UNIVERSIDAD DE CONCEPCIÓN

Centro de Investigación en Ingeniería Matemática (CI²MA)

Locking-free finite element methods for poroelasticity

Ricardo Oyarzúa, Ricardo Ruiz-Baier

PREPRINT 2015-41

SERIE DE PRE-PUBLICACIONES

Locking-free finite element methods for poroelasticity *

Ricardo Oyarzúa[†] Ricardo Ruiz-Baier[‡]

Abstract

We propose a new formulation along with a family of finite element schemes for the approximation of the interaction between fluid motion and linear mechanical response of a porous medium, known as Biot's consolidation problem. The system is recast in terms of displacement, pressure, and volumetric stress, and both continuous and discrete formulations are analyzed as compact perturbations of invertible problems employing a Fredholm argument. Numerical results indicate the satisfactory performance and competitive accuracy of the introduced methods.

Key words: Poroelasticity; Finite element approximation; Volumetric stress formulation; Compact perturbation; Fredholm alternative; Error estimates.

Mathematics subject classifications (2010): 65N30; 76S05; 74F10; 65N15.

1 Introduction

Linear poroelasticity equations consist of a momentum conservation for a porous skeleton, coupled with mass conservation of a diffusive flow within the medium. In its basic form introduced in [4], the system allows to describe physical loading of porous layers and the change of hydraulic equilibrium in a fluid-structure system; and serves as the classical model for subsurface consolidation processes and it has applications in many scenarios of high practical importance, such as petroleum production, geological CO_2 sequestration, waste disposal, pile foundations, perfusion of bones and soft living tissues, etc. The success in accurately replicating poroelasticity solutions using numerical methods is often affected by the presence of two unphysical scenarios: spurious pressure modes and locking phenomena. Here we propose a three-field formulation of the model problem, where classical finite element methods can be employed straightforwardly without the risk of producing the aforementioned phenomena. We remark that the additional third unknown introduced in the model is a scalar field (and contains information about stresses), which makes the formulation very appealing from the computational viewpoint.

*This work was partially supported by CONICYT-Chile through project Anillo ACT1118 (ANANUM), and by Universidad del Bío-Bío through DIUBB project 151408 GI/VC.

†GIMNAP, Departamento de Matemática, Universidad del Bío-Bío, Casilla 5-C, Concepción, Chile, and Centro de Investigación en Ingeniería Matemática (CI^2MA), Universidad de Concepción, Concepción, Chile. E-mail: `royarzua@ubiobio.cl`.

‡Mathematical Institute, Oxford University, Andrew Wiles Building, Woodstock Road, OX2 6GG Oxford, UK. E-mail: `ruizbaier@maths.ox.ac.uk`.

Related work and specifics of this contribution. The stability of a semidiscrete finite element (FE) method applied to linear poroelasticity was studied in the early work [21]. Mixed-primal FE formulations to approximate the solid displacement, the fluid flux and the pore pressure were introduced in [22, 26]. Primal and primal-mixed discontinuous Galerkin (DG) approximations of linear poroelasticity were proposed and analyzed in [7, 19], least-squares mixed FE methods were also applied for Biot's consolidation system in [18], pressure-stabilized methods have been employed in [27, 3], and [28] presents a mixed-mixed formulation for the same problem, where the unknowns are the Cauchy stress, the displacement, the pressure and the fluid flux, and a mixed-mixed FE method follows the same continuous setting.

Our goal is to present a stable and convergent conforming FE method for the discretization of the model problem, where the volumetric contributions to the total stress are merged into an additional unknown, yielding a saddle point formulation that can be analyzed by means of a Fredholm alternative, after realizing that the problem is a compact perturbation of a Stokes-like invertible system. More precisely, in the coupled variational formulation there is a zero order term with a "wrong sign" which causes the loss of invertibility of the associated operator. However, the compactness of the embedding $H^1(\Omega) \hookrightarrow L^2(\Omega)$ allows one to make use of a Fredholm alternative to analyze its solvability (see similar approaches in [8, 13, 14]). In addition, a generic Galerkin scheme is constructed, whose solvability properties follow closely those from the continuous variational form, and more importantly, given that specific FE spaces are chosen adequately, it is stable even in the incompressible limit ($\lambda \to \infty$). We emphasize that the latter means that all constants in the estimates below are independent of the Lamé parameter λ.

Outline. The layout of this paper is as follows. In the remainder of this section we recall some needed notation and general definitions. Section 2 summarizes the model equations of linear poroelasticity, including its strong and weak forms, and boundary conditions considered in the subsequent analysis. The Galerkin scheme is introduced in Section 3, where also the corresponding stability analysis and convergence are derived. In particular, in Section 3.3 we make precise the definition of the involved discrete spaces, recall some approximation properties, and state the theoretical error bounds. Finally, Section 4 collects several numerical results and benchmark test cases illustrating the accuracy of the proposed methods.

Preliminaries. Standard notation will be adopted for Lebesgue and Sobolev spaces. Moreover, $\mathbf{M}$ and $\mathbb{M}$ will denote the corresponding vectorial and tensorial counterparts of the generic scalar functional space M and $\|\cdot\|$, with no subscripts, will stand for the natural norm of either an element or an operator in any product functional space. For instance, if $\Theta \subseteq \mathrm{R}^d$, $d = 2, 3$ is a domain, $\Lambda \subseteq \mathrm{R}^d$ is a Lipschitz surface, and $r \in \mathrm{R}$, we define $\mathbf{H}^r(\Theta) := [\mathrm{H}^r(\Theta)]^d$ and $\mathbf{H}^r(\Lambda) := [\mathrm{H}^r(\Lambda)]^d$. By $\mathbf{0}$ we will refer to the generic null vector (including the null functional and operator), and we will denote by C and c, with or without subscripts, bars, tildes or hats, generic constants independent of the discretization parameters, which may take different values at different occurrences.

2 Governing equations and well-posedness analysis

2.1 Proposed three-field formulation and boundary conditions

Let us consider a homogeneous porous matrix containing a mixture of incompressible grains and interstitial fluid. We assume that this material body occupies a bounded and simply connected domain $\Omega \subset \mathrm{R}^d$, $d = 2, 3$. For all $t > 0$, given a body force $\boldsymbol{f}(t) : \Omega \to \mathrm{R}^d$ and a volumetric fluid source (or sink) $s(t) : \Omega \to \mathrm{R}$, the classical Biot consolidation problem (cf. [4]) consists in finding the displacements of the porous skeleton, $\boldsymbol{u}(t) : \Omega \to \mathrm{R}^d$ and the pore pressure of the fluid, $p(t) : \Omega \to \mathrm{R}$, such that

$$\partial_t\big(c_0 p + \alpha(\operatorname{div} \boldsymbol{u})\big) - \frac{1}{\eta}\operatorname{div}[\kappa(\nabla p - \rho \boldsymbol{g})] = s \qquad \text{in } \Omega, \qquad (2.1)$$

$$\boldsymbol{\sigma} = \lambda(\operatorname{div} \boldsymbol{u})\mathbf{I} + 2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}) - p\mathbf{I} \qquad \text{in } \Omega, \qquad (2.2)$$

$$-\operatorname{\mathbf{div}} \boldsymbol{\sigma} = \boldsymbol{f} \qquad \text{in } \Omega, \qquad (2.3)$$

where $\boldsymbol{\sigma}$ is the total Cauchy solid stress, $\boldsymbol{\varepsilon}(\boldsymbol{u}) = \frac{1}{2}(\nabla\boldsymbol{u} + \nabla\boldsymbol{u}^T)$ is the infinitesimal strain tensor (symmetrized gradient of displacements), κ is the permeability of the porous solid (here assumed isotropic and satisfying $0 < \kappa_1 \le \kappa(\boldsymbol{x}) \le \kappa_2 < \infty$, for all $\boldsymbol{x} \in \Omega$), λ, μ are the Lamé constants of the solid, $c_0 > 0$ is the constrained specific storage coefficient, $\alpha > 0$ is the so-called Biot-Willis parameter, $\boldsymbol{g}$ is the gravity acceleration (constant and aligned with the vertical direction), $\eta > 0, \rho > 0$ are the viscosity and density of the pore fluid, and the term $c_0 p + \alpha(\operatorname{div} \boldsymbol{u})$ represents the total fluid content in the domain (fluid pressure plus the material volume).

Notice that (2.2) states the constitutive law of the solid (differing from the classical linear elastic model in that here p is the fluid pressure), equation (2.3) represents momentum conservation of the porous medium (under the assumption that the solid deformations are much slower than the fluid flow rate), whereas mass conservation of the fluid obeying a Darcy regime, is accounted for by (2.1). Using Hölder continuity assumptions for rather standard boundary and initial data, the solvability of (2.1)-(2.3) has been established in [23].

In order to illustrate the main ideas of the new formulation and its discretization, we will restrict the discussion to a static problem consisting of (2.2),(2.3) coupled with the relation

$$c_0 p + \alpha(\operatorname{div} \boldsymbol{u}) - \frac{1}{\eta}\operatorname{div}[\kappa(\nabla p - \rho\boldsymbol{g})] = s \quad \text{in } \Omega, \qquad (2.4)$$

arising from e.g. Euler time discretization of (2.1) (and making abuse of notation in s). Time dependence of field variables and data can be therefore dropped. Let us further consider the auxiliary unknown representing the volumetric part of the total stress (also may be regarded as a pseudo total pressure) defined as

$$\phi := p - \lambda \operatorname{div} \boldsymbol{u}. \qquad (2.5)$$

Therefore (2.2) and (2.4) read, respectively,

$$\boldsymbol{\sigma} = 2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}) - \phi\mathbf{I}, \qquad \Big(c_0 + \frac{\alpha}{\lambda}\Big)p - \frac{\alpha}{\lambda}\phi - \frac{1}{\eta}\operatorname{div}[\kappa(\nabla p - \rho\boldsymbol{g})] = s \qquad \text{in } \Omega. \qquad (2.6)$$

We assume that the domain boundary is disjointly split into a part where fluid pressure is specified and a part where displacements are imposed $\partial\Omega = \overline{\Gamma}_p \cup \overline{\Gamma}_{\boldsymbol{u}}$, $\Gamma_p \cap \Gamma_{\boldsymbol{u}} = \emptyset$. System (2.5)-(2.6) is then complemented with suitable boundary conditions

$$p = p_\Gamma, \; \boldsymbol{\sigma n} = \boldsymbol{h} \text{ on } \Gamma_p, \quad \text{and} \quad \boldsymbol{u} = \boldsymbol{u}_\Gamma, \; (\kappa \nabla p) \cdot \boldsymbol{n} = j \text{ on } \Gamma_{\boldsymbol{u}}, \tag{2.7}$$

where $\boldsymbol{n}$ is the exterior unit normal vector on $\partial\Omega$, $\boldsymbol{h}$ is a known load vector, and j is an imposed pressure flux.

2.2 Weak formulation

Homogeneous boundary data will be assumed for sake of conciseness of the presentation, but we stress that (2.7) can be incorporated later on, using classical lifting arguments. Let us multiply (2.3),(2.5),(2.6) by adequate test functions and proceed to integrate by parts in such a way that second order derivatives are removed, and the following weak formulation holds: Find $\boldsymbol{u} \in \mathbf{H}, p \in \mathrm{Q}, \phi \in \mathrm{Z}$ such that

$$a_1(\boldsymbol{u}, \boldsymbol{v}) \qquad\qquad + b_1(\boldsymbol{v}, \phi) = F(\boldsymbol{v}) \qquad \forall \boldsymbol{v} \in \mathbf{H}, \tag{2.8}$$

$$a_2(p, q) - b_2(q, \phi) = G(q) \qquad \forall q \in \mathrm{Q}, \tag{2.9}$$

$$b_1(\boldsymbol{u}, \psi) + b_2(p, \psi) - c(\phi, \psi) = 0 \qquad \forall \psi \in \mathrm{Z}, \tag{2.10}$$

where the boundary treatment suggests to define the involved functional spaces as

$$\mathbf{H} := \mathbf{H}^1_{\Gamma_{\boldsymbol{u}}}(\Omega) = \{\boldsymbol{v} \in \mathbf{H}^1(\Omega) : \boldsymbol{v}|_{\Gamma_{\boldsymbol{u}}} = \mathbf{0}\}, \quad \mathrm{Z} := \mathrm{L}^2(\Omega),$$

$$\mathrm{Q} := \mathrm{H}^1_{\Gamma_p}(\Omega) = \{q \in \mathrm{H}^1(\Omega) : q|_{\Gamma_p} = 0\},$$

and the bilinear forms $a_1 : \mathbf{H} \times \mathbf{H} \to \mathrm{R}$, $a_2 : \mathrm{Q} \times \mathrm{Q} \to \mathrm{R}$, $b_1 : \mathbf{H} \times \mathrm{Z} \to \mathrm{R}$, $b_2 : \mathrm{Q} \times \mathrm{Z} \to \mathrm{R}$, $c : \mathrm{Z} \times \mathrm{Z} \to \mathrm{R}$, and linear functionals $F : \mathbf{H} \to \mathrm{R}$, $G : \mathrm{Q} \to \mathrm{R}$ are specified in the following way

$$a_1(\boldsymbol{u}, \boldsymbol{v}) := 2\mu \int_\Omega \boldsymbol{\varepsilon}(\boldsymbol{u}) : \boldsymbol{\varepsilon}(\boldsymbol{v}), \qquad a_2(p, q) := \left(\frac{c_0}{\alpha} + \frac{1}{\lambda}\right) \int_\Omega pq + \frac{1}{\alpha\eta} \int_\Omega \kappa \nabla p \cdot \nabla q, \tag{2.11}$$

$$b_1(\boldsymbol{v}, \psi) := -\int_\Omega \psi \operatorname{div} \boldsymbol{v}, \qquad b_2(q, \psi) := \frac{1}{\lambda} \int_\Omega q\psi, \qquad c(\phi, \psi) := \frac{1}{\lambda} \int_\Omega \phi\psi, \tag{2.12}$$

$$F(\boldsymbol{v}) := \int_\Omega \boldsymbol{f} \cdot \boldsymbol{v}, \qquad G(q) := \frac{\rho}{\alpha\eta} \int_\Omega \kappa \boldsymbol{g} \cdot \nabla q - \frac{\rho}{\alpha\eta} \langle \kappa \boldsymbol{g} \cdot \boldsymbol{n}, q \rangle_{\Gamma_{\boldsymbol{u}}} + \frac{1}{\alpha} \int_\Omega sq, \tag{2.13}$$

where $\langle \cdot, \cdot \rangle_{\Gamma_{\boldsymbol{u}}}$ stands for the duality pairing between $\mathrm{H}^{-1/2}_{00}(\Gamma_{\boldsymbol{u}})$ and $\mathrm{H}^{1/2}_{00}(\Gamma_{\boldsymbol{u}})$ and

$$\mathrm{H}^{1/2}_{00}(\Gamma_{\boldsymbol{u}}) := \{q|_{\Gamma_{\boldsymbol{u}}} : q \in \mathrm{H}^1(\Omega), \quad q = 0 \quad \text{on} \quad \Gamma_p\}.$$

In addition, we also define an auxiliary uncoupled displacement-volumetric stress problem as: find $(\boldsymbol{u}, \phi) \in \mathbf{H} \times \mathrm{Z}$ such that

$$\mathcal{M}^{\pm}((\boldsymbol{u}, \phi), (\boldsymbol{v}, \psi)) = \mathcal{H}(\boldsymbol{v}, \psi), \qquad \forall (\boldsymbol{v}, \psi) \in \mathbf{H} \times \mathrm{Z}, \tag{2.14}$$

where

$$\mathcal{M}^{\pm}((\boldsymbol{u}, \phi), (\boldsymbol{v}, \psi)) := a_1(\boldsymbol{u}, \boldsymbol{v}) + b_1(\boldsymbol{v}, \phi) \pm b_1(\boldsymbol{u}, \psi), \quad \text{and} \quad \mathcal{H}(\boldsymbol{v}, \psi) = F(\boldsymbol{v}),$$

for all $(\boldsymbol{u}, \phi), (\boldsymbol{v}, \psi) \in \mathbf{H} \times \mathrm{Z}$.

2.3 Stability

Let us now discuss the stability properties of the bilinear forms and functionals appearing in (2.8)–(2.10). We start by observing that all the bilinear forms are bounded:

$$
\begin{aligned}
&|a_1(\boldsymbol{u},\boldsymbol{v})| \le 2\mu C_{k,2}\|\boldsymbol{u}\|_{1,\Omega}\|\boldsymbol{v}\|_{1,\Omega}, \quad \boldsymbol{u},\boldsymbol{v}\in\mathbf{H},\\
&|a_2(p,q)| \le \max\left\{\frac{c_0}{\alpha}+\frac{1}{\lambda},\frac{\kappa_2}{\alpha\eta}\right\}\|p\|_{1,\Omega}\|q\|_{1,\Omega}, \quad p,q\in\mathrm{Q},\\
&|b_1(\boldsymbol{v},\psi)| \le \sqrt{n}\|\boldsymbol{v}\|_{1,\Omega}\|\psi\|_{0,\Omega}, \quad \boldsymbol{v}\in\mathbf{H},\ \psi\in\mathrm{Z},\\
&|b_2(q,\psi)| \le \lambda^{-1}\|q\|_{1,\Omega}\|\psi\|_{0,\Omega}, \quad q\in\mathrm{Q},\ \psi\in\mathrm{Z},\\
&|c(\phi,\psi)| \le \lambda^{-1}\|\phi\|_{0,\Omega}\|\psi\|_{0,\Omega}, \quad \phi,\psi\in\mathrm{Z}.
\end{aligned}
\tag{2.15}
$$

Above, $C_{k,2}$ is one of the positive constant satisfying

$$
C_{k,1}\|\boldsymbol{v}\|_{1,\Omega}^2 \le \|\boldsymbol{\varepsilon}(\boldsymbol{v})\|_{0,\Omega}^2 \le C_{k,2}\|\boldsymbol{v}\|_{1,\Omega}^2, \quad \forall\,\boldsymbol{v}\in\mathbf{H}. \tag{2.16}
$$

In turn, the functionals F and G are also bounded:

$$
\begin{aligned}
&|F(\boldsymbol{v})| \le \|\boldsymbol{f}\|_{0,\Omega}\|\boldsymbol{v}\|_{1,\Omega}, \quad \boldsymbol{v}\in\mathbf{H},\\
&|G(q)| \le \alpha^{-1}\left(\rho\eta^{-1}\kappa_2\|\boldsymbol{g}\|_{0,\Omega}+\rho\eta^{-1}\kappa_2 C_\Gamma\|\boldsymbol{g}\cdot\boldsymbol{n}\|_{-1/2,\Gamma_{\boldsymbol{u}}}+\|s\|_{0,\Omega}\right)\|q\|_{1,\Omega}, \quad q\in\mathrm{Q},
\end{aligned}
\tag{2.17}
$$

where $C_\Gamma>0$ is the continuity constant of the trace operator.

We now review the positivity of the forms a_1, a_2 and c. By using the inequality (2.16), the uniform lower bound of κ, and according to the definition of the forms a_1, a_2 and c, it readily follows that

$$
\begin{aligned}
&a_1(\boldsymbol{v},\boldsymbol{v}) \ge 2\mu C_{k,1}\|\boldsymbol{v}\|_{1,\Omega}^2, \quad \forall\,\boldsymbol{v}\in\mathbf{H},\\
&a_2(q,q) \ge \alpha^{-1}\max\{c_0,\kappa_1\eta^{-1}\}\|q\|_{1,\Omega}^2+\lambda^{-1}\|q\|_{0,\Omega}^2, \quad \forall\,q\in\mathrm{Q}\\
&c(\psi,\psi) = \lambda^{-1}\|\psi\|_{0,\Omega}^2, \quad \forall\,\psi\in\mathrm{Z}.
\end{aligned}
\tag{2.18}
$$

Finally, the bilinear form b_1 satisfies the continuous inf-sup condition (see e.g. [17]):

$$
\sup_{\boldsymbol{v}\in\mathbf{H}\setminus\mathbf{0}}\frac{b_1(\boldsymbol{v},\psi)}{\|\boldsymbol{v}\|_{1,\Omega}} \ge \beta\|\psi\|_{0,\Omega} \quad \forall\,\psi\in\mathrm{Z}, \tag{2.19}
$$

with $\beta>0$ only depending on Ω.

2.4 Solvabilty and continuous dependence result

Now, we establish the well-posedness of problem (2.8)–(2.10). We start with the corresponding continuous dependence result.

Lemma 2.1 *Let $(\boldsymbol{u},p,\phi)\in\mathbf{H}\times\mathrm{Q}\times\mathrm{Z}$ be a solution of the system (2.8)–(2.10). Then there exists $C_{stab}>0$ independent of λ, such that*

$$
\|\boldsymbol{u}\|_{1,\Omega}+\|p\|_{1,\Omega}+\|\phi\|_{0,\Omega} \le C_{stab}\left(\|\boldsymbol{f}\|_{0,\Omega}+\|\boldsymbol{g}\|_{0,\Omega}+\|\boldsymbol{g}\cdot\boldsymbol{n}\|_{-1/2,\Gamma_{\boldsymbol{u}}}+\|s\|_{0,\Omega}\right).
$$

Proof. First, choosing $\boldsymbol{v} = \boldsymbol{u}$ in (2.8) and $\psi = \phi$ in (2.10), and combining both equations, we easily obtain

$$a_1(\boldsymbol{u}, \boldsymbol{u}) - b_2(p, \phi) + c(\phi, \phi) = F(\boldsymbol{u}),$$

which together with the positivity of a_1 and c in (2.18), and the continuity of b_2 and F in (2.15) and (2.17), respectively, implies

$$2\mu C_{k,1}\|\boldsymbol{u}\|_{1,\Omega}^2 - \lambda^{-1}\|p\|_{1,\Omega}\|\phi\|_{0,\Omega} + \lambda^{-1}\|\phi\|_{0,\Omega}^2 \le \|\boldsymbol{f}\|_{0,\Omega}\|\boldsymbol{u}\|_{1,\Omega}. \tag{2.20}$$

In turn, choosing $q = p$ in (2.9), we have

$$a_2(p, p) - b_2(p, \phi) = G(p),$$

which in combination with the positivity of a_2 in (2.18) and the continuity of b_2 and G in (2.15) and (2.17), respectively, implies

$$\begin{aligned} &\alpha^{-1}\max\{c_0, \kappa_1\eta^{-1}\}\|p\|_{1,\Omega}^2 + \lambda^{-1}\|p\|_{0,\Omega}^2 - \lambda^{-1}\|p\|_{1,\Omega}\|\phi\|_{0,\Omega} \\ &\le \alpha^{-1}\left(\rho\eta^{-1}\kappa_2\|\boldsymbol{g}\|_{0,\Omega} + \rho\eta^{-1}\kappa_2 C_\Gamma\|\boldsymbol{g}\cdot\boldsymbol{n}\|_{-1/2,\Gamma_{\boldsymbol{u}}} + \|s\|_{0,\Omega}\right)\|p\|_{1,\Omega}. \end{aligned} \tag{2.21}$$

Then, adding (2.20) and (2.21), and utilizing the inequality $-2ab \ge -a^2 - b^2$, we get

$$\begin{aligned} &2\mu C_{k,1}\|\boldsymbol{u}\|_{1,\Omega}^2 + \alpha^{-1}\max\{c_0, \kappa_1\eta^{-1}\}\|p\|_{1,\Omega}^2 \\ &\le \|\boldsymbol{f}\|_{0,\Omega}\|\boldsymbol{u}\|_{1,\Omega} + \alpha^{-1}\left(\rho\eta^{-1}\kappa_2\|\boldsymbol{g}\|_{0,\Omega} + \rho\eta^{-1}\kappa_2 C_\Gamma\|\boldsymbol{g}\cdot\boldsymbol{n}\|_{-1/2,\Gamma_{\boldsymbol{u}}} + \|s\|_{0,\Omega}\right)\|p\|_{1,\Omega}, \end{aligned} \tag{2.22}$$

which readily gives

$$\|\boldsymbol{u}\|_{1,\Omega} + \|p\|_{1,\Omega} \le c\left(\|\boldsymbol{f}\|_{0,\Omega} + \|\boldsymbol{g}\|_{0,\Omega} + \|\boldsymbol{g}\cdot\boldsymbol{n}\|_{-1/2,\Gamma_{\boldsymbol{u}}} + \|s\|_{0,\Omega}\right), \tag{2.23}$$

with $c > 0$ independent of λ.

Now, from the inf-sup condition (2.19) with $\psi = \phi$ and using (2.8), we obtain

$$\beta\|\phi\|_{0,\Omega} \le \sup_{\boldsymbol{v}\in\mathbf{H}\backslash\mathbf{0}} \frac{b_1(\boldsymbol{v}, \phi)}{\|\boldsymbol{v}\|_{1,\Omega}} = \sup_{\boldsymbol{v}\in\mathbf{H}\backslash\mathbf{0}} \frac{F(\boldsymbol{v}) - a_1(\boldsymbol{u}, \boldsymbol{v})}{\|\boldsymbol{v}\|_{1,\Omega}} \le \|\boldsymbol{f}\|_{0,\Omega} + 2\mu C_{k,2}\|\boldsymbol{u}\|_{1,\Omega}, \tag{2.24}$$

which combined with (2.23), implies

$$\|\phi\|_{0,\Omega} \le \tilde{c}\left(\|\boldsymbol{f}\|_{0,\Omega} + \|\boldsymbol{g}\|_{0,\Omega} + \|\boldsymbol{g}\cdot\boldsymbol{n}\|_{-1/2,\Gamma_{\boldsymbol{u}}} + \|s\|_{0,\Omega}\right).$$

The latter bound and inequality (2.23) imply the desired estimate, which concludes the proof. □

Next, we address the unique solvability of (2.8)–(2.10). To that end, we first observe that due to the nonsymmetry of (2.8)-(2.10), its solvability analysis cannot be straightforwardly placed in the framework of the classical Babuška-Brezzi theory. We therefore redefine system (2.8)–(2.10) as the following operator problem: Find $\vec{\boldsymbol{u}} := (\boldsymbol{u}, p, \phi) \in \mathbb{V} := \mathbf{H}\times \mathrm{Q}\times \mathrm{Z}$, such that

$$(\mathcal{A} + \mathcal{K})\vec{\boldsymbol{u}} = \mathcal{F}, \tag{2.25}$$

where $\mathcal{A} : \mathbb{V} \to \mathbb{V}$, $\mathcal{K} : \mathbb{V} \to \mathbb{V}$ and $\mathcal{F} \in \mathbb{V}$ are defined as:

$$
\begin{aligned}
\langle \mathcal{A}(\vec{\boldsymbol{u}}), \vec{\boldsymbol{v}} \rangle_{\mathbb{V}\times\mathbb{V}} &:= a_1(\boldsymbol{u}, \boldsymbol{v}) + b_1(\boldsymbol{v}, \phi) - b_1(\boldsymbol{u}, \psi) + c(\phi, \psi) + a_2(p, q) \\
\langle \mathcal{K}(\vec{\boldsymbol{u}}), \vec{\boldsymbol{v}} \rangle_{\mathbb{V}\times\mathbb{V}} &:= b_2(p, \psi) - b_2(q, \phi) \\
\langle \mathcal{F}, \vec{\boldsymbol{v}} \rangle_{\mathbb{V}\times\mathbb{V}} &:= F(\boldsymbol{v}) + G(q),
\end{aligned}
\tag{2.26}
$$

for all $\vec{\boldsymbol{u}} = (\boldsymbol{u}, p, \phi), \vec{\boldsymbol{v}} = (\boldsymbol{v}, q, \psi) \in \mathbb{V}$.

In this way, similarly to [9, 13], if one proves that $\mathcal{A}$ is invertible, $\mathcal{K}$ is compact and $(\mathcal{A} + \mathcal{K})$ is injective, then the Fredholm alternative theory implies unique solvability of (2.25), and equivalently of (2.8)–(2.10).

We begin by proving the compactness of $\mathcal{K}$.

Lemma 2.2 *The operator $\mathcal{K}$ defined in (2.26) is compact.*

Proof. Let $\mathbb{B}_2 : \mathrm{Q} \to \mathrm{Z}$ be the operator induced by the bilinear form b_2, that is, the operator defined by

$$
\langle \mathbb{B}_2(q), \psi \rangle_{0,\Omega} = b_2(q, \psi) = \frac{1}{\lambda} \int_\Omega q\psi \quad \forall\, q \in \mathrm{Q},\ \forall\, \psi \in \mathrm{Z},
$$

where $\langle \cdot, \cdot \rangle_{0,\Omega}$ denotes the inner product in $L^2(\Omega)$. Moreover, let $I : L^2(\Omega) \to L^2(\Omega)$ be the identity operator and i_c be the compact embedding from $\mathrm{H}^1(\Omega)$ into $\mathrm{L}^2(\Omega)$. Then, it is straightforward to realize that $\mathbb{B}_2 = \lambda^{-1} I \circ i_c$, which implies that $\mathbb{B}_2$ is a compact operator, and so is $\mathbb{B}_2^*$.

Owing to the above, and noting that $\mathcal{K}(\vec{\boldsymbol{u}}) = (\mathbf{0}, \mathbb{B}_2(p), -\mathbb{B}_2^*(\phi))$ for all $\vec{\boldsymbol{u}} = (\boldsymbol{u}, p, \phi)$, we conclude the proof. □

We continue with the invertibility of $\mathcal{A}$.

Lemma 2.3 *The operator $\mathcal{A}$ defined in (2.26) is invertible.*

Proof. Given $\mathcal{F} := (\mathcal{F}_{\mathrm{H}}, \mathcal{F}_{\mathrm{Q}}, \mathcal{F}_{\mathrm{Z}}) \in \mathbb{V}$, we first observe that proving the invertibility of $\mathcal{A}$, is equivalent to proving the existence of a unique $\vec{\boldsymbol{u}} \in \mathbb{V}$, such that

$$
\mathcal{A}(\vec{\boldsymbol{u}}) = \mathcal{F}, \tag{2.27}
$$

which in turn is equivalent to proving the unique solvability of the two uncoupled problems: Find $(\boldsymbol{u}, \phi) \in \mathbf{H} \times \mathrm{Z}$, such that

$$
\begin{aligned}
a_1(\boldsymbol{u}, \boldsymbol{v}) + b_1(\boldsymbol{v}, \phi) &= F_{\mathrm{H}}(\boldsymbol{v}) \quad \forall\, \boldsymbol{v} \in \mathbf{H}, \\
b_1(\boldsymbol{u}, \psi) - c(\phi, \psi) &= F_{\mathrm{Z}}(\psi) \quad \forall\, \psi \in \mathrm{Z},
\end{aligned}
\tag{2.28}
$$

and: Find $p \in \mathrm{Q}$, such that

$$
a_2(p, q) = F_{\mathrm{Q}}(q) \quad \forall\, q \in \mathrm{Q}, \tag{2.29}
$$

where F_{H}, F_{Q} and F_{Z} are the functionals induced by $\mathcal{F}_{\mathrm{H}}$, $\mathcal{F}_{\mathrm{Q}}$ and $\mathcal{F}_{\mathrm{Z}}$, respectively.

According to the stability properties of the forms a_1, b_1 and c discussed above, namely, continuity of a_1, b_1, and c, inf-sup of b_1, ellipticity of a_1 and positive-semidefinitivity of c, the well-posedness of (2.28) follows from a straightforward application of [16, Lemma 3.4]. In turn, owing to the ellipticity and continuity of a_2, the unique solvability of (2.29) holds by virtue of the Lax–Milgram lemma. □

The last step consists in proving injectivity of the full operator $(\mathcal{A} + \mathcal{K})$.

Lemma 2.4 *The map $(\mathcal{A} + \mathcal{K})$ is one-to-one.*

Proof. It suffices to show that the unique solution to the homogeneous problem

$$\begin{aligned} a_1(\boldsymbol{u},\boldsymbol{v}) \qquad\qquad\qquad + b_1(\boldsymbol{v},\phi) &= 0 \qquad \forall \boldsymbol{v} \in \mathbf{H}, && (2.30)\\ a_2(p,q) - b_2(q,\phi) &= 0 \qquad \forall q \in \mathrm{Q}, && (2.31)\\ b_1(\boldsymbol{u},\psi) + b_2(p,\psi) - c(\phi,\psi) &= 0 \qquad \forall \psi \in \mathrm{Z}, && (2.32)\end{aligned}$$

is the null vector in $\mathbb{V}$. To that end, we apply basically the same steps in the proof of Lemma 2.1. In fact, we let $(\boldsymbol{u}, p, \phi) \in \mathbb{V}$ be the solution of (2.30)–(2.32), choose $\boldsymbol{v} = \boldsymbol{u}$ and $\psi = \phi$ in (2.30) and (2.32), respectively, and combine the two equations, to obtain

$$a_1(\boldsymbol{u},\boldsymbol{u}) - b_2(p,\phi) + c(\phi,\phi) = 0. \qquad (2.33)$$

Then, by choosing $q = p$ in (2.30) and adding the resulting equation to (2.33), we obtain

$$a_1(\boldsymbol{u},\boldsymbol{u}) + a_2(p,p) - 2b_2(p,\phi) + c(\phi,\phi) = 0,$$

which, along with the positivity of a_1, a_2 and c in (2.18), the continuity of b_2 in (2.15), and the inequality $-2ab \geq -a^2 - b^2$, implies

$$2\mu C_{k,1}\|\boldsymbol{u}\|^2_{1,\Omega} + \alpha^{-1}\max\{c_0, \kappa_1\eta^{-1}\}\|p\|^2_{1,\Omega} \leq 0.$$

From the previous inequality we readily infer that $\boldsymbol{u} = \mathbf{0}$ and $p = 0$. In turn, by applying the inf-sup condition of b_1 in (2.19) with $\psi = \phi$, and using (2.30) and the continuity of a_1, we easily obtain

$$\beta\|\phi\|_{0,\Omega} \leq \sup_{\boldsymbol{v}\in\mathbf{H}\setminus\mathbf{0}} \frac{|b_1(\boldsymbol{v},\phi)|}{\|\boldsymbol{v}\|_{1,\Omega}} = \sup_{\boldsymbol{v}\in\mathbf{H}\setminus\mathbf{0}} \frac{|a_1(\boldsymbol{u},\boldsymbol{v})|}{\|\boldsymbol{v}\|_{1,\Omega}} \leq 2\mu C_{k,2}\|\boldsymbol{u}\|_{1,\Omega},$$

which implies that $\phi = 0$ and concludes the proof. □

Combination of Lemmas 2.1, 2.2, 2.3 and 2.4 with the Fredholm alternative theory for compact operators, implies the main result of this section, stated in the following theorem.

Theorem 2.5 *Given $\boldsymbol{f} \in \mathbf{L}^2(\Omega)$, $\boldsymbol{g} \in \mathbf{L}^2(\Omega)$ and $s \in L^2(\Omega)$, there exists a unique solution $(\boldsymbol{u}, p, \phi) \in \mathbf{H} \times \mathrm{Q} \times \mathrm{Z}$ to the coupled problem (2.8)-(2.10). Moreover, there exists a positive constant C_{stab}, independent of λ, such that*

$$\|\boldsymbol{u}\|_{1,\Omega} + \|p\|_{1,\Omega} + \|\phi\|_{0,\Omega} \leq C_{stab}\left(\|\boldsymbol{f}\|_{0,\Omega} + \|\boldsymbol{g}\|_{0,\Omega} + \|\boldsymbol{g}\cdot\boldsymbol{n}\|_{-1/2,\Gamma_{\boldsymbol{u}}} + \|s\|_{0,\Omega}\right).$$

3 The Galerkin method

In this section we introduce the Galerkin scheme of (2.8)–(2.10). By considering arbitrary finite dimensional subspaces we analyze its solvability and provide the corresponding Céa's estimate. We begin by introducing the generic discrete spaces

$$\mathbf{H}_h \subseteq \mathbf{H}, \qquad \mathrm{Q}_h \subseteq \mathrm{Q}, \qquad \text{and} \quad \mathrm{Z}_h \subseteq \mathrm{Z},$$

where the subscript h stands for the size of a regular triangulation $\mathcal{T}_h$ of $\bar{\Omega}$ made up of triangles K (when $d = 2$) or tetrahedra K, (when $d = 3$) of diameter h_K, that is $h := \max\{h_K : K \in \mathcal{T}_h\}$.

In this way, the Galerkin scheme associated to (2.8)–(2.10) reads: Find $\boldsymbol{u}_h \in \mathbf{H}_h$, $p_h \in \mathrm{Q}_h$ and $\phi_h \in \mathrm{Z}_h$, such that

$$\begin{aligned} a_1(\boldsymbol{u}_h, \boldsymbol{v}_h) \qquad\qquad\qquad\quad + b_1(\boldsymbol{v}_h, \phi_h) &= F(\boldsymbol{v}_h) \qquad \forall \boldsymbol{v}_h \in \mathbf{H}_h, & (3.1)\\ a_2(p_h, q_h) - b_2(q_h, \phi_h) &= G(q_h) \qquad \forall q_h \in \mathrm{Q}_h, & (3.2)\\ b_1(\boldsymbol{u}_h, \psi_h) + b_2(p_h, \psi_h) - c(\phi_h, \psi_h) &= 0 \qquad \forall \psi_h \in \mathrm{Z}_h, & (3.3) \end{aligned}$$

where the bilinear forms a_1, a_2, b_1, b_2, c and the functionals F and G are defined in (2.11)–(2.13).

3.1 Existence and uniqueness of solution

It is clear that all the bilinear forms and functionals preserve the stability properties (2.15) and (2.17) on the corresponding discrete spaces. In addition, the bilinear forms a_1, a_2 and c preserve the positivity properties (2.18) on $\mathbf{H}_h$, Q_h and Z_h, respectively. However, the inf-sup condition (2.19) is not necessarily inherited at the discrete level, reason why, from now on we assume that there exists a positive constant $\hat{\beta}$, independent of h, such that

$$\sup_{\boldsymbol{v}_h \in \mathbf{H}_h \setminus \mathbf{0}} \frac{b_1(\boldsymbol{v}_h, \psi_h)}{\|\boldsymbol{v}_h\|_{1,\Omega}} \geq \hat{\beta} \|\psi\|_{0,\Omega} \quad \forall\, \psi_h \in \mathrm{Z}_h. \qquad (3.4)$$

As we will see next in Section 3.3, the pair $(\mathbf{H}_h, \mathrm{Z}_h)$ can be chosen as a pair of stable finite element subspaces for the Stokes problem.

The following theorem establishes the well-posedness of the Galerkin scheme (3.1)-(3.3).

Theorem 3.1 *Assume that the discrete inf-sup condition (3.4) holds. Then, given $\boldsymbol{f} \in \mathbf{L}^2(\Omega)$, $\boldsymbol{g} \in \mathbf{L}^2(\Omega)$ and $s \in L^2(\Omega)$, there exists a unique solution $(\boldsymbol{u}_h, p_h, \phi_h) \in \mathbf{H}_h \times \mathrm{Q}_h \times \mathrm{Z}_h$ to the discrete coupled problem (3.1)-(3.3). Moreover, there exists a positive constant $\hat{C}_{stab}$, independent of h and λ, such that*

$$\|\boldsymbol{u}_h\|_{1,\Omega} + \|p_h\|_{1,\Omega} + \|\phi_h\|_{0,\Omega} \leq \hat{C}_{stab} \left(\|\boldsymbol{f}\|_{0,\Omega} + \|\boldsymbol{g}\|_{0,\Omega} + \|\boldsymbol{g} \cdot \boldsymbol{n}\|_{-1/2,\Gamma_{\boldsymbol{u}}} + \|s\|_{0,\Omega} \right). \qquad (3.5)$$

Proof. Since $\mathbf{H}_h$, Q_h and Z_h are finite dimensional spaces, for the solvability analysis it suffices to prove that the solution of the homogeneous problem is the trivial one. To do that, we let $\boldsymbol{u}_h \in \mathbf{H}_h$, $p_h \in \mathrm{Q}_h$ and $\phi_h \in \mathrm{Z}_h$ be the solution of (3.1)-(3.3) with $\boldsymbol{f} = \mathbf{0}$, $\boldsymbol{g} = \mathbf{0}$ and $s = 0$. Then, proceeding identically as in the proof of Lemma 2.4, that is, combining equations (3.1) and (3.3) with $\boldsymbol{v}_h = \boldsymbol{u}_h$ and $\psi_h = \phi_h$, respectively, adding (3.3) with $q_h = p_h$ to the resulting equation, and using the positivity of a_1, a_2 and c in (2.18), the continuity of b_2 in (2.15), and the inequality $-2ab \geq -a^2 - b^2$, we obtain

$$2\mu C_{k,1}\|\boldsymbol{u}_h\|_{1,\Omega}^2 + \alpha^{-1}\max\{c_0, \kappa_1\eta^{-1}\}\|p_h\|_{1,\Omega}^2 \leq 0,$$

from which $\boldsymbol{u}_h = \mathbf{0}$ and $p_h = 0$. Furthermore, from the inf-sup condition (3.4) with $\psi_h = \phi_h$, and the first equation of (3.1), we easily obtain $\phi_h = 0$.

Similarly, the continuous dependence result (3.5) can be derived following exactly the same steps of the proof of Lemma 2.1. We omit further details. □

3.2 A priori error estimate

We now derive the corresponding Céa's estimate. This result is established next.

Theorem 3.2 *Assume that the discrete inf-sup condition* (3.4) *holds. Let* $(\boldsymbol{u}, p, \phi) \in \mathbf{H} \times \mathrm{Q} \times \mathrm{Z}$ *and* $(\boldsymbol{u}_h, p_h, \phi_h) \in \mathbf{H}_h \times \mathrm{Q}_h \times \mathrm{Z}_h$ *be the unique solutions of the continuous and discrete coupled problems* (2.8)-(2.10) *and* (3.1)-(3.3)*, respectively. Then, there exists* $C_{Céa} > 0$*, independent of* h *and* λ*, such that*

$$\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}+\|p-p_h\|_{1,\Omega}+\|\phi-\phi_h\|_{0,\Omega} \leq C_{Céa}\big(\operatorname{dist}(\boldsymbol{u},\mathbf{H}_h)+\operatorname{dist}(p,\mathrm{Q}_h)+\operatorname{dist}(\phi,\mathrm{Z}_h)\big). \quad (3.6)$$

Proof. Let us first introduce the discrete space

$$\mathbf{K}_h := \big\{\boldsymbol{v}_h \in \mathbf{H}_h \,:\, b_1(\boldsymbol{v}_h, \psi_h) = -b_2(p_h, \psi_h) + c(\phi_h, \psi_h), \quad \forall\, \psi_h \in \mathrm{Z}_h\big\},$$

which is clearly non-empty since $\boldsymbol{u}_h \in \mathbf{K}_h$ and since the discrete inf-sup condition (3.4) holds. In addition, it is not difficult to see that the following inequality holds (see for instance [12, Theorem 2.6]):

$$\operatorname{dist}(\boldsymbol{u}, \mathbf{K}_h) \leq C\operatorname{dist}(\boldsymbol{u}, \mathbf{H}_h). \quad (3.7)$$

Next, in order to simplify the subsequent analysis, we write $\mathbf{e}_{\boldsymbol{u}} = \boldsymbol{u} - \boldsymbol{u}_h$, $\mathrm{e}_p = p - p_h$ and $\mathrm{e}_\phi = \phi - \phi_h$. As usual, for arbitrary $\widehat{\boldsymbol{v}}_h \in \mathbf{K}_h$, $\widehat{q}_h \in \mathrm{Q}_h$ and $\widehat{\psi}_h \in \mathrm{Z}_h$, we shall decompose these errors into

$$\mathbf{e}_{\boldsymbol{u}} = \mathbf{r}_{\boldsymbol{u}} + \boldsymbol{\chi}_{\boldsymbol{u}}, \quad \mathrm{e}_p = \mathrm{r}_p + \chi_p, \quad \text{and} \quad \mathrm{e}_\phi = \mathrm{r}_\phi + \chi_\phi, \quad (3.8)$$

with

$$\begin{aligned} \mathbf{r}_{\boldsymbol{u}} &:= \boldsymbol{u} - \widehat{\boldsymbol{v}}_h \in \mathbf{H}, \quad \boldsymbol{\chi}_{\boldsymbol{u}} := \widehat{\boldsymbol{v}}_h - \boldsymbol{u}_h \in \mathbf{H}_h, \\ \mathrm{r}_p &:= p - \widehat{q}_h \in \mathrm{Q}, \qquad \chi_p := \widehat{q}_h - p_h \in \mathrm{Q}_h, \\ \mathrm{r}_\phi &:= \phi - \widehat{\psi}_h \in \mathrm{Z}, \qquad \chi_\phi := \widehat{\psi}_h - \phi_h \in \mathrm{Z}_h. \end{aligned} \quad (3.9)$$

Notice that $\boldsymbol{\chi}_{\boldsymbol{u}} \in \mathrm{Ker}_h(b_1) := \big\{\boldsymbol{v}_h \in \mathbf{H}_h \,:\, b_1(\boldsymbol{v}_h, \psi_h) = 0, \quad \forall\, \psi_h \in \mathrm{Z}_h\big\}$. Observe also that proving the existence of a positive constant C, independent of h and λ, such that

$$\|\boldsymbol{\chi}_{\boldsymbol{u}}\|_{1,\Omega} + \|\chi_p\|_{1,\Omega} + \|\chi_\phi\|_{0,\Omega} \leq C(\|\mathbf{r}_{\boldsymbol{u}}\|_{1,\Omega} + \|\mathrm{r}_p\|_{1,\Omega} + \|\mathrm{r}_\phi\|_{0,\Omega}), \quad (3.10)$$

then we can simply use the triangle inequality and the fact that $\widehat{\boldsymbol{v}}_h$, $\widehat{q}_h$ and $\widehat{\psi}_h$ are arbitrary, to obtain

$$\|\mathbf{e}_{\boldsymbol{u}}\|_{1,\Omega} + \|\mathrm{e}_p\|_{1,\Omega} + \|\mathrm{e}_\phi\|_{0,\Omega} \leq (1 + C)\big(\operatorname{dist}(\boldsymbol{u}, \mathbf{K}_h) + \operatorname{dist}(p, \mathrm{Q}_h) + \operatorname{dist}(\phi, \mathrm{Z}_h)\big),$$

which together to (3.7) implies (3.6). Therefore, in the sequel we focus on proving (3.10). To that end, we first establish the corresponding Galerkin orthogonality property:

$$a_1(\mathbf{e}_{\boldsymbol{u}}, \boldsymbol{v}_h) \qquad\qquad + b_1(\boldsymbol{v}_h, \mathrm{e}_\phi) = 0 \qquad \forall \boldsymbol{v}_h \in \mathbf{H}_h, \tag{3.11}$$
$$a_2(\mathrm{e}_p, q_h) - \ b_2(q_h, \mathrm{e}_\phi) = 0 \qquad \forall q_h \in \mathrm{Q}_h, \tag{3.12}$$
$$b_1(\mathbf{e}_{\boldsymbol{u}}, \psi_h) + b_2(\mathrm{e}_p, \psi_h) - \ c(\mathrm{e}_\phi, \psi_h) = 0 \qquad \forall \psi_h \in \mathrm{Z}_h. \tag{3.13}$$

Then, from (3.11) with $\boldsymbol{v}_h = \boldsymbol{\chi}_{\boldsymbol{u}} \in \mathrm{Ker}_h(b_1)$, and considering the decomposition (3.8), we have

$$a_1(\boldsymbol{\chi}_{\boldsymbol{u}}, \boldsymbol{\chi}_{\boldsymbol{u}}) = -a_1(\mathbf{r}_{\boldsymbol{u}}, \boldsymbol{\chi}_{\boldsymbol{u}}) - b_1(\boldsymbol{\chi}_{\boldsymbol{u}}, \mathrm{r}_\phi),$$

which together to the ellipticity of a_1 (cf. (2.18)) and the continuity of a_1 and b_1 (cf. (2.15)), implies

$$\|\boldsymbol{\chi}_{\boldsymbol{u}}\|_{1,\Omega} \leq C_1\{\|\mathbf{r}_{\boldsymbol{u}}\|_{1,\Omega} + \|\mathrm{r}_\phi\|_{0,\Omega}\}, \tag{3.14}$$

with $C_1 > 0$, independent of h and λ. Notice that the latter inequality implies

$$\|\mathbf{e}_{\boldsymbol{u}}\|_{1,\Omega} \leq (1 + C_1)\|\mathbf{r}_{\boldsymbol{u}}\|_{1,\Omega} + C_1\|\mathrm{r}_\phi\|_{0,\Omega}\}. \tag{3.15}$$

In turn, from the inf-sup condition (3.4), the first equation of (3.11) and the continuity of a_1 and b_1 (cf. (2.15)), we have

$$\begin{aligned} \|\chi_\phi\|_{0,\Omega} &\leq \beta^{-1} \sup_{\boldsymbol{v}_h \in \mathbf{H}_h \setminus \mathbf{0}} \frac{|b_1(\boldsymbol{v}_h, \chi_\phi)|}{\|\boldsymbol{v}_h\|_{1,\Omega}} = \beta^{-1} \sup_{\boldsymbol{v}_h \in \mathbf{H}_h \setminus \mathbf{0}} \frac{|a_1(\mathbf{e}_{\boldsymbol{u}}, \boldsymbol{v}_h) + b_1(\boldsymbol{v}_h, \mathrm{r}_\phi)|}{\|\boldsymbol{v}_h\|_{1,\Omega}} \\ &\leq \beta^{-1}\left(2\mu C_{k,2}\|\mathbf{e}_{\boldsymbol{u}}\|_{1,\Omega} + \sqrt{n}\|\mathrm{r}_\phi\|_{0,\Omega}\right), \end{aligned}$$

which together to (3.15), implies

$$\|\chi_\phi\|_{0,\Omega} \leq C_2(\|\mathbf{r}_{\boldsymbol{u}}\|_{1,\Omega} + \|\mathrm{r}_\phi\|_{0,\Omega}), \tag{3.16}$$

with $C_2 > 0$, independent of h and λ. In addition, similarly as before, we observe that (3.16) and the triangle inequality imply

$$\|\mathrm{e}_\phi\|_{0,\Omega} \leq C_2\|\mathbf{r}_{\boldsymbol{u}}\|_{1,\Omega} + (1 + C_2)\|\mathrm{r}_\phi\|_{0,\Omega}. \tag{3.17}$$

Finally, from (3.12), the ellipticity of a_2 (cf. (2.18)), and the continuity of a_2 and b_2, we obtain

$$\begin{aligned} \alpha^{-1}\max\{c_0, \kappa_1\eta^{-1}\}\|\chi_p\|_{1,\Omega}^2 &\leq a_2(\chi_p, \chi_p) = -a_2(\mathrm{r}_p, \chi_p) + b_2(\chi_p, \mathrm{e}_\phi) \\ &\leq \frac{1}{\alpha}\max\{c_0 + \frac{\alpha}{\lambda}, \frac{\kappa_2}{\eta}\}\|\mathrm{r}_p\|_{1,\Omega}\|\chi_p\|_{1,\Omega} + \lambda^{-1}\|\mathrm{e}_\phi\|_{0,\Omega}\|\chi_p\|_{1,\Omega}, \end{aligned}$$

which together to (3.17), implies

$$\|\chi_p\|_{1,\Omega} \leq C_3\left(\max\{c_0 + \frac{\alpha}{\lambda}, \frac{\kappa_2}{\eta}\}\|\mathrm{r}_p\|_{1,\Omega} + \lambda^{-1}\|\mathbf{r}_{\boldsymbol{u}}\|_{1,\Omega} + \lambda^{-1}\|\mathrm{r}_\phi\|_{0,\Omega}\right). \tag{3.18}$$

Therefore, summing up inequalities (3.14), (3.16) and (3.18), we get

$$\|\boldsymbol{\chi_u}\|_{1,\Omega} + \|\chi_p\|_{1,\Omega} + \|\chi_\phi\|_{0,\Omega} \le (C_1 + C_2 + \lambda^{-1}C_3)\|\mathbf{r_u}\|_{1,\Omega} + C_3 \max\{c_0 + \tfrac{\alpha}{\lambda}, \tfrac{\kappa_2}{\eta}\}\|\mathrm{r}_p\|_{1,\Omega}$$
$$+(C_1 + C_2 + \lambda^{-1}C_3)\|\mathrm{r}_\phi\|_{0,\Omega},$$

which yields the result. □

Remark 3.1 $\max\{c_0 + \frac{\alpha}{\lambda}, \frac{\kappa_2}{\eta}\}$ *and* $(C_1 + C_2 + \lambda^{-1}C_3)$ *in the previous inequality must be understood as constants independent of* λ *since, if* λ *goes to infinity (when the locking phenomenon occurs),* $\lambda^{-1}C_3$ *and* $\lambda^{-1}\alpha$ *are negligible.*

3.3 Particular choice of finite elements

Now, we provide three concrete examples of finite elements subspaces to approximate the solution of (2.8)–(2.10). To do that, given an integer $k \ge 0$ and a set S of R^d, in the sequel we denote by $\mathbb{P}_k(S)$ the space of polynomial functions on S of degree $\le k$.

Hood-Taylor + Lagrange. Let $k \ge 0$ be an integer. Then, the well known Hood-Taylor element (see, e.g. [17]) consists of the pair $(\mathbf{H}_h, \mathrm{Z}_h)$, where

$$\mathbf{H}_h := \left\{ \boldsymbol{v}_h \in [C(\overline{\Omega})]^d \,:\, \boldsymbol{v}_h\big|_K \in [\mathbb{P}_{k+2}(K)]^d \quad \forall\, K \in \mathcal{T}_h, \quad \boldsymbol{v}_h = 0 \text{ on } \Gamma_{\boldsymbol{u}} \right\}$$

and

$$\mathrm{Z}_h := \left\{ \psi_h \in C(\overline{\Omega}) \,:\, \psi_h\big|_K \in \mathbb{P}_{k+1}(K) \quad \forall\, K \in \mathcal{T}_h \right\}.$$

In turn, given an integer $l \ge 0$, to approximate the variable p we can simply choose the discrete space

$$\mathrm{Q}_h := \left\{ q_h \in C(\overline{\Omega}) \,:\, q_h\big|_K \in \mathbb{P}_{l+1}(K) \quad \forall\, K \in \mathcal{T}_h, \quad q_h = 0 \text{ on } \Gamma_p \right\}. \tag{3.19}$$

It is well known that the pair $(\mathbf{H}_h, \mathrm{Z}_h)$ satisfies the inf-sup condition (3.4) (see, for instance [5, 6, 17]). This fact and Theorem 3.1 imply the well-posedness of problem (3.1)–(3.3).

Let us now recall the approximation properties of the subspaces specified above.

($\mathbf{AP}_h^{\boldsymbol{u}}$) There exits $C > 0$, independent of h, such that for all $\boldsymbol{u} \in \mathbf{H}^{k+3}(\Omega)$, there holds

$$\inf_{\boldsymbol{v}_h \in \mathbf{H}_h} \|\boldsymbol{u} - \boldsymbol{v}_h\|_{1,\Omega} \le C h^{k+2} \|\boldsymbol{u}\|_{k+3,\Omega}.$$

($\mathbf{AP}_h^{p}$) There exits $C > 0$, independent of h, such that for all $p \in \mathrm{H}^{l+2}(\Omega)$, there holds

$$\inf_{q_h \in \mathrm{Q}_h} \|p - q_h\|_{1,\Omega} \le C h^{l+1} \|p\|_{l+2,\Omega}.$$

($\mathbf{AP}_h^{\phi}$) There exits $C > 0$, independent of h, such that for all $\phi \in \mathrm{H}^{k+2}(\Omega)$, there holds

$$\inf_{\psi_h \in \mathrm{Z}_h} \|\phi - \psi_h\|_{0,\Omega} \le C h^{k+2} \|\phi\|_{k+2,\Omega}.$$

Owing to these approximation properties, we now can establish the theoretical rate of convergence of our method.

Theorem 3.3 *Let* $(\boldsymbol{u}, p, \phi) \in \mathbf{H} \times \mathrm{Q} \times \mathrm{Z}$ *and* $(\boldsymbol{u}_h, p_h, \phi_h) \in \mathbf{H}_h \times \mathrm{Q}_h \times \mathrm{Z}_h$ *be the unique solutions of* (2.8)–(2.10) *and* (3.1)–(3.3)*, respectively. Given,* k*,* $l \geq 0$*, assume that* $\boldsymbol{u} \in \mathbf{H}^{k+3}(\Omega)$*,* $p \in \mathrm{H}^{l+1}(\Omega)$ *and* $\phi \in \mathrm{H}^{k+2}(\Omega)$*. Then, there exist* C_1*,* C_2*,* > 0*, independent of* h *and* λ*, such that*

$$\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}+\|p-p_h\|_{1,\Omega}+\|\phi-\phi_h\|_{0,\Omega} \leq C_1 h^{k+2}\big\{\|\boldsymbol{u}\|_{k+3,\Omega}+\|\phi\|_{k+2,\Omega}\big\}+C_2 h^{l+1}\|p\|_{l+2,\Omega}. \tag{3.20}$$

Proof. It follows from the Céa estimate (3.6), and the approximation properties $(\mathbf{AP}_h^{\boldsymbol{u}})$, $(\mathbf{AP}_h^{p})$ and $(\mathbf{AP}_h^{\phi})$.

□

MINI–element + Lagrange. In what follows, for the sake of conciseness of the presentation we restrict ourselves to the 2D case. For each $K \in \mathcal{T}_h$, we let $\mathbb{P}_{1,b}(K)$ be the space (see, e.g. [17])

$$\mathbb{P}_{1,b}(K) := [\mathbb{P}_1(K) \oplus \operatorname{span}\{b_K\}]^2,$$

where $b_K := \varphi_1\varphi_2\varphi_3$ is a $\mathbb{P}_3$ bubble function in K, and φ_1, φ_2, φ_3 are the barycentric coordinates of K. Then, the MINI-element (see, e.g. [17]) is the pair $(\mathbf{H}_h, \mathrm{Z}_h)$, where

$$\mathbf{H}_h := \left\{\boldsymbol{v}_h \in [C(\overline{\Omega})]^2 \,:\, \boldsymbol{v}_h\big|_K \in \mathbb{P}_{1,b}(K) \quad \forall\, K \in \mathcal{T}_h, \quad \boldsymbol{v}_h = 0 \text{ on } \Gamma_{\boldsymbol{u}}\right\}$$

and

$$\mathrm{Z}_h := \left\{\psi_h \in C(\overline{\Omega}) \,:\, \psi_h\big|_K \in \mathbb{P}_1(K) \quad \forall\, K \in \mathcal{T}_h\right\}.$$

In addition, to approximate the variable p we now choose the discrete space

$$\mathrm{Q}_h := \left\{q_h \in C(\overline{\Omega}) \,:\, q_h\big|_K \in \mathbb{P}_1(K) \quad \forall\, K \in \mathcal{T}_h, \quad q_h = 0 \text{ on } \Gamma_p\right\}.$$

As for the Hood–Taylor element defined above, it is well-known that the pair $(\mathbf{H}_h, \mathrm{Z}_h)$ satisfies the inf-sup condition (3.4) (see, for instance [10, 17]). Then, owing to Theorem 3.1, the discrete problem (3.1)–(3.3) defined with the subspaces above is clearly well posed.

Let us now recall the approximation properties of these subspaces.

$(\widehat{\mathbf{AP}}_h^{\boldsymbol{u}})$ There exits $C > 0$, independent of h, such that for all $\boldsymbol{u} \in \mathbf{H}^2(\Omega)$, there holds

$$\inf_{\boldsymbol{v}_h \in \mathbf{H}_h} \|\boldsymbol{u}-\boldsymbol{v}_h\|_{1,\Omega} \leq Ch\|\boldsymbol{u}\|_{2,\Omega}.$$

$(\widehat{\mathbf{AP}}_h^{p})$ There exits $C > 0$, independent of h, such that for all $p \in \mathrm{H}^2(\Omega)$, there holds

$$\inf_{q_h \in \mathrm{Q}_h} \|p-q_h\|_{1,\Omega} \leq Ch\|p\|_{2,\Omega}.$$

$(\widehat{\mathbf{AP}}_h^{\phi})$ There exits $C > 0$, independent of h, such that for all $\phi \in \mathrm{H}^1(\Omega)$, there holds

$$\inf_{\psi_h \in \mathrm{Z}_h} \|\phi-\psi_h\|_{0,\Omega} \leq Ch\|\phi\|_{1,\Omega}.$$

Owing to these approximation properties, we now can establish the theoretical rate of convergence of our method.

Theorem 3.4 *Let $(\boldsymbol{u}, p, \phi) \in \mathbf{H} \times \mathrm{Q} \times \mathrm{Z}$ and $(\boldsymbol{u}_h, p_h, \phi_h) \in \mathbf{H}_h \times \mathrm{Q}_h \times \mathrm{Z}_h$ be the unique solutions of (2.8)–(2.10) and (3.1)–(3.3), respectively. Assume that $\boldsymbol{u} \in \mathbf{H}^2(\Omega)$, $p \in \mathrm{H}^2(\Omega)$ and $\phi \in \mathrm{H}^1(\Omega)$. Then, there exist $C > 0$, independent of h and λ, such that*

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,\Omega} + \|p - p_h\|_{1,\Omega} + \|\phi - \phi_h\|_{0,\Omega} \leq Ch\big\{\|\boldsymbol{u}\|_{2,\Omega} + \|p\|_{2,\Omega} + \|\phi\|_{1,\Omega}\big\}. \tag{3.21}$$

Proof. It follows from the Céa estimate (3.6), and the approximation properties $(\widehat{\mathbf{AP}}_h^{\boldsymbol{u}})$, $(\widehat{\mathbf{AP}}_h^{p})$ and $(\widehat{\mathbf{AP}}_h^{\phi})$.

□

Stabilized Lagrange + Lagrange. It is often desirable to provide approximations where the pair $(\mathbf{H}_h, \mathrm{Z}_h)$ would not necessarily fulfil the discrete inf-sup condition (3.4), but it would achieve a more general concept of stability (weak coercivity, see (3.22) below). The stabilization consists in adding terms to the discrete problem to enforce such a condition (see [11]). The most appealing particular advantage is that equal-order discretizations for $\boldsymbol{u}$ and ϕ are allowed. Therefore, for an integer $k \geq 1$ we will consider the spaces

$$\begin{aligned}
\mathbf{H}_h &:= \Big\{\boldsymbol{v}_h \in [C(\overline{\Omega})]^d \,:\, \boldsymbol{v}_h\big|_K \in [\mathbb{P}_k(K)]^d \;\forall\, K \in \mathcal{T}_h,\; \boldsymbol{v}_h = 0 \text{ on } \Gamma_{\boldsymbol{u}}\Big\}, \\
\mathrm{Z}_h &:= \big\{\psi_h \in C(\overline{\Omega}) \,:\, \psi_h\big|_K \in \mathbb{P}_k(K) \;\forall\, K \in \mathcal{T}_h\big\}.
\end{aligned}$$

Lemma 3.5 (See [1]) *Assume $\mathcal{H} : \mathbf{W} \to \mathbb{R}$ is a continuous an linear functional, $\mathbf{W}_h$ is a closed subspace of $\mathbf{W}$, and the bilinear form $\mathcal{M}(\cdot,\cdot)$ is either coercive or it satisfies the discrete weak coercivity conditions:*

$$\sup_{\boldsymbol{s}_h \in \mathbf{W}_h \setminus \mathbf{0}} \frac{\mathcal{M}(\boldsymbol{w}_h, \boldsymbol{s}_h)}{\|\boldsymbol{s}_h\|_{\mathbf{W}}} \geq C_1^{\mathbf{W}} \|\boldsymbol{w}_h\|_{\mathbf{W}} \quad \text{and} \quad \sup_{\boldsymbol{w}_h \in \mathbf{W}_h} \mathcal{M}(\boldsymbol{w}_h, \boldsymbol{s}_h) > 0, \tag{3.22}$$

$\forall\, \boldsymbol{w}_h \in \mathbf{W}_h$ and $\forall\, \boldsymbol{s}_h \in \mathbf{W}_h \setminus \mathbf{0}$, respectively. Then the problem: find $\boldsymbol{w}_h \in \mathbf{W}_h$ such that

$$\mathcal{M}(\boldsymbol{w}_h, \boldsymbol{s}_h) = \mathcal{H}(\boldsymbol{s}_h) \qquad \forall \boldsymbol{s}_h \in \mathbf{W}_h$$

has a unique solution satisfying $\|\boldsymbol{w}_h\|_{\mathbf{W}} \leq C_2^{\mathbf{W}} \|\mathcal{H}\|_{\mathbf{W}'}$. Moreover

$$\|\boldsymbol{w} - \boldsymbol{w}_h\|_{\mathbf{W}} \leq \left(1 + \frac{C_1^{\mathbf{W}}}{C_2^{\mathbf{W}}}\right) \inf_{\boldsymbol{s}_h \in \mathbf{W}_h} \|\boldsymbol{w} - \boldsymbol{s}_h\|_{\mathbf{W}}.$$

In general, embedding the additional terms into expressions that vanish when the solution is sufficiently regular (for instance, residual contributions), leads to strong consistency of the stabilized scheme. A rich variety of stabilized methods targeted for Stokes equations is available from the literature (including e.g. pressure-projection stabilizations, variational multiscale methods, etc) but here we focus only on one family of methods known as Reflected Galerkin Least Squares (RGLS) schemes (see e.g. the review paper [2]). They consist in approximating the problem for displacement and volumetric stress (2.14) by the augmented discrete problem

$$\mathcal{M}^{-}_{\mathrm{RGLS}}\big((\boldsymbol{u}_h, \phi_h), (\boldsymbol{v}_h, \psi_h)\big) = \mathcal{H}^{-}_{\mathrm{RGLS}}(\boldsymbol{v}_h, \psi_h), \quad \forall (\boldsymbol{v}_h, \psi_h) \in \mathbf{H}_h \times \mathrm{Z}_h, \tag{3.23}$$

where

$$\mathcal{M}^{\pm}_{\mathrm{RGLS}}\big((\boldsymbol{u}_h,\phi_h),(\boldsymbol{v}_h,\psi_h)\big) := \mathcal{M}^{\pm}\big((\boldsymbol{u}_h,\phi_h),(\boldsymbol{v}_h,\psi_h)\big)$$
$$+\tau\sum_{K\in\mathcal{T}_h} h_K^2\big(-2\mu\,\mathbf{div}[\boldsymbol{\varepsilon}(\boldsymbol{u}_h)]+\nabla\phi_h,-2\mu\,\mathbf{div}[\boldsymbol{\varepsilon}(\boldsymbol{v}_h)]\mp\nabla\psi_h\big)_{0,K}$$

and

$$\mathcal{H}^{\pm}_{\mathrm{RGLS}}(\boldsymbol{v}_h,\psi_h) := \mathcal{H}(\boldsymbol{v}_h,\psi_h)+\tau\sum_{K\in\mathcal{T}_h} h_K^2\big(\boldsymbol{f},-2\mu\,\mathbf{div}[\boldsymbol{\varepsilon}(\boldsymbol{v}_h)]\mp\nabla\psi_h\big)_{0,K},$$

for a given stabilization constant $\tau > 0$. It can be proved that the nonsymmetric form $\mathcal{M}^{-}_{\mathrm{RGLS}}(\cdot,\cdot)$ is strongly coercive for any positive τ, and therefore problem (3.23) is uniquely solvable and unconditionally stable in the sense of Lemma 3.5 using $\mathbf{W}=\mathbf{H}\times\mathrm{Z}$ (see also [2, Sect. 3.1]). If we take $\mathcal{M}^{+}$ in definition of scheme (3.23), then we end up with the classical Douglas-Wang scheme, and for $k=1$, the problem (3.23) boils down to

$$\begin{aligned} a_1(\boldsymbol{u}_h,\boldsymbol{v}_h)+b_1(\phi_h,\boldsymbol{v}_h)-b_1(\psi_h,\boldsymbol{u}_h)+\tau\sum_{K\in\mathcal{T}_h} h_K^2(\nabla\phi_h,\nabla\psi_h)_{0,K} \\ = F(\boldsymbol{v}_h)+\tau\sum_{K\in\mathcal{T}_h} h_K^2(\nabla\phi_h,\nabla\psi_h)_{0,K}. \end{aligned} \tag{3.24}$$

If one drops the last term in the RHS of (3.24), then we recover the reflected version of the classical Brezzi-Pitkäranta method.

Convergence rates for stabilized methods depend on the stabilization parameters and on the order of the approximations k. For RGLS discretizations, the choice of τ does not affect the expected convergence rates: $O(h^{k+1})$ for displacements in the L^2-norm and $O(h^k)$ in the H^1-norm, whereas a decay of $O(h^k)$ is expected for the volumetric stress error in the L^2-norm (see [11]). Looking now at the pressure approximation, we choose Q_h as in the previous two FE families, and therefore the following convergence result holds

Theorem 3.6 *Let $(\boldsymbol{u},p,\phi)\in\mathbf{H}\times\mathrm{Q}\times\mathrm{Z}$ and $(\boldsymbol{u}_h,p_h,\phi_h)\in\mathbf{H}_h\times\mathrm{Q}_h\times\mathrm{Z}_h$ be the unique solutions of (2.8)–(2.10) and (3.1)–(3.3), respectively. Assume that $\boldsymbol{u}\in\mathbf{H}^{k+1}(\Omega)$, $p\in\mathrm{H}^{k+1}(\Omega)$ and $\phi\in\mathrm{H}^k(\Omega)$. Then, there exists $C>0$, independent of h and λ, such that*

$$\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}+\|p-p_h\|_{1,\Omega}+\|\phi-\phi_h\|_{0,\Omega}\le Ch^k\big\{\|\boldsymbol{u}\|_{k+1,\Omega}+\|p\|_{k+1,\Omega}+\|\phi\|_{k,\Omega}\big\}.$$

Finally, we stress that regarding poroelasticity formulations, only a few stabilization strategies have been applied to Biot consolidation problem, including a Galerkin least squares method [25], and a pressure-projection scheme [3].

Remark 3.2 *We end this section by pointing out that the continuous and discrete inf-sup conditions (resp. (2.19) and (3.4)), are strictly necessary to obtain all the required estimates independent of the parameter λ. In other words, without requiring these inf-sup conditions, it is still possible to prove well-posedness of the continuous and discrete problems and the corresponding Céa estimate. In fact, after simple computations, and without using the inf-sup conditions, one can readily obtain that the operator $\mathcal{A}$ (cf. (2.26)) is invertible and $\mathcal{A}+\mathcal{K}$ is injective. However, by doing so one unfortunately obtain the continuous dependence result and the Céa estimate with constants depending on λ which leads to unstable methods when using, for example, a $[\mathbb{P}_1]^d\times\mathbb{P}_1\times\mathbb{P}_0$ approximation, and λ is large (see Example 1 in Section 4 below).*

4 Numerical tests

We now provide a set of numerical examples putting into evidence some of the features analyzed above. Namely, optimal convergence in the sense of Theorems 3.3,3.4 and 3.6, and the locking-free property.

Example 1: convergence rates for a manufactured solution in 2D. Let us consider a cantilever bracket with curved boundary where we propose the following smooth exact solutions to (2.3),(2.5),(2.6)

$$\boldsymbol{u} = a\begin{pmatrix} \sin^2(\pi x_1)\sin(\pi x_2)\cos(\pi x_2) \\ \sin^2(\pi x_2)\sin(\pi x_1)\cos(\pi x_1) \end{pmatrix}, \quad p = b\sin(\pi x_1)\sin(\pi x_2), \quad \phi = p - \lambda \operatorname{div}\boldsymbol{u}, \tag{4.1}$$

and where the body force $\boldsymbol{f}$ and fluid source s can be simply determined from (4.1). We choose the following set of model parameters: displacement and pressure scalings a =1e-4, $b = \pi$; Young modulus E = 1e4, material permeability κ = 1e-7, Biot-Willis coefficient $\alpha = 0.1$, constrained specific storage c_0 = 1e-5, and the Lamé constants are $\lambda = E\nu(1+\nu)^{-1}(1-2\nu)^{-1}$, $\mu = E/(2+2\nu)$. Here, and in all subsequent tests, we consider zero gravitational forces.

The domain Ω is delimited by four curved boundaries parametrized as

$$\begin{aligned}
\Gamma_1 &= \{\omega \in [0,1]:\ x_1 = \omega + \gamma\cos(\pi\omega)\sin(\pi\omega),\ x_2 = -\gamma\cos(\pi\omega)\sin(\pi\omega)\},\\
\Gamma_2 &= \{\omega \in [0,1]:\ x_1 = 1 + \gamma\cos(\pi\omega)\sin(\pi\omega),\ x_2 = \omega - \gamma\cos(\pi\omega)\sin(\pi\omega)\},\\
\Gamma_3 &= \{\omega \in [1,0]:\ x_1 = \omega + \gamma\cos(\pi\omega)\sin(\pi\omega),\ x_2 = 1 - \gamma\cos(\pi\omega)\sin(\pi\omega)\},\\
\Gamma_4 &= \{\omega \in [1,0]:\ x_1 = \gamma\cos(\pi\omega)\sin(\pi\omega),\ x_2 = \omega - \gamma\cos(\pi\omega)\sin(\pi\omega)\},
\end{aligned}$$

where we take $\gamma = -0.08$. Boundary conditions are assigned as follows: nonhomogeneous Dirichlet displacements and pressure normal fluxes j are set according to (4.1) on $\Gamma_{\boldsymbol{u}} = \Gamma_3 \cup \Gamma_4$; nonhomogeneous Dirichlet pressure and Cauchy normal fluxes $\boldsymbol{h}$ are set according to (4.1) on $\Gamma_p = \Gamma_1 \cup \Gamma_2$.

The accuracy of the numerical approximation using the FE families listed in Section 3.3 (Hood-Taylor, MINI-element, and stabilized scheme (3.24)) can be assessed by partitioning Ω into unstructured triangulations generated putting 2^{n+1} ($n = 0,1,\ldots,8$) vertices on each curve of the domain boundary. Relative errors between exact and approximate solutions are to be computed on each refinement level according to

$$e(w) = \frac{\|w - w_h\|_{i,\Omega}}{\|w\|_{i,\Omega}}, \qquad i = 0,1,$$

where w denotes a generic non-zero scalar or vector field. Two sets of simulations were performed in order to study the influence of the Poisson ratio. The first case corresponds to a mild incompressibility $\nu = 0.4$ and $\lambda = 14285.7$, whereas the second case focuses on a quasi-incompressible regime with $\nu = 0.49999$ and λ = 1.66e8. In the first case, we observe optimal convergence rates for all methods, even for the lowest order discretization (see left panels in Figure 1). While the orders of convergence of pressure and volumetric stress approximations seem invariant to the drastic increase of ν, irrespective of the FE family employed, the displacements are severely affected in the incompressibility limit when using the lowest order approximation. This is clearly observed in the top-right plot of Figure 1, where the relative displacement error does not attain even a $O(h)$ convergence with the lowest order method. In Figure 2 we illustrate the converged numerical solution obtained with the stabilized method (3.24), with stabilization constant $\tau = 1/60$. These snapshots correspond to the case $\nu = 0.49999$ and $\lambda = 1.66\text{e}8$.

Example 2: footing problem. We now focus on the behaviour of the proposed methods when applied to the solution of the 2D *footing test*. The goal is to observe pressure, volumetric stress and displacements incurred after that a rectangular block of porous soil undergoes a load of intensity σ_0 along a strip on top of it. The model parameters are $\Omega = (-50, 50) \times (0, 75)$, E =3e4 N/m^2, κ =1e-4 m^2/Pa, σ_0 =1.5e4 N/m^2 (see a similar test in [15] for moderate Poisson ratios). In addition we put c_0 =1e-3, $\alpha = 0.1$, and here we force the incompressibility limit by setting $\nu = 0.4995$. Boundary conditions are set as follows (see a sketch in the bottom right panel of Figure 3): $\boldsymbol{u} = \boldsymbol{0}$ on Γ_3 (left, right, and bottom sides of the block); $\boldsymbol{\sigma n} = \boldsymbol{h}$ on $\Gamma_1 \cup \Gamma_2$, where $\boldsymbol{h} = (0, -\sigma_0)^T$ on Γ_1 and $\boldsymbol{h} = \boldsymbol{0}$ otherwise; and $p = 0$ on $\partial\Omega$. The domain is partitioned into 71272 unstructured triangles using 35637 vertices.

The value of the Poisson ratio suggests that inf-sup unstable discretizations of displacement and volumetric stress will produce spurious pressure modes. This phenomenon is evidenced in Figure 3, where we depict the numerical solution obtained with the lowest order discretization. Both the volumetric stress and the pressure profiles are populated with oscillations, even with a quite fine mesh. On the other hand, at least in this particular case, the computed displacements do not appear to suffer from locking. Next we perform again the same test, this time using the MINI-element for the discretization of displacement and volumetric stress, whereas the pressure field is approximated with piecewise linear continuous elements. In contrast with the results collected in Figure 3, now in Figure 4 the pressure and volumetric stress fields are stable and completely free from spurious oscillations.

Example 3: swelling of a sponge. Next, the implementation of the proposed schemes in 3D is tested by looking at the displacements incurred by swelling a porous block occupying the domain $\Omega = (0, 1) \times (0, 1) \times (0, \frac{1}{2})$. The driving effect is simply a pressure difference between the sides $x_1 = 0$ and $x_1 = 1$, going from $p = 1\text{e}4$ at $x_1 = 0$ to zero pressure on $x_1 = 1$. Zero-flux conditions are imposed for pressure on the remainder of the boundary. The normal components of the displacements are set to zero on the sides $x_1 = 0$, $x_2 = 0$ and $x_3 = 0$, whereas zero normal stress is considered elsewhere on $\partial\Omega$. Other model and discretization parameters are listed in what follows: $E = 8000$, $\nu = 0.3$, $c_0 = 0.001$, $\kappa = 1\text{e-}5$, $\rho = \alpha = 1$, $\tau = 1/60$. No external or internal forces are considered, and neither fluid sources or sinks. The domain is partitioned into a structured tetrahedral mesh of 62586 elements and 10976 vertices, and a stabilized method using (3.24) is employed for the numerical approximation of displacements, volumetric stress and pressure. The obtained results are depicted in Figure 5, where no pressure oscillations nor unphysically small displacements are observed. We also simulate the swelling of an heterogeneous porous medium, where we consider that the permeability is zero in the strip $0.45 \leq x_1 \leq 0.55$ and otherwise we take $\kappa = 1$ (that is, five orders of magnitude larger than in the previous test). The results are collected in the last row of Figure 5, where a much more pronounced swelling is observed far from the slip-displacement boundaries, whereas on the non-porous region, the material is swelling only due to the elastic compliance behaviour.

Example 4: one-dimensional consolidation benchmark. In our last test we focus on the consolidation behaviour of a thin porous column of height H and cross area W. The top and bottom surfaces of the column are endowed with pervious (zero pore pressure $p = 0$, constant mechanical load in the vertical direction $\boldsymbol{\sigma n} = -\sigma_0 \boldsymbol{e}_3$, and free to drain) and impervious (zero pressure flux $\kappa \nabla p \cdot \boldsymbol{n} = 0$ and zero displacement $\boldsymbol{u} = \boldsymbol{0}$) filtration conditions, respectively. On the lateral walls we enforce zero horizontal displacements (in both x_1 and x_2 directions), therefore Γ_p is the top side of the column, whereas $\Gamma_{\boldsymbol{u}} = \partial\Omega \setminus \Gamma_p$. Moreover, we now consider the general time-dependent system (2.1)-(2.3), and our goal is to compare the obtained numerical approximations against the following exact solutions to the adimensional pseudo-1D version of this problem (see e.g. [24, 20])

$$u^* = 1 - x^* - \sum_{k=0}^{\infty} \frac{2}{M^2} \cos(Mx^*) \exp(-M^2 t^*), \quad p^* = \sum_{k=0}^{\infty} \frac{2}{M} \sin(Mx^*) \exp(-M^2 t^*),$$

where the superscript * denotes adimensional quantities and variables as follows $x^* = x_3/H$, $t^* = (\lambda + 2\mu)\kappa t H^2$, $M = \frac{1}{2}\pi(2k+1)$, $u^* = u_3(\lambda+2\mu)/\sigma_0 H$, $p^* = p/\sigma_0$.

As it stands, our analysis clearly does not cover the original time-dependent system, and our goal is only to illustrate the performance of the proposed schemes applied to (2.1)-(2.3). A semidiscretization of this problem using a backward Euler method with a fixed time-step Δt yields

$$\begin{aligned}
&a_1(\boldsymbol{u}_h^{n+1}, \boldsymbol{v}_h) + b_1(\boldsymbol{v}_h, \phi_h^{n+1}) = F^{n+1}(\boldsymbol{v}_h) && \forall \boldsymbol{v}_h \in \mathbf{H}_h,\\
&\tilde{a}_2(p_h^{n+1}, q_h) - b_2(q_h, \phi_h^{n+1}) = \Delta t\, G^{n+1}(q_h) + \left(\frac{c_0}{\alpha} + \frac{1}{\lambda}\right) \int_\Omega p_h^n q_h - b_2(q_h, \phi_h^n) && \forall q_h \in \mathrm{Q}_h,\\
&b_1(\boldsymbol{u}_h^{n+1}, \psi_h) + b_2(p_h^{n+1}, \psi_h) - c(\phi_h^{n+1}, \psi_h) = 0 && \forall \psi_h \in \mathrm{Z}_h,
\end{aligned}$$

with $\tilde{a}_2(p,q) := \left(\frac{c_0}{\alpha} + \frac{1}{\lambda}\right) \int_\Omega pq + \frac{\Delta t}{\alpha\eta} \int_\Omega \kappa \nabla p \cdot \nabla q$, which implies that at each time-step we need to solve a system of the form (3.1)-(3.3). Notice that the coefficients in the LHS of the system are constant and so only the RHS needs to be re-assembled at each time iteration. We choose the MINI-element + Lagrange approximation of displacement, volumetric stress and pressure, and the thin column with $H = 1$[m], $W = 0.1$[m^2] is discretized into a structured tetrahedral mesh containing 3312 elements. Model and numerical parameters assume the values $\sigma_0 = $ 1e4 [Pa], $E = $ 3e4 [N/m^2], $\nu = 0.2$, $\kappa = $ 1e-10 [m^2], $\eta = $ 1e-3 [Pa s], $c_0 = 0$, $\alpha = 1$, $\rho = 1$, $T = 10$ [s], $\Delta t = 0.1$ [s], and the initial data for displacement and pressure are set according to the idealized 1D solutions with the Fourier series truncated at $k = 350$. Figure 6 presents snapshots of the numerical solutions at early and advanced times, along with profiles of the computed approximations and exact adimensional solutions at the centerline (x_3-axis) of the column, showing good accuracy throughout the time horizon.

References

[1] I. Babuška and A. Aziz, *Survey lectures on the mathematical foundations of the finite element method*. In: The mathematical foundations of the finite element method with applications to PDEs, Academic press, NY (1972) pp. 1–359.

[2] T. Barth, P. Bochev, M. Gunzburger, and J. Shadid, *A taxonomy of consistently stabilized finite element methods for the Stokes problem*. SIAM J. Sci. Comput., **25** (2004) 1585–1607.

[3] L. Berger, R. Bordas, D. Kay, and S. Tavener, *Stabilized lowest-order finite element approximation for linear three-field poroelasticity.* SIAM J. Sci. Comput., **37** (2015) A2222–A2245.

[4] M.A. Biot, *Theory of elasticity and consolidation for a porous anisotropic solid.* J. Appl. Phys., **26** (1955) 182–185.

[5] D. Boffi, *Stability of higher order triangular Hood-Taylor methods for stationary Stokes equations.* Math. Models Methods Appl. Sci., **2** (1994) 223–235.

[6] D. Boffi, *Three-dimensional finite element methods for the Stokes problem.* SIAM J. Numer. Anal., **34** (1997) 664–670.

[7] Y. Chen, Y. Luo, and M. Feng, *Analysis of a discontinuous Galerkin method for the Biot's consolidation problem.* Appl. Math. Comput., **219** (2013) 9043–9056.

[8] C. Domínguez, G.N. Gatica, S. Meddahi, R. Oyarzúa, *A priori error analysis of a fully-mixed finite element method for a two-dimensional fluid-solid interaction problem.* ESAIM: Math. Model. Numer. Anal., **47**(2) (2013) 471–506.

[9] V. Domíngues and F.J. Sayas, *A BEM-FEM overlapping algorithm for the Stokes equation.* Appl. Math. Comput. **182** (2006) 691–710.

[10] A. Ern and J.-L.Guermond, Theory and practice of finite elements, Applied Mathematical Sciences, vol. 159, Springer-Verlag, New York, (2004).

[11] L. Franca, T.J.R. Hughes, and R. Stenberg, *Stabilized finite element methods for the Stokes problem.* In: M. Gunzburger and R.A. Nicolaides, eds. Incompressible computational fluid dynamics, Cambridge University Press (1993), pp. 87–107.

[12] 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).

[13] G.N. Gatica, R. Oyarzúa, and F.J. Sayas, *Convergence of a family of Galerkin discretizations for the Stokes–Darcy coupled problem*, Numer. Methods Part. Diff. Eqns., **27** (2011) 721–748.

[14] G.N. Gatica, A. Márquez, and S. Meddahi, *Analysis of the coupling of primal and dual-mixed finite element methods for a two-dimensional fluid-solid interaction problem.* SIAM J. Numer. Anal. **45** (2007) 2072–2097.

[15] F.J. Gaspar, F.J. Lisbona, and C.W. Oosterlee, *A stabilized difference scheme for deformable porous media and its numerical resolution by multigrid methods.* Comput. Vis. Sci. **11** (2008) 67–76.

[16] G. N. Gatica, R. Oyarzúa, and F.J. Sayas, *Analysis of fully-mixed finite element methods for the Stokes-Darcy coupled problem.* Math. Comp., **80** (2011) 1911–1948.

[17] V. Girault and P.-A. Raviart, Finite Element Approximation of the Navier–Stokes Equations. Lecture Notes in Mathematics, 749. Springer-Verlag, Berlin-New York, (1979).

[18] J. Korsawe and G. Starke, *A least-squares mixed finite element method for Biot's consolidation problem in porous media.* SIAM J. Numer. Anal., **43** (2005) 318–339.

[19] R. Liu, M.F. Wheeler, C.N. Dawson, and R.H. Dean, *On a coupled discontinuous/continuous Galerkin framework and an adaptive penalty scheme for poroelasticity problems.* Comput. Methods Appl. Mech. Engrg., **198** (2009) 3499–3510.

[20] M.A. Murad and A.F.D. Loula, *On stability and convergence of finite element approximations of Biot's consolidation problem.* Int. J. Numer. Methods Engrg., **37** (1994) 645–667.

[21] M.A. Murad, V. Thomée, and A.F.D. Loula, *Asymptotic behavior of semi discrete finite-element approximations of Biot's consolidation problem.* SIAM J. Numer. Anal., **33** (1996) 1065–1083.

[22] P.J. Phillips and M.F. Wheeler, *A coupling of mixed and continuous Galerkin finite element methods for poroelasticity I: the continuous in time case.* Comput. Geosci., **11** (2007) 131–144.

[23] R.E. Showalter, *Diffusion in poro-elastic media.* J. Math. Anal. Appl., **251** (2000) 310–340.

[24] K. Terzaghi, Theoretical Soil Mechanics, Wiley, New York (1943).

[25] A. Truty, *A Galerkin/least-squares finite element formulation for consolidation.* Int. J. Numer. Methods Engrg. **52** (2001) 763–786.

[26] M.F. Wheeler, G. Xue, and I. Yotov, *Coupling multipoint flux mixed finite element methods with continuous Galerkin methods for poroelasticity.* Comput. Geosci., **18** (2014) 57–75.

[27] J.A. White and R.I. Borja, *Stabilized low-order finite elements for coupled solid-deformation/fluid-diffusion and their application to fault zone transients.* Comput. Methods Appl. Mech. Engrg., **197** (2008) 4353–4366.

[28] S.-Y. Yi, *Convergence analysis of a new mixed finite element method for Biot's consolidation model.* Numer. Methods Part. Diff. Eqns., **30** (2014) 1189–1210.

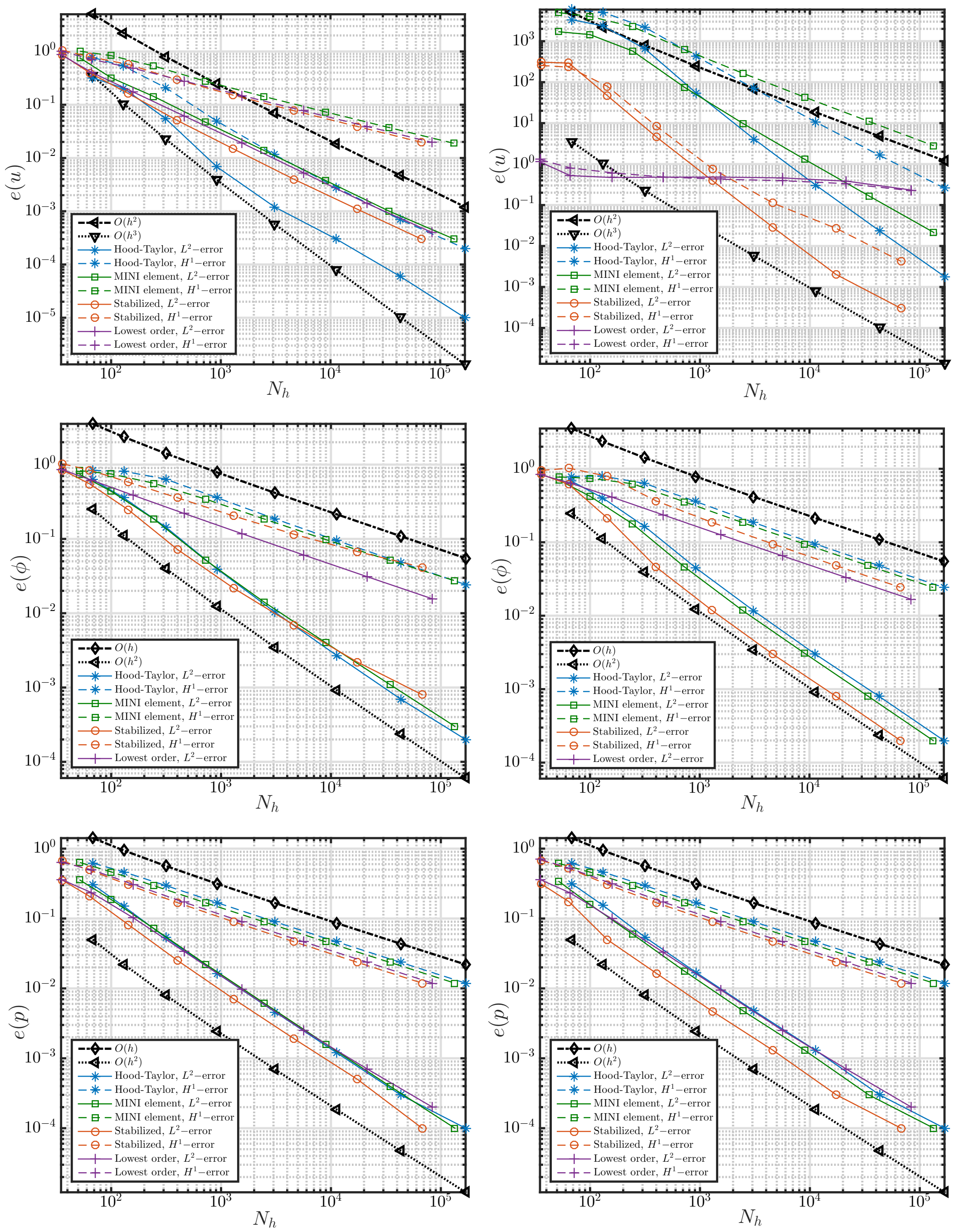


Figure 1: Example 1: error history associated to four different discretizations of the three-field poroelasticity equations. Panels on the left report errors incurred when the Poisson ratio is $\nu = 0.4$, yielding a Lamé constant of $\lambda = 14285.7$. The second round of tests, with $\nu = 0.49999$ and $\lambda = 1.66\text{e}8$ is shown in the right plots.

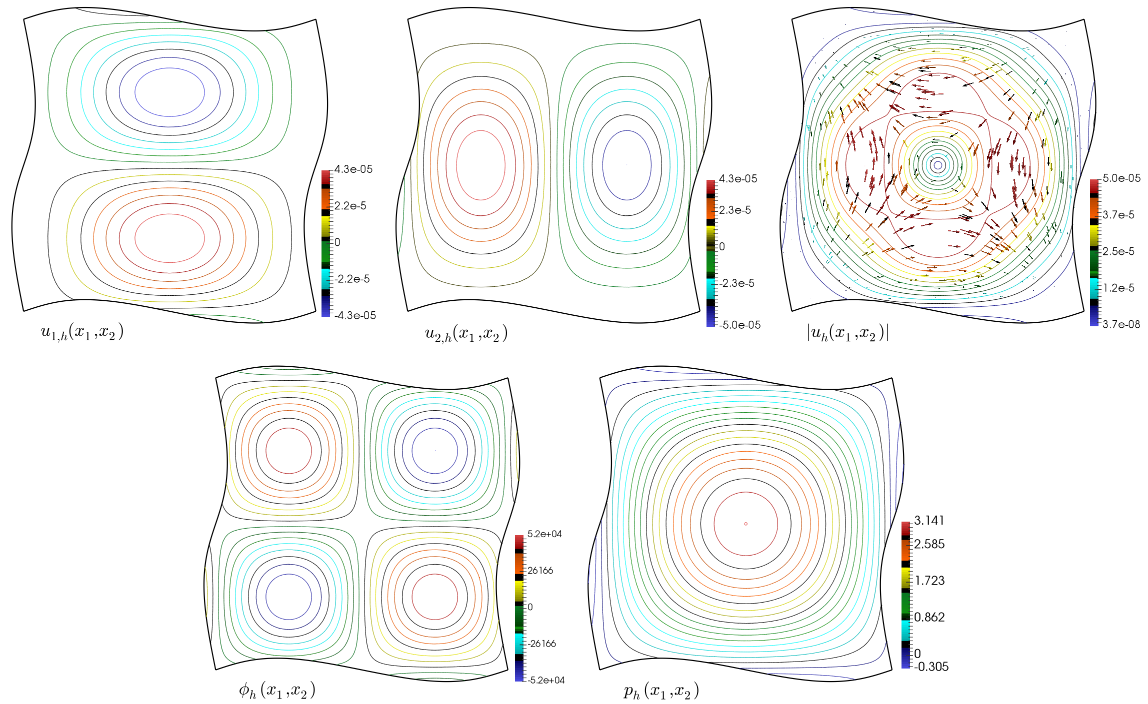


Figure 2: Example 1: three-field poroelasticity equations discretized with a stabilized method. Contour plots of the approximate displacement components, displacement magnitude and vectors, volumetric stress, and pressure profiles in the case where $\eta = 0.49999$ and $\lambda = 1.66\text{e}8$.

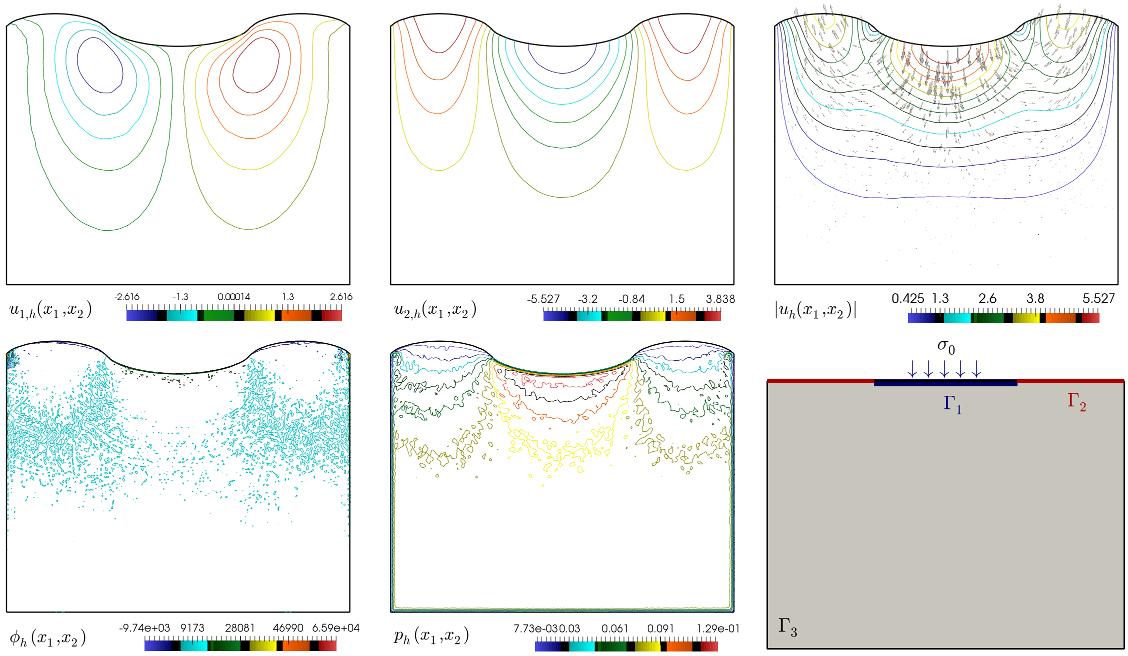


Figure 3: Example 2: footing of a porous block using the lowest order discretization. From top left to bottom right: approximation of displacement components and magnitude, volumetric stress, and pressure; and a sketch of the undeformed domain and boundary splitting.

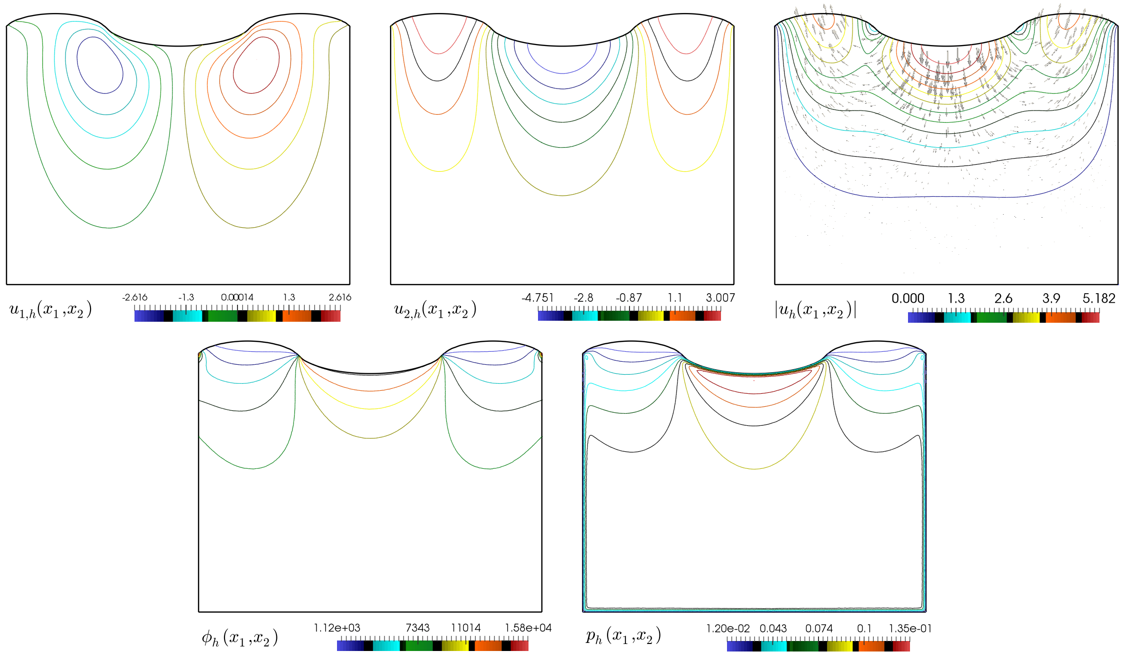


Figure 4: Example 2: footing of a porous block using the MINI-element method. From top left to bottom right: approximation of displacement components and magnitude, volumetric stress, and pressure distribution.

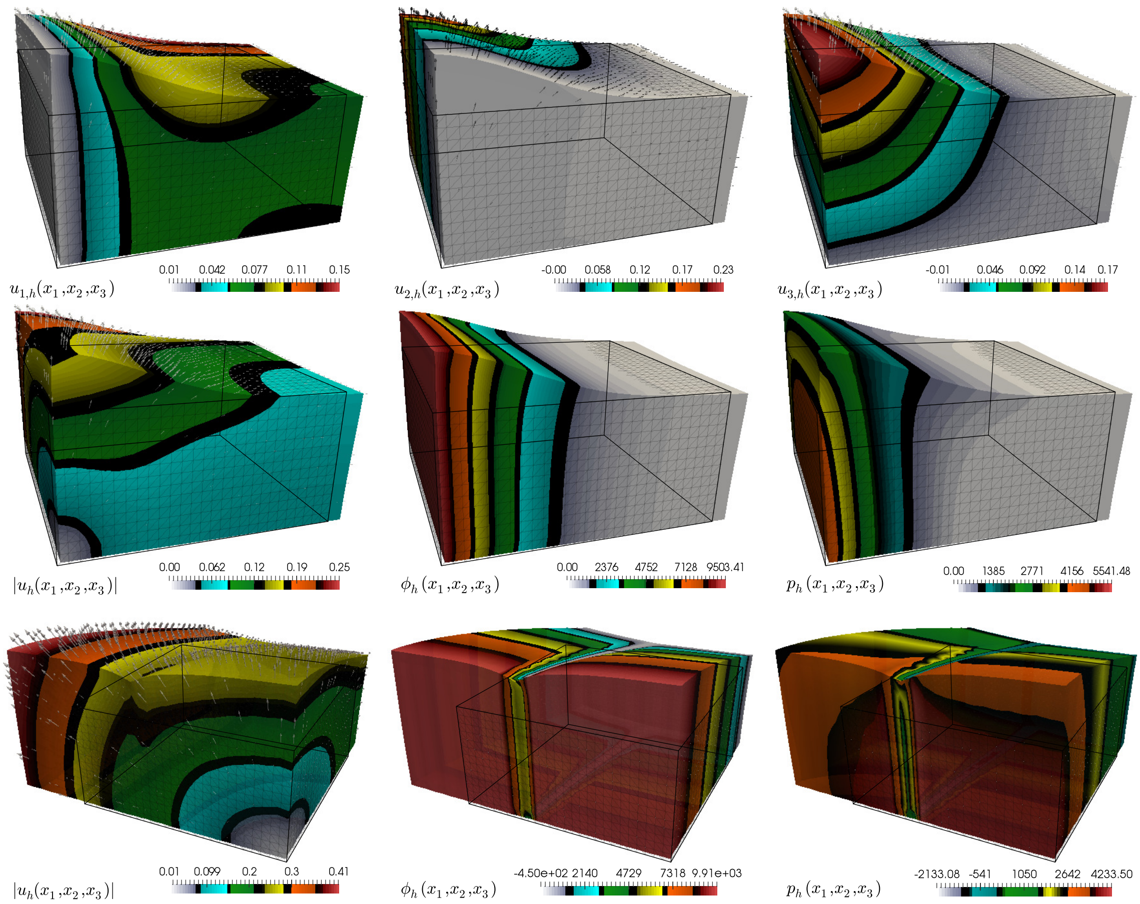


Figure 5: Example 3: swelling of a sponge using a stabilized method. Displacement components and magnitude, volumetric stress, and pressure distribution (top and middle row). The last row shows approximate solutions when a strip of zero permeability is present in the domain. All fields are represented on the deformed configuration, and the skeleton tetrahedral undeformed mesh is also depicted.

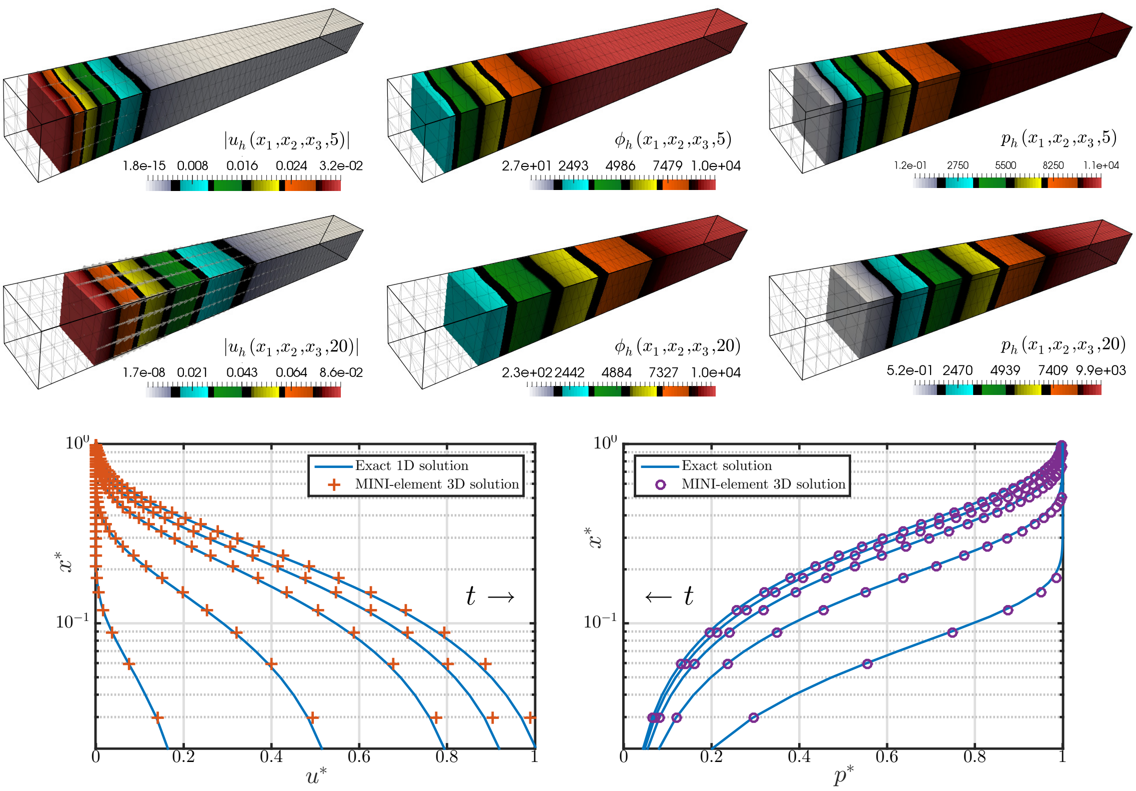


Figure 6: Example 4: consolidation benchmark using the MINI-element + Lagrange approximation together with a backward Euler time stepping. The first two rows show snapshots of the numerical solutions at $t = 5$ [s] (top) and $t = 20$ [s]. The bottom row displays mid-line profiles of the computed vs. exact nondimensional vertical displacement and pressure at five time instants $t^* = 0.2, 0.4, \ldots, 1$.

Centro de Investigación en Ingeniería Matemática (CI²MA)

PRE-PUBLICACIONES 2015

2015-30 Eligio Colmenares, Gabriel N. Gatica, Ricardo Oyarzúa: *An augmented fully-mixed finite element method for the stationary Boussinesq problem*

2015-31 Jessika Camaño, Cristian Muñoz, Ricardo Oyarzúa: *Analysis of a mixed finite element method for the Poisson problem with data in* $L^p, 2n/(n+2) < p < 2, n = 2, 3$

2015-32 Gabriel N. Gatica, Filander A. Sequeira: *A priori and a posteriori error analyses of an augmented HDG method for a class of quasi-Newtonian Stokes flows*

2015-33 Mario Álvarez, Gabriel N. Gatica, Ricardo Ruiz-Baier: *A posteriori error analysis for a viscous flow-transport problem*

2015-34 Andrea Barth, Raimund Bürger, Ilja Kröker, Christian Rohde: *A hybrid stochastic Galerkin method for uncertainty quantification applied to a conservation law modelling a clarifier-thickener unit with several ramdom sources*

2015-35 Raimund Bürger, Julio Careaga, Stefan Diehl, Camilo Mejías, Ingmar Nopens, Peter Vanrolleghem: *A reduced model and simulations of reactive settling of activated sludge*

2015-36 Celso R. B. Cabral, Luis M. Castro, Víctor H. Lachos, Larissa A. Matos: *Multivariate measurement error models based on student-t distribution under censored responses*

2015-37 Luis M. Castro, Víctor H. Lachos, Larissa A. Matos: *Censored mixed-effects models for irregularly observed repeated measures with applications to HIV viral loads*

2015-38 Carlos Garcia, Gabriel N. Gatica, Salim Meddahi: *A new mixed finite element method for elastodynamics with weak symmetry*

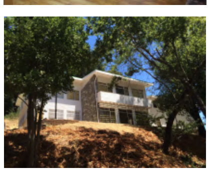
2015-39 Sergio Caucao, Gabriel N. Gatica, Ricardo Oyarzúa, Ivana Sebestova: *A fully-mixed finite element method for the Navier-Stokes/Darcy coupled problem with nonlinear viscosity*

2015-40 Guillaume Chiavassa, María Carmen Martí, Pep Mulet: *Hybrid WENO schemes for polydisperse sedimentation models*

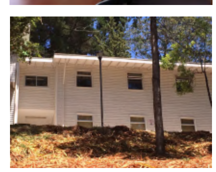
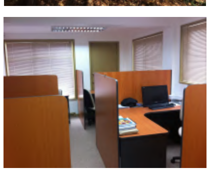
2015-41 Ricardo Oyarzúa, Ricardo Ruiz-Baier: *Locking-free finite element methods for poroelasticity*

Para obtener copias de las Pre-Publicaciones, escribir o llamar a: Director, Centro de Investigación en Ingeniería Matemática, Universidad de Concepción, Casilla 160-C, Concepción, Chile, Tel.: 41-2661324, o bien, visitar la página web del centro: http://www.ci2ma.udec.cl



Centro de Investigación en
Ingeniería Matemática (CI²MA)
Universidad de Concepción

Casilla 160-C, Concepción, Chile
Tel.: 56-41-2661324/2661554/2661316
http://www.ci2ma.udec.cl