf{div}\,\boldsymbol{\sigma} = \mathbf{f}$, as for the Stokes equations, allowing us to derive our conservative scheme. Differently from [6], instead of considering the velocity in H^1, and consequently enriching the formulation with Galerkin least-squares type terms, we introduce non-standard Banach space for both unknowns, the pseudostress $\boldsymbol{\sigma}$ and the velocity $\mathbf{u}$, in such a way well-posedness of the continuous problem can be proved by means of the Banach–Nečas–Babuška theorem, a fixed–point strategy and a small data assumption. For the associated Galerkin scheme we employ Raviart–Thomas elements of degree $k \geq 0$ to approximate $\boldsymbol{\sigma}$ and discontinuous piecewise polynomials of degree k for $\mathbf{u}$ and apply the same arguments utilized for the continuous problem to prove unique solvability. In turn, we derive the corresponding Cea's estimate and prove optimal convergence of our method, which is confirmed with the respective numerical tests below.

The rest of this paper is organized as follows. In Section 2 we present the main aspects of the continuous problem. We reformulate the fluid flow problem as an equivalent first-order set of equations and derive our mixed variational formulation. In Section 3 we introduce the fixed–point strategy and apply, firstly, the classical Banach–Nečas–Babuška theorem, and secondly, the Banach's fixed–point theorems, to prove that the associated fixed–point operator is well defined and that the continuous problem is uniquely solvable, respectively. Next, in Section 4 we introduce and analyze the associated Galerkin scheme by mimicking the theory developed for the continuous problem. In addition, we establish the corresponding Cea's estimate and prove optimal convergence of the method. Finally, several numerical results illustrating the good performance of the method, are reported in Section 5.

2 The model problem and its conservative mixed formulation

2.1 Preliminaries

Let us denote by $\Omega \subseteq \mathbb{R}^n$, $n \in \{2, 3\}$ a given bounded domain with polyhedral boundary Γ, and denote by $\mathbf{n}$ the outward unit normal vector on Γ. Standard notations will be adopted for Lebesgue spaces $\mathrm{L}^p(\Omega)$, with $p \in [1, \infty]$ and Sobolev spaces $\mathrm{W}^{r,p}(\Omega)$ with $r \geq 0$, endowed with the norms $\|\cdot\|_{\mathrm{L}^p(\Omega)}$ and $\|\cdot\|_{\mathrm{W}^{r,p}(\Omega)}$, respectively. Note that $\mathrm{W}^{0,p}(\Omega) = \mathrm{L}^p(\Omega)$ and if $p = 2$, we write $\mathrm{H}^r(\Omega)$ in place of $\mathrm{W}^{r,2}(\Omega)$, with the corresponding Lebesgue and Sobolev norms denoted by $\|\cdot\|_{0,\Omega}$ and $\|\cdot\|_{r,\Omega}$, respectively. We also write $|\cdot|_{r,\Omega}$ for the H^r-seminorm. In addition, $\mathrm{H}^{1/2}(\Gamma)$ is the spaces of traces of functions of $\mathrm{H}^1(\Omega)$ and $\mathrm{H}^{-1/2}(\Gamma)$ denotes its dual. With $\langle\cdot,\cdot\rangle$ we denote the corresponding product of duality between $\mathrm{H}^{1/2}(\Gamma)$ and $\mathrm{H}^{-1/2}(\Gamma)$. By $\mathbf{S}$ and $\mathbb{S}$ we will denote the corresponding vectorial and tensorial counterparts of the generic scalar functional space S. 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

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

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 div acting along the rows of $\boldsymbol{\tau}$, and define the transpose, the trace, the tensor inner product, and the deviatoric tensor, respectively, as

$$\boldsymbol{\tau}^{\mathrm{t}} := (\boldsymbol{\tau}_{ij})_{i,j=1,n}, \quad \operatorname{tr}(\boldsymbol{\tau}) := \sum_{i=1}^{n}\tau_{ii}, \quad \boldsymbol{\tau}:\boldsymbol{\zeta} := \sum_{i,j=1}^{n}\tau_{ij}\zeta_{ij}, \quad \text{and } \boldsymbol{\tau}^{\mathrm{d}} := \boldsymbol{\tau} - \frac{1}{n}\operatorname{tr}(\boldsymbol{\tau})\mathbb{I}, \tag{2.1}$$

where $\mathbb{I}$ is the identity tensor in $\mathbb{R}^{n\times n}$. For simplicity, in what follows we denote

$$(v,w)_\Omega := \int_\Omega vw, \quad (\mathbf{v},\mathbf{w})_\Omega := \int_\Omega \mathbf{v}\cdot\mathbf{w}, \quad (\mathbf{v},\mathbf{w})_\Gamma := \int_\Gamma \mathbf{u}\cdot\mathbf{v} \quad \text{and} \quad (\boldsymbol{\tau},\boldsymbol{\zeta})_\Omega := \int_\Omega \boldsymbol{\tau}:\boldsymbol{\zeta}.$$

Furthermore, we recall that the Hilbert space

$$\mathbf{H}(\mathrm{div};\Omega) := \left\{\tau\in\mathbf{L}^2(\Omega): \quad \operatorname{div}\tau\in\mathrm{L}^2(\Omega)\right\},$$

equipped with the usual norm $\|\tau\|^2_{\mathrm{div},\Omega} := \|\tau\|^2_{0,\Omega} + \|\operatorname{div}\tau\|^2_{0,\Omega}$ is standard in the realm of mixed problems. However, in the sequel we will make use of the tensor version of $\mathbf{H}(\mathrm{div};\Omega)$, namely

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

whose norm will be denoted $\|\cdot\|_{\mathbf{div},\Omega}$. In turn, given $p > \dfrac{2n}{n+2}$, in what follows we will also employ the non-standard Banach space $\mathbb{H}(\mathbf{div}_p\,;\Omega)$ defined by

$$\mathbb{H}(\mathbf{div}_p\,;\Omega) := \{\boldsymbol{\tau}\in\mathbb{L}^2(\Omega) \;:\; \mathbf{div}\,\boldsymbol{\tau}\in\mathbf{L}^p(\Omega)\}, \tag{2.2}$$

endowed with the norm

$$\|\boldsymbol{\tau}\|_{\mathbf{div}_p,\Omega} := \left(\|\boldsymbol{\tau}\|^2_{0,\Omega} + \|\mathbf{div}\,\boldsymbol{\tau}\|^2_{\mathbf{L}^p(\Omega)}\right)^{1/2}.$$

Finally, for any scalar function v, we define the sign function sgn, given by

$$\operatorname{sgn}(v) := \begin{cases} 1 & \text{if} \quad v\geq 0, \\ -1 & \text{if} \quad v<0, \end{cases}$$

It is quite clear that for a given v, there holds $v\operatorname{sgn}(v) = |v|$. For a vector function $\mathbf{v} = (v_i)_{i=1,n}$, we extend the sign function as follows: $\mathbf{sgn}(\mathbf{v})_i = \operatorname{sgn}(v_i)$ and observe that $\mathbf{sgn}(\mathbf{v})\cdot\mathbf{v} = |v_1| + |v_2| + ... + |v_n|$, hence

$$|\mathbf{v}| \leq \mathbf{sgn}(\mathbf{v})\cdot\mathbf{v} \leq \sqrt{n}|\mathbf{v}|. \tag{2.3}$$

Note also that, for all $\mathbf{v}\in\mathbb{R}^n$ we have the property $|\mathbf{sgn}(\mathbf{v})| = \sqrt{n}$. Above, $|\cdot|$ denotes the absolute value or the Euclidian norm in $\mathbb{R}^n$

2.2 The steady-state Navier–Stokes problem

Let $\Omega \subseteq \mathbb{R}^n$, $n \in \{2,3\}$ be a bounded domain with Lipschitz boundary Γ and let $\nu > 0$, $\mathbf{u}$ and p be the viscosity, the velocity and the pressure, respectively, of a viscous fluid occupying the region Ω, whose movement is described by the incompressible steady-state Navier–Stokes equations with Dirichlet boundary condition:

$$\begin{aligned} -\nu\Delta\mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u} + \nabla p &= \mathbf{f} \quad \text{in} \quad \Omega, \\ \operatorname{div}\mathbf{u} &= \mathbf{0} \quad \text{in} \quad \Omega, \\ \mathbf{u} &= \mathbf{u}_D \quad \text{on} \quad \Gamma, \\ (p,1)_\Omega &= 0. \end{aligned} \tag{2.4}$$

Above, $\mathbf{f}$ represents an external force acting on Ω and $\mathbf{u}_D$ is the prescribed velocity on Γ, satisfying the compatibility condition:

$$(\mathbf{u}_D \cdot \mathbf{n}, 1)_\Gamma = 0.$$

As we already mentioned before, we are interested in deriving a conservative mixed method to approximate the solution of (2.4). To that end, we proceed analogously as in [6] and write (2.4) as an equivalent first-order set of equations by introducing the "nonlinear-pseudostress" tensor

$$\boldsymbol{\sigma} := \nu\nabla\mathbf{u} \,-\, p\mathbb{I} \,-\, \mathbf{u}\otimes\mathbf{u} \quad \text{in} \quad \Omega. \tag{2.5}$$

Notice that from the incompressibility condition $\operatorname{div}\mathbf{u} = \operatorname{tr}(\nabla\mathbf{u}) = 0$ in Ω, there hold

$$\mathbf{div}\,(\mathbf{u}\otimes\mathbf{u}) = (\mathbf{u}\cdot\nabla)\mathbf{u} \quad \text{in} \quad \Omega \quad \text{and} \quad \operatorname{tr}(\boldsymbol{\sigma}) = -np - \operatorname{tr}(\mathbf{u}\otimes\mathbf{u}) \quad \text{in} \quad \Omega. \tag{2.6}$$

In particular, the second equation in (2.6) allows us to write the pressure p in terms of the tensor $\boldsymbol{\sigma}$ and the velocity $\mathbf{u}$ as

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

which in turn, together to (2.5), leads us to the equation

$$\boldsymbol{\sigma}^d = \nu\nabla\mathbf{u} - \,(\mathbf{u}\otimes\mathbf{u})^d \quad \text{in} \quad \Omega.$$

On the other hand, from (2.5), and the first equations of (2.4) and (2.6), we easily get the equilibrium equation

$$-\,\mathbf{div}\,\boldsymbol{\sigma} \,=\, \mathbf{f} \quad \text{in} \quad \Omega,$$

which, as we will see later, will allow us to derive our conservative method.

Finally, from (2.7) we observe that the condition $(p,1)_\Omega = 0$, ensuring the uniqueness of solution of problem (2.4), is equivalent to

$$(\operatorname{tr}(\boldsymbol{\sigma}) + \operatorname{tr}(\mathbf{u}\otimes\mathbf{u}), 1)_\Omega \,=\, 0.$$

According to the above, we rewrite equations (2.4) equivalently as follows

$$\begin{aligned} &\boldsymbol{\sigma}^d = \nu\nabla\mathbf{u} - \,(\mathbf{u}\otimes\mathbf{u})^d \quad \text{in} \quad \Omega, && -\,\mathbf{div}\,\boldsymbol{\sigma} \,=\, \mathbf{f} \quad \text{in} \quad \Omega, \\ &\mathbf{u} \,=\, \mathbf{u}_D \quad \text{on} \quad \Gamma, && (\operatorname{tr}(\boldsymbol{\sigma}), 1)_\Omega \,=\, -(\operatorname{tr}(\mathbf{u}\otimes\mathbf{u}), 1)_\Omega, \end{aligned} \tag{2.8}$$

where the unknowns of the system are the tensor $\boldsymbol{\sigma}$ and the velocity $\mathbf{u}$. The pressure p can be easily computed as a postprocess of the solution by using (2.7).

In the sequel we employ the set of equations (2.8) to derive our conservative mixed formulation.

2.3 Derivation of the conservative mixed variational formulation

Here, we derive our conservative mixed problem and define the forms and functionals involved.

First, multiplying the first equation of (2.8) by a tensor $\boldsymbol{\tau}$, living in a suitable space, say $\mathbb{X}$, which will be described next, integrating by parts, utilizing the Dirichlet boundary condition $\mathbf{u} = \mathbf{u}_D$ on Γ, and the identity $\boldsymbol{\sigma}^d : \boldsymbol{\tau} = \boldsymbol{\sigma}^d : \boldsymbol{\tau}^d$, we obtain

$$\frac{1}{\nu}(\boldsymbol{\sigma}^d, \boldsymbol{\tau}^d)_\Omega + (\mathbf{div}\ \boldsymbol{\tau}, \mathbf{u})_\Omega + \frac{1}{\nu}(\mathbf{u}\otimes\mathbf{u}, \boldsymbol{\tau}^d)_\Omega = \langle \boldsymbol{\tau}\mathbf{n}, \mathbf{u}_D\rangle, \quad \forall\, \boldsymbol{\tau} \in \mathbb{X}.$$

In addition, the equilibrium equation $\mathbf{div}\ \boldsymbol{\sigma} = -\mathbf{f}$ is imposed weakly as follows

$$(\mathbf{div}\ \boldsymbol{\sigma}, \mathbf{v})_\Omega = -(\mathbf{f}, \mathbf{v})_\Omega, \quad \forall\, \mathbf{v} \in \mathbf{M},$$

where $\mathbf{M}$ is a suitable space which will be also described next. Then, at first, we have arrived to the following weak problem: Find $(\boldsymbol{\sigma}, \mathbf{u}) \in \mathbb{X}\times\mathbf{M}$, such that $(\mathrm{tr}\,(\boldsymbol{\sigma}), 1)_\Omega = -(\mathrm{tr}\,(\mathbf{u}\otimes\mathbf{u}), 1)_\Omega$, and:

$$\begin{aligned} \frac{1}{\nu}(\boldsymbol{\sigma}^d, \boldsymbol{\tau}^d)_\Omega + (\mathbf{div}\ \boldsymbol{\tau}, \mathbf{u})_\Omega + \frac{1}{\nu}(\mathbf{u}\otimes\mathbf{u}, \boldsymbol{\tau}^d)_\Omega &= \langle \boldsymbol{\tau}\mathbf{n}, \mathbf{u}_D\rangle, \\ (\mathbf{div}\ \boldsymbol{\sigma}, \mathbf{v})_\Omega &= -(\mathbf{f}, \mathbf{v})_\Omega, \end{aligned} \tag{2.9}$$

for all $(\boldsymbol{\tau}, \mathbf{v}) \in \mathbb{X}\times\mathbf{M}$.

Now we turn to specify the spaces $\mathbb{X}$ and $\mathbf{M}$. To that end we notice first that the term $\frac{1}{\nu}(\boldsymbol{\sigma}^d, \boldsymbol{\tau}^d)_\Omega$ is clearly well defined if $\boldsymbol{\sigma}, \boldsymbol{\tau} \in \mathbb{L}^2(\Omega)$. However, if $\boldsymbol{\tau} \in \mathbb{L}^2(\Omega)$ the term $(\mathbf{u}\otimes\mathbf{u}, \boldsymbol{\tau}^d)_\Omega$ forces the velocity $\mathbf{u}$, and consequently the test function $\mathbf{v}$, to live in $\mathbf{L}^4(\Omega)$. Moreover, the latter implies that terms $(\mathbf{div}\ \boldsymbol{\sigma}, \mathbf{v})_\Omega$ and $(\mathbf{div}\,\boldsymbol{\tau}, \mathbf{u})_\Omega$ are well defined if $\mathbf{div}\ \boldsymbol{\sigma}$ and $\mathbf{div}\,\boldsymbol{\tau}$ belong to $\mathbf{L}^{4/3}(\Omega)$. According to the above, and to the definition (2.2), we deduce that the equations of (2.9) make sense if we choose the spaces $\mathbb{X}$ and $\mathbf{M}$ as follows:

$$\mathbb{X} := \mathbb{H}(\mathbf{div}_{4/3}\,; \Omega) \quad \text{and} \quad \mathbf{M} := \mathbf{L}^4(\Omega).$$

Remark 2.1 *We observe that the spaces $\mathbb{X}$ and $\mathbf{M}$ chosen above for the variational problem* (2.16) *are consistent with the spaces of the classical velocity-pressure approach for the Navier-Stokes equations: Given $\mathbf{f} \in \mathbf{L}^2(\Omega)$ and $\mathbf{u}_D \in \mathbf{H}^{1/2}(\Gamma)$, find $(\mathbf{u}, p) \in \mathbf{H}^1(\Omega) \times L^2_0(\Omega)$ such that $\mathbf{u}|_\Gamma = \mathbf{u}_D$ and*

$$\begin{aligned} \nu(\nabla\mathbf{u}, \nabla\mathbf{v})_\Omega + ((\nabla\mathbf{u})\mathbf{u}, \mathbf{v})_\Omega - (p, \mathrm{div}\,\mathbf{v})_\Omega &= (\mathbf{f}, \mathbf{v})_\Omega, \\ (q, \mathrm{div}\,\mathbf{u})_\Omega &= 0, \end{aligned} \tag{2.10}$$

for all $(\mathbf{v}, q) \in \mathbf{H}^1_0(\Omega) \times \mathrm{L}^2_0(\Omega)$. In fact, if $(\mathbf{u}, p) \in \mathbf{H}^1(\Omega) \times \mathrm{L}^2_0(\Omega)$ is the solution of (2.10)*, then clearly $\mathbf{u} \in \mathbf{L}^4(\Omega)$ and*

$$\boldsymbol{\sigma} = \nu\nabla\mathbf{u} - p\mathbb{I} - \mathbf{u}\otimes\mathbf{u} \in \mathbb{L}^2(\Omega). \tag{2.11}$$

In turn, from to the first equation of (2.10) *we have that*

$$(\nu\nabla\mathbf{u} - p\mathbb{I}, \nabla\mathbf{v})_\Omega = -\big((\nabla\mathbf{u})\mathbf{u} - \mathbf{f}, \mathbf{v}\big)_\Omega \quad \forall\, \mathbf{v} \in [C_0^\infty(\Omega)]^n,$$

which in the distributional sense means that $\nu\Delta\mathbf{u} - \nabla p = (\nabla\mathbf{u})\mathbf{u} - \mathbf{f}$ in Ω. Hence, since $\mathbf{u}$ belongs to $\mathbf{L}^4(\Omega)$, it follows that $(\nabla\mathbf{u})\mathbf{u}$ is in $\mathbf{L}^{4/3}(\Omega)$, which implies that

$$\mathbf{div}\,\boldsymbol{\sigma} = \nu\Delta\mathbf{u} - (\nabla\mathbf{u})\mathbf{u} - \nabla p \in \mathbf{L}^{4/3}(\Omega), \tag{2.12}$$

In this way, from (2.11) *and* (2.12) *we find that the definition of the space $\mathbb{H}(\mathbf{div}_{4/3}\,; \Omega)$ for the unknown $\boldsymbol{\sigma}$ is coherent with the classical velocity–pressure formulation of problem* (2.4)*.*

According to the above, defining the forms $\mathbf{a} : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$, $\mathbf{b} : \mathbb{X} \times \mathbf{M} \to \mathbb{R}$ and $\mathbf{c} : \mathbf{M} \times \mathbf{M} \times \mathbb{X} \to \mathbb{R}$ as

$$\mathbf{a}(\boldsymbol{\sigma}, \boldsymbol{\tau}) := \frac{1}{\nu}(\boldsymbol{\sigma}^d, \boldsymbol{\tau}^d)_\Omega \quad \text{and} \quad \mathbf{b}(\boldsymbol{\tau}, \mathbf{v}) := (\mathbf{div}\ \boldsymbol{\tau}, \mathbf{v})_\Omega, \tag{2.13}$$

$$\mathbf{c}(\mathbf{w}; \mathbf{v}, \boldsymbol{\tau}) := \frac{1}{\nu}(\mathbf{w} \otimes \mathbf{v}, \boldsymbol{\tau}^d)_\Omega, \tag{2.14}$$

and the functionals $F \in \mathbb{H}(\mathbf{div}_{4/3}\,; \Omega)'$ and $G \in \mathbf{L}^4(\Omega)'$ as

$$F(\boldsymbol{\tau}) := \langle \boldsymbol{\tau}\mathbf{n}, \mathbf{u}_D \rangle \quad \text{and} \quad G(\mathbf{v}) := -(\mathbf{f}, \mathbf{v})_\Omega, \tag{2.15}$$

which according to Lemma 3.5 below are well defined if the data $\mathbf{f}$ and $\mathbf{u}_D$ belong to $\mathbf{L}^{4/3}(\Omega)$ and $\mathbf{H}^{1/2}(\Gamma)$, respectively, problem (2.9) now reads: Find $(\boldsymbol{\sigma}, \mathbf{u}) \in \mathbb{X} \times \mathbf{M}$, such that $(\mathrm{tr}\,(\boldsymbol{\sigma}), 1)_\Omega = -(\mathrm{tr}\,(\mathbf{u} \otimes \mathbf{u}), 1)_\Omega$, and

$$\begin{aligned} \mathbf{a}(\boldsymbol{\sigma}, \boldsymbol{\tau}) + \mathbf{b}(\boldsymbol{\tau}, \mathbf{u}) + \mathbf{c}(\mathbf{u}; \mathbf{u}, \boldsymbol{\tau}) &= F(\boldsymbol{\tau}) \quad \forall\, \boldsymbol{\tau} \in \mathbb{X}, \\ \mathbf{b}(\boldsymbol{\sigma}, \mathbf{v}) &= G(\mathbf{v}) \quad \forall\, \mathbf{v} \in \mathbf{M}. \end{aligned} \tag{2.16}$$

We end this section by mentioning that, from now on, and for the sake of simplicity, the norms for the spaces $\mathbb{X}$, $\mathbf{M}$ and for the product space $\mathbb{X} \times \mathbf{M}$, will be denoted, respectively by

$$\|\cdot\|_{\mathbb{X}}, \quad \|\cdot\|_{\mathbf{M}} \quad \text{and} \quad \|(\cdot,\cdot)\| = \|\cdot\|_{\mathbb{X}} + \|\cdot\|_{\mathbf{M}}.$$

3 Analysis of the continuous problem

In this section we combine the classical Banach–Nečas–Babuška and Banach fixed–point theorems to prove the well-posedness of (2.16) under a suitable smallness assumption on the data. We begin by introducing the corresponding fixed-point operator.

3.1 The fixed–point operator

Before introducing the associated fixed–point operator, let us first recall that the bilinear form $\mathbf{a}(\cdot,\cdot)$ (cf. (2.13)) is defined in terms of the deviatoric part of the corresponding tensors, and then, an eventual ellipticity of $\mathbf{a}$, which as we will see later is needed to prove the well-definiteness of the associated fixed-point operator, is not feasible in the space $\mathbb{X}$. Therefore, to overcome this drawback we proceed similarly as in [6] and introduce an equivalent version of problem (2.16). To that end, let us define the space

$$\mathbb{X}_0 := \{\boldsymbol{\tau} \in \mathbb{H}(\mathbf{div}_{4/3}\,; \Omega) \,:\, (\mathrm{tr}\,(\boldsymbol{\tau}), 1)_\Omega = 0\},$$

and observe that the following decomposition holds:

$$\mathbb{X} = \mathbb{X}_0 \,\oplus\, P_0(\Omega)\mathbb{I},$$

i.e. $P_0(\Omega)\mathbb{I}$ is a topological supplement for $\mathbb{X}_0$, where $P_0(\Omega)$ is the space of constant polynomials on Ω. More precisely, each $\boldsymbol{\tau} \in \mathbb{X}$ can be decomposed uniquely as:

$$\boldsymbol{\tau} = \boldsymbol{\tau}_0 + c\,\mathbb{I}, \quad \text{with} \quad \boldsymbol{\tau}_0 \in \mathbb{X}_0 \quad \text{and} \quad c := \frac{1}{n\,|\Omega|}(\mathrm{tr}\ \boldsymbol{\tau}, 1)_\Omega \in \mathbb{R}. \tag{3.1}$$

Now, assume that $(\boldsymbol{\sigma},\mathbf{u}) \in \mathbb{X} \times \mathbf{M}$ be a solution of (2.16) and define the tensor

$$\boldsymbol{\sigma}_0 := \boldsymbol{\sigma} + \frac{1}{n|\Omega|}(\mathrm{tr}\,(\mathbf{u}\otimes\mathbf{u}),1)_\Omega \mathbb{I}. \tag{3.2}$$

It is clear that

$$\boldsymbol{\sigma}_0 \in \mathbb{X}_0 \quad \text{if and only if} \quad (\mathrm{tr}\,(\boldsymbol{\sigma}) + \mathrm{tr}\,(\mathbf{u}\otimes\mathbf{u}),1)_\Omega = 0.$$

Then, owing to (3.2) and (3.1), and after simple computations it is not difficult to see that $(\boldsymbol{\sigma}_0,\mathbf{u})$ is a solution to problem: Find $(\boldsymbol{\sigma}_0,\mathbf{u}) \in \mathbb{X}_0 \times \mathbf{M}$, such that

$$\begin{aligned} \mathbf{a}(\boldsymbol{\sigma}_0,\boldsymbol{\tau}) + \mathbf{b}(\boldsymbol{\tau},\mathbf{u}) \,+\, \mathbf{c}(\mathbf{u};\mathbf{u},\boldsymbol{\tau}) \quad &= F(\boldsymbol{\tau}), \\ \mathbf{b}(\boldsymbol{\sigma}_0,\mathbf{v}) \quad &= G(\mathbf{v}), \end{aligned} \tag{3.3}$$

for all $(\boldsymbol{\tau},\mathbf{v}) \in \mathbb{X}_0 \times \mathbf{M}$. Actually, it can be readily seen that problems (2.16) and (3.3) are equivalent. This is established in the following lemma. We omit the proof since it is straightforward.

Lemma 3.1 *If $(\boldsymbol{\sigma},\mathbf{u})$ is a solution of (2.16), then*

$$(\boldsymbol{\sigma}_0,\mathbf{u}) = \left(\boldsymbol{\sigma} + \frac{1}{n|\Omega|}(\mathrm{tr}\,(\mathbf{u}\otimes\mathbf{u}),1)_\Omega\mathbb{I},\mathbf{u}\right) \tag{3.4}$$

is a solution of (3.3). Conversely, if $(\boldsymbol{\sigma}_0,\mathbf{u})$ is a solution of (3.3), then $(\boldsymbol{\sigma},\mathbf{u}) = (\boldsymbol{\sigma}_0 - \frac{1}{n|\Omega|}(\mathrm{tr}\,(\mathbf{u}\otimes\mathbf{u}),1)_\Omega\mathbb{I},\mathbf{u})$ is a solution of (2.16).

Let us now define the fixed-point operator. To do that, let us introduce the bounded set

$$\mathbf{K} := \left\{\mathbf{v} \in \mathbf{M} \,:\, \|\mathbf{v}\|_{\mathbf{M}} \,\leq\, \frac{2}{\gamma}\left(C_F\|\mathbf{u}_D\|_{1/2,\Gamma} + \|\mathbf{f}\|_{\mathbf{L}^{4/3}(\Omega)}\right)\right\}, \tag{3.5}$$

with γ and C_F being the positive constant defined below in (3.29) and (3.24), respectively. Then, we define our fixed-point operator as

$$\mathcal{J} : \mathbf{K} \to \mathbf{K}, \quad \mathbf{w} \to \mathcal{J}(\mathbf{w}) = \mathbf{u}, \tag{3.6}$$

where, given $\mathbf{w} \in \mathbf{K}$, $\mathbf{u}$ is the second component of the solution of the linearized version of problem (3.3): Find $(\boldsymbol{\sigma},\mathbf{u}) \in \mathbb{X}_0 \times \mathbf{M}$, such that

$$\begin{aligned} \mathbf{a}(\boldsymbol{\sigma},\boldsymbol{\tau}) \,+\, \mathbf{b}(\boldsymbol{\tau},\mathbf{u}) \,+\, \mathbf{c}(\mathbf{w};\mathbf{u},\boldsymbol{\tau}) \quad &= F(\boldsymbol{\tau}) \quad \forall\,\boldsymbol{\tau} \in \mathbb{X}_0, \\ \mathbf{b}(\boldsymbol{\sigma},\mathbf{v}) \quad &= G(\mathbf{v}) \quad \forall\,\mathbf{v} \in \mathbf{M}. \end{aligned} \tag{3.7}$$

According to the above, we have the following relations

$$\begin{aligned} \mathcal{J}(\mathbf{u}) = \mathbf{u} \quad &\Leftrightarrow \quad ((\boldsymbol{\sigma}_0,\mathbf{u})) \in \mathbb{X}_0 \times \mathbf{M} \text{ satisfies } (3.3) \\ &\Leftrightarrow \quad (\boldsymbol{\sigma}_0 - \tfrac{1}{n|\Omega|}(\mathrm{tr}\,(\mathbf{u}\otimes\mathbf{u}),1)_\Omega\mathbb{I},\mathbf{u}) \in \mathbb{X} \times \mathbf{M} \text{ satisfies } (2.16). \end{aligned} \tag{3.8}$$

In this way, in establishing the well-posedness of (2.16), or equivalently (3.3), it suffices to prove that $\mathcal{J}$ has a unique fixed-point. Before proceeding with the solvability analysis, we first state the stability of the forms and functionals involved and the well-definiteness of the fixed-point operator $\mathcal{J}$.

3.2 Well–definiteness of $\mathcal{J}$

According to the definition of $\mathcal{J}$ (cf. (3.6)), it is clear that to proving that operator $\mathcal{J}$ is well-defined it suffices to prove that problem (3.7) is well-posed. To do that we first verify the stability properties of the forms and functionals involved. We begin by establishing the continuity of the forms $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$, which can be easily deduced from the Hölder inequality

$$\int_\Omega |fg| \le \|f\|_{\mathrm{L}^p(\Omega)}\|g\|_{\mathrm{L}^q(\Omega)}, \quad \forall\, f \in \mathrm{L}^p(\Omega),\ \forall\, g \in \mathrm{L}^q(\Omega), \quad \text{with} \quad \frac{1}{p}+\frac{1}{q}=1. \tag{3.9}$$

Indeed, using inequality (3.9), we readily deduce that

$$|\mathbf{a}(\boldsymbol{\sigma},\boldsymbol{\tau})| \le \frac{1}{\nu}\|\boldsymbol{\sigma}\|_{\mathbb{X}}\|\boldsymbol{\tau}\|_{\mathbb{X}} \quad \forall\, \boldsymbol{\sigma},\boldsymbol{\tau} \in \mathbb{X}, \tag{3.10}$$

$$|\mathbf{b}(\boldsymbol{\tau},\mathbf{v})| \le \|\boldsymbol{\tau}\|_{\mathbb{X}}\|\mathbf{v}\|_{\mathbf{M}} \quad \forall\, \boldsymbol{\tau} \in \mathbb{X}, \forall\, \mathbf{v} \in \mathbf{M}, \tag{3.11}$$

and

$$|\mathbf{c}(\mathbf{w};\mathbf{v},\boldsymbol{\tau})| \le \frac{1}{\nu}\|\mathbf{w}\|_{\mathbf{M}}\|\mathbf{v}\|_{\mathbf{M}}\|\boldsymbol{\tau}\|_{\mathbb{X}} \quad \forall\, \boldsymbol{\tau} \in \mathbb{X}, \forall\, \mathbf{w},\mathbf{v} \in \mathbf{M}. \tag{3.12}$$

Now we recall the next two classical estimates that will be employed in the upcoming results:

$$\|w\|_{1,\Omega} \le C_P|w|_{1,\Omega} \quad \forall w \in \mathrm{H}^1_0(\Omega) \tag{3.13}$$

and

$$\|w\|_{\mathrm{L}^r(\Omega)} \le C_{Sob}\|w\|_{1,\Omega} \quad \forall\, w \in \mathrm{H}^1(\Omega)\,, \quad \text{for } r \ge 1 \text{ if } n=2 \text{ or } r \in [1,6] \text{ if } n=3, \tag{3.14}$$

with $C_P > 0$ and $C_{Sob} > 0$ depending only on $|\Omega|$. (3.13) can be deduced from [28, Theorem 1.3.3] whereas (3.14) can be found in [28, Theorem 1.3.3]. In turn, after a slight modification of the proof of [17, Lemma 2.3] it is not difficult to see that the following result holds.

Lemma 3.2 *There exists $C_d > 0$, such that*

$$C_d\|\boldsymbol{\tau}\|^2_{0,\Omega} \;\le\; \|\boldsymbol{\tau}^d\|^2_{0,\Omega} \,+\, \|\mathbf{div}\,\boldsymbol{\tau}\|^2_{\mathbf{L}^{4/3}(\Omega)} \qquad \forall\, \boldsymbol{\tau} \,\in\, \mathbb{X}_0. \tag{3.15}$$

Proof. Given $\boldsymbol{\tau} \in \mathbb{X}_0$, and denoting by $\mathbf{z}$ the unique element in $\{\mathbf{w} \in \mathbf{H}^1_0(\Omega) \,:\, \mathrm{div}\,\mathbf{w} = 0 \quad \text{in} \quad \Omega\}^\perp$ such that

$$\mathrm{div}\,\mathbf{z} = \mathrm{tr}\,(\boldsymbol{\tau}) \quad \text{and} \quad \|\mathbf{z}\|_{1,\Omega} \le C\|\mathrm{tr}\,(\boldsymbol{\tau})\|_{0,\Omega}, \tag{3.16}$$

using (3.9), (3.14) with $p = r = 4$ and $q = 4/3$, (3.16) and following the same steps employed in the proof of [17, Lemma 2.3], it can be readily seen that

$$\|\mathrm{tr}\,(\boldsymbol{\tau})\|^2_{0,\Omega} \le nC\|\mathbf{z}\|_{1,\Omega}\left\{\|\mathbf{div}\,\boldsymbol{\tau}\|^2_{\mathbf{L}^{4/3}(\Omega)} + \|\boldsymbol{\tau}^d\|^2_{0,\Omega}\right\}^{1/2},$$

which together to the definition of $\boldsymbol{\tau}^d$ (cf. (2.1)) implies the result. □

We now let $\mathbb{V}$ be the kernel of $\mathbf{b}$, that is

$$\mathbb{V} := \{\boldsymbol{\tau} \in \mathbb{X}_0 : \mathbf{b}(\boldsymbol{\tau},\mathbf{v}) = 0, \forall\, \mathbf{v} \in \mathbf{M}\} = \{\boldsymbol{\tau} \in \mathbb{X}_0 : (\mathbf{div}\;\boldsymbol{\tau},\mathbf{v})_\Omega = 0, \forall\, \mathbf{v} \in \mathbf{M}\}\,.$$

It is clear that $\mathbb{V}$ can be characterized as follows

$$\mathbb{V} = \{\boldsymbol{\tau} \in \mathbb{X}_0 : \quad \mathbf{div}\;\boldsymbol{\tau} = 0 \text{ in } \Omega\}\,.$$

The following lemma establishes the ellipticity of $\mathbf{a}$ on $\mathbb{V}$, whose proof is a direct consequence of (3.15).

Lemma 3.3 *There holds,*

$$\mathbf{a}(\boldsymbol{\tau},\boldsymbol{\tau}) \;\geq\; \alpha\|\boldsymbol{\tau}\|_{\mathbb{X}}^2 \qquad \forall\,\boldsymbol{\tau}\in\mathbb{V}, \tag{3.17}$$

with $\alpha := C_d/\nu$.

Now we provide the corresponding inf-sup condition of the bilinear form $\mathbf{b}$

Lemma 3.4 *There holds,*

$$\sup_{\mathbf{0}\neq\boldsymbol{\tau}\in\mathbb{X}_0}\frac{\mathbf{b}(\boldsymbol{\tau},\mathbf{v})}{\|\boldsymbol{\tau}\|_{\mathbb{X}}} \geq \beta\|\mathbf{v}\|_{\mathbf{M}} \qquad \forall\,\mathbf{v}\,\in\,\mathbf{M}, \tag{3.18}$$

with

$$\beta := (n+nC_P^2C_{Sob}^2)^{-1/2}.$$

Proof. We proceed similarly to the proof of [7, Lemma 2.1]. In fact, given $\mathbf{v}\in\mathbf{M}$, we let $\mathbf{h}(\mathbf{v}) := |\mathbf{v}|^3\mathbf{sgn}(\mathbf{v})$ and observe that

$$(|\mathbf{h}(\mathbf{v})|^{4/3},1)_\Omega = (|\mathbf{v}|^4|\mathbf{sgn}(\mathbf{v})|^{4/3},1)_\Omega = n^{2/3}(|\mathbf{v}|^4,1)_\Omega < +\infty, \tag{3.19}$$

which implies that $\mathbf{h}(\mathbf{v})\in\mathbf{L}^{4/3}(\Omega)$. Then, defining $\tilde{\boldsymbol{\tau}} = -\nabla\mathbf{z} + \dfrac{1}{n|\Omega|}(\operatorname{div}\mathbf{z},1)\mathbb{I}\in\mathbb{L}^2(\Omega)$, with $\mathbf{z}\in\mathbf{H}_0^1(\Omega)$ being the unique solution of the variational problem

$$(\nabla\,\mathbf{z},\nabla\,\mathbf{w})_\Omega = (\mathbf{h}(\mathbf{v}),\mathbf{w})_\Omega \qquad \forall\,\mathbf{w}\in\mathbf{H}_0^1(\Omega), \tag{3.20}$$

it readily follows that

$$\mathbf{div}\,\tilde{\boldsymbol{\tau}} = \mathbf{h}(\mathbf{v})\in\mathbf{L}^{4/3}(\Omega), \qquad (\operatorname{tr}(\tilde{\boldsymbol{\tau}}),1)_\Omega = 0, \tag{3.21}$$

and, consequently, $\tilde{\boldsymbol{\tau}}\in\mathbb{X}_0$. Moreover, from (3.19) we have

$$\|\mathbf{div}\,\tilde{\boldsymbol{\tau}}\|_{\mathbf{L}^{4/3}(\Omega)} = \|\mathbf{h}(\mathbf{v})\|_{\mathbf{L}^{4/3}(\Omega)} = \sqrt{n}\|\mathbf{v}\|_{\mathbf{M}}^3. \tag{3.22}$$

On the other hand, from (3.20) with $\mathbf{w}=\mathbf{z}$, (3.22) and the Hölder inequality (3.9), we obtain

$$|\mathbf{z}|_{1,\Omega}^2 \;=\; (\nabla\,\mathbf{z},\nabla\,\mathbf{z})_\Omega \;=\; (\mathbf{h}(\mathbf{v}),\mathbf{z})_\Omega \;\leq\; \|\mathbf{h}(\mathbf{v})\|_{\mathbf{L}^{4/3}(\Omega)}\|\mathbf{z}\|_{\mathbf{L}^4(\Omega)} \;=\; \sqrt{n}\|\mathbf{v}\|_{\mathbf{M}}^3\|\mathbf{z}\|_{\mathbf{M}},$$

which together to (3.14) with $r=4$ and (3.13), implies

$$\|\mathbf{z}\|_{1,\Omega} \leq \sqrt{n}C_{sob}C_P\|\mathbf{v}\|_{\mathbf{M}}^3.$$

From the latter, it readily follows that

$$\|\tilde{\boldsymbol{\tau}}\|_{0,\Omega} = \left\{|\mathbf{z}|_{1,\Omega}^2 - \frac{1}{n|\Omega|}(\operatorname{div}\mathbf{z},1)_\Omega^2\right\}^{1/2} \leq \|\mathbf{z}\|_{1,\Omega} \leq \sqrt{n}C_{sob}C_P\|\mathbf{v}\|_{\mathbf{M}}^3,$$

which combined with (3.22), yields

$$\|\tilde{\boldsymbol{\tau}}\|_{\mathbb{X}} \leq \beta^{-1}\|\mathbf{v}\|_{\mathbf{M}}^3, \tag{3.23}$$

with $\beta := (n+nC_P^2C_{Sob}^2)^{-1/2}$. In this way, from (2.3), (3.21), and (3.23), it follows that

$$\sup_{\mathbf{0}\neq\boldsymbol{\tau}\in\mathbb{X}_0}\frac{\mathbf{b}(\boldsymbol{\tau},\mathbf{v})}{\|\boldsymbol{\tau}\|_{\mathbb{X}}} \geq \frac{\mathbf{b}(\tilde{\boldsymbol{\tau}},\mathbf{v})}{\|\tilde{\boldsymbol{\tau}}\|_{\mathbb{X}}} \geq \beta\frac{(\mathbf{v},\mathbf{h}(\mathbf{v}))_\Omega}{\|\mathbf{v}\|_{\mathbf{M}}^3} = \beta\frac{(|\mathbf{v}|^3\mathbf{sgn}(\mathbf{v}),\mathbf{v})_\Omega}{\|\mathbf{v}\|_{\mathbf{M}}^3} \geq \beta\|\mathbf{v}\|_{\mathbf{M}},$$

which concludes the proof. $\square$

Finally, we establish the continuity of the functionals F and G.

Lemma 3.5 *There hold:*

$$|F(\boldsymbol{\tau})| = |\langle \boldsymbol{\tau}\mathbf{n}, \mathbf{u}_D\rangle| \le C_F \|\mathbf{u}_D\|_{1/2,\Gamma} \|\boldsymbol{\tau}\|_{\mathbb{X}} \tag{3.24}$$

and

$$|G(\mathbf{v})| = |(\mathbf{f}, \mathbf{v})_\Omega| \le \|\mathbf{f}\|_{\mathbf{L}^{4/3}(\Omega)} \|\mathbf{v}\|_{\mathbf{M}}, \tag{3.25}$$

where C_F is a positive constant depending on C_{Sob} (cf. (3.14)).

Proof. The proof of (3.25) is straightforward. Now, for (3.24) let us first recall that given $\boldsymbol{\tau} \in \mathbb{H}(\mathbf{div}\,;\Omega)$, the normal trace $\boldsymbol{\tau}\mathbf{n}$ is defined as the functional in $\mathbf{H}^{-1/2}(\Gamma)$ given by (see e.g. [17, Section 1.3.4])

$$\langle \boldsymbol{\tau}\mathbf{n}, \boldsymbol{\xi}\rangle = (\boldsymbol{\tau}, \nabla\tilde{\gamma}_0^{-1}(\boldsymbol{\xi}))_\Omega + (\tilde{\gamma}_0^{-1}(\boldsymbol{\xi}), \mathbf{div}\ \boldsymbol{\tau})_\Omega \qquad \forall\, \boldsymbol{\xi} \in \mathbf{H}^{1/2}(\Gamma), \tag{3.26}$$

where $\tilde{\gamma}_0^{-1} : \mathbf{H}^{1/2}(\Gamma) \to [\mathbf{H}_0^1(\Omega)]^\perp$ is the right inverse of the well known trace operator $\gamma_0 : \mathbf{H}^1(\Omega) \to \mathbf{H}^{1/2}(\Gamma)$. Then, since $\tilde{\gamma}_0^{-1}(\boldsymbol{\xi}) \in \mathbf{H}^1(\Omega)$, owing to the Sobolev embedding $\mathbf{H}^1(\Omega) \subset \mathbf{L}^4(\Omega)$, the last term in (3.26) is still well defined if $\mathbf{div}\ \boldsymbol{\tau} \in \mathbf{L}^{4/3}(\Omega)$. This implies that $\boldsymbol{\tau}\mathbf{n} \in \mathbf{H}^{-1/2}(\Gamma)$ for all $\boldsymbol{\tau} \in \mathbb{H}(\mathbf{div}_{4/3}\,;\Omega)$, and then F (cf. (2.15)) is well defined. Moreover, using (3.14) it readily follows that there exists $C_F > 0$, such that

$$|\langle \boldsymbol{\tau}\mathbf{n}, \boldsymbol{\xi}\rangle| \le C_F \|\boldsymbol{\tau}\|_{\mathbb{X}} \|\boldsymbol{\xi}\|_{1/2,\Gamma}, \qquad \forall\, \boldsymbol{\tau} \in \mathbb{H}(\mathbf{div}_{4/3}\,;\Omega),\ \forall\, \boldsymbol{\xi} \in \mathbf{H}^{1/2}(\Gamma),$$

which implies (3.24). □

Let us now define the bilinear form $\mathbf{A} : (\mathbb{X} \times \mathbf{M}) \times (\mathbb{X} \times \mathbf{M}) \to \mathbb{R}$ given by

$$\mathbf{A}((\boldsymbol{\sigma}, \mathbf{u}), (\boldsymbol{\tau}, \mathbf{v})) := \mathbf{a}(\boldsymbol{\sigma}, \boldsymbol{\tau}) \,+\, \mathbf{b}(\boldsymbol{\tau}, \mathbf{u}) \,+\, \mathbf{b}(\boldsymbol{\sigma}, \mathbf{v}). \tag{3.27}$$

Owing to (3.10) and (3.11), it is clear that $\mathbf{A}$ is bounded. Moreover, from (3.17), (3.18) and [11, Proposition 2.36] it is not difficult to see that the following inf-sup condition holds:

$$\sup_{\mathbf{0}\neq(\boldsymbol{\tau},\mathbf{v})\in\mathbb{X}_0\times\mathbf{M}} \frac{\mathbf{A}((\boldsymbol{\zeta}, \mathbf{z}), (\boldsymbol{\tau}, \mathbf{v}))}{\|(\boldsymbol{\tau}, \mathbf{v})\|} \ge \gamma \|(\boldsymbol{\zeta}, \mathbf{z})\| \qquad \forall\, (\boldsymbol{\zeta}, \mathbf{z}) \,\in\, \mathbb{X}_0 \times \mathbf{M}, \tag{3.28}$$

with

$$\gamma := \frac{C_d \beta \min\{1, \nu\beta\}}{2(C_d + 1)(\nu\beta + 1)}. \tag{3.29}$$

Now, we are in position of establishing the well-definiteness of $\mathcal{J}$.

Theorem 3.6 *Assume that*

$$\frac{4}{\nu\gamma^2}\Big(C_F \|\mathbf{u}_D\|_{1/2,\Gamma} + \|\mathbf{f}\|_{\mathbf{L}^{4/3}(\Omega)}\Big) \,\le\, 1. \tag{3.30}$$

Then, given $\mathbf{w} \in \mathbf{K}$, there exists a unique $\mathbf{u} \in \mathbf{K}$ such that $\mathcal{J}(\mathbf{w}) = \mathbf{u}$.

Proof. Given $\mathbf{w} \in \mathbf{K}$, we begin by defining the bilinear form:

$$\mathbf{A}_{\mathbf{w}}((\boldsymbol{\sigma}, \mathbf{u}), (\boldsymbol{\tau}, \mathbf{v})) := \mathbf{A}((\boldsymbol{\sigma}, \mathbf{u}), (\boldsymbol{\tau}, \mathbf{v})) \,+\, \mathbf{c}(\mathbf{w}; \mathbf{u}, \boldsymbol{\tau})\,, \tag{3.31}$$

where $\mathbf{A}$ and $\mathbf{c}$ are the forms defined in (3.27) and (2.14) respectively, that is

$$\mathbf{A}_{\mathbf{w}}((\boldsymbol{\sigma},\mathbf{u}),(\boldsymbol{\tau},\mathbf{v})) = \mathbf{a}(\boldsymbol{\sigma},\boldsymbol{\tau}) + \mathbf{b}(\boldsymbol{\tau},\mathbf{u}) + \mathbf{b}(\boldsymbol{\sigma},\mathbf{v}) + \mathbf{c}(\mathbf{w};\mathbf{u},\boldsymbol{\tau}).$$

Then, evidently problem (3.7) can be rewritten equivalently as: Find $(\boldsymbol{\sigma},\mathbf{u}) \in \mathbb{X}_0 \times \mathbf{M}$, such that

$$\mathbf{A}_{\mathbf{w}}((\boldsymbol{\sigma},\mathbf{u}),(\boldsymbol{\tau},\mathbf{v})) \;=\; F(\boldsymbol{\tau}) + G(\mathbf{v}) \qquad \forall\,(\boldsymbol{\tau},\mathbf{v}) \in \mathbb{X}_0 \times \mathbf{M}. \tag{3.32}$$

Therefore, to prove the well-definiteness of $\mathcal{J}$, in the sequel we equivalently prove that problem (3.32) is well-posed by means of the Banach–Nečas–Babuška theorem (see, for instance [11, Theorem 2.6]).

First, given $(\boldsymbol{\zeta},\mathbf{z}),(\hat{\boldsymbol{\tau}},\hat{\mathbf{v}}) \in \mathbb{X}_0 \times \mathbf{M}$ with $(\hat{\boldsymbol{\tau}},\hat{\mathbf{v}}) \neq \mathbf{0}$, from (3.12) we observe that

$$\begin{aligned}\sup_{\mathbf{0}\neq(\boldsymbol{\tau},\mathbf{v})\in\mathbb{X}_0\times\mathbf{M}} \frac{\mathbf{A}_{\mathbf{w}}((\boldsymbol{\zeta},\mathbf{z}),(\boldsymbol{\tau},\mathbf{v}))}{\|(\boldsymbol{\tau},\mathbf{v})\|} \;&\geq\; \frac{|\mathbf{A}((\boldsymbol{\zeta},\mathbf{z}),(\hat{\boldsymbol{\tau}},\hat{\mathbf{v}}))|}{\|(\hat{\boldsymbol{\tau}},\hat{\mathbf{v}})\|} - \frac{|\mathbf{c}(\mathbf{w};\mathbf{z},\hat{\boldsymbol{\tau}})|}{\|(\hat{\boldsymbol{\tau}},\hat{\mathbf{v}})\|}\\[2ex] &\geq\; \frac{|\mathbf{A}((\boldsymbol{\zeta},\mathbf{z}),(\hat{\boldsymbol{\tau}},\hat{\mathbf{v}}))|}{\|(\hat{\boldsymbol{\tau}},\hat{\mathbf{v}})\|} - \frac{1}{\nu}\|\mathbf{w}\|_{\mathbf{M}}\|(\boldsymbol{\zeta},\mathbf{z})\|\end{aligned}$$

which together to (3.28) and the fact that $(\hat{\boldsymbol{\tau}},\hat{\mathbf{v}})$ is arbitrary, implies

$$\sup_{\mathbf{0}\neq(\boldsymbol{\tau},\mathbf{v})\in\mathbb{X}_0\times\mathbf{M}} \frac{\mathbf{A}_{\mathbf{w}}((\boldsymbol{\zeta},\mathbf{z}),(\boldsymbol{\tau},\mathbf{v}))}{\|(\boldsymbol{\tau},\mathbf{v})\|} \geq \left(\gamma - \frac{1}{\nu}\|\mathbf{w}\|_{\mathbf{M}}\right)\|(\boldsymbol{\zeta},\mathbf{z})\|. \tag{3.33}$$

Hence, from the definition of set $\mathbf{K}$ (cf. (3.5)), and assumption (3.30), we easily get

$$\frac{1}{\nu}\|\mathbf{w}\|_{\mathbf{M}} \;\leq\; \frac{2}{\nu\gamma}\left(C_F\|\mathbf{u}_D\|_{1/2,\Gamma} + \|\mathbf{f}\|_{\mathbf{L}^{4/3}(\Omega)}\right) \leq \frac{\gamma}{2} \tag{3.34}$$

and then, combining (3.33) and (3.34), we obtain

$$\sup_{\mathbf{0}\neq(\boldsymbol{\tau},\mathbf{v})\in\mathbb{X}_0\times\mathbf{M}} \frac{\mathbf{A}_{\mathbf{w}}((\boldsymbol{\zeta},\mathbf{z}),(\boldsymbol{\tau},\mathbf{v}))}{\|(\boldsymbol{\tau},\mathbf{v})\|} \geq \frac{\gamma}{2}\|(\boldsymbol{\zeta},\mathbf{z})\| \quad \forall\,(\boldsymbol{\zeta},\mathbf{z}) \in \mathbb{X}_0 \times \mathbf{M}. \tag{3.35}$$

On the other hand, for a given $(\boldsymbol{\zeta},\mathbf{z}) \in \mathbb{X}_0 \times \mathbf{M}$, we observe that

$$\begin{aligned}\sup_{(\boldsymbol{\tau},\mathbf{v})\in\mathbb{X}_0\times\mathbf{M}} \mathbf{A}_{\mathbf{w}}((\boldsymbol{\tau},\mathbf{v}),(\boldsymbol{\zeta},\mathbf{z})) \;&\geq\; \sup_{\mathbf{0}\neq(\boldsymbol{\tau},\mathbf{v})\in\mathbb{X}_0\times\mathbf{M}} \frac{\mathbf{A}_{\mathbf{w}}((\boldsymbol{\tau},\mathbf{v}),(\boldsymbol{\zeta},\mathbf{z}))}{\|(\boldsymbol{\tau},\mathbf{v})\|}\\ &=\; \sup_{\mathbf{0}\neq(\boldsymbol{\tau},\mathbf{v})\in\mathbb{X}_0\times\mathbf{M}} \frac{\mathbf{A}((\boldsymbol{\tau},\mathbf{v}),(\boldsymbol{\zeta},\mathbf{z})) + \mathbf{c}(\mathbf{w};\mathbf{v},\boldsymbol{\zeta})}{\|(\boldsymbol{\tau},\mathbf{v})\|},\end{aligned}$$

which together to (3.12) implies

$$\sup_{(\boldsymbol{\tau},\mathbf{v})\in\mathbb{X}_0\times\mathbf{M}} \mathbf{A}_{\mathbf{w}}((\boldsymbol{\tau},\mathbf{v}),(\boldsymbol{\zeta},\mathbf{z})) \;\geq\; \sup_{\mathbf{0}\neq(\boldsymbol{\tau},\mathbf{v})\in\mathbb{X}_0\times\mathbf{M}} \frac{\mathbf{A}((\boldsymbol{\tau},\mathbf{v}),(\boldsymbol{\zeta},\mathbf{z}))}{\|(\boldsymbol{\tau},\mathbf{v})\|} - \frac{1}{\nu}\|\mathbf{w}\|_{\mathbf{M}}\|(\boldsymbol{\zeta},\mathbf{z})\|. \tag{3.36}$$

Therefore, using the fact that $\mathbf{A}(\cdot,\cdot)$ is symmetric, from (3.28) and (3.36) we obtain

$$\sup_{(\boldsymbol{\tau},\mathbf{v})\in\mathbb{X}_0\times\mathbf{M}} \mathbf{A}_{\mathbf{w}}((\boldsymbol{\tau},\mathbf{v}),(\boldsymbol{\zeta},\mathbf{z})) \;\geq\; \gamma\|(\boldsymbol{\zeta},\mathbf{z})\| - \frac{1}{\nu}\|\mathbf{w}\|_{\mathbf{M}}\|(\boldsymbol{\zeta},\mathbf{z})\|,$$

which combined with (3.34), yields

$$\sup_{(\boldsymbol{\tau},\mathbf{v})\in\mathbb{X}_0\times\mathbf{M}} \mathbf{A}_{\mathbf{w}}((\boldsymbol{\tau},\mathbf{v}),(\boldsymbol{\zeta},\mathbf{z})) \,\geq\, \frac{\gamma}{2}\|(\boldsymbol{\zeta},\mathbf{z})\| > 0 \quad \forall\,(\boldsymbol{\zeta},\mathbf{z})\in\mathbb{X}_0\times\mathbf{M},\ (\boldsymbol{\zeta},\mathbf{z})\neq\mathbf{0}. \tag{3.37}$$

In this way, from (3.35) and (3.37) we obtain that $\mathbf{A}_{\mathbf{w}}(\cdot,\cdot)$ satisfies the hypotheses of the Banach–Nečas–Babuška theorem (cf. [11, Theorem 2.6]), which allows us to conclude the existence of a unique $(\boldsymbol{\sigma},\mathbf{u})\in\mathbb{X}_0\times\mathbf{M}$ solution to (3.7), or equivalently, the existence of a unique $\mathbf{u}\in\mathbf{M}$ such that $\mathcal{J}(\mathbf{w})=\mathbf{u}$. Finally, from (3.35) with $(\boldsymbol{\zeta},\mathbf{z})=(\boldsymbol{\sigma},\mathbf{u})$ and (3.32) we readily obtain that

$$\|\mathbf{u}\|_{\mathbf{M}} \,\leq\, \|(\boldsymbol{\sigma},\mathbf{u})\| \leq \frac{2}{\gamma}\left(C_F\|\mathbf{u}_D\|_{1/2,\Gamma} + \|\mathbf{f}\|_{\mathbf{L}^{4/3}(\Omega)}\right),$$

which implies that $\mathbf{u}$ belongs to $\mathbf{K}$ and concludes the proof. □

3.3 Well-posedness of the continuous problem

Here we provide the main result of this section, namely, the existence and uniqueness of solution of problem (2.16). This result is established in the following theorem.

Theorem 3.7 *Let $\mathbf{f}\in\mathbf{L}^{4/3}(\Omega)$ and $\mathbf{u}_D\in\mathbf{H}^{1/2}(\Gamma)$ such that*

$$\frac{4}{\nu\gamma^2}\left(C_F\|\mathbf{u}_D\|_{1/2,\Gamma} \,+\, \|\mathbf{f}\|_{\mathbf{L}^{4/3}(\Omega)}\right) \,<\, 1. \tag{3.38}$$

Then, there exists a unique $(\boldsymbol{\sigma},\mathbf{u})\in\mathbb{X}\times\mathbf{M}$ solution to (2.16)*. In addition, there exists $C>0$ such that*

$$\|\mathbf{u}\|_{\mathbf{M}} + \|\boldsymbol{\sigma}\|_{\mathbb{X}} \,\leq\, C\left(\|\mathbf{u}_D\|_{1/2,\Gamma} \,+\, \|\mathbf{f}\|_{\mathbf{L}^{4/3}(\Omega)}\right). \tag{3.39}$$

Proof. As mentioned before, and according to the relations given in (3.8), to proving the well-posedness of (2.16) we equivalently prove that $\mathcal{J}$ possess a unique fixed-point in $\mathbf{K}$ by means of the classical Banach's fixed point theorem. Consequently, in what follows we prove that $\mathcal{J}$ is a contraction mapping.

We start by noticing that assumption (3.38) ensures (see Theorem 3.6) that $\mathcal{J}$ is well–defined. Now, let $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{u}_1$, $\mathbf{u}_2\in\mathbf{K}$, be such that $\mathbf{u}_1=\mathcal{J}(\mathbf{w}_1)$ and $\mathbf{u}_2=\mathcal{J}(\mathbf{w}_2)$. From the definition of $\mathcal{J}$ and (3.32), it follows that there exist unique $\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2\in\mathbb{X}_0$, such that for all $(\boldsymbol{\tau},\mathbf{v})\in\mathbb{X}_0\times\mathbf{M}$, there hold

$$\mathbf{A}_{\mathbf{w}_1}((\boldsymbol{\sigma}_1,\mathbf{u}_1),(\boldsymbol{\tau},\mathbf{v})) \,=\, F(\boldsymbol{\tau})+G(\mathbf{v}), \quad \text{and} \quad \mathbf{A}_{\mathbf{w}_2}((\boldsymbol{\sigma}_2,\mathbf{u}_2),(\boldsymbol{\tau},\mathbf{v})) \,=\, F(\boldsymbol{\tau})+G(\mathbf{v}).$$

Then, subtracting both equations, adding and subtracting suitable terms, and recalling the definition of $\mathbf{A}_{\mathbf{w}}$ in (3.31), we easily arrive at

$$\mathbf{A}_{\mathbf{w}_1}((\boldsymbol{\sigma}_1-\boldsymbol{\sigma}_2,\mathbf{u}_1-\mathbf{u}_2),(\boldsymbol{\tau},\mathbf{v})) \,=\, -\mathbf{c}(\mathbf{w}_1-\mathbf{w}_2;\mathbf{u}_2,\boldsymbol{\tau}) \qquad \forall\,(\boldsymbol{\tau},\mathbf{v})\in\mathbb{X}_0\times\mathbf{M}.$$

Therefore, recalling that $\mathbf{w}_1\in\mathbf{K}$, from the latter identity and from (3.35) and (3.12), we obtain

$$\begin{aligned}
\frac{\gamma}{2}\|\mathbf{u}_1-\mathbf{u}_2\|_{\mathbf{M}} \;&\leq \sup_{\mathbf{0}\neq(\boldsymbol{\tau},\mathbf{v})\in\mathbb{X}_0\times\mathbf{M}} \frac{\mathbf{A}_{\mathbf{w}_1}((\boldsymbol{\sigma}_1-\boldsymbol{\sigma}_2,\mathbf{u}_1-\mathbf{u}_2),(\boldsymbol{\tau},\mathbf{v}))}{\|(\boldsymbol{\tau},\mathbf{v})\|} \\
&= \sup_{\mathbf{0}\neq(\boldsymbol{\tau},\mathbf{v})\in\mathbb{X}_0\times\mathbf{M}} \frac{-\mathbf{c}(\mathbf{w}_1-\mathbf{w}_2;\mathbf{u}_2,\boldsymbol{\tau})}{\|(\boldsymbol{\tau},\mathbf{v})\|} \\
&\leq \frac{1}{\nu}\|\mathbf{w}_1-\mathbf{w}_2\|_{\mathbf{M}}\|\mathbf{u}_2\|_{\mathbf{M}},
\end{aligned}$$

which together to the fact that $\mathbf{u}_2 \in \mathbf{K}$, implies

$$\|\mathbf{u}_1 - \mathbf{u}_2\|_{\mathbf{M}} \,\le\, \frac{4}{\nu\gamma^2}\left(C_F\|\mathbf{u}_D\|_{\mathbf{H}^{1/2}(\Gamma)} \,+\, \|\mathbf{f}\|_{\mathbf{L}^{4/3}(\Omega)}\right)\|\mathbf{w}_1 - \mathbf{w}_2\|_{\mathbf{M}}.$$

The latter and assumption (3.38) readily imply that $\mathcal{J}$ is a contraction mapping.

Now, to derive the estimate (3.39), let $\mathbf{u} \in \mathbf{K}$ be the unique fixed-point of $\mathcal{J}$ and let $\boldsymbol{\sigma}_0 \in \mathbb{X}_0$ be such that $(\boldsymbol{\sigma}_0, \mathbf{u}) \in \mathbb{X}_0 \times \mathbf{M}$ is the unique solution of (3.3). According to the definition of $\mathbf{K}$, evidently $\mathbf{u}$ satisfies

$$\|\mathbf{u}\|_{\mathbf{M}} \,\le\, \frac{2}{\gamma}\left(C_F\|\mathbf{u}_D\|_{1/2,\Gamma} \,+\, \|\mathbf{f}\|_{\mathbf{L}^{4/3}(\Omega)}\right). \tag{3.40}$$

In turn, by applying (3.35) to $\mathbf{A}_{\mathbf{u}}$ with $(\boldsymbol{\zeta}, \mathbf{z}) = (\boldsymbol{\sigma}_0, \mathbf{u})$, recalling again the definition of $\mathbf{A}_{\mathbf{u}}$ in (3.31), and using the fact that $(\boldsymbol{\sigma}_0, \mathbf{u})$ satisfies (3.3), we obtain

$$\|\boldsymbol{\sigma}_0\|_{\mathbb{X}} \,\le \|(\boldsymbol{\sigma}_0, \mathbf{u})\| \,\le \frac{2}{\gamma} \sup_{\mathbf{0}\neq(\boldsymbol{\tau},\mathbf{v})\in\mathbb{X}_0\times\mathbf{M}} \frac{\mathbf{A}_{\mathbf{u}}((\boldsymbol{\sigma}_0, \mathbf{u}), (\boldsymbol{\tau}, \mathbf{v}))}{\|(\boldsymbol{\tau}, \mathbf{v})\|} = \frac{2}{\gamma} \sup_{\mathbf{0}\neq(\boldsymbol{\tau},\mathbf{v})\in\mathbb{X}_0\times\mathbf{M}} \frac{F(\boldsymbol{\tau}) + G(\mathbf{v})}{\|(\boldsymbol{\tau}, \mathbf{v})\|},$$

thus

$$\|\boldsymbol{\sigma}_0\|_{\mathbb{X}} \,\le\, \frac{2}{\gamma}\left(C_F\|\mathbf{u}_D\|_{\mathbf{H}^{1/2}(\Gamma)} \,+\, \|\mathbf{f}\|_{\mathbf{L}^{4/3}(\Omega)}\right). \tag{3.41}$$

In this way, since $\boldsymbol{\sigma} = \boldsymbol{\sigma}_0 - \frac{1}{n|\Omega|}(\mathrm{tr}\,(\mathbf{u}\otimes\mathbf{u}), 1)_\Omega \mathbb{I} = \boldsymbol{\sigma}_0 - \frac{1}{n|\Omega|}\|\mathbf{u}\|_{0,\Omega}^2$ (see Lemma 3.1) from (3.40) and (3.41) we readily obtain the result. $\square$

4 Galerkin scheme

In this section we introduce the Galerkin scheme associated to problem (2.16) and study its solvability and convergence. We mention in advance that, as we shall see in the forthcoming subsections, the well-posedness analysis follows straightforwardly by adapting the results derived for the continuous problem to the discrete case, reason why most of the details are omitted.

4.1 The discrete problem

Let $\mathcal{T}_h$ be a regular family of triangulations of the polyhedral region $\overline{\Omega}$ by triangles T in $\mathbb{R}^2$ or tetrahedra in $\mathbb{R}^3$ of diameter h_T such that $\overline{\Omega} = \cup\{T : T \in \mathcal{T}_h\}$ and define $h := \max\{h_T : T \in \mathcal{T}_h\}$. Now, given an integer $l \ge 0$ and a subset S of $\mathbb{R}^n$, we denote by $P_l(S)$ the space of polynomials of total degree at most l defined on S. Hence, for each integer $k \ge 0$ and for each $T \in \mathcal{T}_h$, we define the local Raviart–Thomas space of order k as (see, for instance [1]):

$$\mathbf{RT}_k(T) := [P_k(T)]^n \oplus \widetilde{P}_k(T)\mathbf{x},$$

where $\mathbf{x} := (x_1, \ldots, x_n)^t$ is a generic vector of $\mathbb{R}^n$ and $\widetilde{P}_k(T)$ is the space of polynomials of total degree equal to k defined on T. In this way, defining the finite element subspaces:

$$\mathbb{X}_h := \left\{\boldsymbol{\tau}_h \in \mathbb{X} : \quad \mathbf{c}^{\mathrm{t}}\,\boldsymbol{\tau}_h|_T \in \mathbf{RT}_k(T) \quad \forall\,\mathbf{c} \in \mathbb{R}^n \quad \forall\, T \in \mathcal{T}_h\right\} \,\subseteq\, \mathbb{X},$$

$$\mathbf{M}_h := \{\mathbf{v}_h \in \mathbf{M} \,:\, \mathbf{v}_h|_K \in [P_k(T)]^n, \quad \forall\, T \in \mathcal{T}_h\} \,\subseteq\, \mathbf{M},$$

the Galerkin scheme associated to problem (2.16) reads: Find $(\boldsymbol{\sigma}_h, \mathbf{u}_h) \in \mathbb{X}_h \times \mathbf{M}_h$, such that $(\mathrm{tr}\,(\boldsymbol{\sigma}_h), 1)_\Omega = -(\mathrm{tr}\,(\mathbf{u}_h \otimes \mathbf{u}_h), 1)_\Omega$, and

$$
\begin{aligned}
\mathbf{a}(\boldsymbol{\sigma}_h, \boldsymbol{\tau}_h) + \mathbf{b}(\boldsymbol{\tau}_h, \mathbf{u}_h) + \mathbf{c}(\mathbf{u}_h; \mathbf{u}_h, \boldsymbol{\tau}_h) &= F(\boldsymbol{\tau}_h) \quad \forall\, \boldsymbol{\tau}_h \in \mathbb{X}_h, \\
\mathbf{b}(\boldsymbol{\sigma}_h, \mathbf{v}_h) &= G(\mathbf{v}_h) \quad \forall\, \mathbf{v}_h \in \mathbf{M}_h,
\end{aligned}
\tag{4.1}
$$

where the forms $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$, as well as the functionals F and G are defined in (2.13), (2.14) and (2.15). In turn, analogously to the continuous case, we observe that

$$
\mathbb{X}_h = \mathbb{X}_{h,0} \oplus P_0(\Omega)\mathbb{I} \qquad \text{with} \qquad \mathbb{X}_{h,0} = \mathbb{X}_h \cap \mathbb{X}_0,
$$

and introduce the problem: Find $(\boldsymbol{\sigma}_{h,0}, \mathbf{u}_h) \in \mathbb{X}_{h,0} \times \mathbf{M}_h$, such that

$$
\begin{aligned}
\mathbf{a}(\boldsymbol{\sigma}_{h,0}, \boldsymbol{\tau}_h) + \mathbf{b}(\boldsymbol{\tau}_h, \mathbf{u}_h) + \mathbf{c}(\mathbf{u}_h; \mathbf{u}_h, \boldsymbol{\tau}_h) &= F(\boldsymbol{\tau}_h) \quad \forall\, \boldsymbol{\tau}_h \in \mathbb{X}_{h,0}, \\
\mathbf{b}(\boldsymbol{\sigma}_{h,0}, \mathbf{v}_h) &= G(\mathbf{v}_h) \quad \forall\, \mathbf{v}_h \in \mathbf{M}_h,
\end{aligned}
\tag{4.2}
$$

which is equivalent to (4.1) in the sense of the following lemma, whose proof is straightforward.

Lemma 4.1 *If $(\boldsymbol{\sigma}_h, \mathbf{u}_h)$ is a solution of* (4.1)*, then*

$$
(\boldsymbol{\sigma}_{h,0}, \mathbf{u}_h) = \left(\boldsymbol{\sigma}_h + \frac{1}{n|\Omega|}(\mathrm{tr}\,(\mathbf{u}_h \otimes \mathbf{u}_h), 1)_\Omega \mathbb{I}, \mathbf{u}_h\right) \tag{4.3}
$$

is a solution of (4.2)*. Conversely, if $(\boldsymbol{\sigma}_{h,0}, \mathbf{u}_h)$ is a solution of* (4.2)*, then $(\boldsymbol{\sigma}_h, \mathbf{u}_h) = (\boldsymbol{\sigma}_{h,0} - \frac{1}{n|\Omega|}(\mathrm{tr}\,(\mathbf{u}_h \otimes \mathbf{u}_h), 1)_\Omega \mathbb{I}, \mathbf{u}_h)$ is a solution of* (4.1)*.*

4.2 Analysis of the discrete problem

Analogously to the analysis developed in Section 3, to prove the well-posedness of problem (4.1), we introduce a fixed-point operator associated to a linearized version of problem (4.2) and equivalently prove that this operator possess a unique fixed-point by means of the Banach fixed-point theorem. To that end we need to introduce some previous results.

4.2.1 Preliminary results

Given $p > \dfrac{2n}{n+2}$, let us define the space

$$
\mathbf{Z}_p := \{\tau \in \mathbf{H}(\mathrm{div}_p; \Omega) \ : \ \tau|_T \in \mathbf{W}^{1,p}(T), \quad \forall\, T \in \mathcal{T}_h\},
$$

and let

$$
\Pi_h^k : \mathbf{Z}_p \to \mathbf{X}_h := \{\tau \in \mathbf{H}(\mathrm{div}; \Omega) \ : \ \tau|_T \in \mathbf{RT}_k(T), \quad \forall\, T \in \mathcal{T}_h\},
$$

be the Raviart–Thomas interpolator operator, which is well defined in $\mathbf{Z}_p$ (see e.g. [11, Section 1.2.7]) and is characterized by the identities

$$
\int_e (\Pi_h^k(\tau) \cdot \nu)\xi = \int_e (\tau \cdot \nu)\xi \qquad \forall\, \xi \in P_k(e),\ \forall \text{ edge or face } e \text{ of } \mathcal{T}_h,
$$

and

$$\int_T \Pi_h^k(\tau)\cdot\psi \,=\, \int_T \tau\cdot\psi \qquad \forall\,\psi\in[P_{k-1}(T)]^n,\ \forall\,T\in\mathcal{T}_h \ (\text{if } k\geq 1)\ .$$

In addition, it is well known (see e.g. [11, Lemma 1.41]) that the following identity holds

$$\mathrm{div}(\Pi_h^k(\tau)) = \mathcal{P}_h^k(\mathrm{div}\tau) \quad \forall\,\tau\in\mathbf{Z}_p, \tag{4.4}$$

where $\mathcal{P}_h^k : L^2(\Omega)\to M_h := \{v\in \mathrm{L}^2(\Omega) \,:\, v|_T\in P_k(T) \quad \forall T\in\mathcal{T}_h\}$ is the usual orthogonal projection, which satisfies the following error estimate (see [11, Proposition 1.135, Section 1.6.3]): For each $0\leq t\leq k+1$ and for each $w\in \mathrm{W}^{t,r}(\Omega)$, with $1\leq r\leq\infty$, there holds

$$\|w-\mathcal{P}_h^k(w)\|_{\mathrm{L}^r(\Omega)} \leq C h^t |w|_{\mathrm{W}^{t,r}(\Omega)}. \tag{4.5}$$

The following lemma establishes the local approximation properties of Π_h^k.

Lemma 4.2 *Let $r>\dfrac{2n}{n+2}$. Then, there exists $C_1>0$, independent of h, such that for each $\tau\in\mathbf{W}^{l+1,r}(T)$ with $0\leq l\leq k$, and for each $0\leq m\leq l+1$, there holds*

$$|\tau-\Pi_h^k(\tau)|_{\mathbf{W}^{m,r}(T)} \leq C_1\frac{h_T^{l+2}}{\rho_T^{m+1}}|\tau|_{\mathbf{W}^{l+1,r}(T)}. \tag{4.6}$$

Moreover, there exists $C_2>0$, independent of h, such that for each $\tau\in\mathbf{W}^{1,r}(T)$, with $\mathrm{div}\,\tau\in \mathrm{W}^{l+1,r}(T)$ and $0\leq l\leq k$, and for each $0\leq m\leq l+1$, there holds

$$|\mathrm{div}\,\tau-\mathrm{div}\,(\Pi_h^k(\tau))|_{\mathrm{W}^{m,r}(T)} \leq C_2\frac{h_T^{l+1}}{\rho_T^{m}}|\mathrm{div}\,\tau|_{\mathrm{W}^{l+1,r}(T)}. \tag{4.7}$$

Proof. Employing the L^p-version of the Deny–Lions Lemma provided in [11, Lemma B.67] and the local estimates given in [11, Lemma 1.101], one can proceed analogously as in [17, Section 3.4.4] and prove that for any $r>\dfrac{2n}{n+2}$ estimates (4.6) and (4.7) hold. We omit further details. □

Owing to the regularity of the mesh and from estimates (4.6) and (4.7), it is not difficult see that the following global estimate holds

$$\|\tau\,-\,\Pi_h^k(\tau)\|_{0,\Omega}+\|\mathrm{div}\,\tau-\mathrm{div}\,(\Pi_h(\tau))\|_{\mathrm{L}^p(\Omega)} \leq ch^{l+1}\Big\{|\tau|_{\mathbf{H}^{l+1}(\Omega)}+|\mathrm{div}\,\tau|_{\mathrm{W}^{l+1,p}(\Omega)}\Big\}, \tag{4.8}$$

for all $0\leq l\leq k+1$, and for all $\tau\in\mathbf{H}^{l+1}(\Omega)$ with $\mathrm{div}\,\tau\in\mathrm{W}^{l+1,p}(\Omega)$.

In the sequel, it will be employed a tensor version of Π_h^k, say $\boldsymbol{\Pi}_h^k:\mathbb{Z}_p\to\mathbb{X}_h$ which is defined row–wise by Π_h^k, and a vector version of $\mathcal{P}_h^k$, say $\mathbf{P}_h^k$, defined element–wise by $\mathcal{P}_h^k$. Obviously, both $\boldsymbol{\Pi}_h^k$ and $\mathbf{P}_h^k$ also satisfy the properties described above

Remark 4.3 *Notice that from the regularity of the mesh and from (4.6) with $m=0$ and $m=1$, one can easily obtain, respectively, that*

$$\|\tau-\Pi_h^k(\tau)\|_{\mathbf{L}^r(T)} \leq C_1\frac{h_T^{l+2}}{\rho_T}|\tau|_{\mathbf{W}^{l+1,r}(T)} \leq \hat{C}_1 h_T^{l+1}|\tau|_{\mathbf{W}^{l+1,r}(T)}$$

and

$$|\tau - \Pi_h^k(\tau)|_{\mathbf{W}^{1,r}(T)} \le C_2 \frac{h_T^{l+2}}{\rho_T^2} |\boldsymbol{\tau}|_{\mathbf{W}^{l+1,r}(T)} \le \hat{C}_2 h_T^l |\tau|_{\mathbf{W}^{l+1,r}(T)},$$

which combined yield

$$\|\tau - \Pi_h^k(\tau)\|_{\mathbf{W}^{1,r}(\Omega)} \le C h^l |\tau|_{\mathbf{W}^{l+1,r}(\Omega)} \quad \forall\, \tau \in \mathbf{W}^{l+1,r}(\Omega). \tag{4.9}$$

The latter will be employed next in the proof of Lemma 4.4.

4.2.2 The discrete fixed-point operator and its well-definiteness

Let us define the set

$$\mathbf{K}_h := \left\{ \mathbf{v} \in \mathbf{M}_h \, : \, \|\mathbf{v}\|_{\mathbf{M}} \,\le\, \frac{2}{\widehat{\gamma}} \Big(C_F \|\mathbf{u}_D\|_{1/2,\Gamma} + \|\mathbf{f}\|_{\mathbf{L}^{4/3}(\Omega)} \Big) \right\},$$

with $\widehat{\gamma} > 0$ being the constant defined in (4.21) below. Then, the discrete fixed-point operator is defined as

$$\mathcal{J}_h : \mathbf{K}_h \to \mathbf{K}_h, \quad \mathbf{w}_h \to \mathcal{J}_h(\mathbf{w}_h) \,=\, \mathbf{u}_h,$$

where, given $\mathbf{w}_h \in \mathbf{K}_h$, $\mathbf{u}_h$ is the second component of the solution of problem: Find $(\boldsymbol{\sigma}_h, \mathbf{u}_h) \in \mathbb{X}_{h,0} \times \mathbf{M}_h$,

$$\begin{aligned} \mathbf{a}(\boldsymbol{\sigma}_h, \boldsymbol{\tau}_h) \,+\, \mathbf{b}(\boldsymbol{\tau}_h, \mathbf{u}_h) \,+\, \mathbf{c}(\mathbf{w}_h; \mathbf{u}_h, \boldsymbol{\tau}_h) \quad &= F(\boldsymbol{\tau}_h), \quad \forall\, \boldsymbol{\tau}_h \in \mathbb{X}_{h,0} \\ \mathbf{b}(\boldsymbol{\sigma}_h, \mathbf{v}_h) \quad &= G(\mathbf{v}_h), \quad \forall\, \mathbf{v}_h \in \mathbf{M}_h. \end{aligned} \tag{4.10}$$

Analogously to the continuous case, we also have the following equivalences

$$\begin{aligned} \mathcal{J}_h(\mathbf{u}_h) = \mathbf{u}_h \quad &\Leftrightarrow \ ((\boldsymbol{\sigma}_{h,0}, \mathbf{u}_h)) \in \mathbb{X}_{h,0} \times \mathbf{M}_h \text{ satisfies } (4.2) \\ &\Leftrightarrow \ (\boldsymbol{\sigma}_{h,0} - \tfrac{1}{n|\Omega|}(\mathrm{tr}\,(\mathbf{u}_h \otimes \mathbf{u}_h), 1)_\Omega \mathbb{I}, \mathbf{u}_h) \in \mathbb{X}_h \times \mathbf{M}_h \text{ satisfies } (4.1), \end{aligned} \tag{4.11}$$

from which we deduce that to proving the well-posedness of problem (4.1), or equivalently (4.2), it suffices to prove that $\mathcal{J}_h$ has a unique fixed-point in $\mathbf{K}_h$. Before doing that, and similarly to the analysis of the continuous problem, first we focus on providing the main ingredients to proving that operator $\mathcal{J}_h$ is well-defined. We begin by observing that, since $\mathbf{div}\ \mathbb{X}_h \subseteq \mathbf{M}_h$, the discrete kernel of $\mathbf{b}$, which is defined by

$$\mathbb{V}_h := \{ \boldsymbol{\tau}_h \in \mathbb{X}_{h,0} : \quad \mathbf{b}(\boldsymbol{\tau}_h, \mathbf{v}_h) = 0, \forall\, \mathbf{v}_h \in \mathbf{M}_h \} \,,$$

can be characterized as follows

$$\mathbb{V}_h = \{ \boldsymbol{\tau}_h \in \mathbb{X}_{h,0} : \quad \mathbf{div}\ \boldsymbol{\tau}_h = 0 \text{ in } \Omega \} \,.$$

Therefore, from (3.15) we readily obtain that the bilinear form $\mathbf{a}$ is elliptic on $\mathbb{V}_h$, that is

$$\mathbf{a}(\boldsymbol{\tau}_h, \boldsymbol{\tau}_h) \,\ge\, \alpha \|\boldsymbol{\tau}_h\|_{\mathbb{X}}^2 \qquad \forall\, \boldsymbol{\tau}_h \in \mathbb{V}_h, \tag{4.12}$$

with α defined as in (3.17).

Now we establish the discrete version of Lemma 3.4.

Lemma 4.4 *There exists $\widehat{\beta} > 0$, such that*

$$\sup_{\mathbf{0}\neq\boldsymbol{\tau}_h\in\mathbb{X}_{h,0}} \frac{\mathbf{b}(\boldsymbol{\tau}_h,\mathbf{v}_h)}{\|\boldsymbol{\tau}_h\|_{\mathbb{X}}} \geq \widehat{\beta}\|\mathbf{v}_h\|_{\mathbf{M}} \qquad \forall\,\mathbf{v}_h\,\in\,\mathbf{M}_h. \tag{4.13}$$

Proof. In what follows we proceed similarly to the proof of [7, Lemma 3.3] and employ the arguments utilized in the proof of [17, Section 4.2] to prove the inf-sup condition (4.13). In fact, given $\mathbf{v}_h \in \mathbf{M}_h$, we set

$$\mathbf{g}(\mathbf{v}_h) := \begin{cases} \mathbf{sgn}(\mathbf{v}_h)|\mathbf{v}_h|^3 & \text{in} \quad \Omega, \\ \mathbf{0} & \text{in} \quad B\backslash\overline{\Omega}, \end{cases}$$

where $B \subseteq \mathbb{R}^n$ is an open and bounded convex set containing $\overline{\Omega}$. Then, since $\mathbf{g}(\mathbf{v}_h) \in \mathbf{L}^{4/3}(\Omega)$, a well known result on regularity of elliptic problems (see e.g. [16]) implies that there exists a unique weak solution $\mathbf{z} \in \mathbf{W}^{2,4/3}(B) \cap \mathbf{W}_0^{1,4/3}(B)$ of the boundary value problem

$$-\Delta\mathbf{z} = \mathbf{g}(\mathbf{v}_h) \quad \text{in} \quad B \quad \text{and} \quad \mathbf{z} = 0 \quad \text{on} \quad \partial B,$$

which satisfies

$$\|\mathbf{z}\|_{\mathbf{W}^{2,4/3}(\Omega)} \leq C\|\mathbf{g}(\mathbf{v}_h)\|_{\mathbf{L}^{4/3}(B)} = C\||\mathbf{v}_h|^3\|_{\mathbf{L}^{4/3}(\Omega)} = C\|\mathbf{v}_h\|_{\mathbf{M}}^3, \tag{4.14}$$

with $C > 0$. Hence, we set $\hat{\boldsymbol{\tau}} = -\nabla\mathbf{z}|_\Omega \in \mathbb{W}^{1,4/3}(\Omega)$, and observe from (4.14) that

$$\|\hat{\boldsymbol{\tau}}\|_{\mathbb{W}^{1,4/3}(\Omega)} \leq C\|\mathbf{v}_h\|_{\mathbf{M}}^3, \tag{4.15}$$

which together to the continuity of the embedding from $\mathrm{W}^{1,4/3}(\Omega)$ into $\mathrm{L}^2(\Omega)$, implies

$$\|\hat{\boldsymbol{\tau}}\|_{0,\Omega} \leq C\|\mathbf{v}_h\|_{\mathbf{M}}^3. \tag{4.16}$$

Then, we define $\hat{\boldsymbol{\tau}}_h = \mathbf{\Pi}_h^k(\hat{\boldsymbol{\tau}}) - \dfrac{1}{n|\Omega|}\Big(\mathrm{tr}\,(\mathbf{\Pi}_h^k(\hat{\boldsymbol{\tau}})), 1\Big)_\Omega \mathbb{I} \in \mathbb{X}_{h,0}$ and observe from (4.4), that

$$\mathbf{div}\,\hat{\boldsymbol{\tau}}_h = \mathbf{P}_h^k(\mathbf{div}\,\hat{\boldsymbol{\tau}}) = \mathbf{P}_h^k(\mathbf{sgn}(\mathbf{v}_h)|\mathbf{v}_h|^3). \tag{4.17}$$

In turn, utilizing the triangle inequality, the continuity of the embedding from $\mathrm{W}^{1,4/3}(\Omega)$ into $\mathrm{L}^2(\Omega)$, and the estimate (4.16), we obtain

$$\begin{aligned}
\|\hat{\boldsymbol{\tau}}_h\|_{0,\Omega} &\leq \left\|\hat{\boldsymbol{\tau}} - \frac{1}{n|\Omega|}\,(\mathrm{tr}\,(\hat{\boldsymbol{\tau}}),1)_\Omega\,\mathbb{I} - \hat{\boldsymbol{\tau}}_h\right\|_{0,\Omega} + \left\|\hat{\boldsymbol{\tau}} - \frac{1}{n|\Omega|}\,(\mathrm{tr}\,(\hat{\boldsymbol{\tau}}),1)_\Omega\,\mathbb{I}\right\|_{0,\Omega} \\
&= \left\|\hat{\boldsymbol{\tau}} - \mathbf{\Pi}_h^k(\hat{\boldsymbol{\tau}}) - \frac{1}{n|\Omega|}\left(\mathrm{tr}\,\left(\hat{\boldsymbol{\tau}} - \mathbf{\Pi}_h^k(\hat{\boldsymbol{\tau}})\right), 1\right)_\Omega \mathbb{I}\right\|_{0,\Omega} + \left\|\hat{\boldsymbol{\tau}} - \frac{1}{n|\Omega|}\,(\mathrm{tr}\,(\hat{\boldsymbol{\tau}}),1)_\Omega\,\mathbb{I}\right\|_{0,\Omega} \\
&\leq \left\|\hat{\boldsymbol{\tau}} - \mathbf{\Pi}_h^k(\hat{\boldsymbol{\tau}})\right\|_{0,\Omega} + \|\hat{\boldsymbol{\tau}}\|_{0,\Omega} \\
&\leq c_1\left\|\hat{\boldsymbol{\tau}} - \mathbf{\Pi}_h^k(\hat{\boldsymbol{\tau}})\right\|_{\mathbb{W}^{1,4/3}(\Omega)} + c_2\|\mathbf{v}_h\|_{\mathbf{M}}^3,
\end{aligned}$$

which together with (4.9) with $r = 4/3$ and $l = 0$, and (4.15), imply

$$\|\hat{\boldsymbol{\tau}}_h\|_{0,\Omega} \,\leq\, C\|\mathbf{v}_h\|_{\mathbf{M}}^3. \tag{4.18}$$

Hence, using the fact that $\mathbf{P}_h^k$ is a continuous operator, from (4.17) and (4.18), we easily obtain

$$\|\hat{\boldsymbol{\tau}}_h\|_{\mathbb{X}} = \left\{\|\hat{\boldsymbol{\tau}}_h\|_{0,\Omega}^2 + \|\mathbf{div}\,(\hat{\boldsymbol{\tau}}_h)\|_{\mathbf{L}^{4/3}(\Omega)}^2\right\}^{1/2} \leq \hat{c}\|\mathbf{v}_h\|_{\mathbf{M}}^3, \tag{4.19}$$

with $\hat{c} > 0$ independent of h.

Therefore, from (2.3), (4.17) and (4.19), we obtain

$$\sup_{\substack{\boldsymbol{\tau}_h\in\mathbb{X}_{h,0}\\ \boldsymbol{\tau}_h\neq\theta}} \frac{\mathbf{b}(\boldsymbol{\tau}_h,\mathbf{v}_h)}{\|\boldsymbol{\tau}_h\|_{\mathbb{X}}} \geq \frac{\mathbf{b}(\hat{\boldsymbol{\tau}}_h,\mathbf{v}_h)}{\|\hat{\boldsymbol{\tau}}_h\|_{\mathbb{X}}} \geq \frac{1}{\hat{c}}\frac{(|\mathbf{v}_h|^3\mathbf{sgn}(\mathbf{v}_h),\mathbf{v}_h)_\Omega}{\|\mathbf{v}_h\|_{\mathbf{M}}^3} \geq \frac{1}{\hat{c}}\frac{\|\mathbf{v}_h\|_{\mathbf{M}}^4}{\|\mathbf{v}_h\|_{\mathbf{M}}^3} = \frac{1}{\hat{c}}\|\mathbf{v}_h\|_{\mathbf{M}},$$

which concludes the proof with $\widehat{\beta} = \dfrac{1}{\hat{c}}$. □

To conclude the derivation of the main tools to proving the well–definiteness of $\mathcal{J}_h$, analogously to the continuous case, from (4.12), (4.13) and [11, Theorem 2.34] we finally deduce that the bilinear form $\mathbf{A}$ defined in (3.27) satisfies the discrete inf-sup condition

$$\sup_{\mathbf{0}\neq(\boldsymbol{\tau}_h,\mathbf{v}_h)\in\mathbb{X}_{h,0}\times\mathbf{M}_h} \frac{\mathbf{A}((\boldsymbol{\sigma}_h,\mathbf{u}_h),(\boldsymbol{\tau}_h,\mathbf{v}_h))}{\|(\boldsymbol{\tau}_h,\mathbf{v}_h)\|} \geq \widehat{\gamma}\|(\boldsymbol{\sigma}_h,\mathbf{u}_h)\| \qquad \forall\,(\boldsymbol{\sigma}_h,\mathbf{u}_h)\,\in\,\mathbb{X}_{h,0}\times\mathbf{M}_h, \tag{4.20}$$

with

$$\widehat{\gamma} := \frac{C_d\widehat{\beta}\min\{1,\nu\widehat{\beta}\}}{2(C_d+1)(\nu\widehat{\beta}+1)}. \tag{4.21}$$

Now, we are in position of establishing the well-definiteness of $\mathcal{J}_h$

Theorem 4.5 *Assume that*

$$\frac{4}{\nu\widehat{\gamma}^2}\left(C_F\|\mathbf{u}_D\|_{1/2,\Gamma} + \|\mathbf{f}\|_{\mathbf{L}^{4/3}(\Omega)}\right) \,\leq\, 1. \tag{4.22}$$

Then, given $\mathbf{w}_h \in \mathbf{K}_h$*, there exists a unique* $\mathbf{u}_h \in \mathbf{K}_h$ *such that* $\mathcal{J}_h(\mathbf{w}_h) \,=\, \mathbf{u}_h$*.*

Proof. Given $\mathbf{w}_h \in \mathbf{K}_h$ we proceed analogously to the proof of Theorem 3.6 and utilize (3.12), (4.20) and (4.22) to deduce that $\mathbf{A}_{\mathbf{w}_h}$ (cf. (3.31)) satisfies the discrete inf-sup condition

$$\sup_{\mathbf{0}\neq(\boldsymbol{\tau}_h,\mathbf{v}_h)\in\mathbb{X}_{h,0}\times\mathbf{M}_h} \frac{\mathbf{A}_{\mathbf{w}_h}((\boldsymbol{\sigma}_h,\mathbf{u}_h),(\boldsymbol{\tau}_h,\mathbf{v}_h))}{\|(\boldsymbol{\tau}_h,\mathbf{v}_h)\|} \geq \frac{\widehat{\gamma}}{2}\|(\boldsymbol{\sigma}_h,\mathbf{u}_h)\|. \tag{4.23}$$

Therefore, owing to the fact that for finite dimensional linear problems, surjectivity and injectivity are equivalent, from (4.23) and the Banach–Nečas–Babuška theorem we obtain that there exists a unique $(\boldsymbol{\sigma}_h,\mathbf{u}_h) \in \mathbb{X}_{h,0}\times\mathbf{M}_h$ satisfying

$$\mathbf{A}_{\mathbf{w}_h}((\boldsymbol{\sigma}_h,\mathbf{u}_h),(\boldsymbol{\tau}_h,\mathbf{v}_h)) = F(\boldsymbol{\tau}_h) + G(\mathbf{v}_h), \quad \forall\,(\boldsymbol{\tau}_h,\mathbf{v}_h)\in\mathbb{X}_{h,0}\times\mathbf{M}_h,$$

or equivalently (4.10), with $\mathbf{u}_h \in \mathbf{K}_h$, which concludes the proof. □

4.2.3 Well-posedness of the discrete problem

The following theorem establishes the well-posedness of the Galerkin scheme (4.1).

Theorem 4.6 *Let* $\mathbf{f} \in \mathbf{L}^{4/3}(\Omega)$ *and* $\mathbf{u}_D \in \mathbf{H}^{1/2}(\Gamma)$ *such that*

$$\frac{4}{\nu \widehat{\gamma}^2} \Big(C_F \|\mathbf{u}_D\|_{1/2,\Gamma} + \|\mathbf{f}\|_{\mathbf{L}^{4/3}(\Omega)} \Big) < 1. \tag{4.24}$$

Then, there exists a unique $(\boldsymbol{\sigma}_h, \mathbf{u}_h) \in \mathbb{X}_h \times \mathbf{M}_h$ *solution to* (4.1). *In addition, there exists* $C > 0$*, independent of* h*, such that*

$$\|\mathbf{u}_h\|_{\mathbf{M}} + \|\boldsymbol{\sigma}_h\|_{\mathbb{X}} \leq C \Big(\|\mathbf{u}_D\|_{1/2,\Gamma} + \|\mathbf{f}\|_{\mathbf{L}^{4/3}(\Omega)} \Big).$$

Proof. Employing (4.11), (4.20) and (4.24), the proof follows repeating exactly the same steps developed in the proof of Theorem 3.7. We omit further details. $\square$

Remark 4.7 *We end this section by emphasizing that from the second equation of* (4.1) *we have that*

$$((\mathbf{div}\, \boldsymbol{\sigma}_h - \mathbf{f}), \mathbf{v}_h)_\Omega = 0 \qquad \forall\, \mathbf{v}_h \in \mathbf{M}_h,$$

which implies that $\mathbf{div}\, \boldsymbol{\sigma}_h = \mathbf{P}_h^k(\mathbf{f})$ *and consequently, if* $\mathbf{f} \in \mathbf{M}_h$ *then our method exactly conserves the discrete equilibrium equation, thus the method is conservative.*

4.3 Cea's estimate and rate of convergence

In this section we aim to provide the convergence of the Galerkin scheme (4.1) and derive the corresponding rate of convergence. We begin by deriving the corresponding Cea's estimate of the equivalent Galerkin scheme (4.2).

Theorem 4.8 *Assume that*

$$\frac{4}{\nu \gamma \widehat{\gamma}} \Big(C_F \|\mathbf{u}_D\|_{1/2,\Gamma} + \|\mathbf{f}\|_{\mathbf{L}^{4/3}(\Omega)} \Big) \leq \frac{1}{2}, \tag{4.25}$$

with γ *and* $\widehat{\gamma}$ *being the positive constants in* (3.29) *and* (4.21)*, respectively. Let* $(\boldsymbol{\sigma}_0, \mathbf{u}) \in \mathbb{X}_0 \times \mathbf{M}$ *and* $(\boldsymbol{\sigma}_h, \mathbf{u}_h) \in \mathbb{X}_{h,0} \times \mathbf{M}_h$ *be the unique solutions of problems* (3.3) *and* (4.2)*, respectively. Then, there exists* $C^0_{cea} > 0$*, independent of* h*, such that*

$$\|(\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}_{h,0}, \mathbf{u} - \mathbf{u}_h)\| \leq C^0_{cea} \inf_{(\boldsymbol{\tau}_h, \mathbf{v}_h) \in \mathbb{X}_{h,0} \times \mathbf{M}_h} \|(\boldsymbol{\sigma}_0 - \boldsymbol{\tau}_h, \mathbf{u} - \mathbf{v}_h)\|. \tag{4.26}$$

Proof. In order to simplify the subsequent analysis, we define $\mathbf{e}_{\boldsymbol{\sigma}_0} := \boldsymbol{\sigma}_0 - \boldsymbol{\sigma}_{h,0}$ and $\mathbf{e}_{\mathbf{u}} := \mathbf{u} - \mathbf{u}_h$, and for any $(\boldsymbol{\zeta}_h, \mathbf{z}_h) \in \mathbb{X}_{h,0} \times \mathbf{M}_h$, we write

$$\mathbf{e}_{\boldsymbol{\sigma}_0} = \boldsymbol{\xi}_{\boldsymbol{\sigma}_0} + \boldsymbol{\chi}_{\boldsymbol{\sigma}_0} := (\boldsymbol{\sigma}_0 - \boldsymbol{\zeta}_h) + (\boldsymbol{\zeta}_h - \boldsymbol{\sigma}_{h,0}), \qquad \mathbf{e}_{\mathbf{u}} = \boldsymbol{\xi}_{\mathbf{u}} + \boldsymbol{\chi}_{\mathbf{u}} := (\mathbf{u} - \mathbf{z}_h) + (\mathbf{z}_h - \mathbf{u}_h). \tag{4.27}$$

Recalling the definition of the bilinear form $\mathbf{A}$ in (3.27), from (3.3) and (4.2) we have that the following identities hold

$$\mathbf{A}((\boldsymbol{\sigma}_0, \mathbf{u}), (\boldsymbol{\tau}, \mathbf{v})) + \mathbf{c}(\mathbf{u}; \mathbf{u}, \boldsymbol{\tau}) = F(\boldsymbol{\tau}) + G(\mathbf{v}) \qquad \forall\, (\boldsymbol{\tau}, \mathbf{v}) \in \mathbb{X}_0 \times \mathbf{M}$$

and

$$\mathbf{A}((\boldsymbol{\sigma}_{h,0},\mathbf{u}_h),(\boldsymbol{\tau}_h,\mathbf{v}_h)) + \mathbf{c}(\mathbf{u}_h;\mathbf{u}_h,\boldsymbol{\tau}_h) = F(\boldsymbol{\tau}_h) + G(\mathbf{v}_h) \qquad \forall\,(\boldsymbol{\tau}_h,\mathbf{v}_h) \in \mathbb{X}_{h,0}\times\mathbf{M}_h,$$

from which we deduce the Galerkin orthogonality property

$$\mathbf{A}((\mathbf{e}_{\boldsymbol{\sigma}_0},\mathbf{e}_{\mathbf{u}}),(\boldsymbol{\tau}_h,\mathbf{v}_h)) \,+\, [\mathbf{c}(\mathbf{u};\mathbf{u},\boldsymbol{\tau}_h) - \mathbf{c}(\mathbf{u}_h;\mathbf{u}_h,\boldsymbol{\tau}_h)] \,=\, 0 \qquad \forall\,(\boldsymbol{\tau}_h,\mathbf{v}_h) \in \mathbb{X}_{h,0}\times\mathbf{M}_h. \tag{4.28}$$

Then, using the decompositions (4.27), the definition of $\mathbf{A}_{\mathbf{w}}$ in (3.31), and the identity

$$\mathbf{c}(\mathbf{u};\mathbf{u},\boldsymbol{\tau}_h) = \mathbf{c}(\mathbf{u}-\mathbf{u}_h;\mathbf{u},\boldsymbol{\tau}_h) + \mathbf{c}(\mathbf{u}_h;\mathbf{u},\boldsymbol{\tau}_h),$$

from (4.28) we obtain that for all $(\boldsymbol{\tau}_h,\mathbf{v}_h) \in \mathbb{X}_{h,0}\times\mathbf{M}_h$, there holds

$$\begin{aligned}\mathbf{A}_{\mathbf{u}_h}((\boldsymbol{\chi}_{\boldsymbol{\sigma}_0},\boldsymbol{\chi}_{\mathbf{u}}),(\boldsymbol{\tau}_h,\mathbf{v}_h)) \;=\; & -\mathbf{A}((\boldsymbol{\xi}_{\boldsymbol{\sigma}_0},\boldsymbol{\xi}_{\mathbf{u}}),(\boldsymbol{\tau}_h,\mathbf{v}_h)) \,-\, \mathbf{c}(\boldsymbol{\xi}_{\mathbf{u}};\mathbf{u},\boldsymbol{\tau}_h) \\ & -\mathbf{c}(\boldsymbol{\chi}_{\mathbf{u}};\mathbf{u},\boldsymbol{\tau}_h) \,-\, \mathbf{c}(\mathbf{u}_h;\boldsymbol{\xi}_{\mathbf{u}},\boldsymbol{\tau}_h),\end{aligned}$$

which together to the definition of $\mathbf{A}$ in (3.27), implies

$$\begin{aligned}\mathbf{A}_{\mathbf{u}_h}((\boldsymbol{\chi}_{\boldsymbol{\sigma}_0},\boldsymbol{\chi}_{\mathbf{u}}),(\boldsymbol{\tau}_h,\mathbf{v}_h)) \;=\; & -\mathbf{a}(\boldsymbol{\xi}_{\boldsymbol{\sigma}_0},\boldsymbol{\tau}_h) - \mathbf{b}(\boldsymbol{\xi}_{\boldsymbol{\sigma}_0},\mathbf{v}_h) \,-\, \mathbf{b}(\boldsymbol{\tau}_h,\boldsymbol{\xi}_{\mathbf{u}}) \\ & -\mathbf{c}(\boldsymbol{\chi}_{\mathbf{u}};\mathbf{u},\boldsymbol{\tau}_h) \,-\, \mathbf{c}(\mathbf{u}_h;\boldsymbol{\xi}_{\mathbf{u}},\boldsymbol{\tau}_h),\end{aligned} \tag{4.29}$$

for all $(\boldsymbol{\tau}_h,\mathbf{v}_h) \in \mathbb{X}_{h,0}\times\mathbf{M}_h$. Then, since $\mathbf{u}_h \in \mathbf{K}_h$, we apply the discrete inf-sup condition (4.23) at the left hand side of (4.29), the continuity properties of $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ (cf. (3.10)-(3.12)), at the right hand side of (4.29), to obtain

$$\|\boldsymbol{\chi}_{\boldsymbol{\sigma}_0}\|_{\mathbb{X}} + \|\boldsymbol{\chi}_{\mathbf{u}}\|_{\mathbf{M}} \;\le \frac{2}{\widehat{\gamma}}\left(\frac{(1+\nu)}{\nu}\|\boldsymbol{\xi}_{\boldsymbol{\sigma}_0}\|_{\mathbb{X}} + \left\{1+\frac{1}{\nu}\|\mathbf{u}_h\|_{\mathbf{M}}\right\}\|\boldsymbol{\xi}_{\mathbf{u}}\|_{\mathbf{M}} + \frac{1}{\nu}\|\boldsymbol{\chi}_{\mathbf{u}}\|_{\mathbf{M}}\|\mathbf{u}\|_{\mathbf{M}}\right),$$

from which

$$\|\boldsymbol{\chi}_{\boldsymbol{\sigma}_0}\|_{\mathbb{X}} + \left(1-\frac{2}{\nu\widehat{\gamma}}\|\mathbf{u}\|_{\mathbf{M}}\right)\|\boldsymbol{\chi}_{\mathbf{u}}\|_{\mathbf{M}} \le \frac{2}{\widehat{\gamma}}\left(\frac{(1+\nu)}{\nu}\|\boldsymbol{\xi}_{\boldsymbol{\sigma}_0}\|_{\mathbb{X}} + \left\{1+\frac{1}{\nu}\|\mathbf{u}_h\|_{\mathbf{M}}\right\}\|\boldsymbol{\xi}_{\mathbf{u}}\|_{\mathbf{M}}\right).$$

Hence, using the fact that $\mathbf{u} \in \mathbf{K}$ and $\mathbf{u}_h \in \mathbf{K}_h$, from assumption (4.25) and the latter inequality, we obtain

$$\|\boldsymbol{\chi}_{\boldsymbol{\sigma}_0}\|_{\mathbb{X}} + \|\boldsymbol{\chi}_{\mathbf{u}}\|_{\mathbf{M}} \,\le\, C\left(\|\boldsymbol{\xi}_{\boldsymbol{\sigma}_0}\|_{\mathbb{X}} \,+\, \|\boldsymbol{\xi}_{\mathbf{u}}\|_{\mathbf{M}}\right), \tag{4.30}$$

with $C > 0$ independent of h. In this way, from (4.27), (4.30) and the triangle inequality we obtain

$$\|(\mathbf{e}_{\boldsymbol{\sigma}_0},\mathbf{e}_{\mathbf{u}})\| \,\le\, \|(\boldsymbol{\chi}_{\boldsymbol{\sigma}_0},\boldsymbol{\chi}_{\mathbf{u}})\| \,+\, \|(\boldsymbol{\xi}_{\boldsymbol{\sigma}_0},\boldsymbol{\xi}_{\mathbf{u}})\| \le\, (1+C)\|(\boldsymbol{\xi}_{\boldsymbol{\sigma}_0},\boldsymbol{\xi}_{\mathbf{u}})\|,$$

which combined to the fact that $(\boldsymbol{\zeta}_h,\mathbf{v}_h) \in \mathbb{X}_{h,0}\times\mathbf{M}_h$ is arbitrary, concludes the proof. $\square$

The following result establishes the corresponding estimate for $\boldsymbol{\sigma}_h \in \mathbb{X}_h$ given by the identity (4.3).

Corollary 4.9 *Let $(\boldsymbol{\sigma}_{h,0},\mathbf{u}_h) \in \mathbb{X}_{h,0}\times\mathbf{M}$ be the unique solution of* (4.2)*, and let $\boldsymbol{\sigma}_h \in \mathbb{X}_h$, given by the identity* (4.3)*, be such that $(\boldsymbol{\sigma}_h,\mathbf{u}_h) \in \mathbb{X}_h\times\mathbf{M}$ is the unique solution of* (4.1)*. Assume that* (4.25) *holds. Then, there exists $C > 0$, independent of h, such that*

$$\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathbb{X}} \,\le\, C \inf_{(\boldsymbol{\tau}_h,\mathbf{v}_h)\in\mathbb{X}_{h,0}\times\mathbf{M}_h} \|(\boldsymbol{\sigma}_0-\boldsymbol{\tau}_h,\mathbf{u}-\mathbf{v}_h)\|, \tag{4.31}$$

where $(\boldsymbol{\sigma}_0,\mathbf{u}) \in \mathbb{X}_0\times\mathbf{M}$ is the unique solution of (3.3) *and $\boldsymbol{\sigma} \in \mathbb{X}$ is given by the identity* (3.4).

Proof. Using the identities (3.4) and (4.3), it is not difficult to see that

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathbb{X}} \,\leq\, \|\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}_{h,0}\|_{\mathbb{X}} \,+\, \frac{1}{n^{1/2}|\Omega|^{1/2}}\,|(\mathrm{tr}\,(\mathbf{u}\otimes\mathbf{u} - \mathbf{u}_h\otimes\mathbf{u}_h), 1)_\Omega|\,. \tag{4.32}$$

In turn, adding and subtracting $\mathbf{u}_h$ and employing suitable algebraic manipulations we obtain

$$\begin{aligned} |(\mathrm{tr}\,(\mathbf{u}\otimes\mathbf{u} - \mathbf{u}_h\otimes\mathbf{u}_h), 1)_\Omega| \;&=\; \big|((|\mathbf{u}|^2 - |\mathbf{u}_h|^2), 1)_\Omega\big| \leq (\|\mathbf{u}\|_{0,\Omega} + \|\mathbf{u}_h\|_{0,\Omega})\|\mathbf{u} - \mathbf{u}_h\|_{0,\Omega} \\ &\leq\; |\Omega|(\|\mathbf{u}\|_{\mathbf{M}} + \|\mathbf{u}_h\|_{\mathbf{M}})\|\mathbf{u} - \mathbf{u}_h\|_{\mathbf{M}}, \end{aligned}$$

which together to (4.32) and the fact that $\mathbf{u} \in \mathbf{K}$ and $\mathbf{u}_h \in \mathbf{K}_h$, yields

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathbb{X}} \,\leq\, \|\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}_{h,0}\|_{\mathbb{X}} + C\|\mathbf{u} - \mathbf{u}_h\|_{\mathbf{M}},$$

with $C > 0$, independent of h. In this way, from the latter inequality and estimate (4.26) we easily obtain (4.31) which concludes the proof. □

Remark 4.10 *From Theorem 4.8 and Corollary 4.9 it can be readily seen that the following Cea's estimate for problem* (4.1) *holds: There exists $C_{cea} > 0$, independent of h, such that*

$$\|(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{u} - \mathbf{u}_h)\| \,\leq\, C_{cea} \inf_{(\boldsymbol{\tau}_h,\mathbf{v}_h)\in\mathbb{X}_{h,0}\times\mathbf{M}_h} \|(\boldsymbol{\sigma}_0 - \boldsymbol{\tau}_h, \mathbf{u} - \mathbf{v}_h)\|.$$

Now we are in position of establishing the rates of convergence associated to the Galerkin schemes (4.1) and (4.2).

Theorem 4.11 *Let $(\boldsymbol{\sigma}_0, \mathbf{u}) \in \mathbb{X}_0 \times \mathbf{M}$ and $(\boldsymbol{\sigma}_{h,0}, \mathbf{u}_h) \in \mathbb{X}_{h,0} \times \mathbf{M}_h$ be the unique solutions of the continuous and discrete problems* (3.3) *and* (4.2)*, respectively, with $\mathbf{f}$ and $\mathbf{u}_D$ satisfying* (4.25)*. Assume further that $\boldsymbol{\sigma}_0 \in \mathbb{H}^{l+1}(\Omega)$, $\mathbf{div}\,\boldsymbol{\sigma}_0 \in \mathbf{W}^{l+1,4/3}(\Omega)$ and $\mathbf{u} \in \mathbf{W}^{l+1,4}(\Omega)$, for $0 \leq l \leq k+1$. Then there exists $C_{rate} > 0$, independent of h, such that*

$$\|(\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}_{h,0}, \mathbf{u} - \mathbf{u}_h)\| \,\leq\, C^0_{rate} h^{l+1} \Big\{|\boldsymbol{\sigma}_0|_{\mathbb{H}^{l+1}(\Omega)} + |\mathbf{div}\,\boldsymbol{\sigma}_0|_{\mathbf{W}^{l+1,4/3}} \,+\, |\mathbf{u}|_{\mathbf{W}^{l+1,4}(\Omega)}\Big\}\,.$$

In addition, if $(\boldsymbol{\sigma}, \mathbf{u}) \in \mathbb{X} \times \mathbf{M}$ and $(\boldsymbol{\sigma}_h, \mathbf{u}_h) \in \mathbb{X}_h \times \mathbf{M}_h$ are the unique solutions of problems (2.16) *and* (4.1)*, then there holds*

$$\|(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{u} - \mathbf{u}_h)\| \,\leq\, C_{rate} h^{l+1} \Big\{|\boldsymbol{\sigma}_0|_{\mathbb{H}^{l+1}(\Omega)} + |\mathbf{div}\,\boldsymbol{\sigma}_0|_{\mathbf{W}^{l+1,4/3}} \,+\, |\mathbf{u}|_{\mathbf{W}^{l+1,4}(\Omega)}\Big\}\,,$$

with $C_{rate} > 0$, independent of h.

Proof. The result is a straightforward application of Theorem 4.8, Corollary 4.9 and estimates (4.5) and (4.8). □

Table 4.1: Local degrees of freedom for the lowest-order method ($k = 0$).

	$\mathbf{RT}_0 - P_0$	Bernardi-Raugel	MINI-element
local Dof	8	10	11

Remark 4.12 *In Table 4.1 we compare the local degrees of freedom (Dof) of our method, considering $k = 0$ and $n = 2$, with the corresponding local Dof of the velocity-pressure formulation discretized by the Bernardi–Raugel element and the MINI-element (see Chapter III in [22]). We observe there that, although our formulation possesses considerably more unknowns (6 unknowns in 2D) than the velocity-pressure formulation (3 unknowns in 2D), the computational cost is not increased.*

4.4 Computing further variables of interest

In this section we introduce suitable approximations for further variables of interest, such that the pressure p, the vorticity $\boldsymbol{\omega} := \frac{1}{2}(\nabla\mathbf{u} - \nabla\mathbf{u}^t)$, the stress $\tilde{\boldsymbol{\sigma}} := \nu(\nabla\mathbf{u} + (\nabla\mathbf{u})^t) - p\mathbb{I}$ and the velocity gradient $\mathbf{G} = \nabla\mathbf{u}$, all of them written in terms of the solution of the discrete problem (4.1). In fact, observing that at the continuous level there hold

$$p \;=\; -\frac{1}{n}(\mathrm{tr}\,(\boldsymbol{\sigma}) + \mathrm{tr}\,(\mathbf{u}\otimes\mathbf{u})), \qquad \tilde{\boldsymbol{\sigma}} \;=\; \boldsymbol{\sigma}^d \;+\; (\mathbf{u}\otimes\mathbf{u})^d + \boldsymbol{\sigma}^t + \mathbf{u}\otimes\mathbf{u},$$
$$\mathbf{G} \;=\; \frac{1}{\nu}(\boldsymbol{\sigma}^d \;+\; (\mathbf{u}\otimes\mathbf{u})^d) \quad \text{and} \quad \boldsymbol{\omega} \;=\; \frac{1}{2\nu}(\boldsymbol{\sigma} - \boldsymbol{\sigma}^t),$$

provided the discrete solution $(\boldsymbol{\sigma}_h, \mathbf{u}_h) \in \mathbb{X}_h \times \mathbf{M}_h$ of problem (4.1), we propose the following approximations for the aforementioned variables:

$$\begin{gathered} p_h \;=\; -\frac{1}{n}(\mathrm{tr}\,(\boldsymbol{\sigma}_h) + \mathrm{tr}\,(\mathbf{u}_h\otimes\mathbf{u}_h)), \qquad \tilde{\boldsymbol{\sigma}}_h \;=\; \boldsymbol{\sigma}_h^d \;+\; (\mathbf{u}_h\otimes\mathbf{u}_h)^d + \boldsymbol{\sigma}_h^t + \mathbf{u}_h\otimes\mathbf{u}_h, \\ \mathbf{G}_h \;=\; \frac{1}{\nu}(\boldsymbol{\sigma}_h^d \;+\; (\mathbf{u}_h\otimes\mathbf{u}_h)^d) \quad \text{and} \quad \boldsymbol{\omega}_h \;=\; \frac{1}{2\nu}(\boldsymbol{\sigma}_h - \boldsymbol{\sigma}_h^t). \end{gathered} \tag{4.33}$$

The following result establishes the corresponding approximation result for this post-processing procedure.

Corollary 4.13 *Let* $(\boldsymbol{\sigma}_0, \mathbf{u}) \in \mathbb{X}_0 \times \mathbf{M}$ *and* $(\boldsymbol{\sigma}_{h,0}, \mathbf{u}_h) \in \mathbb{X}_{h,0} \times \mathbf{M}_h$ *be the unique solutions of the continuous and discrete problems* (3.3) *and* (4.2)*, respectively, with* $\mathbf{f}$ *and* $\mathbf{u}_D$ *satisfying* (4.25)*. Assume further that* $\boldsymbol{\sigma}_0 \in \mathbb{H}^{l+1}(\Omega)$, $\mathbf{div}\,\boldsymbol{\sigma}_0 \in \mathbf{W}^{l+1,4/3}(\Omega)$ *and* $\mathbf{u} \in \mathbf{W}^{l+1,4}(\Omega)$*, for* $0 \le l \le k+1$*. Finally, let* p_h*,* $\tilde{\boldsymbol{\sigma}}_h$*,* $\mathbf{G}_h$ *and* $\boldsymbol{\omega}_h$ *given by* (4.33)*. Then there exists* $\tilde{C} > 0$*, independent of* h*, such that*

$$\begin{aligned} &\|p - p_h\|_{0,\Omega} \;+\; \|\tilde{\boldsymbol{\sigma}} - \tilde{\boldsymbol{\sigma}}_h\|_{0,\Omega} \;+\; \|\mathbf{G} - \mathbf{G}_h\|_{0,\Omega} \;+\; \|\boldsymbol{\omega} - \boldsymbol{\omega}_h\|_{0,\Omega} \\ &\quad \le\; \tilde{C} h^{l+1}\left\{|\boldsymbol{\sigma}_0|_{\mathbb{H}^{l+1}(\Omega)} + |\mathbf{div}\,\boldsymbol{\sigma}_0|_{\mathbf{W}^{l+1,4/3}} \;+\; |\mathbf{u}|_{\mathbf{W}^{l+1,4}(\Omega)}\right\} \end{aligned}$$

Proof. Similarly as in the proof of Corollary 4.9, from Hölder inequality, and recalling that $\mathbf{u} \in \mathbf{K}$ and $\mathbf{u}_h \in \mathbf{K}_h$, it is not difficult to see that

$$\|\mathbf{u}\otimes\mathbf{u} - \mathbf{u}_h\otimes\mathbf{u}_h\|_{0,\Omega} \le C\|\mathbf{u} - \mathbf{u}_h\|_{\mathbf{M}}, \tag{4.34}$$

with $C > 0$, independent of h. Then, using (4.34), the result follows from Theorem 4.11 and Corollary 4.9. We omit further details. □

5 Numerical results

In this section we report two numerical examples that will show the performance of our finite element scheme. Before proceeding with the description of our examples we first point out that, since the condition $(\mathrm{tr}\,(\boldsymbol{\sigma}_h), 1)_\Omega = -(\mathrm{tr}\,(\mathbf{u}_h \otimes \mathbf{u}_h), 1)_\Omega$ in problem (4.1) is not computationally implementable, at least in a simple way, we recall that (4.1) and (4.2) are equivalent (see Lemma 4.1) and consequently we perform the following numerical tests by solving problem (4.2), where the condition $(\mathrm{tr}\,(\boldsymbol{\sigma}_{h,0}), 1)_\Omega = 0$ is imposed through a penalization strategy.

Our implementation is based on a *FreeFem++* code (see [23]), in conjunction with the direct linear solver UMFPACK (see [10]). We apply a Newton's method with a fixed tolerance *tol*= 1E-6 and the iterations are terminated once the relative error of the entire coefficient vectors between two consecutive iterates is sufficiently small, i.e.,

$$\frac{|\mathbf{coeff}^{m+1} - \mathbf{coeff}^m|}{|\mathbf{coeff}^{m+1}|} \leq tol,$$

where $|\cdot|$ is the standard euclidean norm $\mathbb{R}^N$, with N denoting the total number of degrees of freedom defining the finite element subspaces $\mathbb{X}_h$ and $\mathbf{M}_h$. For each example shown below we simply take $(\boldsymbol{\sigma}_{h,0}^0, \mathbf{u}_h^0) = (\mathbf{0}, \mathbf{0})$ as initial guess.

We now introduce some additional notations. The individual errors are denoted by:

$$\mathsf{e}(\boldsymbol{\sigma}_0) := \|\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}_{h,0}\|_{\mathbb{X}}, \quad \mathsf{e}(\mathbf{u}) := \|\mathbf{u} - \mathbf{u}_h\|_{\mathbf{M}}, \quad \mathsf{e}(p) := \|p - p_h\|_{0,\Omega},$$

$$\mathsf{e}(\boldsymbol{\omega}) := \|\boldsymbol{\omega} - \boldsymbol{\omega}_h\|_{0,\Omega}, \quad \mathsf{e}(\nabla\mathbf{u}) := \|\nabla\mathbf{u} - \mathbf{G}_h\|_{0,\Omega}, \quad \mathsf{e}(\tilde{\boldsymbol{\sigma}}) := \|\tilde{\boldsymbol{\sigma}} - \tilde{\boldsymbol{\sigma}}_h\|_{0,\Omega}.$$

where p_h, $\boldsymbol{\omega}_h$, $\mathbf{G}_h$, and $\tilde{\boldsymbol{\sigma}}_h$ are the variables computed through the post-processing formulas (4.33).

In addition, we let $r(\boldsymbol{\sigma}_0)$, $r(\mathbf{u})$, $r(p)$, $r(\boldsymbol{\omega})$, $r(\nabla\mathbf{u})$, and $r(\tilde{\boldsymbol{\sigma}})$ be the experimental rates of convergence given by

$$r(\boldsymbol{\sigma}_0) := \frac{\log(\mathsf{e}(\boldsymbol{\sigma}_0)/\mathsf{e}'(\boldsymbol{\sigma}_0))}{\log(h/h')}, \quad r(\mathbf{u}) := \frac{\log(\mathsf{e}(\mathbf{u})/\mathsf{e}'(\mathbf{u}))}{\log(h/h')}, \quad r(p) := \frac{\log(\mathsf{e}(p)/\mathsf{e}'(p))}{\log(h/h')},$$

$$r(\boldsymbol{\omega}) := \frac{\log(\mathsf{e}(\boldsymbol{\omega})/\mathsf{e}'(\boldsymbol{\omega}))}{\log(h/h')}, \quad r(\nabla\mathbf{u}) := \frac{\log(\mathsf{e}(\nabla\mathbf{u})/\mathsf{e}'(\nabla\mathbf{u}))}{\log(h/h')}, \quad r(\tilde{\boldsymbol{\sigma}}) := \frac{\log(\mathsf{e}(\tilde{\boldsymbol{\sigma}})/\mathsf{e}'(\tilde{\boldsymbol{\sigma}}))}{\log(h/h')},$$

where h and h' denote two consecutive meshsizes with errors e and e'.

The first example focusses on the performance of our method as a function of the viscosity ν, by considering the analytical solution $(\mathbf{u}, p)$ obtained by Kovasznay in [30]. For the domain $\Omega := (-1/2, 3/2) \times (0, 2)$ and for a given ν, this solution is given by

$$\mathbf{u}(x_1, x_2) = \begin{pmatrix} 1 - e^{\lambda x_1}\cos(2\pi x_2) \\ \frac{\lambda}{2\pi} e^{\lambda x_1}\sin(2\pi x_2) \end{pmatrix},$$

$$p(x_1, x_2) = -\frac{1}{2}e^{2\lambda x_1} + \bar{p},$$

where

$$\lambda := \frac{-8\pi^2}{\nu^{-1} + \sqrt{\nu^{-2} + 16\pi^2}},$$

Number of iterations

ν	$h = 0.1905$	$h = 0.0978$	$h = 0.0517$	$h = 0.0316$	$h = 0.0156$
1	4	4	4	4	3
0.1	–	5	5	5	5
0.01	–	–	–	6	6

Table 5.1: EXAMPLE 1: Convergence behavior of the Newton's method with respect to the parameter ν.

and the constant $\bar{p}$ is such that $\int_\Omega p = 0$.

In Table 5.1 we show the behaviour of the Newton's method as a function of the viscosity number, considering different meshsizes. We consider the finite element spaces introduced in Section 4 with $k = 0$. Here we observe that the smaller the parameter ν the higher the number of iterations. Notice is quite moderate on each case. Blank spaces means that the iterative method takes more that 100 iterations. With $\nu = 10^{-3}$, the Newton iteration does not converge anymore, probably because the stationary solution is not stable in this case. Next, in Table 5.2, we summarise the convergence history for a sequence of quasi-uniform triangulations, of two finite element families corresponding to $\mathrm{RT}_0 - \mathrm{P}_0$ and $\mathrm{RT}_1 - \mathrm{P}_1$ and considering the viscosity $\nu = 1$. We see there that the rate of convergence provided by Theorem 4.11 and Corollary 4.13 is attained by the unknowns and all the post-processed variables, that is, $\mathcal{O}(h)$ for the first case, and $\mathcal{O}(h^2)$ for the second one. In addition, the l^∞–norm of $\mathbf{div}\,\boldsymbol{\sigma}_{h,0}$ in each mesh is close to 0 which shows that this method is conservative.

In our second example we perform a classical lid–driven cavity test modelling the steady flow of an immiscible fluid in a box. The domain is given by the unit square $\Omega = (0,1)^2$ and we consider an structured mesh with meshsize $h = 0.0164$. The data are given as in [27, Section 3.2], that is, a null body force $\mathbf{f} = \mathbf{0}$ and the prescribed velocity boundary

$$\mathbf{u}_D(x_1, x_2) = \begin{cases} (10x_1, 0)^t & \text{for} \quad 0 \le x_1 \le 0.1, \quad x_2 =, 1 \\[2mm] (1, 0)^t & \text{for} \quad 0.1 \le x_1 \le 0.9, \quad x_2 = 1, \\[2mm] ((10 - 10x_1), 0)^t & \text{for} \quad 0.9 \le x_1 \le 1, \quad x_2 = 1, \\[2mm] (0,0)^t & \text{for} \quad (x_1, x_2) \in \Gamma \backslash \{(x_1, 1) : 0 \le x_1 \le 1\}. \end{cases}$$

In Figure 5.1 we display the velocity streamlines for $\nu = 1$ (left) and $\nu = 0.0025$ (right) where we can see that, as expected, and similarly to the results obtained in [27], the main vortex moves toward the center of the cavity for increasing Reynolds numbers $Re = \frac{1}{\nu}$.

References

[1] F. BREZZI AND M. FORTIN, *Mixed and Hybrid Finite Element Methods*. Springer Series in Computational Mathematics, 15. Springer-Verlag, New York, (1991).

Errors and rates of convergence for the mixed $\mathrm{RT}_0 - \mathrm{P}_0$ approximation

N	h	$\mathsf{e}(\boldsymbol{\sigma}_0)$	$r(\boldsymbol{\sigma}_0)$	$\mathsf{e}(\mathbf{u})$	$r(\mathbf{u})$	$\lVert\mathbf{div}\,\boldsymbol{\sigma}_{h,0}\rVert_{l^\infty}$	Iterations
3035	0.1905	45.2555	–	2.2882	–	9.0949e-13	4
12199	0.0978	22.5468	1.0457	1.1136	1.0809	1.8190e-12	4
48697	0.0517	11.6288	1.0393	0.5645	1.0665	5.4570e-12	4
196483	0.0316	5.6438	1.4634	0.2752	1.4542	1.4552e-11	4
774345	0.0156	2.8305	0.9815	0.1390	0.9720	2.9104e-11	3

$\mathsf{e}(p)$	$r(p)$	$\mathsf{e}(\boldsymbol{\omega})$	$r(\boldsymbol{\omega})$	$\mathsf{e}(\nabla\mathbf{u})$	$r(\nabla\mathbf{u})$	$\mathsf{e}(\tilde{\boldsymbol{\sigma}})$	$r(\tilde{\boldsymbol{\sigma}})$
21.5953	–	24.0425	–	24.1994	–	46.0334	–
10.7272	1.0502	14.3573	0.7738	12.8855	0.9459	21.9551	1.1112
5.4725	1.0565	7.7320	0.9715	6.7505	1.0148	11.0729	1.0745
2.6113	1.4978	3.8602	1.4062	3.3198	1.4367	5.2839	1.4977
1.2954	0.9970	1.9803	0.9493	1.6791	0.9694	2.6059	1.0053

Errors and rates of convergence for the mixed $\mathrm{RT}_1 - \mathrm{P}_1$ approximation

N	h	$\mathsf{e}(\boldsymbol{\sigma}_0)$	$r(\boldsymbol{\sigma}_0)$	$\mathsf{e}(\mathbf{u})$	$r(\mathbf{u})$	$\lVert\mathbf{div}\,\boldsymbol{\sigma}_{h,0}\rVert_{l^\infty}$	Iterations
9633	0.1905	9.3862	–	0.2653	–	1.0687e-11	4
38881	0.0978	2.2385	2.1515	0.0613	2.1985	2.9559e-11	4
155521	0.0517	0.5926	2.0863	0.0161	2.1012	6.8212e-11	4
628129	0.0316	0.1396	2.9268	0.0038	2.9436	1.6735e-10	3
2476673	0.0156	0.0354	1.9503	0.0010	1.9315	3.6744e-10	3

$\mathsf{e}(p)$	$r(p)$	$\mathsf{e}(\boldsymbol{\omega})$	$r(\boldsymbol{\omega})$	$\mathsf{e}(\nabla\mathbf{u})$	$r(\nabla\mathbf{u})$	$\mathsf{e}(\tilde{\boldsymbol{\sigma}})$	$r(\tilde{\boldsymbol{\sigma}})$
4.2620	–	4.6641	–	5.2331	–	10.1175	–
0.9890	2.1925	1.2550	1.9703	1.2983	2.0922	2.3556	2.1875
0.2594	2.1008	0.3445	2.0295	0.3491	2.0619	0.6202	2.0948
0.0589	2.9996	0.0882	2.7589	0.0851	2.8562	0.1428	2.9724
0.0146	1.9873	0.0234	1.8837	0.0220	1.9259	0.0355	1.9812

Table 5.2: Example 1: Degrees of freedom, meshsizes, errors, rates of convergence, L^∞–norm of $\mathbf{div}\,\boldsymbol{\sigma}_{h,0}$ and number of iterations for the mixed $\mathrm{RT}_0 - \mathrm{P}_0$ and $\mathrm{RT}_1 - \mathrm{P}_1$ approximations of the Navier-Stokes problem, with $\nu = 1$.

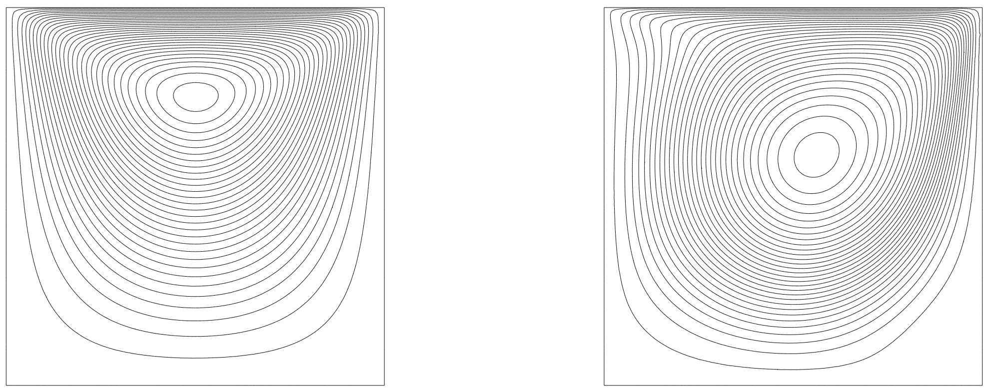

Figure 5.1: Example 2: Velocity streamlines for $Re = 1$ (left) and $Re = 400$ (right) with $h = 0.0164$.

[2] Z. Cai, Ch. Tong, P.S. Vassilevski and Ch. Wang, *Mixed finite element methods for incompressible flow: stationary Stokes equations.* Numerical Methods Partial Differential Equations, vol. 26, pp. 957–978, (2009).

[3] Z. Cai and Y. Wang, *Pseudostress-velocity formulation for incompressible Navier-Stokes equations.* International Journal for Numerical Methods in Fluids, 63, no. 3, 341–356, (2010).

[4] Z. Cai, Ch. Wang and S. Zhang, *Mixed finite element methods for incompressible flow: stationary Navier-Stokes equations.* SIAM Journal on Numerical Analysis, vol 48, no. 1, 79–94, (2010).

[5] Z. Cai and S. Zhang, *Mixed methods for stationary Navier-Stokes equations based on pseudostress-pressure-velocity formulation.* Mathematics of Computation, vol. 81 , no. 280, 1903–1927, (2012).

[6] J. Camaño, R. Oyarzúa and G. Tierra, *Analysis of an augmented mixed-fem for the Navier-Stokes problem.* Mathematics of Computation, vol. 86, no. 304, pp. 589–615, (2017).

[7] J. Camaño, C. Muñoz and R. Oyarzúa, *Numerical analysis of a dual-mixed problem in non-standard Banach spaces.* Electronic Transactions on Numerical Analysis, vol. 48, pp. 114-130, (2018).

[8] A. Cesmelioglu, B. Cockburn and W. Qiu, *Analysis of a hybridizable discontinuous Galerkin method for the steady–state incompressible Navier–Stokes equations.* Mathematics of Computation, vol. 86, pp. 1643–1670, (2017).

[9] B. Cockburn, G. Kanschat, and D. Schötzau, *A locally conservative LDG method for the incompressible Navier-Stokes equations.* Mathematics of Computation, vol. 74, 251, pp. 1067–1095, (2005).

[10] T. Davis, *Algorithm 832: UMFPACK V4.3 - an unsymmetric-pattern multifrontal method.* ACM Transactions on Mathematical Software, vol. 30, pp. 196–199, (2004).

[11] A. Ern and J.-L.Guermond, *Theory and Practice of Finite Elements.* Applied Mathematical Sciences, Springer-Verlag, New-York, vol. 159, (2004).

[12] V.J. Ervin, J.S. Howell and I. Stanculescu, *A dual-mixed approximation method for a three-eld model of a nonlinear generalized Stokes problem.* Computer Methods in Applied Mechanics and Engineering, vol. 197, 33–40, pp. 2886-2900, (2008).

[13] M. Farhloul, S. Nicaise and L. Paquet, *A refined mixed finite-element method for the stationary Navier-Stokes equations with mixed boundary conditions* IMA Journal of Numerical Analysis, vol. 28, no. 1, 25–45, (2008).

[14] M. Farhloul, S. Nicaise and L. Paquet, *A priori and a posteriori error estimations for the dual mixed finite element method of the Navier-Stokes problem*Numerical Methods for Partial Differential Equations, vol. 25, no. 4, 843–869, (2009).

[15] L. Figueroa, G.N. Gatica and A. Márquez, *Augmented mixed finite element methods for the stationary Stokes Equations.* SIAM Journal on Scientific Computing, vol. 31, 2, pp. 1082–1119, (2008).

[16] S. J. Fromm, *Potential space estimates for Green potentials in convex domains.* Proceedings of the American Mathematical Society, vol. 119, no. 1, pp. 225–233, (1993).

[17] G.N. Gatica, *A Simple Introduction to the Mixed Finite Element Method. Theory and Applications* Springer Briefs in Mathematics, Springer, Cham Heidelberg New York Dordrecht London, (2014).

[18] G.N. Gatica, M. González and S. Meddahi, *A low-order mixed finite element method for a class of quasi-Newtonian Stokes flows. I: A priori error analysis.* Computer Methods in Applied Mechanics and Engineering, vol. 193, 9-11, pp. 881-892, (2004).

[19] G.N. Gatica, L.F. Gatica and A. Márquez, *Augmented mixed finite element methods for a vorticity-based velocity–pressure–stress formulation of the Stokes problem in 2D.* International Journal for Numerical Methods in Fluids, vol. 67, 4, pp. 450–477, (2011).

[20] G.N. Gatica, A. Márquez and M.A. Sánchez, *Analysis of a velocity-pressure-pseudostress formulation for incompressible flow.* Computer Methods in Applied Mechanics and Engineering, vol. 199, 17-20, pp. 1064–1079, (2010).

[21] 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.*Computer Methods in Applied Mechanics and Engineering, vol. 200, 17-20, pp. 1619–1636, (2011).

[22] V. Girault and P.-A. Raviart, Finite Element Methods for NavierStokes Equations. Theory and Algorithms. Springer Series in Computational Mathematics, 5. Springer–Verlag, Berlin, (1986).

[23] F. Hecht, *New development in FreeFem++.* Journal of Numerical Mathematics, vol. 20, 3-4, pp. 251–265, (2012).

[24] J.S. Howell, *Dual-mixed nite element approximation of Stokes and nonlinear Stokes problems using trace-free velocity gradients.* Journal of Computational and Applied Mathematics, vol. 231, 2, pp. 780–792, (2009).

[25] J.S. Howell and N. Walkington, *Dual Mixed Finite Element Methods for the Navier Stokes Equations.* ESAIM: Mathematical Modelling and Numerical Analysis, vol. 47, 789–805, (2013).

[26] R.J. Labeur and G.N. Wells, *Energy stable and momentum conserving hybrid finite element method for the incompressible Navier-Stokes equations.* SIAM Journal of Scientific Computing, vol. 34, No. 2, pp. A889–A913, (2012).

[27] A. Montlaur, S. Fernandez-Mendez, J. Peraire and A. Huerta *Discontinuous Galerkin methods for the NavierStokes equations using solenoidal approximations* International Journal for Numerical Methods in Fluids, vol. 64, pp. 549–564, (2010).

[28] A. Quarteroni and A. Valli, *Numerical approximation of partial differential equations.* Springer Series in Computational Mathematics, 23. Springer-Verlag, Berlin, 1994.

[29] A. Sifounakis, S. Lee and D. You, *A conservative finite volume method for incompressible Navier–Stokes equations on locally refined nested Cartesian grids.* Journal of Computational Physics, vol. 326, pp. 845–861, (2016).

[30] L. I. G. Kovasznay, *Laminar flow behind a two-dimensional grid,* Proceedings of the Cambridge Philosophical Society, vol. 44, pp. 58–62, (1948).