Finite volume element methods for three-field saddle-point problems

Zoa de Wijn

St Catherine's College

University of Oxford

A thesis submitted for the degree of

M.Sc. in Mathematical Modelling and Scientific Computing

Trinity Term 2017

I would like to express my sincerest gratitude towards my supervisor Dr. Ricardo Ruiz Baier for his continuous support, invaluable guidance and great enthusiasm for the subject. It was a pleasure and a humbling experience to work with him and I deeply thank him for that. I would also like to thank Dr. Kathryn Gillow for the excellent organisation throughout this year.

Abstract

We propose a family of mixed finite (volume) element methods for the approximation of a novel three-field formulation for linear elastostatics (formulated in terms of the displacement, the rotation vector, and the solid pressure) and a mixed hybrid finite volume element method for the approximation of a new three-field formulation for poroelastostatics (expressed in terms of the solid displacement, the pore pressure of the fluid, and the total pressure). The proposed mixed finite volume element scheme for linear elasticity is of lowest-order and the proposed Petrov-Galerkin method for the approximation of linear poroelastostatics is related to the classical Taylor-Hood finite element pair in the sense that the trial spaces correspond to conforming piecewise quadratic elements for the approximation of the displacements and the fluid pressure, whereas the total pressure is approximated with piecewise linear polynomials. We discuss the unique solvability of both continuous three-field formulations as well as the invertibility and stability of the proposed Galerkin and Petrov-Galerkin formulations. Optimal a priori error estimates are derived using norms that are robust with respect to the Lamé constants, turning these numerical methods particularly appealing for nearly incompressible materials. In particular, for the finite volume element scheme for poroelasticity, we establish the coercivity of the bilinear form associated with the solid viscous stress using spectral properties of symmetrised local stiffness matrices and employ a combination of Céa estimates with standard interpolation properties to prove the overall $\mathcal{O}(h^2)$ convergence of the proposed method. In addition, the predicted accuracy and applicability of the mixed finite (volume) element schemes for linear elasticity is verified numerically by conducting a set of insightful computational tests.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Over the past thirty years, finite volume element (FVE) methods have become one of the main numerical tools for solving partial differential equations (PDEs) for which a numerical method preserving local conservation laws is required. FVE schemes correspond to a specific type of Petrov-Galerkin formulations where the trial space is constructed using a primal partition of the domain, whereas the test space is associated with either a dual mesh or a dual basis. Depending on the particular kind of dual grid, the transfer operator that connects the trial and test spaces possesses different interpolation properties, which are under special circumstances used to recast primarily pure FVE formulations into pure Galerkin formulations. In general, FVE methods enjoy various features shared by finite element (FE) and finite volume (FV) schemes, including local flux conservation properties, liberty to choose different numerical fluxes, flexibility regarding choices for dual partitions associated with the primal partitions of the domain, and several others (see, for instance, [24]).

The early development of FVE schemes was largely influenced by the theory of FV and FE methods and traces back to several publications of the 1980s that study Petrov-Galerkin formulations featuring local conservation properties. In 1982, Li [52, 53,56] employed generalised characteristic functions on dual grids in combination with standard finite elements on the primal grids to obtain so-called generalised difference (GD) methods by using the Petrov-Galerkin method. In the same articles, optimal error estimates in the H^1-norm are derived, provided that the trial space is chosen as a linear finite element space and the test functions are chosen as characteristic functions defined on the dual grid. Afterwards, numerous researchers have contributed to the subject (see, for instance, the survey in [54] and the book [55]). In 1987, Bank et al. [9] reconstructed and analysed the so-called box method (BM) as a Galerkin procedure, where the trial spaces consist of continuous piecewise linear polynomials

on triangular meshes and the associated test spaces are spanned by piecewise constant characteristic functions on dual box meshes. For further studies that focus on the box method, we refer to Hackbusch [42] and Schmidt [67]. Furthermore, in the early 1990s, Cai and others [20,21,23] studied the FVE method, where one utilises a volume integral formulation of the problem in combination with a finite partitioning set of volumes to discretise the equations.

In addition to the brief survey given above, there exists an extensive amount of literature on the numerical analysis and implementation of FVE methods for solving PDEs for which a numerical method preserving local conservation laws is required (see, for instance, [26] and the review in [58]). Observe that all of these methods may be regarded as special types of the Petrov-Galerkin finite element methods where the test space is constructed using dual grids of the domain, such that the discrete equations indeed preserve conservation laws locally.

1.1 Three-field saddle-point formulations for linear (poro)elastostatics

Finite volume element discretisation schemes following the principles outlined above have been systematically employed in numerous fluid flow problems, including Stokes and Navier-Stokes (see, for instance, [43, 51, 59, 64, 75]), and also in coupled flow-transport systems arising from diverse applications (cf. [32, 34, 66]). Although there exists an extensive amount of literature for FVE methods, up to our knowledge, the only contributions addressing FVE-like discretisations for solid mechanics are the hybrid-stress FV method for linear elasticity on quads studied in [77]; and [49], where two alternative stabilisation approaches based on nodal pressure and dual bases and meshes are applied to construct inf-sup stable approximations for nearly incompressible linear elasticity. This observation serves as one of the main motivations for studying FVE methods in the context of elasticity-based problems. In particular, we focus on the FVE approximation of two novel three-field saddle-point formulations for linear elasticity and poroelasticity.

1.1.1 Three-field formulation for linear elasticity

The numerical solution of elasticity-based problems encompasses well-documented difficulties. For instance, for pure-displacement formulations, if one uses classical FE discretisations based on piecewise linear and continuous elements, then accuracy is

ensured only for moderate values of the Poisson ratio ν. As $\nu \to 0.5$ (that is, when the Lamé constant $\lambda \to \infty$ and the elastic material becomes nearly incompressible), the numerical scheme might generate spurious solutions (unphysically small deformations referred to as the locking phenomenon, see for instance [17]). A number of alternative formulations and associated numerical methods are available to overcome this issue. Notably, choosing a mixed scheme would produce accurate solutions even for nearly incompressible materials, and at the same time, one accommodates the direct approximation of auxiliary variables of interest such as pressure, stress, or rotations.

One of the most commonly used mixed approaches for linear elasticity is the Hu-Washizu formulation [44, 74]. Some popular methods based on the Hu-Washizu formulation include the enhanced assumed strain method [68], the assumed stress method [63], the mixed-enhanced strain method [47], the strain gap method [65], and the so-called B-bar method [45]. Some of these methods actually coincide under certain conditions (see the discussions in [4, 18, 30]). The well-posedness for this class of formulations has been established in [50], where it is also shown that a modified version of the Hu-Washizu formulation is more amenable for obtaining uniform convergence with respect to the model parameters. Alternatively, other mixed formulations (such as the Hellinger-Reissner principle) can be employed to obtain methods that are robust with respect to the Lamé constants.

In contrast to the brief literature survey given above, here we advocate to the formulation of the elasticity equations in terms of the displacement, the rotation vector, and the solid pressure. Schemes more closely related to the three-field formulation of interest deal with mixed formulations for elasticity involving pressure, stress, and rotation. We mention for instance mixed formulations based on stress [6, 7, 12, 16], the augmented scheme in [39], a family of pseudostress-based methods from [36, 38], and the displacement-pressure mixed formulations [17, 19]. In addition, we mention the locking-free methods for plate models [10] and the membrane elements introduced in [46], all including rotation tensors as an additional field. In our case, after regarding the pressure together with the rotation vector as a single auxiliary unknown (defined in an appropriate product functional space), we are able to analyse the solvability of the resulting mixed variational formulation using the classical Babuška-Brezzi theory for saddle-point problems. In particular, the well-posedness result and the continuous dependence on the data turn out to be independent of the Lamé constants.

For the approximation of the three-field formulation of interest, we first introduce a family of finite element methods given by piecewise continuous polynomials of degree $k \geq 1$ for the displacement, and piecewise polynomials of degree $k-1$ for the rotation

and the solid pressure. The well-posedness results and optimal convergence estimates (with constants fully independent of the Lamé coefficient λ, guaranteeing robustness in the nearly incompressible limit) for this novel FE method are reviewed, based on [3]. Based on the lowest-order version of this mixed FE method, we proceed to focus on the derivation and analysis of a FVE formulation for the linear elasticity equations written in terms of the displacement, the rotation vector, and the pressure. As in well-established FVE schemes for Stokes equations (cf. [51, 64]), it turns out that the two schemes differ only by the assembly of the forcing term, and therefore a straightforward derivation of stability properties and energy estimates in natural norms can be done by exploiting the results for the family of mixed finite elements.

1.1.2 Three-field formulation for poroelasticity

A further goal of this dissertation is to address the construction and analysis of second-order mixed hybrid FVE methods for the Biot consolidation equations in poromechanics [13]. These equations describe the interaction between a diffusive fluid flow and the solids deformation of a porous skeleton and have been widely applied in numerous practical scenarios of high importance, such as CO_2 sequestration, waste disposal, petroleum production, the perfusion of bones and soft tissue, etc.

Unlike other FVE-based schemes for poroelasticity available from the literature [8,60,69], here we recast the system in terms of the solid displacement, the fluid pore pressure, and the total pressure. Such a three-field formulation (recently proposed in [62]) exhibits the advantage that one obtains stability bounds that are independent from the Lamé modulus of dilation, λ, which approaches infinity in the incompressibility limit. This useful property is inherited by the numerical method as long as an appropriate inf-sup condition is satisfied by the bilinear form relating the discrete spaces where one seeks for the approximate displacements and total pressure. If this is the case, the obtained error estimates are also robust with respect to the Lamé constants and the produced solutions are locking-free.

It is known that the discretisation of various symmetric differential operators (such as the ones defining the Poisson or the Stokes equations) by means of linear FE and linear FVE methods will lead to the same bilinear forms and associated stiffness matrices [78]. However, this correspondence is lost if one utilises quadratic basis functions instead, as pointed out in the early paper by Liebau [57], who proposed second-order FVE methods for general diffusion problems, where the basis functions are explicitly constructed based on geometrical considerations. In particular, it is

known that meshes of high order FVE methods typically need to satisfy certain geometric requirements (cf. [25,26,57,70]). The coercivity of the resulting bilinear forms (which are no longer symmetric) was then analysed by the virtue of inverse inequalities and inspection of the spectral behaviour of symmetrised local block matrices. The same idea was later advanced by Chen [8], and by Chen and collaborators in a series of papers (see [25–28] and the references therein).

In the context of numerical schemes for poroelasticity, we stress that high order methods are currently rare. We mention the mixed high order and weighted interior penalty discretisation recently proposed in [15], which also features local conservativity. Unlike any other numerical scheme available from the literature, the second-order mixed hybrid FVE we focus on utilises a hybrid quadratic Lagrange finite volume element for the displacement, a quadratic Lagrange element for the pore pressure of the fluid, and a linear element for an auxiliary unknown that can be regarded as a pseudo total pressure. It turns out that for the second discrete bilinear form arising in the mixed hybrid FVE formulation of interest, a suitable coercivity condition can be obtained by closely following the approaches from [25–28] (where scalar elliptic boundary value problems are studied instead). In fact, one of the main tasks in the analysis of high order FVE methods is to verify the inf-sup conditions for the discrete bilinear operators of interest.

The main advantages of the proposed FVE scheme are its second-order convergence rate, its local conservativeness, and its robustness when the material properties approach the incompressibility limit, i.e. when the Lamé modulus of dilation $\lambda \to \infty$ and the elastic skeleton becomes nearly incompressible. A further appealing property of the proposed method is the simplistic construction of the dual meshes based on [25–27], which is essentially the same as for standard linear FVE methods. This feature offers a clear advantage over the traditional high order FVE methods [55,57], where one instead introduces a control volume for each node separately, resulting in increasingly complex control volumes that over-complicate the implementation and error analysis.

1.2 Layout

The main goal of this dissertation is to develop new mixed FVE methods for the approximation of two novel three-field formulations for linear elastostatics and poroelastostatics. In particular, we propose a FVE method for the approximation of linear elastostatics (formulated in terms of the displacement, the rotation vector, and the

solid pressure) and introduce a second-order FVE method for a stationary three-field poroelasticity problem on a two-dimensional (2D) domain, where we approximate the solid displacement, the pore pressure of the fluid and the total stress. The structure of the remainder of this dissertation is outlined below.

In Chapter 2, we propose a FVE method for the approximation of a new three-field formulation for linear elastostatics. In Section 2.1, we state the form of the linear elasticity equations that we will focus on, derive a suitable mixed weak formulation, and provide its stability and solvability analysis. A Galerkin method is introduced in Section 2.2, where we also obtain stability properties and a priori error estimates, following [3]. Section 2.3 concentrates on the development of a low-order FVE scheme, and its accuracy is studied in connection with the properties of the associated FE method. The convergence and robustness of the proposed methods is illustrated via a set of insightful computational tests collected in Section 2.4. Finally, we present a variational formulation considering mixed boundary conditions in Appendix A.

In Chapter 3, we propose a second-order FVE method for the approximation of a three-field poroelasticity problem on a 2D domain, formulated in terms of the solid displacement, the pore pressure of the fluid, and an auxiliary unknown that can be regarded as the total pressure or volumetric stress. In Section 3.1, we state the mixed three-field formulation of the poroelasticity equations that we will focus on, introduce the finite dimensional test and trial spaces to be employed, derive the discrete hybrid FVE formulation, and provide various stability results for the bilinear operators and linear functionals arising in the envisioned FVE scheme. In addition, we establish the coercivity of the second discrete bilinear form arising in the hybrid FVE formulation by closely following the approach from [25–28], where scalar elliptic BVPs are considered instead. In Section 3.2, the corresponding stability analysis and convergence estimates are derived. Finally, we motivate why we consider the balance law for the poroelasticity equations proposed in [11, 79] instead of the original setting from [62], see Appendix B.

In Chapter 4, final conclusions are drawn and subjects for further study are briefly introduced. In Section 4.1, we discuss three possible areas for future research, including a three-field elasticity-poroelasticity interface problem, a posteriori estimates of the proposed second-order FVE method for poroelasticity, and the numerical verification of its theoretical second order convergence rate. In Section 4.2, we review the main results of this dissertation in the form of a brief summary and provide concluding remarks.

Original contributions emanating from this dissertation include the particular

three-field formulation for linear elasticity introduced in Section 2.1 as well as the associated FE methods and the lowest-order FVE formulation introduced in Section 2.2 and Section 2.3, respectively. Furthermore, originality is claimed for the second-order mixed hybrid FVE method introduced in Chapter 3, where we remark that the novel mixed three-field formulation we focus on is closely related to the formulation proposed by Oyarzúa et al. [62].

1.3 Preliminaries

To close this chapter, we review some recurrent notations and introduce various general definitions. In addition, we highlight common assumptions made on the geometric properties of the domain and the corresponding domain partitions.

Throughout the remainder of this dissertation, we consider bounded, simply connected, Lipschitz domains $\Omega \subset \mathbb{R}^d$, for $d = 2, 3$, that are polytopic (i.e. polygons or polyhedrons) with polyhedral boundaries $\partial\Omega$. Observe that the latter assumption ensures that we do not have to deal with the issue of approximating the domain as well as the solution, which significantly simplifies the error estimates. In addition, we let $\mathscr{T} := \{\mathcal{T}_h\}_{h>0}$ denote a family of triangulations of the domain Ω, consisting of triangular elements K of diameter h_K, where we let $h := \max\{h_K : K \in \mathcal{T}_h\}$ denote the mesh parameter of the triangulation $\mathcal{T}_h \in \mathscr{T}$. We typically assume that the family $\mathscr{T}$ is shape regular in the sense of [29] and that the triangulations $\mathcal{T}_h$ are quasi-uniform (also in the sense of [29], which ensures that h is a good measurement of the convergence rate). In particular, if $d = 2$, we assume there exist thresholds $\theta_l > 0^\circ$ and $\theta_u < 180^\circ$, independent of h, such that

$$\theta_{\min,K} \geq \theta_l \quad \text{and} \quad \theta_{\max,K} \leq \theta_u \qquad \forall K \in \mathcal{T}_h \in \mathscr{T},$$

where $\theta_{\min,K}$ and $\theta_{\max,K}$ denote the smallest and largest interior angles of the triangular element K, respectively. Equivalently, we may assume there exist constants $\alpha > 1$ and $\beta > 0$, independent of h, such that

$$h_K \leq \alpha \rho_K \quad \text{and} \quad \beta h^2 \leq |K| \leq h^2 \qquad \forall K \in \mathcal{T}_h \in \mathscr{T},$$

where $|K|$ stands for the area of the element K, and

$$\rho_K := \sup\{\text{diam(C)} : \text{C is a disk contained in } K\}.$$

Furthermore, we let $\mathcal{N}_h$ denote the set of vertices in the mesh $\mathcal{T}_h$ and let $|\mathcal{N}_h|$ denote its cardinality.

The standard notation will be employed for Lebesgue spaces $\mathrm{L}^p(\Omega)$, $\mathrm{L}^2_0 := \{f \in \mathrm{L}^2(\Omega) : \int_\Omega f = 0\}$ and Sobolev functional spaces $\mathrm{H}^m(\Omega)$, $\mathrm{H}^1_0(\Omega) := \{f \in \mathrm{H}^1(\Omega) : f = 0 \text{ on } \partial\Omega\}$, which are endowed with their natural norms unless stated otherwise. In addition, $\mathcal{P}_k(U)$ will represent the space of polynomial functions of degree $l \leq k$ on the domain $U \subset \mathbb{R}^d$, we let $\mathbf{0}$ denote the generic null vector (including the null functional, the null vector and the null operator) on the relevant spaces, and the constants C, C_i, for $i \in \mathbb{N}_{\geq 0}$, are generic constants that are independent of the mesh parameter h and may take different values at different occurrences. Lastly, we let $||\cdot||_{m,\Omega}$ and $|\cdot|_{m,\Omega}$ denote the natural norms and seminorms on the Sobolev functional spaces $\mathrm{H}^m(\Omega)$, for $0 \leq m < \infty$, and define the broken $\mathrm{H}^2(\Omega)$-seminorm $|\cdot|_{2,\mathcal{T}_h}$ as

$$|\boldsymbol{u}_h|_{2,\mathcal{T}_h} := \left(\sum_{K \in \mathcal{T}_h} |\boldsymbol{u}_h|^2_{2,K} \right)^{1/2},$$

provided that the vectorial field $\boldsymbol{u}_h$ is sufficiently smooth.

Furthermore, letting d denote spatial dimension, for given vector fields $\boldsymbol{\theta} = (\theta_i)_{i=1}^d$, $\boldsymbol{v} = (v_i)_{i=1}^d$, we recall the following notation for differential operators:

$$\operatorname{div} \boldsymbol{v} := \partial_1 v_1 + \partial_2 v_2 + \partial_3 v_3, \quad \boldsymbol{\theta} \times \boldsymbol{v} := \begin{pmatrix} \theta_2 v_3 - \theta_3 v_2 \\ \theta_3 v_1 - \theta_1 v_3 \\ \theta_1 v_2 - \theta_2 v_1 \end{pmatrix}, \quad \mathbf{curl}\, \boldsymbol{v} := \begin{pmatrix} \partial_2 v_3 - \partial_3 v_2 \\ \partial_3 v_1 - \partial_1 v_3 \\ \partial_1 v_2 - \partial_2 v_1 \end{pmatrix}.$$

We also review a version of Green's formula (see, for instance, [41, Theorem 2.11]):

$$\int_\Omega \mathbf{curl}\, \boldsymbol{\omega} \cdot \boldsymbol{v} = \int_\Omega \boldsymbol{\omega} \cdot \mathbf{curl}\, \boldsymbol{v} + \langle \boldsymbol{\omega} \times \boldsymbol{n}, \boldsymbol{v} \rangle_{\partial\Omega}, \tag{1.3.1}$$

and the following useful identity

$$\mathbf{curl}(\mathbf{curl}\, \boldsymbol{v}) = -\Delta \boldsymbol{v} + \nabla(\operatorname{div} \boldsymbol{v}). \tag{1.3.2}$$

Chapter 2

Displacement-rotation-pressure formulation for linear elasticity

This chapter concentrates on the derivation and analysis of a linear FVE formulation for the linear elasticity equations written in terms of the displacement, the rotation vector, and the solid pressure. The proposed FVE scheme features mass conservativity on the dual control volumes, suitability for irregular domains and unstructured partitions, and robust approximations of displacements. In addition, we introduce a family of mixed FE methods for the approximation of the novel three-field saddle-point formulation of interest. The convergence rates and robustness of the proposed (Petrov-)Galerkin methods is verified numerically via a set of insightful tests.

2.1 The model problem

2.1.1 Displacement-rotation-pressure formulation

We assume that an isotropic and linearly elastic solid occupies a bounded and connected Lipschitz domain Ω of $\mathbb{R}^d$, with boundary $\partial\Omega$. Determining the deformation of a linearly elastic body subject to a volume load and with given boundary conditions, and adopting the hypothesis of small strains, results in the classical linear elasticity problem, formulated as follows. Given an external force $\widetilde{\boldsymbol{f}}$ and a prescribed boundary motion $\boldsymbol{g}$, we seek the displacements $\boldsymbol{u}$ such that

$$\mathbf{div}\big(2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}) + \lambda \operatorname{div}\boldsymbol{u}\,\mathbf{I}\big) = -\widetilde{\boldsymbol{f}} \text{ in } \Omega, \quad \boldsymbol{u} = \boldsymbol{g} \text{ on } \partial\Omega, \tag{2.1.1}$$

where $\boldsymbol{\varepsilon}(\boldsymbol{u}) = \frac{1}{2}(\nabla\boldsymbol{u} + \nabla\boldsymbol{u}^T)$ is the infinitesimal strain tensor, $\mathbf{I}$ denotes the $d \times d-$identity matrix, and μ, λ are the Lamé coefficients (intrinsic to the material prop-

erties of the solid, and here assumed constant). Next, following the pseudostress-based formulation recently introduced in [38] (and motivated by the seminal work [6]), one notices that using the identity

$$\mathbf{div}(\boldsymbol{\varepsilon}(\boldsymbol{u})) = \frac{1}{2}\Delta\boldsymbol{u} + \frac{1}{2}\nabla(\operatorname{div}\boldsymbol{u}),$$

and dividing the momentum equation by $\lambda + \mu$, we can rewrite (2.1.1) in the form of the well-known Cauchy-Navier (or Navier-Lamé) equations

$$\frac{\mu}{\lambda+\mu}\Delta\boldsymbol{u} + \nabla(\operatorname{div}\boldsymbol{u}) = -\boldsymbol{f} \text{ in } \Omega, \quad \boldsymbol{u} = \boldsymbol{g} \text{ on } \partial\Omega, \tag{2.1.2}$$

where the right-hand side has been rescaled as $\boldsymbol{f} = \frac{1}{\lambda+\mu}\widetilde{\boldsymbol{f}}$. We then proceed to define the auxiliary scaling parameter $\eta := \frac{\mu}{\lambda+\mu} > 0$, and recast (2.1.2) in a displacement-pressure formulation (considering $p = -\operatorname{div}\boldsymbol{u}$ as the solid pressure) as follows:

$$\begin{aligned} \eta\Delta\boldsymbol{u} - \nabla p &= -\boldsymbol{f} && \text{in } \Omega, \\ \operatorname{div}\boldsymbol{u} + p &= 0 && \text{in } \Omega, \\ \boldsymbol{u} &= \boldsymbol{g} && \text{on } \partial\Omega. \end{aligned} \tag{2.1.3}$$

At this point, and with the aim of deriving formulations whose stability holds independently of the Lamé coefficient λ, we introduce the field of rescaled rotations, $\boldsymbol{\omega} := \sqrt{\eta}\,\mathbf{curl}\,\boldsymbol{u}$, as an additional unknown. Exploiting (1.3.2) and the definition of the pressure in terms of displacements, we observe that (2.1.3) is fully equivalent to the following set of governing equations (in their pure-Dirichlet case): Find the displacement $\boldsymbol{u}$, the rotation $\boldsymbol{\omega}$ and the pressure p such that (see [22])

$$\sqrt{\eta}\,\mathbf{curl}\,\boldsymbol{\omega} + (1+\eta)\nabla p = \boldsymbol{f} \quad \text{in } \Omega, \tag{2.1.4}$$

$$\boldsymbol{\omega} - \sqrt{\eta}\,\mathbf{curl}\,\boldsymbol{u} = 0 \quad \text{in } \Omega, \tag{2.1.5}$$

$$\operatorname{div}\boldsymbol{u} + p = 0 \quad \text{in } \Omega, \tag{2.1.6}$$

$$\boldsymbol{u} = \boldsymbol{g} \quad \text{on } \partial\Omega. \tag{2.1.7}$$

On the other hand, should one necessitate to incorporate mixed boundary conditions, to e.g. impose a displacement $\boldsymbol{g}$ only on a part of the boundary Γ_{D}, and set a given traction $\widetilde{\boldsymbol{t}}$ on the remainder of the boundary, say $\Gamma_{\mathrm{N}} = \partial\Omega \setminus \Gamma_{\mathrm{D}}$, we can proceed as follows. First we realise that $\boldsymbol{\varepsilon}(\boldsymbol{u})\boldsymbol{n} = (\nabla\boldsymbol{u})\boldsymbol{n} - \frac{1}{2}\,\mathbf{curl}\,\boldsymbol{u} \times \boldsymbol{n}$ (where $\boldsymbol{n}$ denotes the outward unit normal on the boundary). Then, thanks to (2.1.1) and (2.1.4), the normal Cauchy stress can be recast in terms of the strain, rotations and pressure, such that the following set of mixed boundary conditions can be used instead of (2.1.7):

$$\boldsymbol{u} = \boldsymbol{g} \text{ on } \Gamma_{\mathrm{D}}, \quad \text{and} \quad 2\eta(\nabla\boldsymbol{u})\boldsymbol{n} - \sqrt{\eta}\,\boldsymbol{\omega} \times \boldsymbol{n} - (1-\eta)p\,\boldsymbol{n} = \boldsymbol{t} \text{ on } \Gamma_{\mathrm{N}}, \tag{2.1.8}$$

where the rescaled traction is $\boldsymbol{t} = \frac{1}{\lambda+\mu}\widetilde{\boldsymbol{t}}$. However, and for sake of clarity, we will restrict the presentation and analysis to the pure Dirichlet case $\Gamma_{\mathrm{D}} \equiv \partial\Omega$, considering only clamped boundaries $\boldsymbol{g} = \boldsymbol{0}$. A brief comment on how (2.1.8) is set up in a mixed variational formulation is postponed to Appendix A.

2.1.2 Weak form of the governing equations

Let us introduce the functional spaces

$$\mathrm{H} := \mathrm{H}_0^1(\Omega)^d, \quad \mathrm{Z} := \mathrm{L}^2(\Omega)^d, \quad \text{and} \quad \mathrm{Q} := \mathrm{L}^2(\Omega),$$

where Z and Q are endowed with their natural norms, and we recall the definition of the norm in the product space $\mathrm{Z} \times \mathrm{Q}$ as

$$\|(\boldsymbol{\theta}, q)\|_{\mathrm{Z}\times\mathrm{Q}}^2 := \|\boldsymbol{\theta}\|_{0,\Omega}^2 + \|q\|_{0,\Omega}^2.$$

On the other hand, for H we consider the following $\eta-$dependent scaled norm (see for instance, [41, Remark 2.7]):

$$\|\boldsymbol{v}\|_{\mathrm{H}}^2 := \eta\|\,\mathbf{curl}\,\boldsymbol{v}\|_{0,\Omega}^2 + \|\,\mathrm{div}\,\boldsymbol{v}\|_{0,\Omega}^2,$$

We proceed to test equation (2.1.6) against adequate functions, to integrate by parts in two terms, and to take into account the boundary conditions (2.1.7) in such a way that the resulting mixed variational formulation reads as follows: Find $((\boldsymbol{\omega}, p), \boldsymbol{u}) \in (\mathrm{Z} \times \mathrm{Q}) \times \mathrm{H}$ such that

$$a((\boldsymbol{\omega}, p), (\boldsymbol{\theta}, q)) + b((\boldsymbol{\theta}, q), \boldsymbol{u}) = 0 \qquad \forall (\boldsymbol{\theta}, q) \in \mathrm{Z} \times \mathrm{Q}, \tag{2.1.9}$$

$$b((\boldsymbol{\omega}, p), \boldsymbol{v}) = F(\boldsymbol{v}) \qquad \forall \boldsymbol{v} \in \mathrm{H}, \tag{2.1.10}$$

where the bilinear forms $a : (\mathrm{Z} \times \mathrm{Q}) \times (\mathrm{Z} \times \mathrm{Q}) \to \mathbb{R}$, $b : (\mathrm{Z} \times \mathrm{Q}) \times \mathrm{H} \to \mathbb{R}$, together with the linear functional $F : \mathrm{H} \to \mathbb{R}$ are defined as

$$\begin{aligned} a((\boldsymbol{\omega}, p), (\boldsymbol{\theta}, q)) &:= \int_\Omega \boldsymbol{\omega} \cdot \boldsymbol{\theta} + (1+\eta) \int_\Omega pq, \\ b((\boldsymbol{\theta}, q), \boldsymbol{v}) &:= (1+\eta) \int_\Omega q\,\mathrm{div}\,\boldsymbol{v} - \sqrt{\eta} \int_\Omega \boldsymbol{\theta} \cdot \mathbf{curl}\,\boldsymbol{v}, \\ F(\boldsymbol{v}) &:= -\int_\Omega \boldsymbol{f} \cdot \boldsymbol{v}. \end{aligned}$$

2.1.3 Well-posedness

The unique solvability of problem (2.1.9)-(2.1.10), together with the continuous dependence on the data will be established using the well-known Babuška-Brezzi theory.

We first observe that the bilinear forms $a(\cdot,\cdot)$, $b(\cdot,\cdot)$ and the linear functional $F(\cdot)$ are all bounded by positive constants independent of η (and therefore independent of the Lamé coefficient λ). In addition, the bilinear form $a(\cdot,\cdot)$ is $(\mathrm{Z}\times\mathrm{Q})-$elliptic, uniformly with respect to the scaling parameter η, as stated in the following result.

Lemma 2.1.1. *There exists $\alpha>0$, independent of η, such that*

$$a((\boldsymbol{\theta},q),(\boldsymbol{\theta},q)) \geq \alpha\|(\boldsymbol{\theta},q)\|^2_{\mathrm{Z}\times\mathrm{Q}} \qquad \forall\,(\boldsymbol{\theta},q)\in \mathrm{Z}\times\mathrm{Q}.$$

Moreover, the following inf-sup condition holds for the bilinear form $b(\cdot,\cdot)$.

Lemma 2.1.2. *There exists $C>0$, independent of η, such that*

$$\sup_{(\boldsymbol{\theta},q)\in\mathrm{Z}\times\mathrm{Q}} \frac{b((\boldsymbol{\theta},q),\boldsymbol{v})}{\|(\boldsymbol{\theta},q)\|_{\mathrm{Z}\times\mathrm{Q}}} \geq C\|\boldsymbol{v}\|_{\mathrm{H}} \qquad \forall \boldsymbol{v}\in\mathrm{H}.$$

Proof. Let us consider a generic $\boldsymbol{v}\in\mathrm{H}$ and define

$$\widetilde{\boldsymbol{\theta}} := -\sqrt{\eta}\,\mathbf{curl}\,\boldsymbol{v}\in\mathrm{Z}, \quad \text{and} \quad \tilde{q}:=\operatorname{div}\boldsymbol{v}\in\mathrm{Q}.$$

We immediately observe that

$$\|(\widetilde{\boldsymbol{\theta}},\tilde{q})\|_{\mathrm{Z}\times\mathrm{Q}} \leq \|\boldsymbol{v}\|_{\mathrm{H}},$$

and from the definition of $b(\cdot,\cdot)$, we readily obtain

$$\sup_{(\boldsymbol{\theta},q)\in\mathrm{Z}\times\mathrm{Q}} \frac{b((\boldsymbol{\theta},q),\boldsymbol{v})}{\|(\boldsymbol{\theta},q)\|_{\mathrm{Z}\times\mathrm{Q}}} \geq \frac{b((\widetilde{\boldsymbol{\theta}},\tilde{q}),\boldsymbol{v})}{\|(\widetilde{\boldsymbol{\theta}},\tilde{q})\|_{\mathrm{Z}\times\mathrm{Q}}} \geq C\|\boldsymbol{v}\|_{\mathrm{H}} \qquad \forall \boldsymbol{v}\in\mathrm{H},$$

which finishes the proof. □

We are now in a position to state the unique solvability and stability of the continuous problem (2.1.9)-(2.1.10).

Theorem 2.1.3. *There exists a unique solution $((\boldsymbol{\omega},p),\boldsymbol{u})\in(\mathrm{Z}\times\mathrm{Q})\times\mathrm{H}$ to problem (2.1.9)-(2.1.10), which satisfies the following continuous dependence on the data:*

$$\|(\boldsymbol{\omega},p)\|_{\mathrm{Z}\times\mathrm{Q}} + \|\boldsymbol{u}\|_{\mathrm{H}} \leq C\|\boldsymbol{f}\|_{0,\Omega}.$$

Proof. By virtue of the general theory for saddle-point problems (see, for instance, [37]), the desired result follows from a direct application of Lemmas 2.1.1 and 2.1.2. □

2.2 Finite element discretisation

In this section, we introduce a Galerkin scheme associated to problem (2.1.9)-(2.1.10), we specify the finite dimensional subspaces to employ, and analyse the well-posedness of the resulting methods using suitable assumptions on the discrete spaces. We also derive optimal a priori error estimates for the proposed method.

2.2.1 Formulation and solvability

First, recall that $\mathscr{T} = \{\mathcal{T}_h(\Omega)\}_{h>0}$ denotes a shape regular family of quasi-uniform partitions of the domain Ω by tetrahedrons (if $d = 3$, or triangles if $d = 2$) K of diameter h_K, having mesh size $h := \max\{h_K : \ K \in \mathcal{T}_h(\Omega)\}$. Furthermore, given an integer $k \geq 1$ and a set $S \subset \mathbb{R}^d$, recall that the space of polynomial functions defined in S and having total degree $l \leq k$ will be denoted by $\mathcal{P}_k(S)$.

Next, we define the following discrete function spaces:

$$\begin{aligned}
\mathrm{H}_h &:= \{\boldsymbol{v}_h \in \mathrm{H} : \boldsymbol{v}_h|_K \in \mathcal{P}_k(K)^d \quad \forall K \in \mathcal{T}_h(\Omega)\},\\
\mathrm{Z}_h &:= \{\boldsymbol{\theta}_h \in \mathrm{Z} : \boldsymbol{\theta}_h|_K \in \mathcal{P}_{k-1}(K)^d \quad \forall K \in \mathcal{T}_h(\Omega)\},\\
\mathrm{Q}_h &:= \{q_h \in \mathrm{Q} : q_h|_K \in \mathcal{P}_{k-1}(K) \quad \forall K \in \mathcal{T}_h(\Omega)\},
\end{aligned}$$

which are subspaces of H, Z and Q, respectively, and give rise to the following Galerkin scheme associated to the continuous variational formulation (2.1.9)-(2.1.10): Find $((\boldsymbol{\omega}_h, p_h), \boldsymbol{u}_h) \in (\mathrm{Z}_h \times \mathrm{Q}_h) \times \mathrm{H}_h$ such that

$$a((\boldsymbol{\omega}_h, p_h), (\boldsymbol{\theta}_h, q_h)) + b((\boldsymbol{\theta}_h, q_h), \boldsymbol{u}_h) = 0 \qquad \forall (\boldsymbol{\theta}_h, q_h) \in \mathrm{Z}_h \times \mathrm{Q}_h, \quad (2.2.1)$$

$$b((\boldsymbol{\omega}_h, p_h), \boldsymbol{v}_h) = F(\boldsymbol{v}_h) \qquad \forall \boldsymbol{v}_h \in \mathrm{H}_h. \quad (2.2.2)$$

Our next goal is to establish discrete counterparts of Lemmas 2.1.1 and 2.1.2, leading to the solvability and stability of the Galerkin method (2.2.1)-(2.2.2). Their proofs are obtained using the same arguments exploited in the continuous case and are left out for the sake of brevity.

Lemma 2.2.1. *There exists $\alpha > 0$, independent of η, such that*

$$a((\boldsymbol{\theta}_h, q_h), (\boldsymbol{\theta}_h, q_h)) \geq \alpha \|(\boldsymbol{\theta}_h, q_h)\|^2_{\mathrm{Z}\times\mathrm{Q}}.$$

Lemma 2.2.2. *There exists $C > 0$, independent of η, such that*

$$\sup_{(\boldsymbol{\theta}_h, q_h) \in \mathrm{Z}_h \times \mathrm{Q}_h} \frac{b((\boldsymbol{\theta}_h, q_h), \boldsymbol{v}_h)}{\|(\boldsymbol{\theta}_h, q_h)\|_{\mathrm{Z}\times\mathrm{Q}}} \geq C \|\boldsymbol{v}_h\|_{\mathrm{H}} \qquad \forall \boldsymbol{v}_h \in \mathrm{H}_h.$$

We can now state the unique solvability, stability, and convergence properties of the discrete problem (2.2.1)-(2.2.2) in form of the three following theorems.

Theorem 2.2.3. *There exists a unique* $((\boldsymbol{\omega}_h, p_h), \boldsymbol{u}_h) \in (\mathrm{Z}_h \times \mathrm{Q}_h) \times \mathrm{H}_h$ *solution of the discrete problem* (2.2.1)-(2.2.2)*. Moreover, there exists a constant* $C > 0$*, independent of* h *and* η*, such that*

$$\|(\boldsymbol{\omega}_h, p_h)\|_{\mathrm{Z}\times\mathrm{Q}} + \|\boldsymbol{u}_h\|_{\mathrm{H}} \leq C\|\boldsymbol{f}\|_{0,\Omega}.$$

In addition, the following approximation property is satisfied

$$\|(\boldsymbol{\omega}-\boldsymbol{\omega}_h, p-p_h)\|_{\mathrm{Z}\times\mathrm{Q}}+\|\boldsymbol{u}-\boldsymbol{u}_h\|_{\mathrm{H}} \leq C \inf_{\substack{((\boldsymbol{\theta}_h,q_h),\boldsymbol{v}_h)\\ \in(\mathrm{Z}_h\times\mathrm{Q}_h)\times\mathrm{H}_h}} \|(\boldsymbol{\omega}-\boldsymbol{\theta}_h, p-q_h)\|_{\mathrm{Z}\times\mathrm{Q}}+\|\boldsymbol{u}-\boldsymbol{v}_h\|_{\mathrm{H}},$$

where $((\boldsymbol{\omega}, p), \boldsymbol{u}) \in (\mathrm{Z} \times \mathrm{Q}) \times \mathrm{H}$ *is the unique solution of the mixed variational formulation* (2.1.9)-(2.1.10)*.*

Theorem 2.2.4. *Let* $((\boldsymbol{\omega}, p), \boldsymbol{u}) \in (\mathrm{Z} \times \mathrm{Q}) \times \mathrm{H}$ *and* $((\boldsymbol{\omega}_h, p_h), \boldsymbol{u}_h) \in (\mathrm{Z}_h \times \mathrm{Q}_h) \times \mathrm{H}_h$ *be the solutions of the continuous and discrete problems* (2.1.9)-(2.1.10) *and* (2.2.1)-(2.2.2)*, respectively. Then there exists* $C > 0$*, independent of* h *and* η*, such that*

$$\|(\boldsymbol{\omega} - \boldsymbol{\omega}_h, p - p_h)\|_{\mathrm{Z}\times\mathrm{Q}} + \|\boldsymbol{u} - \boldsymbol{u}_h\|_{\mathrm{H}} \leq Ch^k(\|\boldsymbol{\omega}\|_{k,\Omega} + \|p\|_{k,\Omega} + \|\boldsymbol{u}\|_{k+1,\Omega}).$$

Proof. The result follows from Theorem 2.2.3 and the standard error estimates for the Lagrange interpolant of $\boldsymbol{u}$ and the vectorial and scalar L^2-orthogonal projections for $\boldsymbol{\omega}$ and p, respectively. □

To close this section, we observe that the convergence of the displacement error can also be measured in the $\mathrm{L}^2(\Omega)^d-$norm. The next result can be established following a classical duality strategy (for a complete proof, see our unpublished paper [3]), of which details are left out since the main focus lies on FVE discretisations instead.

Theorem 2.2.5. *Let* $((\boldsymbol{\omega}, p), \boldsymbol{u}) \in (\mathrm{Z} \times \mathrm{Q}) \times \mathrm{H}$ *and* $((\boldsymbol{\omega}_h, p_h), \boldsymbol{u}_h) \in (\mathrm{Z}_h \times \mathrm{Q}_h) \times \mathrm{H}_h$ *be the solutions of the continuous and discrete problems* (2.1.9)-(2.1.10) *and* (2.2.1)-(2.2.2)*, respectively. Then, there exists a constant* $C > 0$*, independent of* h *and* η*, such that*

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{0,\Omega} \leq Ch^{k+1}(\|\boldsymbol{\omega}\|_{k,\Omega} + \|p\|_{k,\Omega} + \|\boldsymbol{u}\|_{k+1,\Omega}).$$

2.3 A finite volume element scheme

In addition to the mesh $\mathcal{T}_h$ (from now on, the primal mesh), we introduce another tessellation of Ω, denoted by $\mathcal{T}_h^\star$ and referred to as the dual mesh, where for each element $K \in \mathcal{T}_h$ we create segments joining its barycenter b_K with the midpoints (2D barycenters) m_F of each face $F \subset \partial K$ (or the midpoints of each edge, in 2D), forming four polyhedra (or three quadrilaterals, in the 2D case) Q_z for z in the set of vertices of K, that is, $z \in \mathcal{N}_h \cap K$. Then to each vertex $s_j \in \mathcal{N}_h$, we associate a so-called control volume $K_j^\star$, consisting of the union of the polyhedra (quadrilaterals in 2D) Q_{s_j} sharing the vertex s_j. A sketch of the resulting control volume associated to s_j is depicted in Figure 2.1(a).

Next, we consider the following discrete lowest-order FE spaces:

$$\begin{aligned}
\mathrm{H}_h &:= \{\boldsymbol{v}_h \in \mathrm{H} : \boldsymbol{v}_h|_K \in \mathcal{P}_1(K)^d \quad \forall K \in \mathcal{T}_h(\Omega)\},\\
\mathrm{Z}_h &:= \{\boldsymbol{\theta}_h \in \mathrm{Z} : \boldsymbol{\theta}_h|_K \in \mathcal{P}_0(K)^d \quad \forall K \in \mathcal{T}_h(\Omega)\},\\
\mathrm{Q}_h &:= \{q_h \in \mathrm{Q} : q_h|_K \in \mathcal{P}_0(K) \quad \forall K \in \mathcal{T}_h(\Omega)\},
\end{aligned}$$

which are clearly subspaces of H, Z and Q, respectively.

In its lowest-order version, a FVE method for the approximation of (2.1.9)-(2.1.10) can be constructed by associating discrete spaces to a dual partition of the domain:

$$\begin{aligned}
\mathrm{H}_h^\star := \big\{\boldsymbol{v}_h \in \mathrm{L}^2(\Omega)^d :\ &\boldsymbol{v}_h|_{K_j^\star} \in \mathcal{P}_0(K_j^\star)^d \text{ for all } K_j^\star \in \mathcal{T}_h^\star,\\
&\boldsymbol{v}_h|_{K_j^\star} = \mathbf{0} \text{ if } K_j^\star \text{ is a control volume on } \partial\Omega\big\},
\end{aligned}$$

where we notice that no additional space is introduced for the FVE approximation of $\boldsymbol{\omega}$ or p. Furthermore, we define the $\mathcal{T}_h^\star-$piecewise lumping map $\mathcal{H}_h : \mathrm{H}_h \to \mathrm{H}_h^\star$ which relates the primal and conforming dual meshes by

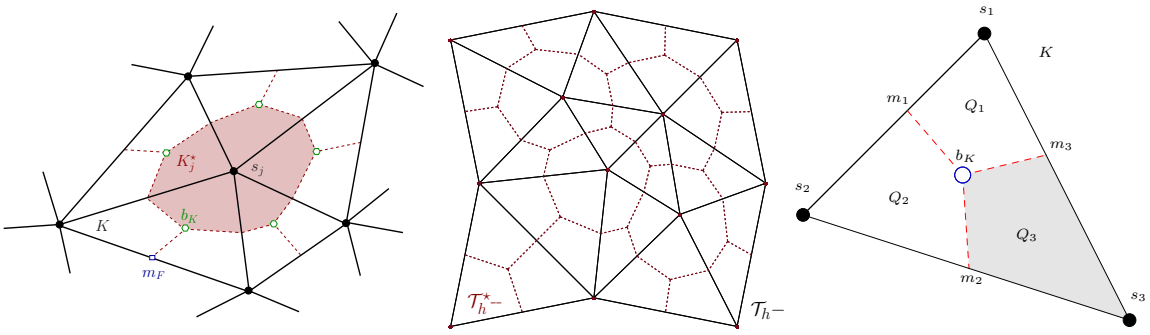
$$\boldsymbol{v}_h(\boldsymbol{x}) = \sum_j \boldsymbol{v}_h(s_j)\boldsymbol{\varphi}_j(\boldsymbol{x}) \mapsto \mathcal{H}_h\boldsymbol{v}_h(\boldsymbol{x}) = \sum_j \boldsymbol{v}_h(s_j)\boldsymbol{\chi}_j(\boldsymbol{x}),$$

for all $\boldsymbol{v}_h \in \mathrm{H}_h$, where $\boldsymbol{\chi}_j$ is the vectorial characteristic function of the control volume $K_j^\star$ and $\{\boldsymbol{\varphi}_j\}_j$ is the canonical FE basis of H_h (cf. [64]). For any $\boldsymbol{v}_h \in \mathrm{H}_h$, this operator satisfies the interpolation bound (see, for instance, [24])

$$\|\boldsymbol{v}_h - \mathcal{H}_h\boldsymbol{v}_h\|_{0,\Omega} \leq Ch|\boldsymbol{v}_h|_{1,\Omega}.$$

In addition, since $\mathrm{H} := \mathrm{H}_0^1(\Omega)^d = \mathrm{H}_0(\mathbf{curl};\Omega) \cap \mathrm{H}_0(\mathrm{div};\Omega)$, remark [41, Remark 2.7] implies that the operator $\mathcal{H}_h(\cdot)$ also satisfies

$$\|\boldsymbol{v}_h - \mathcal{H}_h\boldsymbol{v}_h\|_{0,\Omega} \leq Ch||\boldsymbol{v}_h||_{\mathrm{H}}, \tag{2.3.1}$$



(a) (b) (c)

Figure 2.1: Sketch of five elements in the primal mesh $\mathcal{T}_h$ sharing the vertex s_j and employed to construct the control volume $K_j^\star$ belonging to the dual partition $\mathcal{T}_h^\star$ (a); example of coarse primal and dual meshes (b); and one triangular element $K \in \mathcal{T}_h$ with barycenter b_K, where the m_i's denote the midpoints of the edges, and the Q_i's are the quadrilaterals that form the control volumes (c).

which plays a role in the convergence proof for the envisioned FVE method.

The discrete FVE formulation is obtained by multiplying (2.1.4) by $\boldsymbol{v}_h^\star \in \mathrm{H}_h^\star$ and integrating by parts over each $K_j^\star \in \mathcal{T}_h^\star$, multiplying (2.1.5) by $\boldsymbol{\theta}_h \in \mathrm{Z}_h$ and integrating by parts over each $K \in \mathcal{T}_h$, and multiplying (2.1.6) by $(1+\eta)q_h$, for $q_h \in \mathrm{Q}_h$, and integrating by parts over each $K \in \mathcal{T}_h$. This, along with identity (1.3.1), results in a Petrov-Galerkin formulation that reads as follows: Find $((\hat{\boldsymbol{\omega}}_h, \hat{p}_h), \hat{\boldsymbol{u}}_h) \in (\mathrm{Z}_h \times \mathrm{Q}_h) \times \mathrm{H}_h$ such that

$$a((\hat{\boldsymbol{\omega}}_h, \hat{p}_h), (\boldsymbol{\theta}_h, q_h)) + b((\boldsymbol{\theta}_h, q_h), \hat{\boldsymbol{u}}_h) = 0 \qquad \forall (\boldsymbol{\theta}_h, q_h) \in \mathrm{Z}_h \times \mathrm{Q}_h, \quad (2.3.2)$$

$$\tilde{b}((\hat{\boldsymbol{\omega}}_h, \hat{p}_h), \boldsymbol{v}_h^\star) = \tilde{F}(\boldsymbol{v}_h^\star) \qquad \forall \boldsymbol{v}_h^\star \in \mathrm{H}_h^\star, \quad (2.3.3)$$

where the bilinear form $\tilde{b} : (\mathrm{Z}_h \times \mathrm{Q}_h) \times \mathrm{H}_h^\star \to \mathbb{R}$ and the linear functional $\tilde{F} : \mathrm{H}_h^\star \to \mathbb{R}$ are defined as

$$\tilde{b}((\boldsymbol{\theta}_h, q_h), \boldsymbol{v}_h^\star) := -(1+\eta) \sum_{j=1}^{|\mathcal{N}_h|} \int_{\partial K_j^\star} q_h (\boldsymbol{v}_h^\star \cdot \boldsymbol{n}) - \sqrt{\eta} \sum_{j=1}^{|\mathcal{N}_h|} \int_{\partial K_j^\star} (\boldsymbol{\theta}_h \times \boldsymbol{n}) \cdot \boldsymbol{v}_h^\star,$$

$$\tilde{F}(\boldsymbol{v}_h^\star) := -\sum_{j=1}^{|\mathcal{N}_h|} \int_{K_j^\star} \boldsymbol{f} \cdot \boldsymbol{v}_h^\star.$$

Observe that the bilinear form $\tilde{b}(\cdot,\cdot)$ and the linear functional $\tilde{F}(\cdot)$ are both bounded by positive constants independent of η. We also introduce the bilinear form $B : (\mathrm{Z}_h \times \mathrm{Q}_h) \times \mathrm{H}_h \to \mathbb{R}$, which is defined by

$$B((\boldsymbol{\theta}_h, q_h), \boldsymbol{v}_h) := \tilde{b}((\boldsymbol{\theta}_h, q_h), \mathcal{H}_h \boldsymbol{v}_h)$$

which allows us to recast the Petrov-Galerkin formulation (2.3.2)-(2.3.3) as a standard Galerkin method. If one is interested in imposing mixed boundary conditions (prescribing a displacement $\boldsymbol{g}$ on $\Gamma_{\mathrm{D}} \subset \partial\Omega$, and setting a given traction $\widetilde{\boldsymbol{t}}$ on the remainder of the boundary), we can modify the FVE scheme as detailed in Appendix A.

We proceed to establish a relationship between the bilinear forms $b(\cdot,\cdot)$ and $B(\cdot,\cdot)$, which will be useful to carry out the error analysis in a finite-element-fashion. For the sake of brevity, only the proof for the two-dimensional case is provided. The proof for the three-dimensional case follows in an analogous manner, where we instead consider polyhedral control volumes and boundary surfaces rather than boundary edges.

Lemma 2.3.1. *For any* $(\boldsymbol{\theta}_h, q_h) \in \mathrm{Z}_h \times \mathrm{Q}_h$ *and* $\boldsymbol{v}_h \in \mathrm{H}_h$*, one has*

$$B((\boldsymbol{\theta}_h, q_h), \boldsymbol{v}_h) := \tilde{b}((\boldsymbol{\theta}_h, q_h), \mathcal{H}_h \boldsymbol{v}_h) = b((\boldsymbol{\theta}_h, q_h), \boldsymbol{v}_h). \tag{2.3.4}$$

Proof. First, let g be a function that is continuous on the interior of each quadrilateral Q_j (as shown in Figure 2.1(c)), with $\int_e g = 0$ for any boundary edge e. Using Figure 2.1(c), it is straightforward to show that the following relation holds (cf. [78]):

$$\sum_{j=1}^{|\mathcal{N}_h|} \int_{\partial K_j^\star} g = \sum_{K \in \mathcal{T}_h} \sum_{j=1}^{3} \int_{m_{j+1} b_K m_j} g,$$

where $m_{j+1} b_K m_j$ denotes the union of the line segments $m_{j+1} b_K$ and $b_K m_j$. We take $m_{j+3} = m_j$ in the case that the index is out of bound.

Next, from the definition of the transfer operator $\mathcal{H}_h(\cdot)$, we find that

$$B((\boldsymbol{\theta}_h, q_h), \boldsymbol{v}_h) = -(1+\eta) \sum_{j=1}^{|\mathcal{N}_h|} \int_{\partial K_j^\star} q_h \boldsymbol{v}_h(s_j) \cdot \boldsymbol{n} - \sqrt{\eta} \sum_{j=1}^{|\mathcal{N}_h|} \int_{\partial K_j^\star} (\boldsymbol{\theta}_h \times \boldsymbol{n}) \cdot \boldsymbol{v}_h(s_j).$$

In order to arrive at (2.3.4), we use the definition of $B(\cdot,\cdot)$ in combination with integration by parts and the fact that both q_h and $\boldsymbol{v}_h(s_j)$ are constant in the interior of each quadrilateral Q_j to obtain

$$\begin{aligned}
B((\boldsymbol{\theta}_h, q_h), \boldsymbol{v}_h) &= -(1+\eta) \sum_{K \in \mathcal{T}_h} \sum_{j=1}^{3} \int_{m_{j+1} b_K m_j} q_h \boldsymbol{v}_h(s_{j+1}) \cdot \boldsymbol{n} \\
&\quad - \sqrt{\eta} \sum_{K \in \mathcal{T}_h} \sum_{j=1}^{3} \int_{m_{j+1} b_K m_j} (\boldsymbol{\theta}_h \times \boldsymbol{n}) \cdot \boldsymbol{v}_h(s_{j+1}) \\
&= (1+\eta) \sum_{K \in \mathcal{T}_h} \sum_{j=1}^{3} q_h \int_{s_j m_j \cup m_j s_{j+1}} \boldsymbol{v}_h(s_{j+1}) \cdot \boldsymbol{n} \\
&\quad + \sqrt{\eta} \sum_{K \in \mathcal{T}_h} \sum_{j=1}^{3} \int_{s_j m_j \cup m_j s_{j+1}} (\boldsymbol{\theta}_h \times \boldsymbol{n}) \cdot \boldsymbol{v}_h(s_{j+1}).
\end{aligned}$$

Since q_h and $\boldsymbol{\theta}_h$ are constant on the edges of each element $K \in \mathcal{T}_h$, we can write

$$\begin{aligned} B((\boldsymbol{\theta}_h, q_h), \boldsymbol{v}_h) &= (1+\eta) \sum_{K \in \mathcal{T}_h} \sum_{j=1}^{3} q_h \int_{s_j s_{j+1}} \boldsymbol{v}_h \cdot \boldsymbol{n} + \sqrt{\eta} \sum_{K \in \mathcal{T}_h} \sum_{j=1}^{3} \int_{s_j s_{j+1}} (\boldsymbol{\theta}_h \times \boldsymbol{n}) \cdot \boldsymbol{v}_h \\ &= (1+\eta) \sum_{K \in \mathcal{T}_h} \int_{\partial K} q_h \boldsymbol{v}_h \cdot \boldsymbol{n} + \sqrt{\eta} \sum_{K \in \mathcal{T}_h} \int_{\partial K} (\boldsymbol{\theta}_h \times \boldsymbol{n}) \cdot \boldsymbol{v}_h. \end{aligned}$$

Then, after one application of integration by parts and identity (1.3.1), we arrive at

$$\begin{aligned} B((\boldsymbol{\theta}_h, q_h), \boldsymbol{v}_h) &= (1+\eta) \sum_{K \in \mathcal{T}_h} \int_K q_h \operatorname{div} \boldsymbol{v}_h - \sqrt{\eta} \sum_{K \in \mathcal{T}_h} \int_K \boldsymbol{\theta}_h \cdot \mathbf{curl}\, \boldsymbol{v}_h \\ &= b((\boldsymbol{\theta}_h, q_h), \boldsymbol{v}_h), \end{aligned}$$

which finishes the proof. □

Our next goal is to prove a FVE-counterpart of Lemma 2.2.2, leading to the solvability and stability of (2.3.2)-(2.3.3). Recall that Lemma 2.2.1 establishes that the bilinear form $a(\cdot,\cdot)$ is $(\mathrm{Z}_h \times \mathrm{Q}_h)-$elliptic, uniformly with respect to the auxiliary parameter η. It is also straightforward to show that the bilinear form $B(\cdot,\cdot)$ satisfies the inf-sup condition, as stated in the following result.

Lemma 2.3.2. *There exists $C > 0$, independent of η, such that*

$$\sup_{(\boldsymbol{\theta}_h, q_h) \in \mathrm{Z}_h \times \mathrm{Q}_h} \frac{B((\boldsymbol{\theta}_h, q_h), \boldsymbol{v}_h)}{\|(\boldsymbol{\theta}_h, q_h)\|_{\mathrm{Z}\times\mathrm{Q}}} \geq C \|\boldsymbol{v}_h\|_{\mathrm{H}} \qquad \forall \boldsymbol{v}_h \in \mathrm{H}_h.$$

Proof. This result immediately follows from a straightforward application of Lemma 2.3.1 in combination with Lemma 2.2.2. □

Analogously to the previous section, the following two theorems formulate the unique solvability, stability, quasi-optimality, and convergence properties of the discrete problem (2.3.2)-(2.3.3).

Theorem 2.3.3. *There exists a unique $((\hat{\boldsymbol{\omega}}_h, \hat{p}_h), \hat{\boldsymbol{u}}_h) \in (\mathrm{Z}_h \times \mathrm{Q}_h) \times \mathrm{H}_h$ solution of the discrete problem* (2.3.2)-(2.3.3)*. Moreover, there exists a constant $C > 0$, independent of h and η, such that*

$$\|(\hat{\boldsymbol{\omega}}_h, \hat{p}_h)\|_{\mathrm{Z}\times\mathrm{Q}} + \|\hat{\boldsymbol{u}}_h\|_{\mathrm{H}} \leq C \|\boldsymbol{f}\|_{0,\Omega}.$$

In addition, the following best approximation result is satisfied

$$\|(\boldsymbol{\omega} - \hat{\boldsymbol{\omega}}_h, p - \hat{p}_h)\|_{\mathrm{Z}\times\mathrm{Q}} + \|\boldsymbol{u} - \hat{\boldsymbol{u}}_h\|_{\mathrm{H}} \leq C \inf_{\substack{((\boldsymbol{\theta}_h, q_h), \boldsymbol{v}_h) \\ \in (\mathrm{Z}_h \times \mathrm{Q}_h) \times \mathrm{H}_h}} \|(\boldsymbol{\omega} - \boldsymbol{\theta}_h, p - q_h)\|_{\mathrm{Z}\times\mathrm{Q}} + \|\boldsymbol{u} - \boldsymbol{v}_h\|_{\mathrm{H}},$$

where $((\boldsymbol{\omega}, p), \boldsymbol{u}) \in (\mathrm{Z} \times \mathrm{Q}) \times \mathrm{H}$ is the unique solution of the mixed variational formulation (2.1.9)-(2.1.10)*.*

The next lemma establishes linear convergence, which is expected as the only difference with respect to the FE scheme is the right-hand side.

Theorem 2.3.4. *Let* $((\boldsymbol{\omega}, p), \boldsymbol{u}) \in (\mathrm{Z} \times \mathrm{Q}) \times \mathrm{H}$ *and* $((\hat{\boldsymbol{\omega}}_h, \hat{p}_h), \hat{\boldsymbol{u}}_h) \in (\mathrm{Z}_h \times \mathrm{Q}_h) \times \mathrm{H}_h$ *be the solutions of the continuous and discrete problems* (2.1.9)-(2.1.10) *and* (2.3.2)-(2.3.3)*, respectively. Then there exists* $C > 0$*, independent of* h *and* η*, such that*

$$\|(\boldsymbol{\omega} - \hat{\boldsymbol{\omega}}_h, p - \hat{p}_h)\|_{\mathrm{Z}\times\mathrm{Q}} + \|\boldsymbol{u} - \hat{\boldsymbol{u}}_h\|_{\mathrm{H}} \leq Ch(\|\boldsymbol{\omega}\|_{1,\Omega} + \|p\|_{1,\Omega} + \|\boldsymbol{u}\|_{2,\Omega}).$$

Proof. Let $((\boldsymbol{\omega}_h, p_h), \boldsymbol{u}_h)$ and $((\hat{\boldsymbol{\omega}}_h, \hat{p}_h), \hat{\boldsymbol{u}}_h)$ denote the solutions to the FE formulation (2.2.1)-(2.2.2) and the FVE formulation (2.3.2)-(2.3.3), respectively. In addition, we introduce the notation $\hat{\boldsymbol{r}}_{\boldsymbol{\omega}} := \boldsymbol{\omega}_h - \hat{\boldsymbol{\omega}}_h$, $\hat{r}_p := p_h - \hat{p}_h$ and $\hat{\boldsymbol{r}}_{\boldsymbol{u}} := \boldsymbol{u}_h - \hat{\boldsymbol{u}}_h$.

Observe that Lemma 2.3.1 readily implies that

$$\begin{aligned} a((\hat{\boldsymbol{r}}_{\boldsymbol{\omega}}, \hat{r}_p), (\boldsymbol{\theta}_h, q_h)) + b((\boldsymbol{\theta}_h, q_h), \hat{\boldsymbol{r}}_{\boldsymbol{u}}) &= 0 && \forall (\boldsymbol{\theta}_h, q_h) \in \mathrm{Z}_h \times \mathrm{Q}_h, \\ b((\hat{\boldsymbol{r}}_{\boldsymbol{\omega}}, \hat{r}_p), \boldsymbol{v}_h) &= F(\boldsymbol{v}_h - \mathcal{H}_h \boldsymbol{v}_h) && \forall \boldsymbol{v}_h \in \mathrm{H}_h, \end{aligned}$$

such that substituting $(\boldsymbol{\omega}_h - \hat{\boldsymbol{\omega}}_h, p_h - \hat{p}_h)$ for $(\boldsymbol{\theta}_h, q_h)$ and $\boldsymbol{u}_h - \hat{\boldsymbol{u}}_h$ for $\boldsymbol{v}_h$ gives

$$a((\hat{\boldsymbol{r}}_{\boldsymbol{\omega}}, \hat{r}_p), (\hat{\boldsymbol{r}}_{\boldsymbol{\omega}}, \hat{r}_p)) = -b((\hat{\boldsymbol{r}}_{\boldsymbol{\omega}}, \hat{r}_p), \hat{\boldsymbol{r}}_{\boldsymbol{u}}), \tag{2.3.5}$$

$$b((\hat{\boldsymbol{r}}_{\boldsymbol{\omega}}, \hat{r}_p), \hat{\boldsymbol{r}}_{\boldsymbol{u}}) = F(\hat{\boldsymbol{r}}_{\boldsymbol{u}} - \mathcal{H}_h \hat{\boldsymbol{r}}_{\boldsymbol{u}}). \tag{2.3.6}$$

Next, combining equation (2.3.5) with equation (2.3.6) relates the bilinear form $a(\cdot, \cdot)$ and the linear functional $F(\cdot)$, such that (2.3.1) in combination with Lemma 2.2.1 implies that there exists a constant $C_0 > 0$, independent of h and η, such that

$$||(\boldsymbol{\omega}_h - \hat{\boldsymbol{\omega}}_h, p_h - \hat{p}_h)||_{\mathrm{Z}\times\mathrm{Q}} \leq C_0 h. \tag{2.3.7}$$

Moreover, after applying the inf-sup condition from Lemma 2.3.2 and employing identity (2.3.7), standard arguments imply that there exists a constant $C_1 > 0$, independent of h and η, such that

$$||\boldsymbol{u}_h - \hat{\boldsymbol{u}}_h||_{\mathrm{H}} \leq C_1 h. \tag{2.3.8}$$

Lastly, applying the triangle inequality to the convergence bound for the FE method established in Theorem 2.2.4 in combination with the inequalities (2.3.8) and (2.3.7) finishes the proof. □

To close this section, we prove an L^2-estimate for the displacement error. For this purpose we first state a preliminary result (cf. [64]) that involves the transfer operator $\mathcal{H}_h(\cdot)$.

Lemma 2.3.5. *For any function $\boldsymbol{z}_h \in \mathrm{H}_h$ and any element $K \in \mathcal{T}_h$, one has*

$$\int_K (\boldsymbol{z}_h - \mathcal{H}_h \boldsymbol{z}_h) = 0.$$

Theorem 2.3.6. *Let $((\boldsymbol{\omega}, p), \boldsymbol{u}) \in (\mathrm{Z} \times \mathrm{Q}) \times \mathrm{H}$ and $((\hat{\boldsymbol{\omega}}_h, \hat{p}_h), \hat{\boldsymbol{u}}_h) \in (\mathrm{Z}_h \times \mathrm{Q}_h) \times \mathrm{H}_h$ be the solutions of the continuous and discrete problems* (2.1.9)-(2.1.10) *and* (2.3.2)-(2.3.3)*, respectively. Then there exists a constant $C > 0$, independent of h and η, such that*

$$\|\boldsymbol{u} - \hat{\boldsymbol{u}}_h\|_{0,\Omega} \leq C h^2 \big(\|\boldsymbol{\omega}\|_{1,\Omega} + \|p\|_{1,\Omega} + ||\boldsymbol{f}||_{1,\Omega} + \|\boldsymbol{u}\|_{2,\Omega} \big).$$

Proof. Let $((\boldsymbol{\omega}_h, p_h), \boldsymbol{u}_h)$ and $((\hat{\boldsymbol{\omega}}_h, \hat{p}_h), \hat{\boldsymbol{u}}_h)$ denote the solutions to the FE formulation (2.2.1)-(2.2.2) and the FVE formulation (2.3.2)-(2.3.3), respectively. Following the argument in [3], we resort to a duality argument involving the following problem: Find $((\boldsymbol{\xi}, \phi), \boldsymbol{z}) \in (\mathrm{Z} \times \mathrm{Q}) \times \mathrm{H}$ such that

$$a((\boldsymbol{\theta}, q), (\boldsymbol{\xi}, \phi)) + b((\boldsymbol{\theta}, q), \boldsymbol{z}) = 0 \qquad \forall (\boldsymbol{\theta}, q) \in \mathrm{Z} \times \mathrm{Q}, \tag{2.3.9}$$

$$b((\boldsymbol{\xi}, \phi), \boldsymbol{v}) = \int_\Omega (\boldsymbol{u} - \hat{\boldsymbol{u}}_h) \cdot \boldsymbol{v} \qquad \forall \boldsymbol{v} \in \mathrm{H}. \tag{2.3.10}$$

We assume that the unique solution to (2.3.9)-(2.3.10) satisfies an additional regularity. More precisely, there exists a constant $\tilde{C} > 0$, independent of h and η, such that

$$\|(\boldsymbol{\xi}, \phi)\|_{1,\Omega} + \|\boldsymbol{z}\|_{2,\Omega} \leq \tilde{C} \|\boldsymbol{u} - \hat{\boldsymbol{u}}_h\|_{0,\Omega}. \tag{2.3.11}$$

In addition, we mention that the proof from [3] readily implies that, for all $(\boldsymbol{\xi}_h, \phi_h) \in \mathrm{Z}_h \times \mathrm{Q}_h$, we can write

$$\|\boldsymbol{u} - \hat{\boldsymbol{u}}_h\|_{0,\Omega}^2 = b((\boldsymbol{\xi} - \boldsymbol{\xi}_h, \phi - \phi_h), \boldsymbol{u} - \hat{\boldsymbol{u}}_h) - a((\boldsymbol{\omega} - \hat{\boldsymbol{\omega}}_h, p - \hat{p}_h), (\boldsymbol{\xi}_h, \phi_h)), \tag{2.3.12}$$

as well as

$$\begin{aligned} a((\boldsymbol{\omega} - \hat{\boldsymbol{\omega}}_h, p - \hat{p}_h), (\boldsymbol{\xi}_h, \phi_h)) = & - b((\boldsymbol{\omega} - \hat{\boldsymbol{\omega}}_h, p - \hat{p}_h), \boldsymbol{z}) \\ & - a((\boldsymbol{\omega} - \hat{\boldsymbol{\omega}}_h, p - \hat{p}_h), (\boldsymbol{\xi} - \boldsymbol{\xi}_h, \phi - \phi_h)). \end{aligned} \tag{2.3.13}$$

Next, we employ identity (2.3.12) in order to arrive at the envisioned error bound. Using (2.1.10) and (2.2.2) in combination with (2.3.3) and Lemma 2.3.1, and by adding and subtracting $B((\hat{\boldsymbol{\omega}}_h, \hat{p}_h), \boldsymbol{z}_h)$ we obtain

$$b((\hat{\boldsymbol{e}}_{\boldsymbol{\omega}}, \hat{e}_p), \boldsymbol{z}_h) - b((\hat{\boldsymbol{r}}_{\boldsymbol{\omega}}, \hat{r}_p), \boldsymbol{z}_h) = b((\boldsymbol{\omega}, p), \boldsymbol{z}_h) - b((\boldsymbol{\omega}_h, p_h), \boldsymbol{z}_h) = 0, \tag{2.3.14}$$

where we have introduced

$$\hat{\boldsymbol{e}}_{\boldsymbol{\omega}} := \boldsymbol{\omega} - \hat{\boldsymbol{\omega}}_h, \quad \hat{e}_p := p - \hat{p}_h, \quad \hat{\boldsymbol{e}}_{\boldsymbol{u}} := \boldsymbol{u} - \hat{\boldsymbol{u}}_h,$$
$$\hat{\boldsymbol{r}}_{\boldsymbol{\omega}} := \boldsymbol{\omega}_h - \hat{\boldsymbol{\omega}}_h, \quad \hat{r}_p := p_h - \hat{p}_h, \quad \hat{\boldsymbol{r}}_{\boldsymbol{u}} := \boldsymbol{u}_h - \hat{\boldsymbol{u}}_h.$$

Consequently, identity (2.3.13) in combination with identity (2.3.14) implies that

$$\begin{aligned} a((\hat{\boldsymbol{e}}_{\boldsymbol{\omega}}, \hat{e}_p), (\boldsymbol{\xi}_h, \phi_h)) &= -b((\hat{\boldsymbol{e}}_{\boldsymbol{\omega}}, \hat{e}_p), \boldsymbol{z}) - a((\hat{\boldsymbol{e}}_{\boldsymbol{\omega}}, \hat{e}_p), (\boldsymbol{\xi} - \boldsymbol{\xi}_h, \phi - \phi_h)) \\ &= -b((\hat{\boldsymbol{e}}_{\boldsymbol{\omega}}, \hat{e}_p), \boldsymbol{z} - \boldsymbol{z}_h) - a((\hat{\boldsymbol{e}}_{\boldsymbol{\omega}}, \hat{e}_p), (\boldsymbol{\xi} - \boldsymbol{\xi}_h, \phi - \phi_h)) \\ &\quad - b((\hat{\boldsymbol{r}}_{\boldsymbol{\omega}}, \hat{r}_p), \boldsymbol{z}_h), \end{aligned} \tag{2.3.15}$$

which holds for all $\boldsymbol{z}_h \in \mathrm{H}_h$. In particular, we take the Lagrange interpolant of $\boldsymbol{z}$, denoted by $\boldsymbol{z}_I \in \mathrm{H}_h$. Moreover, for every element $K \in \mathcal{T}_h$, we use $\overline{\boldsymbol{f}}_K$ to denote the average of $\boldsymbol{f}$ on such element. Then, by the virtue of Lemma 2.3.5 and by integrating over elements $K \in \mathcal{T}_h$ instead of over the control volumes $K_j^\star \in \mathcal{T}_h^\star$ and the domain Ω, we find that for some constant $C_0 > 0$, independent of h and η,

$$\begin{aligned} |b((\hat{\boldsymbol{\omega}}_h - \boldsymbol{\omega}_h, \hat{p}_h - p_h), \boldsymbol{z}_I)| &\leq \left| \sum_{j=1}^{|\mathcal{N}_h|} \int_{K_j^\star} \boldsymbol{f} \cdot \mathcal{H}_h \boldsymbol{z}_I - \int_\Omega \boldsymbol{f} \cdot \boldsymbol{z}_I \right| \\ &= \left| \sum_{K \in \mathcal{T}_h} \int_K \boldsymbol{f} \cdot (\mathcal{H}_h \boldsymbol{z}_I - \boldsymbol{z}_I) \right| \\ &= \left| \sum_{K \in \mathcal{T}_h} \int_K (\boldsymbol{f} - \overline{\boldsymbol{f}}_K) \cdot (\mathcal{H}_h \boldsymbol{z}_I - \boldsymbol{z}_I) \right| \\ &= \sum_{K \in \mathcal{T}_h} ||\boldsymbol{f} - \overline{\boldsymbol{f}}_K||_{0,K} ||\mathcal{H}_h \boldsymbol{z}_I - \boldsymbol{z}_I||_{0,K} \\ &\leq C_0 h^2 ||\boldsymbol{f}||_{1,\Omega} |\boldsymbol{z}_I|_{1,\Omega}, \end{aligned}$$

where the last step follows from the interpolation bound satisfied by the transfer operator $\mathcal{H}_h(\cdot)$ in combination with the inequality $||\boldsymbol{f} - \overline{\boldsymbol{f}}_K||_{0,K} \leq Ch||\boldsymbol{f}||_{1,K}$. Furthermore, by one application of the triangle inequality it follows that for some constants $C_1, C_2 > 0$, independent of h and η, one has

$$\begin{aligned} |b((\hat{\boldsymbol{\omega}}_h - \boldsymbol{\omega}_h, \hat{p}_h - p_h), \boldsymbol{z}_I)| &\leq C_0 h^2 ||\boldsymbol{f}||_{1,\Omega} \big(|\boldsymbol{z}_I - \boldsymbol{z}|_{1,\Omega} + |\boldsymbol{z}|_{1,\Omega} \big) \\ &\leq C_1 h^2 ||\boldsymbol{f}||_{1,\Omega} \big(||\boldsymbol{z}||_{2,\Omega} + |\boldsymbol{z}|_{1,\Omega} \big) \\ &\leq C_2 h^2 ||\boldsymbol{f}||_{1,\Omega} ||\boldsymbol{u} - \hat{\boldsymbol{u}}_h||_{0,\Omega}, \end{aligned} \tag{2.3.16}$$

where the last two inequalities follow from the classical error estimates for the Lagrange interpolants and the additional regularity requirement. Hence, by taking the

L^2-projections for $\boldsymbol{\xi}$ and ϕ, using the classical error estimates for the involved spaces, and employing identity (2.3.15) in combination with inequality (2.3.16), we straightforwardly deduce that for some constant $C > 0$, independent of h and η, one has

$$\|\boldsymbol{u} - \hat{\boldsymbol{u}}_h\|_{0,\Omega} \leq Ch^2\big(\|\boldsymbol{\omega}\|_{1,\Omega} + \|p\|_{1,\Omega} + ||\boldsymbol{f}||_{1,\Omega} + \|\boldsymbol{u}\|_{2,\Omega}\big),$$

thus completing the proof. □

2.4 Numerical tests

In this section, the convergence and robustness of the proposed methods is illustrated via a set of insightful computational tests. We remark that in the two-dimensional case, the computational cost of the proposed FE and FVE methods in their lowest order configuration is lower than, for instance, the MINI element for displacement-pressure formulations (see, for instance, [3] and [5]). These and other features turn the proposed discretisations into very appealing methods.

2.4.1 Manufactured solution

For our first computational example, we conduct a convergence test using a sequence of successively refined uniform partitions of the elastic domain $\Omega = (0,1)^2$. We arbitrarily choose the Lamé parameters as $\mu = 50$, $\lambda = 5000$, such that $\eta = 0.0099$. This example focuses on the pure-Dirichlet problem (2.1.4)-(2.1.7), where we propose the following closed-form solutions

$$\boldsymbol{u} = \begin{pmatrix} x(1-x)\cos(\pi x)\sin(2\pi y) \\ \sin(\pi x)\cos(\pi y)y^2(1-y) \end{pmatrix}, \quad \boldsymbol{\omega} = \sqrt{\eta}\,\mathbf{curl}\,\boldsymbol{u}, \quad p = -\operatorname{div}\boldsymbol{u},$$

satisfying the homogeneous Dirichlet datum, and where the forcing term $\boldsymbol{f}$ is constructed using these smooth functions and the linear momentum equation. The convergence study is performed for the mixed FVE method (2.3.2)-(2.3.3) and for the Galerkin schemes (2.2.1)-(2.2.2) of order $k = 1$ and $k = 2$. For a generic scalar or vectorial field $\boldsymbol{v}$, on each nested mesh we will denote computed errors and experimental convergence rates as, for $i = 0, \mathrm{H}$,

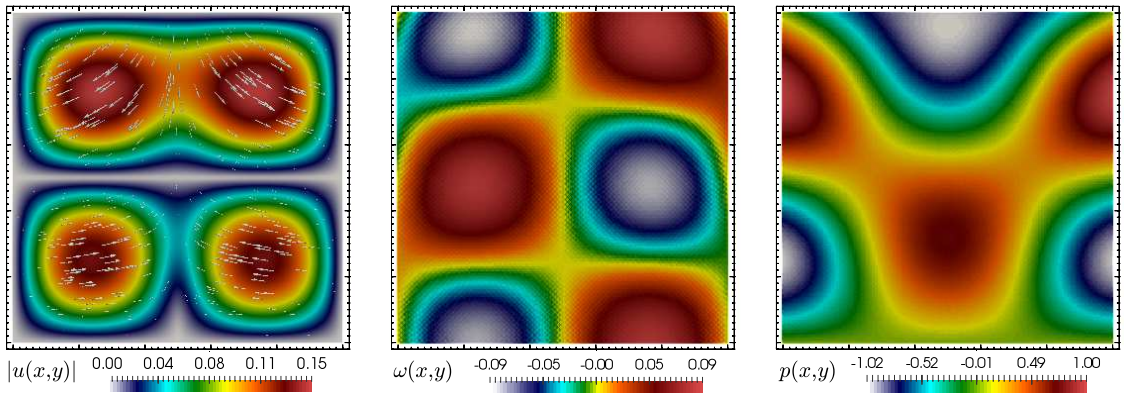
$$e_0(\boldsymbol{v}) = \|\boldsymbol{v} - \boldsymbol{v}_h\|_{0,\Omega}, \quad e_{\mathrm{H}}(\boldsymbol{v}) = \|\boldsymbol{v} - \boldsymbol{v}_h\|_{\mathrm{H}}, \quad r_i(\boldsymbol{v}) = \log\left(\frac{e_i(\boldsymbol{v})}{\widehat{e}_i(\boldsymbol{v})}\right)[\log(h/\widehat{h})]^{-1},$$

where $e_i, \widehat{e}_i$ stand for errors generated by methods defined on meshes with mesh sizes $h, \widehat{h}$, respectively, and we recall that $\|\cdot\|_{\mathrm{H}}$ denotes the $\eta-$dependent norm. These

D.o.f.	h	$e_0(\boldsymbol{u})$	$r_0(\boldsymbol{u})$	$e_{\mathrm{H}}(\boldsymbol{u})$	$r_{\mathrm{H}}(\boldsymbol{u})$	$e_0(\boldsymbol{\omega})$	$r_0(\boldsymbol{\omega})$	$e_0(p)$	$r_0(p)$
Mixed finite volume element with $k = 1$									
34	0.7071	0.079453	–	0.35510	–	0.04578	–	0.42279	–
68	0.4714	0.063802	0.681	0.27558	0.573	0.04244	0.486	0.36318	0.476
172	0.2828	0.023599	1.540	0.21289	1.021	0.03270	0.638	0.20001	1.034
524	0.1571	0.009612	1.960	0.10311	1.065	0.02033	0.798	0.10700	1.074
1804	0.0831	0.001842	2.071	0.05018	1.073	0.01143	0.921	0.05688	1.079
6668	0.0428	0.000819	2.071	0.02001	1.048	0.00613	0.980	0.02730	1.051
25612	0.0217	0.000079	2.053	0.01295	1.027	0.00304	1.001	0.01258	1.029
100364	0.0109	0.000035	2.027	0.00645	1.015	0.00123	1.002	0.00693	1.016
Mixed finite element with $k = 1$									
34	0.7071	0.082091	–	0.44686	–	0.05045	–	0.44405	–
68	0.4714	0.069138	0.623	0.38582	0.562	0.04647	0.202	0.38301	0.464
172	0.2828	0.033144	1.439	0.22359	1.067	0.03639	0.478	0.22061	1.079
524	0.1571	0.009937	2.049	0.11873	1.076	0.02256	0.813	0.11656	1.085
1804	0.0831	0.002630	2.090	0.06008	1.070	0.01247	0.931	0.05877	1.076
6668	0.0428	0.000661	2.082	0.03000	1.047	0.00649	0.984	0.02929	1.050
25612	0.0217	0.000163	2.062	0.01495	1.027	0.00329	1.002	0.01458	1.028
100364	0.0109	0.000041	2.032	0.00745	1.014	0.00165	1.002	0.00727	1.015
Mixed finite element with $k = 2$									
98	0.7071	0.042104	–	0.19409	–	0.06606	–	0.18250	–
206	0.4714	0.020403	1.788	0.10392	1.549	0.03889	1.903	0.09637	1.577
542	0.2828	0.005284	2.647	0.03914	1.918	0.01338	2.027	0.03678	1.885
1694	0.1571	0.000945	2.923	0.01239	1.956	0.00418	1.973	0.01166	1.956
5918	0.0831	0.000138	3.027	0.00349	1.995	0.00114	2.033	0.00330	1.982
22046	0.0428	0.000018	3.042	0.00092	2.002	0.00029	2.023	0.00087	1.994
85022	0.0217	0.000003	3.018	0.00023	2.002	0.00008	2.052	0.00022	1.992
333854	0.0109	0.000001	3.039	0.00006	2.001	0.00002	2.013	0.00006	1.998

Table 2.1: Test 1A. Experimental convergence rates for the mixed Petrov-Galerkin (cf. (2.3.2)-(2.3.3)) and Galerkin (cf. (2.2.1)-(2.2.2)) methods, using $\mu = 50$, $\lambda = 5000$.

errors are tabulated by the corresponding number of degrees of freedom in Table 2.1. Apart from the displacement error measured in the L^2-norm (whose error decays with order h^{k+1} as anticipated by Theorem 2.2.5), each individual error exhibits an



(a) (b) (c)

Figure 2.2: Test 1A (scheme accuracy). Approximate displacement magnitude (a), rotation scalar field (b), and pressure (c); obtained with the lowest-order FVE method.

$O(h^k)$ rate of convergence, as expected from the a priori error estimates stated in Theorems 2.2.4 and 2.3.4. Moreover, the errors produced by the first two methods practically coincide. This is due to the fact that they only differ in the RHS assembly. For reference, in Figure 2.2 we depict approximate solutions generated with the lowest-order FVE scheme. Analogous numerical studies using mixed boundary conditions (see the formulation in Appendix A) produce the same optimal convergence behaviour observed in the pure Dirichlet case.

In addition, these methods are robust with respect to the model parameters, which we confirm by a series of tests where we fix a Young modulus $E = 10000$, we vary the Poisson ratio ν, and measure the errors produced by the first order finite element method on an unstructured mesh of 33282 elements and 100364 D.o.f. (see the first block in Table 2.2). Furthermore, we also construct a different smooth forcing term $\boldsymbol{f} = (10^5 \cos(x), 10^5 \cos(y))^T$, independent of the model parameters, solve the discrete problem for relatively large Lamé constants (we recall that $\lambda = E\nu/[(1+\nu)(1-2\nu)]$ and $\mu = E/(2+2\nu)$), and tabulate the obtained norms of the approximate solutions in the bottom block of Table 2.2. We evidence stable and robust computations even in the regimes of near incompressibility. All linear systems in this example were solved with the Unsymmetric Multi-Frontal sparse LU factorisation (UMFPACK). In addition, we remark that the family of mixed FE schemes and the lowest-order FVE method were implemented using the computational platforms FEniCS and FreeFem++, respectively.

ν	λ	μ	η	$e_0(\boldsymbol{u})$	$e_{\mathrm{H}}(\boldsymbol{u})$	$e_0(\boldsymbol{\omega})$	$e_0(p)$
0.33333	7500	3750.00	0.33333	0.000040	0.01127	0.00807	0.00787
0.40000	14285.71	3571.43	0.20000	0.000038	0.01004	0.00658	0.00758
0.45000	31034.48	3448.28	0.10000	0.000037	0.00886	0.00491	0.00738
0.49000	164429.53	3355.71	0.02000	0.000035	0.00763	0.00232	0.00727
0.49900	1664442.96	3335.56	0.00200	0.000036	0.00730	0.00075	0.00726
0.49990	16664444.30	3333.56	0.00020	0.000039	0.00727	0.00025	0.00726
0.49999	166664444.43	3333.36	0.00002	0.000040	0.00726	0.00008	0.00726
ν	λ	μ	η	$\|\boldsymbol{u}_h\|_{0,\Omega}$	$\|\boldsymbol{u}_h\|_{\mathrm{H}}$	$\|\boldsymbol{\omega}_h\|_{0,\Omega}$	$\|p_h\|_{0,\Omega}$
0.33333	7500	3750.00	0.33333	6.346721	23.1147	13.0760	19.0607
0.40000	14285.71	3571.43	0.20000	7.749333	26.3285	12.6964	23.0649
0.45000	31034.48	3448.28	0.10000	9.346012	29.7836	11.1665	27.6111
0.49000	164429.53	3355.71	0.02000	11.28440	33.7542	6.26612	33.1675
0.49900	1664442.96	3335.56	0.00200	11.85713	34.8893	2.10631	34.8257
0.49990	16664444.30	3333.56	0.00020	11.91832	35.0094	0.67041	35.0030
0.49999	166664444.43	3333.36	0.00002	11.92467	35.0215	0.21214	35.0209

Table 2.2: Test 1B. Accuracy (top rows) and robustness with respect to the Lamé constants (bottom rows) studied for two different benchmark tests, approximated with the lowest-order mixed finite element method.

2.4.2 Cantilevered beam

For the next computational example, we consider the displacement-rotation-pressure patterns of a rectangular beam (with length $L = 10$ and height $l = 2$) subjected to a couple (that is, a prescribed traction $(f(1-y), 0)^T$, with $f = 200$) at one end, as shown in Figure 2.3(a). We assume that the origin O is fully fixed and that the horizontal displacement is zero along the left edge of the domain Ω. Furthermore, on the remainder of the boundary we consider zero normal stresses incorporated through the bilinear form $c(\cdot,\cdot)$ (see (A.0.3) in Appendix A) and we set up a zero body force $\boldsymbol{f} = \boldsymbol{0}$. The availability of an exact solution (cf. [48])

$$\boldsymbol{u}_1(x,y) = \frac{2f(1-\nu)^2}{El} x\left(\frac{l}{2} - y\right), \quad \boldsymbol{u}_2(x,y) = \frac{f(1-\nu)^2}{El}\left(x^2 + \frac{\nu}{1+\nu} y(y-l)\right), \tag{2.4.1}$$

makes that this problem is frequently used as a benchmark. In Figure 2.3, we illustrate the components of the displacement, the rotation and the pressure computed on a

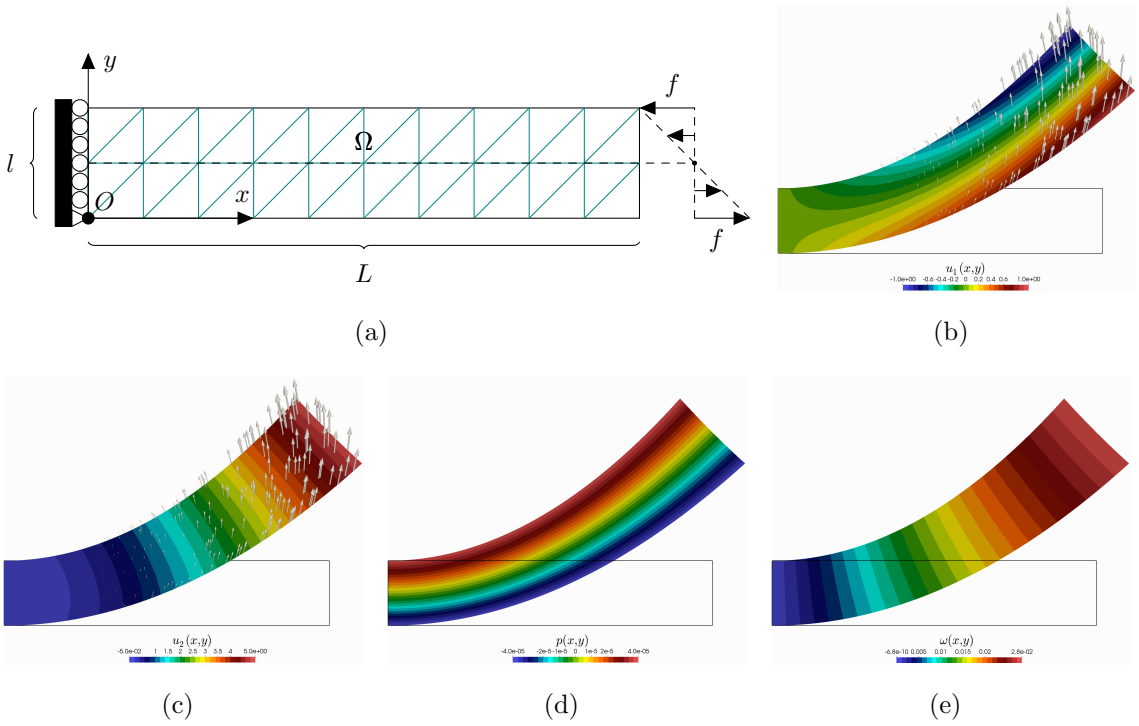
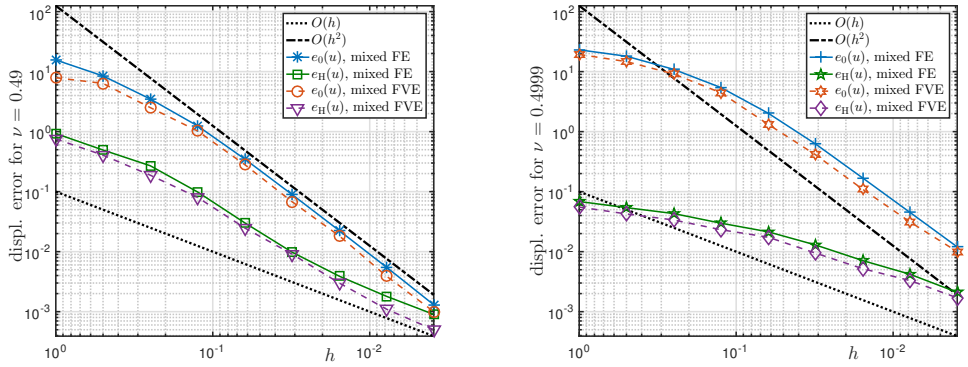


Figure 2.3: Test 2. Rectangular beam fixed at the origin O and with zero horizontal displacement along the left lateral edge, subjected to bending due to a couple at one end. Sketch of the domain configuration with a coarse structured mesh and the imposed boundary conditions (a), displacement components (b,c), pressure distribution (d), and rotation (e); all computed with a second-order mixed FE method on a mesh of 5120 triangular elements.

mesh of 5120 triangular elements, utilising the mixed FE method of order $k = 2$, where the rectangular beam we consider has the following material properties: Young's modulus $E = 1500$, Poisson's ratio $\nu = 0.49$, Lamé constants $\lambda = 24664.4$ and $\mu = 503.356$, such that the model parameter equals $\eta = 0.02$. In addition, we conduct several tests for the lowest-order mixed FE and FVE methods for different mesh sizes and report on the error with respect to the analytic solution (2.4.1). In particular, Figure 2.4 displays the displacement error in the H$-$norm and the L$^2-$norm versus the mesh size, for the FE and FVE schemes, and for two values of the Poisson ratio $\nu = 0.49, 0.4999$, fixing the Young's modulus $E = 1500$. Observe that these results are in agreement with the theoretical results obtained in Sections 2.2-2.3. In addition, although this is in general not true, we mention that the second-order FE scheme ensures extremely rapid convergence (explained by the regularity of the true solution (2.4.1)). For $\nu = 0.4999$, optimal convergence is recovered for finer meshes.



(a) (b)

Figure 2.4: Test 2. Convergence for the displacement approximation using the first order mixed FE and FVE schemes, for $\nu = 0.49$ (a) and $\nu = 0.4999$ (b); fixing the Young's modulus $E = 1500$.

To close this chapter, we highlight that other insightful numerical results (including the approximation of the displacement-rotation-pressure patterns of a 3D cantilevered beam and the well-known Cook's membrane benchmark test) can be found in our unpublished paper [3], where we have also included a comparative study with numerical results obtained with other finite element formulations available from the literature applied to the original linear elasticity equations (2.1.1), such as a classical pure-displacement formulation discretised with piecewise continuous elements of degree $k \geq 1$, the Taylor-Hood finite element for a displacement-pressure formulation, and the MINI-element [5]. These schemes have comparable complexity and we have purposely not included mixed methods based on stress or pseudo-stress formulations, since the associated computational costs would be much higher.

Chapter 3

A second-order hybrid FVE method for poroelasticity

In this chapter, we develop a second-order mixed hybrid FVE method for solving a three-field poroelasticity problem. The main advantages of the proposed Petrov-Galerkin scheme are its second order convergence rate, its local conservativeness, its robustness with respect to the Lamé modulus of dilation, λ, and that the produced solutions are locking-free. A further advantage of the proposed method is the simplistic construction of the dual meshes based on [25–27], which is essentially the same as for standard linear FVE methods.

3.1 Model equations and FVE discretisation

Let us consider Biot's consolidation model for a homogeneous isotropic porous solid saturated with an interstitial fluid. Let $\Omega \subset \mathbb{R}^2$ denote the physical domain occupied by both the solid and fluid phases, which is assumed to be bounded, simply connected and Lipschitz. Given some time interval $[0, T]$, we let $p : \Omega \times (0, T] \to \mathbb{R}$ denote the pore pressure of the fluid, $\boldsymbol{u} : \Omega \times (0, T] \to \mathbb{R}^2$ denote the displacement of the porous linear elastic skeleton, $\boldsymbol{f} : \Omega \times (0, T] \to \mathbb{R}^2$ denote an imposed body force, and write $s : \Omega \times (0, T] \to \mathbb{R}$ for the volumetric fluid source or sink term. Then, the governing equations can be summarised as (cf. [13])

$$\partial_t\big(c_0 p + \alpha \operatorname{div}(\boldsymbol{u})\big) - \frac{1}{\eta} \operatorname{div}\big(\kappa(\nabla p - \rho \boldsymbol{g})\big) = s \qquad \text{in } \Omega, \tag{3.1.1}$$

$$-\mathbf{div}(\boldsymbol{\sigma}) = \boldsymbol{f} \qquad \text{in } \Omega, \tag{3.1.2}$$

where $\boldsymbol{\sigma} = \lambda \operatorname{div}(\boldsymbol{u})\mathbf{I} + 2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}) - p\mathbf{I}$ denotes the total Cauchy solid stress, κ denotes the permeability of the isotropic porous solid (satisfying $0 < \kappa_1 \leq \kappa(\boldsymbol{x}) \leq \kappa_2 < \infty$, for every $\boldsymbol{x} \in \Omega$), λ and μ are the first and second Lamé constants of the solid, $c_0 > 0$ is the constrained specific storage coefficient, α denotes the Biot-Willis parameter, $\boldsymbol{g}$ is the gravity acceleration aligned with the vertical direction, and $\eta > 0$, $\rho > 0$ represent the viscosity and density of the pore fluid, respectively.

In order to demonstrate the main ideas of the formulation introduced in this section and simplify the well-posedness and convergence analysis, we proceed to restrict the presentation to the steady case consisting of equation (3.1.2) in combination with the relation

$$c_0 p + \alpha \operatorname{div}(\boldsymbol{u}) - \frac{1}{\eta} \operatorname{div}\big(\kappa(\nabla p - \rho \boldsymbol{g})\big) = s, \tag{3.1.3}$$

which arises from a time discretisation of equation (3.1.1), with an abuse of notation regarding the volumetric fluid source (or sink) term s.

Following the pseudostress-based mixed formulation for linear elasticity introduced in [38], we observe that the identity $\mathbf{div}(\boldsymbol{\sigma}) = \mu\Delta\boldsymbol{u} + (\lambda + \mu)\nabla \operatorname{div}(\boldsymbol{u}) - \nabla p$ allows us to recast equation (3.1.2) in the form

$$-\mu\Delta\boldsymbol{u} - (\lambda + \mu)\nabla \operatorname{div}(\boldsymbol{u}) + \nabla p = \boldsymbol{f},$$

where we effectively replace the divergence of the total Cauchy solid stress with the divergence of the non-symmetric total pseudo-stress tensor $\boldsymbol{\tau} := \mu\nabla\boldsymbol{u} + (\lambda + \mu)\operatorname{div}(\boldsymbol{u}) - p\mathbf{I}$. This form for the balance law has also been used for poroelasticity equations in [11, 79]. Furthermore, we introduce the following auxiliary parameter as an additional unknown:

$$\phi := p - (\lambda + \mu)\operatorname{div}(\boldsymbol{u}), \tag{3.1.4}$$

which can be regarded as a pseudo total pressure or volumetric stress. This is slightly different from the total pressure introduced in [62]. The main advantage of this modification is that well-posedness of the envisioned second-order FVE scheme can be established. If we instead consider the original setting from [62], then we cannot ensure the coercivity of one particular bilinear operator arising in the mixed hybrid FVE formulation of interest. We refer the reader to Section 3.1 and Appendix B for a more in-depth discussion regarding this technical complication.

Next, exploiting equation (3.1.4), we readily observe that equation (3.1.3) and

equation (3.1.2) are fully equivalent to, respectively,

$$\left(c_0 + \frac{\alpha}{\lambda+\mu}\right)p - \frac{\alpha}{\lambda+\mu}\phi - \frac{1}{\eta}\,\mathrm{div}\left(\kappa(\nabla p - \rho\boldsymbol{g})\right) = s \qquad \text{in } \Omega, \tag{3.1.5}$$

$$-\mu\Delta\boldsymbol{u} + \mathbf{div}(\phi\mathbf{I}) = \boldsymbol{f} \qquad \text{in } \Omega, \tag{3.1.6}$$

which in combination with equation (3.1.4) constitutes the three-field formulation of interest.

To complete the system of equations (3.1.4)-(3.1.6), we need to prescribe suitable boundary conditions. Following Oyarzúa et al. [62], we assume that the boundary $\partial\Omega$ of the domain Ω is split into a part Γ_p on which the fluid pressure is specified and a part $\Gamma_{\boldsymbol{u}}$ where displacements are imposed (such that $\partial\Omega = \bar{\Gamma}_p \cup \bar{\Gamma}_{\boldsymbol{u}}$ and $\Gamma_p \cap \Gamma_{\boldsymbol{u}} = \emptyset$, where it is assumed that both boundary segments have strictly positive measure). Under this assumption, the set of boundary conditions assumes the form

$$p = 0,\ \boldsymbol{\sigma}\boldsymbol{n} = \mathbf{0} \text{ on } \Gamma_p, \quad \text{and} \quad \boldsymbol{u} = \mathbf{0},\ \frac{\kappa}{\eta}(\nabla p - \rho\boldsymbol{g})\cdot\boldsymbol{n} = 0 \text{ on } \Gamma_{\boldsymbol{u}}. \tag{3.1.7}$$

The stability and unique solvability of the continuous problem (3.1.4)-(3.1.6) was addressed in [62], using Fredholm's alternative theory for compact operators. The analysis can be straightforwardly extended to non-homogeneous boundary conditions by employing classical lifting techniques.

3.1.1 Mixed hybrid FVE discretisation

We proceed to establish a high order mixed hybrid FVE scheme for the poroelasticity equations (3.1.4)-(3.1.6).

First, taking into account the homogeneous treatment of the boundary conditions, we introduce the following functional spaces:

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

where we endow H with the natural seminorm $|\cdot|_{1,\Omega}$ and endow Q and Z with the norms $||\cdot||_{1,\Omega}$ and $||\cdot||_{0,\Omega}$, respectively. Furthermore, we consider the following discrete trial function spaces:

$$\mathrm{H}_h := \{\boldsymbol{u}_h \in \mathrm{H} \cap \mathrm{C}(\Omega)^2 : \boldsymbol{u}_h|_K \in \mathcal{P}_2(K)^2 \quad \forall K \in \mathcal{T}_h,\ \boldsymbol{u}_h = \mathbf{0} \text{ on } \Gamma_{\boldsymbol{u}}\},$$
$$\mathrm{Q}_h := \{p_h \in \mathrm{Q} \cap \mathrm{C}(\Omega) : p_h|_K \in \mathcal{P}_2(K) \quad \forall K \in \mathcal{T}_h,\ p_h = 0 \text{ on } \Gamma_p\},$$
$$\mathrm{Z}_h := \{\phi_h \in \mathrm{Z} \cap \mathrm{C}(\Omega) : \phi_h|_K \in \mathcal{P}_1(K) \quad \forall K \in \mathcal{T}_h\},$$

where we highlight that no additional test spaces are introduced for the approximation of the auxiliary total pressure ϕ and the pore pressure of the fluid p.

In order to construct the associated test function space for the displacement, we proceed to construct a dual partition of the domain. Recall that $\mathscr{T} := \{\mathcal{T}_h\}_{h>0}$ denotes a shape regular family of quasi-uniform triangulations of the domain Ω, consisting of triangles K of diameter h_K, where $h := \max\{h_K : K \in \mathcal{T}_h\}$ denotes the mesh parameter. In particular, we assume there exist thresholds $\theta_l > 0^\circ$ and $\theta_u < 180^\circ$, independent of h, such that

$$\theta_{\min,K} \geq \theta_l \quad \text{and} \quad \theta_{\max,K} \leq \theta_u \qquad \forall K \in \mathcal{T}_h \in \mathscr{T},$$

where $\theta_{\min,K}$ and $\theta_{\max,K}$ denote the smallest and largest interior angles of the element K, respectively. Equivalently, we can assume there exist $\alpha > 1$ and $\beta > 0$, independent of h, such that

$$h_k \leq \alpha \rho_k \quad \text{and} \quad \beta h^2 \leq |K| \leq h^2 \qquad \forall K \in \mathcal{T}_h \in \mathscr{T},$$

where we recall that $|K|$ denotes the area of the element K, and

$$\rho_K := \sup\{\text{diam(C)} : \text{C is a disk contained in } K\}.$$

In addition to the primal mesh $\mathcal{T}_h$, we introduce another tessellation of the domain Ω, denoted by $\mathcal{T}_h^\star$ and referred to as the dual mesh. For each $\mathcal{T}_h \in \mathscr{T}$, the corresponding dual mesh $\mathcal{T}_h^\star$ is constructed by joining the barycenters b_K of each element $K \in \mathcal{T}_h$ with the midpoints of each edge $F \subset \partial K$, forming three quadrilaterals Q_z, for z in the set of vertices $K \cap \mathcal{N}_h$. Next, we associate a control volume $K_j^\star$ to each vertex $s_j \in \mathcal{N}_h$, which consists of the union of the quadrilaterals Q_{s_j} sharing the vertex s_j.

In order to construct the test function space for the displacement, observe that the trial space H_h can be decomposed as (cf. [5,7,8])

$$\mathrm{H}_h = \mathrm{V}_{1,h} \oplus \mathrm{W}_{2,h}, \tag{3.1.8}$$

where $\mathrm{V}_{1,h}$ denotes the linear finite element space spanned by the canonical piecewise linear vectorial nodal basis functions $\boldsymbol{\phi}_i$, for $i = 1, \cdots, 2|\mathcal{N}_h|$, and the discrete space $\mathrm{W}_{2,h}$ is spanned by strictly quadratic vectorial basis functions $\boldsymbol{\psi}_i$, for $i = 1, \cdots, 2|\mathcal{N}_h|$. Based on the decomposition (3.1.8), we introduce the following test function space for the displacement associated with the dual partition of the domain:

$$\mathrm{H}_h^\star := \mathrm{V}_{1,h}^\star \oplus \mathrm{W}_{2,h},$$

where the linear space $\mathrm{V}_{1,h}^{\star}$ is spanned by the vectorial characteristic functions $\{\boldsymbol{\chi}_j\}_j$ of the control volumes $K_j^{\star} \in \mathcal{T}_h^{*}$. Observe that the inclusion of these characteristic functions in the test space $\mathrm{H}_h^{\star}$ ensures that conservation laws are preserved locally on each control volume $K_j^{\star} \in \mathcal{T}_h^{\star}$. Effectively, this hybrid characterisation of the test function space can be regarded as a generalisation of both high order FE methods and standard linear FVE methods. Namely, the discrete function space $\mathrm{V}_{1,h}^{\star}$ leads to a global conservation property by linear composition of the control volumes, while the discrete space $\mathrm{W}_{2,h}$ could give rise to high order approximations, as in the corresponding high order FE method.

In addition, we introduce a transfer operator $\mathcal{H}_h : \mathrm{H}_h \to \mathrm{H}_h^{\star}$ that relates the primal and conforming dual meshes, which is defined by

$$\boldsymbol{u}_h(\boldsymbol{x}) = \sum_{j=1}^{2|\mathcal{N}_h|} \boldsymbol{u}_h(s_j)\boldsymbol{\phi}_j(\boldsymbol{x}) + \boldsymbol{u}_h^2(\boldsymbol{x}) \mapsto \mathcal{H}_h\boldsymbol{u}_h(\boldsymbol{x}) = \sum_{j=1}^{2|\mathcal{N}_h|} \boldsymbol{u}_h(s_j)\boldsymbol{\chi}_j(\boldsymbol{x}) + \boldsymbol{u}_h^2(\boldsymbol{x}),$$

where the displacement $\boldsymbol{u}_h \in \mathrm{H}_h$ is written as a linear combination of the basis functions $\{\boldsymbol{\phi}_j\}_j$ and the strictly quadratic component $\boldsymbol{u}_h^2 \in \mathrm{W}_{2,h}$. For any $\boldsymbol{v}_h \in \mathrm{H}_h$, the transfer operator $\mathcal{H}_h(\cdot)$ satisfies the identity (cf. [28])

$$\int_K (\boldsymbol{v}_h - \mathcal{H}_h\boldsymbol{v}_h) = 0,$$

for every triangle $K \in \mathcal{T}_h \in \mathscr{T}$.

The discrete FVE formulation corresponding to the model equations (3.1.4)-(3.1.6) is obtained by multiplying (3.1.6) by $\boldsymbol{v}_h^{\star} \in \mathrm{H}_h^{\star}$ and integrating by parts over each $K_j^{\star} \in \mathcal{T}_h^{\star}$, multiplying (3.1.5) by $q_h \in \mathrm{Q}_h$ and integrating by parts over each $K \in \mathcal{T}_h$, and multiplying (3.1.4) by $\psi_h \in \mathrm{Z}_h$ and integrating by parts over each $K \in \mathcal{T}_h$. This results in the following Petrov-Galerkin formulation: Find $\boldsymbol{u}_h \in \mathrm{H}_h$, $p_h \in \mathrm{Q}_h$, $\phi_h \in \mathrm{Z}_h$ such that

$$\tilde{a}_1(\boldsymbol{u}_h, \boldsymbol{v}_h^{\star}) + \tilde{b}_1(\boldsymbol{v}_h^{\star}, \phi_h) = \tilde{F}(\boldsymbol{v}_h^{\star}) \qquad \forall \boldsymbol{v}_h^{\star} \in \mathrm{H}_h^{\star}, \tag{3.1.9}$$

$$a_2(p_h, q_h) - b_2(q_h, \phi_h) = G(q_h) \qquad \forall q_h \in \mathrm{Q}_h, \tag{3.1.10}$$

$$b_1(\boldsymbol{u}_h, \psi_h) + b_2(p_h, \psi_h) - c(\phi_h, \psi_h) = 0 \qquad \forall \psi_h \in \mathrm{Z}_h, \tag{3.1.11}$$

where the bilinear forms $\tilde{a}_1 : \mathrm{H}_h \times \mathrm{H}_h^{\star} \to \mathbb{R}$, $\tilde{b}_1 : \mathrm{H}_h^{\star} \times \mathrm{Z}_h \to \mathbb{R}$, $b_1 : \mathrm{H}_h \times \mathrm{Z}_h \to \mathbb{R}$, $a_2 : \mathrm{Q}_h \times \mathrm{Q}_h \to \mathbb{R}$, $b_2 : \mathrm{Q}_h \times \mathrm{Z}_h \to \mathbb{R}$, $c : \mathrm{Z}_h \times \mathrm{Z}_h \to \mathbb{R}$ and the linear functionals

$\tilde{F} : \mathrm{H}_h^\star \to \mathbb{R}$, $G : \mathrm{Q}_h \to \mathbb{R}$ are specified in the following way:

$$\tilde{a}_1(\boldsymbol{u}_h, \boldsymbol{v}_h^\star) := \mu \sum_{j=1}^{|\mathcal{N}_h|} \int_{K_j^\star} \nabla \boldsymbol{u}_h : \nabla \boldsymbol{v}_h^\star - \mu \sum_{j=1}^{|\mathcal{N}_h|} \int_{\partial K_j^\star} \boldsymbol{v}_h^\star \cdot (\nabla \boldsymbol{u}_h \boldsymbol{n}),$$

$$b_1(\boldsymbol{u}_h, \psi_h) := -\int_\Omega \psi_h \, \mathrm{div}(\boldsymbol{u}_h), \quad c(\phi_h, \psi_h) := \frac{1}{\lambda + \mu} \int_\Omega \phi_h \psi_h,$$

$$\tilde{b}_1(\boldsymbol{v}_h^\star, \phi_h) := \sum_{j=1}^{|\mathcal{N}_h|} \int_{\partial K_j^\star} \phi_h (\boldsymbol{v}_h^\star \cdot \boldsymbol{n}) - \sum_{j=1}^{|\mathcal{N}_h|} \int_{K_j^\star} \phi_h \, \mathrm{div}(\boldsymbol{v}_h^\star), \quad b_2(q_h, \phi_h) := \frac{1}{\lambda + \mu} \int_\Omega q_h \phi_h,$$

$$a_2(p_h, q_h) := \left(\frac{c_0}{\alpha} + \frac{1}{\lambda + \mu} \right) \int_\Omega p_h q_h + \frac{1}{\alpha \eta} \int_\Omega \kappa \nabla p_h \cdot \nabla q_h,$$

$$\tilde{F}(\boldsymbol{v}_h^\star) := \sum_{j=1}^{|\mathcal{N}_h|} \int_{K_j^\star} \boldsymbol{f} \cdot \boldsymbol{v}_h^\star, \quad G(q_h) := \frac{\rho}{\alpha \eta} \kappa \boldsymbol{g} \cdot \nabla q_h + \frac{1}{\alpha} \int_\Omega s q_h,$$

where $G(\cdot)$ adopts a different form than in [62] (i.e. we do not have the contribution of $\kappa \boldsymbol{g} \cdot \boldsymbol{n}$ on $\Gamma_{\boldsymbol{u}}$) since now our pressure boundary condition on $\Gamma_{\boldsymbol{u}}$ involves the entire normal flux of the fluid. In addition, we define the bilinear forms $A_1 : \mathrm{H}_h \times \mathrm{H}_h \to \mathbb{R}$, $B_1 : \mathrm{H}_h \times \mathrm{Z}_h \to \mathbb{R}$ and the linear functional $F : \mathrm{H}_h \to \mathbb{R}$ as follows:

$$A_1(\boldsymbol{u}_h, \boldsymbol{v}_h) := \tilde{a}_1(\boldsymbol{u}_h, \mathcal{H}_h \boldsymbol{v}_h), \quad B_1(\boldsymbol{v}_h, \psi_h) := \tilde{b}_1(\mathcal{H}_h \boldsymbol{v}_h, \psi_h), \quad F(\boldsymbol{v}_h) := \tilde{F}(\mathcal{H}_h \boldsymbol{v}_h),$$

which allows us to recast the preliminary pure Petrov-Galerkin formulation (3.1.9)-(3.1.11) as a standard Galerkin method.

3.1.2 Stability properties

We proceed to discuss the stability properties of the bilinear operators and linear functionals arising from the mixed hybrid FVE formulation (3.1.9)-(3.1.11). First, one readily observes that the bilinear operators satisfy

$$|a_2(p_h, q_h)| \leq \max\left\{ \frac{c_0}{\alpha} + \frac{1}{\lambda + \mu}, \frac{\kappa_2}{\alpha \eta} \right\} ||p_h||_{1,\Omega} ||q_h||_{1,\Omega} \qquad \forall p_h, q_h \in \mathrm{Q}_h,$$

$$|b_1(\boldsymbol{u}_h, \psi_h)| \leq \sqrt{2} |\boldsymbol{u}_h|_{1,\Omega} ||\psi_h||_{0,\Omega} \qquad \forall \boldsymbol{u}_h \in \mathrm{H}_h, \psi_h \in \mathrm{Z}_h,$$

$$|b_2(p_h, \phi_h)| \leq (\lambda + \mu)^{-1} ||p_h||_{1,\Omega} ||\phi_h||_{0,\Omega} \qquad \forall p_h \in \mathrm{Q}_h, \phi_h \in \mathrm{Z}_h,$$

$$|c(\phi_h, \psi_h)| \leq (\lambda + \mu)^{-1} ||\phi_h||_{0,\Omega} ||\psi_h||_{0,\Omega} \qquad \forall \phi_h, \psi_h \in \mathrm{Z}_h,$$

and that the linear functionals satisfy

$$|F(\boldsymbol{v}_h)| \leq ||\boldsymbol{f}||_{0,\Omega} |\boldsymbol{v}_h|_{1,\Omega} \qquad \forall \boldsymbol{v}_h \in \mathrm{H}_h,$$

$$|G(q_h)| \leq \frac{\rho \kappa_2}{\alpha \eta} \big(||\boldsymbol{g}||_{0,\Omega} + \frac{\eta}{\rho \kappa_2} ||s||_{0,\Omega} \big) ||q_h||_{1,\Omega} \qquad \forall q_h \in \mathrm{Q}_h.$$

In addition, we review the following coercivity conditions (cf. [62]):

$$\begin{aligned} a_2(p_h, p_h) &\geq \alpha^{-1} \min\{c_0, \kappa_1 \eta^{-1}\} ||p_h||_{1,\Omega}^2 + (\lambda + \mu)^{-1} ||p_h||_{0,\Omega}^2 && \forall p_h \in \mathrm{Q}_h, \\ c(\phi_h, \phi_h) &\geq (\lambda + \mu)^{-1} ||\phi_h||_{0,\Omega}^2 && \forall \phi_h \in \mathrm{Z}_h. \end{aligned}$$

Furthermore, it is well-known that the bilinear operator $b_1(\cdot, \cdot)$ satisfies the discrete inf-sup condition (see, for instance, [14])

$$\sup_{\boldsymbol{u}_h \in \mathrm{H}_h \setminus \mathbf{0}} \frac{b_1(\boldsymbol{u}_h, \psi_h)}{|\boldsymbol{u}_h|_{1,\Omega}} \geq c_{b_1} ||\psi_h||_{0,\Omega} \qquad \forall \psi_h \in \mathrm{Z}_h,$$

for some constant $c_{b_1} > 0$. In addition, the next lemma readily follows from [25, 26] and establishes boundedness as well as a discrete inf-sup condition for the bilinear operator $B_1(\cdot, \cdot)$.

Lemma 3.1.1. *Let $\mathscr{T}$ be a shape regular family of quasi-uniform triangulations, then there exist constants $C_{B_1} > 0$ and $c_{B_1} > 0$ such that, for all $\mathcal{T}_h \in \mathscr{T}$,*

$$|B_1(\boldsymbol{v}_h, \phi_h)| \leq C_{B_1} |\boldsymbol{v}_h|_{1,\Omega} ||\phi_h||_{0,\Omega} \qquad \forall \boldsymbol{v}_h \in \mathrm{H}_h, \phi_h \in \mathrm{Z}_h,$$

and

$$\sup_{\boldsymbol{v}_h \in \mathrm{H}_h \setminus \mathbf{0}} \frac{B_1(\boldsymbol{v}_h, \phi_h)}{|\boldsymbol{v}_h|_{1,\Omega}} \geq c_{B_1} ||\phi_h||_{0,\Omega} \qquad \forall \phi_h \in \mathrm{Z}_h.$$

The following boundedness result follows straightforwardly from [25, 28].

Lemma 3.1.2. *Suppose that $\mathscr{T}$ is a shape regular family of quasi-uniform triangulations, then there exists $C_{A_1} > 0$ such that, for all $\mathcal{T}_h \in \mathscr{T}$ and every $\boldsymbol{u}_h, \boldsymbol{v}_h \in \mathrm{H}_h$,*

$$|A_1(\boldsymbol{u}_h, \boldsymbol{v}_h)| \leq C_{A_1} \big(|\boldsymbol{u}_h|_{1,\Omega} + h |\boldsymbol{u}_h|_{2,\mathcal{T}_h} \big) |\boldsymbol{v}_h|_{1,\Omega}.$$

Lastly, we require coercivity of the bilinear operator $A_1(\cdot, \cdot)$ to obtain suitable well-posedness and optimal convergence results for the envisioned mixed hybrid FVE formulation (3.1.9)-(3.1.11). We stress that proving this property is significantly more complicated and that meshes of high order FVE methods typically need to satisfy certain geometric requirements (cf. [25, 26, 57, 70]).

3.1.3 Coercivity and mesh geometric requirements

We proceed to verify coercivity of the bilinear operator $A_1(\cdot, \cdot)$ by employing the hierarchical structure of the trial and test function spaces for the displacement. In

particular, we state a sufficiency condition based on the positive semidefiniteness of symmetrisations of local FVE stiffness matrices and derive suitable mesh geometric requirements for the coercivity of $A_1(\cdot,\cdot)$, closely following the arguments from [25].

First, we introduce an isomorphism $\mathcal{C}_{\mathcal{T}_h} : \mathbb{R}^{2|\mathcal{N}_h|+2|\mathcal{N}_h|} \to \mathrm{H}_h$ that relates the discrete function space H_h with its associated coordinate space:

$$\mathcal{C}_{\mathcal{T}_h}(U^1, U^2) := \sum_{i=1}^{2|\mathcal{N}_h|} U_i^1 \boldsymbol{\phi}_i + \sum_{i=1}^{2|\mathcal{N}_h|} U_i^2 \boldsymbol{\psi}_i,$$

where from now on $\mathbb{V}_{1,h}$ and $\mathbb{W}_{2,h}$ denote the coordinate spaces associated with the discrete function spaces $\mathrm{V}_{1,h}$ and $\mathrm{W}_{2,h}$, respectively.

Furthermore, we introduce the FVE stiffness matrix $\mathcal{A}^{\mathrm{FVE}}$, which can be decomposed as

$$\mathcal{A}^{\mathrm{FVE}} = \begin{bmatrix} M^A & M^C \\ M^B & M^D \end{bmatrix},$$

where the $2|\mathcal{N}_h| \times 2|\mathcal{N}_h|$ sub matrices are defined by

$$M_{i,j}^A := A_1(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j), \quad M_{i,j}^B := A_1(\boldsymbol{\phi}_i, \boldsymbol{\psi}_j), \quad M_{i,j}^C := A_1(\boldsymbol{\psi}_i, \boldsymbol{\phi}_j), \quad M_{i,j}^D := A_1(\boldsymbol{\psi}_i, \boldsymbol{\psi}_j).$$

In addition, we define the symmetrisation of $\mathcal{A}^{\mathrm{FVE}}$ as $\mathcal{A}^{\mathrm{S}} := \frac{1}{2}\big(\mathcal{A}^{\mathrm{FVE}} + (\mathcal{A}^{\mathrm{FVE}})^T\big)$ and let $\mathcal{A}_K^{\mathrm{S}}$ denote its local counterpart. We are now ready to state the main result of this section.

Theorem 3.1.3. *Let $\mathscr{T}$ be a shape regular family of quasi-uniform triangulations and suppose that the matrix $\mathcal{A}_K^{\mathrm{S}}$ is positive semidefinite, for every $K \in \mathcal{T}_h \in \mathscr{T}$. Then, there exists c_{A_1}, depending only on the thresholds θ_l and θ_u, such that, for every $\boldsymbol{u}_h \in \mathrm{H}_h$,*

$$A_1(\boldsymbol{u}_h, \boldsymbol{u}_h) \geq \mu c_{A_1} |\boldsymbol{u}_h|_{1,\Omega}^2.$$

Proof. The proof follows directly by various straightforward extensions of the arguments employed in [25] from the scalar to the vectorial case. □

Evidently, Theorem 3.1.3 states that verifying positive semidefiniteness of the symmetric matrix $\mathcal{A}_K^{\mathrm{S}}$, for every $K \in \mathcal{T}_h \in \mathscr{T}$, is sufficient for establishing coercivity of the bilinear operator $A_1(\cdot,\cdot)$. Therefore, we proceed to express $\mathcal{A}_K^{\mathrm{S}}$ in terms of the angles $\theta_{\min,K}$ and $\theta_{\max,K}$.

Next, recall that the positive semidefiniteness property is invariant under an affine transformation T and multiplication by strictly positive constants, such that we may

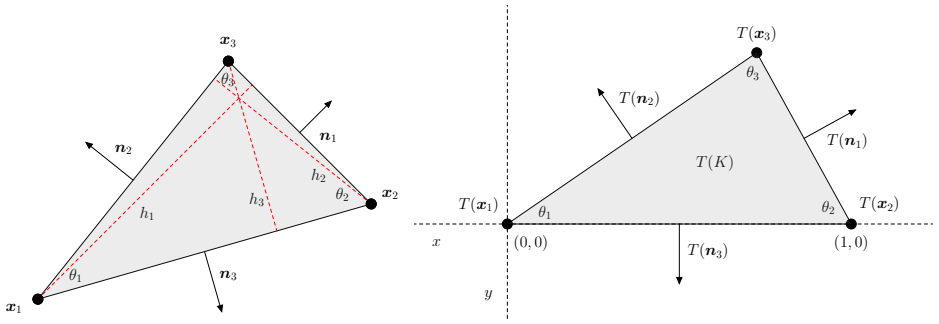


Figure 3.1: Geometric entities of a triangular element $K \in \mathcal{T}_h \in \mathscr{T}$ (a) and the planar configuration of the transposed element $T(K)$ (b); where the outward pointing normal vectors $\boldsymbol{n}_i$ and $T(\boldsymbol{n}_i)$ have magnitude 0.2 and the angles are without loss of generality indexed such that $\theta_1 \leq \theta_2 \leq \theta_3$.

without loss of generality assume that $\mu = 1$, that the ordering of the angles coincides with the ordering displayed in Figure 3.1(b), and that the points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ coincide with the origin and the point $(1, 0)$, respectively. This assumption significantly simplifies the computation of the entries of the symmetric matrix $\mathcal{A}_K^{\mathrm{S}}$

Throughout the remainder of this manuscript, we write $\theta_1 := \theta_{\min,K}$ and $\theta_3 := \theta_{\max,K}$, let θ_i denote the angle at the vertex $\boldsymbol{x}_i$, and index the edge opposite to the vertex $\boldsymbol{x}_i$ by e_i, see Figure 3.1(a). Furthermore, we let h_i denote the distance from the vertex $\boldsymbol{x}_i$ to the edge e_i and write $|e_i|$ for the length of the i-th edge. In addition, we let $\boldsymbol{m}_i = (\boldsymbol{m}_i^1, \boldsymbol{m}_i^2)$ denote the inwards pointing normal vector corresponding the edge e_i with magnitude $|e_i|$. Observe that, for two-dimensional triangular elements K, the normal vector $\boldsymbol{m}_i$ corresponding to edge e_i can be straightforwardly computed by a rotation of the edge vector. More precisely, we define $\boldsymbol{l}_i := \boldsymbol{x}_{i+1} - \boldsymbol{x}_{i-1}$, where the index i is 3-cyclic, such that we can write $\boldsymbol{m}_i = \boldsymbol{l}_i^{\perp}$, where $\boldsymbol{l}_i^{\perp}$ denotes the clockwise rotation of $\boldsymbol{l}_i$ by 90° degrees, i.e. $\boldsymbol{l}_i^{\perp} = [\boldsymbol{l}_i^1, \boldsymbol{l}_i^2]^{\perp} = [\boldsymbol{l}_i^2, -\boldsymbol{l}_i^1]$. We also mention that, when considering a fixed element K, the associated basis functions are with an abuse of notation denoted by $\boldsymbol{\phi}_i$ and $\boldsymbol{\psi}_i$, for $1 \leq i \leq 6$. More precisely, letting φ_i denote the canonical piecewise linear scalar nodal basis functions [41], we define the local basis functions $\boldsymbol{\phi}_i$ and $\boldsymbol{\psi}_i$, for $1 \leq i \leq 6$, as follows:

$$\boldsymbol{\phi}_i := \begin{cases} (\varphi_i, 0) & \text{for } 1 \leq i \leq 3, \\ (0, \varphi_i) & \text{for } 4 \leq i \leq 6, \end{cases} \qquad \text{and} \qquad \boldsymbol{\psi}_i := \begin{cases} (4\varphi_{i-1}\varphi_{i+1}, 0) & \text{for } 1 \leq i \leq 3, \\ (0, 4\varphi_{i-1}\varphi_{i+1}) & \text{for } 4 \leq i \leq 6, \end{cases}$$

where the index i is 3-cyclic. Explicit expressions for the gradients of these basis functions in terms of θ_1 and θ_3 readily follow from the observation that the gradients

of the scalar nodal basis functions $\nabla\varphi_i$, for $1 \leq i \leq 3$, are given by

$$\nabla\varphi_i = \frac{\boldsymbol{m}_i}{h_i|e_i|} = \frac{\boldsymbol{m}_i}{2\,|K|} = \frac{\boldsymbol{l}_i^{\perp}}{2\,|K|}, \tag{3.1.12}$$

where we have employed the identity $|K| = \frac{1}{2}|e_i|h_i$ in the last equality. It is a laborious but elementary exercise to verify that the (i,j)-th entry of the sub matrices M_K^A, M_K^B, M_K^C and M_K^D, for $1 \leq i,j \leq 3$ or $4 \leq i,j \leq 6$, takes the form

$$\big(M_K^A\big)_{i,j} = \frac{\boldsymbol{l}_i \cdot \boldsymbol{l}_j}{4\,|K|}, \qquad \big(M_K^B\big)_{i,j} = \frac{-\boldsymbol{l}_i \cdot \boldsymbol{l}_j}{3\,|K|}, \tag{3.1.13}$$

$$\big(M_K^C\big)_{i,j} = \begin{cases} -\frac{\boldsymbol{l}_{i-1}\cdot\boldsymbol{l}_{i-1}+\boldsymbol{l}_{i+1}\cdot\boldsymbol{l}_{i+1}}{4\,|K|} - \frac{2\boldsymbol{l}_{i+1}\cdot\boldsymbol{l}_{i-1}}{3\,|K|} & \text{if } i = j, \\ -\frac{3\boldsymbol{l}_k\cdot\boldsymbol{l}_k+3\boldsymbol{l}_j\cdot\boldsymbol{l}_k+\boldsymbol{l}_j\cdot\boldsymbol{l}_i}{2\,|K|} - \frac{4\boldsymbol{l}_{j+1}\cdot\boldsymbol{l}_{j-1}}{3\,|K|} & \text{if } i \neq j, \end{cases} \tag{3.1.14}$$

$$\big(M_K^D\big)_{i,j} = \begin{cases} \frac{4\boldsymbol{l}_{i+1}\cdot\boldsymbol{l}_{i+1}+4\boldsymbol{l}_{i-1}\cdot\boldsymbol{l}_{i-1}+4\boldsymbol{l}_{i+1}\boldsymbol{l}_{i-1}}{3\,|K|} & \text{if } i = j, \\ \frac{4\boldsymbol{l}_i\cdot\boldsymbol{l}_j+2\boldsymbol{l}_k\cdot\boldsymbol{l}_k}{3\,|K|} & \text{if } i \neq j, \end{cases} \tag{3.1.15}$$

where $i \neq j \neq k$ and the indices i and j are 3-cyclic. We also remark that one straightforwardly observes that the remaining entries equal zero. Lastly, we mention that one could simplify these expressions further by employing the relation (cf. [27])

$$|l_i|^2 = 2\,|K|\big(\cot(\theta_{i-1}) + \cot(\theta_{i+1})\big), \tag{3.1.16}$$

where the index i is once again 3-cyclic.

Exploiting (3.1.13)-(3.1.16), we proceed to compute the third smallest eigenvalue λ_3 of $\mathcal{A}_K^{\mathrm{S}}$, for different geometric configurations of K. Namely, if λ_3 is strictly positive, then positive semidefiniteness of $\mathcal{A}_K^{\mathrm{S}}$ follows from the fact that the symmetric block matrix has two zero eigenvalues, which correspond to the eigenvectors $V_1 = [1,1,1,0,0,0,\mathbf{0}_6]$ and $V_2 = [0,0,0,1,1,1,\mathbf{0}_6]$, where $\mathbf{0}_6$ denotes the all-zeros vector of length six.

Next, we consider the following set of admissible triangular elements (cf. [25]):

$$\mathcal{K} := \big\{K : 0^\circ < \theta_1 \leq 60^\circ \leq \theta_3 < 180^\circ,\ \theta_1 \leq 180^\circ - \theta_1 - \theta_3 \leq \theta_3\big\}.$$

We discretise the domain $(0,180^\circ) \times (0,180^\circ)$ in the $\theta_1 - \theta_3$ plane by a uniform grid of mesh size 0.05 and compute $\lambda_3 = \lambda_3(\theta_1,\theta_3)$ as a function of the minimum and maximum angles of the element K, where we set $\lambda_3 = 0$ outside of $\mathcal{K}$, see Figure 3.2. Observe that these results coincide with the angular requirements obtained in [25], as expected. Evidently, if the thresholds θ_l and θ_u are chosen such that every triangle in the set $\mathcal{K} \cap \{K : \theta_1 \geq \theta_l,\ \theta_3 \leq \theta_u\}$ is admissible, then coercivity of $A_1(\cdot,\cdot)$ follows by

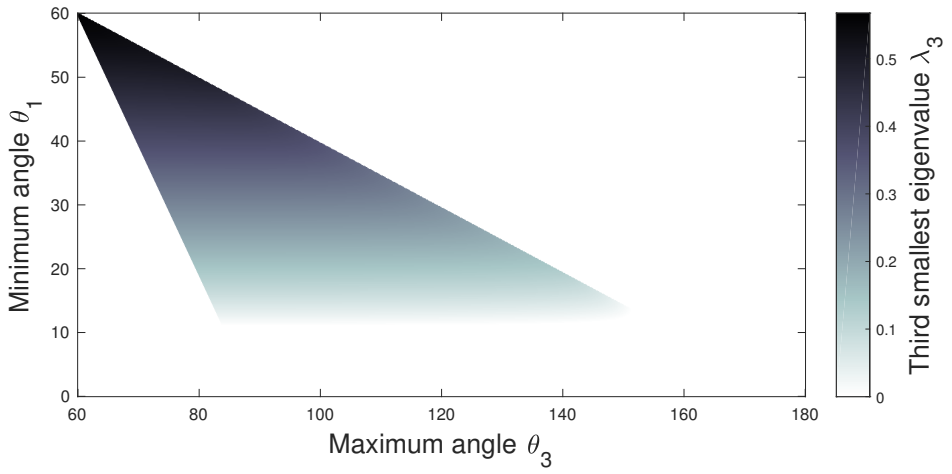


Figure 3.2: The third smallest eigenvalue $\lambda_3(\theta_1, \theta_3)$ of the symmetrised local FVE stiffness matrix plotted against the minimum angle θ_1 (y-axis) and the maximum angle θ_3 (x-axis).

the virtue of Theorem 3.1.3. We will assume this is indeed the case throughout the remainder of this dissertation. However, we also mention that the angle conditions introduced above are by no means necessary, and that the FVE scheme (3.1.9)-(3.1.11) may still have an optimal convergence rate if there are a number of triangular elements $K \in \mathcal{T}_h \in \mathscr{T}$ for which the conditions $\theta_1 \geq \theta_l$ and $\theta_3 \leq \theta_u$ are not satisfied.

To close this section, we remark that replacing the gradients in the definition of the bilinear operator $A_1(\cdot, \cdot)$ with infinitesimal strain tensors $\boldsymbol{\varepsilon}(\cdot)$ drastically changes the spectral properties in the characterisation of coercivity in terms of symmetrised local FVE stiffness matrices. In fact, in Appendix B it is shown that negativity of one eigenvalue restrains us from establishing coercivity when considering strains instead, for every possible geometric configuration. This drawback serves as the main motivation for considering the novel three-field formulation (3.1.4)-(3.1.6) instead of the poroelasticity equations introduced by Oyarzúa et al. [62]

3.2 Well-posedness and convergence

In this section, we address the unique solvability and stability of the mixed hybrid FVE formulation (3.1.9)-(3.1.11) and derive optimal convergence estimates.

3.2.1 Well-posedness

For notational convenience, we first introduce the following discrete product space:

$$\mathrm{X}_h := \mathrm{H}_h \times \mathrm{Q}_h \times \mathrm{Z}_h,$$

endowed with the norm

$$||\vec{\boldsymbol{u}}_h||_{\mathrm{X}_h} := |\boldsymbol{u}_h|_{1,\Omega} + ||p_h||_{1,\Omega} + ||\phi_h||_{0,\Omega},$$

where we write $\vec{\boldsymbol{u}}_h = (\boldsymbol{u}_h, p_h, \phi_h) \in \mathrm{X}_h$. Remark that a generalisation of the Poincaré inequality (see, for instance, [41, Lemma 3.1]) establishes an equivalence between the norm $||\cdot||_{1,\Omega}$ and the seminorm $|\cdot|_{1,\Omega}$ on the space H, which in turn ensures that $||\cdot||_{\mathrm{X}_h}$ indeed defines a norm on the product space X_h, provided that the segment $\Gamma_{\boldsymbol{u}}$ has strictly positive measure. We also review a classical inverse type estimate, which states there exists $C_{\mathrm{inv}} > 0$, independent of h and λ, such that

$$h|\boldsymbol{u}_h|_{2,\mathcal{T}_h} \leq C_{\mathrm{inv}}|\boldsymbol{u}_h|_{1,\Omega}. \tag{3.2.1}$$

In particular, this inequality will be employed to establish well-posedness of the discrete formulation (3.1.9)-(3.1.11).

Next, we review the well-known Babuška-Lax-Milgram theorem. For this purpose, recall that a continuous bilinear functional $D : \mathrm{X} \times \mathrm{Y} \to \mathbb{R}$ is called weakly coercive if there exists a constant $c_D > 0$ such that

$$\sup_{\boldsymbol{v}\in\mathrm{Y}:||\boldsymbol{v}||_{\mathrm{Y}}=1} |D(\boldsymbol{u}, \boldsymbol{v})| \geq c_D||\boldsymbol{u}||_{\mathrm{X}},$$

for all $\boldsymbol{u} \in \mathrm{X}$, and

$$\sup_{\boldsymbol{u}\in\mathrm{X}} |D(\boldsymbol{u}, \boldsymbol{v})| > 0, \tag{3.2.2}$$

for all non-zero $\boldsymbol{v} \in \mathrm{Y}$. If the spaces X and Y are finite dimensional, then the second requirement (3.2.2) can be replaced by the requirement that $\dim \mathrm{X} = \dim \mathrm{Y}$. A proof of the following well-known Babuška-Lax-Milgram theorem can be found in [61].

Theorem 3.2.1. *Let* X *and* Y *be two real Hilbert spaces and let* $D : \mathrm{X} \times \mathrm{Y} \to \mathbb{R}$ *be a weakly coercive continuous bilinear functional. Then, for all* $F \in \mathrm{Y}^\star$*, there exists a unique solution* $\hat{\boldsymbol{u}} \in \mathrm{X}$ *such that*

$$D(\hat{\boldsymbol{u}}, \boldsymbol{v}) = \langle F, \boldsymbol{v}\rangle,$$

for all $\boldsymbol{v} \in \mathrm{Y}$*. Moreover, the solution depends continuously on the given datum:*

$$||\hat{\boldsymbol{u}}||_{\mathrm{X}} \leq \frac{1}{c_D}||F||_{\mathrm{X}^\star}.$$

Next, well-posedness of the mixed FVE formulation (3.1.9)-(3.1.11) is established using Theorem 3.2.1 in combination with appropriate boundedness and coercivity results obtained in Section 3.1.

Theorem 3.2.2. *There exists a unique solution* $(\hat{\boldsymbol{u}}_h, \hat{p}_h, \hat{\phi}_h) \in \mathrm{H}_h \times \mathrm{Q}_h \times \mathrm{Z}_h$ *of the discrete problem* (3.1.9)-(3.1.11). *Moreover, there exists a constant* $C_{stab} > 0$, *independent of* h *and* λ, *such that*

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

Proof. First, observe that the envisioned FVE formulation (3.1.9)-(3.1.11) is fully equivalent to the following discrete formulation: Find $(\hat{\boldsymbol{u}}_h, \hat{p}_h, \hat{\phi}_h) \in \mathrm{H}_h \times \mathrm{Q}_h \times \mathrm{Z}_h$ such that, for all $(\boldsymbol{v}_h, q_h, \psi_h) \in \mathrm{H}_h \times \mathrm{Q}_h \times \mathrm{Z}_h$,

$$D_1\big((\hat{\boldsymbol{u}}_h, \hat{\phi}_h), (\boldsymbol{v}_h, \psi_h)\big) + D_2\big((\hat{p}_h, \hat{\phi}_h), (q_h, \psi_h)\big) = F(\boldsymbol{v}_h) + G(q_h),$$

where we have defined the bilinear forms $D_1(\cdot,\cdot)$ and $D_2(\cdot,\cdot)$ as

$$\begin{aligned} D_1\big((\boldsymbol{u}_h, \phi_h), (\boldsymbol{v}_h, \psi_h)\big) &:= A_1(\boldsymbol{u}_h, \boldsymbol{v}_h) + B_1(\boldsymbol{v}_h, \phi_h) - b_1(\boldsymbol{u}_h, \psi_h), \\ D_2\big((p_h, \phi_h), (q_h, \psi_h)\big) &:= a_2(p_h, q_h) - b_2(q_h, \phi_h) - b_2(p_h, \psi_h) + c(\phi_h, \psi_h). \end{aligned}$$

Furthermore, we introduce another bilinear operator $D : (\mathrm{H}_h \times \mathrm{Q}_h \times \mathrm{Z}_h) \times (\mathrm{H}_h \times \mathrm{Q}_h \times \mathrm{Z}_h) \to \mathbb{R}$, which is defined as

$$D\big((\boldsymbol{u}_h, p_h, \phi_h), (\boldsymbol{v}_h, q_h, \psi_h)\big) := D_1\big((\boldsymbol{u}_h, \phi_h), (\boldsymbol{v}_h, \psi_h)\big) + D_2\big((p_h, \phi_h), (q_h, \psi_h)\big).$$

First, we straightforwardly verify continuity of the bilinear operators $D_1(\cdot,\cdot)$ and $D_2(\cdot,\cdot)$, which in turn implies continuity of the bilinear operator $D(\cdot,\cdot)$. In fact, by the virtue of the boundedness results discussed in Section 3.1 we find that there exists $C_{D_1} > 0$, independent of h and λ, such that

$$\begin{aligned} |D_1\big((\boldsymbol{u}_h, \phi_h), (\boldsymbol{v}_h, \psi_h)\big)| &\leq \mu C_{A_1}|\boldsymbol{u}_h|_{1,\Omega}|\boldsymbol{v}_h|_{1,\Omega} + C_{B_1}|\boldsymbol{v}_h|_{1,\Omega}||\phi_h||_{0,\Omega} + \sqrt{2}|\boldsymbol{u}_h|_{1,\Omega}||\psi_h||_{0,\Omega} \\ &\leq C_{D_1}\big(|\boldsymbol{u}_h|_{1,\Omega} + ||\phi_h||_{0,\Omega}\big)\big(|\boldsymbol{v}_h|_{1,\Omega} + ||\psi_h||_{0,\Omega}\big). \end{aligned}$$

Similarly, there exists $C_{D_2} > 0$, independent of h and λ, such that

$$\begin{aligned} |D_2\big((p_h, \phi_h), (q_h, \psi_h)\big)| &\leq \max\left\{\frac{c_0}{\alpha} + \frac{1}{\lambda+\mu}, \frac{\kappa_2}{\alpha\eta}\right\}||p_h||_{1,\Omega}||q_h||_{1,\Omega} + (\lambda+\mu)^{-1} \\ &\quad \times \big(||q_h||_{1,\Omega}||\phi_h||_{0,\Omega} + ||p_h||_{1,\Omega}||\psi_h||_{0,\Omega} + ||\phi_h||_{0,\Omega}||\psi_h||_{0,\Omega}\big) \\ &\leq C_{D_2}\big(||p_h||_{1,\Omega} + ||\phi_h||_{0,\Omega}\big)\big(||q_h||_{1,\Omega} + ||\psi_h||_{0,\Omega}\big), \end{aligned}$$

for all $(\boldsymbol{u}_h, p_h, \phi_h), (\boldsymbol{v}_h, q_h, \psi_h) \in \mathrm{H}_h \times \mathrm{Q}_h \times \mathrm{Z}_h$. Furthermore, after combining both inequalities, we find that there exists $C_D > 0$, independent of h and λ, such that

$$|D(\vec{\boldsymbol{u}}_h, \vec{\boldsymbol{v}}_h)| \leq C_D\big(|\boldsymbol{u}_h|_{1,\Omega} + ||p_h||_{1,\Omega} + ||\phi_h||_{0,\Omega}\big)\big(|\boldsymbol{v}_h|_{1,\Omega} + ||q_h||_{1,\Omega} + ||\psi_h||_{0,\Omega}\big),$$

which can in turn be written as

$$|D(\vec{\boldsymbol{u}}_h, \vec{\boldsymbol{v}}_h)| \leq C_D||\vec{\boldsymbol{u}}_h||_{\mathrm{X}_h}||\vec{\boldsymbol{v}}_h||_{\mathrm{X}_h}. \tag{3.2.3}$$

We proceed to prove weak coercivity of the bilinear operator $D(\cdot,\cdot)$. For this purpose, observe that Lemma 3.1.1 readily implies that for any $\phi_h \in \mathrm{Z}_h$, there exists $\boldsymbol{w}_h \in \mathrm{H}_h$ such that (cf. [41])

$$B_1(\boldsymbol{w}_h, \phi_h) \geq ||\phi_h||^2_{0,\Omega}, \quad |\boldsymbol{w}_h|_{1,\Omega} \leq c_{B_1}^{-1}||\phi_h||_{0,\Omega}, \tag{3.2.4}$$

where $c_{B_1} > 0$ is independent of h and corresponds to the constant in Lemma 3.1.1. Next, for any $(\boldsymbol{u}_h, p_h, \phi_h) \in \mathrm{H}_h \times \mathrm{Q}_h \times \mathrm{Z}_h$, we choose

$$(\boldsymbol{v}_h, q_h, \psi_h) = (\boldsymbol{u}_h - \tau\boldsymbol{w}_h, p_h, \phi_h), \tag{3.2.5}$$

where the value of the positive constant τ depends on the bilinear operator $D(\cdot,\cdot)$ and will be specified later. For this particular choice of $(\boldsymbol{v}_h, q_h, \psi_h)$, we straightforwardly obtain

$$D_1\big((\boldsymbol{u}_h, \phi_h), (\boldsymbol{u}_h - \tau\boldsymbol{w}_h, \phi_h)\big) = A_1(\boldsymbol{u}_h, \boldsymbol{u}_h) - \tau A_1(\boldsymbol{u}_h, \boldsymbol{w}_h) - \tau B_1(\boldsymbol{w}_h, \phi_h),$$

where we have employed the bilinearity of the involved operators in combination with the fact that $b_1(\cdot,\cdot)$ and $B_1(\cdot,\cdot)$ coincide (see, for instance, [28]). Furthermore, Lemma 3.1.2 in combination with (3.2.4) as well as inequality (3.2.1) implies that there exists $C_0 > 0$, independent of h and λ, such that

$$D_1\big((\boldsymbol{u}_h, \phi_h), (\boldsymbol{u}_h - \tau\boldsymbol{w}_h, \phi_h)\big) \geq \mu c_{A_1}|\boldsymbol{u}_h|^2_{1,\Omega} - \tau C_0|\boldsymbol{u}_h|_{1,\Omega}||\phi_h||_{0,\Omega} + \tau||\phi_h||^2_{0,\Omega}.$$

This inequality can be refined further by one application of Young's inequality

$$ab \leq \frac{a^2}{2\epsilon} + \frac{\epsilon b^2}{2},$$

which holds for any $\epsilon > 0$. More precisely, after one application of Young's inequality with $\epsilon = \tau C_0/\mu c_{A_1}$, we find that

$$\begin{aligned} D_1\big((\boldsymbol{u}_h, \phi_h), (\boldsymbol{u}_h - \tau\boldsymbol{w}_h, \phi_h)\big) &\geq (\mu c_{A_1} - \frac{\tau C_0}{2\epsilon})|\boldsymbol{u}_h|^2_{1,\Omega} + (\tau - \frac{\tau C_0 \epsilon}{2})||\phi_h||^2_{0,\Omega} \\ &= \frac{\mu c_{A_1}}{2}|\boldsymbol{u}_h|^2_{1,\Omega} + \tau(1 - \tau\frac{C_0^2}{2\mu c_{A_1}})||\phi_h||^2_{0,\Omega}. \end{aligned}$$

In particular, setting $\tau < 2\mu c_{A_1}/C_0^2$ ensures that there exists $c_{D_1} > 0$, independent of h and λ, such that

$$D_1\big((\boldsymbol{u}_h, \phi_h), (\boldsymbol{u}_h - \tau \boldsymbol{w}_h, \phi_h)\big) \geq c_{D_1}(|\boldsymbol{u}_h|_{1,\Omega} + ||\phi_h||_{0,\Omega})^2. \tag{3.2.6}$$

We proceed to obtain a lower bound for the second bilinear operator $D_2(\cdot, \cdot)$. For this purpose, we first observe that the boundedness and coercivity results obtained in Section 3.1 readily imply that

$$\begin{aligned} D_2\big((p_h, \phi_h), (p_h, \phi_h)\big) &= a_2(p_h, p_h) - b_2(p_h, \phi_h) - b_2(p_h, \phi_h) + c(\phi_h, \phi_h) \\ &\geq \alpha^{-1} \min\{c_0, \kappa_1 \eta^{-1}\} ||p_h||_{1,\Omega}^2 + (\lambda + \mu)^{-1} ||p_h||_{0,\Omega}^2 \\ &\quad - 2(\lambda + \mu)^{-1} ||p_h||_{1,\Omega} ||\phi_h||_{0,\Omega} + (\lambda + \mu)^{-1} ||\phi_h||_{0,\Omega}^2. \end{aligned}$$

In particular, after employing the standard inequality $-2ab \geq -a^2 - b^2$ and subtracting $(\lambda + \mu)^{-1} ||p_h||_{1,\Omega}^2$ and $(\lambda + \mu)^{-1} ||p_h||_{0,\Omega}^2$, we find that there exists $c_{D_2} > 0$, independent of h and λ, such that

$$D_2\big((p_h, \phi_h), (p_h, \phi_h)\big) \geq c_{D_2} ||p_h||_{1,\Omega}^2. \tag{3.2.7}$$

Hence, after combining inequality (3.2.6) and inequality (3.2.7), we directly establish that

$$\begin{aligned} D\big((\boldsymbol{u}_h, p_h, \phi_h), (\boldsymbol{u}_h - \tau \boldsymbol{w}_h, p_h, \phi_h)\big) &\geq \min\{c_{D_1}, c_{D_2}\}\big((|\boldsymbol{u}_h|_{1,\Omega} + ||\phi_h||_{0,\Omega})^2 + ||p_h||_{1,\Omega}^2\big) \\ &\geq \tfrac{1}{2} \min\{c_{D_1}, c_{D_2}\}(|\boldsymbol{u}_h|_{1,\Omega} + ||p_h||_{1,\Omega} + ||\phi_h||_{0,\Omega})^2, \end{aligned} \tag{3.2.8}$$

where we have employed the inequality $(a + b)^2 + c^2 \geq \frac{1}{2}(a + b + c)^2$ in the last step. Moreover, we observe that the choice of $(\boldsymbol{v}_h, q_h, \psi_h)$ given by (3.2.5) in combination with (3.2.4) implies that

$$\begin{aligned} |\boldsymbol{v}_h|_{1,\Omega} + ||q_h||_{1,\Omega} + ||\psi_h||_{0,\Omega} &= |\boldsymbol{u}_h - \tau \boldsymbol{w}_h|_{1,\Omega} + ||p_h||_{1,\Omega} + ||\phi_h||_{0,\Omega} \\ &\leq |\boldsymbol{u}_h|_{1,\Omega} + ||p_h||_{1,\Omega} + (\tau c_{B_1}^{-1} + 1)||\phi_h||_{0,\Omega} \\ &\leq C_1\big(|\boldsymbol{u}_h|_{1,\Omega} + ||p_h||_{1,\Omega} + ||\phi_h||_{0,\Omega}\big), \end{aligned}$$

for some $C_1 > 0$, independent of h and λ. Hence, after employing the previous inequality in combination with the (3.2.8), we find that there exists $c_D > 0$, independent of h and λ, such that

$$\sup_{(\boldsymbol{v}_h, q_h, \psi_h) \in (\mathrm{H}_h \times \mathrm{Q}_h \times \mathrm{Z}_h) \setminus \mathbf{0}} \frac{D\big((\boldsymbol{u}_h, p_h, \phi_h), (\boldsymbol{v}_h, q_h, \psi_h)\big)}{|\boldsymbol{v}_h|_{1,\Omega} + ||q_h||_{1,\Omega} + ||\psi_h||_{0,\Omega}} \geq c_D\big(|\boldsymbol{u}_h|_{1,\Omega} + ||p_h||_{1,\Omega} + ||\phi_h||_{0,\Omega}\big),$$

which can also be formulated as

$$\sup_{\vec{\boldsymbol{v}}_h \in \mathrm{X}_h \setminus \mathbf{0}} \frac{D(\vec{\boldsymbol{u}}_h, \vec{\boldsymbol{v}}_h)}{||\vec{\boldsymbol{v}}_h||_{\mathrm{X}_h}} \geq c_D ||\vec{\boldsymbol{u}}_h||_{\mathrm{X}_h}. \tag{3.2.9}$$

In other words, inequality (3.2.3) and inequality (3.2.9) imply that the bilinear operator $D(\cdot, \cdot)$ is bounded and weakly coercive. Furthermore, after observing that the Hilbert space X_h is finite dimensional, we find that the Babuška-Lax-Milgram Theorem ensures well-posedness of the mixed hybrid FVE formulation (3.1.9)-(3.1.11), which completes the proof. □

3.2.2 Error estimates

The goal of this section is to obtain a quasi-optimality result and optimal a priori error estimates for the mixed hybrid FVE formulation (3.1.9)-(3.1.11). We will proceed by deriving the corresponding Céa's estimate.

Prior to stating and proving the two main theorems of this section, we introduce the subspace $K_h \subseteq \mathrm{H}_h$ that consists of all feasible vectors $\boldsymbol{v}_h$ satisfying (3.1.11). More precisely, we define (cf. [62])

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

which is clearly non-empty since $\hat{\boldsymbol{u}}_h \in \mathrm{K}_h$ and forms a linear subspace of H_h. We mention that, if $(\boldsymbol{u}, p, \phi)$ denotes the solution to the continuous coupled problem (3.1.4)-(3.1.6), then it is not difficult to derive that the following inequality holds (for a proof, we refer to [37, Thm. 2.6]):

$$\mathrm{dist}(\boldsymbol{u}, \mathrm{K}_h) \leq C_{\mathrm{dist}} \, \mathrm{dist}(\boldsymbol{u}, \mathrm{H}_h) \qquad \forall \boldsymbol{u} \in \mathrm{H}, \tag{3.2.10}$$

for some constant $C_{\mathrm{dist}} > 0$, independent of h and λ. Next, we proceed to derive the corresponding Céa's estimate for the mixed hybrid FVE scheme (3.1.9)-(3.1.11).

Theorem 3.2.3. *Let $(\boldsymbol{u}, p, \phi) \in \mathrm{H} \times \mathrm{Q} \times \mathrm{Z}$ and $(\hat{\boldsymbol{u}}_h, \hat{p}_h, \hat{\phi}_h) \in \mathrm{H}_h \times \mathrm{Q}_h \times \mathrm{Z}_h$ be the solutions of the continuous and discrete problems* (3.1.4)-(3.1.6) *and* (3.1.9)-(3.1.11)*, respectively. Then there exists $C_{Céa} > 0$, independent of h and λ, such that*

$$|\boldsymbol{e}_{\boldsymbol{u}}|_{1,\Omega} + ||e_p||_{1,\Omega} + ||e_\phi||_{0,\Omega} \leq C_{Céa} \big(\mathrm{dist}(\boldsymbol{u}, \mathrm{H}_h) + \mathrm{dist}(p, \mathrm{Q}_h) + \mathrm{dist}(\phi, \mathrm{Z}_h) \big).$$

Proof. First, in order to simplify the subsequent analysis, we write $\boldsymbol{e}_{\boldsymbol{u}} := \boldsymbol{u} - \hat{\boldsymbol{u}}_h$, $e_p := p - \hat{p}_h$, and $e_\phi := \phi - \hat{\phi}_h$ for the approximation errors of interest. Furthermore, for arbitrary $\tilde{\boldsymbol{v}}_h \in \mathrm{K}_h$, $\tilde{q}_h \in \mathrm{Q}_h$, and $\tilde{\psi}_h \in \mathrm{Z}_h$, we can decompose these errors into

$$\boldsymbol{e}_{\boldsymbol{u}} = \boldsymbol{r}_{\boldsymbol{u}} + \boldsymbol{\chi}_{\boldsymbol{u}}, \quad e_p = r_p + \chi_p, \quad \text{and} \quad e_\phi = r_\phi + \chi_\phi,$$

where we write

$$\boldsymbol{r_u} := \boldsymbol{u} - \tilde{\boldsymbol{v}}_h, \quad r_p := p - \tilde{q}_h, \quad r_\phi := \phi - \tilde{\psi}_h,$$
$$\boldsymbol{\chi_u} := \tilde{\boldsymbol{v}}_h - \hat{\boldsymbol{u}}_h, \quad \chi_p := \tilde{q}_h - \hat{p}_h, \quad \chi_\phi := \tilde{\psi}_h - \hat{\phi}_h.$$

Moreover, since both $\hat{\boldsymbol{u}}_h$ and $\tilde{\boldsymbol{v}}_h$ belong to K_h, we find that

$$b_1(\hat{\boldsymbol{u}}_h, \psi_h) = -b_2(\hat{p}_h, \psi_h) + c(\hat{\phi}_h, \psi_h), \quad b_1(\tilde{\boldsymbol{v}}_h, \psi_h) = -b_2(\hat{p}_h, \psi_h) + c(\hat{\phi}_h, \psi_h),$$

for all $\psi_h \in \mathrm{Z}_h$, such that

$$b_1(\boldsymbol{\chi_u}, \psi_h) = b_1(\tilde{\boldsymbol{v}}_h - \hat{\boldsymbol{u}}_h, \psi_h) = 0 \qquad \forall \psi_h \in \mathrm{Z}_h,$$

which in turn implies that $\boldsymbol{\chi_u} \in \mathrm{Ker}_h(b_1) := \{\boldsymbol{v}_h \in \mathrm{H}_h : b_1(\boldsymbol{v}_h, \psi_h) = 0 \quad \forall \psi_h \in \mathrm{Z}_h\}$. Furthermore, observe that the FVE formulation (3.1.9)-(3.1.11) in combination with the corresponding three-field formulation (3.1.4)-(3.1.6) implies that

$$A_1(\boldsymbol{\chi_u}, \boldsymbol{v}_h) + B_1(\boldsymbol{v}_h, \chi_\phi) = -A_1(\boldsymbol{r_u}, \boldsymbol{v}_h) - B_1(\boldsymbol{v}_h, r_\phi) \qquad \forall \boldsymbol{v}_h \in \mathrm{H}_h, \tag{3.2.11}$$

$$a_2(\chi_p, q_h) - b_2(q_h, \chi_\phi) = -a_2(r_p, q_h) + b_2(q_h, r_\phi) \qquad \forall q_h \in \mathrm{Q}_h, \tag{3.2.12}$$

$$b_1(\boldsymbol{\chi_u}, \psi_h) + b_2(\chi_p, \psi_h) - c(\chi_\phi, \psi_h) = -b_1(\boldsymbol{r_u}, \psi_h) - b_2(r_p, \psi_h) + c(r_\phi, \psi_h) \qquad \forall \psi_h \in \mathrm{Z}_h. \tag{3.2.13}$$

In what follows, we focus on obtaining appropriate bounds for $\boldsymbol{\chi_u}$, χ_p and χ_ϕ. First, setting $\boldsymbol{v}_h = \boldsymbol{\chi_u} \in \mathrm{Ker}_h(b_1)$, we readily observe that identity (3.2.11) implies that

$$A_1(\boldsymbol{\chi_u}, \boldsymbol{\chi_u}) = -A_1(\boldsymbol{r_u}, \boldsymbol{\chi_u}) - B_1(\boldsymbol{\chi_u}, r_\phi).$$

By the virtue of Lemma 3.1.1 and Lemma 3.1.2, observe that the first term on the right-hand side satisfies

$$A_1(\boldsymbol{r_u}, \boldsymbol{\chi_u}) \leq C_{A_1}(1 + C_{\mathrm{inv}})|\boldsymbol{r_u}|_{1,\Omega}|\boldsymbol{\chi_u}|_{1,\Omega}, \tag{3.2.14}$$

where we have used inequality (3.2.1). In addition, the results from [28] ensure that there exists $C_0 > 0$, independent of h, for which the bilinear operator $B_1(\cdot, \cdot)$ satisfies

$$B_1(\boldsymbol{\chi_u}, r_\phi) \leq C_0(||r_\phi||_{0,\Omega} + h|r_\phi|_{1,\mathcal{T}_h})|\boldsymbol{\chi_u}|_{1,\Omega}, \tag{3.2.15}$$

such that the estimates (3.2.14) and (3.2.15) in combination coercivity of the bilinear operator $A_1(\cdot, \cdot)$ and an inverse type estimate similar to inequality (3.2.1) imply that there exists $C_1 > 0$, independent of h and λ, such that

$$\mu c_{A_1}|\boldsymbol{\chi_u}|^2_{1,\Omega} \leq A_1(\boldsymbol{\chi_u}, \boldsymbol{\chi_u}) \leq C_1(|\boldsymbol{r_u}|_{1,\Omega} + ||r_\phi||_{0,\Omega})|\boldsymbol{\chi_u}|_{1,\Omega},$$

which in turn implies that there exists $C_2 > 0$, independent of h and λ, such that

$$|\boldsymbol{\chi_u}|_{1,\Omega} \leq C_2(|\boldsymbol{r_u}|_{1,\Omega} + ||r_\phi||_{0,\Omega}). \tag{3.2.16}$$

Next, we obtain an upper bound for the error χ_ϕ by employing Lemma 3.1.1 in combination with identity (3.2.11) and noticing that

$$\begin{aligned}||\chi_\phi||_{0,\Omega} &\leq c_{B_1}^{-1} \sup_{\boldsymbol{v}_h \in \mathrm{H}_h \setminus \mathbf{0}} \frac{|B_1(\boldsymbol{v}_h, \chi_\phi)|}{|\boldsymbol{v}_h|_{1,\Omega}} = c_{B_1}^{-1} \sup_{\boldsymbol{v}_h \in \mathrm{H}_h \setminus \mathbf{0}} \frac{|A_1(\boldsymbol{r_u}, \boldsymbol{v}_h) + A_1(\boldsymbol{\chi_u}, \boldsymbol{v}_h) + B_1(\boldsymbol{v}_h, r_\phi)|}{|\boldsymbol{v}_h|_{1,\Omega}} \\ &\leq c_{B_1}^{-1}(\mu C_{A_1}|\boldsymbol{r_u}|_{1,\Omega} + \mu C_{A_1}|\boldsymbol{\chi_u}|_{1,\Omega} + C_{B_1}||r_\phi||_{0,\Omega}),\end{aligned}$$

which in combination with (3.2.16) implies that there exists $C_3 > 0$, independent of h and λ, such that

$$||\chi_\phi||_{0,\Omega} \leq C_3(|\boldsymbol{r_u}|_{1,\Omega} + ||r_\phi||_{0,\Omega}). \tag{3.2.17}$$

Lastly, we obtain an upper bound for χ_p using equation (3.2.12) and substituting χ_p for q_h. Observe that the coercivity of $a_2(\cdot,\cdot)$ in combination with the continuity of $a_2(\cdot,\cdot)$ and $b_2(\cdot,\cdot)$ implies that

$$\begin{aligned}\min\left\{\frac{c_0}{\alpha}, \frac{\kappa_1\eta}{\alpha}\right\}||\chi_p||_{1,\Omega}^2 &\leq -a_2(r_p, \chi_p) + b_2(\chi_p, r_\phi) + b_2(\chi_p, \chi_\phi) \\ &\leq \frac{1}{\alpha}\max\left\{c_0 + \frac{\alpha}{\lambda+\mu}, \frac{\kappa_2}{\eta}\right\}||r_p||_{1,\Omega}||\chi_p||_{1,\Omega} \\ &\quad + (\lambda+\mu)^{-1}||r_\phi||_{0,\Omega}||\chi_p||_{1,\Omega} + (\lambda+\mu)^{-1}||\chi_\phi||_{0,\Omega}||\chi_p||_{1,\Omega},\end{aligned}$$

which evidently implies that there exists $C_4 > 0$, independent of h and λ, such that

$$||\chi_p||_{1,\Omega} \leq C_4(|\boldsymbol{r_u}|_{1,\Omega} + ||r_p||_{1,\Omega} + ||r_\phi||_{0,\Omega}). \tag{3.2.18}$$

Next, we combine the inequalities (3.2.16), (3.2.17) and (3.2.18) to derive that there exists $C_5 > 0$, independent of h and λ, such that

$$|\boldsymbol{\chi_u}|_{1,\Omega} + ||\chi_p||_{1,\Omega} + ||\chi_\phi||_{0,\Omega} \leq C_5(|\boldsymbol{r_u}|_{1,\Omega} + ||r_p||_{1,\Omega} + ||r_\phi||_{0,\Omega}).$$

Consequently, by one application of the triangle inequality in combination with the previous inequality and the fact that $\tilde{\boldsymbol{v}}_h$, $\tilde{q}_h$ and $\tilde{\psi}_h$ are arbitrary we obtain

$$|\boldsymbol{e_u}|_{1,\Omega} + ||e_p||_{1,\Omega} + ||e_\phi||_{0,\Omega} \leq (1 + C_5)\big(\mathrm{dist}(\boldsymbol{u}, \mathrm{K}_h) + \mathrm{dist}(p, \mathrm{Q}_h) + \mathrm{dist}(\phi, \mathrm{Z}_h)\big),$$

such that (3.2.10) in combination with the last inequality completes the proof. □

To close this section, we observe that Theorem 3.2.3 can be employed to obtain optimal convergence bounds, using classical interpolation and projection estimates.

Theorem 3.2.4. *Let* $(\boldsymbol{u}, p, \phi) \in \mathrm{H} \times \mathrm{Q} \times \mathrm{Z}$ *and* $(\hat{\boldsymbol{u}}_h, \hat{p}_h, \hat{\phi}_h) \in \mathrm{H}_h \times \mathrm{Q}_h \times \mathrm{Z}_h$ *be the solutions of the continuous and discrete problems* (3.1.4)-(3.1.6) *and* (3.1.9)-(3.1.11)*, respectively. Then, there exists a constant* $C > 0$*, independent of* h *and* λ*, such that*

$$|\boldsymbol{u} - \hat{\boldsymbol{u}}_h|_{1,\Omega} + ||p - \hat{p}_h||_{1,\Omega} + ||\phi - \hat{\phi}_h||_{0,\Omega} \leq Ch^2\big(||\boldsymbol{u}||_{3,\Omega} + ||p||_{3,\Omega} + ||\phi||_{2,\Omega}\big).$$

Proof. The result follows from Theorem 3.2.3 and the standard error estimates for the Lagrange interpolant of $\boldsymbol{u}$ and the scalar L^2-orthogonal projections for p and ϕ. □

To close this section, we observe that the well-posedness result obtained in Theorem 3.2.2 and the optimal convergence estimates obtained in Theorem 3.2.4 also hold if the displacement error $\boldsymbol{u} - \hat{\boldsymbol{u}}_h$ is measured in the full H^1-norm instead. Namely, recall that the equivalence between the norm $||\cdot||_{1,\Omega}$ and the seminorm $|\cdot|_{1,\Omega}$ on the function space H (which readily follows from the Poincaré Friedrichs inequality in combination with standard estimates) readily implies that there exist constants $C_0, C_1 > 0$, independent of h and λ, such that

$$C_0||\boldsymbol{u} - \hat{\boldsymbol{u}}_h||^2_{1,\Omega} \leq |\boldsymbol{u} - \hat{\boldsymbol{u}}_h|^2_{1,\Omega} \leq C_1||\boldsymbol{u} - \hat{\boldsymbol{u}}_h||^2_{1,\Omega},$$

which in turn ensures that the results obtained in Section 3.1 and Section 3.2 indeed hold for the full H^1-norm as well.

Chapter 4

Conclusions and future work

In this chapter, we summarise the work presented in this dissertation and discuss three subjects for further study. In particular, we introduce the following three extensions that build upon the mixed FVE methods developed in this dissertation: a three-field elasticity-poroelasticity interface problem, a posteriori estimates, and the numerical implementation of the second-order mixed hybrid FVE method for poroelasticity.

4.1 Future work

4.1.1 A three-field elasticity-poroelasticity interface problem

The competitive FVE methods proposed in this dissertation can be used to set up a three-field elasticity-poroelasticity interface problem, where on one side we consider the three-field formulation for elasticity introduced in Chapter 2 and on the other side we consider the three-field formulation for poroelasticity, see Chapter 3. Following the observation that the pore pressure variations and fluid content within the cap rock (commonly referred to as the non-pay zone) are often unaffected by the injection or extraction of fluids within the reservoir (see, for instance, [40]), we mention that such coupled elastic-poroelastic systems can in particular be used for reservoir modelling (where we consider the classical Biot equations [13] in the reservoir, whereas on the cap rock we solve only the elasticity system). We also stress that a careful set up of the interface conditions is required and refer to [33] for a detailed discussion of the physical intuition behind such transmission conditions.

We highlight that a Galerkin scheme for a coupled elasticity-poroelasticity interface problem has been developed in [40], where the domain decomposition on the interface is established by utilising mortar terms in combination with discontinuous

Galerkin jumps. Elaborating on the results presented in this dissertation, we could introduce a mixed FVE scheme instead, considering a fully monolithic coupling. In particular, we could use the lowest-order FVE method proposed in Chapter 2 in the cap rock, whereas in the reservoir we could utilise the stabilised $\mathbb{P}_1 - \mathbb{P}_1$ FVE method introduced in [64]. One important advantage of this approach is that we could straightforwardly generate numerical solutions for various benchmark tests with already implemented solvers. Lastly, we mention that stabilised, discontinuous and non-conforming FVE schemes have been developed for, for instance, the Stokes-Darcy and Navier-Stokes/Darcy coupled systems (see, for instance, [73, 76]).

4.1.2 A posteriori error estimates

A natural continuation of the a priori error analysis for the second-order mixed hybrid FVE method for poroelasticity developed in Chapter 3 would be to provide a reliable and efficient a posteriori error estimator based on residual terms. The amount of literature that focuses on a posteriori error analysis for (non)linear problems is nowadays quite extensive and includes various important contributions in recent years, but it is important to remark that the main ideas and techniques have been developed in the early works [2, 72] and references therein. Estimators of this kind are typically used to guide adaptive mesh refinement strategies to guarantee adequate convergence behaviour of the Galerkin approximations and play an important role in obtaining accurate numerical solutions for a variety of PDEs. A further advantage of such estimates is that they are fully computable after obtaining $(\hat{\boldsymbol{u}}_h, \hat{p}_h, \hat{\phi}_h)$ (up to some mesh independent constant), can significantly improve the quality of approximations for a given computational cost (cf. [2]), and guarantee convergence in the presence of complex geometries that could potentially lead to spurious solutions (see, for instance, [72]). From this point of view, introducing a reliable and efficient a posteriori estimator would be a valuable addition to the proposed analysis and could significantly increase the numerical accuracy of the proposed second-order mixed hybrid FVE method for poroelasticity.

4.1.3 Implementation of the mixed hybrid FVE scheme

The well-posedness and second order convergence rate of the proposed mixed hybrid FVE scheme for the approximation of a novel three-field formulation for poroelasticity has been established in Section 3.2, where it is also shown that the method is robust when the linear elastic skeleton becomes nearly incompressible. Evidently, a

natural continuation would be to verify its second order convergence rate as well as its robustness with respect to the Lamé constants numerically by performing a set of insightful tests. For instance, one could consider a 2D manufactured solution (see, for instance, [62]), the footing test [35], the swelling of a sponge [62], or the quasistatic leaking flow of unsaturated soil [71], where one could vary the Poisson ratio ν (or equivalently, the Lamé modulus of dilation, λ) to model a nearly incompressible response of the linear elastic skeleton.

We stress that the implementation of the second-order mixed hybrid FVE method introduced in Section 3.1 for the approximation of linear poroelastostatics requires the explicit construction of the control volumes and test spaces for the displacement, whereas numerical results for the linear FVE method introduced in Section 2.3 can be readily obtained by using the code for the associated lowest-order the FE method in combination with a different assembly of the forcing term on the right-hand side. In addition, we remark that meshes $\mathcal{T}_h$ satisfying the angular thresholds $\theta_{\min,K} \geq \theta_l$ and $\theta_{\max,K} \leq \theta_u$, for all $K \in \mathcal{T}_h$, can be generated straightforwardly by using, for instance, the FE mesh generators GMSH and Hypermesh. Although the approach mentioned above is rather laborious and definitely not trivial, it would be interesting to compare the performance of the proposed second-order FVE method with other discrete schemes for poroelastostatics available from the literature.

4.2 Conclusions

In this dissertation, we have developed a linear mixed FVE method for the approximation of a novel displacement-rotation-pressure formulation for linear elastostatics and a second-order mixed hybrid FVE method for the approximation of a new three-field formulation for poroelastostatics. In addition, a family FE methods for the approximation of the three-field displacement-rotation-pressure formulation for elastostatics was presented, based on our unpublished paper [3].

First, the unique solvability and stability of the novel displacement-rotation-pressure formulation for linear elastostatics was established using the well-known Babuška-Brezzi theory for saddle-point problems. Furthermore, invertibility and stability of the proposed family of FE methods and the linear FVE method was established and optimal a priori error estimates were derived, using norms that are robust with respect to the Lamé constants. In addition, an L^2-estimate for the displacement errors of the Galerkin and Petrov-Galerkin methods was obtained. The predicted accuracy and applicability of the proposed discretisation methods was verified by

conducting a set of three insightful computational tests. In particular, we verified that the linear FVE and lowest-order FE method have the same convergence properties and observed that the proposed three-field formulation and the corresponding (Petrov-)Galerkin schemes are particularly appealing for nearly incompressible materials. This outcome in combination with various other numerical results demonstrates that the proposed methods perform well in practice.

Second, we have obtained a novel three-field formulation for the approximation of linear poroelastostatics by recasting the Biot consolidation equations in terms of the solid displacement, the pore pressure of the fluid and the total pressure. The stability and unique solvability of the proposed three-field saddle-point formulation was addressed by Oyarzúa et al. [62], using Fredholm's alternative theory for compact operators. Furthermore, we have developed a second-order mixed hybrid FVE method related to the classical Taylor-Hood finite element pair, in the sense that the trial spaces correspond to conforming piecewise quadratic elements for the approximation of the displacements and the pore pressure of the fluid, whereas the total pressure is approximated using piecewise linear polynomials. Local coercivity of the bilinear form associated with the solid viscous stress has been established using the spectral properties of symmetrised local FVE stiffness matrices, which can only be established under suitable mesh geometric requirements. In addition, well-posedness of the proposed Petrov-Galerkin was established and its optimal second order convergence rate was derived, using a combination of Céa's estimates and standard interpolation properties. In particular, we found that the constants arising in the error bounds are independent of the Lamé modulus of dilation, λ, which must be understood as constants that may be of order $\mathcal{O}(\lambda^{-1})$ and may vanish in the incompressibility limit.

As mentioned previously, this dissertation paves way to further research on FVE methods in linear (poro)elastostatics. We have briefly introduced three possible extensions of the proposed methods, including an interface problem (where on one side of the interface we use the displacement-rotation-pressure for linear elasticity and on the other side of the interface we use the three-field formulation for poroelasticity), a posteriori estimates, and further numerical results for the proposed second-order mixed hybrid FVE method for the approximation of elastostatics. The problems described above in combination with various other questions will be the subject of further research.

Appendix A

The case of mixed boundary conditions

First, let us consider the Cauchy-Navier equations, now furnished with the mixed boundary conditions (2.1.8). Testing equation (2.1.4) against $\boldsymbol{v} \in \widetilde{\mathrm{H}} := \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)^d = \{\boldsymbol{v} \in \mathrm{H}^1(\Omega)^d : \boldsymbol{v}|_{\Gamma_\mathrm{D}} = \boldsymbol{0}\}$, and integrating by parts using (1.3.1), leads to

$$-(1+\eta)\int_\Omega p\,\mathrm{div}\,\boldsymbol{v}+(1+\eta)\int_{\Gamma_\mathrm{N}}\boldsymbol{v}\cdot(p\boldsymbol{n})+\sqrt{\eta}\int_\Omega \boldsymbol{\omega}\cdot\mathbf{curl}\,\boldsymbol{v}+\sqrt{\eta}\int_{\Gamma_\mathrm{N}}(\boldsymbol{\omega}\times\boldsymbol{n})\cdot\boldsymbol{v}=\int_\Omega \boldsymbol{f}\cdot\boldsymbol{v},$$

for all $\boldsymbol{v} \in \widetilde{\mathrm{H}}$. We then employ the definition of the traction $\boldsymbol{t}$ (see the second relation in (2.1.8)) to obtain

$$\begin{aligned}-(1+\eta)\int_\Omega p\,\mathrm{div}\,\boldsymbol{v}+(1+\eta)\int_{\Gamma_\mathrm{N}}\boldsymbol{v}\cdot(p\boldsymbol{n})+\sqrt{\eta}\int_\Omega \boldsymbol{\omega}\cdot\mathbf{curl}\,\boldsymbol{v}\\ +\int_{\Gamma_\mathrm{N}}[2\eta(\nabla\boldsymbol{u})\boldsymbol{n}-(1-\eta)p\boldsymbol{n}-\boldsymbol{t}]\cdot\boldsymbol{v}=\int_\Omega \boldsymbol{f}\cdot\boldsymbol{v},\end{aligned}$$

for all $\boldsymbol{v} \in \widetilde{\mathrm{H}}$.

Therefore, after employing equation (2.1.6) and rearranging terms, we arrive at the following modification of (2.1.9)-(2.1.10) incorporating mixed displacement-traction boundary conditions: Find $((\boldsymbol{\omega}, p), \boldsymbol{u}) \in (\mathrm{Z}\times\mathrm{Q})\times\widetilde{\mathrm{H}}$ such that

$$a((\boldsymbol{\omega},p),(\boldsymbol{\theta},q))+b((\boldsymbol{\theta},q),\boldsymbol{u})=0 \qquad \forall(\boldsymbol{\theta},q)\in\mathrm{Z}\times\mathrm{Q}, \tag{A.0.1}$$

$$b((\boldsymbol{\omega},p),\boldsymbol{v})-c(\boldsymbol{u},\boldsymbol{v})=F(\boldsymbol{v})-\int_{\Gamma_\mathrm{N}}\boldsymbol{t}\cdot\boldsymbol{v} \qquad \forall\boldsymbol{v}\in\widetilde{\mathrm{H}}, \tag{A.0.2}$$

where the additional diagonal bilinear form $c : \widetilde{\mathrm{H}}\times\widetilde{\mathrm{H}} \to \mathbb{R}$ is defined as

$$c(\boldsymbol{u},\boldsymbol{v}) := 2\eta\int_{\Gamma_\mathrm{N}}[\nabla\boldsymbol{u}-(\mathrm{div}\,\boldsymbol{u})\mathbf{I}]\boldsymbol{n}\cdot\boldsymbol{v}, \qquad \forall\boldsymbol{u},\boldsymbol{v}\in\widetilde{\mathrm{H}}. \tag{A.0.3}$$

Note that since $\boldsymbol{u} \in \widetilde{\mathrm{H}}$ and $\mathbf{div}(2\mu\boldsymbol{\varepsilon}(\boldsymbol{u})+\lambda\,\mathbf{div}\,\boldsymbol{u})$ is in $\mathrm{L}^2(\Omega)^d$, then $2\eta(\nabla\boldsymbol{u}\boldsymbol{n}-\mathrm{div}\,\boldsymbol{u}\boldsymbol{n})$ is in $\mathrm{H}^{-1/2}(\Gamma_{\mathrm{N}})^d$. Therefore, the bilinear form $c(\cdot,\cdot)$ is simply a duality pairing between $\mathrm{H}^{-1/2}$ and $\mathrm{H}^{1/2}$. A similar observation can be found in [6].

The Galerkin formulation will then adopt an analogous structure. In turn, the FVE method from Section 2.3 can be modified to incorporate mixed boundary conditions as follows. We consider the discrete space

$$\widetilde{\mathrm{H}}_h^\star := \big\{\boldsymbol{v} \in \mathrm{L}^2(\Omega)^d : \boldsymbol{v}|_{K_j^\star} \in \mathcal{P}_0(K_j^\star)^d \text{ for all } K_j^\star \in \mathcal{T}_h^\star,\ \boldsymbol{v}|_{K_j^\star} = \mathbf{0} \text{ if } s_j \text{ lies on } \Gamma_{\mathrm{D}}\big\}.$$

In view of discretising the Cauchy-Navier equations subject to mixed boundary conditions (2.1.8), we can test (2.1.4) against $\boldsymbol{v}_h^\star \in \widetilde{\mathrm{H}}_h^\star$ and integrate by parts, which leads to

$$(1+\eta)\sum_{j=1}^{|\mathcal{N}_h|}\int_{\partial K_j^\star} p_h\boldsymbol{v}_h^\star\cdot\boldsymbol{n} + \sqrt{\eta}\sum_{j=1}^{|\mathcal{N}_h|}\int_{\partial K_j^\star}(\boldsymbol{\omega}_h\times\boldsymbol{n})\cdot\boldsymbol{v}_h^\star = \sum_{j=1}^{|\mathcal{N}_h|}\int_{K_j^\star}\boldsymbol{f}\cdot\boldsymbol{v}_h^\star \quad \forall \boldsymbol{v}_h^\star \in \widetilde{\mathrm{H}}_h^\star.$$

Next, we employ an argument similar to the one used in the proof of Lemma 2.3.1, considering the edges that coincide with the boundary segment Γ_{N} separately. More precisely, by substituting $\mathcal{H}_h\boldsymbol{v}_h$ and by definition of the traction $\boldsymbol{t}$, we readily obtain

$$\begin{aligned}
&(1+\eta)\Big[\sum_{K\in\mathcal{T}_h}\int_{\partial K/\Gamma_{\mathrm{N}}} p_h\boldsymbol{v}_h\cdot\boldsymbol{n} + \sum_{j=1}^{|\mathcal{N}_h|}\int_{\partial K_j^\star\cap\Gamma_{\mathrm{N}}} p_h\mathcal{H}_h\boldsymbol{v}_h\cdot\boldsymbol{n}\Big]\\
&+\sqrt{\eta}\Big[\sum_{K\in\mathcal{T}_h}\int_{\partial K/\Gamma_{\mathrm{N}}}(\boldsymbol{\omega}_h\times\boldsymbol{n})\cdot\boldsymbol{v}_h + \sum_{j=1}^{|\mathcal{N}_h|}\int_{\partial K_j^\star\cap\Gamma_{\mathrm{N}}}(\boldsymbol{\omega}_h\times\boldsymbol{n})\cdot\mathcal{H}_h\boldsymbol{v}_h\Big]\\
&+\sum_{j=1}^{|\mathcal{N}_h|}\int_{\partial K_j^\star\cap\Gamma_{\mathrm{N}}}[2\eta(\nabla\boldsymbol{u}_h)\boldsymbol{n}-(1-\eta)p_h\boldsymbol{n}-\boldsymbol{t}]\cdot\mathcal{H}_h\boldsymbol{v}_h = \sum_{j=1}^{|\mathcal{N}_h|}\int_{K_j^\star}\boldsymbol{f}\cdot\mathcal{H}_h\boldsymbol{v}_h,
\end{aligned}$$

for all $\boldsymbol{v}_h \in \widetilde{\mathrm{H}}_h$. In order to simplify the previous expression further, we use that $\int_e(\mathcal{H}_h\boldsymbol{v}_h - \boldsymbol{v}_h) = 0$ for every $\boldsymbol{v}_h \in \widetilde{\mathrm{H}}_h$ and every edge e of each element $K \in \mathcal{T}_h$ (cf. [64]) in combination with the assumption that $p_h \in \mathrm{Q}_h$ and $\boldsymbol{\omega}_h \in \mathrm{Z}_h$ are both linear on each element $K \in \mathcal{T}_h$, which in turn implies that

$$\sum_{j=1}^{|\mathcal{N}_h|}\int_{\partial K_j^\star\cap\Gamma_{\mathrm{N}}} p_h\mathcal{H}_h\boldsymbol{v}_h\cdot\boldsymbol{n} = \sum_{K\in\mathcal{T}_h}\int_{\partial K\cap\Gamma_{\mathrm{N}}} p_h\boldsymbol{v}_h\cdot\boldsymbol{n}, \tag{A.0.4}$$

$$\sum_{j=1}^{|\mathcal{N}_h|}\int_{\partial K_j^\star\cap\Gamma_{\mathrm{N}}}(\boldsymbol{\omega}_h\times\boldsymbol{n})\cdot\mathcal{H}_h\boldsymbol{v}_h = \sum_{K\in\mathcal{T}_h}\int_{\partial K\cap\Gamma_{\mathrm{N}}}(\boldsymbol{\omega}_h\times\boldsymbol{n})\cdot\boldsymbol{v}_h, \tag{A.0.5}$$

where we have also used that the union of boundary edges of control volumes and the union of boundary edges of elements coincide. Consequently, after joining integrals by employing (A.0.4)-(A.0.5) and using identity (1.3.1), we obtain the following

modification of the FVE formulation when incorporating mixed displacement-traction boundary conditions: Find $((\hat{\boldsymbol{\omega}}_h, \hat{p}_h), \hat{\boldsymbol{u}}_h) \in (\mathrm{Z}_h \times \mathrm{Q}_h) \times \widetilde{\mathrm{H}}_h$ such that

$$\begin{aligned} a((\hat{\boldsymbol{\omega}}_h, \hat{p}_h), (\boldsymbol{\theta}_h, q_h)) + b((\boldsymbol{\theta}_h, q_h), \hat{\boldsymbol{u}}_h) &= 0, \\ b((\hat{\boldsymbol{\omega}}_h, \hat{p}_h), \boldsymbol{v}_h) - C(\hat{\boldsymbol{u}}_h, \boldsymbol{v}_h) &= \tilde{F}(\boldsymbol{v}_h) - \sum_{j=1}^{|\mathcal{N}_h|} \int_{\partial K_j^\star \cap \Gamma_\mathrm{N}} \boldsymbol{t} \cdot \mathcal{H}_h \boldsymbol{v}_h, \end{aligned}$$

for all $((\boldsymbol{\theta}_h, q_h), \boldsymbol{v}_h) \in (\mathrm{Z}_h \times \mathrm{Q}_h) \times \widetilde{\mathrm{H}}_h$, where the bilinear form $C : \widetilde{\mathrm{H}}_h \times \widetilde{\mathrm{H}}_h \to \mathbb{R}$ is defined as

$$C(\boldsymbol{u}_h, \boldsymbol{v}_h) := 2\eta \sum_{j=1}^{|\mathcal{N}_h|} \int_{\partial K_j^\star \cap \Gamma_\mathrm{N}} [\nabla \boldsymbol{u}_h - (\operatorname{div} \boldsymbol{u}_h)\mathbf{I}]\boldsymbol{n} \cdot \mathcal{H}_h \boldsymbol{v}_h, \qquad \forall \boldsymbol{u}_h, \boldsymbol{v}_h \in \widetilde{\mathrm{H}}_h.$$

Moreover, using a similar argument as before in combination with the assumption that $\boldsymbol{u}_h \in \widetilde{\mathrm{H}}_h$ is linear on each element $K \in \mathcal{T}_h$, we find that

$$C(\boldsymbol{u}_h, \boldsymbol{v}_h) - c(\boldsymbol{u}_h, \boldsymbol{v}_h) = 2\eta \sum_{K \in \mathcal{T}_h} \int_{\partial K \cap \Gamma_\mathrm{N}} [\nabla \boldsymbol{u}_h - (\operatorname{div} \boldsymbol{u}_h)\mathbf{I}]\boldsymbol{n} \cdot (\mathcal{H}_h \boldsymbol{v}_h - \boldsymbol{v}_h) = 0,$$

for all $\boldsymbol{u}_h, \boldsymbol{v}_h \in \widetilde{\mathrm{H}}_h$. In other words, also for mixed displacement-traction boundary conditions we find that the lowest-order FE and FVE schemes only differ by assembly of the right-hand side.

Appendix B

A modified FVE formulation for poroelasticity

The main goal of this section is to motivate why we incorporate gradients instead of strains in the formulation of the bilinear operator $A_1(\cdot,\cdot)$. The proofs obtained throughout this section closely follow the ideas from [25], where scalar elliptic boundary value problems are studied instead.

Let us once again consider the poroelasticity equations (3.1.1)-(3.1.2), but instead focus on the modified three-field formulation introduced in [62]. In this case, testing against adequate functions, integrating by parts, and taking into account the mixed boundary conditions (3.1.7) gives rise to the following mixed Galerkin formulation: Find $\boldsymbol{u}_h \in \mathrm{H}_h$, $p_h \in \mathrm{Q}_h$, $\phi_h \in \mathrm{Z}_h$ such that

$$\widehat{A}_1(\boldsymbol{u}_h, \boldsymbol{v}_h) + \widehat{B}_1(\boldsymbol{v}_h, \phi_h) = \tilde{F}(\boldsymbol{v}_h) \qquad \forall \boldsymbol{v}_h \in \mathrm{H}_h, \tag{B.0.1}$$

$$a_2(p_h, q_h) - b_2(q_h, \phi_h) = G(q_h) \qquad \forall q_h \in \mathrm{Q}_h, \tag{B.0.2}$$

$$b_1(\boldsymbol{u}_h, \psi_h) + b_2(p_h, \psi_h) - c(\phi_h, \psi_h) = 0 \qquad \forall \psi_h \in \mathrm{Z}_h, \tag{B.0.3}$$

where the definitions of the bilinear forms $a_2 : \mathrm{Q}_h \times \mathrm{Q}_h \to \mathbb{R}$, $b_1 : \mathrm{H}_h \times \mathrm{Z}_h \to \mathbb{R}$, $b_2 : \mathrm{Q}_h \times \mathrm{Z}_h \to \mathbb{R}$, $c : \mathrm{Z}_h \times \mathrm{Z}_h \to \mathbb{R}$ coincide with the bilinear forms arising from the corresponding second-order FE formulation (for their precise forms, we refer to [62]) and the bilinear forms $\widehat{A}_1 : \mathrm{H}_h \times \mathrm{H}_h \to \mathbb{R}$, $\widehat{B}_1 : \mathrm{H}_h \times \mathrm{Z}_h \to \mathbb{R}$ are instead defined as

$$\widehat{A}_1(\boldsymbol{u}_h, \boldsymbol{v}_h) := 2\mu \sum_{j=1}^{|\mathcal{N}_h|} \int_{K_j^\star} \boldsymbol{\varepsilon}(\boldsymbol{u}_h) : \boldsymbol{\varepsilon}(\mathcal{H}_h \boldsymbol{v}_h) - 2\mu \sum_{j=1}^{|\mathcal{N}_h|} \int_{\partial K_j^\star} \mathcal{H}_h \boldsymbol{v}_h \cdot (\boldsymbol{\varepsilon}(\boldsymbol{u}_h)\boldsymbol{n}),$$

$$\widehat{B}_1(\boldsymbol{v}_h, \phi_h) := \sum_{j=1}^{|\mathcal{N}_h|} \int_{\partial K_j^\star} \phi_h (\mathcal{H}_h \boldsymbol{v}_h \cdot \boldsymbol{n}) - \sum_{j=1}^{|\mathcal{N}_h|} \int_{K_j^\star} \phi_h \,\mathrm{div}(\mathcal{H}_h \boldsymbol{v}_h).$$

The results from [25, 26] readily imply that the bilinear operators $\widehat{A}_1(\cdot,\cdot)$, $\widehat{B}_1(\cdot,\cdot)$ are bounded and Körns first inequality also ensures that $\widehat{B}_1(\cdot,\cdot)$ satisfies a discrete inf-sup condition. However, it is unclear whether the bilinear operator $\widehat{A}_1(\cdot,\cdot)$ also satisfies a discrete coercivity condition.

In order to investigate this property further, we once again focus on the spectral properties of symmetrised local FVE stiffness matrices corresponding to the bilinear operator $\widehat{A}_1(\cdot,\cdot)$. For this purpose, we first review the definition of the bilinear form $a_1 : \mathrm{H}_h \times \mathrm{H}_h \to \mathbb{R}$ arising from the second-order mixed FE method introduced in [62], which is defined as

$$a_1(\boldsymbol{u}_h, \boldsymbol{v}_h) := 2\mu \int_\Omega \boldsymbol{\varepsilon}(\boldsymbol{u}_h) : \boldsymbol{\varepsilon}(\boldsymbol{v}_h).$$

Next, we decompose the global stiffness matrices $\widehat{\mathcal{A}}^{\mathrm{FVE}}$ and $\mathcal{A}^{\mathrm{FE}}$ corresponding to the mixed hybrid FVE formulation (B.0.1)-(B.0.3) and the associated second-order FE method, respectively, as follows:

$$\widehat{\mathcal{A}}^{\mathrm{FVE}} = \begin{bmatrix} \widehat{M}^{\bar{A}} & \widehat{M}^{C} \\ \widehat{M}^{B} & \widehat{M}^{D} \end{bmatrix}, \qquad \mathcal{A}^{\mathrm{FE}} = \begin{bmatrix} \widehat{M}^{A} & \widehat{M}^{B^T} \\ \widehat{M}^{B} & \widehat{M}^{D} \end{bmatrix},$$

where, after employing the definition of the transfer operator $\mathcal{H}_h(\cdot)$, the $2|\mathcal{N}_h| \times 2|\mathcal{N}_h|$ sub matrices can be obtained by

$$\begin{aligned} \widehat{M}^{A}_{i,j} &:= a_1(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j), \quad \widehat{M}^{\bar{A}}_{i,j} := \widehat{A}_1(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j), \quad \widehat{M}^{B}_{i,j} := \widehat{A}_1(\boldsymbol{\phi}_i, \boldsymbol{\psi}_j) = a_1(\boldsymbol{\phi}_i, \boldsymbol{\psi}_j), \\ \widehat{M}^{B^T}_{i,j} &:= a_1(\boldsymbol{\psi}_i, \boldsymbol{\phi}_j), \quad \widehat{M}^{C}_{i,j} := \widehat{A}_1(\boldsymbol{\psi}_i, \boldsymbol{\phi}_j), \quad \widehat{M}^{D}_{i,j} := \widehat{A}_1(\boldsymbol{\psi}_i, \boldsymbol{\psi}_j) = a_1(\boldsymbol{\phi}_i, \boldsymbol{\psi}_j). \end{aligned}$$

We highlight that the matrices $\widehat{M}^{A}$ and $\widehat{M}^{\bar{A}}$ are indeed equal, which can be easily verified by closely following the arguments from [78]. Furthermore, we once again define the symmetrisation of the FVE stiffness matrix as $\widehat{\mathcal{A}}^{\mathrm{S}} := \frac{1}{2}\big(\widehat{\mathcal{A}}^{\mathrm{FVE}} + (\widehat{\mathcal{A}}^{\mathrm{FVE}})^T\big)$.

The two upcoming lemmas form the basis for the main result of this section.

Lemma B.0.1. *Let $\mathscr{T}$ be a shape regular family of quasi-uniform triangulations, then there exists $\gamma^{\mathrm{FE}} \in [0, 1)$, depending only on the thresholds θ_l and θ_u, such that, for all $\mathcal{T}_h \in \mathscr{T}$ and every $\boldsymbol{u}_h \in \mathrm{H}_h$,*

$$(1 - \gamma^{\mathrm{FE}})\big(U_1^T \widehat{M}^{A} U_1 + U_2^T \widehat{M}^{D} U_2\big) \leq U^T \mathcal{A}^{\mathrm{FE}} U \leq (1 + \gamma^{\mathrm{FE}})\big(U_1^T \widehat{M}^{A} U_1 + U_2^T \widehat{M}^{D} U_2\big).$$

Proof. First, let $U = [U_1, U_2]^T$ denote the vector representation of the function $\boldsymbol{u}_h = \boldsymbol{u}_h^1 + \boldsymbol{u}_h^2$, for $\boldsymbol{u}_h^1 \in \mathrm{V}_{1,h}$ and $\boldsymbol{v}_h^2 \in \mathrm{W}_{2,h}$, such that $\mathcal{C}_{\mathcal{T}_h}(U_1, U_2) = \boldsymbol{u}_h$. We proceed to show that the kernel of $\mathcal{A}_K^{\mathrm{FE}}$ is contained in the localised linear subspace $\mathbb{V}_{1,h,K}$, which

denotes the coordinate space corresponding to the restriction of the discrete function space $\mathrm{V}_{1,h}$ to the triangular element K.

It is well-known that the kernel of the linear operator $\boldsymbol{\varepsilon}(\cdot)$ on the space of twice differentiable functions on some open bounded domain $K \subseteq \mathbb{R}^2$ is the space of rigid deformations, i.e. functions of the form $\boldsymbol{u}_h(\boldsymbol{x}) = U_0 + R\boldsymbol{x}$, where $U_0 \in \mathbb{R}^2$ and $R \in \mathbb{R}^{2\times 2}_{\mathrm{skew}}$ is skew-symmetric. Consequently, letting $a_{1,K}(\cdot,\cdot)$ denote the restriction of the bilinear operator $a_1(\cdot,\cdot)$ to the element K, we find that $a_{1,K}(\boldsymbol{u}_h, \boldsymbol{v}_h) = 0$, for all $\boldsymbol{v}_h \in \mathrm{H}_h$, implies that $\boldsymbol{u}_h$ is contained in the space of rigid deformations on K, which in turn implies that the associated coordinate vector $U = [U_1, U_2]$ is contained in $\mathbb{V}_{1,h,K}$. In particular, this means that the kernel of $\mathcal{A}^{\mathrm{FE}}_K$ is contained in the linear space $\mathbb{V}_{1,h,K}$.

Consequently, the strengthened Cauchy-Bunyakowsky-Schwarz inequality (see, for instance, [1]) implies that, for any $K \in \mathcal{T}_h \in \mathscr{T}$, there exists $\gamma_K \in [0,1)$ such that

$$\big(U_1^T \widehat{M}_K^{B^T} U_1\big)^2 \leq \gamma_K^2 \big(U_1^T \widehat{M}_K^A U_1\big)\big(U_2^T \widehat{M}_K^D U_2\big), \tag{B.0.4}$$

where we highlight that the constant γ_K depends continuously on the geometry of the element K, which can be shown by transferring back to the reference triangle [31].

Next, observe that the set $\mathcal{K}_0 := \mathcal{K} \cap \{K : \theta_1 \geq \theta_l,\ \theta_3 \leq \theta_u\}$ is a compact subset of the set of admissible triangles $\mathcal{K}$, as defined in Section 3.1. Hence, taking $\gamma^{\mathrm{FE}} := \sup_{K\in\mathcal{K}_0} \gamma_K \in [0,1)$ establishes a uniform version of inequality (B.0.4). Next, following the argument from [26], we remark that using the standard Cauchy inequality and summing over all $K \in \mathcal{T}_h$ establishes the second inequality of interest. The first inequality is established using a similar argument. □

The upcoming lemma establishes similar results for the symmetric matrix $\widehat{\mathcal{A}}^{\mathrm{S}}$.

Lemma B.0.2. *Let $\mathscr{T}$ be a shape regular family of quasi-uniform triangulations. If $\widehat{\mathcal{A}}^{\mathrm{S}}_K$ is positive semidefinite for all $K \in \mathcal{T}_h \in \mathscr{T}$, then there exists $\gamma^{\mathrm{S}} \in [0,1)$, depending only on the thresholds θ_l and θ_u, such that, for all $\mathcal{T}_h \in \mathscr{T}$ and every $\boldsymbol{u}_h \in \mathrm{H}_h$, one has*

$$(1-\gamma^{\mathrm{S}})\big(U_1^T \widehat{M}^A U_1 + U_2^T \widehat{M}^D U_2\big) \leq U^T \widehat{\mathcal{A}}^{\mathrm{S}} U \leq (1+\gamma^{\mathrm{S}})\big(U_1^T \widehat{M}^A U_1 + U_2^T \widehat{M}^D U_2\big).$$

Proof. First, recall that the kernel of the linear operator $\boldsymbol{\varepsilon}(\cdot)$ is the space of rigid deformations. In particular, for two-dimensional domains it is straightforward to see that the space of rigid deformations is spanned by the functions $\boldsymbol{e}_1(\boldsymbol{x}) = (1,0)$, $\boldsymbol{e}_2(\boldsymbol{x}) = (0,1)$ and $\boldsymbol{e}_3(\boldsymbol{x}) = (-\boldsymbol{x}^2, \boldsymbol{x}^1)$. This property readily implies that the matrices $\widehat{M}_K^A$, $\widehat{M}_K^B$ have rank three, that the matrix $\widehat{M}_K^C$ has rank two, and that the matrix

$\widehat{M}_K^D$ has full rank. In particular, observe that for any $\boldsymbol{v}_h \in \mathrm{H}_h$, one straightforwardly verifies that

$$\begin{aligned}
&\widehat{A}_1(\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2 + \boldsymbol{\phi}_3, \boldsymbol{v}_h) = \widehat{A}_1(\boldsymbol{\phi}_4 + \boldsymbol{\phi}_5 + \boldsymbol{\phi}_6, \boldsymbol{v}_h) = \widehat{A}_1(\boldsymbol{\phi}_2 - \boldsymbol{\phi}_5, \boldsymbol{v}_h) = 0,\\
&\widehat{A}_1(\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2 + \boldsymbol{\phi}_3, \boldsymbol{v}_h) = \widehat{A}_1(\boldsymbol{\phi}_4 + \boldsymbol{\phi}_5 + \boldsymbol{\phi}_6, \boldsymbol{v}_h) = \widehat{A}_1(\boldsymbol{\phi}_2 - \boldsymbol{\phi}_5, \boldsymbol{v}_h) = 0,\\
&\widehat{A}_1(\boldsymbol{v}_h, \boldsymbol{\phi}_1 + \boldsymbol{\phi}_2 + \boldsymbol{\phi}_3) = \widehat{A}_1(\boldsymbol{v}_h, \boldsymbol{\phi}_4 + \boldsymbol{\phi}_5 + \boldsymbol{\phi}_6) = 0,
\end{aligned}$$

where we have used the fact that the kernel of $\boldsymbol{\varepsilon}(\cdot)$ is spanned by $\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2 + \boldsymbol{\phi}_3 = \boldsymbol{e}_1$, $\boldsymbol{\phi}_4 + \boldsymbol{\phi}_5 + \boldsymbol{\phi}_6 = \boldsymbol{e}_2$ and $\boldsymbol{\phi}_2 - \boldsymbol{\phi}_5 = \boldsymbol{e}_3$. The identities displayed above readily imply that $\mathrm{Ker}(\widehat{\mathcal{A}}_K^{\mathrm{S}})$ is spanned by the vectors $V_1 := [1,1,1,0,0,0,\mathbf{0}_6]^T$ and $V_2 := [0,0,0,1,1,1,\mathbf{0}_6]^T$, which are both contained in the linear space $\mathbb{V}_{1,h,K}$. The remainder of the proof follows in a similar manner as the proof of Lemma B.0.1. □

We are now ready to state the main theorem of this section, which is essentially a modified version of Theorem 3.1.3.

Theorem B.0.3. *Let $\mathscr{T}$ be a shape regular family of quasi-uniform triangulations and suppose that the matrix $\widehat{\mathcal{A}}_K^{\mathrm{S}}$ is positive semidefinite, for every $K \in \mathcal{T}_h \in \mathscr{T}$. Then, there exists a constant c_{A_1}, depending only on the thresholds θ_l and θ_u, such that, for every $\boldsymbol{u}_h \in \mathrm{H}_h$,*

$$\widehat{A}_1(\boldsymbol{u}_h, \boldsymbol{u}_h) \geq 2\mu c_{A_1} |\boldsymbol{u}_h|_{1,\Omega}^2.$$

Proof. Let $\boldsymbol{u}_h = \boldsymbol{u}_h^1 + \boldsymbol{u}_h^2 \in \mathrm{H}_h$, then the corresponding vector representation $U = [U_1, U_2]$ satisfies

$$U^T \widehat{\mathcal{A}}^{\mathrm{FVE}} U = \big((\widehat{\mathcal{A}}^{\mathrm{FVE}})^T U\big)^T U = U^T (\widehat{\mathcal{A}}^{\mathrm{FVE}})^T U = U^T \widehat{\mathcal{A}}^{\mathrm{S}} U,$$

such that, after two straightforward applications of Lemma B.0.1 and Lemma B.0.2, we obtain

$$U^T \widehat{\mathcal{A}}^{\mathrm{FVE}} U = U^T \widehat{\mathcal{A}}^{\mathrm{S}} U \geq (1 - \gamma^{\mathrm{S}})\big(U_1^T \widehat{M}^A U_1 + U_2^T \widehat{M}^D U_2\big) \geq \frac{1 - \gamma^{\mathrm{S}}}{1 + \gamma^{\mathrm{FE}}} U^T \mathcal{A}^{\mathrm{FE}} U.$$

Furthermore, by the coercivity of the bilinear operator $a_1(\cdot,\cdot)$ (see, for instance, [62]), we find that there exists a constant $c_{A_1} > 0$, independent of h and λ, such that

$$A_1(\boldsymbol{u}_h, \boldsymbol{u}_h) = U^T \widehat{\mathcal{A}}^{\mathrm{FVE}} U \geq \frac{1 - \gamma^{\mathrm{S}}}{1 + \gamma^{\mathrm{FE}}} U^T \mathcal{A}^{\mathrm{FE}} U = \frac{1 - \gamma^{\mathrm{S}}}{1 + \gamma^{\mathrm{FE}}} a_1(\boldsymbol{u}_h, \boldsymbol{u}_h) \geq 2\mu c_{A_1} |\boldsymbol{u}_h|_{1,\Omega}^2,$$

which completes the proof. □

Following the argument from Section 3.1, we proceed to describe how one can express $\widehat{\mathcal{A}}_K^{\mathrm{S}}$ in terms of the geometric entities θ_1 and θ_3. First of all, recall that the definition of the infinitesimal strain tensor $\boldsymbol{\varepsilon}(\cdot)$ readily implies that

$$\boldsymbol{\varepsilon}(\boldsymbol{u}_h) : \boldsymbol{\varepsilon}(\boldsymbol{v}_h) = \tfrac{1}{4}\big(\nabla \boldsymbol{u}_h + \nabla \boldsymbol{u}_h^T\big) : \big(\nabla \boldsymbol{v}_h + \nabla \boldsymbol{v}_h^T\big) = \tfrac{1}{2}\nabla \boldsymbol{u}_h : \nabla \boldsymbol{v}_h + \tfrac{1}{2}\nabla \boldsymbol{u}_h : \nabla \boldsymbol{v}_h^T,$$

which enables one to decompose the 6×6 local sub matrices $\widehat{M}_K^A$, $\widehat{M}_K^B$, $\widehat{M}_K^C$ and $\widehat{M}_K^D$ as follows:

$$\begin{aligned} \widehat{M}_K^A &= M_K^A + M_K^{A+}, \qquad & \widehat{M}_K^B &= M_K^B + M_K^{B+}, \\ \widehat{M}_K^C &= M_K^C + M_K^{C+}, \qquad & \widehat{M}_K^D &= M_K^D + M_K^{D+}, \end{aligned}$$

where the sub matrices M_K^A, M_K^B, M_K^C and M_K^D are obtained in Section 3.1 and expressing the entries of the remaining matrices M_K^{A+}, M_K^{B+}, M_K^{C+} and M_K^{D+} in terms of (θ_1, θ_3) involves numerous laborious but elementary applications of identity (3.1.12), where we once again assume without loss of generality that $\mu = 1$, $\boldsymbol{x}_1 = (0, 0)$ and $\boldsymbol{x} = (1, 0)$, see Figure 3.1. Further details on the computation of $\widehat{\mathcal{A}}_K^{\mathrm{S}}$ are omitted for the sake of brevity.

We proceed to test for positive semidefiniteness by discretising $(0, 180°) \times (0, 180°)$ in the $\theta_1 - \theta_3$ plane by a uniform grid of mesh size 0.05. However, this time we compute the smallest eigenvalues $\lambda_1 = \lambda_1(\theta_1, \theta_3)$ of the symmetric matrix $\widehat{\mathcal{A}}_K^{\mathrm{S}}$ as a function of the minimum and maximum angles of the element K, where we set $\lambda_1 = 1$ outside of $\mathcal{K}$. Remark that we consider λ_1 instead of λ_3 to demonstrate that $\widehat{\mathcal{A}}_K^{\mathrm{S}}$ does not satisfy the positive semidefiniteness condition in Theorem B.0.3, for any admissible geometric configuration. Evidently, Figure B.1 indicates that the sufficiency condition stated in Theorem B.0.3 cannot be satisfied. Therefore, if we instead consider the original setting from [62], then we cannot ensure the coercivity of one particular bilinear operator arising in the mixed hybrid FVE formulation of interest.

To close this section, we stress that the condition introduced in Theorem B.0.3 is only sufficient and by no means a necessary condition. In fact, the cancellation of negative and zero eigenvalues could still occur when assembling the local stiffness matrices into the global stiffness matrix.

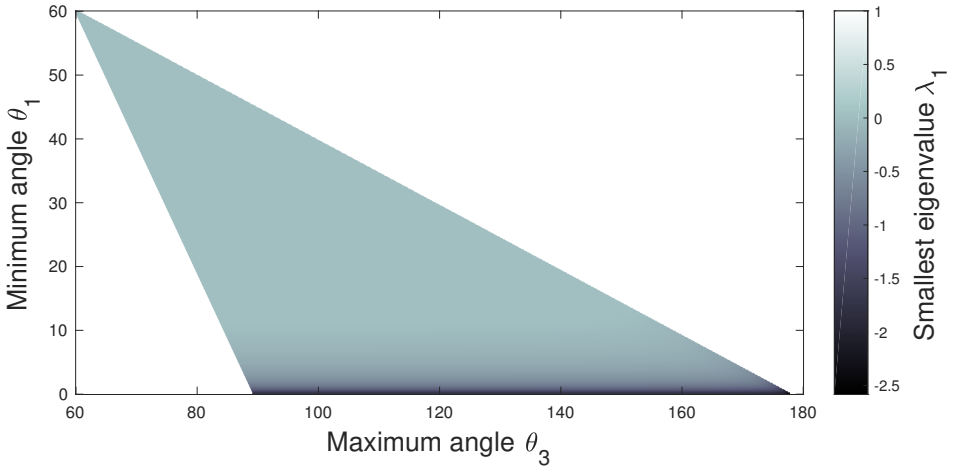


Figure B.1: The smallest eigenvalue $\lambda_1(\theta_1, \theta_3)$ of the symmetric matrix $\widehat{M}_K^{\mathrm{S}}$ plotted against the minimum angle θ_1 (y-axis) and the maximum angle θ_3 (x-axis).

References

[1] B. Achchab, O. Axelsson, L. Laayouni, and A. Souissi. Strengthened Cauchy-Bunyakowski-Schwarz inequality for a three-dimensional elasticity system. *Num. Lin. Alg. Appl.*, 8(3):191–205, 2001.

[2] M. Ainsworth and J. Oden. A posteriori error estimation in finite element analysis. *Comput. Methods Appl. Mech. Engrg.*, 142(1-2):1–88, 1997.

[3] V. Anaya, Z. De Wijn, D. Mora, and R. Ruiz-Baier. Mixed displacement-rotation-pressure formulations for elasticity. *Submitted*, 2017.

[4] U. Andelfinger and E. Ramm. EAS-elements for two-dimensional, three-dimensional, plate and shell structures and their equivalence to HR-elements. *Int. J. Numer. Methods Engrg.*, 36(8):1311–1337, 1993.

[5] D.N. Arnold, F. Brezzi, and M. Fortin. A stable finite element for the Stokes equations. *Calcolo*, 21:337–344, 1984.

[6] D.N. Arnold and R.S. Falk. A new mixed formulation for elasticity. *Numer. Math.*, 53(1-2):13–30, 1988.

[7] D.N. Arnold, R.S. Falk, and R. Winther. Mixed finite element methods for linear elasticity with weakly imposed symmetry. *Math. Comp.*, 76:1699–1723, 2007.

[8] R. Asadi, B. Ataie-Ashtiani, and C.T. Simmons. Finite volume coupling strategies for the solution of a Biot consolidation model. *Comput. Math. Appl.*, 55:494–505, 2014.

[9] R.E. Bank and D.J. Rose. Some error estimates for the box method. *SIAM J. Numer. Anal.*, 24:777–787, 1987.

[10] L. Beirão da Veiga, D. Mora, and R. Rodríguez. Numerical analysis of a locking-free mixed finite element method for a bending moment formulation of Reissner-Mindlin plate model. *Numer. Methods Partial Diff. Eqns.*, 29:40–63, 2013.

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

[12] E. Bertóti. Dual-mixed *hp* finite element methods using first-order stress functions and rotations. *Comput. Mech.*, 26:39–51, 2000.

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

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

[15] D. Boffi, M. Botti, and D. Di Pietro. A Nonconforming High-Order Method for the Biot Problem on General Meshes. *SIAM Journal on Scientific Computing*, 38(3):1508–1537, 2016.

[16] D. Boffi, F. Brezzi, and M. Fortin. Reduced symmetry elements in linear elasticity. *Commun. Pure Appl. Anal.*, 8(1):95–121, 2009.

[17] D. Braess. *Finite Elements. Theory, Fast Solvers, and Applications in Solid Mechanics*. Cambridge University Press, Cambridge, 2nd edition, 2001.

[18] D. Braess, C. Carstensen, and B.D. Reddy. Uniform convergence and a posteriori error estimators for the enhanced strain finite element method. *Numer. Math.*, 96(3):461–479, 2004.

[19] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Springer-Verlag, New York, 1991.

[20] Z. Cai. On the finite volume element method. *Numer. Math.*, 58:713–735, 1991.

[21] Z. Cai, J. Mandel, and S. McCormick. The finite volume element method for diffusion equations on general triangulations. *SIAM J. Numer. Anal.*, 28:392–402, 1991.

[22] Z. Cai, T. Manteuffel, and S. McCormick. First-order system least squares for the Stokes equations, with application to linear elasticity. *SIAM J. Numer. Anal.*, 34(5):1727–1741, 1997.

[23] Z. Cai and S. McCormick. On the accuracy of the finite volume element method for diffusion equations on composite grids. *SIAM J. Numer. Anal.*, 27:636–655, 1990.

[24] C. Carstensen, N. Nataraj, and A.K. Pani. Comparison results and unified analysis for first-order finite volume element methods for a Poisson model problem. *IMA J. Numer. Anal.*, 36(3):1120–1142, 2016.

[25] L. Chen. A New Class of High Order Finite Volume Methods for Second Order Elliptic Equations. *SIAM J. Numer. Anal.*, 47(6):4021–4043, 2010.

[26] Z. Chen, J. Wu, and Y. Xu. Higher-order finite volume methods for elliptic boundary value problems. *Adv. in Comp. Math.*, 37(2):191–253, 2012.

[27] Z. Chen, Y. Xu, and Y. Zhang. A construction of higher-order finite volume methods. *Math. Comp.*, 84(292):599–628, 2015.

[28] Z. Chen, Y. Xu, and Y. Zhang. A second-order hybrid finite volume method for solving the Stokes equation. *Appl. Numer. Math.*, 119:213–224, 2017.

[29] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems, Studies in Mathematics and its Applications*, volume 4. North-Holland, Amsterdam, 1978.

[30] J.K. Djoko, B.P. Lamichhane, B.D. Reddy, and B.I. Wohlmuth. Conditions for equivalence between the Hu-Washizu and related formulations and computational behavior in the incompressible limit. *Comput. Methods Appl. Mech. Engrg.*, 195(33-36):4161–4178, 2006.

[31] V. Eijkhout and P. Vassilevski. The role of the strengthened Cauchy-Buniakowskii-Schwarz inequality in multilevel methods. *SIAM Rev.*, 33(3):405–419, 1991.

[32] R.E. Ewing, R.D. Lazarov, and Y. Lin. Finite volume element approximations of nonlocal reactive flows in porous media. *Numer. Methods Partial Diff. Eqns.*, 16:285–311, 2000.

[33] M. Favino, C. Gross, M. Drolshagen, L. Keilig, J. Deschner, C. Bourauel, and R. Krause. Validation of a heterogeneous elastic-biphasic model for the numerical simulation of the PDL. *Comput. Methods Biomech. Biomed. Engrg.*, 16(5):544–553, 2013.

[34] J. Feistauer, M. Felcman and M. Lukáčová-Medvid'ová. Combined finite element-finite volume solution of compressible flow. *J. Comput. Appl. Math.*, 63(1):179–199, 1995.

[35] 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.*, 11:67–76, 2008.

[36] G.N. Gatica. Analysis of a new augmented mixed finite element method for linear elasticity allowing $\mathbb{RT}_0 - \mathbb{P}_1 - \mathbb{P}_0$ approximations. *ESAIM Math. Model. Numer. Anal.*, 40(1):1–28, 2006.

[37] G.N. Gatica. *A simple introduction to the mixed finite element method: Theory and applications*. Springer, Cham, 2014.

[38] G.N. Gatica, L.F. Gatica, and F. Sequeira. A priori and a posteriori error analyses of a pseudostress-based formulation for linear elasticity. *Comput. Math. Appl.*, 71(2):585–614, 2016.

[39] G.N. Gatica, A. Márquez, and S. Meddahi. An augmented mixed finite element method for 3D linear elasticity problems. *J. Comput. Appl. Math.*, 231:526–540, 2009.

[40] V. Girault, G. Pencheva, M.F. Wheeler, and T. Wildey. Domain decomposition for poroelasticity and elasticity with DG jumps and mortars. *Math. Models Methods Appl. Sci.*, 21(01):21–169, 2011.

[41] V. Girault and P. Raviart. *Finite element methods for Navier-Stokes equations: Theory and algorithms*. Springer-Verlag, Berlin, 1986.

[42] W. Hackbusch. On first and second order box schemes. *Computing*, 41:277–296, 1989.

[43] G. He, Y. He, and X. Feng. Finite volume method based on stabilized finite elements for the nonstationary Navier-Stokes problem. *Numer. Methods Part. Diff. Eqns.*, 23:1167–1191, 2007.

[44] H. Hu. On some variational principles in the theory of elasticity and the theory of plasticity. *Acta Phys. Sin.*, 10(3):259–290, 1954.

[45] T.J.R. Hughes. *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*. Prentice-Hall, New Jersey, 1987.

[46] T.J.R. Hughes, A. Masud, and I. Harari. Numerical assessment of some membrane elements with drilling degrees of freedom. *Comput. Struct.*, 55(2):297–314, 1995.

[47] E.P. Kasper and R.L. Taylor. A mixed-enhanced strain method, Part I: geometrically linear problems. *Comput. Struct.*, 75(3):237–250, 2000.

[48] B.P. Lamichhane. *Higher Order Mortar Finite Elements with Dual Lagrange Multiplier Spaces and Applications*. PhD thesis, Universität Stuttgart, Stuttgart, 2006.

[49] B.P. Lamichhane. Inf-sup stable finite-element pairs based on dual meshes and bases for nearly incompressible elasticity. *IMA J. Numer. Anal.*, 29:404–420, 2009.

[50] B.P. Lamichhane, B.D. Reddy, and B.I. Wohlmuth. Convergence in the incompressible limit of finite element approximations based on the Hu-Washizu formulation. *Numer. Math.*, 104(2):151–175, 2006.

[51] J. Li and Z. Chen. A new stabilized finite volume method for the stationary Stokes equations. *Adv. Comput. Math.*, 30:141–152, 2009.

[52] R. Li. On generalized difference methods for elliptic and parabolic differential equations. In K. Feng and J.L. Lions, editors, *Proceedings of the Symposium on the Finite Element Method Between China and France*, pages 323–360, Beijing, 1982. Science Press.

[53] R. Li. Generalized difference methods for a nonlinear Dirichlet problem. *SIAM J. Numer. Anal.*, 24:77–88, 1987.

[54] R. Li, Z. Chen, and W. Wu. A survey on generalized difference methods and their analysis. In Li Y. Micchelli C.A. Xu Y. Chen, Z., editor, *Advances in Computational Mathematics, Lecture Notes in Pure and Applied Mathematics*, volume 202, pages 321–337. CRC Press, 1999.

[55] R. Li, Z. Chen, and W. Wu. *Generalized Difference Methods for Differential Equations: Numerical Analysis of Finite Volume Methods*. Marcel Dekker, New York, 2000.

[56] R. Li and P. Zhu. Generalized difference methods for second order elliptic partial differential equations (I)-triangle grids. *Numer. Math. J. Chin. Univ.*, 2:140–152, 1982.

[57] F. Liebau. The finite volume element method with quadratic basis functions. *Computing*, 57(4):281–299, 1996.

[58] Y Lin, J Liu, and M Yang. Finite volume element methods : an overview on recent developments. *International journal of numerical analysis and modeling. Series B*, 4(1):14–34, 2013.

[59] Z. Luo and F. Teng. A fully discrete SCNFVE formulation for the non-stationary Navier-Stokes equations. *Comput. Model. Eng. Sci*, 101:33–58, 2014.

[60] A. Naumovich. On finite volume discretization of the three-dimensional Biot poroelasticity system in multilayer domains. *Comput. Methods Appl. Math.*, 6:306–325, 2006.

[61] R. Nicolaides. Existence, Uniqueness and Approximation for Generalized Saddle Point Problems. *SIAM Journal on Numerical Analysis*, 19(2):349–357, 1982.

[62] R. Oyarzúa and R. Ruiz-Baier. Locking-Free Finite Element Methods for Poroelasticity. *SIAM J. Numer. Anal*, 54(5):2951–2973, 2016.

[63] T.H.H. Pian and K. Sumihara. Rational approach for assumed stress finite elements. *Int. J. Numer. Methods Engrg.*, 20(9):1685–1695, 1984.

[64] A. Quarteroni and R. Ruiz-Baier. Analysis of a finite volume element method for the Stokes problem. *Numer. Math.*, 118:737–764, 2011.

[65] G. Romano, F. Marrotti de Sciarra, and M. Diaco. Well-posedness and numerical performances of the strain gap method. *Int. J. Numer. Methods Engrg.*, 51(1):103–126, 2001.

[66] R. Ruiz-Baier and I. Lunati. Mixed finite element – discontinuous finite volume element discretization of a general class of multicontinuum models. *J. Comput. Phys.*, 322:666–688, 2016.

[67] T. Schmidt. Box schemes on quadrilateral meshes. *Computing*, 51:271–292, 1993.

[68] J.C. Simo and M.S. Rifai. A class of assumed strain method and the methods of incompatible modes. *Int. J. Numer. Methods Engrg.*, 29:1595–1638, 1990.

[69] M. Sun and H. Rui. A coupling of weak Galerkin and mixed finite element methods for poroelasticity. *Comput. Math. Appl.*, 73:804–823, 2017.

[70] M. Tian and Z. Chen. Quadratic element generalized differential methods for elliptic equations. *Numer. Math. J. Chin. Univ.*, 13:99–113, 1991.

[71] R. Uzuoka and R.I. Borja. Dynamics of unsaturated poroelastic solids at finite strain. *Int. J. Numer. Anal. Meth. Geomech.*, 36:1535–1573, 2012.

[72] R. Verfürth. *A Review of A-Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*. Wiley-Teubner, Chichester, 1996.

[73] G. Wang, Y. He, and R. Li. Discontinuous finite volume methods for the stationary Stokes-Darcy problem. *Int. J. Numer. Methods Engrg.*, 107(5):395–418, 2015.

[74] K. Washizu. *Variational Methods in Elasticity & Plasticity*. Pergamon Press, New York, 1982.

[75] J. Wen, Y. He, and J. Yang. Multiscale enrichment of a finite volume element method for the stationary Navier–Stokes problem. *Int. J. Comp. Math.*, 90:1938–1957, 2013.

[76] Y. Wu and L. Mei. A non-conforming finite volume element method for the two-dimensional Navier-Stokes/Darcy system. *Comp. Appl. Math.*, 107(5):395–418, 6/2/2016.

[77] Y. Wu, X. Xie, and L. Chen. Hybrid stress finite volume method for linear elasticity problems. *Int. J. Numer. Anal. Model.*, 10(3):634–656, 2013.

[78] X. Ye. On the relationship between finite volume and finite element methods applied to the Stokes equations. *Numer. Methods Partial Diff. Eqns.*, 17:440–453, 2001.

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