Universidad de Concepción
Dirección de Postgrado
Facultad de Ciencias Físicas y Matemáticas
Programa de Doctorado en Ciencias Aplicadas
con Mención en Ingeniería Matemática

MÉTODOS MIXTOS DE ALTO ORDER EN MECÁNICA DEL MEDIO CONTINUO

(HIGH-ORDER MIXED METHODS IN CONTINUUM MECHANICS)

Tesis para optar al grado de Doctor en Ciencias
Aplicadas con mención en Ingeniería Matemática

Paulo Andrés Zúñiga Oyarzo
concepción-chile
2019

Profesor Guía: Manuel Solano Palma
CI^{2}MA y Departamento de Ingeniería Matemática
Universidad de Concepción, Chile

Cotutor: Ricardo Oyarzúa Vargas
GIMNAP–Departamento de Matemática y CI^{2}MA
Universidad del Bío-Bío y Universidad de Concepción, Chile

High-Order Mixed Methods in Continuum Mechanics

Paulo Andrés Zúñiga Oyarzo

Directores de Tesis: Manuel Solano, Universidad de Concepción, Chile.
Ricardo Oyarzúa, Universidad del Bío-Bío, Chile.

Director de Programa: Rodolfo Rodríguez, Universidad de Concepción, Chile.

Comisión evaluadora

Prof. Kent-Andre Mardal, University of Oslo, Norway.

Prof. Weifeng Qiu, City University of Hong Kong, China.

Prof. Ricardo Ruiz-Baier, University of Oxford, UK.

Prof. Tonatiuh Sánchez-Vizuet, Courant Institute of Mathematical Sciences, USA.

Prof. Ivan Yotov, University of Pittsburgh, USA.

Comisión examinadora

Firma: _____
Prof. Sergio Caucao, University of Pittsburgh, USA.

Firma: _____
Prof. Gabriel N. Gatica, Universidad de Concepción, Chile.

Firma: _____
Prof. Luis Gatica, Universidad Católica de la Santísima Concepción, Chile.

Firma: _____
Prof. Ricardo Oyarzúa, Universidad del Bío-Bío, Concepción, Chile.

Firma: _____
Prof. Manuel Solano, Universidad de Concepción, Chile.

Calificación: _____

Concepción, 17 de Diciembre de 2019

Abstract

The aim of this thesis is to develop high order mixed finite element discretizations for the numerical solution of partial differential equations arising from continuum mechanics, focusing on scenarios in which our methods contribute to improve the accuracy of the finite element approximation, namely, the treatment of curved domains and the presence of singularities or high gradients of the solution.

First, we propose a high order mixed finite element method for steady-state diffusion problems with Dirichlet boundary condition on a curved domain. Our approach is based on approximating the domain by a polyhedral computational subdomain where a high order Galerkin method is considered to compute the solution, and on a transferring technique to approximate the Dirichlet data on the computational boundary. Under suitable hypotheses on the distance between the curved and computational boundaries, and the finite dimensional subspaces, we prove the well-posedness of the resulting Galerkin scheme, and derive the corresponding error estimates, as well.

Next, we extend the previous ideas to the Stokes equations in which the pseudostress tensor and the fluid velocity are the only unknowns, whereas the fluid pressure is computed via a postprocessing technique. For the case where the computational boundary is constructed by interpolating the real boundary by a piecewise linear function, we also develop a reliable and quasi-efficient residual-based *a posteriori* error estimator. Its definition employs a more accurate approximate velocity to achieve the same rate of convergence of the method when the solution is smooth enough.

Finally, we present an error analysis of a conforming finite element discretization for a four-field formulation for the stationary Biot's consolidation model in poroelasticity. Assuming standard hypotheses on the discrete spaces, we first prove well-posedness and optimal *a priori* error estimates of the associated Galerkin scheme. Next, we develop a reliable and efficient residual-based *a posteriori* error estimator. We show that both the reliability and efficiency estimates are independent of the modulus of dilatation, even in the incompressible limit.

For all the problems described above, we provide numerical examples validating the theory.

Resumen

El objetivo de esta tesis es desarrollar discretizaciones de elementos finitos mixtos de alto orden para la solución numérica de ecuaciones diferenciales parciales que surgen de la mecánica del medio continuo, centrándose en escenarios en los que nuestros métodos contribuyen a mejorar la precisión de la aproximación de elementos finitos, a saber, el tratamiento de dominios curvos y la presencia de singularidades o altos gradientes de la solución.

Primero, proponemos un método de elementos finitos mixto de alto orden para problemas de difusión de estado estacionario con condición de contorno de Dirichlet sobre un dominio curvo. Nuestro enfoque se basa en aproximar el dominio por un subdominio computacional poliédrico donde se considera un método de Galerkin de alto orden para calcular la solución, y en una técnica de transferencia para aproximar el dato Dirichlet sobre la frontera computacional. Bajo hipótesis adecuadas sobre la distancia entre las fronteras curva y computacional, y los subespacios finito-dimensionales, demostramos el buen planteamiento del esquema de Galerkin resultante, y también obtenemos las estimaciones de error correspondientes.

A continuación, extendemos las ideas anteriores a las ecuaciones de Stokes en las que el tensor de pseudo-esfuerzo y la velocidad del fluido son las únicas incógnitas, mientras que la presión del fluido se calcula mediante una técnica de post-procesamiento. Para el caso en que la frontera computacional se construye interpolando la frontera real por una función lineal a trozos, también desarrollamos un estimador de error *a posteriori* residual, confiable y cuasi-eficiente. Su definición emplea una velocidad aproximada más precisa para lograr la misma tasa de convergencia del método cuando la solución es lo suficientemente suave.

Finalmente, presentamos un análisis de error de una discretización conforme de elementos finitos para una formulación de cuatro campos del modelo de consolidación de Biot estacionario en poroelasticidad. Asumiendo hipótesis estándar sobre los espacios discretos, primero demostramos el buen planteamiento y estimaciones de error *a priori* óptimas del esquema de Galerkin asociado. Luego, desarrollamos un estimador de error *a posteriori* residual, confiable y eficiente. Mostramos que tanto las estimaciones de confiabilidad como de eficiencia son independientes del módulo de dilatación, incluso en el límite incompresible.

Para todos los problemas descritos anteriormente, proporcionamos ejemplos numéricos que validan la teoría.

Agradecimientos

En primer lugar, agradezco a Dios por acompañarme en este camino. A mis padres, Nora y Dagoberto, por apoyarme a lo largo de toda mi vida, a pesar de las dificultades que hemos vivido como familia. Su resilencia ha sido la razón principal por la que nunca me he rendido ante la adversidad. A mis hermanos, Cristóbal y Alan, por aconsejarme y apoyarme a la distancia. A mi pareja, Cecilia, quien ha sido la brújula para llevar este trabajo a buen puerto. Agradezco su amor, alegría, consejos y, por sobre todo, su comprensión en los momentos difíciles.

Agradezco a mi director de tesis, el profesor Manuel Solano, por su paciencia, dedicación y consejos en cada reunión de trabajo. Su capacidad para tomar decisiones rápidas y certeras, además de su gran calidad humana, han influenciado positivamente en mi desarrollo personal y profesional. Gracias por confiar en mí y por ayudarme a concluir de buena forma este trabajo.

A mi co-tutor de tesis, el profesor Ricardo oyarzúa, por seguir colaborando conmigo después de haber sido mi director de tesis de Magíster en la Universidad del Bío-Bío. Su gran calidad humana y su profesionalismo como docente e investigador me inspirarón a seguir el camino que estoy recorriendo. Agradezco su paciencia en cada reunión, sus consejos y por insentivarme a seguir adelante.

My sincere thanks also goes to Prof. Sander Rhebergen (University of Waterloo) for the support he provided in the 3rd chapter of this work. I am particularly grateful for the time he and his research group spent with me during my stay in Canada.

I would like to extend my gratitude to professors Kent-Andre Mardal, Weifeng Qiu, Ricardo Ruiz-Baier, Tonatiuh Sánchez-Vizuet and Ivan Yotov, for being part of my evaluating committee, and to professors Sergio Caucao, Gabriel N. Gatica and Luis Gatica, for being part of my examination committee. Their valuable suggestions and comments helped to improve the quality of this manuscript.

Agradezco de manera especial al profesor Gabriel N. Gatica, en su calidad de director del Centro de Investigación en Ingeniería Matemática, CI^{2}MA, por darme la posibilidad de hacer uso de una oficina y otros espacios físicos con todas las comodidades que cualquier investigador soñaría. También agradezco su dedicación en las tres asignaturas que impartió a mi generación. Su habilidad para reducir ideas complejas a una simple frase me ayudarón a entender los métodos mixtos desarrollados en este trabajo.

A los profesores Raimund Büger y Rodolfo Rodriguez, por las gestiones realizadas como directores del programa de Doctorado en Ciencias Aplicadas c/m Ingeniería Matemática.

Al personal administrativo del CI^{2}MA y Departamento de Ingeniería Matemática de la Universidad de Concepción, Sra. Lorena Carrasco, Sra. Paola Castro, Sra. Cecilia Leiva, Sra. Micaela Ávila, Sr.

José Parra, Sr. Jorge Muñoz y Sr. Iván Tobar, por su buena disposición y buena onda.

A mis amigos(as) y compañeros(as) durante mi estadía en el Doctorado, el compa Ivancho, Eduardo, Mauricio, Víctor, Rodrigo, Rafa, Cristian, Bryan, Paul, Adrian, Néstor, William, Joaquín, Daniel, Camilo, Goga, Elvis, Patrick, Felipe (×2), Yolanda, Yissedt, Cinthya, María Carmen y Nitesh, por compartir conocimientos y conversaciones sobre el acontencer mundial, y por ser parte de extensas jornadas de trabajo académico y no-académico. A mi mejor aproximación a amigo Chileno, Sergio, por ayudarme a instalarme en Concepción en el 2013, cuando recién me alistaba a comenzar mis estudios de Magíster. Extiendo mi gratitud a mis ex-compañeros de Magíster de la Universidad del Bío-Bío, Mauricio Ascencio, José Oyarce y Daniela Carcamo, por apoyarme siempre.

Finalmente, agradezco a la Red Doctoral en Ciencias, Tecnología y Ambiente, REDOC CTA, al proyecto BASAL para apoyo a centros científicos y tecnológicos de excelencia a través del proyecto AFB 170001 del CMM, a CONICYT-Chile a través de la beca PFCHA/Doctorado/2016–21160446, y la dirección de Postgrado de la Universidad de Concepción, por haber financiar mi estadía en el doctorado.

Paulo Andrés Zúñiga Oyarzo

Contents

List of Tables

List of Figures

Introduction

Solving partial differential equations with the help of finite element (FE) methods is of key importance for scientific research and industrial development on a wide range of problems that arise from continuum mechanics. For instance, they have been used to describe two-phase flow [6], fluid flow in porous media [81] and natural convection phenomena [110].

At the beginning, finite element methods were typically associated to low order approximations, using mostly linear or quadratic polynomials. However, Babuška et al. [14] introduced the so-called p version of these methods, in which accuracy is improved by increasing the polynomial degree, p, while keeping a mesh of the domain fixed, opening the door to discussions about the best version of finite elements (see, e.g., [15, 101]). In particular, high order methods seem to be, at first glance, good candidates for applications requiring high-fidelity solutions (see, e.g., [65, 87]) and for large-scale problems where refining the mesh is still computationally quite costly.

Now, let us recall that mixed finite element methods started in the early fifties with papers related to structural engineering (see references in [128, Chapter 9]), and have gained considerable attention because, in addition to the original unknowns, they allow for a direct approximation of further variables of physical interest, such as rotations in linear elasticity (see, e.g., [4, 130]). Before them, obtaining such approximations was only possible through a post-processing of the system solution, although often at the expense of losing accuracy.

Mixed methods have been successfully applied to Stokes equations [46, 50, 78], Navier-Stokes equations [47], Brinkman equations [75], linear elasticity [4, 10, 13] and exterior problems [76, 77, 79], to name a few. In particular, the introduction of the pseudostress tensor (term coined by [45] in the context of least-squares methods) as an additional unknown in incompressible flow problems, has shown benefits from the point of view of implementation. In fact, there is no need for a symmetry condition as in the classical stress-based approach, allowing for an easy discretization via mixed finite elements developed for second-order elliptic partial differential equations (PDEs), i.e., Raviart–Thomas [118] and Brezzi–Douglas–Marini [39] spaces. For previous ideas regarding weakly imposed symmetry (via Lagrange multipliers) and PEERS finite elements, we refer the reader to the pioneering paper by Arnold et al. [10]. It is also worth mentioning that the need of locking-free finite element schemes for nearly incompressible materials (see, e.g., [16]) is the main reason why mixed methods have been introduced to solve elasticity problems. The latter is also important nowadays in the discretization of poroelasticity equations (see, e.g., [107, 114, 147]).

On the other hand, although FE methods are well-known in the numerical treatment of PDEs in regard to *a priori* error estimates to guarantee convergence, the accuracy of the numerical approxima-

tion may be affected by singularities or high gradients of the continuous solution, often as a result of domains with re-entrant corners or boundary layers. Adaptive mesh refinements based on *a posteriori* error estimators constitute a useful technique for error control in such situations (see, e.g., [139]), with much less computational cost than a uniform or quasiuniform refinement. To be precise, starting from a coarse mesh $\mathcal{T}_h$, the critical regions are marked to be refined according to a global estimator Θ in terms of local quantities Θ_T defined for all $T \in \mathcal{T}_h$. The global estimator is called reliable (resp. efficient) if it yields upper (resp. lower) bounds of the error up to high order terms. Both properties together verify that Θ is more or less equivalent to the error. Furthermore, an *a posteriori* error estimator is of residual type if it is obtained from the PDE residuals. We refer to [62, Section 1] for a complete review of residual-based *a posteriori* error estimators for mixed FE methods.

Another reason why standard FE methods lose accuracy arises from the discretization of PDEs in domains $\Omega \subseteq \mathbb{R}^d$ $(d = 2, 3)$ with curved boundary Γ. Indeed, since, in practice, the problem in Ω is solved approximately on a convenient computational domain D_h, the space in which one looks for the discrete solution is no longer a subspace of the continuous space and the price to pay for that is called *a variational crime*. Strang [131] was the first who studied this fact and established how to estimate the consistency error term introduced by the "crime". This term will usually be of low order and dominate the error analysis. To remedy this drawback and recover optimality, different numerical methods have been investigated since the seventies (see, e.g., [8, 20, 27, 28, 32, 42, 86, 97, 131, 132, 134]), all of them with advantages and disadvantages. In general, they can be classified as *fitted* or *unfitted*. In fitted methods, the mesh is matched or "fitted" to Γ with enough accuracy to reduce the error of consistency. For instance, if isoparametric finite elements in two dimensions are considered, each triangle at the computational boundary $\Gamma_h := \partial \mathrm{D}_h$ will have at most one curved side resulting from a local interpolation of Γ. This approach gives high order accuracy when the degree of the interpolation polynomial is large enough [96], but its use may increase the effort required for mesh generation, specially for complicated geometries or moving domains. By contrast, the idea behind unfitted methods, such as CutFEM [42] or immersed boundary methods [97, 111], is to make the construction of D_h as independent of Γ as possible. This can be done by immersing Ω in a *background mesh* and setting D_h to be the union of all the elements of the mesh that lie inside Ω. For this, one only needs an implicit description (via a level set function, for instance) of Ω. However, it is not easy to construct a high order unfitted method, mainly because the boundary data on Γ_h is imposed "away" from the true boundary.

A novel unfitted method for steady-state convection-diffusion with Dirichlet boundary conditions was developed in [59], and later analyzed in [57] in the framework of purely diffusive problems, using hybridizable discontinuous Galerkin (HDG) methods and a *transferring technique* proposed for one-dimensional problems [56]. Assuming that $\sigma := \nabla u$ is part of the system equations and $u = g$ on Γ, the method proposes to rewrite u on Γ_h by performing a line integration of σ along a family of segments, called *transferring paths*, joining both boundaries. Proceeding as for u and integrating the extrapolation of the discrete approximation of σ, an approximation of g on Γ_h is obtained. Thus, the problem is solved in D_h and its solution is extended by local extrapolations to the complementary region $\Omega \setminus \overline{\mathrm{D}_h}$. In [57], it has been shown that the method keeps high order accuracy when the distance $d(\Gamma, \Gamma_h)$ between Γ and Γ_h is only of order of the computational meshsize, say h. Fitted methods obtained through a piecewise linear interpolation of Γ, in which case $d(\Gamma, \Gamma_h)$ is of order of h^2, are also covered by this technique. Moreover, an extension of the method to Neumann boundary

conditions and for an elliptic transmission problem has been proposed by [117].

According to the above, our aim is to develop high order mixed finite element discretizations for the numerical solution of problems arising from continuum mechanics, focusing on scenarios in which our methods contribute to improve the accuracy of the finite element approximation. In particular, since, the transferring technique [56, 57, 59] has only been applied to HDG methods, we are interested in extending its applicability to dual-mixed formulations of elliptic PDEs, starting from purely diffusive problems. We also aim at proposing an adaptive method on curved domains approximated by computational subdomains, which, to our knowledge, has not received much attention until now. On the other hand, as far as polyhedral domains are concerned, the goal is to contribute in the direction of [92] and develop a reliable and efficient residual-based *a posteriori* error estimator for a four-field formulation for the stationary Biot's consolidation model.

This work is organized as follows. In **Chapter** 1, we propose and analyze a high order unfitted mixed method for diffusion problems with Dirichlet boundary condition. For this, the Dirichlet data is approximated on Γ_h by using the transferring technique described above. To deal with the boundedness of the consistency term introduced by the variational crime, we provide suitable hypotheses on the finite dimensional subspaces and the integration segments, ensuring that the resulting Galerkin scheme is well-posed. A feasible choice of discrete spaces is given by Raviart–Thomas elements of order $k \geq 0$ for the vector variable and discontinuous polynomials of degree k for the scalar variable, yielding optimal convergence of order h^{k+1} if the distance $d(\Gamma, \Gamma_h)$ is at most of order h. In addition, the solution is approximated on the complement of D_h and the corresponding error estimates are derived. This first contribution was published in the journal detailed below:

[109] R. Oyarzúa, M. Solano, and P. Zúñiga, *A High Order Mixed-FEM for Diffusion Problems on Curved Domains*, **J. Sci. Comput.**, 79 (2019), pp. 49–78.

Next, in **Chapter** 2, we extend the unfitted mixed method to the incompressible Stokes equations in which the pseudostress tensor and the fluid velocity are the only unknowns, whereas the fluid pressure is computed via a postprocessing technique. It is worth pointing out that, by contrast with related work by Solano and Vargas [126], here the novelties are, on the one hand, the treatment of the pseudostress approximation in D_h and, on the other hand, it is the first time that a residual-based *a posteriori* error estimator resulting from the transferring technique is analyzed, as long as Γ_h is constructed through a piecewise linear interpolation of the boundary Γ. Moreover, unlike the Stokes problem in polyhedral domains [78], our estimator is efficient up to calculable terms involving curved segments and a postprocessed velocity converging with one order higher than the original approximate velocity. This work has been recently accepted for publication in the journal **Computer Methods in Applied Mechanics and Engineering**. The preprint version is detailed below:

[108] R. Oyarzúa, M. Solano, and P. Zúñiga, *A priori and a posteriori error analyses of a high order unfitted mixed-FEM for Stokes flow*, Preprint 2019-15, Centro de Investigación en Ingeniería Matemática (CI^{2}MA), Universidad de Concepción, Chile, Preprint available at `https://ci2ma.udec.cl/publicaciones/prepublicaciones/prepublicacion.php?id=365`.

Finally, in **Chapter** 3, we focus on the numerical approximation of stationary Biot's consolidation

model in the theory of poroelasticity, describing the fluid-structure interaction of an elastic solid infiltrated by an interconnected network of fluid-saturated pores.

Let us briefly comment on the origins of the poroelasticity equations. In 1925, Karl von Terzaghi [133] proposed a simple mechanism to describe consolidation of soils, i.e., the gradual decrease of volume in fully saturated soils under loading. In such a case, the excess pore pressures dissipate and water leaves the soil slowly, resulting in the so-called *settlement of the soil*. Terzaghi's theory, although restricted to the one-dimensional consolidation problem, has several assumptions, including isotropy of the porous skeleton and small solid deformations. For the general theory, Maurice A. Biot later extended this problem to higher dimensions [25] and to the case of anisotropy [26].

In the displacement-pressure formulation of Biot's model, Darcy's law for the motion of the fluid is coupled to Hooke's theory of linear elasticity for the solid deformation. Due to the complex coupling, obtaining analytical solutions for this model is rarely possible (see, e.g., [18]) and, therefore, the development of efficient numerical solvers is indispensable. It is well-known, however, that the main difficulties encountered when developing numerical methods for Biot's model are volumetric locking and spurious, nonphysical pressure oscillations (see, e.g., [114, 147]).

Recently, Oyarzúa et al. [107] proposed and analyzed a three-field formulation for the stationary Biot's model using classical FE methods that are locking-free and free of spurious pressure oscillations. More precisely, in addition to the displacement and fluid pressure, they introduced the total pressure (or volumetric part of the total stress) as an additional unknown. To achieve a numerical scheme that is also mass conserving, they later extended this approach to a four-field formulation by introducing the "fluid flux" as an additional unknown [92]. They propose to approximate the solid displacement in this model by a discontinuous finite volume method while remaining unknowns are approximated by a mixed finite element method.

In **Chapter** 3, we present an *a priori* and *a posteriori* error analysis of a conforming FE discretization for the four-field formulation of stationary Biot's consolidation model [92]. For the *a priori* error analysis we provide suitable hypotheses on the corresponding finite dimensional subspaces ensuring that the associated Galerkin scheme is well-posed. We show that a suitable choice of subspaces is given by the Raviart–Thomas elements of order $k \geq 0$ for the fluid flux, discontinuous polynomials of degree k for the fluid pressure, and any stable pair of Stokes elements for the solid displacements and total pressure. We furthermore show that the scheme is locking-free. Next, we develop a reliable and efficient residual-based *a posteriori* error estimator. Both the reliability and efficiency estimates are shown to be independent of the modulus of dilatation. We remark that, up to our knowledge, this is the first work where efficiency estimates for high order approximations of stationary Biot's consolidation model are proven. The contents of this chapter gave rise to the following preprint:

[106] R. Oyarzúa, S. Rhebergen, M. Solano, and P. Zúñiga, *Error analysis of a conforming and locking-free four-field formulation for the stationary Biot's model.* Preprint 2019-31, Centro de Investigación en Ingeniería Matemática (CI^{2}MA), Universidad de Concepción, Chile, Preprint available at `https://ci2ma.udec.cl/publicaciones/prepublicaciones/prepublicacion.php?id=381`.

Introducción

Resolver ecuaciones diferenciales parciales con la ayuda de métodos de elementos finitos es clave para la investigación científica y el desarrollo industrial en una amplia gama de problemas que surgen de la mecánica del medio continuo. Por ejemplo, ellos han sido utilizados para describir flujos bifásicos [6], el flujo de fluidos en medios porosos [81] y fenómenos de convección natural [110].

En sus origenes, los métodos de elementos finitos se asociaban típicamente a aproximaciones de bajo orden, utilizando principalmente polinomios lineales o cuadráticos. Sin embargo, Babuška y colaboradores [14] introdujeron la versión p de estos métodos, en la cual se mejora la precisión de la aproximación al aumentar el grado polinomial, p, mientras se fija una malla del dominio, abriendo la puerta a discusiones sobre la mejor versión de elementos finitos (ver, por ejemplo, [15, 101]). En particular, los métodos de alto orden parecen ser, a primera vista, buenos candidatos para aplicaciones que requieren soluciones de alta fidelidad (ver, por ejemplo, [65, 87]) y para problemas a gran escala en los que refinar la malla sigue siendo muy costoso desde el punto de vista computacional.

Recordemos ahora que los métodos de elementos finitos mixtos comenzaron a principios de los años cincuenta con publicaciones relacionados con la ingeniería estructural (ver referencias en [128, Chapter 9]), y han ganado considerable atención porque, además de las incógnitas originales, permiten una aproximación directa de otras variables de interés físico, tales como rotaciones en elasticidad lineal (ver, por ejemplo, [4, 130]). Antes de ellos, la obtención de tales aproximaciones sólo era posible a través de un post-proceso de la solución del sistema, aunque a menudo a costa de perder precisión.

Los métodos mixtos han sido aplicados con éxito a las ecuaciones de Stokes [46, 50, 78], ecuaciones de Navier–Stokes [47], ecuaciones de Brinkman [75], elasticidad lineal [4, 10, 13] y problemas exteriores [76, 77, 79], por nombrar algunos. En particular, la introducción del tensor de pseudo-esfuerzo (término acuñado por [45] en el contexto de los métodos de mínimos cuadrados) como una incógnita adicional en problemas de flujo incompresible, ha mostrado beneficios desde el punto de vista de la implementación numérica. En efecto, no hay necesidad de una condición de simetría como en el enfoque clásico basado en el tensor de esfuerzos, lo que permite una fácil discretización a través de elementos finitos mixtos desarrollados para ecuaciones diferenciales parciales (EDPs) elípticas de segundo orden, es decir, espacios de Raviart–Thomas [118] y Brezzi–Douglas–Marini [39]. Para ideas previas sobre simetría débil (a través de multiplicadores de Lagrange) y elementos finitos PEERS, remitimos al lector al artículo pionero de Arnold y colaboradores [10]. También vale la pena mencionar que la necesidad de esquemas de elementos finitos libres de bloqueo para materiales casi incompresibles (ver, por ejemplo, [16]) es la razón principal por la cual los métodos mixtos han sido introducidos para resolver problemas de elasticidad. Esto último es igualmente importante hoy en día en la discretización de las ecuaciones de poroelasticidad (ver, por ejemplo, [107, 114, 147]).

Por otro lado, aunque los métodos de elementos finitos son bien conocidos en el tratamiento numérico de EDPs en lo que respecta a las estimaciones de error a priori para garantizar la convergencia, la precisión de la aproximación numérica puede verse afectada por singularidades o altos gradientes de la solución continua, a menudo como resultado de dominios con esquinas reentrantes o capas límite. Los refinamientos adaptativos basados en estimadores de error *a posteriori* constituyen una técnica útil para el control de errores en tales situaciones (ver, por ejemplo, [139]), con un costo computacional mucho menor que un refinamiento uniforme o cuasi-uniforme. Para ser preciso, a partir de una malla gruesa $\mathcal{T}_h$, las regiones críticas se marcan para ser refinadas de acuerdo a un estimador global Θ en términos de cantidades locales Θ_T definidas para todo $T \in \mathcal{T}_h$. El estimador global se dice confiable (resp. eficiente) si entrega cotas superiores (resp. inferiores) del error salvo, posiblemente, términos de alto orden. Ambas propiedades juntas verifican que Θ es más o menos equivalente al error. Además, se dice que un estimador de error *a posteriori* es del tipo residual si se obtiene de los residuos de la EDP. Remitimos al lector a [62, Section 1] para una revisión completa de los estimadores de error *a posteriori* del tipo residual para métodos de elementos finitos mixtos.

Otra razón por la cual los métodos de elementos finitos estándar pierden precisión surge de la discretización de EDPs en dominios $\Omega \subseteq \mathbb{R}^d$ $(d = 2, 3)$ con frontera curva Γ. En efecto, dado que, en la práctica, el problema en Ω se resuelve aproximadamente en un dominio computacional conveniente D_h, el espacio en el que se busca la solución discreta deja de ser un subespacio del espacio continuo y el precio a pagar se llama *crimen variacional*. Strang [131] fue el primero que estudió este hecho y estableció cómo estimar el término de error de consistencia introducido por el "crimen". Este término generalmente será de bajo orden y dominará el análisis de error. Para remediar este inconveniente y recuperar optimalidad, diferentes métodos numéricos han sido investigados desde los años setenta (ver, por ejemplo, [8, 20, 27, 28, 32, 42, 86, 97, 131, 132, 134]), todos ellos con ventajas y desventajas. En general, pueden ser clasificados como *fitted* o *unfitted*. En los métodos *fitted*, la malla se ajusta a Γ con precisión suficiente para reducir el error de consistencia. Por ejemplo, si se consideran elementos finitos isoparamétricos en dos dimensiones, cada triángulo en la frontera computacional $\Gamma_h := \partial \mathrm{D}_h$ tendrá como máximo un lado curvo resultante de una interpolación local de Γ. Este enfoque proporciona alto orden cuando el grado del polinomio de interpolación es lo suficientemente grande [96], pero su uso puede aumentar el esfuerzo requerido para la generación de mallas, especialmente para geometrías complicadas o dominios en movimiento. Por el contrario, la idea detrás de los métodos *unfitted*, tales como CutFEM [42] o los métodos *immersed boundary* [97, 111], es hacer que la construcción de D_h sea tan independiente de Γ como sea posible. Esto se puede hacer insertando Ω en una malla *background* y configurando D_h para que sea la unión de todos los elementos de la malla que se encuentran contenidos en Ω. Para esto, uno sólo necesita una descripción implícita (a través de una función de conjunto de nivel, por ejemplo) de Ω. Sin embargo, no es fácil construir un método *unfitted* de alto orden, principalmente porque el dato sobre la frontera Γ_h se impone "lejos" de la frontera real.

Un novedoso método *unfitted* para problemas de convección-difusión estacionarios con condiciones de Dirichlet fue propuesto en [59], y luego analizado en [57] en el contexto de problemas puramente difusivos, usando métodos de Galerkin discontinuo hibridizable (HDG) y una *técnica de transferencia* propuesta para problemas unidimensionales [56]. Asumiendo que $\sigma := \nabla u$ es parte de las ecuaciones del sistema y que $u = g$ sobre Γ, el método propone reescribir u sobre Γ_h a través de una integral de línea de σ a lo largo de una familia de segmentos, llamados *caminos de transferencia*, uniendo ambas fronteras. Procediendo como para u e integrando la extrapolación de la aproximación discreta de σ,

se obtiene una aproximación de g sobre Γ_h. De este modo, el problema se resuelve en D_h y su solución se extiende mediante extrapolaciones locales a la región restante de Ω. En [57], se ha demostrado que el método mantiene el alto orden de la aproximación cuando la distancia $d(\Gamma, \Gamma_h)$ entre Γ y Γ_h es sólo del orden del tamaño de la malla computacional, digamos h. Los métodos *fitted* obtenidos a través de una interpolación lineal a trozos de Γ, en cuyo caso $d(\Gamma, \Gamma_h)$ es de orden h^2, también están cubiertos por esta técnica. Además, una extensión del método a condiciones de contorno de Neumann y para un problema de transmisión elíptico ha sido propuesto por [117].

De acuedo a lo anterior, nuestro objetivo es desarrollar discretizaciones de elementos finitos mixtos de alto orden para la solución numérica de problemas que surgen de la mecánica del medio continuo, centrándose en escenarios en los que nuestros métodos contribuyen a mejorar la precisión de la aproximación de elementos finitos. En particular, dado que la técnica de transferencia [56, 57, 59] sólo se ha aplicado a los métodos HDG, estamos interesados en extender su aplicabilidad a las formulaciones duales-mixtas de EDPs elípticas, comenzando por problemas puramente difusivos. También pretendemos proponer un método adaptativo en dominios curvos aproximados por subdominios computacionales, el cual, hasta donde sabemos, no ha recibido mucha atención hasta ahora. Por otro lado, en lo que respecta a los dominios poliédricos, el objetivo es contribuir en la dirección de [92] y desarrollar un estimador de error *a posteriori* residual, confiable y eficiente, para una formulación de cuatro campos del modelo de consolidación de Biot estacionario.

Este trabajo está organizado de la siguiente manera. En el **Capítulo** 1, proponemos y analizamos un método mixto *unfitted* de alto orden para problemas de difusión con condición de contorno de Dirichlet. Para esto, el dato Dirichlet se aproxima sobre Γ_h utilizando la técnica de transferencia descrita anteriormente. Para lidiar con el acotamiento del término de consistencia introducido por el crimen variacional, proporcionamos hipótesis adecuadas sobre los subespacios de dimensión finita y los segmentos de integración, asegurando que el esquema de Galerkin resultante está bien puesto. Una opción factible de espacios discretos está dada por Raviart–Thomas de orden $k \geq 0$ para la variable vectorial y polinomios discontinuos de grado k para la variable escalar, entregando convergencia óptima de orden h^{k+1} si la distancia $d(\Gamma, \Gamma_h)$ es como máximo de orden h. Además, la solución se aproxima sobre el complemento de D_h y se derivan las estimaciones de error correspondientes. Esta primera contribución fue publicada en la revista que se detalla a continuación:

[109] R. Oyarzúa, M. Solano, and P. Zúñiga, *A High Order Mixed-FEM for Diffusion Problems on Curved Domains*, **J. Sci. Comput.**, 79 (2019), pp. 49–78.

A continuación, en el **Capítulo** 2, extendemos el método mixto *unfitted* a las ecuaciones de Stokes en las que el tensor de pseudo-esfuerzo y la velocidad del fluido son las únicas incógnitas, mientras que la presión del fluido se calcula mediante una técnica de post-procesamiento. Vale la pena señalar que, en contraste con el trabajo de Solano y Vargas [126], aquí las novedades son, por un lado, el tratamiento de la aproximación del pseudo-esfuerzo en D_h y, por otro lado, que es la primera vez que se analiza un estimador de error *a posteriori* del tipo residual que resulta de la técnica de transferencia, mientras que Γ_h se construye a través de una interpolación lineal a trozos de la frontera Γ. Además, a diferencia del problema de Stokes en dominios poliédricos [78], nuestro estimador es eficiente salvo términos calculables que involucran segmentos curvos y una velocidad post-procesada que converge con un orden más alto que la velocidad aproximada original. Este trabajo ha sido recientemente aceptado para publicación en la revista **Computer Methods in Applied Mechanics and Engineering**.

La pre-publicación se detalla a continuación:

[108] R. Oyarzúa, M. Solano, and P. Zúñiga, *A priori and a posteriori error analyses of a high order unfitted mixed-FEM for Stokes flow*, Preprint 2019-15, Centro de Investigación en Ingeniería Matemática (CI^2MA), Universidad de Concepción, Chile, Preprint available at `https://ci2ma.udec.cl/publicaciones/prepublicaciones/prepublicacion.php?id=365`.

Finalmente, en el **Capítulo** 3, nos enfocamos en la aproximación numérica del modelo de consolidación de Biot estacionario en la teoría de la poroelasticidad, describiendo la interacción fluido-estructura de un sólido elástico infiltrado por una red interconectada de poros saturados de fluido.

Comentemos brevemente sobre los orígenes de las ecuaciones de poroelasticidad. En 1925, Karl von Terzaghi [133] propuso un mecanismo simple para describir la consolidación de suelos, esto es, la disminución gradual del volumen en suelos (completamente saturados) debido a una carga aplicada. En tal caso, el exceso de presión del agua en los poros se disipa y el agua abandona el suelo lentamente, lo que resulta en el llamado *asentamiento del suelo*. La teoría de Terzaghi, aunque restringida al problema de consolidación unidimensional, tiene varios supuestos, incluyendo la isotropía del esqueleto poroso y pequeñas deformaciones en el sólido. Para la teoría general, Maurice A. Biot luego extendió este problema a tres dimensiones [25] y al caso de anisotropía [26].

En la formulación desplazamiento-presión del modelo de Biot, la ley de Darcy para el movimiento del fluido está acoplada a la teoría de la elasticidad lineal de Hooke para la deformación del sólido. Debido al complejo acoplamiento, rara vez es posible obtener soluciones analíticas para este modelo (ver, por ejemplo, [18]) y, por lo tanto, el desarrollo de *solvers* numéricos eficientes es indispensable. Sin embargo, es bien sabido que las principales dificultades encontradas al desarrollar métodos numéricos para el modelo de Biot son el bloqueo volumétrico y las oscilaciones de la presión (ver, por ejemplo, [114, 147]).

Recientemente, Oyarzúa y colaboradores [107] propusieron y analizaron una formulación de tres campos para el modelo de Biot estacionario utilizando métodos clásicos de elementos finitos que son libres de bloqueo y de oscilaciones de la presión. Más precisamente, además del desplazamiento y la presión del fluido, introdujeron la presión total (o parte volumétrica del esfuerzo total) como una incógnita adicional. Para lograr un esquema numérico que también conserve masa, más tarde los autores extendieron este enfoque a una formulación de cuatro campos al introducir el "flujo de fluido" como una incógnita adicional [92]. Ellos proponen aproximar el desplazamiento del sólido en este modelo por un método de volumenes finitos discontinuo, mientras que las incógnitas restantes se aproximan por un método mixto.

En el **Capítulo** 3, presentamos un análisis de error *a priori* y *a posteriori* de una discretización de elementos finitos conforme para la formulación de cuatro campos del modelo de consolidación de Biot estacionario [92]. Para el análisis de error *a priori* establecemos hipótesis adecuadas sobre los subespacios finito-dimensionales correspondientes, asegurando que el esquema de Galerkin asociado está bien puesto. Demostramos que una elección adecuada de subspacios está dada por Raviart–Thomas de orden $k \geq 0$ para el flujo de fluido, polinomios discontinuos de grado k para la presión de fluido, y cualquier par de elementos estables para Stokes en el caso de los desplazamientos del sólido y la presión total. Además, mostramos que el esquema es libre bloqueo. A continuación, desarrollamos un

estimador de error *a posteriori* del tipo residual y se demuesta que tanto las estimaciones de confiabilidad como de eficiencia son independientes del módulo de dilatación. Observamos que, hasta donde sabemos, este es el primer trabajo donde se demuestra la eficiencia del estimador para approximaciones de alto orden del modelo de consolidación de Biot estacionario. El contenido de este capítulo dio lugar a la siguiente pre-publicación:

[106] R. Oyarzúa, S. Rhebergen, M. Solano, and P. Zúñiga, *Error analysis of a conforming and locking-free four-field formulation for the stationary Biot's model.* Preprint 2019-31, Centro de Investigación en Ingeniería Matemática (CI^{2}MA), Universidad de Concepción, Chile, Preprint available at `https://ci2ma.udec.cl/publicaciones/prepublicaciones/prepublicacion.php?id=381`.

CHAPTER 1

A high order mixed-FEM for diffusion problems on curved domains

In this chapter we propose and analyze a high order mixed finite element method for diffusion problems with Dirichlet boundary condition on a domain Ω with curved boundary. The method is based on approximating Ω by a polyhedral subdomain where a high order Galerkin scheme is considered to compute the solution. We provide suitable hypotheses on the mesh near the curved boundary and the corresponding finite dimensional subspaces to achieve a well-posed discrete formulation with optimal *a priori* error estimates. Numerical experiments illustrate the performance of the scheme and validate the theory.

1.1 Introduction

Given $f \in \mathrm{L}^d(\Omega)$ and $g \in \mathrm{H}^{1/2}(\Gamma)$ we are interested in approximating, by a mixed finite element discretization, the vector field $\boldsymbol{\sigma}$ and the scalar field u satisfying the following first-order system of equations:

$$\boldsymbol{\sigma} = \nabla u \quad \text{in} \quad \Omega, \quad \operatorname{div} \boldsymbol{\sigma} = -f \quad \text{in} \quad \Omega, \quad u = g \quad \text{on} \quad \Gamma, \tag{1.1}$$

where $\Gamma := \partial\Omega$ is the boundary of Ω, which is assumed to be piecewise $\mathcal{C}^2$ and Lipschitz. Our approach is based on a technique originally developed in the context of high order hybridizable discontinuous Galerkin (HDG) methods [56, 57, 59]. It consists of approximating Ω by a polyhedral subdomain D_h, with boundary Γ_h, and transferring the Dirichlet boundary datum g from Γ to the computational boundary Γ_h, in such a way that the method keeps high order accuracy when D_h does not necessarily fit Ω. As we will detail below in Section 1.2.1, the transferred boundary datum on Γ_h, denoted by $\widetilde{g}$, is obtained by integrating $\boldsymbol{\sigma} = \nabla u$ along a family of segments joining Γ_h and Γ, which will be referred to as *transferring paths*. At the discrete level, $\widetilde{g}$ is approximated by a boundary datum $\widetilde{g}_h$ obtained by integrating the extrapolation of the discrete approximation of $\boldsymbol{\sigma}$ along the transferring paths. Thus, the problem is solved in D_h by means of any standard mixed method for polyhedral domains.

This technique, as mentioned before, has been introduced for HDG methods. It was first proposed and analyzed for the one-dimensional case in [56]. The approach was extended in [59] to two dimensions where numerical evidence indicated that the method performs optimally. Later, the authors in [57] proved that the method converges with optimal order in two and three dimensions under assumptions regarding the transferring paths. In addition, this technique has been successfully applied to

convection-diffusion problems [60], exterior diffusion equations [58], the Stokes flow problem [126], the semi-linear Grad–Shafranov equation [122], and the Oseen equations [125]. We point out that in all these papers the distance $d(\Gamma_h, \Gamma)$ between Γ_h and Γ is only of the order of the meshize h and there is no need of fitting the domain Ω. On the other hand, also in the context of HDG methods, [117] applied this technique to a diffusion problem with mixed boundary conditions and to an elliptic transmission problem where the interface is not piecewise flat. In these two cases, the boundary/interface needs to be interpolated by a piecewise linear computational boundary/interface in order to obtain high order accuracy, which means that the distance between the computational boundary/interface and the true boundary/interface has to be at most of order h^2. The reason why this approach works for the Dirichlet problem under a less restrictive assumption than the Neumann problem ($d(\Gamma_h, \Gamma)$ of order h versus order h^2) relies on the fact that the PDE provides a way to determine the Dirichlet data at the computational boundary through performing a line integration of the equation $\boldsymbol{\sigma} = \nabla u$. An appropriate transferring procedure of the Neumann datum, allowing $d(\Gamma_h, \Gamma)$ to be of order h, remains as an open problem.

On the other hand, a variety of numerical methods dealing with curved boundaries or interfaces have been proposed since the seventies; most of them provide low order approximations. In general, they can be classified in two groups: *fitted* and *unfitted* methods. Fitted methods adjust the computational boundary to Γ. For example, Γ_h can be constructed by a linear interpolation of Γ and the boundary data is transferred in a *natural way*, i.e., if $x \in \Gamma_h$ and $\bar{x} \in \Gamma$ is a projection of x in Γ, then $\widetilde{g}(x) := g(\bar{x})$. We recall that $\widetilde{g}$ denotes the boundary data on Γ_h. This idea, which was first introduced in [35] and then extended to interface problems in [36], leads to a low order approximation. To achieve a high order approximation in the context of fitted methods, an alternative procedure is to use isoparametric finite elements (see, e.g., [96]). However, these meshes are not easy to construct, especially for complicated geometries or when dealing with moving domains. On the contrary, unfitted methods, such as the immersed boundary method, allow us to work with background meshes, which is useful in complicated geometries. Nevertheless, since the boundary of the resulting polygonal domain is "far" from the curved boundary, the boundary data must be incorporated differently from the classical approaches. We refer the reader to [57, Section 1] for a review of unfitted methods, including the work [33, 97, 102, 112].

The method presented in this chapter can be classified as an unfitted method, where the boundary data is transferred in such a way that optimal high order accuracy is achieved. To the best of our knowledge, this technique has only been applied to HDG methods. Therefore, the purpose of our work is to consider this approach to the context of dual–mixed formulations of elliptic problems. The literature regarding mixed methods in polygonal/polyhedral domains is extensive. For instance, we refer the reader to [41] and [73] for a detailed analysis of mixed methods applied to different problems. However, in the context of curved domains the literature is scarce. Up to the author's knowledge, probably the only works dealing with mixed methods in curved domains are [22] and [23], where a parametric Raviart–Thomas finite elements for domains with curved boundaries is employed. In particular, the authors in [22] take an approximate polygonal domain Ω_h instead of Ω, where the paramatric Raviart–Thomas space of order $k \geq 0$ is constructed through the standard Piola transformation (see, e.g., [73]) and a piecewise polynomial mapping F_h of degree $k+1$ from a reference domain $\widehat{\Omega}_h$ to Ω_h. They then define the parametric Raviart–Thomas interpolation operator from the classical interpolation operator associated to $\widehat{\Omega}_h$ and the mapping F_h. Furthermore, they extended that operator to the complement of D_h by using a piecewise polynomial representation, providing

high order approximation properties under suitable regularity assumptions on the exact solution. The last approach was successfully applied to a mixed formulation of the Poisson problem with Neumman boundary condition, retaining the optimal convergence in the high order case for domains with piecewise $\mathcal{C}^{k+2}$ boundary (for $k \geq 0$), provided that the solution and the data are regular enough. On the contrary, in our method we can relax the smoothness requirement of the boundary to piecewise $\mathcal{C}^2$ only, and, besides, we take into account Dirichlet boundary conditions, in which case, to the best of our knowledge, a parametric-type method allowing high order approximation for mixed problems has not been presented yet.

This chapter is organized as follows. In the remainder of this section we recall notation and general definitions. Then, the domain Ω is approximated by a polyhedral subdomain where a Galerkin scheme is introduced and analyzed in Section 1.2. In Section 1.3, we derive the corresponding *a priori* error analysis whenever the distance $\mathrm{d}(\Gamma, \Gamma_h)$ is at most of $\mathcal{O}(h)$. Next, in Section 1.4 we make precise the definition of the involved discrete spaces and recall some approximation properties. In Section 1.5, several numerical examples illustrating the good performance of the method, are reported.

We end this section by introducing definitions and notation. In the sequel, when no confusion arises, $|\cdot|$ will denote the Euclidean norm in $\mathbb{R}^n$, $n = 2, 3$. Additionally, in what follows we utilize standard simplified terminology for Sobolev spaces and norms, where spaces of vector-valued functions are denoted in bold face. In particular, if $\mathcal{O}$ is a domain in $\mathbb{R}^n$, Σ is an open or closed Lipschitz curve (respectively surface in $\mathbb{R}^3$), and $s \in \mathbb{R}$, we define $\mathbf{H}^s(\mathcal{O}) := [\mathrm{H}^s(\mathcal{O})]^n$ and $\mathbf{H}^s(\Sigma) := [\mathrm{H}^s(\Sigma)]^n$. However, when $s = 0$ we write $\mathbf{L}^2(\mathcal{O})$ and $\mathbf{L}^2(\Sigma)$ instead of $\mathbf{H}^0(\mathcal{O})$ and $\mathbf{H}^0(\Sigma)$, respectively. The corresponding norms are denoted by $\|\cdot\|_{s,\mathcal{O}}$ for $\mathrm{H}^s(\mathcal{O})$, $\mathbf{H}^s(\mathcal{O})$, and $\|\cdot\|_{s,\Sigma}$ for $\mathrm{H}^s(\Sigma)$ and $\mathbf{H}^s(\Sigma)$. For $s \geq 0$, we write $|\cdot|_{s,\mathcal{O}}$ for the $\mathbf{H}^s$-seminorm and H^s-seminorm. In addition, we define the Sobolev space (see, e.g., [41, 73, 84]):

$$\mathbf{H}(\mathrm{div}; \mathcal{O}) := \left\{ \boldsymbol{\tau} \in \mathbf{L}^2(\mathcal{O}) : \ \mathrm{div}\, \boldsymbol{\tau} \in \mathrm{L}^2(\mathcal{O}) \right\},$$

equipped with the norm $\|\boldsymbol{\tau}\|_{\mathrm{div},\mathcal{O}} := \left(\|\boldsymbol{\tau}\|_{0,\mathcal{O}}^2 + \|\mathrm{div}\, \boldsymbol{\tau}\|_{0,\mathcal{O}}^2 \right)^{1/2}$, where the divergence operator, div, is understood in the sense of distributions, that is,

$$\langle \mathrm{div}\, \boldsymbol{\tau}, \varphi \rangle_{\mathscr{D}'(\mathcal{O}) \times \mathscr{D}(\mathcal{O})} := - \int_{\mathcal{O}} \boldsymbol{\tau} \cdot \nabla \varphi \, d\mathbf{x} \quad \forall\, \varphi \in \mathscr{D}(\mathcal{O}) := \mathcal{C}_0^\infty(\mathcal{O}),$$

with $\langle \cdot, \cdot \rangle_{\mathscr{D}'(\mathcal{O}) \times \mathscr{D}(\mathcal{O})}$ being the distributional paring between $\mathscr{D}'(\mathcal{O})$ and $\mathscr{D}(\mathcal{O})$. Note that if $\boldsymbol{\tau} \in \mathbf{H}(\mathrm{div}; \mathcal{O})$, then $\boldsymbol{\tau} \cdot \boldsymbol{\nu}_{\partial\mathcal{O}} \in \mathrm{H}^{-1/2}(\partial\mathcal{O})$, where $\boldsymbol{\nu}_{\partial\mathcal{O}}$ denotes the outward unit vector normal to the boundary $\partial\mathcal{O}$ and $\mathrm{H}^{-1/2}(\partial\mathcal{O})$ corresponds to the dual space of $\mathrm{H}^{1/2}(\mathcal{O})$. Hereafter, $\langle \cdot, \cdot \rangle_{\partial\mathcal{O}}$ denotes the duality pairing between $\mathrm{H}^{-1/2}(\partial\mathcal{O})$ and $\mathrm{H}^{1/2}(\partial\mathcal{O})$ with respect to the $\mathrm{L}^2(\partial\mathcal{O})$-inner product.

Finally, by $\mathbf{0}$ we will refer to the generic null vector (including the null functional and operator), and we will denote by C and c, with or without subscripts, bars, tildes or hats, generic constants independent of the meshsize, but might depend on the polynomial degree, the shape-regularity of the triangulation and the domain. Moreover, for quantities A and B, we write $A \lesssim B$, whenever there exists $C > 0$ such that $A \leq CB$.

1.2 The Galerkin method

In this section we derive our numerical scheme and analyze its well-posedness. Throughout this section, by the sake of simplicity of notation and exposition, we will consider the two-dimensional case. Most of the results are straightforward for three dimensions, but some of them require technicalities that will be addressed in Appendix B. We begin by introducing some notation and auxiliary results.

1.2.1 Notation and preliminaries

For the sake of completeness and easy presentation of the main ideas, we start by briefly recalling the mixed formulation of the Poisson problem, which reads: Find $(\boldsymbol{\sigma}, u) \in \mathbf{H}(\mathrm{div};\Omega) \times \mathrm{L}^2(\Omega)$ such that

$$\begin{aligned} a(\boldsymbol{\sigma},\boldsymbol{\tau}) + b(\boldsymbol{\tau},u) &= G(\boldsymbol{\tau}) \quad \forall\, \boldsymbol{\tau} \in \mathbf{H}(\mathrm{div};\Omega), \\ b(\boldsymbol{\sigma},v) &= F(v) \quad \forall\, v \in \mathrm{L}^2(\Omega), \end{aligned} \tag{1.2}$$

where the bilinear forms $a : \mathbf{H}(\mathrm{div};\Omega) \times \mathbf{H}(\mathrm{div};\Omega) \to \mathbb{R}$, $b : \mathbf{H}(\mathrm{div};\Omega) \times \mathrm{L}^2(\Omega) \to \mathbb{R}$, and the linear functionals $G : \mathbf{H}(\mathrm{div};\Omega) \to \mathbb{R}$, $F : \mathrm{L}^2(\Omega) \to \mathbb{R}$ are defined by

$$a(\boldsymbol{\sigma},\boldsymbol{\tau}) := \int_\Omega \boldsymbol{\sigma}\cdot\boldsymbol{\tau}\, d\mathbf{x}, \quad b(\boldsymbol{\tau},v) := \int_\Omega v\, \mathrm{div}\,\boldsymbol{\tau}\, d\mathbf{x}, \quad G(\boldsymbol{\tau}) := \langle \boldsymbol{\tau}\cdot\boldsymbol{\nu}_\Gamma, g\rangle_\Gamma, \quad F(v) := -\int_\Omega f\, v\, d\mathbf{x}.$$

Here $\boldsymbol{\nu}_\Gamma$ stands for the outward unit normal to Γ. For the well-posedness analysis of this problem we refer the reader to [73, Chapter 2].

Next, to derive our numerical method, from now on we suppose that Ω can be approximated by a family of polygonal subdomains D_h. In doing so, the most natural choice, guided by [59, Section 2.1], consists of considering a *background domain* $\mathcal{B} \supset \Omega$ easy to triangulate. More precisely, given a mesh T_h of $\overline{\mathcal{B}}$ made up of triangles K of diameter h_K, we use a level set function φ to determine which elements are inside of Ω in order to set our subdomain D_h; see an illustration in Figure 1.1. Here $\varphi : \mathcal{B} \to \mathbb{R}$ is a continuous function such that $\varphi < 0$ in Ω, $\varphi = 0$ in Γ and $\varphi > 0$ in $\mathcal{B}\setminus\overline{\Omega}$. We then set $\mathcal{T}_h := \{K \in \mathrm{T}_h :\ \varphi(\mathbf{x}) \le 0\ \ \forall\, \mathbf{x} \in K\}$ and $\mathrm{D}_h := \left(\cup_{K\in\mathcal{T}_h} \overline{K}\right)^\circ$. We also set $\Gamma_h := \partial\mathrm{D}_h$ and $\mathrm{D}_h^c := \Omega \setminus \overline{\mathrm{D}_h}$.

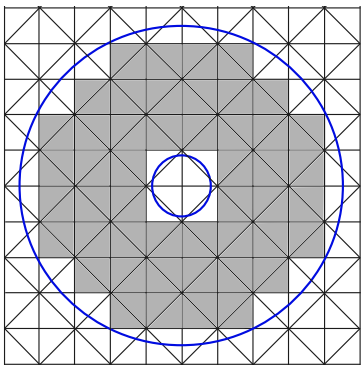


Figure 1.1: Example of curved domain Ω (annulus of boundary Γ), a background domain $\mathcal{B}$, and corresponding polygonal subdomain D_h. (figure produced by the author)

Now, we introduce notation associated with the sets introduced above. Hereafter, h denotes the meshsize of the triangulation $\mathcal{T}_h$ of $\overline{\mathrm{D}_h}$, that is, $h := \max\{h_K : K \in \mathcal{T}_h\}$. In addition, we denote by $\mathcal{E}_h$ the set of all edges of $\mathcal{T}_h$, subdivided as follows:

$$\mathcal{E}_h = \mathcal{E}_h^0 \cup \mathcal{E}_h^\partial,$$

where $\mathcal{E}_h^0 := \{e \in \mathcal{E}_h : e \subseteq \mathrm{D}_h\}$ and $\mathcal{E}_h^\partial := \{e \in \mathcal{E}_h : e \subseteq \Gamma_h\}$. Finally, for all K, $\boldsymbol{\nu}_K$ will denote the the unit outward normal vector on the boundary ∂K. However, to emphasize that a unit vector is normal to Γ_h or to an edge e of K, we will write $\boldsymbol{\nu}_{\Gamma_h}$ or $\boldsymbol{\nu}_e$, respectively. In the remainder of the chapter, we will drop the subscripts when referring to outward normal vectors whenever no confusion will occur.

In the computational domain D_h, the solution of (1.2) satisfies in a distributional sense,

$$\boldsymbol{\sigma} = \nabla u \quad \text{in} \quad \mathrm{D}_h, \quad \operatorname{div}\boldsymbol{\sigma} = -f \quad \text{in} \quad \mathrm{D}_h. \tag{1.3}$$

Moreover, by the first equation in (1.3), the trace of u on Γ_h, denoted by $\widetilde{g}$, can be written as

$$\widetilde{g}(\mathbf{x}) := \overline{g}(\mathbf{x}) - \int_{\mathscr{C}(\mathbf{x})} \boldsymbol{\sigma}\cdot\mathbf{m}(\mathbf{x})\, dr, \tag{1.4}$$

where $\mathscr{C}(\mathbf{x})$ is, in principle, any path starting at $\mathbf{x} \in \Gamma_h$ and ending at $\widetilde{\mathbf{x}} \in \Gamma$, $\mathbf{m}(\mathbf{x})$ is the unit tangent vector of $\mathscr{C}(\mathbf{x})$, and $\overline{g}(\mathbf{x}) := g(\widetilde{\mathbf{x}}(\mathbf{x}))$. In Section 1.2.2 we specify a construction of a suitable family of paths. Note that the value of $\widetilde{g}$ is independent of the integration path since it comes from integrating $\boldsymbol{\sigma} = \nabla u$. In addition, it is easy to see that the solution of (1.2) also satisfies

$$\begin{aligned} a_h(\boldsymbol{\sigma},\boldsymbol{\tau}) + b_h(\boldsymbol{\tau},u) - \langle \boldsymbol{\tau}\cdot\boldsymbol{\nu}_{\Gamma_h}, \widetilde{g}\rangle_{\Gamma_h} &= 0 \qquad \forall\, \boldsymbol{\tau} \in \mathbf{H}(\operatorname{div};\mathrm{D}_h), \\ b_h(\boldsymbol{\sigma}, v) &= F_h(v) \quad \forall\, v \in \mathrm{L}^2(\mathrm{D}_h), \end{aligned} \tag{1.5}$$

where the bilinear forms $a_h : \mathbf{H}(\operatorname{div};\mathrm{D}_h)\times\mathbf{H}(\operatorname{div};\mathrm{D}_h) \to \mathbb{R}$ and $b_h : \mathbf{H}(\operatorname{div};\mathrm{D}_h)\times\mathrm{L}^2(\mathrm{D}_h) \to \mathbb{R}$, and the functional $F_h : \mathbf{H}(\operatorname{div};\mathrm{D}_h) \to \mathbb{R}$, are given by

$$a_h(\boldsymbol{\sigma},\boldsymbol{\tau}) := \int_{\mathrm{D}_h} \boldsymbol{\sigma}\cdot\boldsymbol{\tau}\, d\mathbf{x}, \quad b_h(\boldsymbol{\tau}, v) := \int_{\mathrm{D}_h} v \operatorname{div}\boldsymbol{\tau}\, d\mathbf{x}, \quad F_h(v) := -\int_{\mathrm{D}_h} f\, v\, d\mathbf{x}. \tag{1.6}$$

We end this section by mentioning that, instead of using standard mixed methods to provide a Galerkin scheme for (1.2), we aim to propose a Galerkin scheme for (1.5), under a suitable approximation of the Dirichlet data on the boundary Γ_h, denoted by $\widetilde{g}_h$, allowing a high order approximation and keeping high order accuracy when the distance between Γ and Γ_h is of only order h. Before doing that, we proceed analogously to [57] and construct the aforementioned family of transferring paths.

1.2.2 Family of transferring paths

We now summarize the procedure introduced in [59] to construct the family of transferring paths $\{\mathscr{C}(\mathbf{x})\}_{\mathbf{x}\in\Gamma_h}$ connecting Γ_h and Γ. Let $\mathbf{u}$ and $\mathbf{v}$ be the vertices of a boundary edge e, $\mathbf{x}$ be a point on e and K^e the only element of $\mathcal{T}_h$ where e belongs. We first determine points $\widetilde{\mathbf{u}}$ and $\widetilde{\mathbf{v}}$ in Γ associated to $\mathbf{u}$ and $\mathbf{v}$, respectively:

***Step 1*:** For the vertex $\mathbf{u}$, we suggest two approaches to define $\tilde{\mathbf{u}}$.

- One possibility is to use the algorithm proposed in [59, Section 2.4.1] that uniquely determines a point $\tilde{\mathbf{u}}$ as the closest point to $\mathbf{u}$ such that $\mathscr{C}(\mathbf{u})$ does not intersect any other path and does not intersect the interior of the domain D_h. In Figure 1.2 (left) we display an illustration where $\tilde{\mathbf{u}}$ is the point in Γ associated to $\mathbf{u}$.
- An alternative is to assume that Γ is $\mathcal{C}^2$ and the mesh is fine enough. In this case $\tilde{\mathbf{u}}$ can be set as the orthogonal projections of $\mathbf{u}$ onto Γ.

Let $\widehat{\mathbf{m}}^{\mathbf{u}} := \tilde{\mathbf{u}} - \mathbf{u}$. We set $\mathbf{m}^{\mathbf{u}} := \widehat{\mathbf{m}}^{\mathbf{u}}/|\widehat{\mathbf{m}}^{\mathbf{u}}|$ if $|\widehat{\mathbf{m}}^{\mathbf{u}}| \neq 0$ and $\mathbf{m}^{\mathbf{u}} = \boldsymbol{\nu}_e$, otherwise. To define $\tilde{\mathbf{v}}$ and $\mathbf{m}^{\mathbf{v}}$ we proceed similarly.

Then, for a point $\mathbf{x} \in e$, which is not a vertex,

***Step 2*:** $\mathscr{C}(\mathbf{x})$ is determined as a convex combination of those paths originated from the vertices of e. More precisely, for $\theta \in [0,1]$, we write $\mathbf{x} = \mathbf{u}(1-\theta) + \theta\mathbf{v}$ and define $\widehat{\mathbf{m}} := \mathbf{m}^{\mathbf{u}}(1-\theta) + \theta\mathbf{m}^{\mathbf{v}}$. Then, we write $\mathbf{m} := \widehat{\mathbf{m}}/|\widehat{\mathbf{m}}|$ if $|\widehat{\mathbf{m}}| \neq 0$ and $\mathbf{m} := \boldsymbol{\nu}_e$, otherwise. Thus, we set $\tilde{\mathbf{x}}$ as the intersection between the boundary Γ and the ray starting at $\mathbf{x}$ whose unit tangent vector is $\mathbf{m}$; see Figure 1.2 (right) for an illustration.

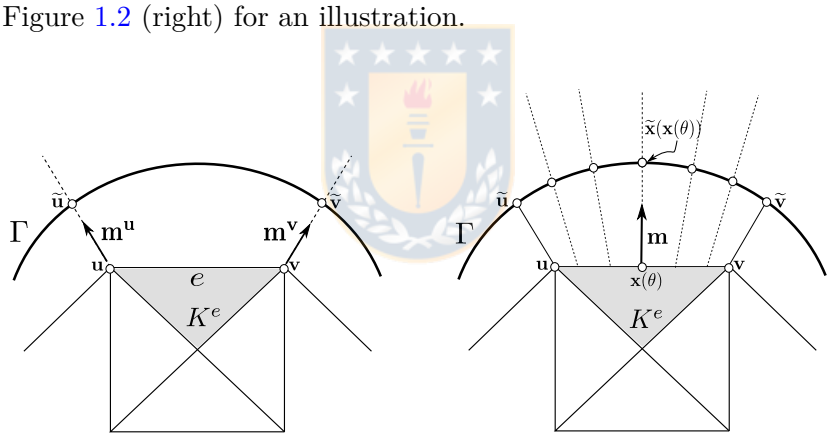


Figure 1.2: *Transferring paths* from a boundary edge e. (figure produced by the author)

Consequently, the transferring path connecting a point $\mathbf{x} \in \Gamma_h$ to the point $\tilde{\mathbf{x}} := \mathbf{x} + \ell(\mathbf{x})\mathbf{m} \in \Gamma$, where $\ell(\mathbf{x}) := |\tilde{\mathbf{x}} - \mathbf{x}|$, is given by

$$\mathscr{C}(\mathbf{x}) := \{\mathbf{x} + t\mathbf{m} : \ t \in [0, \ell(\mathbf{x})]\}\,. \tag{1.7}$$

Furthermore, for each edge $e \in \mathcal{E}_h^{\partial}$ with vertices $\mathbf{u}$ and $\mathbf{v}$, we define $\widetilde{K}_{ext}^e$ as the region enclosed by the intersection of D_h^c with the cones (see Figure 1.3):

$$C_1 := \left\{\mathbf{u} + \eta_1(\tilde{\mathbf{u}} - \mathbf{u}) + \eta_2(\mathbf{v} - \mathbf{u}) : \ \eta_1,\, \eta_2 \in \mathbb{R}^+\right\},$$
$$C_2 := \left\{\mathbf{v} + \eta_1(\tilde{\mathbf{v}} - \mathbf{v}) + \eta_2(\mathbf{u} - \mathbf{v}) : \ \eta_1,\, \eta_2 \in \mathbb{R}^+\right\},$$

and denote by $\widetilde{\mathcal{T}}_h := \left\{\widetilde{K}_{ext}^e : \ e \in \mathcal{E}_h^{\partial}\right\}$ the partition of D_h^c, satisfying $\overline{\mathrm{D}_h^c} = \bigcup\left\{\widetilde{K}_{ext}^e : \ e \in \mathcal{E}_h^{\partial}\right\}$.

1.2.3 Statement of the Galerkin scheme

Let us introduce generic finite dimensional subspaces $\mathbf{H}_h(\mathrm{D}_h)$ and $\mathrm{Q}_h(\mathrm{D}_h)$ of $\mathbf{H}(\mathrm{div};\mathrm{D}_h)$ and $\mathrm{L}^2(\mathrm{D}_h)$, respectively. On each $K \in \mathcal{T}_h$, we let $(\mathbf{M}(K), \mathrm{W}(K))$ be a pair of arbitrary finite dimensional subspaces, where $\mathbf{M}(K)$ is the space of two-dimensional vector functions on K, and $\mathrm{W}(K)$ is the space of scalar functions on K. Then, our approach consists of approximating the exact solution $(\boldsymbol{\sigma}, u)$ by a pair $(\boldsymbol{\sigma}_h, u_h)$ belonging to the product space $\mathbf{H}_h(\mathrm{D}_h) \times \mathrm{Q}_h(\mathrm{D}_h)$, where

$$\begin{aligned} \mathbf{H}_h(\mathrm{D}_h) &:= \left\{ \boldsymbol{\tau}_h \in \mathbf{H}(\mathrm{div};\mathrm{D}_h) : \ \boldsymbol{\tau}_h\big|_K \in \mathbf{M}(K) \ \ \forall\, K \in \mathcal{T}_h \right\}, \\ \mathrm{Q}_h(\mathrm{D}_h) &:= \left\{ v_h \in \mathrm{L}^2(\mathrm{D}_h) : \ v_h\big|_K \in \mathrm{W}(K) \ \ \forall\, K \in \mathcal{T}_h \right\}. \end{aligned} \tag{1.8}$$

A feasible choice of $(\mathbf{M}(K), \mathrm{W}(K))$ will be specified in Section 1.4. Inspired now by (1.4), for any $\mathbf{x}$ lying in $e \in \mathcal{E}_h^\partial$, $\widetilde{g}$ can be approximated by

$$\widetilde{g}_h(\mathbf{x}) := \overline{g}(\mathbf{x}) - \int_0^{\ell(\mathbf{x})} \mathbf{E}_h(\boldsymbol{\sigma}_h)(\mathbf{x} + t\mathbf{m}) \cdot \mathbf{m}\, dt, \tag{1.9}$$

where $\mathbf{E}_h(\boldsymbol{\sigma}_h)$ is a local extension operator from K^e to $\widetilde{K}^e_{ext}$ acting on $\boldsymbol{\sigma}_h$. In practice, since $\mathbf{M}(K)$ is a space of polynomials, given $\boldsymbol{\zeta}_h \in \mathbf{M}(K)$, we consider $\mathbf{E}_h(\boldsymbol{\zeta}_h)$ as the extrapolation of $\boldsymbol{\zeta}_h$ from K^e to $\widetilde{K}^e_{ext}$. In this way, defining now

$$d_h(\boldsymbol{\zeta}_h, \boldsymbol{\tau}_h) := \sum_{e \in \mathcal{E}_h^\partial} \int_e \left(\int_0^{\ell(\mathbf{x})} \mathbf{E}_h(\boldsymbol{\zeta}_h)(\mathbf{x} + t\mathbf{m}) \cdot \mathbf{m}\, dt \right) \boldsymbol{\tau}_h \cdot \boldsymbol{\nu}_e\, dS_{\mathbf{x}} \tag{1.10}$$

and

$$G_h(\boldsymbol{\tau}_h) := \sum_{e \in \mathcal{E}_h^\partial} \int_e \overline{g}\, \boldsymbol{\tau}_h \cdot \boldsymbol{\nu}_e\, dS_{\mathbf{x}} \tag{1.11}$$

for all $\boldsymbol{\zeta}_h$, $\boldsymbol{\tau}_h \in \mathbf{H}_h(\mathrm{D}_h)$, the Galerkin scheme of (1.5) reads: Find $(\boldsymbol{\sigma}_h, u_h) \in \mathbf{H}_h(\mathrm{D}_h) \times \mathrm{Q}_h(\mathrm{D}_h)$ such that

$$\begin{aligned} (a_h + d_h)(\boldsymbol{\sigma}_h, \boldsymbol{\tau}_h) + b_h(\boldsymbol{\tau}_h, u_h) &= G_h(\boldsymbol{\tau}_h) \quad \forall\, \boldsymbol{\tau}_h \in \mathbf{H}_h(\mathrm{D}_h), \\ b_h(\boldsymbol{\sigma}_h, v_h) &= F_h(v_h) \quad \forall\, v_h \in \mathrm{Q}_h(\mathrm{D}_h), \end{aligned} \tag{1.12}$$

where the bilinear forms a_h, b_h and the functional F_h were defined in Section 1.2.1. We remark that problem (1.12) can be seen as the discrete version of problem (1.5) where $\widetilde{g}$ has been approximated by $\widetilde{g}_h$ in (1.9). Moreover, if Ω were a polygonal domain coinciding with D_h, the term $d_h(\boldsymbol{\zeta}_h, \boldsymbol{\tau}_h)$ would be zero for all $\boldsymbol{\zeta}_h, \boldsymbol{\tau}_h \in \mathbf{H}_h(\mathrm{D}_h)$, and then problem (1.12) would become well-posed provided the Babuška–Brezzi conditions are proved, namely, the coercivity of a_h on the kernel of b_h, the discrete inf-sup condition for b_h and the boudedness of all the forms involved.

We would like to comment that mixed boundary conditions could be considered. However, it is not straightforward how to deal with this situation in the discrete case since the Neumann data cannot be treated in the same way as the Dirichlet data. This is subject of ongoing work.

1.2.4 Solvability analysis

We now aim to prove the well-posedness of problem (1.12). We begin by stating the assumptions regarding the Galerkin method, the triangulation and the *closeness* between Γ_h and Γ. Let us first introduce some assumptions on the boundary Γ and the mesh $\mathcal{T}_h$.

Assumptions A. For some technical results concerning inverse inequalities, we first assume that the elements K in $\mathcal{T}_h$ are shape-regular in the sense of Ciarlet [53]:

(A.2) There is a constant γ_K such that $h_K \leq \gamma_K \rho_K$, where ρ_K is the radius of the largest ball contained in K.

Next, in order to give sense to the integrals involved in (1.10) and (1.11), we need $\widetilde{g}_h$ (cf. (1.9)) to be a measurable function. This certainly holds under the following assumptions on the boundary Γ (see [57, Lemma 3.1]):

(A.2) Γ is a compact Lipschitz boundary,

(A.3) There exists $\widetilde{\Gamma} \subset \Gamma$ closed in Γ such that $|\widetilde{\Gamma}| = 0$ and $\Gamma \backslash \widetilde{\Gamma}$ is $\mathcal{C}^2$.

Furthermore, we can introduce an extension operator from Ω to $\mathbb{R}^2$. In fact, relaxing the smoothness requirement in Assumption **(A.3)** to $\mathcal{C}^1$ only, the following extension theorem holds. For its proof we refer to [129, Chapter VI].

Theorem 1.1. *There is an extension mapping $\mathscr{E} : \mathrm{H}^m(\Omega) \to \mathrm{H}^m(\mathbb{R}^2)$ defined for all non-negative integers m satisfying $\mathscr{E}(\zeta)\big|_\Omega = \zeta$ for all $\zeta \in \mathrm{H}^m(\Omega)$ and*

$$\|\mathscr{E}(\zeta)\|_{m,\mathbb{R}^2} \leq C \|\zeta\|_{m,\Omega},$$

where C is independent of ζ.

In order to simplify the technicalities of the analysis on the region D_h^c, for every $e \in \mathcal{E}_h^\partial$ and $\mathbf{x} \in e$, we assume that

(A.4) the intersection of the ray $\{\mathbf{x} + \eta(\widetilde{\mathbf{x}} - \mathbf{x}),\ \eta \in \mathbb{R}^+\}$ with Γ is unique.

This prevents situations like the one shown at the right of Figure 1.3.

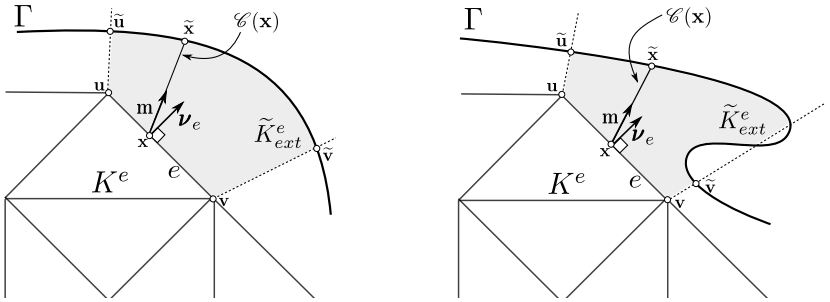


Figure 1.3: Examples of sets $\widetilde{K}^e_{ext}$. (figure produced by the author)

Next, we describe two sets of hypotheses establishing the constraints on the choice of the discrete subspaces in (1.8).

Assumptions B. Let $\mathbf{V}^{\mathrm{D}_h}$ be the discrete kernel of b_h, i.e.,

$$\mathbf{V}^{\mathrm{D}_h} = \{\boldsymbol{\tau}_h \in \mathbf{H}_h(\mathrm{D}_h) : \ b_h(\boldsymbol{\tau}_h, v_h) = 0 \ \ \forall\, v_h \in \mathrm{Q}_h(\mathrm{D}_h)\}\,.$$

In order to have a more explicit definition of $\mathbf{V}^{\mathrm{D}_h}$ we introduce the following assumption:

(B.1) $\operatorname{div} \mathbf{H}_h(\mathrm{D}_h) \subseteq \mathrm{Q}_h(\mathrm{D}_h)$.

In fact, by Assumption **(B.1)**, the subspace $\mathbf{V}^{\mathrm{D}_h}$ can be characterized as follows:

$$\mathbf{V}^{\mathrm{D}_h} = \{\boldsymbol{\tau}_h \in \mathbf{H}_h(\mathrm{D}_h) : \ \operatorname{div} \boldsymbol{\tau}_h \equiv 0 \quad \text{in} \quad \mathrm{D}_h\}\,.$$

Consequently, the bilinear form a_h satisfies the identity

$$a_h(\boldsymbol{\tau}_h, \boldsymbol{\tau}_h) \ = \ \|\boldsymbol{\tau}_h\|^2_{\mathrm{div},\mathrm{D}_h} \quad \forall\, \boldsymbol{\tau}_h \in \mathbf{V}^{\mathrm{D}_h}.$$

This shows that a_h is coercive on $\mathbf{V}^{\mathrm{D}_h}$ with constant $\hat{\alpha} = 1$.

In addition, we assume that the following inf-sup condition holds:

(B.2) There exists $\hat{\beta} > 0$, independent of h, such that

$$\sup_{\substack{\boldsymbol{\tau}_h \in \mathbf{H}_h(\mathrm{D}_h) \\ \boldsymbol{\tau}_h \neq \mathbf{0}}} \frac{b_h(\boldsymbol{\tau}_h, v_h)}{\|\boldsymbol{\tau}_h\|_{\mathrm{div},\mathrm{D}_h}} \geq \hat{\beta} \|v_h\|_{0,\mathrm{D}_h} \quad \forall\, v_h \in \mathrm{Q}_h(\mathrm{D}_h).$$

For the subsequent analysis we will also need the following hypotheses on the local discrete spaces.

Assumptions C. Given an integer $k \geq 0$ and a region $\mathcal{O} \subset \mathbb{R}^2$, we denote by $\mathrm{P}_k(\mathcal{O})$ the space of polynomials of degree at most k defined on $\mathcal{O}$, and let $\mathbf{P}_k(\mathcal{O}) := [\mathrm{P}_k(\mathcal{O})]^2$. Let n_1, n_2 and n_3 be integers such that n_1, $n_2 \geq 1$ and $n_3 \geq 0$. For every $e \in \mathcal{E}_h^{\partial}$, we assume that

(C.1) $\mathbf{M}(K^e) \subseteq \mathbf{P}_{n_1}(K^e)$,

(C.2) $\mathbf{M}(K^e) \cdot \boldsymbol{\nu}_{K^e}|_{\widetilde{e}} \subseteq \mathrm{P}_{n_2}(\widetilde{e})$ for all edge $\widetilde{e} \subset \partial K^e$,

(C.3) $\mathrm{W}(K^e) \subseteq \mathrm{P}_{n_3}(K^e)$,

Next, in Section 1.4 we specify suitable choices of finite element subspaces satisfying **(B.1)**, **(B.2)** and **(C.1)**-**(C.3)**.

In what follows, we introduce assumptions related to the sets $\widetilde{K}^e_{ext}$ and the bilinear form d_h. They are *smallness* assumptions on certain quantities that will appear in the analysis of our method when approximating the L^2-norm of functions defined on $\widetilde{K}^e_{ext}$. These conditions determine how close the boundaries Γ and Γ_h must be.

Assumptions D. Let e be any edge in $\mathcal{E}_h^\partial$. We define $\widetilde{r}_e := \widetilde{H}_e/h_e^\perp$, where $\widetilde{H}_e := \max_{\boldsymbol{x}\in e}\ell(\boldsymbol{x})$ and $h_e^\perp$ is the distance between the vertex of K^e, opposite to e, and the plane determined by e. We assume

(D.1) $\widetilde{r}_e \leq R$,

where R denotes a constant that does not dependent on the meshsize h. This hypothesis indicates that the distance $d(\Gamma,\Gamma_h)$ must be at most of $\mathcal{O}(h)$. We note that, by construction, the family of paths (Σ_h) presented in Section 1.2.2 satisfies **(D.1)**.

To establish the remaining hypotheses, for each $K\in\mathcal{T}_h$ we denote

$$\mathrm{N}_h(\partial K) = \left\{ w\in \mathrm{L}^2(\partial K) :\ w\big|_e \in \mathrm{P}_{n_2}(e) \ \text{ for all edges } e \text{ of } \ K \right\},$$

and introduce the following constant:

$$C_{eq}^e := h_{K^e}^{1/2} \sup_{\substack{w_h\in \mathrm{N}_h(\partial K^e)\\ w_h\neq 0}} \frac{\|w_h\|_{0,\partial K^e}}{\|w_h\|_{-1/2,\partial K^e}}. \tag{1.13}$$

This definition can be inferred using the equivalence of the norms $\|\cdot\|_{0,\partial K}$ and $\|\cdot\|_{-1/2,\partial K}$ on the space $\mathrm{N}_h(\partial K)$ for all $K\in\mathcal{T}_h$; see [62, Lemma 3.2] for further details. Moreover, the value of C_{eq}^e depends solely on the shape-regularity constant γ_{K^e} and the polynomial degree of the space $\mathrm{N}_h(\partial K^e)$.

We shall also make frequent use of the quantity

$$|||\mathbf{p}|||_e := \left(\int_e \int_0^{\ell(\mathbf{x})} |\mathbf{p}(\mathbf{x}+t\mathbf{m}(\mathbf{x}))|^2\, dt\, dS_{\mathbf{x}} \right)^{1/2}, \tag{1.14}$$

where $e\in\mathcal{E}_h^\partial$ and $\mathbf{p}$ is smooth enough in order to make the integral well-defined. In addition, we define

$$\widetilde{C}_{ext}^e := \widetilde{r}_e^{-1/2} \sup_{\substack{\boldsymbol{\zeta}_h\in\mathbf{M}(K^e)\\ \boldsymbol{\zeta}_h\neq\mathbf{0}}} \frac{|||\mathbf{E}_h(\boldsymbol{\zeta}_h)|||_e}{\|\boldsymbol{\zeta}_h\|_{0,K^e}}. \tag{1.15}$$

We recall that $\mathbf{E}_h(\boldsymbol{\zeta}_h)$ is the extrapolation of the polynomial $\boldsymbol{\zeta}_h$ from K^e to $\widetilde{K}_{ext}^e$, since $\mathbf{M}(K^e)$ is a space of polynomials thanks to **(C.1)**. The constant $\widetilde{C}_{ext}^e$ is independent of the meshsize h, but depends on the shape-regularity constant γ_{K^e} and on the polynomial degree; see Appendix A.

We are now in a position of discussing the boundedness of the bilinear form d_h. Let $\boldsymbol{\zeta}_h\in\mathbf{H}_h(\mathrm{D}_h)$. According to the notation introduced in Section 1.2.2, for any $\mathbf{x}$ lying on a boundary edge e, we set

$$\widetilde{w}_h(\mathbf{x}) := \int_0^{\ell(\mathbf{x})} \mathbf{E}_h(\boldsymbol{\zeta}_h)(\mathbf{x}+t\,\mathbf{m}(\mathbf{x}))\cdot\mathbf{m}(\mathbf{x})\,dt.$$

Applying now the Cauchy–Schwarz inequality, considering definitions (1.14) and (1.15), and the fact that, for all $\mathbf{x}\in e$, $\ell(\mathbf{x})\leq \widetilde{H}_e = \widetilde{r}_e h_e^\perp \leq \widetilde{r}_e h_{K^e}$, we obtain

$$\begin{aligned}
\|\widetilde{w}_h\|_{0,e}^2 &\leq \int_e \ell(\mathbf{x}) \int_0^{\ell(\mathbf{x})} |\mathbf{E}_h(\boldsymbol{\zeta}_h)|^2(\mathbf{x}+t\mathbf{m}(\mathbf{x}))\,dt\,dS_{\mathbf{x}} \\
&\leq \widetilde{r}_e\widetilde{H}_e \left(\widetilde{C}_{ext}^e\right)^2 \|\boldsymbol{\zeta}_h\|_{0,K^e}^2 \\
&\leq \widetilde{r}_e^2 h_{K_e} \left(\widetilde{C}_{ext}^e\right)^2 \|\boldsymbol{\zeta}_h\|_{0,K^e}^2.
\end{aligned} \tag{1.16}$$

Then, by definition of d_h (cf. (1.10)), applying the Cauchy–Schwarz inequality, and using (1.13) and Assumption **(C.2)**, it follows that

$$|d_h(\boldsymbol{\zeta}_h,\boldsymbol{\tau}_h)| \leq \sum_{e\in\mathcal{E}_h^\partial} \|\widetilde{w}_h\|_{0,e}\|\boldsymbol{\tau}_h\cdot\boldsymbol{\nu}_{K^e}\|_{0,\partial K^e} \leq \max_{e\in\mathcal{E}_h^\partial}\left\{\widetilde{r}_e\widetilde{C}^e_{ext}C^e_{eq}\right\}\|\boldsymbol{\zeta}_h\|_{\mathrm{div},\mathrm{D}_h}\|\boldsymbol{\tau}_h\|_{\mathrm{div},\mathrm{D}_h}, \tag{1.17}$$

for all $\boldsymbol{\zeta}_h,\boldsymbol{\tau}_h\in\mathbf{H}_h(\mathrm{D}_h)$, where the continuity of the normal trace operator acting from $\mathbf{H}(\mathbf{div};K^e)$ onto $\mathrm{H}^{-1/2}(\partial K^e)$ (see, e.g., [73, Theorem 1.7]) has been applied to bound $\|\boldsymbol{\tau}\cdot\boldsymbol{\nu}_{K^e}\|_{0,\partial K^e}$. Therefore, the boundedness of d_h is satisfied if we assume that

(D.2) $\max_{e\in\mathcal{E}_h^\partial}\left\{\widetilde{r}_e\widetilde{C}^e_{ext}C^e_{eq}\right\}\leq 1/2.$

In most cases the above condition is not entirely verifiable because the left-hand side might be not a fully computable quantity. Certainly it holds if $\widetilde{r}_e$ is of order h and h is small enough, as it happens when the boundary is interpolated by a piecewise linear function.

Having introduced the aforementioned hypotheses, we are now in a position of establishing the main result of this section, namely, the well-posedness of problem (1.12). To allow for a more compact notation, in the sequel we employ the norm

$$\|(\boldsymbol{\tau},\mathbf{v})\|_{\mathbb{H}(\mathbf{div};\Omega)\times\mathbf{L}^2(\Omega)} := \left(\|\boldsymbol{\tau}\|^2_{\mathbf{div},\Omega}+\|\mathbf{v}\|^2_{0,\Omega}\right)^{1/2} \quad \forall\,(\boldsymbol{\tau},\mathbf{v})\in\mathbb{H}(\mathbf{div};\Omega)\times\mathbf{L}^2(\Omega).$$

Theorem 1.2. *Suppose that Assumptions* **A**, **B**, **C** *and* **D** *are satisfied. Then, given* $f\in\mathrm{L}^2(\Omega)$ *and* $g\in\mathrm{H}^{1/2}(\Gamma)$, *there exists a unique* $(\boldsymbol{\sigma}_h,u_h)\in\mathbf{H}_h(\mathrm{D}_h)\times\mathrm{Q}_h(\mathrm{D}_h)$ *solution to problem* (1.12), *satisfying*

$$\|(\boldsymbol{\sigma}_h,u_h)\|_{\mathbf{H}(\mathrm{div};\mathrm{D}_h)\times\mathrm{L}^2(\mathrm{D}_h)} \lesssim \|F_h\|_{[\mathrm{Q}(\mathrm{D}_h)]'}+\|G_h\|_{[\mathbf{H}_h(\mathrm{D}_h)]'}.$$

Proof. We first discuss the stability of the forms involved in (1.12). It is clear that a_h and b_h are bounded with constants less or equal to 1. Moreover, F_h is bounded with $\|F_h\|_{[\mathbf{H}(\mathrm{D}_h)]'}\leq\|f\|_{0,\mathrm{D}_h}$. To obtain a bound for $\|G_h\|_{[\mathbf{H}_h(\mathrm{D}_h)]'}$, we first note that the composition $\overline{g}(\cdot):=g(\widetilde{\mathbf{x}}(\cdot))$ is a function in $\mathrm{H}^{1/2}(\Gamma_h)$, since $\widetilde{\mathbf{x}}:\Gamma_h\to\Gamma$ is a continuous mapping and $g\in\mathrm{H}^{1/2}(\Gamma)$. Then, we apply the boundedness of the normal trace operator acting from $\mathbf{H}(\mathrm{div};\mathrm{D}_h)$ onto $\mathrm{H}^{-1/2}(\Gamma_h)$ (see [73, Theorem 1.7]) and obtain G_h is bounded with $\|G_h\|_{[\mathrm{Q}(\mathrm{D}_h)]'}\leq\|\overline{g}\|_{1/2,\Gamma_h}$.

On the other hand, the bilinear form a_h+d_h is coercive on $\mathbf{V}^{\mathrm{D}_h}$. In fact, the result follows from (1.17) and Assumptions **(B.1)** and **(D.2)**, that is,

$$(a_h+d_h)(\boldsymbol{\tau}_h,\boldsymbol{\tau}_h)\geq\frac{1}{2}\|\boldsymbol{\tau}_h\|^2_{\mathrm{div},\mathrm{D}_h}\quad\boldsymbol{\tau}_h\in\mathbf{V}^{\mathrm{D}_h}.$$

Finally, the discrete inf-sup condition for b_h is fulfilled by virtue of Assumption **(B.2)**. Therefore, the result is a straightforward consequence of the classical Babuška–Brezzi theory. □

1.3 Error analysis

In this section we carry out the error analysis for our Galerkin scheme (1.12). We first derive error estimates on D_h by considering the arbitrary finite element subspaces satisfying the assumptions in

Section 1.2.4, and well-known Strang-type estimates for saddle point problems. Then, we will follow the procedure in [57, Section 5.2] to control the errors on D_h^c. Moreover, we use the aforementioned analysis to state the theoretical rates of convergence when using the specific discrete spaces provided in Section 1.4.

1.3.1 Error estimates on D_h

Let $(\boldsymbol{\sigma}, u) \in \mathbf{H}(\mathrm{div};\Omega) \times \mathrm{L}^2(\Omega)$ be the solution of (1.2) satisfying (1.5), and let $(\boldsymbol{\sigma}_h, u_h) \in \mathbf{H}_h(\mathrm{D}_h) \times \mathrm{Q}_h(\mathrm{D}_h)$ be the solution of (1.12). Firstly, we are interested in obtaining upper bounds for

$$\|(\boldsymbol{\sigma}, u) - (\boldsymbol{\sigma}_h, u_h)\|_{\mathbf{H}(\mathrm{div};\mathrm{D}_h)\times\mathrm{L}^2(\mathrm{D}_h)}.$$

To this end, we rearrange (1.5) and (1.12) as the following pair of continuous and discrete formulations:

$$\begin{aligned} a_h(\boldsymbol{\sigma}, \boldsymbol{\tau}) + b_h(\boldsymbol{\tau}, u) &= \langle \boldsymbol{\tau}\cdot\boldsymbol{\nu}_{\Gamma_h}, \widetilde{g}\rangle_{\Gamma_h} \quad \forall\, \boldsymbol{\tau} \in \mathbf{H}(\mathrm{div};\mathrm{D}_h), \\ b_h(\boldsymbol{\sigma}, v) &= F_h(v) \qquad\qquad\quad \forall\, v \in \mathrm{L}^2(\mathrm{D}_h), \end{aligned} \tag{1.18}$$

and

$$\begin{aligned} a_h(\boldsymbol{\sigma}_h, \boldsymbol{\tau}_h) + b_h(\boldsymbol{\tau}_h, u_h) &= G_h(\boldsymbol{\tau}_h) - d_h(\boldsymbol{\sigma}_h, \boldsymbol{\tau}_h) \quad \forall\, \boldsymbol{\tau}_h \in \mathbf{H}_h(\mathrm{D}_h), \\ b_h(\boldsymbol{\sigma}_h, v_h) &= F_h(v_h) \qquad\qquad\quad \forall\, v_h \in \mathrm{Q}_h(\mathrm{D}_h). \end{aligned} \tag{1.19}$$

Thus, as we have already pointed out before and as suggested by the structure of the foregoing systems, in what follows we proceed similarly to [82] (see also [63]) and apply a Strang-type estimate for saddle point problems whose continuous and discrete schemes differ only in the functionals involved. For the sake of completeness, this result is recalled next. We refer the reader to [121, Theorem 11.2] for more details.

Theorem 1.3. *Let $\mathbf{H}$ and Q be two Hilbert spaces, $\mathcal{G} \in \mathbf{H}'$, $\mathcal{F} \in \mathrm{Q}'$, and let $a : \mathbf{H}\times\mathbf{H} \to \mathbb{R}$ and $b : \mathbf{H}\times\mathrm{Q} \to \mathbb{R}$ be bounded bilinear forms satisfying the Babuška–Brezzi conditions, that is,*

i) There exists $\alpha > 0$ such that

$$a(\boldsymbol{\tau}, \boldsymbol{\tau}) \geq \alpha \|\boldsymbol{\tau}\|_{\mathbf{H}}^2 \quad \forall\, \boldsymbol{\tau} \in \mathbf{V},$$

where $\mathbf{V} := \{\boldsymbol{\tau} \in \mathbf{H} : \ b(\boldsymbol{\tau}, v) = 0 \ \ \forall\, v \in \mathrm{Q}\}$.

ii) There exists $\beta > 0$ such that

$$\sup_{\substack{\boldsymbol{\tau}\in\mathbf{H}\\ \boldsymbol{\tau}\neq\mathbf{0}}} \frac{b(\boldsymbol{\tau}, v)}{\|\boldsymbol{\tau}\|_{\mathbf{H}}} \geq \beta \|v\|_{\mathrm{Q}} \quad \forall\, v \in \mathrm{Q}.$$

In addition, let $\mathbf{H}_h$ and Q_h be two finite dimensional subspaces of $\mathbf{H}$ and Q, respectively, and for each $h > 0$ consider functionals $\mathcal{G}_h \in \mathbf{H}_h'$ and $\mathcal{F}_h \in \mathrm{Q}_h'$. Assume that:

iii) There exists $\hat{\alpha} > 0$, independent of the discretization parameter h, such that

$$a(\boldsymbol{\tau}_h, \boldsymbol{\tau}_h) \geq \hat{\alpha} \|\boldsymbol{\tau}_h\|_{\mathbf{H}}^2 \quad \forall\, \boldsymbol{\tau}_h \in \mathbf{V}_h,$$

where $\mathbf{V}_h := \{\boldsymbol{\tau}_h \in \mathbf{H}_h : \ b(\boldsymbol{\tau}_h, v_h) = 0 \ \ \forall\, v_h \in \mathrm{Q}_h\}$.

iv) There exists $\hat{\beta} > 0$, independent of the discretization parameter h, such that

$$\sup_{\substack{\boldsymbol{\tau}_h \in \mathbf{H}_h \\ \boldsymbol{\tau}_h \neq \mathbf{0}}} \frac{b(\boldsymbol{\tau}_h, v_h)}{\|\boldsymbol{\tau}_h\|_{\mathbf{H}}} \geq \hat{\beta} \|v_h\|_{\mathrm{Q}} \quad \forall\, v_h \in \mathrm{Q}_h.$$

Furthermore, let $(\boldsymbol{\sigma}, u) \in \mathbf{H} \times \mathrm{Q}$ and $(\boldsymbol{\sigma}_h, u_h) \in \mathbf{H}_h \times \mathrm{Q}_h$ be such that

$$\begin{aligned} a(\boldsymbol{\sigma}, \boldsymbol{\tau}) + b(\boldsymbol{\tau}, u) &= \mathcal{G}(\boldsymbol{\tau}) \qquad \forall\, \tau \in \mathbf{H}, \\ b(\boldsymbol{\sigma}, v) &= \mathcal{F}(v) \qquad \forall\, v \in \mathrm{Q}, \end{aligned} \tag{1.20}$$

and

$$\begin{aligned} a(\boldsymbol{\sigma}_h, \boldsymbol{\tau}_h) + b(\boldsymbol{\tau}_h, u_h) &= \mathcal{G}_h(\boldsymbol{\tau}_h) \qquad \forall\, \boldsymbol{\tau}_h \in \mathbf{H}_h, \\ b(\boldsymbol{\sigma}_h, v_h) &= \mathcal{F}_h(v_h) \qquad \forall\, v_h \in \mathrm{Q}_h. \end{aligned} \tag{1.21}$$

Then, for each $h > 0$, the following estimates hold:

$$\begin{aligned} \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathrm{div},\mathrm{D}_h} \leq{}& \left(1 + \frac{\|a\|}{\hat{\alpha}}\right)\left(1 + \frac{\|b\|}{\hat{\beta}}\right) \inf_{\boldsymbol{\zeta}_h \in \mathbf{H}_h} \|\boldsymbol{\sigma} - \boldsymbol{\zeta}_h\|_{\mathbf{H}} + \frac{\|b\|}{\hat{\alpha}} \inf_{w_h \in \mathrm{Q}_h} \|u - w_h\|_{\mathrm{Q}} \\ &+ \frac{1}{\hat{\beta}}\left(1 + \frac{\|a\|}{\hat{\alpha}}\right) \sup_{\substack{w_h \in \mathrm{Q}_h \\ w_h \neq 0}} \frac{|(\mathcal{F} - \mathcal{F}_h)(w_h)|}{\|w_h\|_{\mathrm{Q}}} + \left(\frac{1}{\hat{\alpha}}\right) \sup_{\substack{\boldsymbol{\tau}_h \in \mathbf{H}_h \\ \boldsymbol{\tau}_h \neq \mathbf{0}}} \frac{|(\mathcal{G} - \mathcal{G}_h)(\boldsymbol{\tau}_h)|}{\|\boldsymbol{\tau}_h\|_{\mathbf{H}}}, \end{aligned} \tag{1.22}$$

and

$$\begin{aligned} \|u - u_h\|_{0,\mathrm{D}_h} \leq{}& \frac{\|a\|}{\hat{\beta}}\left(1 + \frac{\|a\|}{\hat{\alpha}}\right)\left(1 + \frac{\|b\|}{\hat{\beta}}\right) \inf_{\boldsymbol{\zeta}_h \in \mathbf{H}_h} \|\boldsymbol{\sigma} - \boldsymbol{\zeta}_h\|_{\mathbf{H}} \\ &+ \left(1 + \frac{\|b_h\|}{\hat{\beta}} + \frac{\|b\|}{\hat{\beta}}\frac{\|a\|}{\hat{\alpha}}\right) \inf_{w_h \in \mathrm{Q}_h} \|u - w_h\|_{\mathrm{Q}} \\ &+ \frac{\|a\|}{\hat{\beta}^2}\left(1 + \frac{\|a\|}{\hat{\alpha}}\right) \sup_{\substack{w_h \in \mathrm{Q}_h \\ w_h \neq 0}} \frac{|(\mathcal{F} - \mathcal{F}_h)(w_h)|}{\|w_h\|_{\mathrm{Q}}} \\ &+ \frac{1}{\hat{\beta}}\left(1 + \frac{\|a\|}{\hat{\alpha}}\right) \sup_{\substack{\boldsymbol{\tau}_h \in \mathbf{H}_h \\ \boldsymbol{\tau}_h \neq \mathbf{0}}} \frac{|(\mathcal{G} - \mathcal{G}_h)(\boldsymbol{\tau}_h)|}{\|\boldsymbol{\tau}_h\|_{\mathbf{H}}}. \end{aligned} \tag{1.23}$$

Hence, applying (1.22) and (1.23) to (1.18) and (1.19), and noting that in our case $\hat{\alpha} = 1$ and $\|a\|, \|b\| \leq 1$, we deduce that

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathrm{div},\mathrm{D}_h} \leq C_S^1 \inf_{\boldsymbol{\zeta}_h \in \mathbf{H}_h(\mathrm{D}_h)} \|\boldsymbol{\sigma} - \boldsymbol{\zeta}_h\|_{\mathrm{div},\mathrm{D}_h} + C_S^2 \inf_{w_h \in \mathrm{Q}_h(\mathrm{D}_h)} \|u - w_h\|_{0,\mathrm{D}_h} + \mathbb{T}^{\boldsymbol{\sigma}} \tag{1.24}$$

and

$$\|u - u_h\|_{0,\mathrm{D}_h} \leq C_S^3 \inf_{\boldsymbol{\zeta}_h \in \mathbf{H}_h(\mathrm{D}_h)} \|\boldsymbol{\sigma} - \boldsymbol{\zeta}_h\|_{\mathrm{div},\mathrm{D}_h} + C_S^4 \inf_{w_h \in \mathrm{Q}_h(\mathrm{D}_h)} \|u - w_h\|_{0;\mathrm{D}_h} + \frac{2}{\hat{\beta}}\mathbb{T}^{\boldsymbol{\sigma}}, \tag{1.25}$$

with C_S^1, C_S^2, C_S^3 and C_S^4 being positive constants independent of the discretization parameters and

$$\mathbb{T}^{\boldsymbol{\sigma}} := \sup_{\substack{\boldsymbol{\tau}_h \in \mathbf{H}_h(\mathrm{D}_h) \\ \boldsymbol{\tau}_h \neq \mathbf{0}}} \frac{|\langle \boldsymbol{\tau}_h \cdot \boldsymbol{\nu}_{\Gamma_h}, \widetilde{g} \rangle_{\Gamma_h} - (G_h(\boldsymbol{\tau}_h) - d_h(\boldsymbol{\sigma}_h, \boldsymbol{\tau}_h))|}{\|\boldsymbol{\tau}_h\|_{\mathrm{div},\mathrm{D}_h}}. \tag{1.26}$$

We now proceed to bound $\mathbb{T}^{\boldsymbol{\sigma}}$.

Lemma 1.4. *There exists a positive constant C, independent of h, such that*

$$\mathbb{T}^{\boldsymbol{\sigma}} \leq \inf_{\boldsymbol{\zeta}_h \in \mathbf{H}_h(\mathrm{D}_h)} \left(\sum_{e \in \mathcal{E}_h^{\partial}} (\widetilde{r}_e)^{1/2} C_{eq}^e |||\boldsymbol{\sigma} - \mathbf{E}_h(\boldsymbol{\zeta}_h)|||_e + \frac{1}{2} \|\boldsymbol{\sigma} - \boldsymbol{\zeta}_h\|_{\mathrm{D}_h} \right) + \frac{1}{2} \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathrm{D}_h}. \tag{1.27}$$

Proof. Firstly, using the Cauchy–Schwarz inequality, (1.9) and (1.13), we obtain

$$\mathbb{T}^{\boldsymbol{\sigma}} \leq \sum_{e \in \mathcal{E}_h^{\partial}} C_{eq}^e h_{K^e}^{-1/2} \|\widetilde{g} - \widetilde{g}_h\|_{0,e}. \tag{1.28}$$

On the other hand, from definitions (1.4) and (1.9) we have, for each $e \in \mathcal{E}_h^{\partial}$ and $\mathbf{x} \in e$,

$$(\widetilde{g} - \widetilde{g}_h)(\mathbf{x}) = -\int_0^{\ell(\mathbf{x})} (\boldsymbol{\sigma} - \mathbf{E}_h(\boldsymbol{\sigma}_h))(\mathbf{x} + t\mathbf{m}(\mathbf{x})) \cdot \mathbf{m}(\mathbf{x})\, dt.$$

Applying now the Cauchy–Schwarz inequality,

$$\|\widetilde{g} - \widetilde{g}_h\|_{0,e}^2 \leq \widetilde{H}_e |||\boldsymbol{\sigma} - \mathbf{E}_h(\boldsymbol{\sigma}_h)|||_e^2 \leq \widetilde{r}_e h_{K^e} |||\boldsymbol{\sigma} - \mathbf{E}_h(\boldsymbol{\sigma}_h)|||_e^2.$$

Combined with (1.28) this implies

$$\mathbb{T}^{\boldsymbol{\sigma}} \leq \sum_{e \in \mathcal{E}_h^{\partial}} (\widetilde{r}_e)^{1/2} C_{eq}^e |||\boldsymbol{\sigma} - \mathbf{E}_h(\boldsymbol{\sigma}_h)|||_e.$$

Let now $\boldsymbol{\zeta}_h \in \mathbf{H}_h(\mathrm{D}_h)$. Adding and subtracting $\mathbf{E}_h(\boldsymbol{\zeta}_h)$ to the term on the right-hand side of the last inequality, and using definition (1.15) and Assumption **(D.2)**, we obtain

$$\begin{aligned} \mathbb{T}^{\boldsymbol{\sigma}} &\leq \sum_{e \in \mathcal{E}_h^{\partial}} (\widetilde{r}_e)^{1/2} C_{eq}^e |||\boldsymbol{\sigma} - \mathbf{E}_h(\boldsymbol{\zeta}_h)|||_e + \sum_{e \in \mathcal{E}_h^{\partial}} (\widetilde{r}_e)^{1/2} C_{eq}^e |||\mathbf{E}_h(\boldsymbol{\zeta}_h) - \mathbf{E}_h(\boldsymbol{\sigma}_h)|||_e \\ &\leq \sum_{e \in \mathcal{E}_h^{\partial}} (\widetilde{r}_e)^{1/2} C_{eq}^e |||\boldsymbol{\sigma} - \mathbf{E}_h(\boldsymbol{\zeta}_h)|||_e + \frac{1}{2} \sum_{e \in \mathcal{E}_h^{\partial}} \|\boldsymbol{\zeta}_h - \boldsymbol{\sigma}_h\|_{0,K^e}. \end{aligned}$$

Thus, adding and subtracting $\boldsymbol{\sigma}$ we obtain (1.27). □

In summary, (1.24), (1.25) and (1.27), yield the following result.

Theorem 1.5. *Suppose that assumptions of Theorem 1.2 are satisfied. Let $(\boldsymbol{\sigma}, u) \in \mathbf{H}(\mathrm{div}; \Omega) \times \mathrm{L}^2(\Omega)$ be the solution of (1.2) satisfying (1.5), and $(\boldsymbol{\sigma}_h, u_h) \in \mathbf{H}_h(\mathrm{D}_h) \times \mathrm{Q}_h(\mathrm{D}_h)$ be the solution of (1.12). Then, there holds*

$$\begin{aligned} &\|(\boldsymbol{\sigma}, u) - (\boldsymbol{\sigma}_h, u_h)\|_{\mathbf{H}(\mathrm{div};\mathrm{D}_h) \times \mathrm{L}^2(\mathrm{D}_h)} \\ &\lesssim \inf_{w_h \in \mathrm{Q}_h(\mathrm{D}_h)} \|u - w_h\|_{0,\mathrm{D}_h} + \inf_{\boldsymbol{\zeta}_h \in \mathbf{H}_h(\mathrm{D}_h)} \left(\|\boldsymbol{\sigma} - \boldsymbol{\zeta}_h\|_{\mathrm{div},\mathrm{D}_h} + \sum_{e \in \mathcal{E}_h^{\partial}} (\widetilde{r}_e)^{1/2} C_{eq}^e |||\boldsymbol{\sigma} - \mathbf{E}_h(\boldsymbol{\zeta}_h)|||_e \right). \end{aligned} \tag{1.29}$$

1.3.2 Approximating $\boldsymbol{\sigma}$ and u in D_h^c

In this section we provide error estimates outside the computational domain. Before doing so, we need to show that, under certain conditions, the norms $\|\cdot\|_{0,\widetilde{K}^e_{ext}}$ and $|||\cdot|||_e$ are equivalent.

Let $\mathbf{u}$ and $\mathbf{v}$ be the vertices of a boundary edge e, and $\widetilde{\mathbf{u}}$ and $\widetilde{\mathbf{v}}$ be their corresponding points in Γ described in Section 1.2.2. We recall that $\widetilde{K}^e_{ext}$ is the region determined by $\mathbf{u}$, $\mathbf{v}$, $\widetilde{\mathbf{u}}$ and $\widetilde{\mathbf{v}}$ as Figure 1.3 (left) shows. Then, a point $\mathbf{x}$ on e can be represented as $\mathbf{x}(\theta) = \mathbf{u} + \theta(\mathbf{v}-\mathbf{u})$ for $\theta \in [0,1]$. According to Section 1.2.2, the tangent vector of the path associated to $\mathbf{x}$ can be then written as $\widehat{\mathbf{m}}(\theta) := \mathbf{m}^{\mathbf{u}} + \theta(\mathbf{m}^{\mathbf{v}} - \mathbf{m}^{\mathbf{u}})$. Moreover, $\mathbf{m}(\theta) := \widehat{\mathbf{m}}(\theta)/|\widehat{\mathbf{m}}(\theta)|$ if $\widehat{\mathbf{m}}(\theta) \neq \mathbf{0}$; and $\mathbf{m}(\theta) = \boldsymbol{\nu}_e$, otherwise.

Thus, for $\mathbf{y} \in \widetilde{K}^e_{ext}$ we have:

$$\mathbf{y}(\theta, s) = \mathbf{x}(\theta) + \mathbf{m}(\theta)s \quad s \in [0, \ell(\theta)],\ \theta \in [0,1], \tag{1.30}$$

where $\ell(\theta)$ is the length of the transferring path associated to $\mathbf{x}(\theta)$.

Now, for a vector $\mathbf{w} = (w_1, w_2)$, we define $\mathbf{w}^{\perp} := (-w_2, w_1)$ and the Jacobian of the above transformation is given by

$$\mathbf{J}(s,\theta) \quad = \quad \left| |e|\mathbf{m}(\theta)\cdot\boldsymbol{\nu}_e + \frac{s}{\alpha(\theta)}\mathbf{m}(\theta)\cdot(\mathbf{m}^{\mathbf{v}} - \mathbf{m}^{\mathbf{u}})^{\perp} \right|, \tag{1.31}$$

where $\alpha(\theta) = |\widehat{\mathbf{m}}(\theta)|$ if $\widehat{\mathbf{m}}(\theta) \neq \mathbf{0}$; and $\alpha(\theta) = 1$, otherwise. In turn, considering the parametrization (1.30), we have that

$$\|\mathbf{p}\|^2_{0,\widetilde{K}^e_{ext}} = \int_{\widetilde{K}^e_{ext}} |\mathbf{p}(\mathbf{y})|^2\, d\mathbf{y} = \int_0^1 \int_0^{\ell(\theta)} |\mathbf{p}(\mathbf{y}(s,\theta))|^2 |\mathbf{J}(s,\theta)|\, ds\, d\theta. \tag{1.32}$$

Therefore, the equivalence of norms holds if $|\mathbf{J}(s,\theta)|$ is bounded from above and below for which specific conditions must be satisfied by the vectors appearing in (1.31). More precisely, we have,

Lemma 1.6. *Let $\mathbf{p} \in \mathrm{L}^2(\widetilde{K}^e_{ext})$ and suppose that Assumptions* **A** *are satisfied. In addition, let us consider the following conditions:*

i) $\mathbf{m}^{\mathbf{u}} \cdot \mathbf{m}^{\mathbf{v}} \geq 0$,

ii) there exists constant β_e, independent of h, such that $\mathbf{m}(\theta)\cdot\boldsymbol{\nu}_e \geq \beta_e > 0$ for all $\theta \in [0,1]$; and

iii) $\mathbf{m}^{\mathbf{u}} \cdot (\mathbf{m}^{\mathbf{v}})^{\perp} \geq 0$.

If i) holds, then

$$\|\mathbf{p}\|_{0,\widetilde{K}^e_{ext}} \leq C_2^e |||\mathbf{p}|||_e, \tag{1.33}$$

where $C_2^e := \left(1 + 2\gamma_{K^e}\widetilde{r}_e\sqrt{2}\right)^{1/2}$. Moreover, if ii) and iii) hold, then

$$C_1^e |||\mathbf{p}|||_e \leq \|\mathbf{p}\|_{0,\widetilde{K}^e_{ext}}, \tag{1.34}$$

with $C_1^e := \beta_e^{1/2}$.

We point out that, if $\mathbf{m}^{\mathbf{u}}$ and $\mathbf{m}^{\mathbf{v}}$ are parallel to $\boldsymbol{\nu}_e$, then $|\mathbf{J}(s,\theta)| = |e|$, which means that $|||\mathbf{p}|||_e = \|\mathbf{p}\|_{0,\widetilde{K}^e_{ext}}$ and conditions i)-iii) are not required.

Proof. By assumption i), we obtain

$$\alpha(\theta)^2 = \theta^2 + (\theta-1)^2 + 2\theta(1-\theta)\mathbf{m}^{\mathbf{u}}\cdot\mathbf{m}^{\mathbf{v}} \geq \theta^2 + (\theta-1)^2 \geq 1/2.$$

Since $\ell(\theta) \leq \widetilde{H}_e \leq \widetilde{r}_e h_{K_e} \leq \gamma_{K^e}\widetilde{r}_e|e|$ for all $\theta\in[0,1]$, then

$$|\mathbf{J}(s,\theta)| \leq |e| + \frac{\ell(\theta)}{\alpha(\theta)}(|\mathbf{m}^{\mathbf{u}}| + |\mathbf{m}^{\mathbf{v}}|) \leq |e| + 2\gamma_{K^e}\widetilde{r}_e\sqrt{2}|e|.$$

Thus,

$$\|\mathbf{p}\|^2_{0,\widetilde{K}^e_{ext}} \leq \left(1+2\gamma_{K^e}\widetilde{r}_e\sqrt{2}\right)|e|\int_0^1\int_0^{\ell(\theta)} |\mathbf{p}(\mathbf{y}(s,\theta))|^2\,ds\,d\theta = \left(1+2\gamma_{K^e}\widetilde{r}_e\sqrt{2}\right)|||\mathbf{p}|||^2_e,$$

which implies (1.33).

On the other hand, we note that the Jacobian (1.31) can be written as

$$\mathbf{J}(s,\theta) = \left| |e|\mathbf{m}(\theta)\cdot\boldsymbol{\nu}_e + \frac{s}{\alpha(\theta)}\mathbf{m}^{\mathbf{u}}\cdot(\mathbf{m}^{\mathbf{v}})^{\perp}\right|. \tag{1.35}$$

Then, by assumptions ii) and iii), we have that $\mathbf{J}(s,\theta)\geq\beta_e|e|$. By (1.32), we obtain (1.34). □

Then we have the following intermediate result.

Lemma 1.7. *In addition to the hypotheses of Theorem 1.2 and assumption (i) in Lemma 1.6, we suppose that there exists an integer $m\geq 0$ such that $\boldsymbol{\sigma}\in\mathbf{H}^{m+1}(\Omega)$, with $\operatorname{div}\boldsymbol{\sigma}\in\mathrm{H}^{m+1}(\Omega)$. Then, for any $\boldsymbol{\zeta}_h\in\mathbf{H}_h(\mathrm{D}_h)$, there holds*

$$\sum_{e\in\mathcal{E}_h^{\partial}} \|\boldsymbol{\sigma}-\mathbf{E}_h(\boldsymbol{\zeta}_h)\|_{0,\widetilde{K}^e_{ext}} \lesssim h^{m+1}\|\boldsymbol{\sigma}\|_{m+1,\Omega} + \|\boldsymbol{\sigma}-\boldsymbol{\zeta}_h\|_{0,\mathrm{D}_h} \tag{1.36}$$

and

$$\sum_{e\in\mathcal{E}_h^{\partial}} \|\boldsymbol{\sigma}-\mathbf{E}_h(\boldsymbol{\zeta}_h)\|_{\operatorname{div},\widetilde{K}^e_{ext}} \lesssim \|\boldsymbol{\sigma}-\boldsymbol{\zeta}_h\|_{\operatorname{div},\mathrm{D}_h} + h^{m+1}\Big(\|\boldsymbol{\sigma}\|_{m+1,\Omega} + \|\operatorname{div}\boldsymbol{\sigma}\|_{m+1,\Omega}\Big). \tag{1.37}$$

Proof. Let $\boldsymbol{\zeta}_h\in\mathbf{H}_h(\mathrm{D}_h)$ and $\mathscr{E}:\mathrm{H}^{m+1}(\Omega)\to\mathrm{H}^{m+1}(\mathbb{R}^2)$ be the extension operator introduced in Theorem 1.1. Since Γ is Lipschitz continuous thanks to Assumption **(A.1)**, we define

$$\boldsymbol{\psi}_e := \left(\mathrm{T}_e^{m+1}(\mathscr{E}\sigma_1), \mathrm{T}_e^{m+1}(\mathscr{E}\sigma_2)\right)^T, \tag{1.38}$$

where, for each $i\in\{1,2\}$ and for any $e\in\mathcal{E}_h^{\partial}$, $\mathrm{T}_e^{m+1}(\mathscr{E}\sigma_i)$ is the Taylor polynomial of degree $m+1$ of the function $\mathscr{E}\sigma_i$ around the center of the ball $\widetilde{B}^e$ (see [37, Chapter IV] for details), with $\widetilde{B}^e$ being the ball of radius $h_{\widetilde{B}^e}$ (equal to the diameter of $\widetilde{K}^e_{ext}\cup K^e$) centered at the middle point of the edge e; see Figure 1.4. Thus, by definition, $\boldsymbol{\psi}_e\in\mathbf{P}_s(\widetilde{B}^e)$ with $s<m+1$.

Then, applying the triangle inequality, and using (1.15) and Lemma 1.6, we obtain

$$\begin{aligned}
\|\boldsymbol{\sigma}-\mathbf{E}_h(\boldsymbol{\zeta}_h)\|_{0,\widetilde{K}^e_{ext}} &\leq \|\boldsymbol{\sigma}-\boldsymbol{\psi}_e\|_{0,\widetilde{K}^e_{ext}} + \|\boldsymbol{\psi}_e-\mathbf{E}_h(\boldsymbol{\zeta}_h)\|_{0,\widetilde{K}^e_{ext}}\\
&\leq \|\boldsymbol{\sigma}-\boldsymbol{\psi}_e\|_{0,\widetilde{K}^e_{ext}} + C_2^e(\widetilde{r}_e)^{1/2}\widetilde{C}^e_{ext}\|\boldsymbol{\psi}_e-\boldsymbol{\zeta}_h\|_{0,K^e}\\
&\leq \left(1+C_2^e(\widetilde{r}_e)^{1/2}\widetilde{C}^e_{ext}\right)\|\boldsymbol{\sigma}-\boldsymbol{\psi}_e\|_{0,\widetilde{K}^e_{ext}} + C_2^e(\widetilde{r}_e)^{1/2}\widetilde{C}^e_{ext}\|\boldsymbol{\sigma}-\boldsymbol{\zeta}_h\|_{0,K^e},
\end{aligned} \tag{1.39}$$

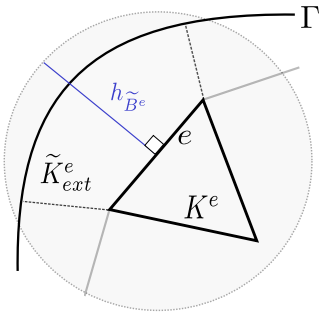


Figure 1.4: Example of ball $\widetilde{B}^e$ associated with a boundary edge e. (figure produced by the author)

where in the last line we added and subtracted $\boldsymbol{\sigma}$. Furthermore, using the approximation properties of Taylor polynomials [37], we have

$$\|\boldsymbol{\sigma} - \boldsymbol{\psi}_e\|_{0,\widetilde{K}^e_{ext}} \leq h^{m+1} |\boldsymbol{\mathscr{E}}\boldsymbol{\sigma}|_{m+1,\widetilde{B}^e}, \tag{1.40}$$

where $\boldsymbol{\mathscr{E}}\boldsymbol{\sigma} := (\mathscr{E}\sigma_1, \mathscr{E}\sigma_2)^T$. Thus, substituting (1.40) into (1.39), summing over all $e \in \mathcal{E}_h^\partial$, and using the continuity of $\boldsymbol{\mathscr{E}}$ and Assumptions **D**, we obtain (1.36).

On the other hand, we note that $\operatorname{div} \mathbf{E}_h(\boldsymbol{\zeta}_h)(\mathbf{y}) = \mathbf{E}_h(\operatorname{div} \boldsymbol{\zeta}_h)(\mathbf{y})$ for all $\mathbf{y} \in \widetilde{K}^e_{ext}$. Then, mimicking the steps that led us to (1.36), but this time taking $w_e := \mathrm{T}_e^m(\mathscr{E}(\operatorname{div} \boldsymbol{\sigma})) \in \mathrm{P}_s(\widetilde{B}^e)$, with $s < m$, instead of $\boldsymbol{\psi}_e$, we readily deduce that

$$\begin{aligned}\sum_{e\in\mathcal{E}_h^\partial} \|\operatorname{div}\,(\boldsymbol{\sigma} - \mathbf{E}_h(\boldsymbol{\zeta}_h))\,\|_{0,\widetilde{K}^e_{ext}} &\leq \sum_{e\in\mathcal{E}_h^\partial} \Big(\|\operatorname{div} \boldsymbol{\sigma} - w_e\|_{0,\widetilde{K}^e_{ext}} + \|\operatorname{div} \mathbf{E}_h(\boldsymbol{\zeta}_h) - w_e\|_{0,\widetilde{K}^e_{ext}} \Big) \\ &\lesssim h^{m+1} \|\operatorname{div} \boldsymbol{\sigma}\|_{m+1,\Omega} + \|\operatorname{div}\,(\boldsymbol{\sigma} - \boldsymbol{\zeta}_h)\|_{0,\mathrm{D}_h}.\end{aligned}$$

Combined with (1.36) this implies (1.37). □

We now propose suitable approximations for $\boldsymbol{\sigma}$ and u in D_h^c. These approximations, in abuse of notation, will also be named $\boldsymbol{\sigma}_h$ and u_h. For this, we let $(\boldsymbol{\sigma}_h, u_h) \in \mathbf{H}_h(\mathrm{D}_h) \times \mathrm{Q}_h(\mathrm{D}_h)$ be the unique solution of (1.12).

First, to approximate $\boldsymbol{\sigma}$ in D_h^c, we proceed analogously to [57, Section 2.1.3] and simply take the extrapolation of $\boldsymbol{\sigma}_h$ in D_h^c, that is, for any $e \in \mathcal{E}_h^\partial$ and any $\mathbf{y} \in \widetilde{K}^e_{ext}$, we define

$$\boldsymbol{\sigma}_h(\mathbf{y}) := \mathbf{E}_h(\boldsymbol{\sigma}_h)(\mathbf{y}). \tag{1.41}$$

Note that, for each edge $e \in \mathcal{E}_h^\partial$, the extrapolation of $\boldsymbol{\sigma}_h|_{K^e}$ to $\widetilde{K}^e_{ext}$ belongs to $\mathbf{H}(\operatorname{div}; \widetilde{K}^e_{ext})$, but not necessarily to $\mathbf{H}(\operatorname{div}; \mathrm{D}_h^c)$. Consequently, for the subsequent analysis we introduce the broken space (see, for instance [67])

$$\mathbf{H}(\operatorname{div}; \widetilde{\mathcal{T}}_h) := \prod_{e\in\mathcal{E}_h^\partial} \mathbf{H}(\operatorname{div}; \widetilde{K}^e_{ext}),$$

endowed with the broken norm $\|\boldsymbol{\xi}\|_{\operatorname{div},\widetilde{\mathcal{T}}_h} := \left(\sum_{e\in\mathcal{E}_h^\partial} \|\boldsymbol{\xi}\|^2_{\operatorname{div},\widetilde{K}^e_{ext}} \right)^{1/2}$.

The following result establishes the estimate for $(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)$ in D_h^c.

Lemma 1.8. *Suppose that assumptions of Lemma 1.7 are satisfied. Then, there hold*

$$\begin{aligned}\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathrm{div},\widetilde{\mathcal{T}}_h} \lesssim & \inf_{\boldsymbol{\zeta}_h\in\mathbf{H}_h(\mathrm{D}_h)}\|\boldsymbol{\sigma}-\boldsymbol{\zeta}_h\|_{\mathrm{div},\mathrm{D}_h}+\inf_{w_h\in\mathrm{Q}_h(\mathrm{D}_h)}\|u-w_h\|_{0,\mathrm{D}_h}\\ &+\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathrm{div},\mathrm{D}_h}+h^{m+1}\left(\|\boldsymbol{\sigma}\|_{m+1,\Omega}+\|\mathrm{div}\,\boldsymbol{\sigma}\|_{m+1,\Omega}\right)\end{aligned} \tag{1.42}$$

and

$$\begin{aligned}\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{0,\mathrm{D}_h^c} \lesssim & \inf_{\boldsymbol{\zeta}_h\in\mathbf{H}_h(\mathrm{D}_h)}\|\boldsymbol{\sigma}-\boldsymbol{\zeta}_h\|_{0,\mathrm{D}_h}+\inf_{w_h\in\mathrm{Q}_h(\mathrm{D}_h)}\|u-w_h\|_{0,\mathrm{D}_h}\\ &+\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{0,\mathrm{D}_h}+h^{m+1}\|\boldsymbol{\sigma}\|_{m+1,\Omega}.\end{aligned} \tag{1.43}$$

Proof. Let $\boldsymbol{\zeta}_h\in\mathbf{H}_h(\mathrm{D}_h)$. Applying estimate (1.37) and using definition (1.15), we obtain

$$\begin{aligned}&\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathrm{div},\widetilde{\mathcal{T}}_h}\\ &\quad\le\sum_{e\in\mathcal{E}_h^\partial}\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathrm{div},\widetilde{K}_{ext}^e}\le\sum_{e\in\mathcal{E}_h^\partial}\left(\|\boldsymbol{\sigma}-\mathbf{E}_h(\boldsymbol{\zeta}_h)\|_{\mathrm{div},\widetilde{K}_{ext}^e}+\|\mathbf{E}_h(\boldsymbol{\zeta}_h)-\boldsymbol{\sigma}_h\|_{\mathrm{div},\widetilde{K}_{ext}^e}\right)\\ &\quad\lesssim\|\boldsymbol{\sigma}-\boldsymbol{\zeta}_h\|_{\mathrm{div},\mathrm{D}_h}+h^{m+1}\left(\|\boldsymbol{\sigma}\|_{m+1,\Omega}+\|\mathrm{div}\,\boldsymbol{\sigma}\|_{m+1,\Omega}\right)+\sum_{e\in\mathcal{E}_h^\partial}\widetilde{C}_{ext}^e\,(\widetilde{r}_e)^{1/2}\,\|\boldsymbol{\zeta}_h-\boldsymbol{\sigma}_h\|_{\mathrm{div},T^e}.\end{aligned} \tag{1.44}$$

Thanks to Assumption **(D.1)**, the last term on the right-hand side of (1.44) is bounded as follows:

$$\sum_{e\in\mathcal{E}_h^\partial}\widetilde{C}_{ext}^e\,(\widetilde{r}_e)^{1/2}\,\|\boldsymbol{\zeta}_h-\boldsymbol{\sigma}_h\|_{\mathrm{div},T^e}\lesssim\|\boldsymbol{\zeta}_h-\boldsymbol{\sigma}_h\|_{\mathrm{div},\mathrm{D}_h}. \tag{1.45}$$

From (1.44) and (1.45), adding an subtracting $\boldsymbol{\sigma}$ to $\|\boldsymbol{\zeta}_h-\boldsymbol{\sigma}_h\|_{\mathrm{div},\mathrm{D}_h}$, and applying estimate (1.29), we obtain

$$\begin{aligned}\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathrm{div},\widetilde{\mathcal{T}}_h}\lesssim & \inf_{w_h\in\mathrm{Q}_h(\mathrm{D}_h)}\|u-w_h\|_{0,\mathrm{D}_h}+\|\boldsymbol{\sigma}-\boldsymbol{\zeta}_h\|_{\mathrm{div},\mathrm{D}_h}\\ &+\sum_{e\in\mathcal{E}_h^\partial}|||\boldsymbol{\sigma}-\mathbf{E}_h(\boldsymbol{\zeta}_h)|||_e+h^{m+1}\left(\|\boldsymbol{\sigma}\|_{m+1,\Omega}+\|\mathrm{div}\,\boldsymbol{\sigma}\|_{m+1,\Omega}\right).\end{aligned} \tag{1.46}$$

Furthermore, by the equivalence of norms given by Lemma 1.6, and considering once more the estimate (1.37), we find

$$\begin{aligned}\sum_{e\in\mathcal{E}_h^\partial}|||\boldsymbol{\sigma}-\mathbf{E}_h(\boldsymbol{\zeta}_h)|||_e&\lesssim\sum_{e\in\mathcal{E}_h^\partial}\|\boldsymbol{\sigma}-\mathbf{E}_h(\boldsymbol{\zeta}_h)\|_{0,\widetilde{K}_{ext}^e}\\ &\lesssim\|\boldsymbol{\sigma}-\boldsymbol{\zeta}_h\|_{\mathrm{div},\mathrm{D}_h}+h^{m+1}\left(\|\boldsymbol{\sigma}\|_{m+1,\Omega}+\|\mathrm{div},\boldsymbol{\sigma}\|_{m+1,\Omega}\right),\end{aligned}$$

and then (1.46) implies

$$\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathrm{div},\widetilde{\mathcal{T}}_h}\lesssim\inf_{w_h\in\mathrm{Q}_h(\mathrm{D}_h)}\|u-w_h\|_{0,\mathrm{D}_h}+\|\boldsymbol{\sigma}-\boldsymbol{\zeta}_h\|_{\mathrm{div},\mathrm{D}_h}+h^{m+1}\left(\|\boldsymbol{\sigma}\|_{m+1,\Omega}+\|\mathrm{div}\,\boldsymbol{\sigma}\|_{m+1,\Omega}\right),$$

which clearly gives (1.42). The estimate (1.43) is obtained analogously, but considering the estimate (1.36) instead of (1.37). □

Now, to define the approximation of u in D_h^c, we proceed again analogously to [57, Section 2.1.3] and adopt the same ideas as for $\widetilde{g}_h$ defined in (1.9). More precisely, given $e\in\mathcal{E}_h^\partial$, for any point $\mathbf{y}\in\widetilde{K}_{ext}^e$ there is a path $\mathscr{C}(\mathbf{x})$, starting at $\mathbf{x}\in\Gamma_h$ and ending at $\widetilde{\mathbf{x}}\in\Gamma$, such that we can write $\mathbf{y}=\mathbf{x}+(\eta/\ell(\mathbf{x}))(\widetilde{\mathbf{x}}-\mathbf{x})$ for some $\eta\in[0,\ell(\mathbf{x})]$. Then, for any $e\in\mathcal{E}_h^\partial$ and $\mathbf{y}\in\widetilde{K}_{ext}^e$, we set

$$u_h(\mathbf{y}):=u(\widetilde{\mathbf{y}})-\int_0^{|\widetilde{\mathbf{y}}-\mathbf{y}|}\boldsymbol{\sigma}_h(\mathbf{y}+s\mathbf{w}(\mathbf{y}))\cdot\mathbf{w}(\mathbf{y})ds, \tag{1.47}$$

where $\widetilde{\mathbf{y}} := \widetilde{\mathbf{x}}$, $\mathbf{w}(\mathbf{y}) := (\widetilde{\mathbf{y}} - \mathbf{y})/|\widetilde{\mathbf{y}} - \mathbf{y}|$ and $\boldsymbol{\sigma}_h$ is defined as in (1.41).

We now address the estimate for $(u - u_h)$ by using the L^2-norm on D_h^c.

Lemma 1.9. *Suppose that assumptions of Lemmas 1.6 and 1.7 are satisfied. Then, there holds*

$$\|u - u_h\|_{0,\mathrm{D}_h^c} \lesssim h^{m+2}\|\boldsymbol{\sigma}\|_{m+1,\Omega} + h\left(\inf_{w_h \in \mathrm{Q}_h(\mathrm{D}_h)} \|u - w_h\|_{0,\mathrm{D}_h} + \inf_{\boldsymbol{\zeta}_h \in \mathbf{H}_h(\mathrm{D}_h)} \|\boldsymbol{\sigma} - \boldsymbol{\zeta}_h\|_{\mathrm{div},\mathrm{D}_h}\right). \quad (1.48)$$

Proof. We first use (1.33) to obtain

$$\|u - u_h\|^2_{0,\mathrm{D}_h^c} \leq \sum_{e \in \mathcal{E}_h^\partial} (C_2^e)^2 \, |||u - u_h|||_e^2 = \sum_{e \in \mathcal{E}_h^\partial} (C_2^e)^2 \int_e \int_0^{\ell(\mathbf{x})} |u - u_h|^2(\mathbf{x} + t\mathbf{m}(\mathbf{x}))\, dt\, dS_{\mathbf{x}}. \quad (1.49)$$

Let $\mathbf{y} = \mathbf{x} + t\mathbf{m}(\mathbf{x})$. Using the definition of $u_h(\mathbf{y})$ (cf. (1.47)) and the fact that $\widetilde{\mathbf{y}} = \widetilde{\mathbf{x}}$ and $\mathbf{w}(\mathbf{y}) = \mathbf{m}(\mathbf{x})$, we find

$$\begin{aligned}(u - u_h)(\mathbf{y}) &= -\int_0^{|\widetilde{\mathbf{y}} - \mathbf{y}|} (\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)(\mathbf{y} + s\mathbf{w}(\mathbf{y})) \cdot \mathbf{w}(\mathbf{y})\, ds \\ &= -\int_0^{(\ell(\mathbf{x}) - t)} (\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)(\mathbf{x} + (t+s)\mathbf{m}(\mathbf{x})) \cdot \mathbf{m}(\mathbf{x})\, ds.\end{aligned}$$

From this, applying the Cauchy–Schwarz inequality and a simple change of variables, yields

$$\begin{aligned}|u - u_h|^2(\mathbf{y}) &\leq (\ell(\mathbf{x}) - t) \int_t^{\ell(\mathbf{x})} |(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)(\mathbf{x} + r\mathbf{m}(\mathbf{x}))|^2\, dr \\ &\leq \ell(\mathbf{x}) \int_0^{\ell(\mathbf{x})} |(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)(\mathbf{x} + r\mathbf{m}(\mathbf{x}))|^2\, dr.\end{aligned} \quad (1.50)$$

Substituting (1.50) into (1.49), it follows that

$$\|u - u_h\|^2_{0,\mathrm{D}_h^c} \leq \sum_{e \in \mathcal{E}_h^\partial} (C_2^e)^2 \int_e \ell(\mathbf{x})^2 \int_0^{\ell(\mathbf{x})} |(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)(\mathbf{x} + r\mathbf{m}(\mathbf{x}))|^2\, dr. \quad (1.51)$$

Since $\ell(\mathbf{x}) \leq \widetilde{H}_e = \widetilde{r}_e h_e^\perp \leq \widetilde{r}_e h_{K^e}$, we obtain, thanks to Assumption **(D.1)** and (1.34), that

$$\|u - u_h\|^2_{0,\mathrm{D}_h^c} \leq \sum_{e \in \mathcal{E}_h^\partial} (C_2^e)^2\, \widetilde{r}_e h_{K^e} |||\boldsymbol{\sigma} - \boldsymbol{\sigma}_h|||_e^2 \leq (Rh)^2 \max_{e \in \mathcal{E}_h^\partial} (C_2^e)^2\, (C_1^e)^{-2} \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|^2_{0,\mathrm{D}_h^c},$$

and then the result follows from (1.43). □

Remark 1.1. *The solvability and error analyses in previous sections depend on how the computational domain and transferring paths are constructed since the Assumption* **A**, **D** *and the assumptions of Lemma 1.6 need to be satisfied. We recall that in our present setting the domain is immersed in a background mesh and* D_h *is the union of all the elements inside* Ω*, and hence* $d(\Gamma, \Gamma_h) \lesssim h$*. Although the Assumption* **(D.2)** *cannot be verified in practice for this case, we will see in Section 1.5 that it provides optimal performance of the method.*

Remark 1.2. *We now illustrate an alternative way to construct the computational domain* D_h*. If* Ω *is convex, we can construct* Γ_h *interpolating* Γ *by a piecewise linear function. Thus, the subdomain* D_h *is the region enclosed by* Γ_h *and the transferring paths associated to the interior points of a boundary*

edge e can be chosen so that they are perpendicular to e. In this setting, Assumptions **A** *and* **D** *hold and actually $\widetilde{r}_e$ is of order h. Moreover, the norms $\|\cdot\|_{0,\widetilde{K}^e_{ext}}$ and $|||\cdot|||$ coincide, and hence Assumptions i)-iii) of Lemma 1.6 are not necessary. If Ω is not convex, we can proceed similarly and our analysis still holds under the additional assumption that the solution $(\boldsymbol{\sigma}, u)$ of (1.2) can be extended to the region $\Omega^c \cap \mathrm{D}_h$.*

1.4 Particular choice of finite elements

Given an integer $k \geq 0$ and a set $\mathcal{O}$ in $\mathbb{R}^2$, we denote by $\widetilde{\mathrm{P}}_k(\mathcal{O}) \subset \mathrm{P}_k(\mathcal{O})$ the space of polynomials of total degree equal to k. In addition, we define the local Raviart–Thomas space of order k, for each $K \in \mathcal{T}_h$, as

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

where $\mathbf{x} := (x_1, x_2)^T$ is a generic vector of $\mathbb{R}^2$, and $\mathbf{P}_k(K)$ stands for the space of vector-valued polynomials of degree at most k on the element K. A concrete example of discrete spaces is then given by the sets:

$$\begin{aligned} \mathbf{H}_h(\mathrm{D}_h) &:= \left\{ \boldsymbol{\tau}_h \in \mathbf{H}(\mathrm{div}; \mathrm{D}_h) : \ \boldsymbol{\tau}_h|_K \in \mathbf{RT}_k(K) \ \ \forall K \in \mathcal{T}_h \right\}, \\ \mathrm{Q}_h(\mathrm{D}_h) &:= \left\{ v_h \in \mathrm{L}^2(\mathrm{D}_h) : \ v_h|_K \in \mathrm{P}_k(K) \ \ \forall K \in \mathcal{T}_h \right\}. \end{aligned} \tag{1.52}$$

It is well-known that these spaces satisfy Assumptions **B** and **C** in Section 1.2.4. Moreover, they have the following approximation properties (see, e.g., [73, 88]):

($\mathbf{AP}^{\boldsymbol{\sigma}}_h$) For each $r \in (0, k+1]$, and for each $\boldsymbol{\sigma} \in \mathbf{H}^r(\mathrm{D}_h) \cap \mathbf{H}(\mathrm{div}; \mathrm{D}_h)$ with $\mathrm{div}\,\boldsymbol{\sigma} \in \mathrm{H}^r(\mathrm{D}_h)$, there holds

$$\inf_{\boldsymbol{\zeta}_h \in \mathbf{H}_h(\mathrm{D}_h)} \|\boldsymbol{\sigma} - \boldsymbol{\zeta}_h\|_{\mathrm{div}, \mathrm{D}_h} \lesssim h^r \left(\|\boldsymbol{\sigma}\|_{r, \mathrm{D}_h} + \|\mathrm{div}\,\boldsymbol{\sigma}\|_{r, \mathrm{D}_h} \right).$$

($\mathbf{AP}^{u}_h$) For each $r \in (0, k+1]$, and for each $u \in \mathrm{H}^r(\mathrm{D}_h)$, there holds

$$\inf_{w_h \in \mathrm{Q}_h(\mathrm{D}_h)} \|u - w_h\|_{0, \mathrm{D}_h} \lesssim h^r \|u\|_{r, \mathrm{D}_h}.$$

The following theorem establishes the *a priori* error estimates for the Galerkin scheme (1.12) under suitable regularity assumptions on the exact solution. It also provides estimates of the error in the non-meshed region D^c_h.

Theorem 1.10. *In addition to the hypotheses of Theorem 1.5, Lemmas 1.6 and 1.7, suppose that the exact solution $(\boldsymbol{\sigma}, u)$ satisfies $\boldsymbol{\sigma} \in \mathbf{H}^{k+1}(\Omega) \cap \mathbf{H}(\mathrm{div}; \Omega)$, with $\mathrm{div}\,\boldsymbol{\sigma} \in \mathrm{H}^{k+1}(\Omega)$, and $u \in \mathrm{H}^{k+1}(\Omega)$. Then, there hold*

$$\begin{aligned} \|(\boldsymbol{\sigma}, u) - (\boldsymbol{\sigma}_h, u_h)\|_{\mathbf{H}(\mathrm{div}; \mathrm{D}_h) \times \mathrm{L}^2(\mathrm{D}_h)} &\lesssim h^{k+1} \left(\|\boldsymbol{\sigma}\|_{k+1, \Omega} + \|\mathrm{div}\,\boldsymbol{\sigma}\|_{k+1, \Omega} + \|u\|_{k+1, \Omega} \right), \\ \|\boldsymbol{\sigma} - \mathbf{E}_h(\boldsymbol{\sigma}_h)\|_{\mathrm{div}, \widetilde{\mathcal{T}}_h} &\lesssim h^{k+1} \left(\|\boldsymbol{\sigma}\|_{k+1, \Omega} + \|\mathrm{div}\,\boldsymbol{\sigma}\|_{k+1, \Omega} + \|u\|_{k+1, \Omega} \right), \\ \|u - u_h\|_{0, \mathrm{D}^c_h} &\lesssim h^{k+2} \left(\|\boldsymbol{\sigma}\|_{k+1, \Omega} + \|\mathrm{div}\,\boldsymbol{\sigma}\|_{k+1, \Omega} + \|u\|_{k+1, \Omega} \right). \end{aligned}$$

Proof. The result follows from Theorem 1.5, Lemmas 1.8, 1.9, and the approximations properties $\mathbf{(AP}_h^u\mathbf{)}$ and $\mathbf{(AP}_h^{\boldsymbol{\sigma}}\mathbf{)}$ specified above. □

Remark 1.3. *The theory developed above covers other similar finite element subspaces available in the literature, such as the local Brezzi–Douglas–Marini space of order $k \geq 1$:*

$$\mathbf{BDM}_k(K) := \mathbf{P}_k(K).$$

More precisely, one can also choose the discrete spaces in (1.8) as

$$\mathbf{H}_h(\mathrm{D}_h) := \left\{ \boldsymbol{\tau}_h \in \mathbf{H}(\mathrm{div};\mathrm{D}_h) : \ \boldsymbol{\tau}_h|_K \in \mathbf{BDM}_k(K) \ \ \forall\, K \in \mathcal{T}_h \right\},$$
$$\mathrm{Q}_h(\mathrm{D}_h) := \left\{ v_h \in \mathrm{L}^2(\mathrm{D}_h) : \ v_h|_K \in \mathrm{P}_{k-1}(K) \quad \forall\, K \in \mathcal{T}_h \right\},$$

and obtain the well–posedness of the discrete problem and optimal error estimates, as well.

1.5 Numerical results

In this section we present numerical experiments in two dimensions illustrating the performance of the discrete scheme introduced and analized in Section 1.2. The numerical results shown below were obtained using a MATLAB code. As a direct solver we used UMFPACK [64]. In all the computations we consider the specific finite element subspaces $\mathbf{H}_h(\mathrm{D}_h)$ and $\mathrm{Q}_h(\mathrm{D}_h)$ in (1.52) with $k = 0, \ldots, 3$. With regard to this, an important issue is the computational implementation of specific basis functions providing high order approximations. This is facilitated through the use of *hierarchical basis* for the local Raviart–Thomas space of order k, as was introduced in [24], and *Dubiner basis* for the local polynomial space of degree at most k (see, e.g., [66]).

We begin by introducing additional notation. Firstly, we must take into account that, in all our examples, the computational domain D_h and the region D_h^c change with h. That is why we compute the relative errors:

$$\mathrm{e}_{\mathrm{int}}(u) := \frac{\|u - u_h\|_{0,\mathrm{D}_h}}{\|u\|_{0,\mathrm{D}_h}}, \qquad \mathrm{e}_{\mathrm{int}}(\boldsymbol{\sigma}) := \frac{\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathrm{div},\mathrm{D}_h}}{\|\boldsymbol{\sigma}\|_{\mathrm{div},\mathrm{D}_h}},$$
$$\mathrm{e}_{\mathrm{ext}}(u) := \frac{\|u - u_h\|_{0,\mathrm{D}_h^c}}{\|u\|_{0,\mathrm{D}_h^c}}, \qquad \mathrm{e}_{\mathrm{ext}}(\boldsymbol{\sigma}) := \frac{\|\boldsymbol{\sigma} - \mathbf{E}_h(\boldsymbol{\sigma}_h)\|_{\mathrm{div},\widetilde{\mathcal{T}}_h}}{\|\boldsymbol{\sigma}\|_{\mathrm{div},\widetilde{\mathcal{T}}_h}}.$$

Furthermore, we define the experimental rates of convergence as

$$\mathrm{r}_{\mathrm{int}}(\cdot) := -2[\log(\mathrm{e}_{\mathrm{int}}(\cdot)/\mathrm{e'}_{\mathrm{int}}(\cdot))/\log(N/N')], \qquad \mathrm{r}_{\mathrm{ext}}(\cdot) := -2[\log(\mathrm{e}_{\mathrm{ext}}(\cdot)/\mathrm{e'}_{\mathrm{ext}}(\cdot))/\log(N/N')],$$

where N and N' denote the number of elements of two consecutive meshes with their respective errors $\mathrm{e}_{\mathrm{int}}$ and $\mathrm{e}'_{\mathrm{int}}$ (resp. $\mathrm{e}_{\mathrm{ext}}$ and $\mathrm{e}'_{\mathrm{ext}}$).

Example 1. We take $u(x_1, x_2) := \sin(\pi x_1)\sin(\pi x_2)$ as exact solution, and choose Ω to be the annular domain consisting in two concentric circles of radius 1.5 and 0.7, respectively. As required by Assumption **(D.1)**, the subdomain D_h is constructed in such a way that the distance $d(\Gamma, \Gamma_h)$ is at most of $\mathcal{O}(h)$. In doing so, we consider a *background triangulation* T_h of the square $\mathcal{B} \supset \Omega$, obtained

by subdividing the squares of the Cartesian grid into four congruent triangles, and then follow the process in Section 1.2.1 to choose those elements of $\mathcal{T}_h$ inside of Ω. In Table 1.1 we present the history of convergence and observe that the convergence rates predicted by Theorem 1.10 are attained by all the unknowns, namely $\mathcal{O}(h^{k+1})$ for $\mathrm{e}_{\mathrm{int}}(\boldsymbol{\sigma})$, $\mathrm{e}_{\mathrm{ext}}(\boldsymbol{\sigma})$ and $\mathrm{e}_{\mathrm{int}}(u)$, and $\mathcal{O}(h^{k+2})$ for $\mathrm{e}_{\mathrm{ext}}(u)$. Next, in Figure 1.5 we display the approximate value of the second component of $\boldsymbol{\sigma}$, denoted by $\sigma_{h,2}$, obtained for the approximation $\mathbf{RT}_3 - \mathrm{P}_3$ with total number of degrees of freedom (d.o.f) equal to 32560 and $N = 1152$ elements. The corresponding extrapolated solution on the set D_h^c is also displayed there.

				Errors on D_h				Errors on D_h^c			
k	N	h	d.o.f	$\mathrm{e}_{\mathrm{int}}(u)$	$\mathrm{r}_{\mathrm{int}}(u)$	$\mathrm{e}_{\mathrm{int}}(\boldsymbol{\sigma})$	$\mathrm{r}_{\mathrm{int}}(\boldsymbol{\sigma})$	$\mathrm{e}_{\mathrm{ext}}(u)$	$\mathrm{r}_{\mathrm{ext}}(u)$	$\mathrm{e}_{\mathrm{ext}}(\boldsymbol{\sigma})$	$\mathrm{r}_{\mathrm{ext}}(\boldsymbol{\sigma})$
0	248	0.262	664	$2.28e-01$	–	$2.30e-01$	–	$9.84e-03$	–	$2.99e-01$	–
	1152	0.131	2956	$1.08e-01$	0.96	$1.10e-01$	0.96	$2.28e-03$	1.90	$1.24e-01$	1.14
	4840	0.065	12260	$5.31e-02$	0.99	$5.39e-02$	0.99	$5.52e-04$	1.97	$6.50e-02$	0.90
	22028	0.031	55352	$2.20e-02$	1.16	$2.26e-02$	1.14	$1.62e-04$	1.61	$3.28e-02$	0.90
	89384	0.015	224020	$1.09e-02$	0.99	$1.12e-02$	0.99	$3.22e-05$	2.30	$1.58e-02$	1.03
1	248	0.262	2072	$2.79e-02$	–	$2.37e-02$	–	$3.62e-03$	–	$1.08e-01$	–
	1152	0.131	9368	$5.44e-03$	2.13	$5.51e-03$	1.90	$4.72e-04$	2.65	$2.43e-02$	1.94
	4840	0.065	39040	$1.32e-03$	1.96	$1.36e-03$	1.95	$5.35e-05$	3.03	$6.70e-03$	1.79
	22028	0.031	176790	$2.95e-04$	1.97	$3.03e-04$	1.97	$5.02e-06$	3.12	$1.79e-03$	1.74
	89384	0.015	716200	$7.36e-05$	1.98	$7.56e-05$	1.98	$5.86e-07$	3.06	$4.41e-04$	1.99
2	248	0.262	4224	$6.51e-03$	–	$2.88e-03$	–	$9.75e-04$	–	$2.16e-02$	5.87
	1152	0.131	19236	$2.74e-04$	4.12	$2.58e-04$	3.14	$4.57e-05$	3.98	$1.76e-03$	3.26
	4840	0.065	80340	$3.01e-05$	3.07	$3.13e-05$	2.93	$4.51e-06$	3.22	$2.97e-04$	2.48
	22028	0.031	364310	$2.32e-06$	3.38	$2.42e-06$	3.37	$1.18e-07$	4.80	$2.53e-05$	3.24
	89384	0.015	1476500	$2.84e-07$	2.99	$2.96e-07$	3.00	$7.08e-09$	4.02	$2.92e-06$	3.08
3	248	0.262	7120	$2.27e-03$	–	$7.27e-04$	–	$2.25e-04$	–	$5.59e-03$	–
	1152	0.131	32560	$2.83e-05$	5.70	$1.70e-05$	4.88	$3.48e-06$	5.43	$2.82e-04$	3.88
	4840	0.065	136160	$1.16e-06$	4.44	$1.22e-06$	3.67	$2.25e-07$	3.81	$2.56e-05$	3.34
	22028	0.031	617910	$1.67e-08$	5.59	$2.18e-08$	5.31	$1.72e-09$	6.43	$1.52e-06$	3.72
	89384	0.015	2505000	$9.51e-10$	4.09	$1.14e-09$	4.20	$4.27e-11$	5.28	$8.87e-08$	4.05

Table 1.1: History of convergence of the approximation in Example 1. (table produced by the author)

Example 2. We set f and g such that $u(x_1, x_2) := x_1^2 \exp(2(x_2 - 1))$, and consider the kidney-shaped domain Ω whose boundary satisfies the equation

$$(2[(x_1 + 0.5)^2 + x_2^2] - x_1 - 0.5)^2 - [(x_1 + 0.5)^2 + x_2^2] + 0.1 = 0.$$

The way to construct D_h is the same as in the previous example. In Table 1.2 we present the corresponding convergence history and again observe there that optimal convergence rates predicted by Theorem 1.10 are reached by all the unknowns. In Figure 1.6 we display the approximate value of the first component of $\boldsymbol{\sigma}$, denoted by $\sigma_{h,1}$, obtained for the approximation $\mathbf{RT}_3 - \mathrm{P}_3$ with total number of degrees of freedom (d.o.f) equal to 18480 and $N = 654$ elements.

Example 3. We consider exactly the same domain Ω as in Example 2, but this time we choose $u(x_1, x_2) := \sin(10\pi x_1 - 5\pi x_2)$ as exact solution instead. The goal is to explore how the error of our method is affected when we consider keeping a triangulation of $\overline{\mathrm{D}_h}$ fixed and varying the degree k of the finite element spaces in (1.52). In Figure 1.7 we show the results for three fixed meshes with $N = 146$, $N = 654$ and $N = 3068$ elements, respectively. As expected, it can be appreciated there that the quality of the approximations improves as h diminishes or k increases.

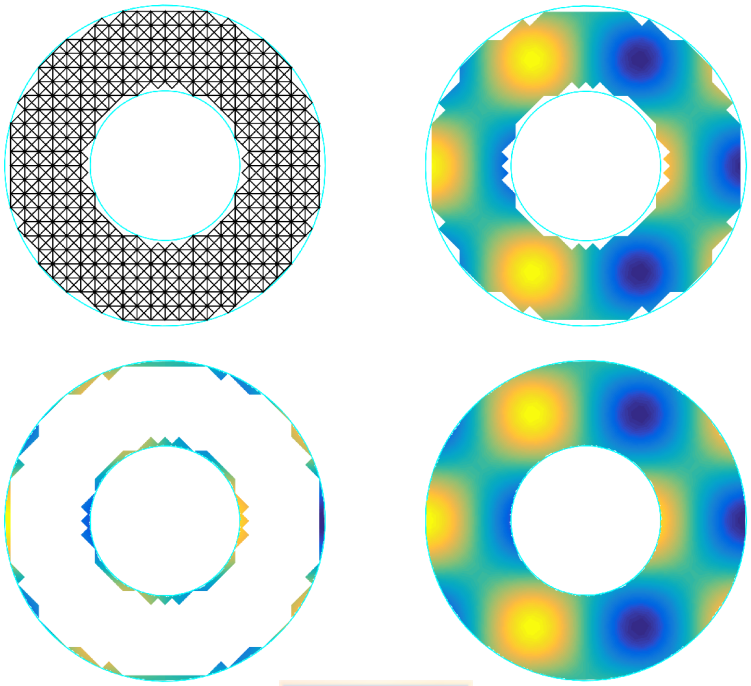


Figure 1.5: Example 1: $\sigma_{h,2}$ for the approximation $\mathbf{RT}_3 - \mathrm{P}_3$ with $N = 1152$ elements. (figure produced by the author)

				Errors on D_h				Errors on D_h^c			
k	N	h	d.o.f	$\mathrm{e}_{\mathrm{int}}(u)$	$\mathrm{r}_{\mathrm{int}}(u)$	$\mathrm{e}_{\mathrm{int}}(\boldsymbol{\sigma})$	$\mathrm{r}_{\mathrm{int}}(\boldsymbol{\sigma})$	$\mathrm{e}_{\mathrm{ext}}(u)$	$\mathrm{r}_{\mathrm{ext}}(u)$	$\mathrm{e}_{\mathrm{ext}}(\boldsymbol{\sigma})$	$\mathrm{r}_{\mathrm{ext}}(\boldsymbol{\sigma})$
0	146	0.131	384	$1.65e-01$	–	$5.12e-02$	–	$3.51e-03$	–	$1.01e-01$	–
	654	0.065	1677	$7.88e-02$	0.98	$2.61e-02$	0.89	$1.51e-03$	1.12	$5.25e-02$	0.88
	3068	0.031	7748	$3.84e-02$	0.93	$1.12e-02$	1.09	$2.91e-04$	2.13	$2.63e-02$	0.89
	12579	0.015	31602	$1.89e-02$	0.99	$5.64e-03$	0.97	$7.13e-05$	1.99	$1.32e-02$	0.96
	50877	0.007	127500	$9.44e-03$	0.99	$2.82e-03$	0.98	$1.66e-05$	2.08	$6.68e-03$	0.98
1	146	0.131	1206	$1.22e-02$	–	$2.19e-03$	–	$4.48e-04$	–	$8.88e-03$	–
	654	0.065	5316	$2.68e-03$	2.02	$5.23e-04$	1.91	$6.60e-05$	2.55	$2.40e-03$	1.74
	3068	0.031	24700	$6.84e-04$	1.76	$1.17e-04$	1.93	$7.47e-06$	2.81	$6.89e-04$	1.61
	12579	0.015	100940	$1.67e-04$	1.99	$2.89e-05$	1.98	$9.24e-07$	2.96	$1.78e-04$	1.91
	50877	0.007	4076400	$4.12e-05$	2.00	$7.20e-06$	1.99	$1.29e-07$	2.81	$4.29e-05$	2.03
2	146	0.131	2466	$2.74e-04$	–	$6.18e-05$	–	$1.59e-05$	–	$5.59e-04$	–
	654	0.065	10917	$3.16e-05$	2.88	$1.23e-05$	2.14	$2.66e-06$	2.38	$9.84e-05$	2.31
	3068	0.031	50856	$2.83e-06$	3.12	$5.62e-07$	4.00	$6.59e-08$	4.78	$1.09e-05$	2.83
	12579	0.015	208020	$3.40e-07$	3.00	$6.31e-08$	3.10	$2.96e-09$	4.39	$1.36e-06$	2.95
	50877	0.007	840400	$4.20e-08$	2.99	$7.67e-09$	3.01	$2.00e-10$	3.85	$1.66e-07$	3.01
3	146	0.131	4164	$4.76e-06$	–	$1.58e-06$	4.94	$6.26e-07$	–	$2.62e-05$	–
	654	0.065	18480	$4.73e-07$	3.07	$2.78e-07$	2.31	$6.11e-08$	3.10	$3.00e-06$	2.89
	3068	0.031	86216	$9.83e-09$	5.01	$4.52e-09$	5.33	$6.07e-10$	5.96	$1.40e-07$	3.96
	12579	0.015	352830	$5.27e-10$	4.14	$1.66e-10$	4.67	$1.42e-11$	5.32	$8.00e-09$	4.06
	50877	0.007	1425800	$3.15e-11$	4.03	$8.91e-12$	4.19	$4.87e-13$	4.82	$5.05e-10$	3.95

Table 1.2: History of convergence of the approximation in Example 2. (table produced by the author)

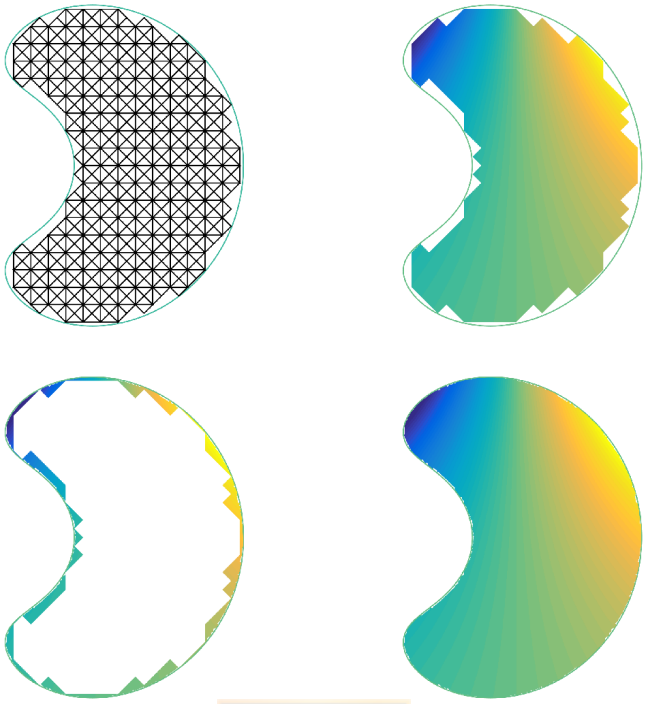


Figure 1.6: Example 2: $\sigma_{h,2}$ for the approximation $\mathbf{RT}_3 - \mathrm{P}_3$ with $N = 654$ elements. (figure produced by the author)

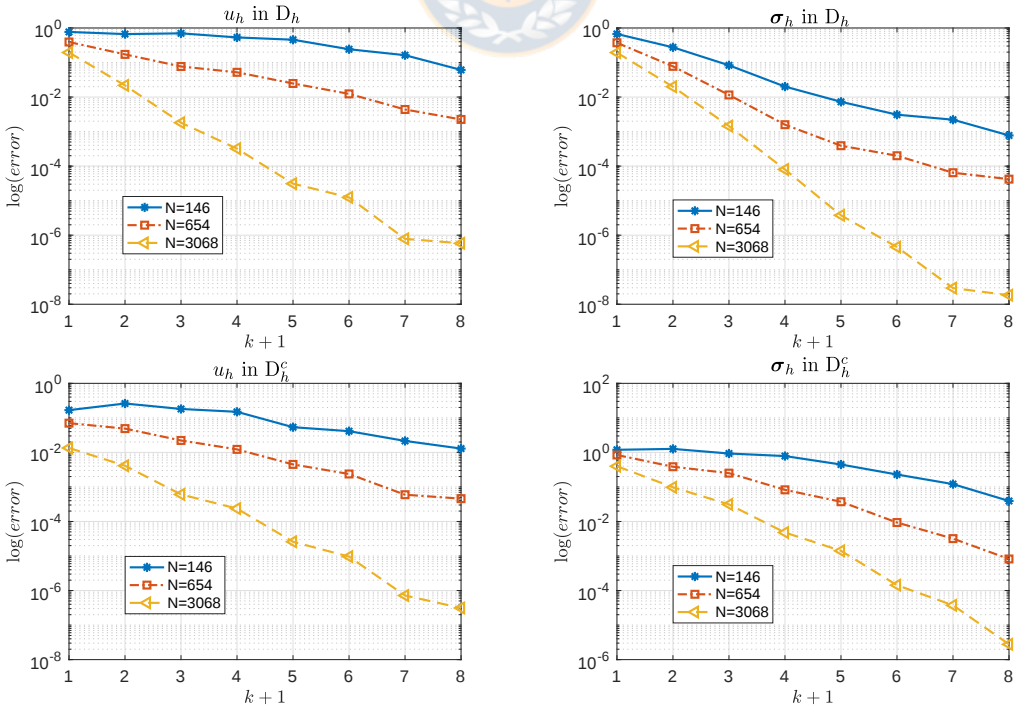


Figure 1.7: Example 3: Log of the error vs $(k+1)$ for $k = 0, \ldots, 7$ and three fixed meshes. (figure produced by the author)

Example 4. In our last experiment, we observe the performance of the method considering another type of computational domain, as Remark 1.2 mentioned. We take $u(x_1, x_2) := \sin(x_1)\sin(x_2)$ as exact solution and consider Ω to be the annular domain consisting of two concentric circles of radius 2 and 0.5, respectively. In this case, the computational boundary Γ_h is defined through a piecewise linear interpolation of Γ as Figure 1.8 shows. Here the distance $\mathrm{d}(\Gamma, \Gamma_h)$ is at most of $\mathcal{O}(h^2)$. Table 1.3 shows that the experimental rates of convergence for $\mathrm{e}_{\mathrm{int}}(\boldsymbol{\sigma})$, $\mathrm{e}_{\mathrm{ext}}(\boldsymbol{\sigma})$ and $\mathrm{e}_{\mathrm{int}}(u)$ are optimal, i.e., $\mathcal{O}(h^{k+1})$. In addition, the convergence rate of $\mathrm{e}_{\mathrm{ext}}(u)$ is of $\mathcal{O}(h^{k+3})$. This behavior can be explained by the proof of Lemma 1.9. In fact, since $\widetilde{r}_e$ is now of order h, the estimate (1.48) yields

$$\|u - u_h\|_{0,\mathrm{D}_h^c} \lesssim h^2 \|\boldsymbol{\sigma} - \mathbf{E}_h(\boldsymbol{\sigma}_h)\|_{0,\mathrm{D}_h^c} \lesssim h^{k+3}.$$

				Errors on D_h				Errors on D_h^c			
k	N	h	d.o.f	$\mathsf{e}_{\mathtt{int}}(u)$	$\mathsf{r}_{\mathtt{int}}(u)$	$\mathsf{e}_{\mathtt{int}}(\boldsymbol{\sigma})$	$\mathsf{r}_{\mathtt{int}}(\boldsymbol{\sigma})$	$\mathsf{e}_{\mathtt{ext}}(u)$	$\mathsf{r}_{\mathtt{ext}}(u)$	$\mathsf{e}_{\mathtt{ext}}(\boldsymbol{\sigma})$	$\mathsf{r}_{\mathtt{ext}}(\boldsymbol{\sigma})$
0	150	0.660	395	$1.39e-01$	–	$1.26e-01$	–	$8.03e-05$	–	$9.38e-02$	–
	608	0.355	1560	$6.89e-02$	1.00	$6.25e-02$	1.00	$9.51e-06$	3.04	$4.62e-02$	1.01
	2396	0.187	6070	$3.50e-02$	0.98	$3.16e-02$	0.99	$1.16e-06$	3.05	$2.35e-02$	0.98
	9358	0.095	23555	$1.78e-02$	0.99	$1.61e-02$	0.99	$1.46e-07$	3.04	$1.18e-02$	1.00
	37798	0.050	94815	$8.98e-03$	0.98	$8.05e-03$	0.99	$1.87e-08$	2.94	$5.99e-03$	0.97
1	150	0.660	1240	$8.68e-03$	–	$1.10e-02$	–	$2.42e-05$	–	$1.39e-02$	–
	608	0.355	4944	$2.23e-03$	1.93	$2.73e-03$	1.99	$1.68e-06$	3.81	$3.36e-03$	2.03
	2396	0.187	19328	$5.69e-04$	1.99	$6.98e-04$	1.99	$1.02e-07$	4.07	$9.36e-04$	1.86
	9358	0.095	75184	$1.46e-04$	1.98	$1.79e-04$	1.99	$7.27e-09$	3.89	$2.44e-04$	1.97
	37798	0.050	30302	$3.63e-05$	2.00	$4.44e-05$	1.99	$4.72e-10$	3.91	$6.31e-05$	1.94
2	150	0.660	2535	$5.86e-04$	–	$5.65e-04$	–	$1.44e-06$	–	$6.96e-04$	–
	608	0.355	10152	$6.99e-05$	3.03	$7.02e-05$	2.98	$5.43e-08$	4.68	$1.01e-04$	2.74
	2396	0.187	39774	$8.92e-06$	3.00	$9.17e-06$	2.96	$1.70e-09$	5.04	$1.40e-05$	2.88
	9358	0.095	154890	$1.14e-06$	3.01	$1.18e-06$	2.99	$5.64e-11$	5.00	$1.73e-06$	3.06
	37798	0.050	624630	$1.42e-07$	2.98	$1.47e-07$	2.99	$1.90e-12$	4.85	$2.40e-07$	2.83
3	150	0.660	4280	$1.81e-05$	–	$2.36e-05$	–	$6.37e-08$	–	$4.11e-05$	–
	608	0.355	17184	$1.29e-06$	3.77	$1.49e-06$	3.94	$8.55e-10$	6.16	$2.80e-06$	3.84
	2396	0.187	67408	$8.32e-08$	4.00	$9.66e-08$	3.99	$1.58e-11$	5.81	$1.78e-07$	4.01
	9358	0.095	262660	$5.54e-09$	3.97	$6.37e-09$	3.99	$2.55e-13$	6.05	$1.21e-08$	3.93
	37798	0.050	1059600	$3.38e-10$	4.00	$3.90e-10$	3.99	$4.13e-15$	5.90	$8.40e-10$	3.83

Table 1.3: History of convergence of the approximation in Example 4. (table produced by the author)

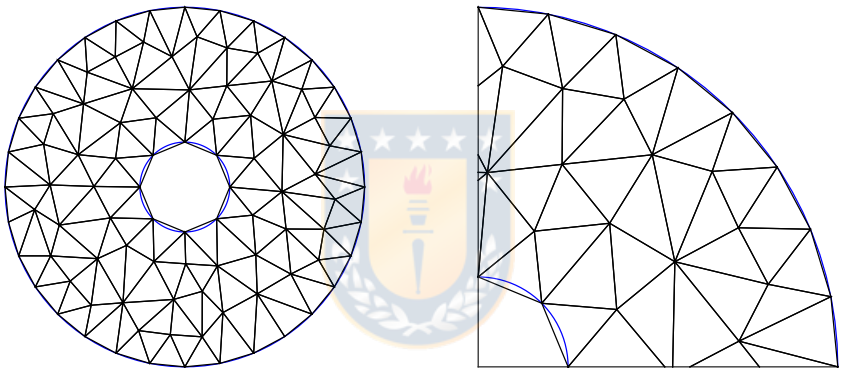


Figure 1.8: Example 4: Left, mesh with $N = 150$ elements where Γ_h is constructed through a piecewise linear interpolation of the boundary Γ (blue line) and right, part of the domain Ω that lies in the first quadrant of the Cartesian plane. (figure produced by the author)

CHAPTER 2

A priori and *a posteriori* error analyses of a high order unfitted mixed-FEM for Stokes flow

In this chapter we continue the investigation on unfitted mixed finite element methods by extending the theory presented in Chapter 1 to the pseudostress-velocity formulation of the incompressible Stokes equations. For the case when the computational boundary is constructed through a piecewise linear interpolation of the curved boundary, we furthermore introduce a reliable and quasi-efficient residual-based *a posteriori* error estimator. Numerical experiments illustrate the performance of the scheme, show the behaviour of the associated adaptive algorithm and validate the theory.

2.1 Introduction

It is well-known that standard Galerkin procedures devised to solve PDEs on curved domains Ω do not achieve high order accuracy whenever Ω is approximated by a nearby domain D_h. In principle, neither the regularity of the solution nor the smoothness of the curved boundary Γ are the reasons behind the loss of accuracy. Instead, the difficulties arise from the *variational crimes* (see, e.g., [37, Chapter 10]) given by an eventual noncoforming method. For stationary problems, isoparametric elements can be efficiently implemented without much difficulty. However, in evolving domains, remeshing is a major issue for body-fitted approaches.

Alternatively, unfitted methods, minimize the complexity of mesh generation by, for instance, immersing Ω in a background uniform mesh and setting, for example, D_h as the union of all the elements of the mesh that lie inside Ω. However, one of the main challenges in this case is the imposition of the boundary data on the computational domain. One of the first contribution in this context was developed in the seventies by [34], where the boundary-value correction is based on Nitsche's approach [105] combined with the polygonal domain approximation method of [134]. Since then, a vast literature related to unfitted methods using this type of penalty approach for interface problems can be found, mostly for low order approximations. During the last years, the development of cut finite element method (CutFEM) [42, 43, 44] has strengthened the capabilities of non-body fitted approaches, even in the high order case. In fact, there are numerical techniques to improve the cut cell integration for level set domains and achieve higher order accuracy [103]. Theoretical framework in this context has

been also developed [95].

Provided a domain Ω with Lipschitz continuous and picewise $\mathcal{C}^2$ boundary Γ, a novel high order unfitted method for Dirichlet boundary value problems has been proposed in the context of hybridizable discontinuous Galerkin (HDG) methods [56, 57, 59]. More precisely, denoting by u the variable such that $\sigma := \nabla u$ in Ω and $u = g$ on Γ, it consists of transferring the Dirichlet datum g from Γ to Γ_h by integrating σ along a family of segments joining both boundaries, which are usually referred to as *transferring paths*. At the discrete level, the transferred data, say $\widetilde{g}$, is approximated by $\widetilde{g}_h$ obtained by integrating the extrapolation of the discrete approximation of σ along the transferring paths. Thus, the problem is solved in D_h and its solution is extended by local extrapolations to D_h^c. It is remarkable that the method keeps high order accuracy when the distance $d(\Gamma, \Gamma_h)$ between Γ and Γ_h is only of order of the meshsize h. Also, it covers the case where Γ_h is constructed through a picewiese linear interpolation of Γ. In addition, also in the context of HDG methods, this *transferring technique* has been successfully applied to a wide variety of problems in continuum mechanics, including exterior diffusion equations [58], convection-diffusion problems [60], the semi-linear Grad–Shafranov equation [122], the Stokes equations for incompressible flow [126], and the Oseen equations [125]. It has been also extended to a diffusion problem with mixed boundary conditions and to an elliptic transmission problem where the interface is not piecewise flat, for which we refer to [117].

Another method based on the idea of imposing the boundary data on a computational boundary which is order h away from Γ is the shifted boundary method [99, 100]. There, the approximate data $\widetilde{g}_h$ is obtained by a Taylor expansion of the solution at the boundary points. Finally, we would like to mention a recent approach developed by [52] based on polynominal extrapolations. Roughly speaking, that method forces a polynomial extension of the approximate solution to match the prescribed boundary data on Γ.

On the other hand, in Chapter 1 we proposed and analyzed a high order unfitted mixed finite element method for diffusion problems where the Dirichlet datum is transferred according to the above transferring technique. Considering general finite dimensional subspaces, we showed the well-posed of the discrete formulation by means of the classical Babuška–Brezzi theory (see, e.g., [73]). In particular, we showed that Raviart–Thomas elements of order order $k \geq 0$ for the vectorial variable and discontinuous polynomials of degree k for the scalar variable, ensure unique solvability and optimal convergence of $\mathcal{O}(h^{k+1})$ of the associated Galerkin scheme, which rely only on some hypotheses involving the *closeness* between Γ and Γ_h.

According to the above, our first goal in this chapter is to additionally contribute in the direction of Chapter 1 and provide a high order unfitted mixed-FEM for the incompressible Stokes equations in which the pseudostress tensor [45] and the fluid velocity are the only unknowns, whereas the pressure is computed via a postprocessing procedure. We refer the reader to the early work of Gatica et al. [78] (see also [46]), for the analysis of this problem in polyhedral domains. A few points for this choice deserve comments. First, the pseudostress tensor has been widely used to overcome the well-known disadvantages of considering the symmetric stress tensor (see, e.g., [10, 13, 15, 41]). Indeed, in the modeling equations the pseudostress takes the place of the stress without requiring symmetry. In addition, an accurate direct calculation of further physical quantities such as the velocity gradient, the vorticity and the stress, can be expressed in terms of the pseudostress discretization via a simple postprocessing procedure, and with the same accuracy. Finally, we remark that, different

from the work by Solano and Vargas [126], here the novelty lies on the treatment of the pseudostress approximation in D_h.

Now, in addition to the loss of accuracy over curved domains, the numerical approximation could be deteriorated by singularities or high gradients of the solution, often as a result of domains with re-entrant corners or solutions having interior/boundary layers. In order to guarantee a good convergence behaviour in those cases, one usually needs to apply an adaptive mesh refinement near the critical region; for a survey, we refer the reader to [141]. The elements to be refined are marked according to a global estimator Θ given in terms of local indicators Θ_T on each element T of a given mesh. The estimator Θ is said to be efficient (resp. reliable) if there exists $C_{\mathrm{eff}} > 0$ (resp. $C_{\mathrm{rel}} > 0$), independent of the meshsizes, such that

$$C_{\mathrm{eff}}\Theta + \text{h.o.t.} \leq \|\text{error}\| \leq C_{\mathrm{rel}}\Theta + \text{h.o.t.},$$

where h.o.t. is a generic expression denoting one or several high order terms. In particular, concerning our problem of interest, a residual-based *a posteriori* error estimator has been developed by [78]. However, in all the proofs, Ω has been assumed to be polyhedral.

In this chapter, provided Γ is interpolated by a piecewise linear function, we further contribute in developing the first residual-based *a posteriori* error analysis for Stokes flow where the above mentioned transferring technique is employed. Unlike the polygonal case, our estimator is efficient up to calculable terms involving curved segments and a postprocessed velocity with enhanced accuracy. It is important to remark that the literature regarding high order approximations and adaptive mesh refinement on curved domains is scarce. Up to the authors's knowledge, probably the only work treating this matter was carried out in [8], where the Poisson problem was solved by using the *hp* finite element method [15], combined with isoparametric elements fitting a Lipschitz continuous and piecewise $\mathcal{C}^{k+2}$ boundary Γ (for $k \geq 0$). However, the associated *hp* adaptivity strategy is difficult to implement. Indeed, at each refinement step and on each marked element, it must be decided whether to refine the mesh (h-version of FEM) or increase the polynomial degree (p-version of FEM). In our analysis the assumption on Γ can be relaxed to piecewise $\mathcal{C}^2$ only. Moreover, our adaptive algorithm keeps the polynomial degree fixed and improves the accuracy of the approximation by refining the mesh without the need of using isoparametric elements.

We have organized this chapter as follows. In the remainder of this section we recall recurrent notation and general definitions. Next, in Section 2.2 we present the model problem and recall its classical dual-mixed formulation, having the pseudostress tensor and the fluid velocity as main unknowns. In Section 2.3, the fluid domain Ω is approximated by a polyhedral subdomain D_h where a high order Galerkin scheme is introduced and analyzed. Next, an *a priori* error analysis, involving hypotheses of *closeness* between Γ and Γ_h, is derived in Section 2.4. Moreover, in Section 2.5 we derive our *a posteriori* error estimator and establish its main properties, as long as Γ is interpolated by a piecewise linear function. Finally, in Section 2.6 we present numerical experiments validating the theory.

In the sequel, when no confusions arises, $|\cdot|$ will denote the Euclidean norm in $\mathbb{R}^n$, $n = 2, 3$. In turn, given tensor fields $\boldsymbol{\sigma} := (\sigma_{ij})_{1\leq i,j\leq n}$ and $\boldsymbol{\tau} := (\tau_{ij})_{1\leq i,j\leq n}$, we let $\mathbf{div}\,\boldsymbol{\tau}$ be the divergence operator, div, acting along the rows of $\boldsymbol{\tau}$, and define the trace $\mathrm{tr}\,(\boldsymbol{\tau}) := \sum_{i=1}^n \boldsymbol{\tau}_{ii}$, the inner product $\boldsymbol{\sigma} : \boldsymbol{\tau} := \sum_{i,j=1}^n \sigma_{ij}\tau_{ij}$, and the deviatoric tensor $\boldsymbol{\tau}^{\mathtt{d}} := \boldsymbol{\tau} - \frac{1}{n}\mathrm{tr}\,(\boldsymbol{\tau})\mathbb{I}$, where $\mathbb{I}$ stand for the identity tensor in $\mathbb{R}^{n\times n}$. Also, we adopt standard simplified terminology for Sobolev spaces and norms, where

spaces of vector-valued and tensor-valued functions are denoted in bold face and blackboard bold face, respectively. For instance, if $\mathcal{O}$ is a domain in $\mathbb{R}^n$, $\mathscr{C}$ is an open or closed Lipschitz curve (resp. surface in $\mathbb{R}^3$), and $s \in \mathbb{R}$, we define

$$\mathbf{H}^s(\mathcal{O}) := [\mathrm{H}^s(\mathcal{O})]^n\,, \quad \mathbb{H}^s(\mathcal{O}) := [\mathrm{H}^s(\mathcal{O})]^{n\times n} \quad \text{and} \quad \mathbf{H}^s(\mathscr{C}) := [\mathrm{H}^s(\mathscr{C})]^n\,,$$

with the convention that $\mathbf{H}^0(\mathcal{O}) = \mathbf{L}^2(\mathcal{O})$, $\mathbb{L}^2(\mathcal{O}) = \mathbb{H}^0(\mathcal{O})$ and $\mathbf{L}^2(\mathscr{C}) = \mathbf{H}^0(\mathscr{C})$. The corresponding norms are denoted by $\|\cdot\|_{s,\mathcal{O}}$ and $\|\cdot\|_{s,\mathscr{C}}$, whereas the seminorm is denoted by $|\cdot|_{s,\mathcal{O}}$. Furthermore, we recall that

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

equipped with the norm $\|\cdot\|_{\mathbf{div},\mathcal{O}} := \left(\|\cdot\|^2_{0,\mathcal{O}} + \|\mathbf{div}\,(\cdot)\|^2_{0,\mathcal{O}} \right)^{1/2}$, is a Hilbert space. Note that if $\boldsymbol{\tau} \in \mathbb{H}(\mathbf{div};\mathcal{O})$, then $\boldsymbol{\tau}\mathbf{n}_{\partial\mathcal{O}} \in \mathbf{H}^{-1/2}(\partial\mathcal{O})$, where $\mathbf{H}^{1/2}(\partial\mathcal{O})$ is the space of traces of $\mathbf{H}^1(\mathcal{O})$, $\mathbf{H}^{-1/2}(\partial\mathcal{O})$ corresponds to the dual space of $\mathbf{H}^{1/2}(\partial\mathcal{O})$, and $\mathbf{n}_{\partial\mathcal{O}}$ denotes the outward unit normal vector on $\partial\mathcal{O}$. Hereafter, $\langle\cdot,\cdot\rangle_{\partial\mathcal{O}}$ denotes the duality pairing between $\mathbf{H}^{-1/2}(\partial\mathcal{O})$ and $\mathbf{H}^{1/2}(\partial\mathcal{O})$ with respect to the $\mathbf{L}^2(\mathcal{O})$-inner product. The following estimate (see, e.g., [73, Theorem 1.7]) holds:

$$\|\boldsymbol{\tau}\mathbf{n}\|_{-1/2,\partial\mathcal{O}} \leq \|\boldsymbol{\tau}\|_{\mathbf{div},\mathcal{O}} \quad \forall\, \boldsymbol{\tau} \in \mathbb{H}(\mathbf{div};\mathcal{O}). \tag{2.1}$$

In addition, by $\mathbf{0}$ we will refer to the generic null vector (including the null functional and operator), and we will denote by C and c, with or without subscripts, bars, tildes or hats, generic constants independent of the meshsize, but might depend on the polynomial degree, the shape-regularity of the triangulation and the domain. Furthermore, for quantities A and B, we write $A \lesssim B$, whenever there exists $C > 0$ such that $A \leq CB$. Finally, $A \simeq B$ stands for both $A \lesssim B$ and $B \lesssim A$ being satisfied.

2.2 The continuous problem

2.2.1 Governing equations

Let Ω be a bounded and open, not necessarily polyhedral region with boundary Γ, which we assume to be piecewise $\mathcal{C}^2$ and Lipschitz continuous. We are interested in approximating, by a mixed finite element method, the Stokes equations describing a steady viscous incompressible fluid flow occupying Ω, under the action of external forces, given by

$$\begin{aligned} &\boldsymbol{\sigma} = 2\mu\nabla\mathbf{u} - p\mathbb{I} \quad \text{in} \quad \Omega, \quad \mathbf{div}\,\boldsymbol{\sigma} = -\mathbf{f} \quad \text{in} \quad \Omega, \\ &\mathrm{div}\,\mathbf{u} = 0 \quad \text{in} \quad \Omega, \quad \mathbf{u} = \mathbf{g} \quad \text{on} \quad \Gamma, \quad \int_\Omega p = 0. \end{aligned} \tag{2.2}$$

Here, the unknowns are the fluid velocity $\mathbf{u}$, the fluid pressure p, and the so-called pseudostress tensor $\boldsymbol{\sigma}$; the given data are a volume force $\mathbf{f} \in \mathbf{L}^2(\Omega)$ and the boundary velocity $\mathbf{g} \in \mathbf{H}^{1/2}(\Gamma)$, while the kinematic viscosity μ is a positive constant. Note that the incompressibility constraint $\mathrm{div}\,\mathbf{u} = 0$ in Ω, which expresses the conservation of mass, enforces that $\mathbf{g}$ must satisfy the compatibility condition

$$\int_\Gamma \mathbf{g}\cdot\mathbf{n}_\Gamma = 0, \tag{2.3}$$

where $\mathbf{n}_\Gamma$ stands for the outward unit normal vector to Γ. Furthermore, the last condition in (2.2) is added to ensure uniqueness of solution, and this will lead us to the introduction of the space $\mathrm{L}^2_0(\Omega) := \{q \in \mathrm{L}^2(\Omega) : \ \int_\Omega q = 0\}$.

2.2.2 The pseudostress-velocity formulation

In what follows, we briefly recall the pseudostress-velocity formulation employed in [46] and [78] for the Stokes problem described in the previous section. Let us first remark that taking the matrix trace operator in the first equation and using the incompressibility condition, we easily obtain the postprocessing formula

$$p = -\frac{1}{2}\mathrm{tr}\,(\boldsymbol{\sigma}) \quad \text{in} \quad \Omega. \tag{2.4}$$

In this way, using (2.4) we can eliminate p from (2.2), obtaining

$$\begin{gathered} \frac{1}{2\mu}\boldsymbol{\sigma}^{\mathtt{d}} = \nabla \mathbf{u} \quad \text{in} \quad \Omega, \quad \mathbf{div}\,\boldsymbol{\sigma} = -\boldsymbol{f} \quad \text{in} \quad \Omega, \\ \mathbf{u} = \mathbf{g} \quad \text{on} \quad \Gamma, \quad \int_\Omega \mathrm{tr}\,(\boldsymbol{\sigma}) = 0. \end{gathered} \tag{2.5}$$

Notice that the last condition is a consequence of (2.4) and of the requirement on the pressure space, and this therefore suggests the introduction of the space $\mathbb{H}_0(\mathbf{div};\Omega) := \{\boldsymbol{\tau} \in \mathbb{H}(\mathbf{div};\Omega) : \int_\Omega \mathrm{tr}\,(\boldsymbol{\tau}) = 0\}$ satisfying $\mathbb{H}(\mathbf{div};\Omega) = \mathbb{H}_0(\mathbf{div};\Omega) \oplus \mathrm{P}_0(\Omega)\mathbb{I}$, where $\mathrm{P}_0(\Omega)$ is the space of constant polynomials defined on Ω. More precisely, each $\boldsymbol{\tau} \in \mathbb{H}(\mathbf{div};\Omega)$ can be decomposed uniquely as $\boldsymbol{\tau} = \boldsymbol{\tau}_0 + c\mathbb{I}$, with

$$\boldsymbol{\tau}_0 := \boldsymbol{\tau} - \left(\frac{1}{2|\Omega|}\int_\Omega \mathrm{tr}\,(\boldsymbol{\tau})\right)\mathbb{I} \in \mathbb{H}_0(\mathbf{div};\Omega) \quad \text{and} \quad c := \frac{1}{2|\Omega|}\int_\Omega \mathrm{tr}\,(\boldsymbol{\tau}) \in \mathbb{R}.$$

As a consequence of the above, from (2.5) it is not difficult to obtain the following variational formulation of (2.5): Find $(\boldsymbol{\sigma}, \mathbf{u}) \in \mathbb{H}_0(\mathbf{div};\Omega) \times \mathbf{L}^2(\Omega)$ such that

$$\begin{aligned} a(\boldsymbol{\sigma}, \boldsymbol{\tau}) + b(\boldsymbol{\tau}, \mathbf{u}) &= \langle \boldsymbol{\tau}\mathbf{n}_\Gamma, \mathbf{g}\rangle_\Gamma \quad \forall\, \boldsymbol{\tau} \in \mathbb{H}_0(\mathbf{div};\Omega), \\ b(\boldsymbol{\sigma}, \mathbf{v}) &= -\int_\Omega \mathbf{f}\cdot\mathbf{v} \quad \forall\, \in \mathbf{L}^2(\Omega), \end{aligned} \tag{2.6}$$

where $\mathbf{n}_\Gamma$ stands for the outward unit normal vector on Γ, whereas the bounded bilinear forms $a : \mathbb{H}(\mathbf{div};\Omega) \times \mathbb{H}(\mathbf{div};\Omega) \to \mathbb{R}$ and $b : \mathbb{H}(\mathbf{div};\Omega) \times \mathbf{L}^2(\Omega)$ are given, respectively, by

$$a(\boldsymbol{\sigma}, \boldsymbol{\tau}) := \frac{1}{2\mu}\int_\Omega \boldsymbol{\sigma}^{\mathtt{d}} : \boldsymbol{\tau}^{\mathtt{d}} \quad \text{and} \quad b(\boldsymbol{\tau}, \mathbf{v}) := \int_\Omega \mathbf{v}\cdot\mathbf{div}\,\boldsymbol{\tau}.$$

We refer the reader to [78, Theorem 2.1] for the well-posedness analysis of this problem. In particular, the respective continuous dependence result provided by the classical Babuška–Brezzi theorem (see, for instance [78, Theorem 2.3]), implies that the following global inf-sup conditions holds:

$$\|(\boldsymbol{\zeta}, \mathbf{w})\|_{\mathbb{H}(\mathbf{div};\Omega)\times\mathbf{L}^2(\Omega)} \lesssim \sup_{\substack{(\boldsymbol{\tau},\mathbf{v}) \in \mathbb{H}_0(\mathbf{div};\Omega)\times\mathbf{L}^2(\Omega) \\ (\boldsymbol{\tau},\mathbf{v}) \neq \mathbf{0}}} \frac{|a(\boldsymbol{\zeta}, \boldsymbol{\tau}) + b(\boldsymbol{\tau}, \mathbf{w}) + b(\boldsymbol{\zeta}, \mathbf{v})|}{\|(\boldsymbol{\tau}, \mathbf{v})\|_{\mathbb{H}(\mathbf{div};\Omega)\times\mathbf{L}^2(\Omega)}} \tag{2.7}$$

for all $(\boldsymbol{\zeta}, \mathbf{w}) \in \mathbb{H}_0(\mathbf{div};\Omega) \times \mathbf{L}^2(\Omega)$, where $\|(\boldsymbol{\zeta}, \mathbf{w})\|_{\mathbb{H}(\mathbf{div};\Omega)\times\mathbf{L}^2(\Omega)} := \left(\|\boldsymbol{\zeta}\|^2_{\mathbf{div},\Omega} + \|\mathbf{w}\|^2_{0,\Omega}\right)^{1/2}$. The specific purpose of this estimate will become clear below in Section 2.5 when dealing with the *a posteriori* error analysis.

To end this section, we remark that the solution of (2.6) solves the original problem (2.5) in the sense of the following lemma. The proof is omitted because it is straightforward.

Lemma 2.1. *Let $(\boldsymbol{\sigma}, \mathbf{u}) \in \mathbb{H}_0(\mathbf{div};\Omega) \times \mathbf{L}^2(\Omega)$ be the unique solution of (2.6). It satisfies in a distributional sense, $(2\mu)^{-1}\boldsymbol{\sigma}^{\mathtt{d}} = \nabla\mathbf{u}$ in Ω, and $\mathbf{div}\,\boldsymbol{\sigma} = -\mathbf{f}$ in Ω. Moreover, $\mathbf{u} \in \mathbf{H}^1(\Omega)$ and satisfies the boundary condition described in (2.5).*

2.3 The Galerkin scheme

Throughout this section, by the sake of simplicity, we will develop the theory for the two-dimensional case. The results that we will present can be extended to the three-dimensional case, but some of them require technicalities that will be addressed in Appendix B.

2.3.1 Preliminary results

In the context of curved domains, we now proceed as in Chapter 1 (see also [57, 59] for HDG methods) and suppose that there exists a family of subdomains D_h of the fluid region Ω having a polygonal boundary $\Gamma_h := \partial\mathrm{D}_h$, which may not necessarily fit the true boundary Γ. The index h will refer to the size of a given triangulation of $\overline{\mathrm{D}_h}$. For ease of presentation, in this section we develop the theory and postpone the construction of D_h to Sections 2.5 and 2.6.

As a consequence of Lemma 2.1, we can infer that the solution of (2.6) satisfies in a distributional sense,

$$\frac{1}{2\mu}\boldsymbol{\sigma}^{\mathtt{d}} = \nabla\mathbf{u} \quad \text{in} \quad \mathrm{D}_h, \quad \mathbf{div}\,\boldsymbol{\sigma} = -\mathbf{f} \quad \text{in} \quad \mathrm{D}_h. \tag{2.8}$$

In turn, following the approach of [59], the trace of $\mathbf{u}$ on Γ_h, denoted by $\widetilde{\mathbf{g}}$, can be conveniently rewritten in terms of $\boldsymbol{\sigma}$. Indeed, integrating componentwise $\frac{1}{2\mu}\boldsymbol{\sigma}^{\mathtt{d}} = \nabla\mathbf{u}$ along a segment, say $\mathscr{C}(\mathbf{x})$, starting at $\mathbf{x} \in \Gamma_h$ and ending at $\widetilde{\mathbf{x}} \in \Gamma$, which is often referred to as *transferring path* and whose definition will be detailed in Section 2.3.2, we get

$$\widetilde{\mathbf{g}}(\mathbf{x}) = \overline{\mathbf{g}}(\mathbf{x}) - \frac{1}{2\mu}\int_{\mathscr{C}(\mathbf{x})} \boldsymbol{\sigma}^{\mathtt{d}}\,\mathbf{m}(\mathbf{x})\,d\eta, \tag{2.9}$$

where $\overline{\mathbf{g}}(\mathbf{x}) := \mathbf{g}(\widetilde{\mathbf{x}}(\mathbf{x}))$ and $\mathbf{m}(\mathbf{x})$ is the unit tangent vector to $\mathscr{C}(\mathbf{x})$. Clearly, this definition coincides with the trace of $\mathbf{u}$ on Γ_h, as it does not depend on the integration path. Moreover, when high order accuracy is required, the line integral in (2.9) allows us to obtain a better approximation of $\widetilde{\mathbf{g}}$ than the trace of the finite element solution associated to $\mathbf{u}$ on Γ_h.

Next, after reducing the equations of (2.8) to a weak form and using (2.9), it readily follows that the solution of (2.6) satisfies

$$\int_{\mathrm{D}_h} \mathrm{tr}\,(\boldsymbol{\sigma}) = -\int_{\mathrm{D}_h^c} \mathrm{tr}\,(\boldsymbol{\sigma}) \quad \text{with} \quad \mathrm{D}_h^c := \Omega \setminus \overline{\mathrm{D}_h}, \tag{2.10}$$

and

$$\begin{aligned} a_h(\boldsymbol{\sigma},\boldsymbol{\tau}) + b_h(\boldsymbol{\tau},\mathbf{u}) &= \langle \boldsymbol{\tau}\mathbf{n}_{\Gamma_h}, \widetilde{\mathbf{g}}\rangle_{\Gamma_h} \quad \forall\,\boldsymbol{\tau} \in \mathbb{H}(\mathbf{div};\mathrm{D}_h), \\ b_h(\boldsymbol{\sigma},\mathbf{v}) &= F_h(\mathbf{v}) \qquad\qquad \forall\,\mathbf{v} \in \mathbf{L}^2(\mathrm{D}_h), \end{aligned} \tag{2.11}$$

where $\mathbf{n}_{\Gamma_h}$ denotes the unit vector pointing in the outward normal direction of Γ_h with respect to D_h, and $a_h(\cdot,\cdot)$ on $\mathbb{H}(\mathbf{div};\mathrm{D}_h)\times\mathbb{H}(\mathbf{div};\mathrm{D}_h)$, $b_h(\cdot,\cdot)$ on $\mathbb{H}(\mathbf{div};\mathrm{D}_h)\times\mathbf{L}^2(\mathrm{D}_h)$, and $F_h(\cdot)$ on $\mathbb{H}(\mathbf{div};\mathrm{D}_h)$, denote the forms defined, respectively, by

$$a_h(\boldsymbol{\sigma},\boldsymbol{\tau}) := \frac{1}{2\mu}\int_{\mathrm{D}_h} \boldsymbol{\sigma}^{\mathtt{d}} : \boldsymbol{\tau}^{\mathtt{d}}, \quad b_h(\boldsymbol{\tau},\mathbf{v}) := \int_{\mathrm{D}_h} \mathbf{v}\cdot\mathbf{div}\,\boldsymbol{\tau}, \quad F_h(\mathbf{v}) := -\int_{\mathrm{D}_h} \mathbf{f}\cdot\mathbf{v}. \tag{2.12}$$

However, defining $\boldsymbol{\sigma}_0 \in \mathbb{H}(\mathbf{div}; \mathrm{D}_h)$ by

$$\boldsymbol{\sigma}_0 := \boldsymbol{\sigma}|_{\mathrm{D}_h} - \left(\frac{\gamma}{2|\mathrm{D}_h|}\right)\mathbb{I} \quad \text{with} \quad \gamma := -\int_{\mathrm{D}_h^c} \mathrm{tr}\,(\boldsymbol{\sigma}), \tag{2.13}$$

it is not difficult to see that $\boldsymbol{\sigma}_0 \in \mathbb{H}_0(\mathbf{div}; \mathrm{D}_h)$ if and only if (2.10) holds, and therefore, the equations (2.10)-(2.11) can be rewritten, equivalently, as

$$\begin{aligned} a_h(\boldsymbol{\sigma}_0, \boldsymbol{\tau}) + b_h(\boldsymbol{\tau}, \mathbf{u}) &= \langle \boldsymbol{\tau}\mathbf{n}_{\Gamma_h}, \widetilde{\mathbf{g}}\rangle_{\Gamma_h} \quad \forall\, \boldsymbol{\tau} \in \mathbb{H}_0(\mathbf{div}; \mathrm{D}_h), \\ b_h(\boldsymbol{\sigma}_0, \mathbf{v}) &= F_h(\mathbf{v}) \qquad\qquad \forall\, \mathbf{v} \in \mathbf{L}^2(\mathrm{D}_h), \end{aligned} \tag{2.14}$$

provided that the compatibility condition $\int_{\Gamma_h} \widetilde{\mathbf{g}} \cdot \mathbf{n}_{\Gamma_h} = 0$ is satisfied. The latter is, indeed, a consequence of Gauss' divergence theorem and the equation $\mathrm{div}\,\mathbf{u} = 0$ in D_h, obtained from the first equation of (2.8) by applying the matrix trace operator. In addition, let us observe that since $\boldsymbol{\sigma}^{\mathtt{d}} = \boldsymbol{\sigma}_0^{\mathtt{d}}$, (2.9) can be written in terms of $\boldsymbol{\sigma}_0$ as

$$\widetilde{\mathbf{g}}(\mathbf{x}) = \overline{\mathbf{g}}(\mathbf{x}) - \frac{1}{2\mu}\int_{\mathscr{C}(\mathbf{x})} \boldsymbol{\sigma}_0^{\mathtt{d}}\,\mathbf{m}(\mathbf{x})\,d\eta, \tag{2.15}$$

Therefore, in what follows we propose a Galerkin scheme for (2.14). Before discussing further this matter, in the next section we introduce notation that will be useful to define our approximation in the region D_h^c.

2.3.2 Meshes and transferring paths

We consider a shape-regular family of triangulations $\{\mathcal{T}_h\}_{h>0}$ that subdivides the polygonal region $\overline{\mathrm{D}_h}$ into triangles T of diameter h_T and outward unit normal vector $\mathbf{n}_T$. Here, the index $h > 0$, refers to the meshsize $h := \max\{h_T : T \in \mathcal{T}_h\}$. Furthermore, we denote by $\mathcal{E}_h^i$ and $\mathcal{E}_h^\partial$ the sets of interior and boundary edges, respectively, and denote $\mathcal{E}_h = \mathcal{E}_h^i \cup \mathcal{E}_h^\partial$. Given $e \in \mathcal{E}_h$, we denote by T^e the element of $\mathcal{T}_h$ having e as an edge. In addition, to emphasize that a unit vector is normal to Γ_h or to an edge $e \in \mathcal{E}_h^\partial$, we will write $\mathbf{n}_{\Gamma_h}$ and $\mathbf{n}_e$, respectively.

We now turn to specify the family of transferring paths connecting Γ_h and Γ, and follow similar steps as in [59, Section 2.4] (see also Section 1.2.2). Given $e \in \mathcal{E}_h^\partial$, let $\mathbf{p}_1$ and $\mathbf{p}_2$ its two vertices. To each of them, we assign a unique point in Γ, denoted by $\widetilde{\mathbf{p}}_1$ and $\widetilde{\mathbf{p}}_2$, respectively. In the numerical experiment section we will describe how $\widetilde{\mathbf{p}}_i$ $(i = 1, 2)$ can be obtained. Now, let $\widehat{\mathbf{m}}^{\mathbf{p}_i} := \widetilde{\mathbf{p}}_i - \mathbf{p}_i$. We set $\mathbf{m}^{\mathbf{p}_i} := \widehat{\mathbf{m}}^{\mathbf{p}_i}/|\widehat{\mathbf{m}}^{\mathbf{p}_i}|$ if $|\widehat{\mathbf{m}}^{\mathbf{p}_i}| \neq 0$ and $\mathbf{m}^{\mathbf{p}_i} = \mathbf{n}_e$, otherwise. Given $\mathbf{x} \in e$, $\mathscr{C}(\mathbf{x})$ is determined as a convex combination of those paths originated from the vertices of e. More precisely, for $\theta \in [0, 1]$, we write $\mathbf{x} = \mathbf{p}_1(1-\theta) + \theta\mathbf{p}_2$ and define $\widehat{\mathbf{m}} := \mathbf{m}^{\mathbf{p}_1}(1-\theta) + \theta\mathbf{m}^{\mathbf{p}_2}$. Then, we write $\mathbf{m} := \widehat{\mathbf{m}}/|\widehat{\mathbf{m}}|$ if $|\widehat{\mathbf{m}}| \neq 0$ and $\mathbf{m} := \mathbf{n}_e$, otherwise. Thus, we set $\widetilde{\mathbf{x}}$ as the closest intersection between the boundary Γ and the ray starting at $\mathbf{x}$ whose unit tangent vector is $\mathbf{m}$. In other words, the transferring path connecting a point $\mathbf{x} \in \Gamma_h$ to a point $\widetilde{\mathbf{x}} \in \Gamma$, is given by

$$\mathscr{C}(\mathbf{x}) := \{\mathbf{x} + \eta\mathbf{m}(\mathbf{x}) : \ 0 \leq \eta \leq \ell(\mathbf{x}) := |\widetilde{\mathbf{x}} - \mathbf{x}|\}\,.$$

In addition, concerning our approximate solution outside D_h, we consider, for each boundary edge e with vertices $\mathbf{p}_1$ and $\mathbf{p}_2$, the cones

$$\begin{aligned} \mathrm{C}_{\mathbf{p}_1}^e &:= \left\{\mathbf{p}_1 + \varepsilon_1(\widetilde{\mathbf{p}}_1 - \mathbf{p}_1) + \varepsilon_2(\mathbf{p}_2 - \mathbf{p}_1) : \ \varepsilon_1, \varepsilon_2 \in \mathbb{R}^+\right\}, \\ \mathrm{C}_{\mathbf{p}_2}^e &:= \left\{\mathbf{p}_2 + \varepsilon_1(\widetilde{\mathbf{p}}_2 - \mathbf{p}_2) + \varepsilon_2(\mathbf{p}_1 - \mathbf{p}_2) : \ \varepsilon_1, \varepsilon_2 \in \mathbb{R}^+\right\}, \end{aligned}$$

and define, for $e \in \mathcal{E}_h^\partial$,

$$\widetilde{T}_{ext}^e := \{\mathscr{C}(\mathbf{x}) : \ \mathbf{x} \in e\} \cap \mathrm{C}_{\mathbf{p}_1}^e \cap \mathrm{C}_{\mathbf{p}_2}^e \cap \mathrm{D}_h^c.$$

Also, it will be convenient to write Γ_e to denote the intersection between Γ and the closure of the region $\widetilde{T}_{ext}^e$.

Finally, given $e \in \mathcal{E}_h^\partial$, the exterior region $\widetilde{T}_{ext}^e$ is said to be an *admissible set* if for every $\mathbf{x} \in e$, the intersection of the ray $\{\mathbf{x} + \varepsilon(\widetilde{\mathbf{x}} - \mathbf{x}) : \ \varepsilon \in \mathbb{R}^+\}$ with Γ is a single point (see left panel of Figure 2.1). According to the above and for the sake for simplicity, from now on we assume that $\widetilde{T}_{ext}^e$ is an *admissible set*, and denote by $\widetilde{\mathcal{T}}_h$ the partition of D_h^c into those sets. Therefore, for instance, cases like the one on the right of Figure 2.1 are not considered.

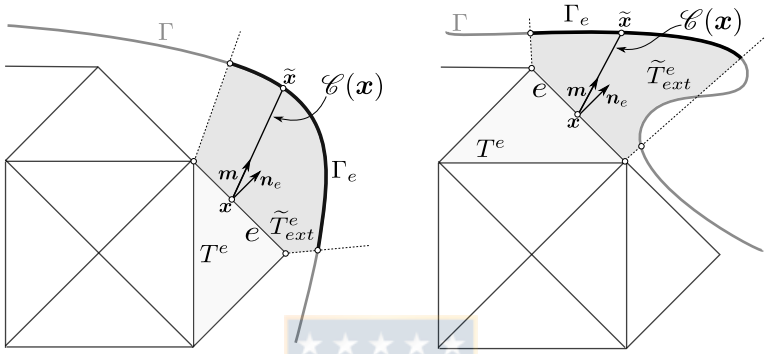


Figure 2.1: Examples of sets $\widetilde{T}_{ext}^e$. The *admissible case* is the one on the left. (figure produced by the author)

2.3.3 Statement of the Galerkin scheme

In this section we specify the Galerkin approximation of (2.14). It requires first some definitions. Given an integer $l \geq 0$ and a subset $\mathcal{O}$ of $\mathbb{R}^2$, we let $\mathrm{P}_l(\mathcal{O})$ (resp. $\widetilde{\mathrm{P}}_l(\mathcal{O})$) be the space of polynomials of degree at most l defined on $\mathcal{O}$ (resp. of degree equal to l) and according to the terminology described in Section 2.1, we set $\mathbf{P}_l(\mathcal{O}) := [\mathrm{P}_l(\mathcal{O})]^2$ and $\mathbb{P}_l(\mathcal{O}) := [\mathrm{P}_l(\mathcal{O})]^{2\times 2}$. Then, for each integer $k \geq 0$ and for each $T \in \mathcal{T}_h$, we define the local Raviart–Thomas space of order k (see, e.g., [41, 73]) as

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

where $\mathbf{x} := (x_1 x_2)^T$ is a generic vector of $\mathbb{R}^2$. Furthermore, in agreement with the previous notation, the space of matrix-valued functions whose rows belong to $\mathbf{RT}_k(T)$ will be denoted by $\mathbb{RT}_k(T)$. Also, we let

$$\mathbb{H}_h(\mathrm{D}_h) := \left\{ \boldsymbol{\tau} \in \mathbb{H}(\mathbf{div}\,; \mathrm{D}_h) : \ \boldsymbol{\tau}\big|_T \in \mathbb{RT}_k(T) \ \ \forall\, T \in \mathcal{T}_h \right\},$$

and

$$\mathbf{Q}_h(\mathrm{D}_h) := \left\{ \mathbf{v} \in \mathbf{L}^2(\mathrm{D}_h) : \ \mathbf{v}\big|_T \in \mathbf{P}_k(T) \ \ \forall\, T \in \mathcal{T}_h \right\}.$$

Notice that $\mathbb{H}(\mathrm{D}_h) = \mathbb{H}_{0,h}(\mathrm{D}_h) \oplus \mathbb{R}\mathbb{I}$, where $\mathbb{H}_{0,h}(\mathrm{D}_h) := \mathbb{H}_h(\mathrm{D}_h) \cap \mathbb{H}_0(\mathbf{div}; \Omega)$. In this way, we propose to approximate the solution of (2.14) by $(\boldsymbol{\sigma}_{0,h}, \mathbf{u}_h) \in \mathbb{H}_{0,h}(\mathrm{D}_h) \times \mathbf{Q}_h(\mathrm{D}_h)$, satisfying

$$\begin{aligned} (a_h + d_h)\,(\boldsymbol{\sigma}_{0,h}, \boldsymbol{\tau}_h) + b_h(\boldsymbol{\tau}_h, \mathbf{u}_h) &= G_h(\boldsymbol{\tau}_h) \quad \forall\, \boldsymbol{\tau}_h \in \mathbb{H}_{0,h}(\mathrm{D}_h), \\ b_h(\boldsymbol{\sigma}_{0,h}, \mathbf{v}_h) &= F_h(\mathbf{v}_h) \quad \forall\, \mathbf{v}_h \in \mathbf{Q}_h(\mathrm{D}_h), \end{aligned} \tag{2.16}$$

where a_h, b_h and F_h are given by (2.12),

$$G_h(\boldsymbol{\tau}_h) := \sum_{e\in\mathcal{E}_h^\partial}\int_e \overline{\mathbf{g}}\cdot(\boldsymbol{\tau}_h\mathbf{n}_e)(\mathbf{x})\,d\mathcal{S}_\mathbf{x}, \tag{2.17}$$

and

$$d_h(\boldsymbol{\xi}_h,\boldsymbol{\tau}_h) := \frac{1}{2\mu}\sum_{e\in\mathcal{E}_h^\partial}\int_e\left(\int_0^{\ell(\mathbf{x})}\mathbf{E}_h(\boldsymbol{\xi}_h^{\mathtt{d}})(\mathbf{x}+\eta\mathbf{m}(\mathbf{x}))\,\mathbf{m}(\mathbf{x})\,d\eta\right)\cdot(\boldsymbol{\tau}_h\mathbf{n}_e)(\mathbf{x})\,d\mathcal{S}_\mathbf{x} \tag{2.18}$$

for $\boldsymbol{\xi}_h,\boldsymbol{\tau}_h\in\mathbb{H}_h(\mathrm{D}_h)$, where we recall that $\overline{\mathbf{g}}(\mathbf{x}) := \mathbf{g}(\widetilde{\mathbf{x}}(\mathbf{x}))$. Above, $\mathbf{E}_h$ is the extrapolation operator given by

$$\mathbf{E}_h : \mathbb{P}_n(\mathcal{T}_h)\ni\boldsymbol{\tau}_h\longrightarrow\mathbf{E}_h(\boldsymbol{\tau}_h)(\mathbf{y}) := \begin{cases}\boldsymbol{\tau}_h(\mathbf{y}) & \forall\,\mathbf{y}\in T, \quad \forall\,T\in\mathcal{T}_h,\\ \boldsymbol{\tau}_h|_{T^e}(\mathbf{y}) & \forall\,\mathbf{y}\in\widetilde{T}^e_{ext}, \quad \forall\,e\in\mathcal{E}_h^\partial,\end{cases} \tag{2.19}$$

where for any integer $n\geq 0$, $\mathbb{P}_n(\mathcal{T}_h) := \prod_{T\in\mathcal{T}_h}\mathbb{P}_n(T)$. We observe that $\mathbf{E}_h(\boldsymbol{\sigma}_{0,h}^{\mathtt{d}})$ is well-defined since $\boldsymbol{\sigma}_{0,h}|_T\in\mathbb{RT}_k(T)\subseteq\mathbb{P}_{k+1}(T)$ for all $T\in\mathcal{T}_h$. We also observe that above we are implicitly using the following approximation of $\widetilde{\mathbf{g}}$ (cf. (2.15)):

$$\widetilde{\mathbf{g}}_h(\mathbf{x}) := \overline{\mathbf{g}}(\mathbf{x}) - \frac{1}{2\mu}\int_0^{\ell(\mathbf{x})}\mathbf{E}_h(\boldsymbol{\sigma}_{0,h}^{\mathtt{d}})(\mathbf{x}+\eta\mathbf{m}(\mathbf{x}))\,\mathbf{m}(\mathbf{x})\,d\eta, \tag{2.20}$$

for any edge $e\in\mathcal{E}_h^\partial$ and for each $\mathbf{x}\in e$. Note that if $\Omega=\mathrm{D}_h$ is a polygonal domain, then $\widetilde{\mathbf{g}}_h=\mathbf{g}$ and $d_h\equiv 0$. Hence, (2.16) would be reduced to the standard approach to approximate the saddle-point problem (2.6).

We end this section by recalling the approximation properties of the corresponding discrete spaces. To that end, we first introduce the $\mathbf{L}^2(\mathrm{D}_h)$-orthogonal projector onto $\mathbf{Q}_h(\mathrm{D}_h)$, $\boldsymbol{\mathcal{P}}_h^k:\mathbf{L}^2(\mathrm{D}_h)\to\mathbf{Q}_h(\mathrm{D}_h)$, which for each $\mathbf{v}\in\mathbf{H}^l(\mathrm{D}_h)$, with $0\leq l\leq k+1$, satisfies the approximation property

$$\|\mathbf{v}-\boldsymbol{\mathcal{P}}_h^k(\mathbf{v})\|_{0,T}\lesssim h_T^l|\mathbf{v}|_{l,T}\quad\forall\,T\in\mathcal{T}_h. \tag{2.21}$$

Furthermore, we recall the classical Raviart–Thomas interpolation operator $\boldsymbol{\Pi}_h^k:\mathbb{H}^1(\mathrm{D}_h)\to\mathbb{H}_h(\mathrm{D}_h)$, which, given $\boldsymbol{\tau}\in\mathbb{H}^1(\mathrm{D}_h)$, is characterized by the identities

$$\begin{aligned}\int_T\boldsymbol{\Pi}_h^k(\boldsymbol{\tau}):\boldsymbol{\xi}_h &= \int_T\boldsymbol{\tau}:\boldsymbol{\xi}_h \qquad \forall\,\boldsymbol{\xi}_h\in\mathbb{P}_{k-1}(T),\ \forall\,T\in\mathcal{T}_h,\quad \text{when } k\geq 1,\\ \int_e\left(\boldsymbol{\Pi}_h^k(\boldsymbol{\tau})\mathbf{n}_e\right)\cdot\boldsymbol{\psi}_h &= \int_e(\boldsymbol{\tau}\mathbf{n}_e)\cdot\boldsymbol{\psi}_h \quad \forall\,\boldsymbol{\psi}_h\in\mathbf{P}_k(e),\ \forall\,e\in\mathcal{E}_h,\quad \text{when } k\geq 0,\end{aligned}$$

whence it is easy to show that $\mathbf{div}\,(\boldsymbol{\Pi}_h^k(\boldsymbol{\tau})) = \boldsymbol{\mathcal{P}}_h^k(\mathbf{div}\,\boldsymbol{\tau})$ for all $\boldsymbol{\tau}\in\mathbb{H}^1(\mathrm{D}_h)$. Moreover, the local approximation properties of $\boldsymbol{\Pi}_h^k$ (see, e.g., [41, 121]) satisfy

- For each $\boldsymbol{\tau}\in\mathbb{H}^l(\mathrm{D}_h)$, with $1\leq l\leq k+1$, there holds

$$\|\boldsymbol{\tau}-\boldsymbol{\Pi}_h^k(\boldsymbol{\tau})\|_{0,T}\lesssim h_T^l|\boldsymbol{\tau}|_{l,T}\quad\forall\,T\in\mathcal{T}_h. \tag{2.22}$$

- For each $\boldsymbol{\tau}\in\mathbb{H}^1(\mathrm{D}_h)$ such that $\mathbf{div}\,\boldsymbol{\tau}\in\mathbf{H}^l(\mathrm{D}_h)$, with $0\leq l\leq k+1$,

$$\|\mathbf{div}\,(\boldsymbol{\tau}-\boldsymbol{\Pi}_h^k(\boldsymbol{\tau}))\|_{0,T}\lesssim h_T^l|\mathbf{div}\,\boldsymbol{\tau}|_{l,T}\quad\forall\,T\in\mathcal{T}_h. \tag{2.23}$$

- For each $\boldsymbol{\tau} \in \mathbb{H}^1(\mathrm{D}_h)$, there holds

$$\|(\boldsymbol{\tau} - \boldsymbol{\Pi}_h^k(\boldsymbol{\tau}))\mathbf{n}_e\|_{0,e} \lesssim h_e^{1/2}\|\boldsymbol{\tau}\|_{1,T^e} \quad \forall\, e \in \mathcal{E}_h. \tag{2.24}$$

We also recall that the interpolation operator $\boldsymbol{\Pi}_h^k$ can be defined as a bounded linear operator from the larger space $\mathbb{H}^s(\mathrm{D}_h) \cap \mathbb{H}(\mathbf{div}, \mathrm{D}_h)$ into $\mathbb{H}_h(\mathrm{D}_h)$ for all $s \in (0, 1]$ (see, e.g., [88, Theorem 3.1]) and in that case, the approximation property reduces to

$$\|\boldsymbol{\tau} - \boldsymbol{\Pi}_h^k(\boldsymbol{\tau})\|_{\mathbf{div},T} \lesssim h_T^s \left(\|\boldsymbol{\tau}\|_{s,T} + \|\mathbf{div}\,\boldsymbol{\tau}\|_{s,T}\right) \quad \forall\, T \in \mathcal{T}_h. \tag{2.25}$$

2.3.4 Well-posedness

We first introduce some hypotheses regarding the *closeness* between Γ and Γ_h. We remark that most of the notation and ideas here are closely connected to the ones of [57] and Section 1.2.4.

Let $e \in \mathcal{E}_h^\partial$. We define $\widetilde{r}_e := \widetilde{H}_e/h_e^\perp$, where $\widetilde{H}_e := \max_{\mathbf{x}\in e} \ell(\mathbf{x})$ and $h_e^\perp$ is the distance between the vertex, opposite to e, and the plane determined by e. In turn, for each $T \in \mathcal{T}_h$, we introduce $\mathbf{S}_k(\partial T) := \prod_{e\in\mathcal{E}(T)} \mathbf{P}_k(e)$, and for each edge $e \in \mathcal{E}_h^\partial$, we set

$$C_{eq}^e := h_{T^e}^{1/2} \sup_{\substack{\mathbf{v}_h \in \mathbf{S}_k(\partial T^e) \\ \mathbf{v}_h \neq \mathbf{0}}} \frac{\|\mathbf{v}_h\|_{0,\partial T^e}}{\|\mathbf{v}_h\|_{-1/2,\partial T^e}}, \tag{2.26}$$

$$\widetilde{C}_{ext}^e := (\widetilde{r}_e)^{-1/2} \sup_{\substack{\boldsymbol{\tau}_h \in \mathbb{P}_n(T^e) \\ \boldsymbol{\tau}_h \neq \mathbf{0}}} \frac{|||\mathbf{E}_h(\boldsymbol{\tau}_h)|||_e}{\|\boldsymbol{\tau}_h\|_{0,T^e}}, \tag{2.27}$$

where the mapping

$$\boldsymbol{\xi} \longmapsto |||\boldsymbol{\xi}|||_e := \left(\int_e \int_0^{\ell(\mathbf{x})} |\boldsymbol{\xi}(\mathbf{x} + \eta\mathbf{m}(\mathbf{x}))|^2 \, d\eta \, \mathcal{S}_\mathbf{x} \right)^{1/2} \tag{2.28}$$

defines a norm over the space $\mathbb{L}^2(\widetilde{T}_{ext}^e)$, which is equivalent to the standard $\mathbb{L}^2(\widetilde{T}_{ext}^e)$-norm (see Lemma 1.6) if we assume that

i) $\mathbf{m}^{\mathbf{p}_1} \cdot \mathbf{m}^{\mathbf{p}_2} \geq 0$,

ii) there exists a constant δ_e, independent of h, such that $\mathbf{m}(\theta) \cdot \mathbf{n}_e \geq \delta_e > 0$ for all $\theta \in [0, 1]$; and

iii) $\mathbf{m}^{\mathbf{p}_1} \cdot (\mathbf{m}^{\mathbf{p}_2})^\perp \geq 0$, with $(\mathbf{m}^{\mathbf{p}_2})^\perp$ being the vector obtained from $\mathbf{m}^{\mathbf{p}_2}$ through a counterclockwise rotation by $\pi/2$ about the origin.

We notice that both norms coincide when $\mathbf{m}^{\mathbf{p}_1}$ and $\mathbf{m}^{\mathbf{p}_2}$ are parallel to $\mathbf{n}_e$, and in such a case, conditions *i*)-*iii*) are no longer required. On the other hand, (2.26) is inspired by the equivalence between the norms $\|\cdot\|_{-1/2,\partial T^e}$ and $\|\cdot\|_{0,\partial T^e}$ (see, e.g., [61, Lemma 3.2]), whereas (2.27) was originally introduced by [57] with the $\mathbb{L}^2(\widetilde{T}_{ext}^e)$-norm, and later generalized to the norm $|||\cdot|||_e$ by Lemma 1.6. We also recall that both C_{eq}^e and $\widetilde{C}_{ext}^e$ are independent of the meshsize h, but depend on the shape-regularity constant and the polynomial degree. In turn, we denote $R := \max_{e\in\mathcal{E}_h^\partial} \widetilde{r}_e$ and assume

(A1) $R \leq C$, where $C > 0$ is independent of h; and

(A2) $\max_{e\in\mathcal{E}_h^\partial}\left\{\widetilde{r}_e\widetilde{C}_{ext}^e C_{eq}^e\right\} \leq C_1/4C_2$.

Above, C_1 and C_2 are positive constants, depending only on D_h, such that

$$C_1\|\boldsymbol{\tau}\|_{0,\mathrm{D}_h}^2 \leq \|\boldsymbol{\tau}^{\mathtt{d}}\|_{0,\mathrm{D}_h}^2 + \|\mathbf{div}\,\boldsymbol{\tau}\|_{0,\mathrm{D}_h}^2 \quad \forall\,\boldsymbol{\tau}\in\mathbb{H}_0(\mathbf{div};\mathrm{D}_h) \tag{2.29}$$

and

$$\|\boldsymbol{\tau}^{\mathtt{d}}\|_{0,\mathrm{D}_h} \leq C_2\|\boldsymbol{\tau}\|_{\mathbf{div},\mathrm{D}_h} \quad \forall\,\boldsymbol{\tau}\in\mathbb{H}(\mathbf{div},\mathrm{D}_h). \tag{2.30}$$

In particular, the proof of (2.29) can be found in [12, Lemma 3.1] (see also [41, Proposition 3.1]).

Let us briefly discuss the implications of these constraints. We are interested in two scenarios where we consider that our method can be used. This first one is the case where Γ_h is constructed by interpolating Γ by a picewise linear function, hence $d(\Gamma,\Gamma_h)\lesssim h^2$. In this case, Assumption **(A1)** holds since R is of order h and h is upper-bounded. Assumption **(A2)** is also satisfied for h small enough. The second scenario corresponds to the case where the domain is immersed in a background mesh and D_h is the union of all elements inside Ω. In this situation, $d(\Gamma,\Gamma_h)\lesssim h$ and R is of order one, which means **(A1)** is satisfied, but for a general domain Ω we cannot guaranty **(A2)** holds. However, we think this theoretical assumption could be relaxed because our numerical experiments show optimal rates of convergence even in the latter case.

We have then the following result.

Lemma 2.2. *Suppose that* **(A1)** *and* **(A2)** *hold. There exist positive constants $\widetilde{C}_d$ and $\widetilde{\alpha}$, independent of the meshsize h, such that*

$$|d_h(\boldsymbol{\xi}_h,\boldsymbol{\tau}_h)| \leq \widetilde{C}_d\|\boldsymbol{\xi}_h\|_{\mathbf{div},\mathrm{D}_h}\|\boldsymbol{\tau}_h\|_{\mathbf{div},\mathrm{D}_h} \quad \forall\,\boldsymbol{\xi}_h,\boldsymbol{\tau}_h\in\mathbb{H}_{0,h}(\mathrm{D}_h), \tag{2.31}$$

$$(a_h+d_h)(\boldsymbol{\tau}_h,\boldsymbol{\tau}_h) \geq \widetilde{\alpha}\|\boldsymbol{\tau}_h\|_{\mathbf{div},\mathrm{D}_h}^2 \quad \forall\boldsymbol{\tau}_h\in\mathbb{V}_h(\mathrm{D}_h), \tag{2.32}$$

where $\mathbb{V}_h(\mathrm{D}_h) := \{\boldsymbol{\tau}_h\in\mathbb{H}_{0,h}(\mathrm{D}_h):\ b_h(\boldsymbol{\tau}_h,\mathbf{v}_h)=0\ \ \forall\,\mathbf{v}_h\in\mathbf{Q}_h(\mathrm{D}_h)\}$.

Proof. We proceed analogously to Section 1.2.4. In fact, having in mind the estimation of d_h, let us first define for any $\boldsymbol{\xi}_h\in\mathbb{H}_{0,h}(\mathrm{D}_h)$ and any edge $e\in\mathcal{E}_h^\partial$,

$$\mathbf{w}_h(\mathbf{x}) := \int_0^{\ell(\mathbf{x})} \mathbf{E}_h(\boldsymbol{\xi}_h^{\mathtt{d}})(\mathbf{x}+\eta\mathbf{m}(\mathbf{x}))\,\mathbf{m}(\mathbf{x})\,d\eta \quad \forall\,\mathbf{x}\in e.$$

Integrating this vector-valued function over the edge e, applying the Cauchy–Schwarz inequality, using the constants $\widetilde{C}_{ext}^e$ (cf. (2.27)) to bound the norm $|||\cdot|||_e$, the fact that $h_e^\perp\leq h_{T^e}$, and the estimate (2.30), yields

$$\begin{aligned}\|\mathbf{w}_h\|_{0,e}^2 &\leq \int_e \ell(\mathbf{x})\int_0^{\ell(\mathbf{x})}|\mathbf{E}_h(\boldsymbol{\xi}_h^{\mathtt{d}})(\mathbf{x}+\eta\mathbf{m}(\mathbf{x}))|^2\,d\eta\,d\mathcal{S}_{\mathbf{x}} \leq \widetilde{r}_e\widetilde{H}_e\left(\widetilde{C}_{ext}^e\right)^2\|\boldsymbol{\xi}_h^{\mathtt{d}}\|_{0,T^e}^2\\ &\leq h_{T^e}\left(\widetilde{r}_e\widetilde{C}_{ext}^e\right)^2\|\boldsymbol{\xi}_h^{\mathtt{d}}\|_{0,T^e}^2 \leq h_{T^e}\left(\widetilde{r}_e\widetilde{C}_{ext}^e C_2\right)^2\|\boldsymbol{\xi}_h\|_{\mathbf{div},T^e}^2.\end{aligned} \tag{2.33}$$

Then, applying the Cauchy–Schwarz inequality together with (2.33), we have

$$|d_h(\boldsymbol{\xi}_h,\boldsymbol{\tau}_h)| \leq \frac{1}{2\mu}\sum_{e\in\mathcal{E}_h^\partial}\|\mathbf{w}_h\|_{0,e}\|\boldsymbol{\tau}_h\mathbf{n}_{T^e}\|_{0,\partial T^e} \leq \frac{C_2}{2\mu}\|\boldsymbol{\xi}_h\|_{\mathbf{div},\mathrm{D}_h}\sum_{e\in\mathcal{E}_h^\partial}\widetilde{r}_e\widetilde{C}_{ext}^e h_{T^e}^{1/2}\|\boldsymbol{\tau}_h\mathbf{n}_{T^e}\|_{0,\partial T^e}$$

for all $\boldsymbol{\xi}_h, \boldsymbol{\tau}_h \in \mathbb{H}_{0,h}(\mathrm{D}_h)$. From this, by the definition of C^e_{eq} (cf. (2.26)) and estimate (2.1), and after some algebraic manipulations, we have from Assumption **(A1)** that (2.31) holds. On the other hand, to obtain the coercivity of $(a_h + d_h)$ on $\mathbb{V}_h(\mathrm{D}_h)$, we note that $\boldsymbol{\tau}_h \in \mathbb{V}_h(\mathrm{D}_h)$ implies $\mathbf{div}\,\boldsymbol{\tau}_h \equiv \mathbf{0}$ in D_h, since $\mathbf{div}\,\mathbb{H}_{0,h}(\mathrm{D}_h) \subseteq \mathbf{Q}_h(\mathrm{D}_h)$. Consequently, from the inequality (2.29), the boundedness of d_h and Assumption **(A2)**, it follows that

$$(a_h + d_h)(\boldsymbol{\tau}_h, \boldsymbol{\tau}_h) \geq \frac{1}{2\mu}\|\boldsymbol{\tau}_h^{\mathtt{d}}\|^2_{0,\Omega} - \frac{C_1}{8\mu}\|\boldsymbol{\tau}_h\|^2_{\mathbf{div};\mathrm{D}_h} \geq \frac{3C_1}{8\mu}\|\boldsymbol{\tau}_h\|^2_{\mathbf{div},\mathrm{D}_h}$$

for all $\boldsymbol{\tau}_h \in \mathbb{V}_h(\mathrm{D}_h)$, showing that (2.32) is satisfied with $\widetilde{\alpha} = 3C_1/8\mu$ and concluding the proof. □

Remark 2.1. *The boundedness of the functional G_h in (2.17) can be deduced from the continuity of the mapping $\widetilde{\mathbf{x}} : \Gamma_h \to \Gamma$ (cf. Section 2.3.2). In fact, since $\overline{\mathbf{g}}(\cdot) = \mathbf{g}(\widetilde{\mathbf{x}}(\cdot))$ belongs to $\mathbf{H}^{1/2}(\Gamma_h)$, we apply the continuity of the normal trace operator and estimate (2.1), to obtain $|G_h(\boldsymbol{\tau}_h)| \leq \|\overline{\mathbf{g}}\|_{1/2,\Gamma_h}\|\boldsymbol{\tau}_h\|_{\mathbf{div},\mathrm{D}_h}$ for all $\boldsymbol{\tau}_h \in \mathbb{H}_{0,h}(\mathrm{D}_h)$, as required.*

Furthermore, we recall that the pair $(\mathbb{H}_{h,0}(\mathrm{D}_h), \mathbf{Q}_h(\mathrm{D}_h))$ satisfies the following discrete inf-sup condition (see, for instance [78, Lemma 3.2]):

$$\inf_{\substack{\boldsymbol{\tau}_h \in \mathbb{H}_{0,h}(\mathrm{D}_h) \\ \boldsymbol{\tau}_h \neq \mathbf{0}}} \frac{b_h(\boldsymbol{\tau}_h, \mathbf{v}_h)}{\|\boldsymbol{\tau}_h\|_{\mathbf{div},\Omega}} \geq \hat{\beta}\|\mathbf{v}_h\|_{0,\Omega} \quad \forall\, \mathbf{v}_h \in \mathrm{Q}_h(\mathrm{D}_h), \tag{2.34}$$

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

We are now ready to state the main result concerning the well-posedness of (2.16).

Theorem 2.3. *Suppose that* **(A1)** *and* **(A2)** *hold. Given $\mathbf{f} \in \mathbf{L}^2(\Omega)$ and $\mathbf{g} \in \mathbf{H}^{1/2}(\Gamma)$, there exists a unique $(\boldsymbol{\sigma}_{0,h}, \mathbf{u}) \in \mathbb{H}_{h,0}(\mathrm{D}_h) \times \mathbf{Q}_h(\mathrm{D}_h)$ solution to the problem (2.16), which satisfies*

$$\|(\boldsymbol{\sigma}_{0,h}, \mathbf{u}_h)\|_{\mathbb{H}(\mathbf{div};\mathrm{D}_h)\times\mathbf{L}^2(\mathrm{D}_h)} \lesssim \|F_h\|_{[\mathbf{Q}_h(\mathrm{D}_h)]'} + \|G_h\|_{\left[\mathbb{H}_{0,h}(\mathrm{D}_h)\right]'}.$$

Proof. The proof follows from the discrete version of the Babuška–Brezzi theorem (see, e.g., [73, Section 2.5]). □

We end this section by providing a postprocessing technique for approximating the pseudostress $\boldsymbol{\sigma}$ and the pressure p in the computational domain D_h. For this, we let $(\boldsymbol{\sigma}_{0,h}, \mathbf{u}_h) \in \mathbb{H}_{h,0}(\mathrm{D}_h) \times \mathbf{Q}_h(\mathrm{D}_h)$ be the unique solution of (2.16) and based on the definition (2.13), we propose the following approximations of $\boldsymbol{\sigma}$ and p:

$$\boldsymbol{\sigma}_h := \boldsymbol{\sigma}_{0,h} + \left(\frac{\gamma_h}{2|\mathrm{D}_h|}\right)\mathbb{I}, \tag{2.35}$$

and

$$p_h := -\frac{1}{2}\mathrm{tr}\,(\boldsymbol{\sigma}_h), \tag{2.36}$$

where

$$\gamma_h := -\int_{\mathrm{D}_h^c} \mathrm{tr}\left(\mathbf{E}_h(\boldsymbol{\sigma}_{0,h}) - \left(\frac{1}{2|\Omega|}\int_{\mathrm{D}_h^c}\mathrm{tr}\,(\mathbf{E}_h(\boldsymbol{\sigma}_{0,h}))\right)\mathbb{I}\right), \tag{2.37}$$

and $\mathbf{E}_h(\boldsymbol{\sigma}_{0,h})$ denotes the extrapolation of $\boldsymbol{\sigma}_{0,h}$ (cf. (2.19)). Notice that the following identity holds:

$$\int_{\mathrm{D}_h}\mathrm{tr}\,(\boldsymbol{\sigma}_h) = \gamma_h. \tag{2.38}$$

2.4 *A priori* error bounds

Given $(\boldsymbol{\sigma},\mathbf{u})\in\mathbb{H}_0(\mathbf{div};\Omega)\times\mathbf{L}^2(\Omega)$ and $(\boldsymbol{\sigma}_{0,h},\mathbf{u}_h)\in\mathbb{H}_{0,h}(\mathrm{D}_h)\times\mathbf{Q}_h(\mathrm{D}_h)$ solutions of (2.6) and (2.16), respectively, we are now interested in obtaining upper bounds for

$$\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathbf{div},\mathrm{D}_h},\quad \|\mathbf{u}-\mathbf{u}_h\|_{0,\mathrm{D}_h}\quad\text{and}\quad \|p-p_h\|_{0,\mathrm{D}_h},$$

where $\boldsymbol{\sigma}_h$ and p_h are given by (2.35) and (2.36), respectively. These errors, as we shall see below, depend on a Céa-type estimate for $\|\boldsymbol{\sigma}_0-\boldsymbol{\sigma}_{0,h}\|_{\mathbf{div},\mathrm{D}_h}$, with $\boldsymbol{\sigma}_0$ defined as in (2.13). For this reason, we follow the strategy of Section 1.3: we first derive the corresponding Céa estimate, then apply it to derive error bounds for the main variables, even on the complement D_h^c, and finally infer the theoretical rate of convergence. Most of the arguments that we will employ to obtain the Céa estimate consist of an application of the same components of the method in Chapter 1. However, as we will see, in our case we need to take into account the influence of the term $(\gamma-\gamma_h)$ arising from the fact that we are looking for the first component of the solution in the space $\mathbb{H}_0(\mathbf{div};\mathrm{D}_h)$.

2.4.1 Estimates on D_h

We begin with a Céa-type estimate for our Galerkin scheme (2.16). For its proof we proceed similarly to the proof of Theorem 1.5.

Theorem 2.4. *Let $(\boldsymbol{\sigma},\mathbf{u})\in\mathbb{H}_0(\mathbf{div};\Omega)\times\mathbf{L}^2(\Omega)$ and $(\boldsymbol{\sigma}_{0,h},\mathbf{u}_h)$ be the unique solutions of (2.6) and (2.16), respectively. Let $\boldsymbol{\sigma}_0$ be defined as in (2.13) and suppose that hypotheses of Theorem 2.3 are satisfied. Then, there holds*

$$\begin{aligned}&\|(\boldsymbol{\sigma}_0-\boldsymbol{\sigma}_{0,h},\mathbf{u}-\mathbf{u}_h)\|_{\mathbb{H}(\mathbf{div};\mathrm{D}_h)\times\mathbf{L}^2(\mathrm{D}_h)}\\ &\lesssim \inf_{\mathbf{v}_h\in\mathbf{Q}_h(\mathrm{D}_h)}\|\mathbf{u}-\mathbf{v}_h\|_{0,\mathrm{D}_h}+\inf_{\boldsymbol{\xi}_{0,h}\in\mathbb{H}_{0,h}(\mathrm{D}_h)}\left(\|\boldsymbol{\sigma}_0-\boldsymbol{\xi}_{0,h}\|_{\mathbf{div},\mathrm{D}_h}+\sum_{e\in\mathcal{E}_h^\partial}\left\|\boldsymbol{\sigma}^{\mathtt{d}}-\mathbf{E}_h(\boldsymbol{\xi}_{0,h}^{\mathtt{d}})\right\|_e\right).\end{aligned}$$

Proof. Recalling that $(\boldsymbol{\sigma}_0,\mathbf{u})$ solves (2.14), and rearranging conveniently (2.16), it follows that

$$\begin{aligned}a_h(\boldsymbol{\sigma}_0,\boldsymbol{\tau})+b_h(\boldsymbol{\tau},\mathbf{u})&=\langle\boldsymbol{\tau}\mathbf{n}_{\Gamma_h},\widetilde{\mathbf{g}}\rangle_{\Gamma_h}\quad\forall\,\boldsymbol{\tau}\in\mathbb{H}_0(\mathbf{div};\mathrm{D}_h),\\ b_h(\boldsymbol{\sigma}_0,\mathbf{v})&=F_h(\mathbf{v})\qquad\forall\,\mathbf{v}\in\mathbf{L}^2(\mathrm{D}_h),\end{aligned}$$

and

$$\begin{aligned}a_h(\boldsymbol{\sigma}_{0,h},\boldsymbol{\tau}_h)+b_h(\boldsymbol{\tau}_h,\mathbf{u}_h)&=G_h(\boldsymbol{\tau}_h)-d_h(\boldsymbol{\sigma}_{0,h},\boldsymbol{\tau}_h)\quad\forall\,\boldsymbol{\tau}_h\in\mathbb{H}_{0,h}(\mathrm{D}_h),\\ b_h(\boldsymbol{\sigma}_{0,h},\mathbf{v}_h)&=F_h(\mathbf{v}_h)\qquad\forall\,\mathbf{v}_h\in\mathbf{Q}_h(\mathrm{D}_h).\end{aligned}$$

It should be noted that the structure of these problems differs only in the functionals concerning the Dirichlet boundary condition. This leads us to apply the well-known Strang-type estimate to obtain our preliminary error bounds as done in Section 1.3.1 (see also [63, Lemma 5.2] or [82, Section 4.1]):

$$\begin{aligned}\|\boldsymbol{\sigma}_0-\boldsymbol{\sigma}_{0,h}\|_{\mathbf{div},\mathrm{D}_h}\leq\left(1+\frac{\|a_h\|}{\hat{\alpha}}\right)\left(1+\frac{\|b_h\|}{\hat{\beta}}\right)\inf_{\boldsymbol{\xi}_{0,h}\in\mathbb{H}_{0,h}(\mathrm{D}_h)}\|\boldsymbol{\sigma}_0-\boldsymbol{\xi}_{0,h}\|_{\mathbf{div},\mathrm{D}_h}\\ +\frac{\|b_h\|}{\hat{\alpha}}\inf_{\mathbf{w}_h\in\mathbf{Q}_h(\mathrm{D}_h)}\|\mathbf{u}-\mathbf{w}_h\|_{0,\mathrm{D}_h}+\frac{1}{\hat{\alpha}}\mathbb{T}^{\boldsymbol{\sigma}},\end{aligned}\tag{2.39}$$

and

$$\begin{aligned}\|\mathbf{u}-\mathbf{u}_h\|_{0,\mathrm{D}_h} \leq \frac{\|a_h\|}{\hat{\beta}}\left(1+\frac{\|a_h\|}{\hat{\alpha}}\right)\left(1+\frac{\|b_h\|}{\hat{\beta}}\right)\inf_{\boldsymbol{\xi}_{0,h}\in\mathbb{H}_{0,h}(\mathrm{D}_h)}\|\boldsymbol{\sigma}_0-\boldsymbol{\xi}_{0,h}\|_{\mathbf{div},\mathrm{D}_h}\\ +\left(1+\frac{\|b_h\|}{\hat{\beta}}+\frac{\|b_h\|\,\|a_h\|}{\hat{\beta}\hat{\alpha}}\right)\inf_{\mathbf{w}_h\in\mathbf{Q}_h(\mathrm{D}_h)}\|\mathbf{u}-\mathbf{w}_h\|_{0,\mathrm{D}_h}+\frac{1}{\hat{\beta}}\left(1+\frac{\|a_h\|}{\hat{\alpha}}\right)\mathbb{T}^{\boldsymbol{\sigma}},\end{aligned} \tag{2.40}$$

where $\hat{\alpha}$ is the coercivity constant of the bilinear form a_h (actually, $\hat{\alpha}=C_1/2\mu$), $\hat{\beta}$ is the positive constant satisfying (2.34), $\|\cdot\|$ denotes the norm of the corresponding bilinear forms, and $\mathbb{T}^{\boldsymbol{\sigma}}$ is the error of the boundary condition on Γ_h given by

$$\mathbb{T}^{\boldsymbol{\sigma}}:=\sup_{\substack{\boldsymbol{\tau}_h\in\mathbb{H}_{0,h}(\mathrm{D}_h)\\ \boldsymbol{\tau}_h\neq\mathbf{0}}}\frac{|\langle\boldsymbol{\tau}_h\mathbf{n}_{\Gamma_h},\widetilde{\mathbf{g}}\rangle_{\Gamma_h}-(G_h(\boldsymbol{\tau}_h)-d_h(\boldsymbol{\sigma}_{0,h},\boldsymbol{\tau}_h))|}{\|\boldsymbol{\tau}_h\|_{\mathbf{div},\mathrm{D}_h}}=\sup_{\substack{\boldsymbol{\tau}_h\in\mathbb{H}_{0,h}(\mathrm{D}_h)\\ \boldsymbol{\tau}_h\neq\mathbf{0}}}\frac{|\langle\boldsymbol{\tau}_h\mathbf{n}_{\Gamma_h},\widetilde{\mathbf{g}}-\widetilde{\mathbf{g}}_h\rangle_{\Gamma_h}|}{\|\boldsymbol{\tau}_h\|_{\mathbf{div},\mathrm{D}_h}}.$$

It remains therefore to bound $\mathbb{T}^{\boldsymbol{\sigma}}$ from above. To this end, using the Cauchy–Schwarz inequality, the constant C^e_{eq} of (2.26), the definition of $\widetilde{r}_e$, and the norm given in (2.28), it follows that

$$\mathbb{T}^{\boldsymbol{\sigma}}\leq\frac{1}{2\mu}\sum_{e\in\mathcal{E}_h^\partial}C^e_{eq}h_{T^e}^{-1/2}\|\widetilde{\mathbf{g}}-\widetilde{\mathbf{g}}_h\|_{0,e}\leq\frac{1}{2\mu}\sum_{e\in\mathcal{E}_h^\partial}C^e_{eq}(\widetilde{r}_e)^{1/2}\left|\!\left|\!\left|\boldsymbol{\sigma}^{\mathtt{d}}-\mathbf{E}_h\left(\boldsymbol{\sigma}_h^{\mathtt{d}}\right)\right|\!\right|\!\right|_e, \tag{2.41}$$

where we recall that $\boldsymbol{\sigma}_h$ has been defined in (2.35) and $\boldsymbol{\sigma}_h^{\mathtt{d}}=\boldsymbol{\sigma}_{0,h}^{\mathtt{d}}$.

Now, we will establish an upper bound for $\left|\!\left|\!\left|\boldsymbol{\sigma}^{\mathtt{d}}-\mathbf{E}_h(\boldsymbol{\sigma}_h^{\mathtt{d}})\right|\!\right|\!\right|_e$. Inspired by (2.35), let $\boldsymbol{\xi}_{0,h}\in\mathbb{H}_{0,h}(\mathrm{D}_h)$ and

$$\boldsymbol{\xi}_h:=\boldsymbol{\xi}_{0,h}+\left(\frac{c_h}{2|\mathrm{D}_h|}\right)\mathbb{I}, \tag{2.42}$$

with constant c_h being defined as γ_h in (2.37) by replacing $\boldsymbol{\sigma}_{0,h}$ by $\boldsymbol{\xi}_{0,h}$. Then, adding and subtracting $\mathbf{E}_h(\boldsymbol{\xi}_h^{\mathtt{d}})$ in (2.41), using the constants $\widetilde{C}^e_{ext}$ and C_2 (cf. (2.27) and (2.30), respectively), and also employing Assumption **(A2)**, we have

$$\mathbb{T}^{\boldsymbol{\sigma}}\leq\frac{1}{2\mu}\sum_{e\in\mathcal{E}_h^\partial}C^e_{eq}(\widetilde{r}_e)^{1/2}\left|\!\left|\!\left|\boldsymbol{\sigma}^{\mathtt{d}}-\mathbf{E}_h(\boldsymbol{\xi}_h^{\mathtt{d}})\right|\!\right|\!\right|_e+\frac{C_1}{8\mu}\|\boldsymbol{\sigma}_h-\boldsymbol{\xi}_h\|_{\mathbf{div},\mathrm{D}_h},$$

from which, adding and subtracting $\boldsymbol{\sigma}$ and considering the identity $\boldsymbol{\xi}_h^{\mathtt{d}}=\boldsymbol{\xi}_{0,h}^{\mathtt{d}}$, it holds

$$\mathbb{T}^{\boldsymbol{\sigma}}\leq\frac{1}{2\mu}\sum_{e\in\mathcal{E}_h^\partial}C^e_{eq}(\widetilde{r}_e)^{1/2}\left|\!\left|\!\left|\boldsymbol{\sigma}^{\mathtt{d}}-\mathbf{E}_h(\boldsymbol{\xi}_{0,h}^{\mathtt{d}})\right|\!\right|\!\right|_e+\frac{C_1}{8\mu}\left(\|\boldsymbol{\sigma}-\boldsymbol{\xi}_h\|_{\mathbf{div},\mathrm{D}_h}+\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathbf{div},\mathrm{D}_h}\right). \tag{2.43}$$

Furthermore, according to definition (2.13), we know that $\boldsymbol{\sigma}|_{\mathrm{D}_h}=\boldsymbol{\sigma}_0+\left(\frac{\gamma}{2|\mathrm{D}_h|}\right)\mathbb{I}$ and $\int_{\mathrm{D}_h}\mathrm{tr}\,(\boldsymbol{\sigma})=\gamma$. Thus, concerning the last term in (2.43), we use (2.38) to infer

$$\begin{aligned}\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathbf{div},\mathrm{D}_h}&\leq\|\boldsymbol{\sigma}_0-\boldsymbol{\sigma}_{0,h}\|_{\mathbf{div},\mathrm{D}_h}+\left\|\left(\frac{\gamma-\gamma_h}{2|\mathrm{D}_h|}\right)\mathbb{I}\right\|_{0,\mathrm{D}_h}\\ &=\|\boldsymbol{\sigma}_0-\boldsymbol{\sigma}_{0,h}\|_{\mathbf{div},\mathrm{D}_h}+\left\|\frac{1}{2|\mathrm{D}_h|}\left(\int_{\mathrm{D}_h}\mathrm{tr}\,(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\right)\mathbb{I}\right\|_{0,\mathrm{D}_h}\\ &\leq\|\boldsymbol{\sigma}_0-\boldsymbol{\sigma}_{0,h}\|_{\mathbf{div},\mathrm{D}_h}+\frac{1}{\sqrt{2}}\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathbf{div},\mathrm{D}_h},\end{aligned}$$

and hence,

$$\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathbf{div},\mathrm{D}_h} \leq \left(\frac{2}{2-\sqrt{2}}\right)\|\boldsymbol{\sigma}_0-\boldsymbol{\sigma}_{0,h}\|_{\mathbf{div},\mathrm{D}_h}. \tag{2.44}$$

Similarly, we have

$$\|\boldsymbol{\sigma}-\boldsymbol{\xi}_h\|_{\mathbf{div},\mathrm{D}_h} \leq \left(\frac{2}{2-\sqrt{2}}\right)\|\boldsymbol{\sigma}_0-\boldsymbol{\xi}_{0,h}\|_{\mathbf{div},\mathrm{D}_h}. \tag{2.45}$$

Therefore, from (2.39), (2.43), (2.44) and (2.45), we deduce, after simple algebraic manipulations and recalling that $\hat{\alpha}=C_1/2\mu$, that

$$\begin{aligned}\left(\frac{3-2\sqrt{2}}{4-2\sqrt{2}}\right)&\|\boldsymbol{\sigma}_0-\boldsymbol{\sigma}_{0,h}\|_{\mathbf{div},\mathrm{D}_h}\\ &\lesssim \inf_{\mathbf{w}_h\in\mathbf{Q}_h(\mathrm{D}_h)}\|\mathbf{u}-\mathbf{w}_h\|_{0,\mathrm{D}_h}+\|\boldsymbol{\sigma}_0-\boldsymbol{\xi}_{0,h}\|_{\mathbf{div},\mathrm{D}_h}+\sum_{e\in\mathcal{E}_h^\partial}\left\|\left|\boldsymbol{\sigma}^{\mathtt{d}}-\mathbf{E}_h\left(\boldsymbol{\xi}_{0,h}^{\mathtt{d}}\right)\right|\right\|_e.\end{aligned} \tag{2.46}$$

Finally, dividing (2.46) by $\left(\frac{3-2\sqrt{2}}{4-2\sqrt{2}}\right)>0$, placing the resulting inequality together with (2.40), one easily arrives at the claimed result. □

Corollary 2.5. *Suppose that hypotheses of Theorem 2.4 hold. Then,*

$$\begin{aligned}&\|p-p_h\|_{0,\mathrm{D}_h}+\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathbf{div},\mathrm{D}_h}\\ &\lesssim \inf_{\mathbf{v}_h\in\mathbf{Q}_h(\mathrm{D}_h)}\|\mathbf{u}-\mathbf{v}_h\|_{0,\mathrm{D}_h}+\inf_{\boldsymbol{\xi}_{0,h}\in\mathbb{H}_{0,h}(\mathrm{D}_h)}\left(\left\|\boldsymbol{\sigma}_0-\boldsymbol{\xi}_{0,h}\right\|_{\mathbf{div},\mathrm{D}_h}+\sum_{e\in\mathcal{E}_h^\partial}\left\|\left|\boldsymbol{\sigma}^{\mathtt{d}}-\mathbf{E}_h\left(\boldsymbol{\xi}_{0,h}^{\mathtt{d}}\right)\right|\right\|_e\right).\end{aligned}$$

Proof. A direct application of definitions (2.4) and (2.36), and estimate (2.44), implies

$$\|p-p_h\|_{0,\mathrm{D}_h}+\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathbf{div},\mathrm{D}_h} \leq \left(\frac{3}{2-\sqrt{2}}\right)\|\boldsymbol{\sigma}_0-\boldsymbol{\sigma}_{0,h}\|_{\mathbf{div},\mathrm{D}_h}.$$

The rest of the proof follows from Theorem 2.4. □

2.4.2 Approximation in D_h^c and rate of convergence

We now turn to provide approximations of the pseudostress $\boldsymbol{\sigma}$, the velocity $\mathbf{u}$ and the pressure p outside D_h. To alleviate the notation, these approximations will be also denoted by $\boldsymbol{\sigma}_h$, $\mathbf{u}_h$ and p_h, respectively.

In order to approximate $\boldsymbol{\sigma}$ in D_h^c, we follow the idea in [57, Section 2.1.3]. To that end, given $(\boldsymbol{\sigma}_{0,h},\mathbf{u}_h)\in\mathbb{H}_{0,h}(\mathrm{D}_h)\times\mathbf{Q}_h(\mathrm{D}_h)$ the unique solution of (2.16), let $\boldsymbol{\sigma}_h$ be the tensor defined in (2.35). Then, for each $e\in\mathcal{E}_h^\partial$ and any $\mathbf{y}\in\widetilde{T}_{ext}^e$, we set

$$\boldsymbol{\sigma}_h(\mathbf{y}) := \mathbf{E}_h(\boldsymbol{\sigma}_h)(\mathbf{y}). \tag{2.47}$$

Remark 2.2. *From (2.37), we have that* $\int_{\mathrm{D}_h^c}\mathrm{tr}\left(\mathbf{E}_h(\boldsymbol{\sigma}_{0,h})\right)=-\frac{|\Omega|}{|\mathrm{D}_h|}\gamma_h$, *thus*

$$\int_{\mathrm{D}_h^c}\mathrm{tr}\left(\boldsymbol{\sigma}_h\right)=\int_{\mathrm{D}_h^c}\mathrm{tr}\,\left(\mathbf{E}_h(\boldsymbol{\sigma}_{0,h})\right)+\left(\frac{\gamma_h}{|\mathrm{D}_h|}\right)|\mathrm{D}_h^c|=-\gamma_h,$$

and by (2.38) *we conclude that* $\int_\Omega \mathrm{tr}\,(\boldsymbol{\sigma}_h) = 0$. *In addition, we can write*

$$\boldsymbol{\sigma}_h = \mathbf{E}_h(\boldsymbol{\sigma}_{0,h}) - \left(\frac{1}{2|\Omega|}\int_{\mathrm{D}_h^c} \mathrm{tr}\,\left(\mathbf{E}_h(\boldsymbol{\sigma}_{0,h})\right)\right)\mathbb{I} \quad \textit{in} \quad \Omega. \tag{2.48}$$

When Assumption **(A1)** and definition (2.47) (or equivalently, (2.48)) are considered, it is important to point out that since, in D_h^c, the normal component of the extrapolated $\boldsymbol{\sigma}_h$ is, in general, discontinuous across the transferring paths $\{\mathscr{C}(\mathbf{x})\}_{\mathbf{x}\in\Gamma_h}$ (cf. Section 2.3.2), the method ensures that, at least, $\boldsymbol{\sigma}_h$ belongs to the broken Sobolev space (see, e.g., [67, Section 1.2.6])

$$\mathbb{H}(\mathbf{div};\widetilde{\mathcal{T}}_h) := \left\{\boldsymbol{\tau}\in\mathbb{L}^2(\mathrm{D}_h^c):\ \boldsymbol{\tau}|_{\widetilde{T}^e_{ext}} \in \mathbb{H}(\mathbf{div};\widetilde{T}^e_{ext})\ \ \forall\, e\in\mathcal{E}_h^\partial\right\}$$

endowed with the broken norm $\|\cdot\|_{\mathbf{div},\widetilde{\mathcal{T}}_h} := \left(\sum_{e\in\mathcal{E}_h^\partial}\|\cdot\|^2_{\mathbf{div},\widetilde{T}^e_{ext}}\right)^{1/2}$, where $\widetilde{\mathcal{T}}_h$ is the mesh defined in Section 2.3.2.

On the other hand, by defining

$$p_h := -\frac{1}{2}\mathrm{tr}\,(\boldsymbol{\sigma}_h) \quad \text{in} \quad \mathrm{D}_h^c, \tag{2.49}$$

it is clear from Remark 2.2 that $\int_\Omega p_h = 0$. Moreover, from definitions (2.4) and (2.49), we have

$$\|p - p_h\|_{0,\mathrm{D}_h^c} \le \frac{1}{2}\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{0,\mathrm{D}_h^c}. \tag{2.50}$$

The latter suggests to establish firstly the error estimate associated to the pseudostress.

Let us start by introducing the following intermediate result.

Lemma 2.6. *Let* $(\boldsymbol{\sigma},\mathbf{u}) \in \mathbb{H}_0(\mathbf{div};\Omega)\times\mathbf{L}^2(\Omega)$ *be the unique solution of* (2.6) *and assume that hypotheses of Theorem 2.4 hold. Suppose further that there exists an integer* $l \ge 0$ *such that* $\boldsymbol{\sigma}\in\mathbb{H}^{l+1}(\Omega)$, *with* $\mathbf{div}\,\boldsymbol{\sigma}\in\mathbf{H}^{l+1}(\Omega)$. *Then, for any* $\boldsymbol{\xi}_h\in\mathbb{H}_h(\mathrm{D}_h)$, *we have*

$$\sum_{e\in\mathcal{E}_h^\partial}\|\boldsymbol{\sigma} - \mathbf{E}_h(\boldsymbol{\xi}_h)\|_{0,\widetilde{T}^e_{ext}} \lesssim \|\boldsymbol{\sigma}-\boldsymbol{\xi}_h\|_{0,\mathrm{D}_h} + h^{l+1}\|\boldsymbol{\sigma}\|_{l+1,\Omega}, \tag{2.51}$$

and

$$\sum_{e\in\mathcal{E}_h^\partial}\|\boldsymbol{\sigma} - \mathbf{E}_h(\boldsymbol{\xi}_h)\|_{\mathbf{div},\widetilde{T}^e_{ext}} \lesssim \|\boldsymbol{\sigma}-\boldsymbol{\xi}_h\|_{\mathbf{div},\mathrm{D}_h} + h^{l+1}\left(\|\boldsymbol{\sigma}\|_{l+1,\Omega} + \|\mathbf{div}\,\boldsymbol{\sigma}\|_{l+1,\Omega}\right). \tag{2.52}$$

Proof. The proof makes use of averaged Taylor polynomials (cf. [37, Chapter 4]) in the neighborhood of the curved boundary Γ, and their well-known approximation properties. For details of the proof we refer to Lemma 1.7. □

The following lemma allows us to deduce upper bounds for $(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)$ in the $\mathbf{L}^2$-norm, as well as in the broken $\mathbb{H}(\mathbf{div})$-norm on D_h^c. The general idea of the proof is inspired by Lemma 1.8.

Lemma 2.7. *Assume the same hypotheses of Theorem 2.4. Let* $(\boldsymbol{\sigma},\mathbf{u})\in\mathbb{H}_0(\mathbf{div};\Omega)\times\mathbf{L}^2(\Omega)$ *and* $(\boldsymbol{\sigma}_{0,h},\mathbf{u}_h)\in\mathbb{H}_{0,h}(\mathrm{D}_h)\times\mathbf{Q}_h(\mathrm{D}_h)$ *be the unique solutions of* (2.6) *and* (2.16), *respectively. Let* $\boldsymbol{\sigma}_h$ *be*

defined as in (2.47). *Suppose further that there exists an integer* $l \geq 0$ *such that* $\boldsymbol{\sigma} \in \mathbb{H}^{l+1}(\Omega)$, *with* $\mathbf{div}\,\boldsymbol{\sigma} \in \mathbf{H}^{l+1}(\Omega)$. *Then, we have*

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{0,\mathrm{D}_h^c} \lesssim \inf_{\mathbf{w}_h \in \mathbf{Q}_h(\mathrm{D}_h)} \|\mathbf{u} - \mathbf{w}_h\|_{0,\mathrm{D}_h} + \inf_{\boldsymbol{\xi}_{0,h} \in \mathbb{H}_{0,h}(\mathrm{D}_h)} \|\boldsymbol{\sigma}_0 - \boldsymbol{\xi}_{0,h}\|_{0,\mathrm{D}_h} + h^{l+1}\|\boldsymbol{\sigma}\|_{l+1,\Omega}, \tag{2.53}$$

and

$$\begin{aligned}&\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathbf{div},\widetilde{\mathcal{T}}_h}\\ &\quad\lesssim \inf_{\mathbf{w}_h \in \mathbf{Q}_h(\mathrm{D}_h)} \|\mathbf{u} - \mathbf{w}_h\|_{0,\mathrm{D}_h} + \inf_{\boldsymbol{\xi}_{0,h} \in \mathbb{H}_{0,h}(\mathrm{D}_h)} \|\boldsymbol{\sigma}_0 - \boldsymbol{\xi}_{0,h}\|_{\mathbf{div},\mathrm{D}_h} + h^{l+1}\left(\|\boldsymbol{\sigma}\|_{l+1,\Omega} + \|\mathbf{div}\,\boldsymbol{\sigma}\|_{l+1,\Omega}\right).\end{aligned} \tag{2.54}$$

Proof. Let $\boldsymbol{\xi}_h$ be given by (2.42). Adding and subtracting convenient terms, applying the estimate (2.51), using the definition of $\widetilde{C}^e_{ext}$ in (2.27), and making use of Assumption **(A1)**, we obtain

$$\begin{aligned}\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{0,\mathrm{D}_h^c} &\leq \sum_{e \in \mathcal{E}_h^\partial} \left(\|\boldsymbol{\sigma} - \mathbf{E}_h(\boldsymbol{\xi}_h)\|_{0,\widetilde{T}^e_{ext}} + \|\mathbf{E}_h(\boldsymbol{\xi}_h) - \boldsymbol{\sigma}_h\|_{0,\widetilde{T}^e_{ext}}\right)\\ &\lesssim h^{l+1}\|\boldsymbol{\sigma}\|_{l+1,\Omega} + \|\boldsymbol{\sigma} - \boldsymbol{\xi}_h\|_{0,\mathrm{D}_h} + \sum_{e \in \mathcal{E}_h^\partial} \widetilde{C}^e_{ext}(\widetilde{r}_e)^{1/2}\|\boldsymbol{\sigma}_h - \boldsymbol{\xi}_h\|_{0,T^e}\\ &\lesssim h^{l+1}\|\boldsymbol{\sigma}\|_{l+1,\Omega} + \|\boldsymbol{\sigma} - \boldsymbol{\xi}_h\|_{0,\mathrm{D}_h} + \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{0,\mathrm{D}_h}.\end{aligned} \tag{2.55}$$

On the other hand, the same arguments as for (2.44) and (2.45) imply

$$\|\boldsymbol{\sigma} - \boldsymbol{\xi}_h\|_{0,\mathrm{D}_h} + \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{0,\mathrm{D}_h} \leq \left(\frac{2}{2-\sqrt{2}}\right)\left(\|\boldsymbol{\sigma}_0 - \boldsymbol{\sigma}_{0,h}\|_{0,\mathrm{D}_h} + \|\boldsymbol{\sigma}_0 - \boldsymbol{\xi}_{0,h}\|_{0,\mathrm{D}_h}\right). \tag{2.56}$$

Combining (2.55) and (2.56), and employing the error estimate given by Lemma 2.4, it yields

$$\begin{aligned}\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{0,\mathrm{D}_h^c} \lesssim h^{l+1}\|\boldsymbol{\sigma}\|_{l+1,\Omega} &+ \sum_{e \in \mathcal{E}_h^\partial} \left|\!\left|\!\left|\boldsymbol{\sigma}^{\mathtt{d}} - \mathbf{E}_h(\boldsymbol{\xi}^{\mathtt{d}}_{0,h})\right|\!\right|\!\right|_e\\ &+ \|\boldsymbol{\sigma}_0 - \boldsymbol{\xi}_{0,h}\|_{0,\mathrm{D}_h} + \inf_{\mathbf{v}_h \in \mathbf{Q}_h(\mathrm{D}_h)} \|\mathbf{u} - \mathbf{v}_h\|_{0,\mathrm{D}_h}.\end{aligned} \tag{2.57}$$

Furthermore, by using the fact that $\|\cdot\|_{0,\widetilde{T}^e_{ext}}$ and $|||\cdot|||_e$ are equivalent norms over $\mathbb{L}^2(\widetilde{T}^e_{ext})$ (cf. Lemma 1.6), and noting that $\|\boldsymbol{\tau}^{\mathtt{d}}\|_{0,\widetilde{T}^e_{ext}} \lesssim \|\boldsymbol{\tau}\|_{0,\widetilde{T}^e_{ext}}$ holds for all $\boldsymbol{\tau} \in \mathbb{H}(\mathbf{div};\widetilde{T}^e_{ext})$, we find

$$\begin{aligned}\sum_{e \in \mathcal{E}_h^\partial} \left|\!\left|\!\left|\boldsymbol{\sigma}^{\mathtt{d}} - \mathbf{E}_h(\boldsymbol{\xi}^{\mathtt{d}}_{0,h})\right|\!\right|\!\right|_e &= \sum_{e \in \mathcal{E}_h^\partial} \left|\!\left|\!\left|\boldsymbol{\sigma}^{\mathtt{d}} - \mathbf{E}_h(\boldsymbol{\xi}^{\mathtt{d}}_{h})\right|\!\right|\!\right|_e\\ &\lesssim \sum_{e \in \mathcal{E}_h^\partial} \|\boldsymbol{\sigma}^{\mathtt{d}} - \mathbf{E}_h(\boldsymbol{\xi}^{\mathtt{d}}_h)\|_{0,\widetilde{T}^e_{ext}} \lesssim \sum_{e \in \mathcal{E}_h^\partial} \|\boldsymbol{\sigma} - \mathbf{E}_h(\boldsymbol{\xi}_h)\|_{0,\widetilde{T}^e_{ext}}\\ &\lesssim h^{l+1}\|\boldsymbol{\sigma}\|_{l+1,\Omega} + \|\boldsymbol{\sigma} - \boldsymbol{\xi}_h\|_{0,\mathrm{D}_h} \lesssim h^{l+1}\|\boldsymbol{\sigma}\|_{l+1,\Omega} + \|\boldsymbol{\sigma}_0 - \boldsymbol{\xi}_{0,h}\|_{0,\mathrm{D}_h}.\end{aligned} \tag{2.58}$$

Therefore, (2.53) is obtained by gathering (2.57) and (2.58), and by noting, thanks to the identity (2.13), that $\|\boldsymbol{\sigma}_0\|_{l+1,\mathrm{D}_h} \lesssim \|\boldsymbol{\sigma}\|_{l+1,\Omega}$. The estimate (2.54) is obtained analogously to (2.53), but considering the estimate (2.52) instead of (2.51). □

The following result is a direct consequence of inequalities (2.50) and (2.53).

Corollary 2.8. *Let us suppose that hypotheses of Lemma 2.7 are satisfied. Let p and p_h be defined as in (2.4) and (2.49), respectively. There holds*

$$\|p-p_h\|_{0,\mathrm{D}_h^c} \lesssim \inf_{\mathbf{w}_h\in\mathbf{Q}_h(\mathrm{D}_h)} \|\mathbf{u}-\mathbf{w}_h\|_{0,\mathrm{D}_h} + \inf_{\boldsymbol{\xi}_{0,h}\in\mathbb{H}_{0,h}(\mathrm{D}_h)} \|\boldsymbol{\sigma}_0-\boldsymbol{\xi}_{0,h}\|_{0,\mathrm{D}_h} + h^{l+1}\|\boldsymbol{\sigma}\|_{l+1,\Omega}.$$

To conclude this section, it remains to specify $\mathbf{u}_h$ in D_h^c. In doing so, we proceed exactly as in [57, Section 2.1.3]. In fact, given an edge $e\in\mathcal{E}_h^{\partial}$, it is easy to see that for each point $\mathbf{y}\in\widetilde{T}_{ext}^e$ there exists a transferring path $\mathscr{C}(\mathbf{x})$, starting at $\mathbf{x}\in\Gamma_h$ and ending at $\widetilde{\mathbf{x}}\in\Gamma$, such that $\mathbf{y}=\mathbf{x}+(\varepsilon/\ell(\mathbf{x}))(\widetilde{\mathbf{x}}-\mathbf{x})$ for some $\varepsilon\in[0,\ell(\mathbf{x})]$. As a result, the definition of $\mathbf{u}_h$ in D_h^c can be stated similarly to the one of $\widetilde{\mathbf{g}}_h$, that is,

$$\mathbf{u}_h(\mathbf{y}) := \mathbf{u}(\widetilde{\mathbf{y}}) - \frac{1}{2\mu}\int_0^{|\widetilde{\mathbf{y}}-\mathbf{y}|} \boldsymbol{\sigma}_h^{\mathtt{d}}(\mathbf{y}+\eta\mathbf{k}(\mathbf{y}))\mathbf{k}(\mathbf{y})\,d\eta, \tag{2.59}$$

where $\boldsymbol{\sigma}_h$ is defined as in (2.47), $\widetilde{\mathbf{y}}:=\widetilde{\mathbf{x}}$ and $\mathbf{k}(\mathbf{y}):=(\widetilde{\mathbf{y}}-\mathbf{y})/|\widetilde{\mathbf{y}}-\mathbf{y}|$. Actually, it is possible to define $\mathbf{u}_h$ with either $\boldsymbol{\sigma}_h$ or $\boldsymbol{\sigma}_{0,h}$ upon taking into account the identity $\boldsymbol{\sigma}_h^{\mathtt{d}}=\boldsymbol{\sigma}_{0,h}^{\mathtt{d}}$.

The next lemma provides an upper bound for $(\mathbf{u}-\mathbf{u}_h)$ in the $\mathbf{L}^2$-norm on D_h^c. The proof, which involves the estimate (2.53), is basically the same as for Lemma 1.9, and for this reason is omitted.

Lemma 2.9. *Suppose that the hypotheses of Lemma 2.7 are satisfied. Then, there holds*

$$\|\mathbf{u}-\mathbf{u}_h\|_{0,\mathrm{D}_h^c} \lesssim Rh\left(\inf_{\mathbf{w}_h\in\mathbf{Q}_h(\mathrm{D}_h)} \|\mathbf{u}-\mathbf{w}_h\|_{0,\mathrm{D}_h} + \inf_{\boldsymbol{\xi}_{0,h}\in\mathbb{H}_{0,h}(\mathrm{D}_h)} \|\boldsymbol{\sigma}_0-\boldsymbol{\xi}_{0,h}\|_{0,\mathrm{D}_h}\right) + Rh^{l+2}\|\boldsymbol{\sigma}\|_{l+1,\Omega}.$$

Finally, the following theorem provides the theoretical rate of convergence of our Galerkin scheme (2.16) and the main unknowns, provided the usual regularity assumptions on the exact solution.

Theorem 2.10. *In addition to the hypotheses of Theorem 2.4 and Lemma 2.7, suppose that there exists $s\in(0,k+1]$ such that $\boldsymbol{\sigma}\in\mathbb{H}^s(\Omega)$, $\mathbf{div}\,\boldsymbol{\sigma}\in\mathbf{H}^s(\Omega)$ and $\mathbf{u}\in\mathbf{H}^s(\Omega)$. Then, there hold*

$$\|(\boldsymbol{\sigma}_0-\boldsymbol{\sigma}_{0,h},\mathbf{u}-\mathbf{u}_h)\|_{\mathbb{H}(\mathbf{div};\mathrm{D}_h)\times\mathbf{L}^2(\mathrm{D}_h)} \lesssim h^s\left(\|\boldsymbol{\sigma}\|_{s,\Omega}+\|\mathbf{div}\,\boldsymbol{\sigma}\|_{s,\Omega}+\|\mathbf{u}\|_{s,\Omega}\right)$$

and

$$\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathbf{div},\mathrm{D}_h}+\|p-p_h\|_{0,\mathrm{D}_h} \lesssim h^s\left(\|\boldsymbol{\sigma}\|_{s,\Omega}+\|\mathbf{div}\,\boldsymbol{\sigma}\|_{s,\Omega}+\|\mathbf{u}\|_{s,\Omega}\right).$$

Furthermore, in the non-meshed region D_h^c, we have

$$\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathbf{div},\widetilde{\mathcal{T}}_h}+\|p-p_h\|_{0,\mathrm{D}_h^c} \lesssim h^s\left(\|\boldsymbol{\sigma}\|_{s,\Omega}+\|\mathbf{div}\,\boldsymbol{\sigma}\|_{s,\Omega}+\|\mathbf{u}\|_{s,\Omega}\right)$$

and

$$\|\mathbf{u}-\mathbf{u}_h\|_{0,\mathrm{D}_h^c} \lesssim Rh^{s+1}\left(\|\boldsymbol{\sigma}\|_{s,\Omega}+\|\mathbf{div}\,\boldsymbol{\sigma}\|_{s,\Omega}+\|\mathbf{u}\|_{s,\Omega}\right).$$

Proof. It is concluded from Theorem 2.4, Corollary 2.5, Lemma 2.7, Corollary 2.8, Lemma 2.9, the approximation properties (2.21)-(2.23), and (2.25), and the usual interpolation estimates. □

It is interesting to note here that the extra power of h related to $\|\mathbf{u}-\mathbf{u}_h\|_{0,\mathrm{D}_h^c}$ follows exclusively from Assumption **(A1)**, i.e., from the fact that in (2.59) the maximum length of the integration segments is of order of Rh. However, the convergence rate of the method is entirely determined by the error estimates on the computational domain D_h.

2.5 A residual-based *a posteriori* error analysis

In this section we develop a reliable and quasi-efficient residual-based *a posteriori* error estimator for the Galerkin scheme (2.16). For simplicity, however, we restrict ourselves to the problem in two dimensions and to the case where Γ_h is constructed by interpolating Γ by a piecewise linear function and D_h is contained in Ω. In that case, the distance between Γ_h and Γ is of order h^2. We emphasize that the *a priori* error analysis in previous sections holds under the less restrictive assumption that $d(\Gamma_h,\Gamma)$ is of only order h, as long as Assumption **(A.1)** and **(A.2)** are satisfied. However, the corresponding *a posteriori* error analysis of the latter case is not trivial and is subject of ongoing work. In Section 2.5.3 we will comment how to deal with the case when D_h is not necessarily contained in Ω. We will discuss the *a posteriori* error estimator in three dimensions in Section 2.5.4.

We start by introducing some useful notation and previous results. In what follows, h_e stands for the length of a given edge $e\in\mathcal{E}_h$. Moreover, for every $e\in\mathcal{E}_h$ we fix a unit normal vector $\mathbf{n}_e:=(n_{e,1},n_{e,2})^T$ to the edge e, and let $\mathbf{t}_e:=(-n_{e,2},n_{e,1})^T$ be the unit tangential vector along e. We define $\mathbf{n}_{\Gamma_e}$ and $\mathbf{t}_{\Gamma_e}$ similarly. In particular, for every $e\in\mathcal{E}_h^{\partial}$ (resp. $\Gamma_e\subset\Gamma$), we take $\mathbf{n}_e$ (resp. $\mathbf{n}_{\Gamma_e}$) as the vector pointing in the outward direction of Γ_h (resp. Γ) from D_h (resp. Ω). However, when no confusion arises we will simply write $\mathbf{n}$ and $\mathbf{t}$ instead of $\mathbf{n}_e$ and $\mathbf{t}_e$ (or, $\mathbf{n}_{\Gamma_e}$ and $\mathbf{t}_{\Gamma_e}$), respectively. Now, given an edge $e\in\mathcal{E}_h$, $\mathbf{v}\in\mathbf{L}^2(\Omega)$ and $\boldsymbol{\tau}\in\mathbb{L}^2(\Omega)$, such that $\mathbf{v}|_T\in[\mathcal{C}(T)]^2$ and $\boldsymbol{\tau}|_T\in[\mathcal{C}(T)]^{2\times 2}$ on each $T\in\mathcal{T}_h$, we let $[\![\mathbf{v}]\!]$ and $[\![\boldsymbol{\tau}\mathbf{t}]\!]$ be the corresponding jumps across e, that is,

$$[\![\mathbf{v}]\!]:=\left(\mathbf{v}|_{T^+}\right)|_e-\left(\mathbf{v}|_{T^-}\right)|_e\quad\text{and}\quad[\![\boldsymbol{\tau}\mathbf{t}]\!]:=\left\{\left(\boldsymbol{\tau}|_{T^+}\right)|_e-\left(\boldsymbol{\tau}|_{T^-}\right)|_e\right\}\mathbf{t},$$

where T^+ and T^- are two triangles of $\mathcal{T}_h$ sharing a common edge e. Finally, if $\mathbf{v}:=(v_i)_{i=1,2}$ and $\boldsymbol{\tau}:=(\tau_{ij})_{i,j=1,2}$ are sufficiently smooth vector-valued and tensor-valued functions, respectively, we let

$$\underline{\mathbf{curl}}\,(\mathbf{v}):=\begin{pmatrix}\dfrac{\partial v_1}{\partial x_2} & -\dfrac{\partial v_1}{\partial x_1}\\[2mm] \dfrac{\partial v_2}{\partial x_2} & -\dfrac{\partial v_2}{\partial x_1}\end{pmatrix}\quad\text{and}\quad\mathrm{curl}\,(\boldsymbol{\tau}):=\begin{pmatrix}\dfrac{\partial \tau_{12}}{\partial x_1}-\dfrac{\partial \tau_{11}}{\partial x_2}\\[2mm] \dfrac{\partial \tau_{22}}{\partial x_1}-\dfrac{\partial \tau_{21}}{\partial x_2}\end{pmatrix}.$$

Let $(\boldsymbol{\sigma}_{0,h},\mathbf{u}_h)\in\mathbb{H}_{0,h}(\mathrm{D}_h)\times\mathbf{Q}_h(\mathrm{D}_h)$ be the unique solution of (2.16) and $\boldsymbol{\sigma}_h$ be defined as in (2.35). For the forthcoming analysis we introduce an element-by-element postprocessed velocity $\mathbf{u}_h^\star$ being the unique function in $\prod_{T\in\mathcal{T}_h}\mathbf{P}_{k+1}(T)$, such that, for all $T\in\mathcal{T}_h$,

$$\begin{aligned}\int_T\nabla\mathbf{u}_h^\star:\nabla\mathbf{q}&=\frac{1}{2\mu}\int_T\boldsymbol{\sigma}_h^{\mathtt{d}}:\nabla\mathbf{q}\quad\forall\,\mathbf{q}\in\mathbf{P}_{k+1}(T),\\ \int_T\mathbf{u}_h^\star&=\int_T\mathbf{u}_h.\end{aligned}\tag{2.60}$$

It is immediate to check that $\mathbf{u}_h^\star$ is well-defined. Moreover, if we assume that $\mathbf{u}\in\mathbf{H}^{m+1}(\mathrm{D}_h)$ and $\boldsymbol{\sigma}\in\mathbb{H}^l(\mathrm{D}_h)$, with $m,l\in[1,k+1]$, it is not difficult to verify (see, e.g., [55, Theorem 5.2]) that

$$\|\mathbf{u}-\mathbf{u}_h^\star\|_{0,\mathrm{D}_h}\lesssim h^{\min\{l+1,m+1\}}\left(\|\boldsymbol{\sigma}\|_{l,\mathrm{D}_h}+\|\mathbf{u}\|_{m+1,\mathrm{D}_h}\right).\tag{2.61}$$

Therefore, the pair $(\boldsymbol{\sigma}_h,\mathbf{u}_h^\star)$ is an optimal convergent approximation of $(\boldsymbol{\sigma},\mathbf{u})\in\mathbb{H}_h(\mathrm{D}_h)\times\mathbf{Q}_h(\mathrm{D}_h)$. For the sake of simplicity, the extrapolation of $\mathbf{u}_h^\star$ on D_h^c, in the sense of (2.19), will be denoted simply as $\mathbf{u}_h^\star$.

We introduce the following global a posteriori error estimator:

$$\Theta := \left(\sum_{T \in \mathcal{T}_h} \Theta_T^2 \right)^{1/2}, \tag{2.62}$$

where Θ_T is the local error indicator defined, for each $T \in \mathcal{T}_h$, as

$$\begin{aligned} \Theta_T^2 := h_T^2 \left\| \mathrm{curl} \left\{ \frac{1}{2\mu} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\} \right\|_{0,T}^2 + \sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h^i} \left(h_e \left\| \left[\!\!\left[\frac{1}{2\mu} \boldsymbol{\sigma}_h^{\mathtt{d}} \mathbf{t} \right]\!\!\right] \right\|_{0,e}^2 + h_e^{-1} \| [\![\mathbf{u}_h^\star]\!] \|_{0,e}^2 \right) \\ + \left\| \frac{1}{2\mu} \boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla \mathbf{u}_h^\star \right\|_{0,T}^2 + \| \mathbf{f} + \mathbf{div}\, \boldsymbol{\sigma}_h \|_{0,T}^2 + \sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h^\partial} \| \mathbf{f} + \mathbf{div}\, \boldsymbol{\sigma}_h \|_{0, \widetilde{T}_{ext}^e}^2 \\ + \| \mathbf{u}_h - \mathbf{u}_h^\star \|_{0,T}^2 + \sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h^\partial} \| \mathbf{u}_h - \mathbf{u}_h^\star \|_{0, \widetilde{T}_{ext}^e}^2 + \mathbb{J}_T^2 + \mathbb{K}_T^2. \end{aligned} \tag{2.63}$$

Here, $\mathbb{J}_T$ and $\mathbb{K}_T$ are computable terms concerning the curved boundary Γ, which take the form

$$\mathbb{J}_T := \left(\sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h^\partial} h_e^{-1} \| \mathbf{g} - \mathbf{u}_h^\star \|_{0,\Gamma_e}^2 \right)^{1/2}, \tag{2.64}$$

and

$$\mathbb{K}_T := \left(\sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h^\partial} h_{T^e} \left\| \frac{d\mathbf{g}}{d\mathbf{t}} - \frac{1}{2\mu} \boldsymbol{\sigma}_h^{\mathtt{d}} \mathbf{t} \right\|_{0,\Gamma_e}^2 \right)^{1/2}. \tag{2.65}$$

Note that, from the strong equations (2.5) and the regularity of the continuous weak solution, the residual character of each term defining (2.63) becomes clear. Note also that (2.65) requires that $d\mathbf{g}/d\mathbf{t} \in \mathbf{L}^2(\Gamma_e)$ for each curved edge Γ_e being part of the boundary Γ, which is overcome below by simply assuming that $\mathbf{g} \in \mathbf{H}^1(\Gamma)$. Moreover, since by (2.61) with $l = m = k+1$ the postprocessed $\mathbf{u}_h^\star$ converges to $\mathbf{u}$ with order $\mathcal{O}(h^{k+2})$ in the $\mathbf{L}^2(\mathrm{D}_h)$-norm, it should be expected, and this is verified in practice (cf. Section 2.6), that the global a posteriori error estimator Θ retains the rate of convergence of our method, i.e., $\mathcal{O}(h^{k+1})$, if the solution is smooth enough.

We are now in position of establishing the main result of this section.

Theorem 2.11. *Assume that $\mathbf{g} \in \mathbf{H}^1(\Gamma)$. Then, there exist positive constant C_{rel} and C_{eff}, both independent of the meshsizes and the continuous and discrete solutions, such that*

$$\| (\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{u} - \mathbf{u}_h) \|_{\mathbb{H}(\mathbf{div};\Omega) \times \mathbf{L}^2(\Omega)} \leq C_{\mathrm{rel}} \Theta, \tag{2.66}$$

and

$$C_{\mathrm{eff}} \Theta \leq \| (\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{u} - \mathbf{u}_h) \|_{\mathbb{H}(\mathbf{div};\Omega) \times \mathbf{L}^2(\Omega)} + \mathbb{B}, \tag{2.67}$$

where

$$\mathbb{B} := \left(\sum_{T \in \mathcal{T}_h} \mathbb{J}_T^2 \right)^{1/2} + \left(\sum_{T \in \mathcal{T}_h} \mathbb{K}_T^2 \right)^{1/2}, \tag{2.68}$$

and $\mathbb{J}_T$ and $\mathbb{K}_T$ are given by (2.64) and (2.65), respectively.

We recall from Section 2.4.2 that $\boldsymbol{\sigma}_h$ in D_h^c is obtained by (2.47) and satisfies $\int_\Omega \mathrm{tr}\,(\boldsymbol{\sigma}_h) = 0$. Then, since Γ_h is constructed by a picewise linear interpolation of Γ, it is clear that $\boldsymbol{\sigma}_h \in \mathbb{H}_0(\mathbf{div};\Omega)$, and hence the norm in the left-hand side of (2.66) makes sense. In addition, we notice from (2.67) that Θ is efficient up to the term $\mathbb{B}$, which is usually referred as quasi-efficiency (see, e.g., [5, 83]). More importantly, the terms $\mathbb{J}_T$ and $\mathbb{K}_T$ lie on both sides of the inequality (2.67), which does not represent any problem since they provides computable estimates for the approximations $\mathbf{u}_h^\star$ and $(2\mu)^{-1}\boldsymbol{\sigma}_h^{\mathtt{d}}\mathbf{t}$ of the boundary data $\mathbf{g}$ and its tangential derivative along Γ, respectively. It should be noted, however, that $\mathbb{B}$ must have at least the same rate of convergence of the global error if the exact solution is smooth enough. In section 2.5.2 we treat this matter in more detail.

The proof of Theorem 2.11 is separated into several steps. In Section 2.5.1 we prove that Θ satisfies the reliability property (2.66), whereas the corresponding quasi-efficiency property (2.67) is derived in Section 2.5.2.

2.5.1 Reliability of the a posteriori error estimator

We proceed similarly as in [80] (see also [78, 81]), that is, we start by using the global inf-sup condition in (2.7). In fact, we have

$$\begin{aligned}\|(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,\mathbf{u}-\mathbf{u}_h)\|_{\mathbb{H}(\mathbf{div};\Omega)\times\mathbf{L}^2(\Omega)} &\lesssim \|\mathbf{u}_h-\mathbf{u}_h^\star\|_{0,\Omega}+\|(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,\mathbf{u}-\mathbf{u}_h^\star)\|_{\mathbb{H}(\mathbf{div};\Omega)\times\mathbf{L}^2(\Omega)}\\ &\lesssim \|\mathbf{u}_h-\mathbf{u}_h^\star\|_{0,\Omega}+\sup_{\substack{(\boldsymbol{\tau},\mathbf{v})\in\mathbb{H}_0(\mathbf{div};\Omega)\times\mathbf{L}^2(\Omega)\\(\boldsymbol{\tau},\mathbf{v})\neq\mathbf{0}}}\frac{|a(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,\boldsymbol{\tau})+b(\boldsymbol{\tau},\mathbf{u}-\mathbf{u}_h^\star)+b(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,\mathbf{v})|}{\|(\boldsymbol{\tau},\mathbf{v})\|_{\mathbb{H}(\mathbf{div};\Omega)\times\mathbf{L}^2(\Omega)}},\end{aligned}$$

from which

$$\|(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,\mathbf{u}-\mathbf{u}_h)\|_{\mathbb{H}(\mathbf{div};\Omega)\times\mathbf{L}^2(\Omega)} \lesssim \|\mathbf{u}_h-\mathbf{u}_h^\star\|_{0,\Omega}+\|\mathbf{f}+\mathbf{div}\,\boldsymbol{\sigma}_h\|_{0,\Omega}+\|\mathcal{R}\|_{\mathbb{H}_0(\mathbf{div};\Omega)'}, \tag{2.69}$$

where $\mathcal{R}:\mathbb{H}_0(\mathbf{div};\Omega)\to\mathbb{R}$ is the linear and bounded functional defined as

$$\mathcal{R}(\boldsymbol{\tau}) := \langle\boldsymbol{\tau}\mathbf{n}_\Gamma,\mathbf{g}\rangle_\Gamma - a(\boldsymbol{\sigma}_h,\boldsymbol{\tau}) - b(\boldsymbol{\tau},\mathbf{u}_h^\star) \quad \forall\,\boldsymbol{\tau}\in\mathbb{H}_0(\mathbf{div};\Omega). \tag{2.70}$$

To obtain the reliability estimate (2.66) it suffices to bound (2.70). We notice that in the case of mixed methods with Ω being polygonal, this is typically accomplished by using a stable Helmholtz decomposition of $\boldsymbol{\tau}$. In what follows, with the help of an auxiliary polygon different from D_h, we shall extend that idea to domains Ω with curved boundary.

Given $e\in\mathcal{E}_h^\partial$ such that $e\neq\Gamma_e$, we suppose that there exists an *auxiliary triangle* $\widetilde{T}_{aux}^e$, with diameter $h_{\widetilde{T}_{aux}^e}$, satisfying

(B1) $\widetilde{T}_{aux}^e$ has e as a boundary edge, $\Gamma_e\subset\widetilde{T}_{aux}^e$, $h_{\widetilde{T}_{aux}^e}\simeq h_{T^e}$, $|\Gamma_e|\simeq h_e$; and if $F=\overline{\widetilde{T}_{aux}^{e_i}}\cap\overline{\widetilde{T}_{aux}^{e_j}}$, with $e_i,e_j\in\mathcal{E}_h^\partial$, $i\neq j$, then F is either a common vertex or a common edge of $\widetilde{T}_{aux}^{e_i}$ and $\widetilde{T}_{aux}^{e_j}$; see an illustration in Figure 2.2.

We observe that in the case of $e=\Gamma_e$, we can simply take $\widetilde{T}_{aux}^e$ as T^e. For this reason, from now on we assume, without loss of generality, that for all $e\in\mathcal{E}_h^\partial$, $e\neq\Gamma_e$. By defining $\widetilde{\mathcal{T}}_h^{aux}:=\left\{\widetilde{T}_{aux}^e:\ e\in\mathcal{E}_h^\partial\right\}$, we further assume that

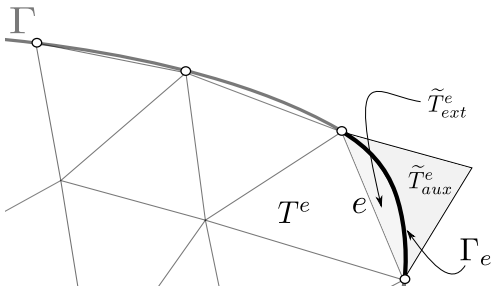


Figure 2.2: Example of an *auxiliary triangle* $\widetilde{T}^e_{aux}$ (gray region). (figure produced by the author)

(B2) the triangulation $\mathcal{T}_h^* := \mathcal{T}_h \cup \widetilde{\mathcal{T}}_h^{aux}$ is shape-regular.

These hypotheses are expected to be satisfied on sufficiently fine meshes since Γ_h is constructed through a picewise linear interpolation of Γ, even though, as we shall see later, the auxiliary triangles will not be used to compute our a posteriori error estimator. It is then straightforward to extend the Raviart–Thomas interpolation operator (cf. Section 2.3) to the polygonal region D_h^* induced by the triangularization $\mathcal{T}_h^*$, say $\overline{\mathrm{D}_h^*} = \bigcup \{T : \ T \in \mathcal{T}_h^*\}$. Therefore, the approximation properties of this operator also hold in $\mathcal{T}_h^*$.

Next, we introduce the Clément interpolation operator [54]

$$\mathcal{I}_h : \mathrm{H}^1(\mathrm{D}_h^*) \to \left\{ v \in \mathcal{C}(\overline{\mathrm{D}_h^*}) : \ v\big|_T \in \mathrm{P}_1(T) \ \ \forall\, T \in \mathcal{T}_h^* \right\}.$$

From this operator we recall the following classical approximation properties.

Lemma 2.12. *Assume that* **(B1)**-**(B2)** *are satisfied. Then, for all* $v \in \mathrm{H}^1(\mathrm{D}_h^*)$ *there hold*

$$\|v - \mathcal{I}_h(v)\|_{0,T} \lesssim h_T |v|_{1,\Delta(T)} \quad \forall\, T \in \mathcal{T}_h^* \tag{2.71}$$

and

$$\|v - \mathcal{I}_h(v)\|_{0,e} \lesssim h_e |v|_{1,\Delta(e)} \quad \forall \text{ edge } e \text{ of } \mathcal{T}_h^*, \tag{2.72}$$

where $\Delta(T) := \bigcup \{T' \in \mathcal{T}_h^* : \ T \cap T' \neq \varnothing\}$ *and* $\Delta(e) := \bigcup \{T' \in \mathcal{T}_h^* : \ e \cap T' \neq \varnothing\}$.

Let us continue with the estimation of $\|\mathcal{R}\|_{[\mathbb{H}_0(\mathbf{div};\Omega)]'}$. For this, we let (see, e.g., [78, Setion 4])

$$\boldsymbol{\tau} = \boldsymbol{\zeta} + \underline{\mathbf{curl}}(\boldsymbol{\varphi}) \quad \text{in} \quad \Omega, \tag{2.73}$$

with $\boldsymbol{\chi} \in \mathbb{H}^1(\Omega)$ and $\boldsymbol{\varphi} \in \mathbf{H}^1(\Omega)$ satisfying the stability property

$$\|\boldsymbol{\zeta}\|_{1,\Omega} + \|\boldsymbol{\varphi}\|_{1,\Omega} \lesssim \|\boldsymbol{\tau}\|_{\mathbf{div},\Omega}. \tag{2.74}$$

Furthermore, following essentially the ideas in [78, Section 4.1] (see also the proof of Lemma 3.8 in [80]), we specify the discrete version of the identity (2.73). First, we recall from [129] that for any $v \in \mathrm{H}^1(\Omega)$ there exists an extension $\mathscr{E}(v) \in \mathrm{H}^1(\mathbb{R}^2)$ such that $\mathscr{E}(v)|_\Omega = v$ and $\|\mathscr{E}(v)\|_{1,\mathbb{R}^2} \lesssim \|v\|_{1,\Omega}$. Then, we let

$$\boldsymbol{\zeta}_h := \boldsymbol{\Pi}_h^k \left(\boldsymbol{\mathscr{E}}(\boldsymbol{\zeta})\big|_{\mathrm{D}_h^*} \right) \quad \text{and} \quad \boldsymbol{\varphi}_h := \boldsymbol{\mathcal{I}}_h \left(\boldsymbol{\mathscr{E}}(\boldsymbol{\varphi})\big|_{\mathrm{D}_h^*} \right),$$

where $\mathbf{\Pi}_h^k$ is the Raviart–Thomas interpolation operator described before, whereas $\boldsymbol{\mathscr{E}}$ and $\boldsymbol{\mathcal{I}}_h$ are defined componentwise by the extension operator $\mathscr{E}$ and the Clément interpolant $\mathcal{I}_h$, respectively. The discrete Helmholtz decomposition is then defined as

$$\boldsymbol{\tau}_h := \boldsymbol{\zeta}_h + \underline{\mathbf{curl}}\,(\boldsymbol{\varphi}_h) + c_0\mathbb{I} \quad \text{in} \quad \mathrm{D}_h^*, \tag{2.75}$$

with $c_0 := -\frac{1}{2|\Omega|}\int_\Omega \mathrm{tr}\,(\boldsymbol{\zeta}_h + \underline{\mathbf{curl}}\,(\boldsymbol{\varphi}_h))$ chosen such that $\int_\Omega \mathrm{tr}\,(\boldsymbol{\tau}_h) = 0$.

In this way, adding and subtracting $\boldsymbol{\tau}_h$ in the argument of $\mathcal{R}$ defined in (2.70), using the identities (2.73) and (2.75), noting that $c_0\mathbb{I}$ vanishes in the definition of $\mathcal{R}$ due to the compatibility condition (2.3), we obtain

$$\mathcal{R}(\boldsymbol{\tau}) = \mathcal{R}(\boldsymbol{\tau}_h) + \mathcal{R}(\boldsymbol{\zeta} - \boldsymbol{\zeta}_h) + \mathcal{R}(\underline{\mathbf{curl}}\,(\boldsymbol{\varphi} - \boldsymbol{\varphi}_h)). \tag{2.76}$$

In particular, from (2.70) and the identity $\boldsymbol{\sigma}_h^{\mathtt{d}} : \boldsymbol{\tau}_h^{\mathtt{d}} = \boldsymbol{\sigma}_h^{\mathtt{d}} : \boldsymbol{\tau}_h$, it follows that

$$\mathcal{R}(\boldsymbol{\tau}_h) = \sum_{e\in\mathcal{E}_h^\partial}\int_{\Gamma_e} \mathbf{g}\cdot(\boldsymbol{\tau}_h\mathbf{n}_{\Gamma_e}) - \frac{1}{2\mu}\int_\Omega \boldsymbol{\sigma}_h^{\mathtt{d}} : \boldsymbol{\tau}_h - \int_\Omega \mathbf{u}_h^\star\cdot\mathbf{div}\,\boldsymbol{\tau}_h. \tag{2.77}$$

Then, splintting the integrals over Ω into D_h and D_h^c, integrating by parts elementwise on each of these regions, and recalling that, for every $e\in\mathcal{E}_h^\partial$, the vector $\mathbf{n}_e$ is pointing outwards from D_h, there hold

$$\begin{aligned}
&\frac{1}{2\mu}\int_{\mathrm{D}_h}\boldsymbol{\sigma}_h^{\mathtt{d}} : \boldsymbol{\tau}_h + \int_{\mathrm{D}_h}\mathbf{u}_h^\star\cdot\mathbf{div}\,\boldsymbol{\tau}_h \\
&\quad= \sum_{T\in\mathcal{T}_h}\left(\int_T\left(\frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla\mathbf{u}_h^\star\right) : \boldsymbol{\tau}_h + \int_{\partial T}\mathbf{u}_h^\star\cdot(\boldsymbol{\tau}_h\mathbf{n})\right) \\
&\quad= \sum_{T\in\mathcal{T}_h}\int_T\left(\frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla\mathbf{u}_h^\star\right) : \boldsymbol{\tau}_h + \sum_{e\in\mathcal{E}_h^i}\int_e[\![\mathbf{u}_h^\star]\!]\cdot(\boldsymbol{\tau}_h\mathbf{n}_e) + \sum_{e\in\mathcal{E}_h^\partial}\int_e\mathbf{u}_h^\star\cdot(\boldsymbol{\tau}_h\mathbf{n}_e),
\end{aligned} \tag{2.78}$$

and

$$\begin{aligned}
&\frac{1}{2\mu}\int_{\mathrm{D}_h^c}\boldsymbol{\sigma}_h^{\mathtt{d}} : \boldsymbol{\tau}_h + \int_{\mathrm{D}_h^c}\mathbf{u}_h^\star\cdot\mathbf{div}\,\boldsymbol{\tau}_h \\
&\quad= \sum_{e\in\mathcal{E}_h^\partial}\left(\int_{\widetilde{T}_{ext}^e}\left(\frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla\mathbf{u}_h^\star\right) : \boldsymbol{\tau}_h + \int_{\partial\widetilde{T}_{ext}^e}\mathbf{u}_h^\star\cdot(\boldsymbol{\tau}_h\mathbf{n})\right) \\
&\quad= \sum_{e\in\mathcal{E}_h^\partial}\left(\int_{\widetilde{T}_{ext}^e}\left(\frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla\mathbf{u}_h^\star\right) : \boldsymbol{\tau}_h - \int_e\mathbf{u}_h^\star\cdot(\boldsymbol{\tau}_h\mathbf{n}_e) + \int_{\Gamma_e}\mathbf{u}_h^\star\cdot(\boldsymbol{\tau}_h\mathbf{n}_{\Gamma_e})\right).
\end{aligned} \tag{2.79}$$

Combining (2.77), (2.78) and (2.79), and observing that $\mathbf{u}_h^\star$ coincides with its extrapolation along every edge $e\in\mathcal{E}_h^\partial$, we obtain

$$\begin{aligned}
\mathcal{R}(\boldsymbol{\tau}_h) = &-\sum_{T\in\mathcal{T}_h}\int_T\left(\frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla\mathbf{u}_h^\star\right) : \boldsymbol{\tau}_h - \sum_{e\in\mathcal{E}_h^i}\int_e[\![\mathbf{u}_h^\star]\!]\cdot(\boldsymbol{\tau}_h\mathbf{n}_e) \\
&+ \sum_{e\in\mathcal{E}_h^\partial}\left(\int_{\Gamma_e}(\mathbf{g}-\mathbf{u}_h^\star)\cdot(\boldsymbol{\tau}_h\mathbf{n}_{\Gamma_e}) - \int_{\widetilde{T}_{ext}^e}\left(\frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla\mathbf{u}_h^\star\right) : \boldsymbol{\tau}_h\right).
\end{aligned} \tag{2.80}$$

The following result plays an important role when estimating $|\mathcal{R}(\boldsymbol{\tau}_h)|$.

Lemma 2.13. *Suppose that* **(B1)-(B2)** *hold. Then, for every edge* $e \in \mathcal{E}_h^\partial$ *and each* $\boldsymbol{\tau} \in \mathbb{H}^1(\widetilde{T}^e_{aux})$, *there hold*

$$\|\boldsymbol{\tau}\mathbf{n}_{\Gamma_e}\|_{0,\Gamma_e} \lesssim h_{T^e}^{-1/2}\|\boldsymbol{\tau}\|_{1,\widetilde{T}^e_{aux}}, \tag{2.81}$$

$$\|(\boldsymbol{\tau} - \boldsymbol{\Pi}_h^k(\boldsymbol{\tau}))\mathbf{n}_{\Gamma_e}\|_{0,\Gamma_e} \lesssim h_{T^e}^{1/2}\|\boldsymbol{\tau}\|_{1,\widetilde{T}^e_{aux}}, \tag{2.82}$$

$$\|\boldsymbol{\tau} - \boldsymbol{\Pi}_h^k(\boldsymbol{\tau})\|_{0,\Gamma_e} \lesssim h_{T^e}^{1/2}\|\boldsymbol{\tau}\|_{1,\widetilde{T}^e_{aux}}. \tag{2.83}$$

Proof. Given $e \in \mathcal{E}_h^\partial$, let F^e_{aux} be the usual invertible affine mapping satisfying $F^e_{aux}(T_{ref}) = \widetilde{T}^e_{aux}$, with T_{ref} denoting the reference element. Let then Γ_{ref} be the corresponding inverse image of Γ_e. For $\boldsymbol{\tau} \in \mathbf{H}^1(\widetilde{T}^e_{aux})$, we set $\widehat{\boldsymbol{\tau}} := \boldsymbol{\tau} \circ F^e_{aux}$. According to [86, Lemma 3], it follows that

$$\|\widehat{\boldsymbol{\tau}}\|^2_{0,\Gamma_{ref}} \lesssim \|\widehat{\boldsymbol{\tau}}\|_{0,T_{ref}}\|\widehat{\boldsymbol{\tau}}\|_{1,T_{ref}}.$$

A simple scaling argument yields

$$h_{\widetilde{T}^e_{aux}}\|\boldsymbol{\tau}\mathbf{n}_\Gamma\|^2_{0,\Gamma_e} \le h_{\widetilde{T}^e_{aux}}\|\boldsymbol{\tau}\|^2_{0,\Gamma_e} \lesssim \|\boldsymbol{\tau}\|^2_{0,\widetilde{T}^e_{aux}} + \left(h_{\widetilde{T}^e_{aux}}\right)^2\|\boldsymbol{\tau}\|^2_{1,\widetilde{T}^e_{aux}}. \tag{2.84}$$

Combined with the assumption that $h_{\widetilde{T}^e_{aux}} \simeq h_{T^e}$, this implies (2.81).

The remaining two estimates follow from (2.84), the approximation properties of the Raviart–Thomas interpolation operator, and the fact that $h_{\widetilde{T}^e_{aux}} \simeq h_{T^e}$, by replacing $\boldsymbol{\tau}$ by $\boldsymbol{\tau} - \boldsymbol{\Pi}_h^k(\boldsymbol{\tau})$. □

Similarly, for every $e \in \mathcal{E}_h^\partial$ and all $\mathbf{v} \in \mathbf{H}^1(\mathrm{D}_h^*)$, we have

$$\|\mathbf{v} - \boldsymbol{\mathcal{I}}_h(\mathbf{v})\|_{0,\Gamma_e} \lesssim h_{T^e}^{1/2}\|\mathbf{v}\|_{1,\Delta(\widetilde{T}^e_{aux})}, \tag{2.85}$$

where $\boldsymbol{\mathcal{I}}_h$ is the vector Clément interpolant introduced above and $\Delta(\widetilde{T}^e_{aux})$ is the union of all the elements of $\mathcal{T}_h^*$ intersecting with $\widetilde{T}^e_{aux}$.

Under Assumptions **(B1)-(B2)** the following three lemmas provide upper bounds for $|\mathcal{R}(\boldsymbol{\tau}_h)|$, $|\mathcal{R}(\boldsymbol{\zeta} - \boldsymbol{\zeta}_h)|$ and $|\mathcal{R}(\underline{\mathbf{curl}}(\boldsymbol{\varphi} - \boldsymbol{\varphi}_h))|$ arising from (2.76).

Lemma 2.14. *There holds*

$$|\mathcal{R}(\boldsymbol{\tau}_h)| \lesssim \left(\sum_{T\in\mathcal{T}_h}\Theta^2_{0,T}\right)^{1/2}\|\boldsymbol{\tau}\|_{\mathbf{div},\Omega}, \tag{2.86}$$

where

$$\Theta_{0,T} := \left\|\frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla\mathbf{u}_h^\star\right\|^2_{0,T} + \sum_{e\in\mathcal{E}(T)\cap\mathcal{E}_h^\partial}(h_{T^e})^{-1}\|\mathbf{g} - \mathbf{u}_h^\star\|^2_{0,\Gamma_e} + \sum_{e\in\mathcal{E}(T)\cap\mathcal{E}_h^i}(h_{T^e})^{-1}\|[\![\mathbf{u}_h^\star]\!]\|^2_{0,e}.$$

Proof. Applying the Cauchy–Schwarz inequality to each term in (2.80), and using (2.26), (2.81), the estimate $\|\cdot\|_{0,\widetilde{T}^e_{ext}} \lesssim |||\cdot|||_e$ (cf. Lemma 1.6), and the extrapolation constant (2.27), it follows that

$$\begin{aligned}|\mathcal{R}(\boldsymbol{\tau}_h)| &\lesssim \sum_{T\in\mathcal{T}_h}\left\|\frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla\mathbf{u}_h^\star\right\|_{0,T}\|\boldsymbol{\tau}_h\|_{0,T} + \sum_{e\in\mathcal{E}_h^i}C^e_{eq}\,(h_{T^e})^{-1/2}\,\|[\![\mathbf{u}_h^\star]\!]\|_{0,e}\|\boldsymbol{\tau}_h\|_{\mathbf{div},\mathcal{K}(e)}\\ &+ \sum_{e\in\mathcal{E}_h^\partial}\left((h_{T^e})^{-1/2}\,\|\mathbf{g} - \mathbf{u}_h^\star\|_{0,\Gamma_e}\,\|\boldsymbol{\tau}_h\|_{1,\widetilde{T}^e_{aux}} + \widetilde{C}^e_{ext}(\widetilde{r}_e)^{1/2}\left\|\frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla\mathbf{u}_h^\star\right\|_{0,T^e}\|\boldsymbol{\tau}_h\|_{0,\widetilde{T}^e_{ext}}\right),\end{aligned} \tag{2.87}$$

where

$$\mathcal{K}(e) := \bigcup \{T' \in \mathcal{T}_h : \ e \in \mathcal{E}(T')\}\,. \tag{2.88}$$

Notice that $\|\boldsymbol{\tau}_h\|_{0,\widetilde{T}^e_{ext}}$ can be bounded by $\|\boldsymbol{\tau}_h\|_{0,\widetilde{T}^e_{aux}}$ thanks to Assumption **(B1)**. Combining it with (2.87), using again the Cauchy–Schwarz inequality, and finally observing that (2.74) and (2.75) give $\|\boldsymbol{\tau}_h\|_{1,\mathrm{D}^*_h} \lesssim \|\boldsymbol{\tau}\|_{\mathbf{div},\Omega}$, we have

$$|\mathcal{R}(\boldsymbol{\tau}_h)| \lesssim \|\boldsymbol{\tau}\|_{\mathbf{div};\Omega} \left(\sum_{T\in\mathcal{T}_h} \left\| \frac{1}{2\mu}\boldsymbol{\sigma}^{\mathtt{d}}_h - \nabla \mathbf{u}^\star_h \right\|^2_{0,T} + \sum_{e\in\mathcal{E}^i_h} \left(C^e_{eq}\right)^2 (h_{T^e})^{-1} \|[\![\mathbf{u}^\star_h]\!]\|^2_{0,e} \right.$$
$$\left. + \sum_{e\in\mathcal{E}^\partial_h} \left((h_{T^e})^{-1} \|\mathbf{g} - \mathbf{u}^\star_h\|^2_{0,\Gamma_e} + \left(\widetilde{C}^e_{ext}\right)^2 \widetilde{r}_e \left\| \frac{1}{2\mu}\boldsymbol{\sigma}^{\mathtt{d}}_h - \nabla \mathbf{u}^\star_h \right\|^2_{0,T^e} \right) \right)^{1/2},$$

where, by Assumption **(A1)**, $\widetilde{r}_e \le C$ for all $e \in \mathcal{E}^\partial_h$. This completes the proof. □

Lemma 2.15. *There holds*

$$|\mathcal{R}(\boldsymbol{\zeta} - \boldsymbol{\zeta}_h)| \lesssim \left(\sum_{T\in\mathcal{T}_h} \Theta^2_{1,T} \right)^{1/2} \|\boldsymbol{\zeta}\|_{1,\Omega}, \tag{2.89}$$

where

$$\Theta_{1,T} := h^2_T \left\| \frac{1}{2\mu}\boldsymbol{\sigma}^{\mathtt{d}}_h - \nabla \mathbf{u}^\star_h \right\|^2_{0,T} + \sum_{e\in\mathcal{E}(T)\cap\mathcal{E}^\partial_h} h_{T^e} \|\mathbf{g} - \mathbf{u}^\star_h\|^2_{0,\Gamma_e} + \sum_{e\in\mathcal{E}(T)\cap\mathcal{E}^i_h} h_{T^e} \|[\![\mathbf{u}^\star_h]\!]\|^2_{0,e}.$$

Proof. We first observe that, making use of the approximation properties of the Raviart–Thomas interpolation operator, we obtain, for every edge $e \in \mathcal{E}^\partial_h$,

$$\|\boldsymbol{\zeta} - \boldsymbol{\zeta}_h\|_{0,\widetilde{T}^e_{ext}} \le \|\boldsymbol{\mathscr{E}}(\boldsymbol{\zeta}) - \boldsymbol{\zeta}_h\|_{0,\widetilde{T}^e_{aux}} \lesssim h_{T^e} \|\boldsymbol{\mathscr{E}}(\boldsymbol{\zeta})\|_{1,\widetilde{T}^e_{aux}}, \tag{2.90}$$

since, by Assumption **(B1)**, $\widetilde{T}^e_{ext} \subset \widetilde{T}^e_{aux}$ and $h_{\widetilde{T}^e_{aux}} \simeq h_{T^e}$. Then, after replacing $\boldsymbol{\tau}_h$ by $(\boldsymbol{\zeta} - \boldsymbol{\zeta}_h)$ in (2.80), we use similar arguments as in the previous lemma to obtain

$$|\mathcal{R}(\boldsymbol{\zeta} - \boldsymbol{\zeta}_h)| \lesssim \sum_{T\in\mathcal{T}_h} h_T \left\| \frac{1}{2\mu}\boldsymbol{\sigma}^{\mathtt{d}}_h - \nabla \mathbf{u}^\star_h \right\|_{0,T} \|\boldsymbol{\zeta}\|_{1,T} + \sum_{e\in\mathcal{E}^i_h} h^{1/2}_e \|[\![\mathbf{u}^\star_h]\!]\|_{0,e} \|\boldsymbol{\zeta}\|_{1,\mathcal{K}(e)}$$
$$+ \sum_{e\in\mathcal{E}^\partial_h} \left((h_{T^e})^{1/2} \|\mathbf{g} - \mathbf{u}^\star_h\|_{0,\Gamma_e} + \widetilde{C}^e_{ext} h_{T^e} (\widetilde{r}_e)^{1/2} \left\| \frac{1}{2\mu}\boldsymbol{\sigma}^{\mathtt{d}}_h - \nabla \mathbf{u}^\star_h \right\|_{0,T^e} \right) \|\boldsymbol{\mathscr{E}}(\boldsymbol{\zeta})\|_{1,\widetilde{T}^e_{aux}}.$$

The result follows from the continuity of the extension operator $\boldsymbol{\mathscr{E}}$, Assumption **(A1)** and the Cauchy–Schwarz inequality. □

Lemma 2.16. *Assume that* $\mathbf{g} \in \mathbf{H}^1(\Gamma)$. *Then, there holds*

$$|\mathcal{R}(\underline{\mathbf{curl}}\,(\boldsymbol{\varphi} - \boldsymbol{\varphi}_h))| \lesssim \left(\sum_{T\in\mathcal{T}_h} \Theta^2_{2,T} \right)^{1/2} \|\boldsymbol{\varphi}\|_{1,\Omega}, \tag{2.91}$$

where

$$\Theta_{2,T} := h_T^2 \left\| \mathrm{curl} \left\{ \frac{1}{2\mu} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\} \right\|_{0,T}^2 + \sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h^{\partial}} h_{T^e} \left\| \frac{d\mathbf{g}}{d\mathbf{t}} - \frac{1}{2\mu} \boldsymbol{\sigma}_{0,h}^{\mathtt{d}} \mathbf{t} \right\|_{0,\Gamma_e}^2 + \sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h^{i}} h_e \left\| \left[\!\!\left[\frac{1}{2\mu} \boldsymbol{\sigma}_h^{\mathtt{d}} \mathbf{t} \right]\!\!\right] \right\|_{0,e}^2 .$$

Proof. We follow [78, Lemma 4.3] and use the integration by parts formula, but more precisely the identities from [84, eq. (2.17) and Theorem 2.11], and the fact that $\underline{\mathbf{curl}}(\mathbf{v})\mathbf{n}_\Gamma = d\mathbf{v}/d\mathbf{t}$ for a sufficiently smooth vector-valued function $\mathbf{v}$, to obtain

$$\langle \underline{\mathbf{curl}}(\boldsymbol{\varphi} - \boldsymbol{\varphi}_h)\mathbf{n}_\Gamma, \mathbf{g} \rangle_\Gamma = - \sum_{e \in \mathcal{E}_h^{\partial}} \int_{\Gamma_e} \frac{d\mathbf{g}}{d\mathbf{t}} \cdot (\boldsymbol{\varphi} - \boldsymbol{\varphi}_h), \tag{2.92}$$

which holds true because $\mathbf{g} \in \mathbf{H}^1(\Gamma)$ has been assumed. Moreover, from $\mathcal{R}(\underline{\mathbf{curl}}(\boldsymbol{\varphi} - \boldsymbol{\varphi}_h))$, using the identity $\mathbf{div}\,(\underline{\mathbf{curl}}(\boldsymbol{\varphi} - \boldsymbol{\varphi}_h)) = \mathbf{0}$, applying [84, Theorem 2.11] to integrate by parts elementwise the integrals over D_h and D_h^c separately, and combining the resulting terms with (2.92), it follows that

$$\begin{aligned}
\mathcal{R}(\underline{\mathbf{curl}}(\boldsymbol{\varphi} - \boldsymbol{\varphi}_h)) &= - \sum_{e \in \mathcal{E}_h^{\partial}} \int_{\Gamma_e} \frac{d\mathbf{g}}{d\mathbf{t}} \cdot (\boldsymbol{\varphi} - \boldsymbol{\varphi}_h) - \frac{1}{2\mu} \int_\Omega \boldsymbol{\sigma}_h^{\mathtt{d}} : \underline{\mathbf{curl}}\,(\boldsymbol{\varphi} - \boldsymbol{\varphi}_h) \\
&= - \sum_{T \in \mathcal{T}_h} \int_T \mathrm{curl} \left\{ \frac{1}{2\mu} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\} \cdot (\boldsymbol{\varphi} - \boldsymbol{\varphi}_h) + \sum_{e \in \mathcal{E}_h^{i}} \int_e \left[\!\!\left[\frac{1}{2\mu} \boldsymbol{\sigma}_h^{\mathtt{d}} \mathbf{t} \right]\!\!\right] \cdot (\boldsymbol{\varphi} - \boldsymbol{\varphi}_h) \\
&\quad - \sum_{e \in \mathcal{E}_h^{\partial}} \left(\int_{\widetilde{T}_{ext}^e} \mathrm{curl} \left\{ \frac{1}{2\mu} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\} \cdot (\boldsymbol{\varphi} - \boldsymbol{\varphi}_h) + \int_{\Gamma_e} \left(\frac{d\mathbf{g}}{d\mathbf{t}} - \frac{1}{2\mu} \boldsymbol{\sigma}_h^{\mathtt{d}} \mathbf{t} \right) \cdot (\boldsymbol{\varphi} - \boldsymbol{\varphi}_h) \right).
\end{aligned}$$

Next, applying the Cauchy-Schwarz inequality to each term above, noting that similarly to (2.90), one has

$$\|\boldsymbol{\varphi} - \boldsymbol{\varphi}_h\|_{0,\widetilde{T}_{ext}^e} \lesssim h_{T^e} \|\boldsymbol{\mathcal{E}}(\boldsymbol{\varphi})\|_{1,\Delta(\widetilde{T}_{aux}^e)} \quad \forall\, e \in \mathcal{E}_h^{\partial},$$

using the extrapolation constant (2.27) in the same fashion as in the proof of Lemma 2.14, and making use of the approximation properties (2.71)-(2.72) and (2.85), we obtain

$$\begin{aligned}
|\mathcal{R}(\underline{\mathbf{curl}}(\boldsymbol{\varphi} - \boldsymbol{\varphi}_h))| &\lesssim \sum_{T \in \mathcal{T}_h} h_T \left\| \mathrm{curl} \left\{ \frac{1}{2\mu} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\} \right\|_{0,T} \|\boldsymbol{\varphi}\|_{1,\Delta(T)} + \sum_{e \in \mathcal{E}_h^{i}} h_e^{1/2} \left\| \left[\!\!\left[\frac{1}{2\mu} \boldsymbol{\sigma}_h^{\mathtt{d}} \mathbf{t} \right]\!\!\right] \right\| \|\boldsymbol{\varphi}\|_{1,\Delta(e)} \\
&\quad + \sum_{e \in \mathcal{E}_h^{\partial}} \left(\widetilde{C}_{ext}^e h_{T^e} (\widetilde{r}_e)^{1/2} \left\| \mathrm{curl} \left\{ \frac{1}{2\mu} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\} \right\|_{0,T^e} + (h_{T^e})^{1/2} \left\| \frac{d\mathbf{g}}{d\mathbf{t}} - \frac{1}{2\mu} \boldsymbol{\sigma}_h^{\mathtt{d}} \mathbf{t} \right\|_{0,\Gamma_e} \right) \|\boldsymbol{\mathscr{E}}(\boldsymbol{\varphi})\|_{1,\Delta(\widetilde{T}_{aux}^e)}.
\end{aligned}$$

Since the number of triangles in $\Delta(\widetilde{T}_{aux}^e)$, $\Delta(T)$ and $\Delta(e)$ is bounded (due to shape-regularity of $\mathcal{T}_h^*$), the proof ends by using the same arguments as in the proofs of the last two lemmas. □

Finally, from the identity (2.76), estimates (2.86), (2.89), and (2.91), and the stability of the Helmholtz decomposition (cf. (2.74)), we have

$$\|\mathcal{R}\|_{[\mathbb{H}_0(\mathbf{div};\Omega)]'} \lesssim \left(\sum_{T \in \mathcal{T}_h} \sum_{i=0}^{2} \Theta_{i,T}^2 \right)^{1/2}.$$

Combined with (2.69), this yields the reliability estimate (2.66), since $h_e \leq h_{T^e}$ for all $e \in \mathcal{E}_h^{\partial}$.

2.5.2 Quasi-efficiency of the *a posteriori* error estimator

In order to prove the quasi-efficiency of our estimator Θ, in what follows we derive suitable upper bounds for each term defining the local error indicator Θ_T defined in (2.63). In particular, we briefly discuss at the end of this section the situation of $\mathbb{B}$ (cf. (2.68)) involving the Dirichlet datum $\mathbf{g}$ and the postprocessed velocity $\mathbf{u}_h^\star$.

We first notice that, using $\mathbf{div}\,\boldsymbol{\sigma} = -\mathbf{f}$ in Ω (see Lemma 2.1), there holds

$$\|\mathbf{f} + \mathbf{div}\,\boldsymbol{\sigma}_h\|_{0,T}^2 = \|\mathbf{div}\,(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\|_{0,T}^2 \leq \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathbf{div},T}^2 \quad \forall\, T \in \mathcal{T}_h, \tag{2.93}$$

and similarly,

$$\|\mathbf{f} + \mathbf{div}\,\boldsymbol{\sigma}_h\|_{0,\widetilde{T}_{ext}^e}^2 \leq \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathbf{div},\widetilde{T}_{ext}^e}^2 \quad \forall\, e \in \mathcal{E}_h^\partial. \tag{2.94}$$

On the other hand, we have the following result for the terms involving the curl operator and the tangential jumps across the interior edges of $\mathcal{T}_h$.

Lemma 2.17. *There hold*

$$h_e \left\| \left[\!\!\left[\frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}}\mathbf{t} \right]\!\!\right] \right\|_{0,e}^2 \lesssim \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{0,\mathcal{K}(e)}^2 \quad \forall\, e \in \mathcal{E}_h^i, \tag{2.95}$$

and

$$h_T^2 \left\| \mathrm{curl}\left\{ \frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} \right\} \right\|_{0,T}^2 \lesssim \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{0,T}^2 \quad \forall\, T \in \mathcal{T}_h, \tag{2.96}$$

where $\mathcal{K}(e)$ is given by (2.88).

Proof. It follows by using similar arguments as in the proofs of Lemmas 6.3 and 6.4 in [49] (see also [17, Lemmas 4.3 and 4.4] or [78, Lemma 4.11]). We omit further details. □

Next, we exploit the properties of the postprocessed velocity $\mathbf{u}_h^\star$ (cf. (2.60)) and derive the local efficiency of $h_e^{-1}|[\![\mathbf{u}_h^\star]\!]\|_{0,e}^2$ for all $e \in \mathcal{E}_h^i$. In doing so, we follow here the approach of [62, Section 3.2]. Denoting by $\boldsymbol{\mathcal{P}}_h^0$ the $\mathbf{L}^2(\mathrm{D}_h)$-projection onto the piecewise constant functions on each edge, and then adding and subtracting a convenient term, we easily get

$$h_e^{-1}\|[\![\mathbf{u}_h^\star]\!]\|_{0,e}^2 \lesssim h_e^{-1}\|(\boldsymbol{I} - \boldsymbol{\mathcal{P}}_h^0)([\![\mathbf{u}_h^\star]\!])\|_{0,e}^2 + h_e^{-1}\|\boldsymbol{\mathcal{P}}_h^0([\![\mathbf{u}_h^\star]\!])\|_{0,e}^2, \tag{2.97}$$

where $\boldsymbol{I}$ denotes the identity operator. We will bound each term on the right-hand side of (2.97). The first of them is provided next.

Lemma 2.18. *For every edge $e \in \mathcal{E}_h^i$, we have*

$$h_e^{-1}\|(\boldsymbol{I} - \boldsymbol{\mathcal{P}}_h^0)[\![\mathbf{u}_h^\star]\!]\|_{0,e}^2 \lesssim \sum_{T \in \mathcal{K}(e)} \|\nabla(\mathbf{u} - \mathbf{u}_h^\star)\|_{0,T}^2. \tag{2.98}$$

Moreover, for each $T \in \mathcal{T}_h$, there holds

$$\left\| \frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla\mathbf{u}_h^\star \right\|_{0,T}^2 \lesssim \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{0,T}^2. \tag{2.99}$$

Proof. With minor modifications the result follows from the proofs of Lemmas 3.5 and 3.7 in [62]. □

The next result establishes an upper bound for the last term in (2.97). The proof is similar to the proof of Lemma 3.4 in [62], where the equations of the proposed hybridized Raviart-Thomas method and the posptocessed velocity are used to establish a relation between the residuals on elements and edges. For the sake of completeness and since we are not using hybrid-based methods, we include a detailed proof.

Lemma 2.19. *For each $e \in \mathcal{E}_h^i$, there holds*

$$h_e^{-1/2}\|\boldsymbol{\mathcal{P}}_h^0([\![\mathbf{u}_h^\star]\!])\|_{0,e} \lesssim \sum_{T\in\mathcal{K}(e)} \left\|\frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla\mathbf{u}_h^\star\right\|_{0,T}. \tag{2.100}$$

Proof. From the Galerkin scheme (2.16) and the equations defining the postprocessed velocity $\mathbf{u}_h^\star$ (cf. (2.60)), it is easy to check that

$$\begin{aligned}\sum_{e\in\mathcal{E}_h^\partial}\int_e \widetilde{\mathbf{g}}_h\cdot(\boldsymbol{\tau}_h\mathbf{n}_e) &= \int_{T\in\mathcal{T}_h}\left(\frac{1}{2\mu}\int_T \boldsymbol{\sigma}_h^{\mathtt{d}} : \boldsymbol{\tau}_h + \int_T \mathbf{u}_h\cdot\mathbf{div}\,\boldsymbol{\tau}_h\right)\\ &= \sum_{T\in\mathcal{T}_h}\left(\int_T\left(\frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla\mathbf{u}_h^\star\right) : \boldsymbol{\tau}_h + \int_{\partial T}\mathbf{u}_h^\star\cdot(\boldsymbol{\tau}_h\mathbf{n})\right)\end{aligned}$$

for all $\boldsymbol{\tau}_h$ in the space given by $\mathbb{H}_{0,h}(\mathrm{D}_h)$ with $k=0$ (cf. Section 2.3.3). After some algebraic manipulations, it yields

$$\sum_{T\in\mathcal{T}_h}\int_T\left(\frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla\mathbf{u}_h^\star\right) : \boldsymbol{\tau}_h = -\sum_{e\in\mathcal{E}_h^i}\int_e [\![\mathbf{u}_h^\star]\!]\cdot(\boldsymbol{\tau}_h\mathbf{n}_e) + \sum_{e\in\mathcal{E}_h^\partial}\int_e(\widetilde{\mathbf{g}}_h - \mathbf{u}_h^\star)\cdot(\boldsymbol{\tau}_h\mathbf{n}_e)\,. \tag{2.101}$$

In particular, taking $\boldsymbol{\tau}_h$ such that, for a given edge $e'\in\mathcal{E}_h^i$ and each $T\in\mathcal{K}(e')$,

$$\begin{aligned}\int_e \boldsymbol{\tau}_h\mathbf{n}_T &= \mathbf{0} && \forall\, e\in\mathcal{E}(T),\ e\neq e',\\ \int_{e'}\boldsymbol{\tau}_h\mathbf{n}_T &= \int_{e'}\boldsymbol{\mathcal{P}}_h^0([\![\mathbf{u}_h^\star]\!]) && \text{for the edge } e',\end{aligned}$$

and for all $T\in\mathcal{T}_h\setminus\mathcal{K}(e')$,

$$\int_e\boldsymbol{\tau}_h\mathbf{n}_T = \mathbf{0}\quad\forall\, e\in\mathcal{E}(T),$$

we have that $\boldsymbol{\tau}_h|_T\equiv\mathbf{0}$ for all $T\in\mathcal{T}_h\setminus\mathcal{K}(e')$, and then (2.101) gives

$$\sum_{T\in\mathcal{K}(e')}\int_T\left(\frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla\mathbf{u}_h^\star\right) : \boldsymbol{\tau}_h = \int_{e'}[\![\mathbf{u}_h^\star]\!]\cdot\boldsymbol{\mathcal{P}}_h^0([\![\mathbf{u}_h^\star]\!]) = \|\boldsymbol{\mathcal{P}}_h^0([\![\mathbf{u}_h^\star]\!])\|_{0,e'}^2.$$

Furthermore, applying the Cauchy–Schwarz inequality and observing that $\|\boldsymbol{\tau}_h\|_{0,T}\lesssim h_{e'}^{1/2}\|\boldsymbol{\tau}_h\mathbf{n}_{e'}\|_{0,e'}$ for all $T\in\mathcal{K}(e')$ (see, e.g., [62, Lemma A.1]), we obtain

$$\begin{aligned}\|\boldsymbol{\mathcal{P}}_h^0([\![\mathbf{u}_h^\star]\!])\|_{0,e'}^2 &\le \sum_{T\in\mathcal{K}(e')} h_{e'}^{1/2}\left\|\frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla\mathbf{u}_h^\star\right\|_{0,T}\|\boldsymbol{\tau}_h\mathbf{n}_{e'}\|_{0,e'}\\ &= \sum_{T\in\mathcal{K}(e')} h_{e'}^{1/2}\left\|\frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla\mathbf{u}_h^\star\right\|_{0,T}\|\boldsymbol{\mathcal{P}}_h^0([\![\mathbf{u}_h^\star]\!])\|_{0,e'}.\end{aligned}$$

Clearly, this implies the claimed result. □

Consequently, gathering (2.98) and (2.100) into (2.97) and employing estimate (2.99), we conclude that, for each edge $e \in \mathcal{E}_h^i$,

$$h_e^{-1}\|[\![\mathbf{u}_h^\star]\!]\|_{0,e}^2 \lesssim \sum_{T\in\mathcal{K}(e)} \|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{0,T}^2. \tag{2.102}$$

The following lemma deals with the corresponding upper bound for the estimator terms involving only the two velocity approximations.

Lemma 2.20. *For each $T \in \mathcal{T}_h$ and $h < 1$, there holds*

$$\|\mathbf{u}_h-\mathbf{u}_h^\star\|_{0,T}^2 \lesssim \|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{0,T}^2 + \|\mathbf{u}-\mathbf{u}_h\|_{0,T}^2. \tag{2.103}$$

Furthermore, for all $e \in \mathcal{E}_h^\partial$, we have

$$\|\mathbf{u}_h-\mathbf{u}_h^\star\|_{0,\widetilde{T}_{ext}^e}^2 \lesssim \|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{0,T^e}^2 + \|\mathbf{u}-\mathbf{u}_h\|_{0,T^e}^2. \tag{2.104}$$

Proof. Let $\mathbf{w}_h := \mathbf{u}_h - \mathbf{u}_h^\star$. Denoting by $\boldsymbol{\mathcal{P}}_T^0$ the $\mathbf{L}^2(T)$-projection onto $\mathbf{P}_0(T)$, we have $\boldsymbol{\mathcal{P}}_T^0(\mathbf{w}_h|_T) = \mathbf{0}$ for all $T \in \mathcal{T}_h$, since $\mathbf{u}_h^\star$ solves (2.60). Applying now the approximation property (2.21) with $k = 0$ and $l = 1$, and using the fact that $\boldsymbol{\mathcal{P}}_T^0(\mathbf{w}_h|_T) = \boldsymbol{\mathcal{P}}_h^0(\mathbf{w}_h)|_T$, we obtain

$$\|\mathbf{w}_h\|_{0,T} = \left\|\mathbf{w}_h - \boldsymbol{\mathcal{P}}_h^0(\mathbf{w}_h)\right\|_{0,T} \leq h_T|\mathbf{w}_h|_{1,T}.$$

Adding and subtracting $(2\mu)^{-1}\boldsymbol{\sigma}_h^{\mathtt{d}}$, applying the triangle inequality and assuming $h < 1$, it follows that

$$\|\mathbf{w}_h\|_{0,T}^2 \lesssim \left\|\frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla\mathbf{u}_h^\star\right\|_{0,T}^2 + h_T^2\left\|\frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla\mathbf{u}_h\right\|_{0,T}^2.$$

The first term on the right-hand side of the above inequality can be bounded using (2.99). Moreover, following the proof of [48, Lemma 6.3] (see also [78, Lemma 4.13]), we find

$$h_T^2\left\|\frac{1}{2\mu}\boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla\mathbf{u}_h\right\|_{0,T}^2 \lesssim \|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{0,T}^2 + \|\mathbf{u}-\mathbf{u}_h\|_{0,T}^2,$$

concluding (2.103).

On the other hand, using the equivalence of the norms $\|\cdot\|_{0,\widetilde{T}_{ext}^e}$ and $|||\cdot|||_{0,e}$ (cf. Lemma 1.6), and the extrapolation constant (2.27), we obtain

$$\|\mathbf{w}_h\|_{0,\widetilde{T}_{ext}^e} \lesssim (\widetilde{r}_e)^{1/2}\widetilde{C}_{ext}^e\|\mathbf{w}_h\|_{0,T^e},$$

which, combined with the estimate (2.103), implies (2.104) thanks to Assumption **(A1)**. □

Therefore, the quasi-efficiency property of the estimator Θ is a consequence of the upper bounds given by (2.93)-(2.96) and (2.102)-(2.104).

Having established (2.67), as already mentioned at the beginning of this section, the mayor issue is the convergence rate of $\mathbb{B}$ given by (2.68). If $\mathbf{g}$ were piecewise polynomial on a polygonal boundary Γ, it would be possible to apply the results given by Lemmas 4.14 and 4.15 in [78], which are based on standard tools including the usual localization technique of bubble functions and inverse inequalities,

to deduce that the convergence order of $\mathbb{B}$ is at least $\mathcal{O}(h^{k+1})$ owing to the approximation properties of the postprocessed velocity $\mathbf{u}_h^\star$. Otherwise, assuming that $\mathbf{g}$ is sufficiently smooth, the previous estimate is actually valid with possible further high order terms arising from Taylor approximations of the data. The extension of this idea to curved domains is an ongoing work. However, our numerical results below allow us to conjecture that $\mathbb{B}$ has the above mentioned optimal convergence property.

2.5.3 Extension of the estimator to more complicated geometries

When defining the computational boundary as in the previous section, it would be possible to have $\omega := \Omega^c \cap \mathrm{D}_h \neq \varnothing$. Indeed, this certainly happens if we consider nonconvex curved domains Ω, even though some regions having boundaries that are not completely curved, as for instance the pacman-shaped domain, could be the exception. Therefore, our intention here is to propose a way of extending the previous analysis to that situation. For this, we assume that the solution $(\boldsymbol{\sigma}, \mathbf{u})$ of (2.6) can be extended to ω, with $\boldsymbol{\sigma} \in \mathbb{H}(\mathbf{div}; \omega \cup \Omega)$, but not necessarily satisfying $\int_{\omega \cup \Omega} \mathrm{tr}\,(\boldsymbol{\sigma}) = 0$.

Now, since $\boldsymbol{\sigma}$ solves (2.6), which ensures that $\int_\Omega \mathrm{tr}\,(\boldsymbol{\sigma}) = 0$, we can write

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}_0 - \frac{1}{2|\mathrm{D}_h|}\left(\int_{\mathrm{D}_h^c} \mathrm{tr}\,(\boldsymbol{\sigma}) - \int_\omega \mathrm{tr}\,(\boldsymbol{\sigma})\right)\mathbb{I} \quad \text{in} \quad \mathrm{D}_h,$$

where $\boldsymbol{\sigma}_0 \in \mathbb{H}_0(\mathbf{div}; \mathrm{D}_h)$. Similarly, the tensor $\boldsymbol{\sigma}_h$ could be defined as in (2.35), by replacing γ_h in (2.37) by

$$\gamma_h := -\int_{\mathrm{D}_h^c} \mathrm{tr}\left(\mathbf{E}_h(\boldsymbol{\sigma}_{0,h}) - \frac{1}{2|\Omega|}\left(\int_{\mathrm{D}_h^c} \mathrm{tr}\,\left(\mathbf{E}_h(\boldsymbol{\sigma}_{0,h})\right) - \int_\omega \mathrm{tr}\,\left(\mathbf{E}_h(\boldsymbol{\sigma}_{0,h})\right)\right)\mathbb{I}\right), \tag{2.105}$$

from which we easily obtain that $\boldsymbol{\sigma}_h \in \mathbb{H}_0(\mathbf{div}; \Omega)$, provided $\boldsymbol{\sigma}_{0,h} \in \mathbb{H}_{0,h}(\mathrm{D}_h)$. As a consequence, the *a priori* error bounds in Section 2.4 are still valid on the larger region $\omega \cup \Omega$. Moreover, whenever $T^e \cap \omega \neq \varnothing$ we set $\widetilde{T}^e_{aux}$ in Section 2.5.1 to $\widetilde{T}^e_{aux} = T^e$. We then define the global a posteriori error estimator Θ as in (2.62), except that now $\boldsymbol{\sigma}_h$ is computed in terms of (2.105).

2.5.4 Extension of the estimator to three dimensions

We start by introducing additional notation. Given a sufficiently smooth vector field $\mathbf{v} := (v_i)_{1\leq i\leq 3}$, we set the differential operator

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

Furthermore, we take a tetrahedralization $\mathcal{T}_h$ of $\overline{\mathrm{D}_h}$ and consider the same notation as in the introduction of Section 2.5, but now replacing the word "edge" by "face". For any tensor $\boldsymbol{\tau} := (\tau_{ij})_{1\leq i,j\leq 3}$, we let $\mathrm{curl}\,(\boldsymbol{\tau})$ and $\boldsymbol{\tau} \times \mathbf{n}$ denote the tensors whose ith rows $(i = 1, 2, 3)$ are given by $\underline{\mathbf{curl}}\,(\tau_{i1}, \tau_{i2}, \tau_{i3})$ and $(\tau_{i1}, \tau_{i2}, \tau_{i3}) \times \mathbf{n}$, respectively. Given a face $e \in \mathcal{E}_h$, $\mathbf{v} \in \mathbf{L}^2(\Omega)$ and $\boldsymbol{\tau} \in \mathbb{L}^2(\Omega)$, such that $\mathbf{v}|_T \in [\mathcal{C}(T)]^3$ and $\boldsymbol{\tau}|_T \in [\mathcal{C}(T)]^{3\times 3}$ on each $T \in \mathcal{T}_h$, we let $[\![\mathbf{v}]\!]$ and $[\![\boldsymbol{\tau} \times \mathbf{n}]\!]$ be the corresponding jumps across e, that is, $[\![\mathbf{v}]\!] := (\mathbf{v}|_{T^+})\,|_e - (\mathbf{v}|_{T^-})\,|_e$ and $[\![\boldsymbol{\tau} \times \mathbf{n}]\!] := \{(\boldsymbol{\tau}|_{T^+})\,|_e - (\boldsymbol{\tau}|_{T^-})\,|_e\} \times \mathbf{n}$, respectively, where T^+ and T^- are two elements of $\mathcal{T}_h$ sharing a face e.

Now, the estimator term $\mathbb{K}_T$ reads

$$\mathbb{K}_T := \left(\sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h^{\partial}} h_{T^e} \left\| \left(\nabla \mathbf{u} - \frac{1}{2\mu} \boldsymbol{\sigma}_h^{\mathtt{d}} \right) \times \mathbf{n} \right\|_{0,\Gamma_e}^2 \right)^{1/2},$$

whereas $\mathbb{J}_T$ remains as in (2.64). We then set the global indicator as in the two-dimensional case, by replacing Θ_T^2 by

$$\begin{aligned} \Theta_T^2 := h_T^2 \left\| \mathrm{curl} \left\{ \frac{1}{2\mu} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\} \right\|_{0,T}^2 + \sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h^{i}} \left(h_e \left\| \left[\!\!\left[\frac{1}{2\mu} \boldsymbol{\sigma}_h^{\mathtt{d}} \times \mathbf{n} \right]\!\!\right] \right\|_{0,e}^2 + h_e^{-1} \| [\![\mathbf{u}_h^{\star}]\!] \|_{0,e}^2 \right) \\ + \left\| \frac{1}{2\mu} \boldsymbol{\sigma}_h^{\mathtt{d}} - \nabla \mathbf{u}_h^{\star} \right\|_{0,T}^2 + \| \mathbf{f} + \mathbf{div}\, \boldsymbol{\sigma}_h \|_{0,T}^2 + \sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h^{\partial}} \| \mathbf{f} + \mathbf{div}\, \boldsymbol{\sigma}_h \|_{0, \widetilde{T}_{ext}^e}^2 \\ + \| \mathbf{u}_h - \mathbf{u}_h^{\star} \|_{0,T}^2 + \sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h^{\partial}} \| \mathbf{u}_h - \mathbf{u}_h^{\star} \|_{0, \widetilde{T}_{ext}^e}^2 + \mathbb{J}_T^2 + \mathbb{K}_T^2. \end{aligned}$$

Let us briefly comment on the extension of Theorem 2.11 to three dimensions. First, notice that surface interpolation techniques allow to construct D_h satisfying $d(\Gamma, \Gamma_h) \lesssim h^2$. Then the three-dimensional analogues of Assumptions **(A1)-(A2)** and **(B1)-(B2)** of Sections 2.3.4 and 2.5.1, respectively, make sense for h small enough. In this case, all the results for the reliability estimate of Section 2.5.1 can be obtained, as in the two-dimensional case, from standard integration by parts formulae, the global inf-sup condition (2.7), local approximation properties of Clément and Raviart–Thomas interpolation operators, provided $\mathbb{H}_0(\mathbf{div}\,;\Omega)$ admits a stable Helmholtz decomposition. The latter is known to hold for arbitrary polyhedral regions and $\mathcal{C}^{1,1}$ domains; see [74] for further discussion. On the other hand, the quasi-efficiency estimate in three dimensions proceeds along the same lines as in Section 2.5.2.

2.6 Numerical results

We now present a series of numerical examples in two dimensions devised to illustrate the good performance of our discrete scheme (2.16), to validate the reliability and quasi-efficiency of the a posteriori error estimator Θ defined in (2.62), and to show the behavior of the associated adaptive algorithm. Our implementation is based on a MATLAB code along with the direct linear solver UMFPACK [64]. All our examples were carried out using the finite element spaces $\mathbb{H}_{0,h}(\mathrm{D}_h)$ and $\mathbf{Q}_h(\mathrm{D}_h)$ of Section 2.3.3 with $k = 0, \ldots, 3$. In turn, the condition $\int_{\mathrm{D}_h} \mathrm{tr}\,(\boldsymbol{\tau}_h) = 0$ for $\boldsymbol{\tau}_h \in \mathbb{H}_{0,h}(\mathrm{D}_h)$ was imposed as usual, that is, via a real Lagrange multiplier.

We emphasize that the error estimates presented in this work are independent of the construction of basis functions. For the numerical simulations, we consider *hierarchical basis* for the local Raviart–Thomas space of order k, as presented in [24], and the *Dubiner basis* (see, e.g., [66]) for the local polynomial space of degree less or equal to k.

In what follows, we denote by N the total number of elements defining the mesh $\mathcal{T}_h$ associated to the computational domain D_h. Denoting by $\mathbf{u}_h$ the solution of the problem (2.11), and by $\boldsymbol{\sigma}_h$, p_h and

$\mathbf{u}_h^\star$ the postprocessed solutions given by (2.35), (2.36) and (2.60), respectively, the individual errors are defined as

$$\mathrm{e}(\mathbf{u}) := \|\mathbf{u}-\mathbf{u}_h\|_{0,\Omega}, \quad \mathrm{e}^\star(\mathbf{u}) := \|\mathbf{u}-\mathbf{u}_h^\star\|_{0,\Omega},$$
$$\mathrm{e}(p) := \|p-p_h\|_{0,\Omega}, \quad \text{and} \quad \mathrm{e}(\boldsymbol{\sigma}) := \left(\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|^2_{\mathbf{div},\mathrm{D}_h} + \|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|^2_{\mathbf{div},\widetilde{\mathcal{T}}_h}\right)^{1/2},$$

where the approximations in D_h^c are those specified in Section 2.4.2. According to Theorem 2.11, the global error is computed as

$$\mathrm{e}(\boldsymbol{\sigma},\mathbf{u}) := \left(\mathrm{e}(\mathbf{u})^2 + \mathrm{e}(\boldsymbol{\sigma})^2\right)^{1/2},$$

whereas the quality of the *a posteriori* error estimator Θ is measured by using the effectivity index $\mathtt{eff}(\Theta) := \Theta/\mathrm{e}(\boldsymbol{\sigma},\mathbf{u})$. In order to explore the convergence properties of $\mathbb{B}$ defined in (2.68), we also introduce the estimator terms

$$\mathbb{J} := \left(\sum_{T\in\mathcal{T}_h} \mathbb{J}_T^2\right)^{1/2} \quad \text{and} \quad \mathbb{K} := \left(\sum_{T\in\mathcal{T}_h} \mathbb{K}_T^2\right)^{1/2},$$

where $\mathbb{J}_T$ and $\mathbb{K}_T$ are given by (2.64) and (2.65), respectively. In addition, suppose that e and e' are any of the above quantities for two consecutive meshes with N and N' number of elements, respectively. Then, by using the fact that $h \simeq N^{-1/2}$, we consider the experimental rate of convergence given by

$$\mathtt{rate} := -2[\log(\mathrm{e}/\mathrm{e}')/\log(N/N')].$$

The examples to be considered in this section are summarized in Table 2.1. For the examples that include adaptivity, we use the following algorithm:

1. Start with a coarse mesh $\mathcal{T}_h$ of $\overline{\mathrm{D}_h}$.
2. Solve the discrete problem (2.16) on the current mesh $\mathcal{T}_h$.
3. Compute Θ_T for each $T \in \mathcal{T}_h$.
4. Check the stopping criterion and decide whether to finish or go to next step.
5. Use *red-green-blue* procedure to refine each $T' \in \mathcal{T}_h$ satisfying:

$$\Theta_{T'} \geq 0.5 \max\{\Theta_T : \ T \in \mathcal{T}_h\}.$$

6. Project every new vertex $\mathbf{x}$ of Γ_h onto the closest point $\widetilde{\mathbf{x}}$ of Γ by using the transferring paths.
7. Define the resulting mesh as the current mesh $\mathcal{T}_h$, and go to step 2.

While Steps 1-5 are applied to refine polygonal meshes (see, e.g., [139]), the $6th$ step is added to improve the approximation of the curved boundary (see, e.g., [137]) and also to expect the Assumption **(A2)** of Section 2.3.4 to hold. In fact, without including the $6th$ step, the region D_h^c remains unchanged when updating $\mathcal{T}_h$.

Example 1. This test is aimed at evaluating the performance of the method when the computational boundary is as far from Γ as the theory allows. To that end, we consider the kidney-shaped domain

Example	$d(\Gamma,\Gamma_h)$	Exact solution	Ω	$\Omega^c \cap \mathrm{D}_h$	Adaptivity
1	$\mathcal{O}(h)$	smooth	nonconvex	$\varnothing$	no
2	$\mathcal{O}(h^2)$	smooth	convex	$\varnothing$	no
3	$\mathcal{O}(h^2)$	smooth	convex	$\varnothing$	yes
4	$\mathcal{O}(h^2)$	with a singularity	nonconvex	$\varnothing$	yes
5	$\mathcal{O}(h^2)$	with a singularity	nonconvex	$\varnothing$	yes
6*	$\mathcal{O}(h^2)$	smooth	nonconvex	$\neq \varnothing$	yes

Table 2.1: *It is carried out with the help of the considerations made in Section 2.5.3. (table produced by the author)

Ω whose boundary satisfies $\left(2\left[(x_1+0.5)^2+x_2^2\right]-x_1-0.5\right)^2-\left[(x_1+0.5)^2+x_2^2\right]+0.1=0$. In turn, we take the viscosity $\mu=1$, and $\mathbf{f}$ and $\mathbf{g}$ such that the solution of problem (2.5) is given by $\mathbf{u}:=(u_1,u_2)^T$, where $u_1(x_1,x_2):=-2x_2\sin(x_1)$ and $u_2(x_1,x_2):=(x_1^2+x_2^2)\cos(x_1)+2x_1\sin(x_1)$, and $p(x_1,x_2):=\sin(x_1^2+x_2^2)-p_0(x_1,x_2)$, where $p_0\in\mathbb{R}$ is chosen such that $p\in \mathrm{L}^2_0(\Omega)$. In practice, p_0 is computed numerically employing a extremely fine polygonal mesh approximating Ω. The precise construction of D_h is given next. Following [59, Section 2.1] or Section 1.2.1, we consider a uniform Cartesian background grid $\mathcal{B}_h$ of a square domain $\mathcal{B}$ such that $\Omega\subset\mathcal{B}$, and then set D_h as the union of all elements that are inside Ω; see an example in the left panel of Figure 2.3. Here, the index $h>0$, refers to the meshsize of $\mathcal{B}_h$. By construction, the distance $d(\Gamma_h,\Gamma)$ is only of order h, which increases the complexity for the implementation of the transferring paths. However, as we have already seen in Section 2.3.2, this task is reduced to find those paths associated to the vertices $\mathbf{p}_1$ and $\mathbf{p}_2$ of every edge $e\in\mathcal{E}_h^{\partial}$. To that end, we use the algorithm proposed in [59, Section 2.4.1] that uniquely determines a point $\tilde{\mathbf{p}}_i$ $(i=1,2)$ in Γ as the closest point to $\mathbf{p}_i$ such that $\mathscr{C}(\mathbf{p}_i)$ does not intersect any other path and does not intersect the interior of the domain D_h; computed paths are shown in the right panel of Figure 2.3.

In Table 2.2 we present the convergence history obtained for this example under a sequence of uniform triangulations of the background mesh detailed before. We observe there that the convergence rate predicted by Theorem 2.10, namely $\mathcal{O}(h^{k+1})$, is attained by $\mathrm{e}(\mathbf{u})$, $\mathrm{e}(\boldsymbol{\sigma})$ and $\mathrm{e}(p)$. In addition, the error $\mathrm{e}^\star(\mathbf{u})$ is clearly converging like $\mathcal{O}(h^{k+2})$, that is, it is superconvergent, which corresponds to the theoretical error bound (2.61) with $l=m=k+1$. On the other hand, the approximate pseudostress component $\sigma_{11,h}$ obtained with $N=654$ and $k=2$ is depicted in Figure 2.4. The good accuracy of the approximation suggests that Assumption **(A2)** (cf. Section 2.5) holds true, event though it is not entirely verifiable because some of the quantities involved cannot be calculable explicitly.

Example 2. Next, the accuracy of the proposed scheme (2.16) is tested under a sequence of quasi-uniform triangulations satisfying the hypotheses in Section 2.5. The main goal is to asses the properties of the posteriori error estimator Θ (cf. (2.62)) via the effectivity index $\mathrm{eff}(\Theta)$. We choose Ω as a disc centered at the origin with radius 2, the viscosity $\mu=1$ and the smooth solution to the problem (2.5) given by $\mathbf{u}:=(u_1,u_2)^T$, where $u_1(x_1,x_2):=\pi\cos(\pi x_2)\sin(\pi x_1)$ and $u_2(x_1,x_2):=\pi\cos(\pi x_1)\sin(\pi x_2)$, and $p(x_1,x_2):=x_2\exp(x_1)-p_0(x_1,x_2)$, with p_0 satisfying the same as that required by the previous example, in terms of which we define the corresponding source term $\mathbf{f}$ and the Dirichlet data $\mathbf{g}$. Let us now specify the domain D_h. Given $h>0$, let Γ_h be the computational boundary constructed through a piecewise linear interpolation of Γ, such that the length of each segment is of order h. We define D_h as the region enclosed by Γ_h and then set $\mathcal{T}_h$ as a quasi-uniform triangulation of D_h with meshsize

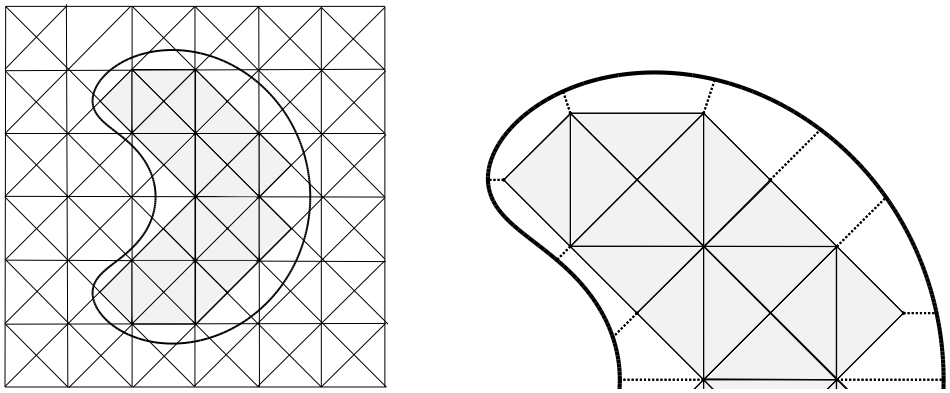


Figure 2.3: Left: the domain Ω defined in Example 1, its boundary Γ (solid line), the first background mesh $\mathcal{B}_h$ under consideration, and corresponding computational domain D_h. Right: computed transferring paths (dotted lines) associated to the vertices of the computational boundary; they were obtained by using the algorithm introduced in [59, Section 2.4.1]. (figure produced by the author)

k	N	h	d.o.f	$\mathrm{e}(\mathbf{u})$	rate	$\mathrm{e}^\star(\mathbf{u})$	rate	$\mathrm{e}(\boldsymbol{\sigma})$	rate	$\mathrm{e}(p)$	rate
0	28	0.263	159	$3.98e-02$	–	$2.02e-02$	–	$8.40e-01$	–	$3.09e-01$	–
	146	0.131	769	$2.50e-02$	0.56	$3.87e-03$	2.00	$3.81e-01$	0.96	$1.21e-01$	1.14
	654	0.066	3355	$1.39e-02$	0.78	$8.99e-04$	1.95	$1.76e-01$	1.03	$5.16e-02$	1.14
	3068	0.031	15497	$6.90e-03$	0.90	$2.88e-04$	1.47	$7.57e-02$	1.10	$2.02e-02$	1.21
	12579	0.016	63205	$3.52e-03$	0.96	$7.66e-05$	1.88	$3.63e-02$	1.04	$9.28e-03$	1.10
	50877	0.008	255007	$1.78e-03$	0.98	$1.90e-05$	1.99	$1.77e-02$	1.03	$4.51e-03$	1.03
1	28	0.263	485	$5.23e-03$	–	$2.06e-03$	–	$2.08e-01$	–	$1.09e-01$	–
	146	0.131	2413	$1.42e-03$	1.58	$1.80e-04$	2.95	$2.64e-02$	2.50	$7.78e-03$	3.20
	654	0.066	10633	$3.70e-04$	1.79	$4.22e-05$	1.94	$7.05e-03$	1.76	$2.88e-03$	1.33
	3068	0.031	49401	$9.54e-05$	1.75	$2.52e-06$	3.65	$1.20e-03$	2.29	$4.18e-04$	2.50
	12579	0.016	201883	$2.40e-05$	1.95	$2.65e-07$	3.19	$2.58e-04$	2.18	$7.49e-05$	2.44
	50877	0.008	815275	$6.04e-06$	1.98	$2.57e-08$	3.34	$5.35e-05$	2.25	$1.21e-05$	2.61
2	28	0.263	979	$3.15e-04$	–	$3.09e-04$	–	$2.13e-02$	–	$1.41e-02$	–
	146	0.131	4933	$1.48e-05$	3.70	$1.29e-05$	3.85	$1.41e-03$	3.29	$8.30e-04$	3.43
	654	0.066	21835	$6.52e-06$	1.10	$6.03e-06$	1.02	$1.03e-03$	0.42	$6.89e-04$	0.25
	3068	0.031	101713	$1.40e-07$	4.97	$6.56e-08$	5.85	$1.40e-05$	5.56	$8.44e-06$	5.70
	12579	0.016	416035	$1.62e-08$	3.06	$3.48e-09$	4.16	$1.35e-06$	3.31	$7.88e-07$	3.36
	50877	0.008	1680805	$2.02e-09$	2.98	$2.04e-10$	4.06	$1.03e-07$	3.69	$5.63e-08$	3.78
3	28	0.263	1641	$6.78e-05$	–	$6.74e-05$	–	$5.78e-03$	–	$3.88e-03$	–
	146	0.131	8329	$1.68e-06$	4.48	$1.66e-06$	4.49	$1.73e-04$	4.25	$1.11e-04$	4.31
	654	0.066	36961	$1.35e-07$	3.36	$1.35e-07$	3.35	$2.58e-05$	2.54	$1.67e-05$	2.52
	3068	0.031	172433	$1.27e-09$	6.04	$3.97e-10$	7.54	$1.77e-07$	6.45	$1.06e-07$	6.55
	12579	0.016	705661	$7.68e-11$	3.97	$1.18e-11$	4.99	$8.33e-09$	4.33	$4.69e-09$	4.42
	50877	0.008	2851597	$4.77e-12$	3.98	$2.86e-13$	5.32	$3.46e-10$	4.55	$2.01e-10$	4.51

Table 2.2: Example 1: History of convergence of the individual errors under a uniform refinement. (table produced by the author)

Figure 2.4: Example 1: Approximate pseudostress component $\sigma_{11,h}$ obtained with $N = 654$ and $k = 2$. (figure produced by the author)

h. The transferring paths associated to the interior points of a boundary edge e can be chosen so that they are perpendicular to e, we have $d(\Gamma, \Gamma_h) = \mathcal{O}(h^2)$ and actually the assumptions of Section 2.3.4 hold for h small enough. Also, all the geometrical hypotheses required by the a posteriori error analysis (cf. Section 2.5) are satisfied.

The results reported in Table 2.3 are in accordance with the theoretical bounds established in (2.61) and Theorem 2.10. In addition, from Table 2.4, we can conclude that both estimator terms $\mathbb{J}$ and $\mathbb{K}$ yield a convergence $\mathcal{O}(h^{k+3/2})$, which, together with the fact that, for each $k \in \{0, 1, 2, 3\}$, the effectivity index $\mathrm{eff}(\Theta)$ remains bounded, verifies not only the reliability of the a posteriori error estimator Θ, but also suggests its efficiency. In turn, the effectivity index increases as k does, which is not surprising since, according to Theorem 2.11, the reliability constant depends on the polynomial degree, and more specifically on the extrapolation constant defined in (2.27).

Example 3. We set the fluid domain Ω, the computational domain D_h, the transferring paths and the viscosity as in the previous example. However, this time, the manufactured exact solution adopts the form

$$\mathbf{u}(x_1, x_2) := \begin{pmatrix} x_1 \sin(x_2) - \sin(x_1) \\ \cos(x_2) + x_2 \cos(x_1) \end{pmatrix} \quad \text{and} \quad p(x, y) := \frac{1}{[x_1^2 + x_2^2 - 2.05^2]} - p_0(x_1, x_2),$$

with $p_0 \in \mathbb{R}$ being chosen as before. Notice that p has high gradients near the boundary Γ and thus, in addition to the accuracy of the method, we now asses the performance of the a posteriori error estimator Θ by using both quasi-uniform and adaptive refinement strategies. In Figure 2.5, we display the total error decay with respect to the total number of elements using both refinement strategies and different polynomial degrees. In all cases, the errors of the adaptive refinement are considerably smaller than the quasi-uniform ones considering the same number of elements $N > 500$, and it is also able to achieve the optimal convergence order for the total error $\mathrm{e}(\boldsymbol{\sigma}, \mathbf{u})$, namely $\mathcal{O}(h^{k+1})$. Some snapshots of the adapted meshes obtained with $k = 0$ and $k = 2$ are depicted in Figure 2.6, and it is concluded from there that the adaptive procedure is marking where is needed. Moreover, it is clear that the case $k = 2$ produces a very accurate approximate pseudostress component $\sigma_{22,h}$ with a considerable less number of triangles than its counterpart of lowest order.

Example 4. The next example is on the pacman-shaped domain

$$\Omega := \left\{ (x_1, x_2) \in \mathbb{R}^2 : \ x_1^2 + x_2^2 < 1 \right\} \setminus]0, 1[\times]-1, 0[,$$

k	N	d.o.f	$\mathrm{e}(\mathbf{u})$	rate	$\mathrm{e}^\star(\mathbf{u})$	rate	$\mathrm{e}(\boldsymbol{\sigma})$	rate	$\mathrm{e}(p)$	rate
0	36	191	$5.82e+00$	–	$4.89e+00$	–	$2.31e+02$	–	$2.15e+01$	–
	138	721	$3.51e+00$	0.75	$1.55e+00$	1.72	$1.40e+02$	0.75	$1.23e+01$	0.83
	528	2721	$1.78e+00$	1.01	$4.31e-01$	1.90	$7.20e+01$	0.99	$6.15e+00$	1.03
	2120	10713	$8.86e-01$	1.01	$1.01e-01$	2.09	$3.59e+01$	1.00	$3.03e+00$	1.02
	8696	43737	$4.43e-01$	0.98	$2.49e-02$	1.99	$1.79e+01$	0.98	$1.53e+00$	0.97
	34612	173573	$2.21e-01$	1.00	$6.30e-03$	1.99	$8.98e+00$	1.00	$7.63e-01$	1.01
1	36	597	$3.30e+00$	–	$1.69e+00$	–	$1.27e+02$	–	$1.19e+01$	–
	138	2269	$8.85e-01$	1.96	$2.12e-01$	3.09	$3.40e+01$	1.96	$4.41e+00$	1.48
	528	8609	$2.46e-01$	1.91	$2.88e-02$	2.98	$9.98e+00$	1.83	$1.16e+00$	1.99
	2120	34145	$5.86e-02$	2.06	$3.34e-03$	3.10	$2.41e+00$	2.04	$2.75e-01$	2.07
	8696	139649	$1.44e-02$	1.99	$4.12e-04$	2.97	$5.95e-01$	1.98	$6.85e-02$	1.97
	34612	554817	$3.60e-03$	2.00	$5.12e-05$	3.02	$1.49e-01$	2.00	$1.71e-02$	2.01
2	36	1219	$1.08e+00$	–	$4.44e-01$	–	$4.51e+01$	–	$5.66e+00$	–
	138	4645	$1.67e-01$	2.78	$2.10e-02$	4.54	$6.73e+00$	2.83	$7.78e-01$	2.95
	528	17665	$2.17e-02$	3.04	$1.66e-03$	3.78	$8.96e-01$	3.01	$1.15e-01$	2.85
	2120	70297	$2.61e-03$	3.05	$8.91e-05$	4.21	$1.08e-01$	3.04	$1.36e-02$	3.07
	8696	287737	$3.29e-04$	2.93	$5.46e-06$	3.96	$1.36e-02$	2.94	$1.65e-03$	2.99
	34612	1143733	$4.10e-05$	3.02	$3.39e-07$	4.02	$1.70e-03$	3.02	$2.07e-04$	3.00
3	36	2057	$3.08e-01$	–	$1.43e-01$	–	$1.41e+01$	–	$1.98e+00$	–
	138	7849	$2.09e-02$	4.01	$3.43e-03$	5.56	$8.69e-01$	4.15	$1.26e-01$	4.10
	528	29889	$1.89e-03$	3.58	$1.24e-04$	4.94	$7.68e-02$	3.62	$8.06e-03$	4.10
	2120	119169	$1.01e-04$	4.22	$3.64e-06$	5.08	$4.13e-03$	4.21	$4.57e-04$	4.13
	8696	488001	$6.34e-06$	3.92	$1.19e-07$	4.85	$2.61e-04$	3.91	$2.95e-05$	3.88
	34612	1940321	$3.93e-07$	4.03	$3.66e-09$	5.04	$1.61e-05$	4.03	$1.80e-06$	4.05

Table 2.3: Example 2: History of convergence of the individual errors with under a quasi-uniform refinement. (table produced by the author)

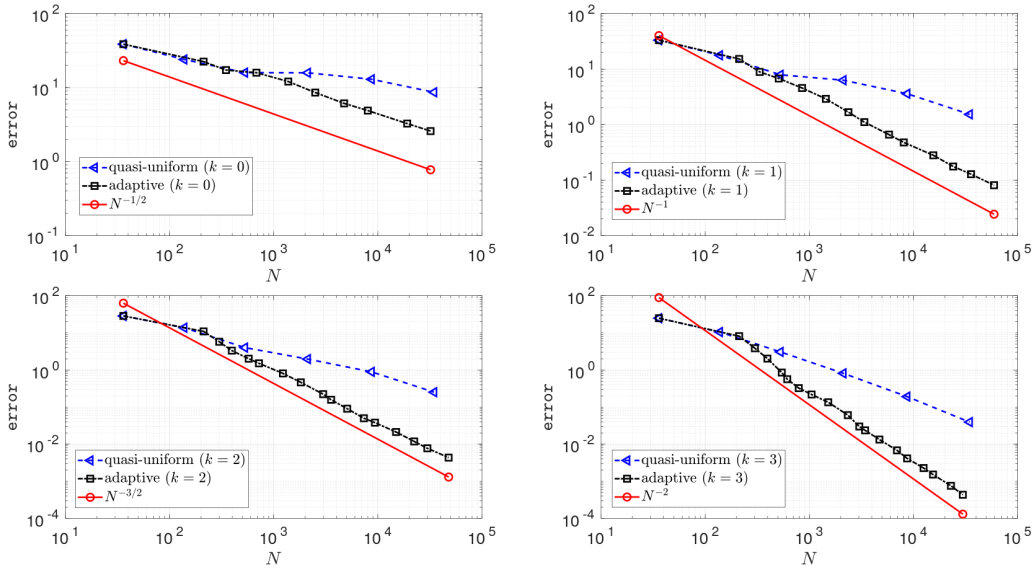


Figure 2.5: Example 3: Log-log plot of $\mathrm{e}(\boldsymbol{\sigma}, \mathbf{u})$ vs N for both refinement strategies and $k = 0, \ldots, 3$. (figure produced by the author)

k	N	d.o.f	$\mathbb{J}$	rate	$\mathbb{K}$	rate	$\mathrm{e}(\boldsymbol{\sigma},\mathbf{u})$	rate	Θ	rate	eff(Θ)
0	36	191	$6.23e+00$	–	$2.51e+01$	–	$2.31e+02$	–	$2.26e+02$	–	0.981
	138	721	$2.20e+00$	1.55	$1.02e+01$	1.34	$1.40e+02$	0.75	$1.43e+02$	0.69	1.021
	528	2721	$7.30e-01$	1.65	$4.76e+00$	1.14	$7.20e+01$	0.99	$7.42e+01$	0.97	1.031
	2120	10713	$4.00e-01$	0.87	$1.90e+00$	1.32	$3.59e+01$	1.00	$3.70e+01$	1.00	1.032
	8696	43737	$1.23e-01$	1.67	$6.63e-01$	1.49	$1.80e+01$	0.98	$1.85e+01$	0.98	1.031
	34612	173573	$4.38e-02$	1.50	$2.36e-01$	1.50	$8.98e+00$	1.00	$9.26e+00$	1.00	1.031
1	36	597	$2.13e+00$	–	$4.89e+01$	–	$1.27e+02$	–	$1.86e+02$	–	1.468
	138	2269	$4.45e-01$	2.33	$5.66e+00$	3.21	$3.40e+01$	1.96	$5.73e+01$	1.75	1.687
	528	8609	$6.60e-02$	2.84	$8.89e-01$	2.76	$9.98e+00$	1.83	$1.60e+01$	1.91	1.599
	2120	34145	$1.51e-02$	2.12	$1.81e-01$	2.29	$2.41e+00$	2.04	$3.88e+00$	2.04	1.607
	8696	139649	$2.20e-03$	2.73	$2.79e-02$	2.65	$5.95e-01$	1.98	$9.96e-01$	1.93	1.673
	34612	554817	$4.08e-04$	2.44	$5.26e-03$	2.42	$1.49e-01$	2.00	$2.47e-01$	2.02	1.657
2	36	1219	$7.90e-01$	–	$2.83e+01$	–	$4.51e+01$	–	$9.32e+01$	–	2.067
	138	4645	$4.44e-02$	4.28	$1.01e+00$	4.95	$6.73e+00$	2.83	$1.29e+01$	2.95	1.911
	528	17665	$2.94e-03$	4.05	$9.89e-02$	3.47	$8.96e-01$	3.01	$2.17e+00$	2.65	2.418
	2120	70297	$3.98e-04$	2.88	$1.19e-02$	3.05	$1.08e-01$	3.04	$2.47e-01$	3.12	2.287
	8696	287737	$2.75e-05$	3.79	$8.85e-04$	3.68	$1.36e-02$	2.94	$3.11e-02$	2.94	2.279
	34612	1143733	$2.48e-06$	3.48	$7.87e-05$	3.50	$1.70e-03$	3.02	$3.88e-03$	3.01	2.285
3	36	2057	$1.98e-01$	–	$9.05e+00$	–	$1.41e+01$	–	$3.08e+01$	–	2.175
	138	7849	$5.06e-03$	5.46	$1.93e-01$	5.73	$8.70e-01$	4.15	$2.76e+00$	3.59	3.176
	528	29889	$1.56e-04$	5.18	$6.55e-03$	5.05	$7.68e-02$	3.62	$1.99e-01$	3.92	2.594
	2120	119169	$1.05e-05$	3.88	$4.38e-04$	3.89	$4.13e-03$	4.21	$1.13e-02$	4.13	2.729
	8696	488001	$2.83e-07$	5.12	$1.46e-05$	4.82	$2.61e-04$	3.91	$7.61e-04$	3.82	2.920
	34612	1940321	$1.30e-08$	4.46	$7.46e-07$	4.31	$1.61e-05$	4.03	$4.61e-05$	4.06	2.855

Table 2.4: Example 2: History of convergence of some estimator terms and the total error under a quasi-uniform refinement. (table produced by the author)

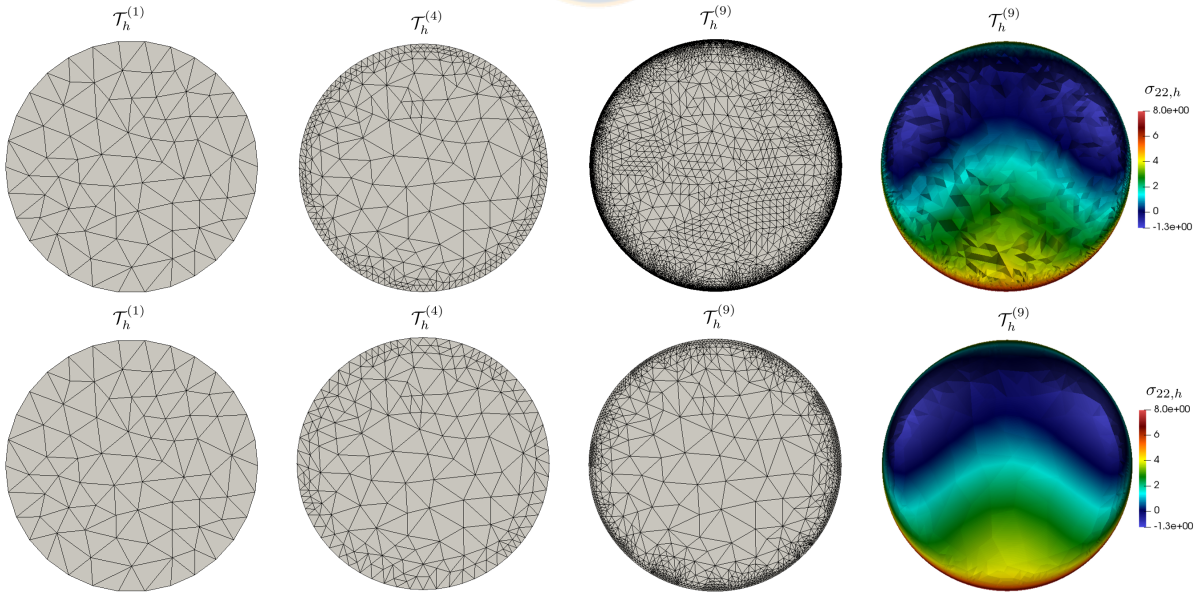


Figure 2.6: Example 3: Initial mesh and two adapted meshes according to the residual-based a posteriori error estimator Θ with $k=0$ (first row) and $k=2$ (second row), and comparative view of the approximate pseudostress component $\sigma_{22,h}$ obtained at the $9th$ iteration. (figure produced by the author)

with $\mu = 1/2$ and the manufactured exact solution given, in polar coordinates, by $p(r,\lambda) := 0$ and $\mathbf{u} := (u_1, u_2)^{\mathtt{t}}$, where $u_1(r,\lambda) := r^{2/3}\sin(2\lambda/3)$ and $u_2(r,\lambda) := r^{2/3}\cos(2\lambda/3)$, satisfying $\mathbf{f} = \mathbf{0}$. We notice that the partial derivatives of u_1 and u_2 have a singularity at the origin, and then a convergence of $\mathcal{O}(h^{2/3-\delta})$ ($\delta > 0$) should be expected from Theorem 2.10. The construction of the domain D_h and transferring paths are the same as that indicated in the last two examples. We point out that, since the nonconvex part of $\overline{\Omega}$ is only including the straight segments $\{0\}\times]-1,0[$ and $]0,1[\times\{0\}$, on which the boundaries Γ_h and Γ coincide, the requirement $\mathrm{D}_h \subset \Omega$ holds true. Furthermore, note that the singularity at the origin is not a concern in our numerical implementation, because the partial derivatives of u_1 and u_2 are only used to compute the error $\mathrm{e}(\boldsymbol{\sigma})$ and none of the quadrature points falls on the computational boundary.

In Table 2.5 we report the convergence history of the total error for $k = 0$ using a quasi-uniform refinement strategy, where the total error is converging like $\mathcal{O}(h^{2/3})$, as expected. In turn, it can be observed from Figure 2.7 that in all cases the adaptive algorithm reduces significantly the magnitude of the total error and also restores the optimal convergence order. Again, very accurate approximations are obtained with a few elements when the polynomial degree is increased as Figure 2.8 shows.

k	N	d.o.f	$\mathrm{e}(\boldsymbol{\sigma},\mathbf{u})$	r
0	65	349	2.62e-01	–
	257	1331	1.58e-01	0.73
	1037	5273	1.01e-01	0.65
	4143	20873	6.32e-02	0.67
	16583	83333	4.08e-02	0.63

Table 2.5: Example 4: History of convergence of the total error under a quasi-uniform refinement strategy. (table produced by the author)

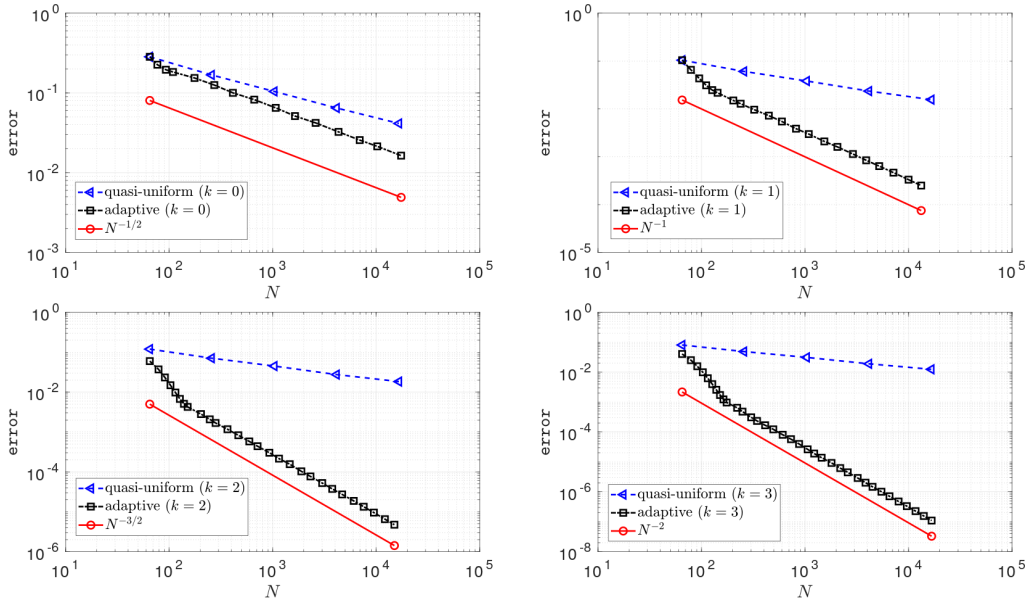


Figure 2.7: Example 4: Log-log plot of $\mathrm{e}(\boldsymbol{\sigma},\mathbf{u})$ vs N for both refinement strategies and $k = 0,\ldots,3$. (figure produced by the author)

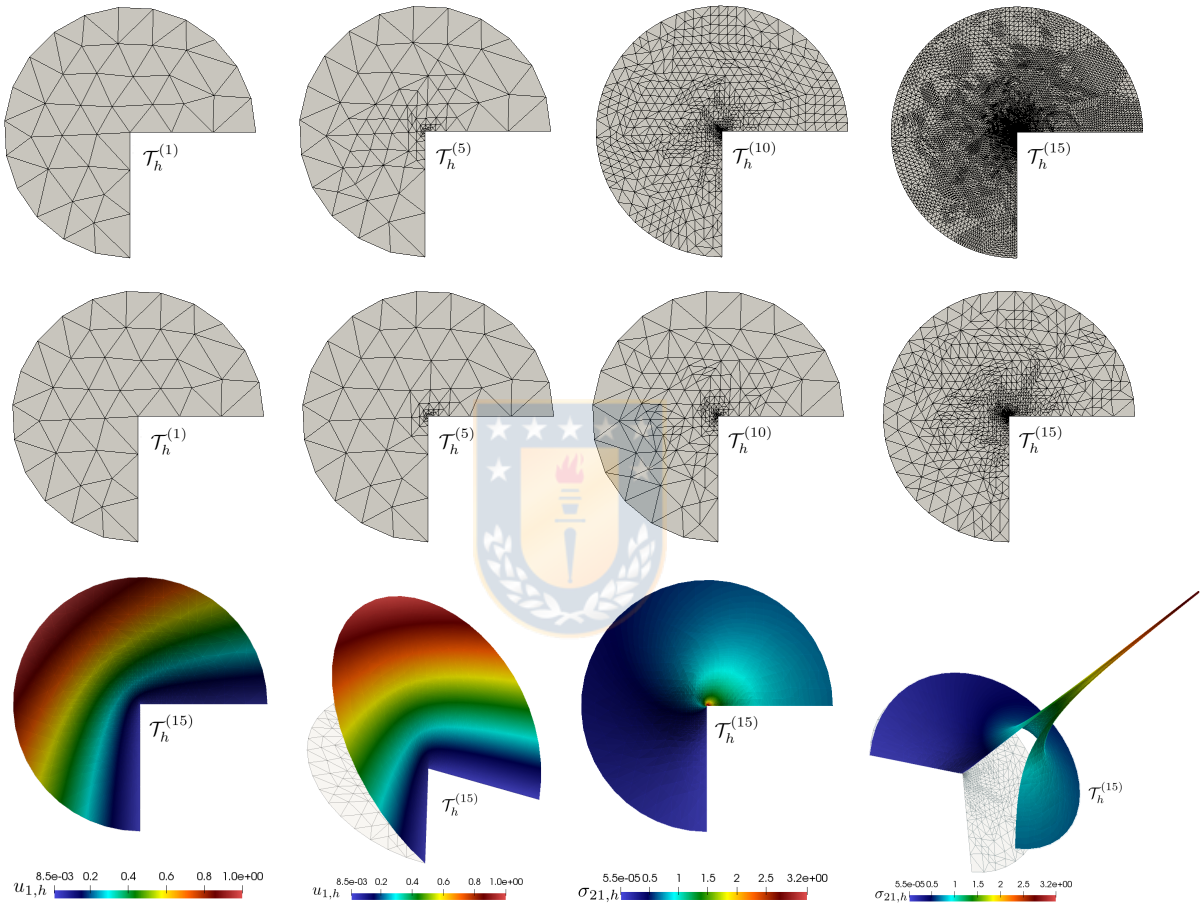


Figure 2.8: Example 4: Initial mesh and three adapted meshes according to residual-based a posteriori error estimator Θ with $k = 0$ (first row) and $k = 1$ (second row), approximate velocity component $u_{1,h}$ and approximate pseudostress component $\sigma_{21,h}$ obtained at the $15th$ iteration with $k = 1$ and $N = 2055$ (third row). (figure produced by the author)

Example 5. We are now interested in evaluating the capability of the *a posteriori* error estimator of detecting a singularity on a fully curved domain. We consider exactly the same manufactured solution and model parameters of Example 4 in order to enforce a singularity of the partial derivatives of u_1 and u_2 when Ω is chosen as the circle of center $(-0.5, 0)$ and radius 0.5. Figure 2.9 shows the initial coarse mesh we used in the adaptive algorithm. Notice that the singularity at the origin is not intersecting the initial Γ_h, as it did in the previous example.

Two snapshots of adapted meshes obtained with $k = 1$ are shown in Figure 2.10, where it is clear that the adaptive algorithm is marking near the singularity on Γ. Moreover, in Figure 2.10 we report the convergence history of the total error in the case $k = 1$, showing that the adaptive procedure reduces the magnitude of $\mathrm{e}(\boldsymbol{\sigma}, \mathbf{u})$ with the expected optimal convergence of $\mathcal{O}(h^2)$.

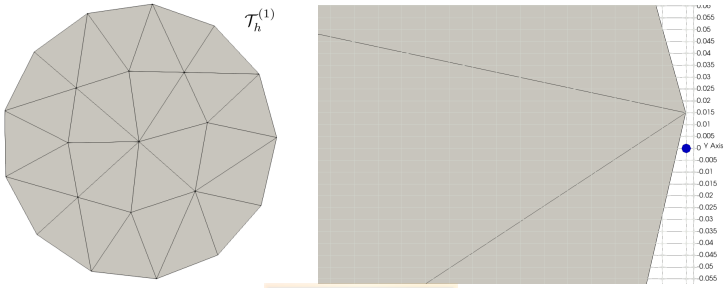


Figure 2.9: Example 5: Left, initial mesh and right, part of this mesh near the blue point at the origin. (figure produced by the author)

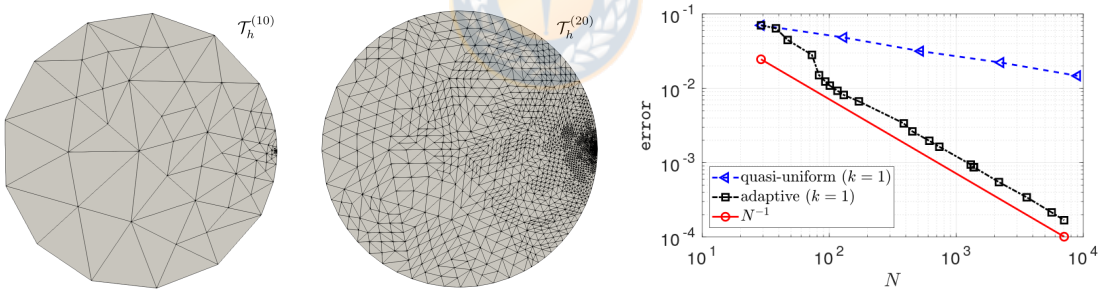


Figure 2.10: Example 5: Two adapted meshes according to the residual-based *a posteriori* error estimator Θ with $k = 1$, and log-log plot of $\mathrm{e}(\boldsymbol{\sigma}, \mathbf{u})$ vs N for both refinement strategies and $k = 1$. (figure produced by the author)

Example 6. To conclude, we choose Ω to be the annular domain consisting in two concentric circles of radius 0.5 and 2, respectively. Here the computational boundary is also constructed through a piecewise linear interpolation of Γ, implying $\omega := \Omega^c \cap \mathrm{D}_h \neq \varnothing$. In order to asses the accuracy of the Galerkin scheme (2.16) we use both quasi-uniform and adaptive refinement strategies and adopt the considerations made in Section 2.5.3. To that end, we take $\mu = 1$ and the manufactured exact solution as follows: $\mathbf{u} := (u_1, u_2)^T$, where $u_1(x_1, x_2) := \frac{x_2}{[x_1^2+x_2^2-2.2^2]} - \pi \cos(\pi x_2) \sin(\pi x_1)$ and $u_2(x_1, x_2) := -\frac{x_1}{[x_1^2+x_2^2-2.2^2]} + \pi \cos(\pi x_1) \sin(\pi x_2)$, and $p(x_1, x_2) = \frac{1}{[\exp(x_1^2+x_2^2-0.45^2)-1]} - p_0(x_1, x_2)$, with $p_0 \in \mathbb{R}$ being chosen so that $p \in \mathrm{L}_0^2(\Omega)$. As a result, the fluid pressure has high gradients near the boundary of the circle of radius 0.5, whereas the components of the fluid velocity have high gradients near the circle of radius 2.

The decay of the total error with respect to the total number of elements using both refinement strategies is depicted in Figure 2.11. In all cases, although the adaptive procedure is able to recognize the regions where there exist high gradients of the solution, the error convergence is oscillatory for small values of N, which could be explained by the fact that the region ω is too big when starting the mesh refinement process as shown in Figure 2.12. After that, the adaptive refinement strategy is much superior that the quasi-uniform one because it reduces the magnitude of the total error with optimal convergence of $\mathcal{O}(h^{k+1})$. We also present in Figure 2.12 the approximate velocity component $u_{1,h}$ and the approximate pressure p_h obtained with the adaptive procedure and $k = 2$.

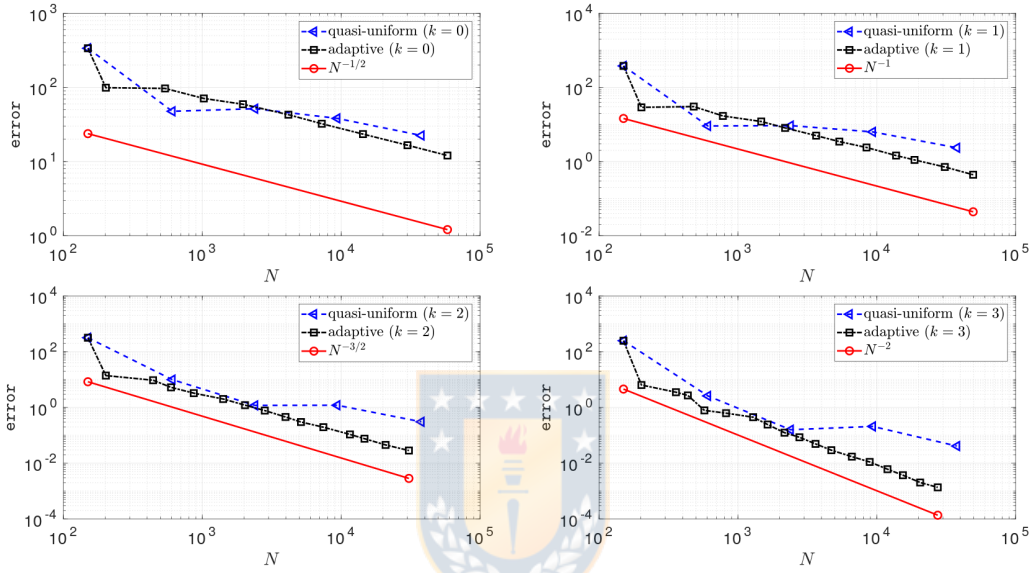


Figure 2.11: Example 6: Log-log plot of $\mathrm{e}(\boldsymbol{\sigma}, \mathbf{u})$ vs N for both refinement strategies and $k = 0, \ldots, 3$. (figure produced by the author)

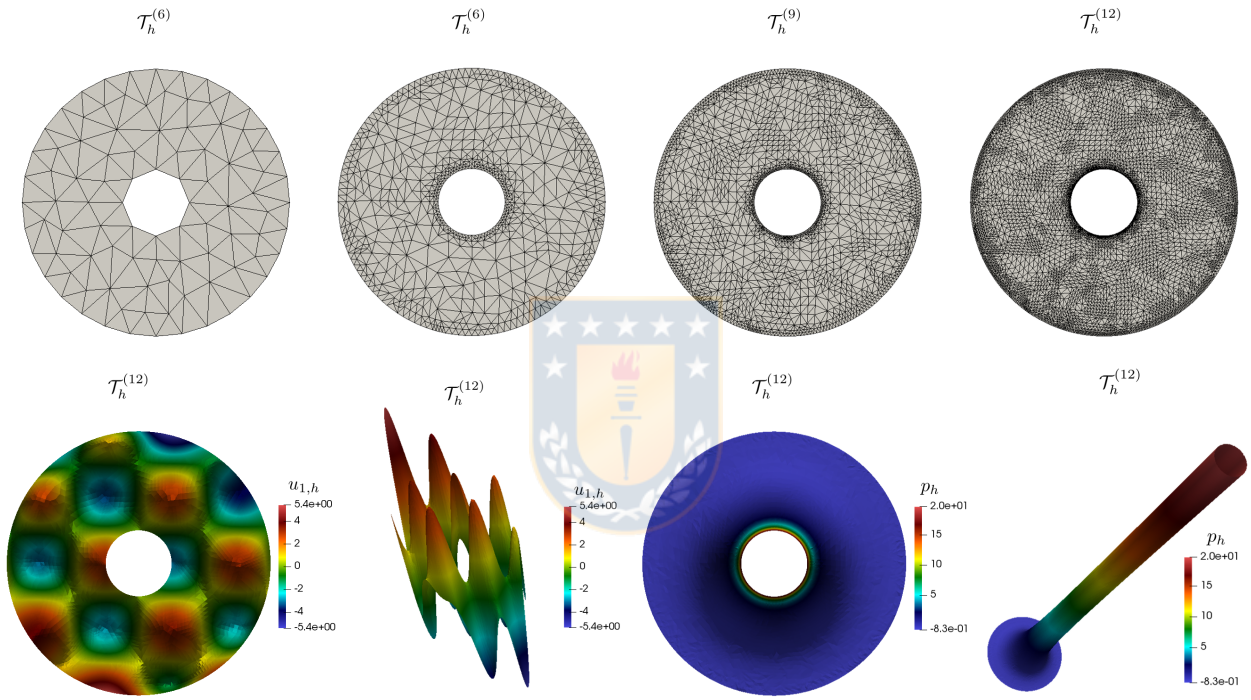


Figure 2.12: Example 6: Initial mesh and three adapted meshes according to the residual-based a posteriori error estimator Θ with $k = 2$ (first row), approximate velocity component $u_{1,h}$ and approximate pressure p_h obtained at the $12th$ iteration with $k = 2$ and $N = 11571$ (second row). (figure produced by the author)

CHAPTER 3

Error analysis of a conforming and locking-free four-field formulation for the stationary Biot's model

In this chapter we present an *a priori* and *a posteriori* error analysis of a conforming finite element method for a four-field formulation of the steady-state Biot's consolidation model. For the *a priori* error analysis we provide suitable hypotheses on the corresponding finite dimensional subspaces ensuring that the associated Galerkin scheme is well-posed. Next, we develop a reliable and efficient residual-based *a posteriori* error estimator. Both the reliability and efficiency estimates are shown to be independent of the modulus of dilatation. Numerical examples in 2D and 3D verify our analysis and illustrate the performance of the proposed *a posteriori* error indicator.

3.1 Introduction

Linear poroelasticity refers to fluid-structure interaction of an elastic solid infiltrated by an interconnected network of fluid-saturated pores. The modeling equations can be traced back to the pioneering theory of soil consolidation by Terzaghi [133] and Biot [25, 26], in which Darcy's law for the motion of a fluid is coupled to Hooke's theory of linear elasticity for the solid deformation. Advances in the understanding of the mechanical and physical aspects of Biot's consolidation model are of key importance in many applications. For instance, it has been used to predict the mechanics of groundwater withdrawals [91], earthquake fault zones [144], CO_2 sequestration [142] and biological systems (brain [19, 98], bones [71], arteries [89], intestines [148], etc.).

There is an extensive literature on theoretical results for this problem. A well-accepted mathematical analysis of existence, uniqueness, and regularity of the solution for the displacement-pressure formulation of Biot's model was carried out by Showalter [123, 124]. Moreover, many different numerical schemes have been proposed for this formulation with varying success, e.g., [21, 51, 72, 104, 113, 120, 143, 144, 145, 146, 147] and references therein. The main difficulties encountered when developing numerical methods for this model are volumetric locking and spurious, nonphysical pressure oscillations. While volumetric locking is similar to the locking phenomenon in linear elasticity (see, e.g., [16]), spurious pressure oscillations occur when the displacement of the porous skeleton is driven to a divergence-free state, the permeability of the porous solid is low and the so-called "constrained

specific storage constant" is close to zero (see, e.g., [114]).

Recently, Oyarzúa et al. [107] (see also [93]) proposed and analyzed a three-field formulation for the stationary Biot's model using classical finite element methods that are locking-free and free of spurious pressure oscillations. More precisely, in addition to the displacement and the pore pressure of the fluid, they introduced a "total pressure", showing existence, uniqueness, and stability of the discrete solution with constants independent of the modulus of dilatation, even in the incompressible limit. To achieve a numerical scheme that is also mass conserving, they later extended this formulation to a four-field formulation by introducing also the "fluid flux" as an unknown [92]. They propose to approximate the solid displacement in this model by a finite volume method (FVM) while remaining unknowns are approximated by a mixed finite element method (MFE).

In this chapter, we consider a conforming finite element discretization of the four-field formulation of the stationary Biot's consolidation model [92]. Assuming standard hypotheses on the discrete spaces, we first show well-posedness and optimal *a priori* error estimates of the Galerkin scheme. In particular, we show that any pair of stable Stokes element, such as the Hood–Taylor elements, for solid displacements and total pressure, combined with Raviart–Thomas elements of degree $k \geq 0$ for the fluid flux, and discontinuous polynomials of degree k for the pore pressure, are suitable finite element subspaces for this problem. We furthermore show that the scheme is locking-free.

The main contribution of this chapter, however, is a reliable and efficient residual-based *a posteriori* error estimator for the four-field formulation of Biot's consolidation model. In this direction, an *a posteriori* error analysis for a conforming finite element method (with Backward Euler time stepping) of the displacement-pressure formulation for poroelasticity was presented by Ern and Meunier [69]. They proved reliability and efficiency estimates related to energy norms through direct arguments (dual problems, local properties of Clément-type interpolation operators, and localization techniques), and showed an overall convergence of $\mathcal{O}(h)$. To show higher order accuracy, an elliptic reconstruction approach was applied but without efficiency of the estimator. Later, a reliable *a posteriori* error estimator based on stress and flux reconstructions was proposed by Riedlbeck et al. [119], while a reliable space-time *a posteriori* error estimator for a four-field system, in terms of the total stress tensor, displacement, fluid flux, and pressure, was derived in [1]. To the best of our knowledge, however, no efficiency estimates for poroelasticity have been proven for higher order accurate approximations.

In this chapter, we will prove efficiency estimates for higher order accurate approximations of the four-field formulation of Biot's consolidation model by using a localization technique by bubble functions and inverse inequalities. Such an approach was previously used, for example, in the *a posteriori* analysis of the Stokes-Darcy problem in [80], and of the elasticity problem in [49] and [140]. By inf-sup conditions on the involved finite element spaces, Helmholtz decompositions, and standard local approximation properties of Clément and Raviart–Thomas interpolation operators, we furthermore prove a reliability estimate and propose an adaptive algorithm for our problem.

This chapter is organized as follows. The governing equations, corresponding weak formulation and well-posedness of the problem are discussed in Section 3.2. In Section 3.3 we introduce the Galerkin scheme and derive the stability result and corresponding Céa's estimate. We derive a reliable and efficient residual-based *a posteriori* error estimator in Section 3.4 and present numerical results in Section 3.5.

3.2 A four-field formulation of Biot's equations

3.2.1 Notation

Let $\Omega \subseteq \mathbb{R}^d$, $d \in \{2,3\}$, denote a bounded and simply connected domain with Lipschitz boundary $\Gamma = \overline{\Gamma}_{\boldsymbol{u}} \cup \overline{\Gamma}_p$ such that $|\Gamma_{\boldsymbol{u}}| > 0$ and $\Gamma_{\boldsymbol{u}} \cap \Gamma_p \neq \varnothing$. In what follows we use standard notation for Sobolev spaces and norms, and denote spaces of vector-valued functions in boldface. For example, if $r \in \mathbb{R}$, we denote $\mathbf{H}^r(\Omega) := [\mathrm{H}^r(\Omega)]^d$ and $\mathbf{H}^r(\Gamma) := [\mathrm{H}^r(\Gamma)]^d$, with the convention that $\mathbf{H}^0(\Omega) = \mathbf{L}^2(\Omega)$ and $\mathbf{H}^0(\Gamma) = \mathbf{L}^2(\Gamma)$. For vector-valued functions we also require the Hilbert space

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

equipped with the norm

$$\|\cdot\|_{\mathrm{div},\Omega} := \left(\|\cdot\|_{0,\Omega}^2 + \|\mathrm{div}\,(\cdot)\|_{0,\Omega}^2 \right)^{1/2}.$$

Furthermore, we denote by $\mathrm{H}_{00}^{-1/2}(\Gamma_p)$ the dual space of $\mathrm{H}_{00}^{1/2}(\Gamma_p) := \left\{ q|_{\Gamma_p} : \ q \in \mathrm{H}^1_{\Gamma_{\boldsymbol{u}}}(\Omega) \right\}$, with

$$\mathrm{H}^1_{\Gamma_{\boldsymbol{u}}}(\Omega) := \left\{ v \in \mathrm{H}^1(\Omega) : \ v|_{\Gamma_{\boldsymbol{u}}} = 0 \right\}. \tag{3.1}$$

The space $\mathrm{H}_{00}^{1/2}(\Gamma_p)$ is endowed with the norm

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

Finally, by $\mathbf{0}$ we will refer to the generic null vector (including the null functional and operator), and we will denote by C, with or without subscripts, bar, tildes, or hats, generic constants independent of the discretization parameters.

3.2.2 Governing equations

For all $t > 0$, given a body force $\boldsymbol{f}(t) : \Omega \to \mathbb{R}^d$ and a volumetric fluid source $\ell(t) : \Omega \to \mathbb{R}$, the classical Biot's consolidation problem, describing the interaction between fluid motion and linear mechanical response of a porous medium occupying Ω, consists in finding the displacement of the porous skeleton $\boldsymbol{u}(t) : \Omega \to \mathbb{R}^d$, and the total pore pressure of the fluid $p(t) : \Omega \to \mathbb{R}$, satisfying

$$\partial_t(c_0 p + \alpha(\mathrm{div}\,\boldsymbol{u})) - \frac{1}{\eta}\mathrm{div}\,[\kappa(\nabla p - \rho \boldsymbol{g})] = \ell \qquad \text{in } \Omega, \tag{3.2}$$

$$\boldsymbol{\sigma} = \lambda(\mathrm{div}\,\boldsymbol{u})\boldsymbol{I} + 2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}) - p\boldsymbol{I} \qquad \text{in } \Omega, \tag{3.3}$$

$$-\mathbf{div}\,\boldsymbol{\sigma} = \boldsymbol{f} \qquad \text{in } \Omega, \tag{3.4}$$

and suitable boundary and initial conditions. Above $\boldsymbol{\sigma}$ is the total Cauchy solid stress, $\boldsymbol{\varepsilon}(\boldsymbol{u}) := \frac{1}{2}\left(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T\right)$ is the total strain rate tensor, $\boldsymbol{I}$ is the identity tensor in $\mathbb{R}^{d\times d}$, and $\mathbf{div}$ stands for the divergence operator div acting along the rows of a given tensor. Furthermore, $\boldsymbol{g}$ is the gravity acceleration (constant and aligned with the vertical direction), $\alpha > 0$ is the so-called Biot–Willis parameter (which is close to 1), $c_0 > 0$ is the constrained specific storage coefficient, $\eta, \rho > 0$ are the viscosity and density of the pore fluid, λ, μ are the Lamé parameters of the solid (dilation and

shear moduli of the solid), and κ is the permeability of the porous solid, here assumed to be uniformly bounded: $0 < \kappa_1 \leq \kappa(\boldsymbol{x}) \leq \kappa_2$ for all $\boldsymbol{x} \in \Omega$.

Using Backward Euler time stepping to discretize (3.2)–(3.4) in time, we obtain

$$(c_0 p^{n+1} + \alpha(\operatorname{div} \boldsymbol{u}^{n+1})) - \frac{1}{\eta}\operatorname{div}\left[\kappa(\nabla p^{n+1} - \rho \boldsymbol{g})\right] = \ell^{n+1} + (c_0 p^n + \alpha(\operatorname{div} \boldsymbol{u}^n)) \quad \text{in } \Omega, \tag{3.5}$$

$$\boldsymbol{\sigma}^{n+1} = \lambda(\operatorname{div} \boldsymbol{u}^{n+1})\boldsymbol{I} + 2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}^{n+1}) - p^{n+1}\boldsymbol{I} \quad \text{in } \Omega, \tag{3.6}$$

$$-\mathbf{div}\, \boldsymbol{\sigma}^{n+1} = \boldsymbol{f}^{n+1} \quad \text{in } \Omega, \tag{3.7}$$

where we absorbed the discrete time step into the constants c_0 and α. Re-defining $\ell^{n+1} \leftarrow \ell^{n+1} + (c_0 p^n + \alpha(\operatorname{div} \boldsymbol{u}^n))$ and dropping the superscript, we obtain the system of equations that needs to be solved at each time step:

$$(c_0 p + \alpha(\operatorname{div} \boldsymbol{u})) - \frac{1}{\eta}\operatorname{div}\left[\kappa(\nabla p - \rho \boldsymbol{g})\right] = \ell \quad \text{in } \Omega, \tag{3.8}$$

$$\boldsymbol{\sigma} = \lambda(\operatorname{div} \boldsymbol{u})\boldsymbol{I} + 2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}) - p\boldsymbol{I} \quad \text{in } \Omega, \tag{3.9}$$

$$-\mathbf{div}\, \boldsymbol{\sigma} = \boldsymbol{f} \quad \text{in } \Omega, \tag{3.10}$$

In this paper we will analyze this "stationary" case of Biot's consolidation problem. In particular, following [92] by introducing the total fluid-structure pressure (or total volumetric stress) $\phi = \alpha p - \lambda \operatorname{div} \boldsymbol{u}$ and the fluid flux $\boldsymbol{\sigma} = -\frac{\kappa}{\eta}(\nabla p - \rho \boldsymbol{g})$ as new unknowns, we study a conforming discretization of the following system

$$-\mathbf{div}\,(2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}) - \phi\boldsymbol{I}) = \boldsymbol{f} \quad \text{in} \quad \Omega, \tag{3.11a}$$

$$\phi = \alpha p - \lambda \operatorname{div} \boldsymbol{u} \quad \text{in} \quad \Omega, \tag{3.11b}$$

$$\boldsymbol{\sigma} = -\frac{\kappa}{\eta}(\nabla p - \rho \boldsymbol{g}) \quad \text{in} \quad \Omega, \tag{3.11c}$$

$$\left(c_0 + \frac{\alpha^2}{\lambda}\right) p - \frac{\alpha}{\lambda}\phi + \operatorname{div} \boldsymbol{\sigma} = \ell \quad \text{in} \quad \Omega, \tag{3.11d}$$

complemented with suitable boundary conditions

$$p = p_\Gamma, \quad (2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}) - \phi\boldsymbol{I})\boldsymbol{n} = \boldsymbol{m}_\Gamma \quad \text{on} \quad \Gamma_p, \tag{3.12a}$$

$$\boldsymbol{u} = \boldsymbol{0}, \quad \boldsymbol{\sigma} \cdot \boldsymbol{n} = 0 \quad \text{on} \quad \Gamma_{\boldsymbol{u}}, \tag{3.12b}$$

where $\boldsymbol{m}_\Gamma \in \mathbf{H}_{00}^{-1/2}(\Gamma_p)$ and $p_\Gamma \in \mathrm{H}^{1/2}(\Gamma_{\boldsymbol{u}})$.

3.2.3 Weak formulation

The weak formulation of the coupled problem (3.11) is given by [92, Section 2]: Find $(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p) \in \mathbf{H} \times \mathrm{Q} \times \mathbf{Z} \times \mathrm{Q}$ such that

$$a_s(\boldsymbol{u}, \boldsymbol{v}) + b_s(\boldsymbol{v}, \phi) = F(\boldsymbol{v}) \quad \forall\, \boldsymbol{v} \in \mathbf{H}, \tag{3.13a}$$

$$b_s(\boldsymbol{u}, \psi) - c_s(\phi, \psi) + b_{sf}(\psi, p) = 0 \quad \forall\, \psi \in \mathrm{Q}, \tag{3.13b}$$

$$a_f(\boldsymbol{\sigma}, \boldsymbol{\tau}) + b_f(\boldsymbol{\tau}, p) = G(\boldsymbol{\tau}) \quad \forall\, \boldsymbol{\tau} \in \mathbf{Z}, \tag{3.13c}$$

$$b_{sf}(\phi,q)+ b_f(\boldsymbol{\sigma},q)-c_f(p,q) \;\; = H(q) \quad \forall\, q \in \mathrm{Q}, \tag{3.13d}$$

where, by the boundary conditions (3.12b), the functional spaces are defined as

$$\mathbf{H} := \mathbf{H}^1_{\Gamma_{\boldsymbol{u}}}(\Omega), \quad \mathrm{Q} := \mathrm{L}^2(\Omega), \quad \text{and} \quad \mathbf{Z} := \{\boldsymbol{\tau} \in \mathbf{H}(\mathrm{div};\Omega) : \; \boldsymbol{\tau}\cdot\boldsymbol{n} = 0 \quad \text{on} \quad \Gamma_{\boldsymbol{u}}\},$$

and the corresponding forms are defined as

$$\begin{aligned}
&a_s(\boldsymbol{u},\boldsymbol{v}) := 2\mu\int_\Omega \boldsymbol{\varepsilon}(\boldsymbol{u}) : \boldsymbol{\varepsilon}(\boldsymbol{v}), \quad b_s(\boldsymbol{v},\psi) := -\int_\Omega \psi\,\mathrm{div}\,\boldsymbol{v}, \quad c_s(\phi,\psi) := \frac{1}{\lambda}\int_\Omega \phi\psi, \quad b_{sf}(\psi,q) := \frac{\alpha}{\lambda}\int_\Omega \psi q,\\
&a_f(\boldsymbol{\sigma},\boldsymbol{\tau}) := \int_\Omega \frac{\eta}{\kappa}\boldsymbol{\sigma}\cdot\boldsymbol{\tau}, \quad b_f(\boldsymbol{\tau},q) := -\int_\Omega q\,\mathrm{div}\,\boldsymbol{\tau}, \quad c_f(p,q) := \left(c_0 + \frac{\alpha^2}{\lambda}\right)\int_\Omega pq,\\
&F(\boldsymbol{v}) := \int_\Omega \boldsymbol{f}\cdot\boldsymbol{v} + \langle \boldsymbol{m}_\Gamma, \boldsymbol{v}\rangle_{\Gamma_p}, \quad G(\boldsymbol{\tau}) := \int_\Omega \rho\boldsymbol{g}\cdot\boldsymbol{\tau} - \langle \boldsymbol{\tau}\cdot\boldsymbol{n}, p_\Gamma\rangle_{\Gamma_p}, \quad H(q) := -\int_\Omega \ell q.
\end{aligned} \tag{3.14}$$

The subscripts "s" or "f" are introduced to emphasize that a bilinear form is only related to structure or fluid variables, respectively.

Let us discuss the stability properties of the forms involved in (3.13). Firstly, it is easy to check that

$$\begin{aligned}
&|a_s(\boldsymbol{u},\boldsymbol{v})| \le 2\mu C_{k,2}\|\boldsymbol{u}\|_{1,\Omega}\|\boldsymbol{v}\|_{1,\Omega}, && |a_f(\boldsymbol{\sigma},\boldsymbol{\tau})| \le \eta\kappa_1^{-1}\|\boldsymbol{\sigma}\|_{\mathrm{div},\Omega}\|\boldsymbol{\tau}\|_{\mathrm{div},\Omega},\\
&|b_s(\boldsymbol{v},\psi)| \le \sqrt{d}\|\boldsymbol{v}\|_{1,\Omega}\|\psi\|_{0,\Omega}, && |b_f(\boldsymbol{\tau},q)| \le \|\boldsymbol{\tau}\|_{\mathrm{div},\Omega}\|q\|_{0,\Omega},\\
&|b_{sf}(\psi,q)| \le \alpha\lambda^{-1}\|\psi\|_{0,\Omega}\|q\|_{0,\Omega}, && |c_s(\phi,\psi)| \le \lambda^{-1}\|\phi\|_{0,\Omega}\|\psi\|_{0,\Omega},\\
&|c_f(p,q)| \le \left(c_0 + \alpha^2\lambda^{-1}\right)\|p\|_{0,\Omega}\|q\|_{0,\Omega},
\end{aligned} \tag{3.15}$$

for all $\boldsymbol{u},\boldsymbol{v} \in \mathbf{H}$, $p,q,\phi,\psi \in \mathrm{Q}$, and $\boldsymbol{\sigma},\boldsymbol{\tau} \in \mathbf{Z}$. Above, $C_{k,2}$ is one of the positive constants satisfying

$$C_{k,1}\|\boldsymbol{v}\|^2_{1,\Omega} \le \|\boldsymbol{\varepsilon}(\boldsymbol{u})\|^2_{0,\Omega} \le C_{k,2}\|\boldsymbol{v}\|^2_{1,\Omega} \quad \forall\, \boldsymbol{v} \in \mathbf{H}. \tag{3.16}$$

Also, the functionals F, G, and H can be bounded as follows:

$$\begin{aligned}
&|F(\boldsymbol{v})| \le \left(\|\boldsymbol{f}\|_{0,\Omega} + \|\boldsymbol{m}_\Gamma\|_{-1/2,00,\Gamma_p}\right)\|\boldsymbol{v}\|_{1,\Omega} && \forall\, \boldsymbol{v} \in \mathbf{H},\\
&|G(\boldsymbol{\tau})| \le \left(\rho\|\boldsymbol{g}\|_{0,\Omega} + \|p_\Gamma\|_{1/2,00,\Gamma_p}\right)\|\boldsymbol{\tau}\|_{\mathrm{div},\Omega} && \forall\, \boldsymbol{\tau} \in \mathbf{Z},\\
&|H(q)| \le \|\ell\|_{0,\Omega}\|q\|_{0,\Omega} && \forall\, q \in \mathrm{Q}.
\end{aligned}$$

On the other hand, the positivity of the bilinear forms a_s and a_f is immediate from the lower bound for κ and the inequality (3.16). More precisely, we have

$$a_s(\boldsymbol{v},\boldsymbol{v}) \ge 2\mu C_{k,1}\|\boldsymbol{v}\|^2_{1,\Omega} \quad \forall\, \boldsymbol{v} \in \mathbf{H}, \quad \text{and} \quad a_f(\boldsymbol{\tau},\boldsymbol{\tau}) \ge \eta\kappa_2^{-1}\|\boldsymbol{\tau}\|^2_{\mathrm{div},\Omega} \quad \forall\, \boldsymbol{\tau} \in \mathbf{K}_f, \tag{3.17}$$

where

$$\mathbf{K}_f := \{\boldsymbol{\tau} \in \mathbf{Z} : \; b_f(\boldsymbol{\tau},q) = 0 \;\; \forall\, q \in \mathrm{Q}\} = \{\boldsymbol{\tau} \in \mathbf{Z} : \; \mathrm{div}\,\boldsymbol{\tau} = 0 \;\; \text{in} \;\; \Omega\}. \tag{3.18}$$

Finally, the following inf-sup conditions are well-known to hold (see, e.g., [84]):

$$\sup_{\substack{\boldsymbol{v}_h \in \mathbf{H}\\ \boldsymbol{v}_h \neq \mathbf{0}}} \frac{b_s(\boldsymbol{v},\psi)}{\|\boldsymbol{v}\|_{1,\Omega}} \ge \beta_s\|\psi\|_{0,\Omega} \quad \forall\, \psi \in \mathrm{Q}, \quad \text{and} \quad \sup_{\substack{\boldsymbol{\tau} \in \mathbf{Z}\\ \boldsymbol{\tau}_h \neq \mathbf{0}}} \frac{b_f(\boldsymbol{\tau},q)}{\|\boldsymbol{\tau}\|_{\mathrm{div},\Omega}} \ge \beta_f\|q\|_{0,\Omega} \quad \forall\, q \in \mathrm{Q},$$

where $\beta_s, \beta_f > 0$ depend on $|\Omega|$.

Let us now briefly comment on the well-posedness of the problem (3.13). To this end, we follow the approach of [92, Section 2]. We start by recalling the following continuous dependence result for (3.13) with arbitrary functionals. This will also be useful later on when deriving our *a priori* and *a posteriori* error bounds (cf. Sections 3.3 and 3.4, respectively). To alleviate the notation, in the sequel we use the norm

$$|||(\boldsymbol{v}, \psi, \boldsymbol{\tau}, q)||| := \|\boldsymbol{v}\|_{1,\Omega} + \|\psi\|_{0,\Omega} + \|\boldsymbol{\tau}\|_{\mathrm{div},\Omega} + \|q\|_{0,\Omega} \tag{3.19}$$

for all $\boldsymbol{v} \in \mathbf{H}$, $\psi \in \mathrm{Q}$, $\boldsymbol{\tau} \in \mathbf{Z}$, $p \in \mathrm{Q}$.

Lemma 3.1. *Given $F_1 \in \mathbf{H}'$, $G_1 \in \mathrm{Q}'$, $F_2 \in \mathbf{Z}'$ and $G_2 \in \mathrm{Q}'$, let $(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p) \in \mathbf{H} \times \mathrm{Q} \times \mathbf{Z} \times \mathrm{Q}$ be such that*

$$\begin{aligned}
a_s(\boldsymbol{u}, \boldsymbol{v}) + b_s(\boldsymbol{v}, \phi) &= F_1(\boldsymbol{v}) \quad \forall\, \boldsymbol{v} \in \mathbf{H}, && \text{(3.20a)}\\
b_s(\boldsymbol{u}, \psi) - c_s(\phi, \psi) + b_{sf}(\psi, p) &= G_1(\psi) \quad \forall\, \psi \in \mathrm{Q}, && \text{(3.20b)}\\
a_f(\boldsymbol{\sigma}, \boldsymbol{\tau}) + b_f(\boldsymbol{\tau}, p) &= F_2(\boldsymbol{\tau}) \quad \forall\, \boldsymbol{\tau} \in \mathbf{Z}, && \text{(3.20c)}\\
b_{sf}(\phi, q) + b_f(\boldsymbol{\sigma}, q) - c_f(p, q) &= G_2(q) \quad \forall\, q \in \mathrm{Q}, && \text{(3.20d)}
\end{aligned}$$

where the bilinear forms a_s, b_s, c_s, a_f, b_f, c_s and b_{sf} are given by (3.14). There exists a constant $C > 0$, independent of λ, such that

$$|||(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p)||| \leq C \left(\|F_1\|_{\mathbf{H}'} + \|G_1\|_{\mathrm{Q}'} + \|F_2\|_{\mathbf{Z}'} + \|G_2\|_{\mathrm{Q}'} \right), \tag{3.21}$$

Now, let $\mathcal{M} : \mathbf{H} \times \mathrm{Q} \times \mathbf{Z} \times \mathrm{Q} \to \mathbf{H} \times \mathrm{Q} \times \mathbf{Z} \times \mathrm{Q}$ be the mapping induced by the left-hand side of (3.20). Then, if $(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p)$ satisfies (3.20), it follows that

$$\mathcal{M}(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p) = (\mathcal{R}_{\mathbf{H}}(F_1), \mathcal{R}_{\mathrm{Q}}(G_1), \mathcal{R}_{\mathbf{Z}}(F_2), \mathcal{R}_{\mathrm{Q}}(G_2)),$$

where $\mathcal{R}_{\mathbf{H}} : \mathbf{H}' \to \mathbf{H}$, $\mathcal{R}_{\mathrm{Q}} : \mathrm{Q}' \to \mathrm{Q}$ and $\mathcal{R}_{\mathbf{Z}} : \mathbf{Z}' \to \mathbf{Z}$ are the corresponding Riesz operators. Moreover, from (3.21) we have

$$|||(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p)||| \leq C |||\mathcal{M}(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p)|||,$$

which implies that $\mathcal{M}$ has closed range and its kernel is the null vector, or equivalently, $\mathcal{M}^*$ is surjective (see, e.g., [38, Theorem 2.20]). Since $\mathcal{M}$ is self-adjoint, it becomes clear that the unique solvability of (3.13) follows from the estimate (3.21) by setting $F_1 = F$, $G_1 = 0$, $F_2 = G$ and $G_2 = H$, that is, the following result holds.

Theorem 3.2. *There exists a unique $(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p) \in \mathbf{H} \times \mathrm{Q} \times \mathbf{Z} \times \mathrm{Q}$ satisfying (3.13). Moreover, there exists $C_{\mathrm{stab}} > 0$, independent of λ, such that*

$$|||(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p)||| \leq C_{\mathrm{stab}} \left(\|\boldsymbol{f}\|_{0,\Omega} + \|\boldsymbol{g}\|_{0,\Omega} + \|\ell\|_{0,\Omega} + \|\boldsymbol{m}_\Gamma\|_{-1/2,00,\Gamma_p} + \|p_\Gamma\|_{1/2,00,\Gamma_p} \right).$$

We close this section by observing that the solution of (3.13) solves the original problem (3.11) in the sense of the following lemma.

Lemma 3.3. *Let $(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p) \in \mathbf{H} \times \mathrm{Q} \times \mathbf{Z} \times \mathrm{Q}$ be the unique solution of (3.13). It satisfies in a distributional sense, $-\mathbf{div}\,(2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}) - \phi\boldsymbol{I}) = \boldsymbol{f}$ in Ω, $\frac{1}{\lambda}(\alpha p - \phi) - \mathrm{div}\,\boldsymbol{u} = 0$ in Ω, $\frac{\eta}{\kappa}\boldsymbol{\sigma} + \nabla p - \rho\boldsymbol{g} = \mathbf{0}$ in Ω, $\left(c_0 + \frac{\alpha^2}{\lambda}\right) p - \frac{\alpha}{\lambda}\phi + \mathrm{div}\,\boldsymbol{\sigma} - \ell = 0$ in Ω. Additionally, $\boldsymbol{u}$, ϕ, $\boldsymbol{\sigma}$ and p satisfy the boundary conditions described in (3.12a)-(3.12b).*

Proof. The result follows by applying integration by parts in (3.13) and using suitable test functions. We omit the details. □

3.3 The Galerkin method

In this section we introduce the Galerkin approximation of the problem (3.13), analyze its well-posedness and establish the associated Céa's estimate. For this, we consider arbitrary finite dimensional subspaces, denoted by

$$\mathbf{H}_h \subseteq \mathbf{H}, \quad \mathrm{Q}_h, \mathrm{W}_h \subseteq \mathrm{Q}, \quad \text{and} \quad \mathbf{Z}_h \subseteq \mathbf{Z}. \tag{3.22}$$

Hereafter, the index $h > 0$, refers to the meshsize of a shape-regular triangulation $\mathcal{T}_h$ of $\overline{\Omega}$ made of triangles T (when $d = 2$) or tetrahedra (when $d = 3$) of diameter h_T, i.e., $h := \max\{h_T : T \in \mathcal{T}_h\}$.

In this way, the Galerkin scheme associated to (3.13) reads: Find $(\boldsymbol{u}_h, \phi_h, \boldsymbol{\sigma}_h, p_h) \in \mathbf{H}_h \times \mathrm{Q}_h \times \mathbf{Z}_h \times \mathrm{W}_h$ such that

$$\begin{aligned}
a_s(\boldsymbol{u}_h, \boldsymbol{v}_h) + b_s(\boldsymbol{v}_h, \phi_h) &= F(\boldsymbol{v}_h) \quad \forall\, \boldsymbol{v}_h \in \mathbf{H}_h, && \text{(3.23a)}\\
b_s(\boldsymbol{u}_h, \psi_h) - c_s(\phi_h, \psi_h) + b_{sf}(\psi_h, p_h) &= 0 \quad \forall\, \psi_h \in \mathrm{Q}_h, && \text{(3.23b)}\\
a_f(\boldsymbol{\sigma}_h, \boldsymbol{\tau}_h) + b_f(\boldsymbol{\tau}_h, p_h) &= G(\boldsymbol{\tau}_h) \quad \forall\, \boldsymbol{\tau}_h \in \mathbf{Z}_h, && \text{(3.23c)}\\
b_{sf}(\phi_h, q_h) + b_f(\boldsymbol{\sigma}_h, q_h) - c_f(p_h, q_h) &= H(q_h) \quad \forall\, q_h \in \mathrm{W}_h, && \text{(3.23d)}
\end{aligned}$$

where the bilinear forms and the functionals are as in (3.14).

Next, we proceed as in [92] and make use of the discrete analogue of Lemma 3.1 to prove the well-posedness of the Galerkin scheme (3.23). Before doing so, in order to ensure the stability properties of the bilinear forms that are not inherited from the continuous case, we derive general hypotheses on the subspaces in (3.22).

Let us first look at the discrete kernel of the bilinear form b_f, which is given by

$$\mathbf{K}_{f,h} := \{\boldsymbol{\tau}_h \in \mathbf{Z}_h : \; b_f(\boldsymbol{\tau}_h, q_h) = 0 \;\; \forall\, q_h \in \mathrm{W}_h\}\,.$$

A more explicit definition of this space can be obtained if we assume that

(H0) $\operatorname{div} \mathbf{Z}_h \subseteq \mathrm{W}_h$.

In fact, this implies that $\mathbf{K}_{f,h} = \{\boldsymbol{\tau}_h \in \mathbf{Z}_h : \operatorname{div} \boldsymbol{\tau}_h = 0 \text{ in } \Omega\}$. Moreover, since $\mathbf{K}_{f,h} \subseteq \mathbf{K}_f$ (cf. (3.18)), the ellipticity of bilinear form a_f on $\mathbf{K}_{f,h}$ is deduced from (3.17), and with the same constant.

Let us also assume that the following discrete inf-sup conditions hold:

(H1) There exists $\widehat{\beta}_f > 0$, independent of h, such that

$$\sup_{\substack{\boldsymbol{\tau}_h \in \mathbf{Z}_h \\ \boldsymbol{\tau}_h \neq \mathbf{0}}} \frac{b_f(\boldsymbol{\tau}_h, q_h)}{\|\boldsymbol{\tau}_h\|_{\mathrm{div},\Omega}} \geq \widehat{\beta}_f \|q_h\|_{0,\Omega} \quad \forall\, q_h \in \mathrm{W}_h.$$

(H2) There exists $\widehat{\beta}_s > 0$, independent of h, such that

$$\sup_{\substack{\boldsymbol{v}_h \in \mathbf{H}_h \\ \boldsymbol{v}_h \neq \mathbf{0}}} \frac{b_s(\boldsymbol{v}_h, \psi_h)}{\|\boldsymbol{v}_h\|_{1,\Omega}} \geq \widehat{\beta}_s \|\psi_h\|_{0,\Omega} \quad \forall\, \psi_h \in \mathrm{Q}_h.$$

In Section 3.3.1 we specify suitable choices of finite element subspaces satisfying the above hypotheses. We remark in advance that $(\mathbf{H}_h, \mathrm{Q}_h)$ can be taken as a pair of stable finite element subspaces for the Stokes problem, whereas $\mathbf{Z}_h$ and W_h are given by, but are not limited to, the Raviart–Thomas element and the space of discontinuous polynomials, respectively.

The following result is the discrete analogue of Lemma 3.1 and can be proven by a similar technique.

Lemma 3.4. *Given $\widehat{F}_1 \in \mathbf{H}_h'$, $\widehat{G}_1 \in \mathrm{Q}_h'$, $\widehat{F}_2 \in \mathbf{Z}_h'$ and $\widehat{G}_2 \in \mathrm{W}_h'$, let $(\boldsymbol{u}_h, \phi_h, \boldsymbol{\sigma}_h, p_h) \in \mathbf{H}_h \times \mathrm{Q}_h \times \mathbf{Z}_h \times \mathrm{W}_h$ be such that*

$$\begin{aligned}
a_s(\boldsymbol{u}_h, \boldsymbol{v}_h) + b_s(\boldsymbol{v}_h, \phi_h) &= \widehat{F}_1(\boldsymbol{v}_h) \quad \forall\, \boldsymbol{v}_h \in \mathbf{H}_h, && \text{(3.24a)}\\
b_s(\boldsymbol{u}_h, \psi_h) - c_s(\phi_h, \psi_h) + b_{sf}(\psi_h, p_h) &= \widehat{G}_1(\psi_h) \quad \forall\, \psi_h \in \mathrm{Q}_h, && \text{(3.24b)}\\
a_f(\boldsymbol{\sigma}_h, \boldsymbol{\tau}_h) + b_f(\boldsymbol{\tau}_h, p_h) &= \widehat{F}_2(\boldsymbol{\tau}_h) \quad \forall\, \boldsymbol{\tau}_h \in \mathbf{Z}_h, && \text{(3.24c)}\\
b_{sf}(\phi_h, q_h) + b_f(\boldsymbol{\sigma}_h, q_h) - c_f(p_h, q_h) &= \widehat{G}_2(q_h) \quad \forall\, q_h \in \mathrm{W}_h, && \text{(3.24d)}
\end{aligned}$$

where the bilinear forms are defined as in (3.14), and suppose that hypotheses **(H0)**-**(H2)** *hold. There exists a constant $C > 0$, independent of λ and h, such that*

$$|||(\boldsymbol{u}_h, \phi_h, \boldsymbol{\sigma}_h, p_h)||| \leq C \left(\|\widehat{F}_1\|_{\mathbf{H}_h'} + \|\widehat{G}_1\|_{\mathrm{Q}_h'} + \|\widehat{F}_2\|_{\mathbf{Z}_h'} + \|\widehat{G}_2\|_{\mathrm{W}_h'} \right). \tag{3.25}$$

We are now in a position of stating the well-posedness of the Galerkin scheme (3.23) and the associated Céa's estimate.

Theorem 3.5. *Suppose that* **(H0)**-**(H2)** *hold. Then, there exists a unique $(\boldsymbol{u}_h, \phi_h, \boldsymbol{\sigma}_h, p_h) \in \mathbf{H}_h \times \mathrm{Q}_h \times \mathbf{Z}_h \times \mathrm{W}_h$ satisfying (3.23). Moreover, there exists a constant $\widehat{C}_{\mathrm{stab}}$, independent of λ and h, such that*

$$|||(\boldsymbol{u}_h, \phi_h, \boldsymbol{\sigma}_h, p_h)||| \leq \widehat{C}_{\mathrm{stab}} \left(\|\boldsymbol{f}\|_{0,\Omega} + \|\boldsymbol{g}\|_{0,\Omega} + \|\ell\|_{0,\Omega} + \|\boldsymbol{m}_\Gamma\|_{-1/2,00,\Gamma_p} + \|p_\Gamma\|_{1/2,00,\Gamma_p} \right). \tag{3.26}$$

In addition, there exists $C_{\mathrm{cea}} > 0$, also independent of λ and h, such that

$$\begin{aligned}
&|||(\boldsymbol{u} - \boldsymbol{u}_h, \phi - \phi_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h, p - p_h)|||\\
&\quad \leq C_{\mathrm{cea}} \left(\inf_{\boldsymbol{v}_h \in \mathbf{H}_h} \|\boldsymbol{u} - \boldsymbol{v}_h\|_{1,\Omega} + \inf_{\psi_h \in \mathrm{Q}_h} \|\phi - \psi_h\|_{0,\Omega} + \inf_{\boldsymbol{\tau}_h \in \mathbf{Z}_h} \|\boldsymbol{\sigma} - \boldsymbol{\tau}_h\|_{\mathrm{div},\Omega} + \inf_{q_h \in \mathrm{W}_h} \|p - q_h\|_{0,\Omega} \right).
\end{aligned} \tag{3.27}$$

Proof. We first observe that (3.26) is a particular case of estimate (3.25). Consequently, the unique solvability of problem (3.23) can be readily deduced. In fact, since in finite dimensional linear problems existence and uniqueness of the solution are equivalent, it suffices to note, thanks to (3.26), that the solution of the Galerkin scheme (3.23) with homogeneous data will be the trivial one.

It remains to prove (3.27), for which we proceed as in the proof of [92, Theorem 5.1]. Firstly, testing equations (3.13a)-(3.13d) with $(\boldsymbol{v},\psi,\boldsymbol{\tau},q)=(\boldsymbol{v}_h,\psi_h,\boldsymbol{\tau}_h,q_h)\in\mathbf{H}_h\times\mathrm{Q}_h\times\mathbf{Z}_h\times\mathrm{W}_h$ and subtracting the resulting system from (3.23), we get the Galerkin orthogonality equations

$$
\begin{aligned}
a_s(\boldsymbol{u}-\boldsymbol{u}_h,\boldsymbol{v}_h)+b_s(\boldsymbol{v}_h,\phi-\phi_h) &= 0 \quad \forall\,\boldsymbol{v}_h\in\mathbf{H}_h, && \text{(3.28a)}\\
b_s(\boldsymbol{u}-\boldsymbol{u}_h,\psi_h)-c_s(\phi-\phi_h,\psi_h)+b_{sf}(\psi_h,p-p_h) &= 0 \quad \forall\,\psi_h\in\mathrm{Q}_h, && \text{(3.28b)}\\
a_f(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,\boldsymbol{\tau}_h)+b_f(\boldsymbol{\tau}_h,p-p_h) &= 0 \quad \forall\,\boldsymbol{\tau}_h\in\mathbf{Z}_h, && \text{(3.28c)}\\
b_{sf}(\phi-\phi_h,q_h)+b_f(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,q_h)-c_f(p-p_h,q_h) &= 0 \quad \forall\,q_h\in\mathrm{W}_h. && \text{(3.28d)}
\end{aligned}
$$

Next, given $\widehat{\boldsymbol{v}}_h\in\mathbf{H}_h$, $\widehat{\psi}_h\in\mathrm{Q}_h$, $\widehat{\boldsymbol{\tau}}_h\in\mathbf{Z}_h$ and $\widehat{q}_h\in\mathrm{W}_h$, we let $\widehat{F}_1\in\mathbf{H}_h'$, $\widehat{G}_1\in\mathrm{Q}_h'$, $\widehat{F}_2\in\mathbf{Z}_h'$ and $\widehat{G}_2\in\mathrm{W}_h'$ be the functionals defined as follows:

$$
\begin{aligned}
\widehat{F}_1(\boldsymbol{v}_h) &:= -a_s(\boldsymbol{u}-\widehat{\boldsymbol{v}}_h,\boldsymbol{v}_h)-b_s(\boldsymbol{v}_h,\phi-\widehat{\psi}_h),\\
\widehat{G}_1(\psi_h) &:= -b_s(\boldsymbol{u}-\widehat{\boldsymbol{v}}_h,\psi_h)+c_s(\phi-\widehat{\psi}_h,\psi_h)-b_{sf}(\psi_h,p-\widehat{q}_h),\\
\widehat{F}_2(\boldsymbol{\tau}_h) &:= -a_f(\boldsymbol{\sigma}-\widehat{\boldsymbol{\tau}}_h,\boldsymbol{\tau}_h)-b_f(\boldsymbol{\tau}_h,p-\widehat{q}_h),\\
\widehat{G}_2(q_h) &:= -b_{sf}(\phi-\widehat{\psi}_h,q)-b_f(\boldsymbol{\sigma}-\widehat{\boldsymbol{\tau}}_h,q)+c_f(p-\widehat{q}_h,q_h).
\end{aligned}
$$

Then, adding and subtracting convenient terms to the individual errors in system (3.28), and using Lemma 3.4, it follows that

$$
\left|\!\left|\!\left|\left(\widehat{\boldsymbol{v}}_h-\boldsymbol{u}_h,\widehat{\psi}_h-\phi_h,\widehat{\boldsymbol{\tau}}_h-\boldsymbol{\sigma}_h,\widehat{q}_h-p_h\right)\right|\!\right|\!\right| \leq C\left(\|\widehat{F}_1\|_{\mathbf{H}_h'}+\|\widehat{G}_1\|_{\mathrm{Q}_h'}+\|\widehat{F}_2\|_{\mathbf{Z}_h'}+\|\widehat{G}_2\|_{\mathrm{W}_h'}\right). \tag{3.29}
$$

Using the boundedness of the above bilinear forms (cf. (3.15)), we have

$$
\begin{aligned}
\|\widehat{F}_1\|_{\mathbf{H}_h'} &\leq 2\mu C_{k,2}\|\boldsymbol{u}-\widehat{\boldsymbol{v}}_h\|_{1,\Omega}+\sqrt{d}\|\phi-\widehat{\psi}_h\|_{0,\Omega},\\
\|\widehat{G}_1\|_{\mathrm{Q}_h'} &\leq \sqrt{d}\|\boldsymbol{u}-\widehat{\boldsymbol{v}}_h\|_{1,\Omega}+\frac{1}{\lambda}\|\phi-\widehat{\psi}_h\|_{0,\Omega}+\frac{\alpha}{\lambda}\|p-\widehat{q}_h\|_{0,\Omega},\\
\|\widehat{F}_2\|_{\mathbf{Z}_h'} &\leq \frac{\eta}{\kappa_1}\|\boldsymbol{\sigma}-\widehat{\boldsymbol{\tau}}_h\|_{\mathrm{div},\Omega}+\|p-\widehat{q}_h\|_{0,\Omega},\\
\|\widehat{G}_2\|_{\mathrm{W}_h'} &\leq \frac{\alpha}{\lambda}\|\phi-\widehat{\psi}_h\|_{0,\Omega}+\|\boldsymbol{\sigma}-\widehat{\boldsymbol{\tau}}_h\|_{\mathrm{div},\Omega}+\left(c_0+\frac{\alpha^2}{\lambda}\right)\|p-\widehat{q}_h\|_{0,\Omega}.
\end{aligned}
$$

Therefore, we obtain using the triangle inequality and estimate (3.29),

$$
|||(\boldsymbol{u}-\boldsymbol{u}_h,\phi-\phi_h,\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,p-p_h)||| \leq \left(1+\widetilde{C}\right)\left|\!\left|\!\left|(\boldsymbol{u}-\widehat{\boldsymbol{v}}_h,\phi-\widehat{\psi}_h,\boldsymbol{\sigma}-\widehat{\boldsymbol{\tau}}_h,p-\widehat{q}_h)\right|\!\right|\!\right|,
$$

where

$$
\widetilde{C}:=C\max\left\{2\mu C_{k,2}+\sqrt{d},\ \frac{1}{\lambda}(1+\alpha)+\sqrt{d},\ \frac{\eta}{\kappa_1}+1,\ \frac{\alpha}{\lambda}(\alpha+1)+c_0+1\right\}.
$$

Above, $\widetilde{C}$ can be bounded by a constant independent of λ because $\lambda^{-1}(1+\alpha)$ and $\alpha\lambda^{-1}(1+\alpha)$ are bounded. In particular, they are negligible when volumetric locking occurs (i.e., as $\lambda\to\infty$). The proof ends by observing that $\widehat{\boldsymbol{v}}_h$, $\widehat{\psi}_h$, $\widehat{\boldsymbol{\tau}}_h$ and $\widehat{q}_h$ are arbitrary. □

3.3.1 Specific finite element subspaces

The aim of this section is to take advantage of the flexibility of conforming methods to provide concrete finite element subspaces satisfying the crucial hypotheses **(H0)**-**(H2)**. To that end, given an integer

$l \geq 0$ and a subset S of $\mathbb{R}^d$, we let $\mathrm{P}_l(S)$ (resp. $\widetilde{\mathrm{P}}_l(S)$) denote the space of polynomials of degree at most l on S (resp. of degree equal to l on S). We also set $\mathbf{P}_l(S) := [\mathrm{P}_l(S)]^d$.

Let $k \geq 0$ be an integer. The generalized Hood–Taylor element (see, e.g., [31, Section 8.8.2]) consists of the pair $(\mathbf{H}_h, \mathrm{Q}_h)$ specified by

$$\mathbf{H}_h := \left\{ \boldsymbol{v}_h \in [\mathcal{C}(\overline{\Omega})]^d : \ \boldsymbol{v}_h|_T \in \mathbf{P}_{k+2}(T) \ \ \forall T \in \mathcal{T}_h, \ \boldsymbol{v}_h = \mathbf{0} \quad \text{on} \quad \Gamma_{\boldsymbol{u}} \right\} \tag{3.30}$$

and

$$\mathrm{Q}_h := \left\{ \psi_h \in \mathcal{C}(\overline{\Omega}) : \ \psi_h|_T \in \mathrm{P}_{k+1}(T) \ \ \forall T \in \mathcal{T}_h \right\}. \tag{3.31}$$

This pair satisfies the inf-sup condition in hypothesis **(H2)**. We refer the reader to [29] for the proof (see also [31, 40]). In addition, the following approximation properties are well-known to hold:

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

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

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

$$\inf_{\psi_h \in \mathrm{Q}_h} \|\phi - \psi_h\|_{0,\Omega} \leq C h^{s+1} \|\phi\|_{s+1,\Omega}.$$

Furthermore, the local Raviart–Thomas space of order k, for each $T \in \mathcal{T}_h$, is defined as

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

where $\boldsymbol{x} := (x_1, \ldots, x_d)^T$ is a generic vector in $\mathbb{R}^d$. To approximate the fluid flux $\boldsymbol{\sigma}$ we consider the global Raviart–Thomas space of order k which is given by

$$\mathbf{Z}_h := \{ \boldsymbol{\tau}_h \in \mathbf{H}(\mathrm{div}; \Omega) : \ \boldsymbol{\tau}_h|_T \in \mathbf{RT}_k(T) \ \ \forall T \in \mathcal{T}_h, \ \ \boldsymbol{\tau}_h \cdot \boldsymbol{n} = 0 \quad \text{on} \quad \Gamma_{\boldsymbol{u}} \}. \tag{3.32}$$

We consider discontinuous polynomials of order k for the fluid pressure:

$$\mathrm{W}_h := \left\{ q_h \in \mathrm{L}^2(\Omega) : \ q_h|_T \in \mathrm{P}_k(T) \ \ \forall T \in \mathcal{T}_h \right\}. \tag{3.33}$$

It is well-known that the pair $(\mathbf{Z}_h, \mathrm{W}_h)$ satisfies the hypotheses **(H0)** and **(H1)** (see, e.g., [41, 73]). This fact completes the requirements of Theorem 3.5, and therefore the well-posedness of (3.23) holds for the above subspaces.

Let us now recall the approximation properties of $\mathbf{Z}_h$ and W_h.

($\mathbf{AP}_h^{\boldsymbol{\sigma}}$) There exists $C > 0$, independent of h, such that for each $m \in (0, k+1]$ and each $\boldsymbol{\sigma} \in \mathbf{H}^m(\Omega) \cap \mathbf{Z}$, with $\mathrm{div}\, \boldsymbol{\sigma} \in \mathrm{H}^m(\Omega)$, there holds

$$\inf_{\boldsymbol{\tau}_h \in \mathbf{Z}_h} \|\boldsymbol{\sigma} - \boldsymbol{\tau}_h\|_{\mathrm{div},\Omega} \leq C h^m (\|\boldsymbol{\sigma}\|_{m,\Omega} + \|\mathrm{div}\, \boldsymbol{\sigma}\|_{m,\Omega}).$$

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

$$\inf_{q_h \in \mathrm{W}_h} \|p - q_h\|_{0,\Omega} \leq C h^m \|p\|_{m,\Omega}.$$

From the above discussion, the following theorem provides the theoretical rate of convergence of the Galerkin scheme (3.23) under suitable regularity assumptions on the exact solution.

Theorem 3.6. *Given $s, m \in (0, k+1]$, assume that $\boldsymbol{u} \in \mathbf{H}^{s+2}(\Omega)$, $\phi \in \mathrm{H}^{s+1}(\Omega)$, $\boldsymbol{\sigma} \in \mathbf{H}^m(\Omega) \cap \mathbf{Z}$ such that $\mathrm{div}\,\boldsymbol{\sigma} \in \mathbf{H}^m(\Omega)$, and $p \in \mathrm{H}^m(\Omega)$. There exists $C_{\mathrm{rate}} > 0$, independent of λ and h, such that*

$$
\begin{aligned}
&|||(\boldsymbol{u} - \boldsymbol{u}_h, \phi - \phi_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h, p - p_h)||| \\
&\qquad \leq C_{\mathrm{rate}} h^{\min\{s+1,m\}} \left(\|\boldsymbol{u}\|_{s+2,\Omega} + \|\phi\|_{s+1,\Omega} + \|\boldsymbol{\sigma}\|_{m,\Omega} + \|\mathrm{div}\,\boldsymbol{\sigma}\|_{m,\Omega} + \|p\|_{m,\Omega} \right).
\end{aligned}
$$

Proof. The result is a straightforward application of Céa's estimate (3.27), and the approximation properties $(\mathbf{AP}_h^{\boldsymbol{u}})$, $(\mathbf{AP}_h^{\phi})$, $(\mathbf{AP}_h^{\boldsymbol{\sigma}})$ and $(\mathbf{AP}_h^{p})$. □

Remark 3.1. *To approximate the solution of problem (3.13), one may consider other finite element subspaces available in the literature. For example, for each $T \in \mathcal{T}_h$, consider the Brezzi–Douglas–Marini space $\mathbf{BDM}_k(T) := \mathbf{P}_k(T)$ of order $k \geq 1$ (see, e.g., [41]), and the enriched space $\mathbf{P}_{1,b}(T) := [\mathrm{P}_1(T) \oplus \mathrm{span}\{b_T\}]^d$, where b_T is the bubble function defined as $b_T := \prod_{i=1}^{d+1} \lambda_i$ and $\{\lambda_i\}$, $1 \leq i \leq d+1$, are the barycentric coordinates of T. The following finite element spaces,*

$$
\begin{aligned}
\mathbf{H}_h &:= \left\{ \boldsymbol{v}_h \in [\mathcal{C}(\overline{\Omega})]^d : \ \boldsymbol{v}_h|_T \in \mathbf{P}_{1,b}(T) \ \ \forall T \in \mathcal{T}_h, \ \boldsymbol{v}_h = \mathbf{0} \quad \text{on} \quad \Gamma_{\boldsymbol{u}} \right\}, \\
\mathrm{Q}_h &:= \left\{ \psi_h \in \mathcal{C}(\overline{\Omega}) : \ \psi_h|_T \in \mathrm{P}_1(T) \ \ \forall T \in \mathcal{T}_h \right\}, \\
\mathbf{Z}_h &:= \left\{ \boldsymbol{\tau}_h \in \mathbf{H}(\mathrm{div};\Omega) : \ \boldsymbol{\tau}_h|_T \in \mathbf{BDM}_k(T) \ \ \forall T \in \mathcal{T}_h, \ \boldsymbol{\tau}_h \cdot \boldsymbol{n} = 0 \quad \text{on} \quad \Gamma_{\boldsymbol{u}} \right\}, \\
\mathrm{W}_h &:= \left\{ q_h \in \mathrm{L}^2(\Omega) : \ q_h|_T \in \mathrm{P}_{k-1}(T) \ \ \forall T \in \mathcal{T}_h \right\},
\end{aligned}
\tag{3.34}
$$

result also in a well–posed Galerkin scheme (3.23) with optimal error bounds. In particular, we recall that $(\mathbf{H}_h, \mathrm{Q}_h)$, which is usually referred to as the MINI-element [11], satisfies the hypothesis **(H2)**. *For its proof in two dimensions, we refer to [11] (see also [41]). The stability of this element in three dimensions follows, as in the two-dimensional case, by using a suitable Fortin operator (see, e.g., [30]).*

The theory developed in this section holds for combinations of the pairs $(\mathbf{H}_h, \mathrm{Q}_h)$ and $(\mathbf{Z}_h, \mathrm{W}_h)$ resulting from the finite element subspaces (3.30)-(3.33) and (3.34).

3.4 A residual-based *a posteriori* error estimator

We now develop a reliable and efficient residual-based *a posteriori* error estimator for the Galerkin scheme (3.23). In doing so, we may use any choice of finite dimensional subspaces satisfying the hypotheses of Section 3.3. For simplicity, however, we consider the finite dimensional subspaces (3.30)-(3.33), and restrict ourselves to the problem in two dimensions. In Section 3.4.3 we will comment on the main consideration for extending the estimator to three dimensions. We begin by introducing further notation and definitions.

For each $T \in \mathcal{T}_h$, we let $\mathcal{E}(T)$ be the set of all edges of T, and denote by $\mathcal{E}_h$ the set of all edges of $\mathcal{T}_h$, that is, $\mathcal{E}_h = \mathcal{E}_h(\Omega) \cup \mathcal{E}_h(\Gamma_{\boldsymbol{u}}) \cup \mathcal{E}_h(\Gamma_p)$, where $\mathcal{E}_h(\Omega) := \{e \in \mathcal{T}_h : \ e \subseteq \Omega\}$, $\mathcal{E}_h(\Gamma_{\boldsymbol{u}}) := \{e \in \mathcal{T}_h : \ e \subseteq \Gamma_{\boldsymbol{u}}\}$ and $\mathcal{E}_h(\Gamma_p) := \{e \in \mathcal{T}_h : \ e \subseteq \Gamma_p\}$. In what follows, h_e stands for the diameter of a given edge

$e \in \mathcal{E}_h$. For every edge $e \in \mathcal{E}_h$ we fix a unit normal vector $\boldsymbol{n}_e := (n_1, n_2)^T$ to the edge e, and let $\boldsymbol{s}_e := (-n_2, n_1)^T$ be the fixed unit tangential vector along e. However, when no confusion arises we will simply write $\boldsymbol{n}$ and $\boldsymbol{s}$ instead of $\boldsymbol{n}_e$ and $\boldsymbol{s}_e$, respectively. Given an edge $e \in \mathcal{E}_h(\Omega)$, $\boldsymbol{\tau} \in \mathbf{L}^2(\Omega)$ and $\boldsymbol{\xi} \in [\mathrm{L}(\Omega)]^{2\times 2}$, such that $\boldsymbol{\tau} \in [\mathcal{C}(T)]^2$ and $\boldsymbol{\xi} \in [\mathcal{C}(T)]^{2\times 2}$ for all $T \in \mathcal{T}_h$, we let $[\![\boldsymbol{\tau} \cdot \boldsymbol{s}]\!]$ and $[\![\boldsymbol{\xi}\boldsymbol{n}]\!]$ be the corresponding jumps across e, i.e., $[\![\boldsymbol{\tau} \cdot \boldsymbol{s}]\!] := \{(\boldsymbol{\tau}|_T)|_e - (\boldsymbol{\tau}|_{T'})|_e\} \cdot \boldsymbol{s}$ and $[\![\boldsymbol{\xi}\boldsymbol{n}]\!] := \{(\boldsymbol{\xi}|_T)|_e - (\boldsymbol{\xi}|_{T'})|_e\}\boldsymbol{n}$, respectively, where T and T' are two triangles of $\mathcal{T}_h$ sharing a common edge e. Finally, given scalar and vector-valued fields ψ and $\boldsymbol{\tau} := (\tau_i)_{1\le i\le 2}$, respectively, we set

$$\mathrm{rot}\,\boldsymbol{\tau} := \frac{\partial \tau_2}{\partial x_1} - \frac{\partial \tau_1}{\partial x_2} \quad \text{and} \quad \mathbf{curl}\,\psi := \begin{pmatrix} \frac{\partial \psi}{\partial x_2} \\ -\frac{\partial \psi}{\partial x_1} \end{pmatrix}.$$

Now, let $(\boldsymbol{u}_h, \phi_h, \boldsymbol{\sigma}_h, p_h) \in \mathbf{H}_h \times \mathrm{Q}_h \times \mathbf{Z}_h \times \mathrm{W}_h$ be the unique solution of problem (3.23) and introduce the global a posteriori error estimator

$$\Theta := \left(\sum_{T \in \mathcal{T}_h} \left\{ \Theta_{s,T}^2 + \Theta_{f,T}^2 + \Theta_{sf,T}^2 \right\} \right)^{1/2}, \tag{3.35}$$

where $\Theta_{s,T}$, $\Theta_{f,T}$ and $\Theta_{sf,T}$ are the local error indicators defined for each $T \in \mathcal{T}_h$ as follows:

$$\begin{aligned} \Theta_{s,T}^2 := {} & h_T^2 \|\boldsymbol{f} + \mathbf{div}\,(2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}_h) - \phi_h \boldsymbol{I})\|_{0,T}^2 + \sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h(\Omega)} h_e \|[\![(2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}_h) - \phi_h \boldsymbol{I})\boldsymbol{n}]\!]\|_{0,e}^2 \\ & + \sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h(\Gamma_p)} h_e \|\boldsymbol{m}_\Gamma - [\![(2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}_h) - \phi_h \boldsymbol{I})\boldsymbol{n}]\!]\|_{0,e}^2, \end{aligned} \tag{3.36}$$

$$\begin{aligned} \Theta_{f,T}^2 := {} & h_T^2 \left\| \nabla p_h - \rho \boldsymbol{g} + \frac{\eta}{\kappa} \boldsymbol{\sigma}_h \right\|_{0,T}^2 + h_T^2 \left\| \mathrm{rot} \left(\frac{\eta}{\kappa} \boldsymbol{\sigma}_h - \rho \boldsymbol{g} \right) \right\|_{0,T}^2 \\ & + \sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h(\Gamma_p)} \left(h_e \|p_\Gamma - p_h\|_{0,e}^2 + h_e \left\| \left(\frac{\eta}{\kappa} \boldsymbol{\sigma}_h - \rho \boldsymbol{g} \right) \cdot \boldsymbol{s} + \frac{dp_\Gamma}{d\boldsymbol{s}} \right\|_{0,e}^2 \right) \\ & + \sum_{e \in \mathcal{E}(T) \cap \mathcal{E}_h(\Omega)} h_e \left\| \left[\!\!\left[\left(\frac{\eta}{\kappa} \boldsymbol{\sigma}_h - \rho \boldsymbol{g} \right) \cdot \boldsymbol{s} \right]\!\!\right] \right\|_{0,e}^2, \end{aligned} \tag{3.37}$$

$$\Theta_{sf,T}^2 := \left\| \frac{1}{\lambda}(\phi_h - \alpha p_h) + \mathrm{div}\,\boldsymbol{u}_h \right\|_{0,T}^2 + \left\| \left(c_0 + \frac{\alpha^2}{\lambda} \right) p_h - \frac{\alpha}{\lambda} \phi_h + \mathrm{div}\,\boldsymbol{\sigma}_h - \ell \right\|_{0,T}^2. \tag{3.38}$$

The residual character of each term defining $(\Theta_{s,T} + \Theta_{f,T} + \Theta_{sf,T})$ is a consequence of the strong problem (3.11) and the regularity of the weak solution at the continuous level. It is important to remark that the third term of $\Theta_{s,T}$ requires $\boldsymbol{m}_\Gamma \in \mathbf{L}^2(e)$ for all $e \in \mathcal{E}_h(\Gamma_p)$, which will be assumed from now on. Similarly, as we will see in Lemma 3.11 (see, in particular, equation (3.58)), we need to assume that $p_\Gamma \in \mathrm{H}^1(\Gamma_p)$. The latter implies that the fourth and fifth terms of $\Theta_{f,T}$ are well-defined.

In what follows we prove the main properties of Θ, namely its reliability and efficiency.

3.4.1 Reliability of the a posteriori error estimator

In this section we focus on the proof of the following result.

Theorem 3.7. *There exists a constant $C_{\rm rel} > 0$, independent of λ and h, such that*

$$|||(\boldsymbol{u}-\boldsymbol{u}_h, \phi-\phi_h, \boldsymbol{\sigma}-\boldsymbol{\sigma}_h, p-p_h)||| \leq C_{\rm rel}\Theta, \tag{3.39}$$

where $|||\cdot|||$ was defined in (3.19).

The proof of Theorem 3.7 will be separated into several steps. We start by providing a preliminary upper bound for the total error, as done in [80]. The idea is to bound the global error by dual norms of the residuals associated with problem (3.23). The following result holds the key to this.

Lemma 3.8. *Let $(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p) \in \mathbf{H}\times \mathrm{Q}\times \mathbf{Z}\times \mathrm{Q}$ and $(\boldsymbol{u}_h, \phi_h, \boldsymbol{\sigma}_h, p_h) \in \mathbf{H}_h\times \mathrm{Q}_h\times \mathbf{Z}_h\times \mathrm{W}_h$ be the unique solutions of problems (3.13) and (3.23), respectively. There exists a constant $C > 0$, independent of λ and h, such that*

$$|||(\boldsymbol{u}-\boldsymbol{u}_h, \phi-\phi_h, \boldsymbol{\sigma}-\boldsymbol{\sigma}_h, p-p_h)||| \leq C\left(\|\mathcal{F}_1\|_{\mathbf{H}'} + \|\mathcal{G}_1\|_{\mathrm{Q}'} + \|\mathcal{F}_2\|_{\mathbf{Z}'} + \|\mathcal{G}_2\|_{\mathrm{Q}'}\right),$$

where $\mathcal{F}_1(\cdot)$ on $\mathbf{H}$, $\mathcal{G}_1(\cdot)$ on Q, $\mathcal{F}_2(\cdot)$ on $\mathbf{Z}$ and $\mathcal{G}_2(\cdot)$ on Q denote the linear functionals defined, respectively, by

$$\mathcal{F}_1(\boldsymbol{v}) := F(\boldsymbol{v}) - a_s(\boldsymbol{u}_h, \boldsymbol{v}) - b_s(\boldsymbol{v}, \phi_h), \tag{3.40}$$

$$\mathcal{G}_1(\psi) := -b_s(\boldsymbol{u}_h, \psi) + c_s(\phi_h, \psi) - b_{sf}(\psi, p_h), \tag{3.41}$$

$$\mathcal{F}_2(\boldsymbol{\tau}) := G(\boldsymbol{\tau}) - a_f(\boldsymbol{\sigma}_h, \boldsymbol{\tau}) - b_f(\boldsymbol{\tau}, p_h), \tag{3.42}$$

$$\mathcal{G}_2(q) := H(q) - b_{sf}(\phi_h, q) - b_f(\boldsymbol{\sigma}_h, q) + c_f(p_h, q). \tag{3.43}$$

Proof. Adding and subtracting $(\boldsymbol{u}_h, \phi_h, \boldsymbol{\sigma}_h, p_h)$ to the continuous solution in system (3.13), the conclusion follows directly from the estimate (3.21) by taking $F_1 = \mathcal{F}_1$, $G_1 = \mathcal{G}_1$, $F_2 = \mathcal{F}_2$ and $G_2 = \mathcal{G}_2$. □

Having proved Lemma 3.8, and noting that $\mathcal{G}_1, \mathcal{G}_2 \in \mathrm{Q}'$ satisfy

$$\|\mathcal{G}_1\|_{\mathrm{Q}'} \leq \left\|\frac{1}{\lambda}(\phi_h - \alpha p_h) + \operatorname{div}\boldsymbol{u}_h\right\|_{0,\Omega} \quad \text{and} \quad \|\mathcal{G}_2\|_{\mathrm{Q}'} \leq \left\|\left(c_0 + \frac{\alpha^2}{\lambda}\right)p_h - \frac{\alpha}{\lambda}\phi_h + \operatorname{div}\boldsymbol{\sigma}_h - \ell\right\|_{0,\Omega}, \tag{3.44}$$

it is clear that in order to show (3.39), we need to obtain suitable upper bounds for $\|\mathcal{F}_1\|_{\mathbf{H}'}$ and $\|\mathcal{F}_2\|_{\mathbf{Z}'}$. From the Galerkin scheme (3.23) we note that $\mathcal{F}_1(\boldsymbol{v}_h) = 0$ for all $\boldsymbol{v}_h \in \mathbf{H}_h$, and $\mathcal{F}_2(\boldsymbol{\tau}_h) = 0$ for all $\boldsymbol{\tau}_h \in \mathbf{Z}_h$. We can therefore write

$$\|\mathcal{F}_1\|_{\mathbf{H}'} := \sup_{\substack{\boldsymbol{v}\in\mathbf{H}\\ \boldsymbol{v}_h\neq\mathbf{0}}} \frac{|\mathcal{F}_1(\boldsymbol{v}-\boldsymbol{v}_h)|}{\|\boldsymbol{v}\|_{1,\Omega}} \tag{3.45}$$

and

$$\|\mathcal{F}_2\|_{\mathbf{Z}'} := \sup_{\substack{\boldsymbol{\tau}\in\mathbf{Z}\\ \boldsymbol{\tau}_h\neq\mathbf{0}}} \frac{|\mathcal{F}_2(\boldsymbol{\tau}-\boldsymbol{\tau}_h)|}{\|\boldsymbol{\tau}\|_{\operatorname{div},\Omega}}, \tag{3.46}$$

with $\boldsymbol{v}_h \in \mathbf{H}_h$ and $\boldsymbol{\tau}_h \in \mathbf{Z}_h$ suitably chosen functions that will be defined later.

Upper bound for $\|\mathcal{F}_1\|_{\mathbf{H}'}$

To satisfy homogeneous Dirichlet boundary conditions, we introduce the Clément-type interpolant

$$\mathcal{I}_{h,\Gamma_{\boldsymbol{u}}} : \mathrm{H}^1_{\Gamma_{\boldsymbol{u}}}(\Omega) \to \mathrm{X}_{h,\Gamma_{\boldsymbol{u}}},$$

where

$$\mathrm{X}_{h,\Gamma_{\boldsymbol{u}}} := \left\{ v \in \mathcal{C}(\overline{\Omega}) : \ v|_T \in \mathrm{P}_1(T) \ \ \forall T \in \mathcal{T}_h, \ \ v = 0 \quad \text{on} \quad \Gamma_{\boldsymbol{u}} \right\} \subseteq \mathrm{H}^1_{\Gamma_{\boldsymbol{u}}}(\Omega),$$

with $\mathrm{H}^1_{\Gamma_{\boldsymbol{u}}}(\Omega)$ defined as in (3.1). It can be shown that this operator satisfies the same approximation properties as the standard Clément interpolant [54], i.e.,

$$\|v - \mathcal{I}_{h,\Gamma_{\boldsymbol{u}}}(v)\|_{0,T} \leq C_1 h_T |v|_{1,\Delta(T)} \quad \forall T \in \mathcal{T}_h, \quad \text{and} \quad \|v - \mathcal{I}_{h,\Gamma_{\boldsymbol{u}}}(v)\|_{0,e} \leq C_2 h_e^{1/2} |v|_{1,\Delta(e)} \quad \forall e \in \mathcal{E}_h, \tag{3.47}$$

where $\Delta(T)$ and $\Delta(e)$ are the union of all the elements intersecting with T and e, respectively. Furthermore, we denote by $\boldsymbol{\mathcal{I}}_{h,\Gamma_{\boldsymbol{u}}}$ the vector operator defined componentwise by $\mathcal{I}_{h,\Gamma_{\boldsymbol{u}}}$.

Next, proceeding analogously to [140, Section 6], we state the main result of this section.

Lemma 3.9. *Assuming that $\boldsymbol{m}_\Gamma \in \mathbf{L}^2(e)$ for all $\mathcal{E}_h(\Gamma_p)$, there exists a constant $C > 0$, independent of λ and h, such that*

$$\|\mathcal{F}_1\|_{\mathbf{H}'} \leq C \left(\sum_{T \in \mathcal{T}_h} \Theta^2_{s,T} \right)^{1/2},$$

where $\Theta_{s,f}$ is defined in (3.36).

Proof. Integrating by parts (3.40) on each $T \in \mathcal{T}_h$ yields for all $\boldsymbol{w} \in \mathbf{H}$,

$$\begin{aligned}
\mathcal{F}_1(\boldsymbol{w}) &= \int_\Omega \boldsymbol{f} \cdot \boldsymbol{w} + \int_{\Gamma_p} \boldsymbol{m}_\Gamma \cdot \boldsymbol{w} - 2\mu \int_\Omega \boldsymbol{\varepsilon}(\boldsymbol{u}_h) : \boldsymbol{\varepsilon}(\boldsymbol{w}) + \int_\Omega \phi_h \, \mathrm{div}\, \boldsymbol{w} \\
&= \sum_{T \in \mathcal{T}_h} \int_T \boldsymbol{f} \cdot \boldsymbol{w} + \sum_{e \in \mathcal{E}_h(\Gamma_p)} \int_e \boldsymbol{m}_\Gamma \cdot \boldsymbol{w} - \sum_{T \in \mathcal{T}_h} \int_T (2\mu \boldsymbol{\varepsilon}(\boldsymbol{u}_h) - \phi_h \boldsymbol{I}) : \nabla \boldsymbol{w} \\
&= \sum_{e \in \mathcal{E}_h(\Gamma_p)} \int_e \boldsymbol{m}_\Gamma \cdot \boldsymbol{w} + \sum_{T \in \mathcal{T}_h} \left(\int_T (\boldsymbol{f} + \mathbf{div}\,(2\mu \boldsymbol{\varepsilon}(\boldsymbol{u}_h) - \phi_h \boldsymbol{I})) \cdot \boldsymbol{w} - \int_{\partial T} (2\mu \boldsymbol{\varepsilon}(\boldsymbol{u}_h) - \phi_h \boldsymbol{I}) \boldsymbol{n} \cdot \boldsymbol{w} \right) \\
&= \sum_{T \in \mathcal{T}_h} \int_T (\boldsymbol{f} + \mathbf{div}\,(2\mu \boldsymbol{\varepsilon}(\boldsymbol{u}_h) - \phi_h \boldsymbol{I})) \cdot \boldsymbol{w} + \sum_{e \in \mathcal{E}_h(\Gamma_p)} \int_e (\boldsymbol{m}_\Gamma - (2\mu \boldsymbol{\varepsilon}(\boldsymbol{u}_h) - \phi_h \boldsymbol{I}) \boldsymbol{n}) \cdot \boldsymbol{w} \\
&\quad - \sum_{e \in \mathcal{E}_h(\Omega)} \int_e [\![(2\mu \boldsymbol{\varepsilon}(\boldsymbol{u}_h) - \phi_h \boldsymbol{I}) \boldsymbol{n}]\!] \cdot \boldsymbol{w}.
\end{aligned}$$

Given $\boldsymbol{v} \in \mathbf{H}$, set $\boldsymbol{v}_h$ in (3.45) to $\boldsymbol{v}_h := \boldsymbol{\mathcal{I}}_{h,\Gamma_{\boldsymbol{u}}}(\boldsymbol{v})$ and let $\boldsymbol{w} := \boldsymbol{v} - \boldsymbol{v}_h$. Then, applying the Cauchy–Schwarz inequality to each term above, and by the approximation properties of $\boldsymbol{\mathcal{I}}_{h,\Gamma_{\boldsymbol{u}}}$ (cf. (3.47)), we obtain

$$|\mathcal{F}_1(\boldsymbol{w})| \leq C \left(\sum_{T \in \mathcal{T}_h} \Theta^2_{s,T} \right)^{1/2} \left(\sum_{T \in \mathcal{T}_h} \|\boldsymbol{v}\|^2_{1,\Delta(T)} + \sum_{e \in \mathcal{E}_h(\Omega)} \|\boldsymbol{v}\|^2_{1,\Delta(e)} + \sum_{e \in \mathcal{E}_h(\Gamma_p)} \|\boldsymbol{v}\|^2_{1,\Delta(e)} \right)^{1/2}.$$

The result follows by using the definition of $\mathcal{F}_1$, and noting, by the shape-regularity of the mesh, that the number of triangles in $\Delta(T)$ and $\Delta(e)$ is bounded. □

Upper bound for $\|\mathcal{F}_2\|_{\mathbf{Z}'}$

In this section, a stable Helmholtz decomposition of $\mathbf{Z}$ and suitable interpolation operators will be of paramount importance to define $\boldsymbol{\tau}_h$ appearing in definition (3.46). This term is necessary to provide an upper bound for $\|\mathcal{F}_2\|_{\mathbf{Z}'}$. The approach we follow has been widely used in *a posteriori* error estimators for mixed methods, see for instance [3, 49, 78].

We start by introducing the $\mathrm{L}^2(\Omega)$-orthogonal projection onto W_h (cf. (3.33)), $\mathcal{P}_h^k : \mathrm{L}^2(\Omega) \to \mathrm{W}_h$, which, for each $q \in \mathrm{H}^m(\Omega)$, with $0 \leq m \leq k+1$, satisfies the approximation property

$$|q - \mathcal{P}_h^k(q)|_{s,T} \leq C h^{m-s} |q|_{m,T} \quad \forall\, T \in \mathcal{T}_h,\ \forall\, s \in \{0, \ldots, m\}. \tag{3.48}$$

In addition, letting

$$\mathbf{Z}_h^{RT} := \left\{ \boldsymbol{\tau}_h \in \mathbf{H}(\mathrm{div}; \Omega) : \ \boldsymbol{\tau}_h|_T \in \mathbf{RT}_k(T) \ \ \forall\, T \in \mathcal{T}_h \right\},$$

we recall the classical Raviart–Thomas interpolation operator $\boldsymbol{\Pi}_h^k : \mathbf{H}^1(\Omega) \to \mathbf{Z}_h^{RT}$, which, given $\boldsymbol{\tau} \in \mathbf{H}^1(\Omega)$, is characterized by the identities

$$\int_T \boldsymbol{\Pi}_h^k(\boldsymbol{\tau}) \cdot \boldsymbol{\zeta} = \int_T \boldsymbol{\tau} \cdot \boldsymbol{\zeta} \qquad \forall\, \boldsymbol{\zeta} \in \mathbf{P}_{k-1}(T),\ \forall\, T \in \mathcal{T}_h, \ \text{ when } \ k \geq 1, \tag{3.49}$$

$$\int_e (\boldsymbol{\Pi}_h^k(\boldsymbol{\tau}) \cdot \boldsymbol{n}) \psi = \int_e (\boldsymbol{\tau} \cdot \boldsymbol{n}) \psi \quad \forall\, \psi \in \mathrm{P}_k(e),\ \forall\, e \in \mathcal{E}_h, \ \text{ when } \ k \geq 0. \tag{3.50}$$

Consequently, it is not difficult to check (see, e.g., [73, Lemma 3.7]) that

$$\mathrm{div}\,(\boldsymbol{\Pi}_h^k(\boldsymbol{\tau})) = \mathcal{P}_h^k(\mathrm{div}\,\boldsymbol{\tau}) \quad \forall\, \boldsymbol{\tau} \in \mathbf{H}^1(\Omega). \tag{3.51}$$

Moreover, the following local approximation properties hold [41, 53, 73]:

- For each $\boldsymbol{\tau} \in \mathbf{H}^m(\Omega)$, with $0 \leq m \leq k+1$,

$$\|\boldsymbol{\tau} - \boldsymbol{\Pi}_h^k(\boldsymbol{\tau})\|_{0,T} \leq C h_T^m |\boldsymbol{\tau}|_{m,T} \quad \forall\, T \in \mathcal{T}_h, \tag{3.52}$$

- For each $\boldsymbol{\tau} \in \mathbf{H}^1(\Omega)$ such that $\mathrm{div}\,\boldsymbol{\tau} \in \mathrm{H}^m(\Omega)$, with $0 \leq m \leq k+1$,

$$\|\mathrm{div}\,(\boldsymbol{\tau} - \boldsymbol{\Pi}_h^k(\boldsymbol{\tau}))\|_{0,T} \leq C h_T^m |\mathrm{div}\,\boldsymbol{\tau}|_{m,T} \quad \forall\, T \in \mathcal{T}_h, \tag{3.53}$$

- For each $\boldsymbol{\tau} \in \mathbf{H}^1(\Omega)$, there holds

$$\|(\boldsymbol{\tau} - \boldsymbol{\Pi}_h^k(\boldsymbol{\tau})) \cdot \boldsymbol{n}\|_{0,e} \leq C h_e^{1/2} |\boldsymbol{\tau}|_{1,T_e}, \tag{3.54}$$

  where T_e denotes an element of $\mathcal{T}_h$ having e as an edge.

We now introduce a stable Helmholtz decomposition of $\mathbf{Z}$. This will require $\Gamma_{\boldsymbol{u}}$ to lie on the boundary of a convex domain containing Ω. We refer to [3, Lemma 3.9] for the proof of this result in the tensorial case.

Lemma 3.10. *Assume that there exists a convex domain Ξ such that $\overline{\Omega} \subseteq \Xi$ and $\Gamma_{\boldsymbol{u}} \subseteq \partial\Xi$. Then, for each $\boldsymbol{\tau} \in \mathbf{Z}$ there exist $\boldsymbol{\zeta} \in \mathbf{H}^1(\Omega)$ and $\varphi \in \mathrm{H}^1_{\Gamma_{\boldsymbol{u}}}(\Omega)$, such that*

$$\boldsymbol{\tau} = \boldsymbol{\zeta} + \mathbf{curl}\,\varphi \quad \text{in} \quad \Omega, \quad \text{and} \quad \|\boldsymbol{\zeta}\|_{1,\Omega} + \|\varphi\|_{1,\Omega} \leq C\|\boldsymbol{\tau}\|_{\mathrm{div},\Omega}, \tag{3.55}$$

where C is a positive constant independent of $\boldsymbol{\tau}$, $\boldsymbol{\zeta}$ and φ.

We now introduce the discrete version of (3.55) and follow similar steps as in [80, Lemma 3.8] (see also [78, Section 4.1]). Given $\boldsymbol{\tau} \in \mathbf{Z}$ and its Helmholtz decomposition (3.55), we let $\boldsymbol{\zeta}_h := \boldsymbol{\Pi}^k_h(\boldsymbol{\zeta})$ and $\varphi_h := \mathcal{I}_{h,\Gamma_{\boldsymbol{u}}}(\varphi)$, where $\mathcal{I}_{h,\Gamma_{\boldsymbol{u}}}$ is the Clément-type interpolant given in Section 3.4.1. We then set the discrete Helmholtz decomposition as $\boldsymbol{\tau}_h := \boldsymbol{\zeta}_h + \mathbf{curl}\,\varphi_h \in \mathbf{Z}_h$.

From the above discussion and by definition of $\mathcal{F}_2$ (cf. (3.42)), we can write

$$\mathcal{F}_2(\boldsymbol{\tau} - \boldsymbol{\tau}_h) = \mathcal{F}_2(\boldsymbol{\zeta} - \boldsymbol{\zeta}_h) + \mathcal{F}_2(\mathbf{curl}\,(\varphi - \varphi_h)). \tag{3.56}$$

We will bound each term on the right-hand side of (3.56) separately.

Proceeding as in the proof of [78, Lemma 4.4], applying the Cauchy–Schwarz inequality, using the identities (3.49)-(3.51), the approximation properties (3.52) and (3.54), and the fact that the number of triangles in $\Delta(T)$ and $\Delta(e)$ is bounded (due to shape-regularity of the mesh), we obtain, after some algebraic manipulations,

$$|\mathcal{F}_2(\boldsymbol{\zeta} - \boldsymbol{\zeta}_h)| \leq C \left(\sum_{T \in \mathcal{T}_h} h_T^2 \left\| \nabla p_h - \rho \boldsymbol{g} + \frac{\eta}{\kappa} \boldsymbol{\sigma}_h \right\|^2_{0,T} + \sum_{e \in \mathcal{E}_h(\Gamma_p)} h_e \|p_\Gamma - p_h\|^2_{0,e} \right)^{1/2} \|\boldsymbol{\zeta}\|_{1,\Omega}. \tag{3.57}$$

The upper bound for $|\mathcal{F}_2(\mathbf{curl}\,(\varphi - \varphi_h))|$ follows by similar arguments as in [78, Lemma 4.3]. Indeed, using the identity $\mathbf{curl}\,(\varphi - \varphi_h) \cdot \boldsymbol{n} = \frac{d}{d\boldsymbol{s}}(\varphi - \varphi_h)$, assuming $\frac{dp_\Gamma}{d\boldsymbol{s}} \in \mathrm{L}^2(\Gamma_p)$, and integrating by parts on Γ_p (see [68, Lemma 3.5, eq. (3.34)]), we obtain

$$\langle \mathbf{curl}\,(\varphi - \varphi_h) \cdot \boldsymbol{n}, p_\Gamma \rangle_{\Gamma_p} = -\left\langle \frac{dp_\Gamma}{d\boldsymbol{s}}, \varphi - \varphi_h \right\rangle_{\Gamma_p} = -\sum_{e \in \mathcal{E}_h(\Gamma_p)} \int_e (\varphi - \varphi_h) \frac{dp_\Gamma}{d\boldsymbol{s}}. \tag{3.58}$$

We can then write $\mathcal{F}_2(\mathbf{curl}\,(\varphi - \varphi_h))$, using (3.58) and applying [84, Theorem 2.11] to integrate by parts elementwise, as

$$\begin{aligned}
\mathcal{F}_2(\mathbf{curl}\,(\varphi - \varphi_h)) &= -\int_\Omega \left(\frac{\eta}{\kappa} \boldsymbol{\sigma}_h - \rho \boldsymbol{g} \right) \cdot \mathbf{curl}\,(\varphi - \varphi_h) - \langle \mathbf{curl}\,(\varphi - \varphi_h) \cdot \boldsymbol{n}, p_\Gamma \rangle_{\Gamma_p} \\
&= -\sum_{T \in \mathcal{T}_h} \int_T \mathrm{rot} \left(\frac{\eta}{\kappa} \boldsymbol{\sigma}_h - \rho \boldsymbol{g} \right) (\varphi - \varphi_h) + \sum_{e \in \mathcal{E}_h(\Omega)} \int_e \left[\!\!\left[\left(\frac{\eta}{\kappa} \boldsymbol{\sigma}_h - \rho \boldsymbol{g} \right) \cdot \boldsymbol{s} \right]\!\!\right] (\varphi - \varphi_h) \\
&\quad + \sum_{e \in \mathcal{E}_h(\Gamma_p)} \int_e \left\{ \left(\frac{\eta}{\kappa} \boldsymbol{\sigma}_h - \rho \boldsymbol{g} \right) \cdot \boldsymbol{s} + \frac{dp_\Gamma}{d\boldsymbol{s}} \right\} (\varphi - \varphi_h).
\end{aligned}$$

Next, applying the Cauchy–Schwarz inequality, using (3.47), and the shape-regularity of the mesh, it follows that

$$\begin{aligned}
|\mathcal{F}_2(\mathbf{curl}\,(\varphi - \varphi_h))| \leq C \Bigg(&\sum_{T \in \mathcal{T}_h} h_T^2 \left\| \mathrm{rot} \left(\frac{\eta}{\kappa} \boldsymbol{\sigma}_h - \rho \boldsymbol{g} \right) \right\|^2_{0,T} + \sum_{e \in \mathcal{E}_h(\Omega)} h_e \left\| \left[\!\!\left[\left(\frac{\eta}{\kappa} \boldsymbol{\sigma}_h - \rho \boldsymbol{g} \right) \cdot \boldsymbol{s} \right]\!\!\right] \right\|^2_{0,e} \\
&+ \sum_{e \in \mathcal{E}_h(\Gamma_p)} h_e \left\| \left(\frac{\eta}{\kappa} \boldsymbol{\sigma}_h - \rho \boldsymbol{g} \right) \cdot \boldsymbol{s} + \frac{dp_\Gamma}{d\boldsymbol{s}} \right\|^2_{0,e} \Bigg)^{1/2} \|\varphi\|_{1,\Omega}.
\end{aligned} \tag{3.59}$$

Finally, combining (3.57) and (3.59), and using the stability of the Helmholtz decomposition (3.55), we obtain the desired bound as summarized in the next lemma.

Lemma 3.11. *Suppose that the hypotheses of Lemma 3.10 hold. Assume further that* $p_\Gamma \in \mathrm{H}^1(\Gamma_p)$. *Then, there exists* $C > 0$, *independent of* λ *and* h, *such that*

$$\|\mathcal{F}_2\|_{\mathbf{Z}'} \leq C \left(\sum_{T \in \mathcal{T}_h} \Theta_{f,T}^2 \right)^{1/2},$$

with $\Theta_{f,T}$ *defined in* (3.37).

We end this section by noting that the reliability estimate (3.39) is a direct consequence of Lemmas 3.9 and 3.11, and the estimates given by (3.44)

3.4.2 Efficiency of the a posteriori error estimator

The main result of this section reads as follows.

Theorem 3.12. *There exists a constant* $C_{\mathrm{eff}} > 0$, *independent of* λ *and* h, *such that*

$$C_{\mathrm{eff}}\Theta \leq |||(\boldsymbol{u} - \boldsymbol{u}_h, \phi - \phi_h, \boldsymbol{\sigma} - \boldsymbol{\sigma}_h, p - p_h)||| + \text{h.o.t.}, \tag{3.60}$$

where h.o.t. *is a generic expression denoting one or several terms of higher order.*

To obtain (3.60), we will find upper bounds for each estimator term in (3.36), (3.37) and (3.38), separately. We can immediately deduce the estimates for the zero-order terms appearing in the definition of $\Theta_{sf,T}$ (cf. (3.38)), as done in the following lemma.

Lemma 3.13. *For all* $T \in \mathcal{T}_h$, *there hold*

$$\left\| \frac{1}{\lambda}(\phi_h - \alpha p_h) + \operatorname{div} \boldsymbol{u}_h \right\|_{0,T} \leq \sqrt{2}\|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,T} + \frac{1}{\lambda}\|\phi - \phi_h\|_{0,T} + \frac{\alpha}{\lambda}\|p - p_h\|_{0,T},$$

and

$$\left\| \left(c_0 + \frac{\alpha^2}{\lambda} \right) p_h - \frac{\alpha}{\lambda}\phi_h + \operatorname{div} \boldsymbol{\sigma}_h - \ell \right\|_{0,T} \leq \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\operatorname{div},T} + \frac{\alpha}{\lambda}\|\phi - \phi_h\|_{0,T} + \left(c_0 + \frac{\alpha^2}{\lambda} \right) \|p - p_h\|_{0,T}.$$

Note that volumetric locking is not a concern in the above two inequalities, because at least one term on the right-hand side does not vanish when $\lambda \to \infty$.

To bound the remaining terms, we introduce further notation and preliminary results. Given $T \in \mathcal{T}_h$ and $e \in \mathcal{E}(T)$, we let Φ_T and Φ_e be the usual element-bubble and edge-bubble functions [138], respectively. In particular, Φ_T satisfies $\Phi_T \in \mathrm{P}_3(T)$, $\sup \Phi_T \subseteq T$, $\Phi_T = 0$ on ∂T and $0 \leq \Phi_T \leq 1$ in T. Similarly, one has $\Phi_e|_T \in \mathrm{P}_2(T)$, $\sup \Phi_e \subseteq \omega_e := \cup \{T' \in \mathcal{T}_h : \ e \in \mathcal{E}(T')\}$, $\Phi_e = 0$ on $\partial T \setminus \{e\}$ and $0 \leq \Phi_e \leq 1$ in ω_e. We then have the following useful result.

Lemma 3.14. *Given an integer $k \geq 0$, there exists an extension operator $\mathcal{L} : \mathcal{C}(e) \to \mathcal{C}(T)$ such that $\mathcal{L}(q)|_e = q$ for all $q \in \mathrm{P}_k(e)$. Moreover, there exist positive constants γ_i, $i \in \{1,2,3,4\}$, which only depend on k and on the shape-regularity parameter of the mesh, such that for each $T \in \mathcal{T}_h$ and each $e \in \mathcal{E}(T)$,*

$$\|\Phi_T \psi\|_{0,T}^2 \leq \|\psi\|_{0,T}^2 \leq \gamma_1 \|\Phi_T^{1/2} \psi\|_{0,T}^2 \quad \forall\, \psi \in \mathrm{P}_k(T), \tag{3.61}$$

$$\|\Phi_e \mathcal{L}(q)\|_{0,e}^2 \leq \|q\|_{0,e}^2 \leq \gamma_2 \|\Phi_e^{1/2} q\|_{0,e}^2 \quad \forall\, q \in \mathrm{P}_k(e), \tag{3.62}$$

and

$$\gamma_3 h_e^{1/2} \|q\|_{0,e} \leq \|\Phi_e^{1/2} \mathcal{L}(q)\|_{0,T} \leq \gamma_4 h_e^{1/2} \|q\|_{0,e} \quad \forall\, q \in \mathrm{P}_k(e). \tag{3.63}$$

Proof. See [138, Lemma 4.1] or [139, Lemma 3.3] for details. □

The following inverse estimate will also be used.

Lemma 3.15. *Let $k, m, l \in \mathbb{N} \cup \{0\}$ such that $l \leq m$. There exists a constant $C > 0$, depending only on k, m, l and the shape-regularity constant of the mesh, such that for each $T \in \mathcal{T}_h$ there holds*

$$|q|_{m,T} \leq C_{\mathrm{inv}} h_T^{l-m} |q|_{l,T} \quad \forall\, q \in \mathrm{P}_k(T). \tag{3.64}$$

Proof. See [53, Theorem 3.2.6]. □

Furthermore, we will need the following trace inequality (see, e.g., [9]):

$$\|v\|_{0,e} \leq C_{\mathrm{tr}} \left(h_e^{-1/2} \|v\|_{0,T_e} + h_e^{1/2} |v|_{1,T_e} \right) \quad \forall\, v \in \mathrm{H}^1(T_e). \tag{3.65}$$

Above, T_e is the mesh element introduced in (3.54). Moreover, the constant $C_{\mathrm{tr}} > 0$ depends only on the minimum angle of T_e.

In what follows, considering $\boldsymbol{\sigma}_h$ the approximate fluid flux in problem (3.23), we often write $\boldsymbol{\xi} := \frac{\eta}{\kappa} \boldsymbol{\sigma}_h$ and assume, for simplicity, that for $r, m \geq k+2$, the permeability satisfies: $\kappa^{-1}|_T \in \mathrm{H}^{r+1}(T)$ for all $T \in \mathcal{T}_h$, and $\kappa^{-1}|_e \in \mathrm{H}^{m+1}(e)$ for all $e \in \mathcal{E}_h$. Furthermore, the vector counterpart of the projection operator $\mathcal{P}_h^k$ (cf. (3.48)) will be denoted in boldface.

The following three lemmas provide upper bounds for the estimator terms in (3.37). We present here proofs inspired by the proofs of Lemmas 6.10, 6.11 and 6.12 in [50]. Similar ideas can be found in [48].

Lemma 3.16. *There exists a constant $c_1 > 0$, independent of λ and h, such that for all $T \in \mathcal{T}_h$,*

$$h_T \left\| \mathrm{rot} \left(\frac{\eta}{\kappa} \boldsymbol{\sigma}_h - \rho \boldsymbol{g} \right) \right\|_{0,T} \leq c_1 \left(\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathrm{div},T} + \mathrm{h.o.t.} \right). \tag{3.66}$$

Proof. Adding and subtracting $\boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi})$, and using the triangle inequality, there holds

$$\|\mathrm{rot}\ (\boldsymbol{\xi} - \rho \boldsymbol{g})\|_{0,T} \leq C |\boldsymbol{\xi} - \boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi})|_{1,T} + \|\mathrm{rot}\ (\boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi}) - \rho \boldsymbol{g})\|_{0,T}. \tag{3.67}$$

Applying now (3.61) to the second term on the right-hand side of (3.67), and noting, by Lemma 3.3, that $\rho\boldsymbol{g} = \nabla p + \boldsymbol{\xi} + \frac{\eta}{\kappa}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)$ in Ω, we obtain

$$\begin{aligned}\|\mathrm{rot}\,(\boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi}) - \rho\boldsymbol{g})\|_{0,T}^2 &\leq \gamma_1 \left\|\Phi_T^{1/2}\,\mathrm{rot}\,(\boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi}) - \rho\boldsymbol{g})\right\|_{0,T}^2 = \gamma_1 \int_T \Phi_T\,(\mathrm{rot}\,(\boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi}) - \rho\boldsymbol{g}))^2 \\ &= \gamma_1 \int_T \Phi_T\,\mathrm{rot}\,(\boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi}) - \rho\boldsymbol{g})\;\mathrm{rot}\left(\boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi}) - \boldsymbol{\xi} - \frac{\eta}{\kappa}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\right) \\ &= \gamma_1 \int_T \mathbf{curl}\,(\Phi_T\,\mathrm{rot}\,(\boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi}) - \rho\boldsymbol{g}))\cdot\left(\boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi}) - \boldsymbol{\xi} - \frac{\eta}{\kappa}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\right).\end{aligned}$$

It then follows from (3.61) and (3.64) that

$$\|\mathrm{rot}\,(\boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi}) - \rho\boldsymbol{g})\|_{0,\Omega} \leq C_{\mathrm{inv}}\gamma_1 h_T^{-1}\left(\|\boldsymbol{\xi} - \boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi})\|_{0,T} + \left\|\frac{\eta}{\kappa}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\right\|_{0,T}\right). \tag{3.68}$$

Substituting (3.68) into (3.67), using the lower bound for κ, and applying the approximation property of $\boldsymbol{\mathcal{P}}_h^r$ in (3.48), yields

$$h_T\,\|\mathrm{rot}\,(\boldsymbol{\xi} - \rho\boldsymbol{g})\|_{0,T} \leq \widetilde{C}\left(\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathrm{div},T} + h_T^{r+1}|\boldsymbol{\xi}|_{r+1,T}\right).$$

Since $r \geq k+2$, the result follows. □

Lemma 3.17. *There exists a constant $c_2 > 0$, independent of λ and h, such that for all $T \in \mathcal{T}_h$,*

$$h_T\left\|\nabla p_h - \rho\boldsymbol{g} + \frac{\eta}{\kappa}\boldsymbol{\sigma}_h\right\|_{0,T} \leq c_2\,(h_T\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathrm{div},T} + \|p - p_h\|_{0,T} + \text{h.o.t.}). \tag{3.69}$$

Proof. First, adding and subtracting $\boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi})$, it follows that

$$\|\nabla p_h - \rho\boldsymbol{g} + \boldsymbol{\xi}\|_{0,T} \leq \|\nabla p_h - \rho\boldsymbol{g} + \boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi})\|_{0,T} + \|\boldsymbol{\xi} - \boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi})\|_{0,T}. \tag{3.70}$$

To bound the first term on the right-hand side of (3.70), we apply estimate (3.61), integrate by parts, and use the identity $\rho\boldsymbol{g} = \nabla p + \boldsymbol{\xi} + \frac{\eta}{\kappa}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)$ in Ω, to obtain

$$\begin{aligned}\|\nabla p_h - \rho\boldsymbol{g} + \boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi})\|_{0,T}^2 &\leq \gamma_1\left\|\Phi_T^{1/2}\,(\nabla p_h - \rho\boldsymbol{g} + \boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi}))\right\|_{0,T}^2 \\ &= \gamma_1\int_T \Phi_T\,(\nabla p_h - \rho\boldsymbol{g} + \boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi}))\cdot\nabla(p_h - p) \\ &\quad - \gamma_1\int_T \Phi_T\,(\nabla p_h - \rho\boldsymbol{g} + \boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi}))\cdot\left(\frac{\eta}{\kappa}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h) + \boldsymbol{\xi} - \boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi})\right) \\ &= -\gamma_1\int_T (p_h - p)\,\mathrm{div}\,(\Phi_T\,(\nabla p_h - \rho\boldsymbol{g} + \boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi}))) \\ &\quad - \gamma_1\int_T \Phi_T\,(\nabla p_h - \rho\boldsymbol{g} - \boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi}))\cdot\left(\frac{\eta}{\kappa}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h) + \boldsymbol{\xi} - \boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi})\right).\end{aligned}$$

Using the Cauchy–Schwarz inequality and the estimates (3.61) and (3.64), it follows that

$$\|\nabla p_h - \rho\boldsymbol{g} + \boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi})\|_{0,T} \leq C\left(h_T^{-1}\,\|p_h - p\|_{0,T} + \left\|\frac{\eta}{\kappa}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\right\|_{0,T} + \|\boldsymbol{\xi} - \boldsymbol{\mathcal{P}}_h^r(\boldsymbol{\xi})\|_{0,T}\right),$$

where $C > 0$ is independent of λ and h. Combined with (3.70) we obtain estimate (3.69). □

Lemma 3.18. *There exists a constant $c_3 > 0$, independent of λ and h, such that for all $e \in \mathcal{E}_h(\Omega)$,*

$$h_e^{1/2} \left\| \left[\!\!\left[\left(\frac{\eta}{\kappa} \boldsymbol{\sigma}_h - \rho \boldsymbol{g} \right) \cdot \boldsymbol{s} \right]\!\!\right] \right\|_{0,e} \leq c_3 \sum_{T \subseteq \omega_e} \left(\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathrm{div},T} + \text{h.o.t.} \right). \tag{3.71}$$

Furthermore, assuming that p_Γ is a piecewise polynomial, there exist constants $c_4, c_5 > 0$, also independent of λ and h, such that for all $e \in \mathcal{E}_h(\Gamma_p)$,

$$h_e^{1/2} \left\| \left(\frac{\eta}{\kappa} \boldsymbol{\sigma}_h - \rho \boldsymbol{g} \right) \cdot \boldsymbol{s} + \frac{dp_\Gamma}{d\boldsymbol{s}} \right\|_{0,e} \leq c_4 \left(\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathrm{div},T} + \text{h.o.t.} \right), \tag{3.72}$$

$$h_e^{1/2} \|p_\Gamma - p_h\|_{0,e} \leq c_5 \left(\|p - p_h\|_{0,T} + (1 + h_T) \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathrm{div},T} + \text{h.o.t} \right). \tag{3.73}$$

Proof. Let us first prove (3.71). In order to simplify notation, given $e \in \mathcal{E}_h(\Omega)$, we decompose $[\![(\boldsymbol{\xi} - \rho \boldsymbol{g}) \cdot \boldsymbol{s}]\!]$ into $\chi_e := [\![(\boldsymbol{\xi} - \boldsymbol{\mathcal{P}}_h^m(\boldsymbol{\xi})) \cdot \boldsymbol{s}]\!]$ and $\zeta_e := [\![(\boldsymbol{\mathcal{P}}_h^m(\boldsymbol{\xi}) - \rho \boldsymbol{g}) \cdot \boldsymbol{s}]\!]$. Applying now the estimate (3.65) and using similar arguments as in the previous two lemmas,

$$\begin{aligned} \|[\![(\boldsymbol{\xi} - \rho \boldsymbol{g}) \cdot \boldsymbol{s}]\!]\|_{0,e} &\leq \|\chi_e\|_{0,e} + \|\zeta_e\|_{0,e} \\ &\leq \sum_{T \subseteq \omega_e} C_{\mathrm{tr}} \left(h_e^{-1/2} \|\boldsymbol{\xi} - \boldsymbol{\mathcal{P}}_h^m(\boldsymbol{\xi})\|_{0,T} + h_e^{1/2} |\boldsymbol{\xi} - \boldsymbol{\mathcal{P}}_h^m(\boldsymbol{\xi})|_{1,T} \right) + \|\zeta_e\|_{0,e} \\ &\leq h_e^{-1/2} \sum_{T \subseteq \omega_e} C_{\mathrm{tr}} \left(\|\boldsymbol{\xi} - \boldsymbol{\mathcal{P}}_h^m(\boldsymbol{\xi})\|_{0,T} + h_e |\boldsymbol{\xi} - \boldsymbol{\mathcal{P}}_h^m(\boldsymbol{\xi})|_{1,T} \right) + \|\zeta_e\|_{0,e} \\ &\leq C h_e^{-1/2} \sum_{T \subseteq \omega_e} h_T^{m+1} |\boldsymbol{\xi}|_{m+1,T} + \|\zeta_e\|_{0,e}, \end{aligned} \tag{3.74}$$

where we recall that $\omega_e := \bigcup \{T' \in \mathcal{T}_h : \ e \in \mathcal{E}(T')\}$. To estimate $\|\zeta_e\|_{0,e}$, we use the second inequality in (3.62), integrate by parts, and use the identity $\rho \boldsymbol{g} = \nabla p + \boldsymbol{\xi} + \frac{\eta}{\kappa}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)$ in Ω. This yields

$$\begin{aligned} \|\zeta_e\|_{0,e}^2 &\leq \gamma_2 \|\Phi_T^{1/2} \zeta_e\|_{0,e}^2 = \gamma_2 \int_e \left(\Phi_e \mathcal{L}(\zeta_e) \right) \zeta_e \\ &= \sum_{T \subseteq \omega_e} \left(\int_T \Phi_e \mathcal{L}(\zeta_e) \, \mathrm{rot} \, (\boldsymbol{\mathcal{P}}_h^m(\boldsymbol{\xi}) - \rho \boldsymbol{g}) - \int_T (\boldsymbol{\mathcal{P}}_h^m(\boldsymbol{\xi}) - \rho \boldsymbol{g}) \cdot \mathbf{curl} \left(\Phi_e \mathcal{L}(\zeta_e) \right) \right) \\ &= \sum_{T \subseteq \omega_e} \left(\int_T \Phi_e \mathcal{L}(\zeta_e) \, \mathrm{rot} \, (\boldsymbol{\mathcal{P}}_h^m(\boldsymbol{\xi}) - \rho \boldsymbol{g}) - \int_T \left(\boldsymbol{\mathcal{P}}_h^m(\boldsymbol{\xi}) - \boldsymbol{\xi} - \frac{\eta}{\kappa}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h) - \nabla p \right) \cdot \mathbf{curl} \left(\Phi_e \mathcal{L}(\zeta_e) \right) \right), \end{aligned}$$

where clearly $\int_T \nabla p \cdot \mathbf{curl} \left(\Phi_e \mathcal{L}(\zeta_e) \right) = 0$ for all $T \subseteq \omega_e$. Using the Cauchy–Schwarz inequality and the inverse estimate (3.64), it follows that

$$\|\zeta_e\|_{0,e}^2 \leq \widetilde{C} \sum_{T \subseteq \omega_e} h_T^{-1} \left(h_T \, \|\mathrm{rot} \, (\boldsymbol{\mathcal{P}}_h^m(\boldsymbol{\xi}) - \rho \boldsymbol{g})\|_{0,T} + \|\boldsymbol{\xi} - \boldsymbol{\mathcal{P}}_h^m(\boldsymbol{\xi})\|_{0,T} + \left\| \frac{\eta}{\kappa}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h) \right\|_{0,T} \right) \|\Phi_e \mathcal{L}(\zeta_e)\|_{0,T}. \tag{3.75}$$

Furthermore, by (3.63) and by construction of Φ_e, we obtain

$$\|\Phi_e \mathcal{L}(\zeta_e)\|_{0,T} \leq \|\Phi_e^{1/2} \mathcal{L}(\zeta_e)\|_{0,T} \leq \gamma_4 h_e^{1/2} \|\zeta_e\|_{0,e}.$$

This, together with estimates (3.48), (3.66) and (3.75), and the fact that $h_e \leq h_T$ for all $T \subset \omega_e$, gives

$$\|\zeta_e\|_{0,e} \leq \widehat{C} h_e^{-1/2} \sum_{T \subseteq \omega_e} \left(\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathrm{div},T} + h_T^{m+1} |\boldsymbol{\xi}|_{m+1,T} \right). \tag{3.76}$$

The result (3.71) follows by combining (3.74) and (3.76).

To prove (3.72), we proceed as in the proof of (3.71). Given $e \in \mathcal{E}_h(\Gamma_p)$, we let $\varrho_e := \boldsymbol{\mathcal{P}}_h^m(\boldsymbol{\xi}) - \rho\boldsymbol{g} - \frac{dp_\Gamma}{ds}$. Since p_Γ is assumed to be a piecewise polynomial, we use similar arguments as in (3.74) to obtain

$$\begin{aligned}\|\varrho_e\|_{0,e}^2 \leq \gamma_2 \|\Phi_T^{1/2}\varrho_e\|_{0,e}^2 &= \gamma_2 \int_e (\Phi_e \mathcal{L}(\varrho_e))\,\varrho_e \\ &= \int_{T^e} \Phi_e \mathcal{L}(\varrho_e)\,\mathrm{rot}\left(\boldsymbol{\mathcal{P}}_h^m(\boldsymbol{\xi}) - \boldsymbol{\xi} - \frac{\eta}{\kappa}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\right) \\ &\quad - \int_{T_e} \left(\boldsymbol{\mathcal{P}}_h^m(\boldsymbol{\xi}) - \boldsymbol{\xi} - \frac{\eta}{\kappa}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h) - \nabla p\right) \cdot \mathbf{curl}\,(\Phi_e \mathcal{L}(\varrho_e)),\end{aligned}$$

where T_e denotes the only element of $\mathcal{T}_h$ having e as an edge. Therefore, (3.72) follows by mimicking the steps in the proof of (3.71).

Finally, proceeding exactly as in the proof of [78, Lemma 4.14], we find

$$\begin{aligned}\|p_\Gamma - p_h\|_{0,e} &\leq C_{\mathrm{tr}} \left(h_e^{-1/2}\|p - p_h\|_{0,T} + h_e^{1/2}|p - p_h|_{1,T}\right) \\ &= C_{\mathrm{tr}} \left(h_e^{-1/2}\|p - p_h\|_{0,T} + h_e^{1/2}\left\|\rho\boldsymbol{g} - \frac{\eta}{\kappa}\boldsymbol{\sigma}_h - \frac{\eta}{\kappa}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h) - \nabla p_h\right\|_{1,T}\right) \\ &\leq C_{\mathrm{tr}} \left(h_e^{-1/2}\|p - p_h\|_{0,T} + h_e^{1/2}\left\|\nabla p_h - \rho\boldsymbol{g} + \frac{\eta}{\kappa}\boldsymbol{\sigma}_h\right\|_{0,T} + h_e^{1/2}\left\|\frac{\eta}{\kappa}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\right\|_{1,T}\right).\end{aligned}$$

The result (3.73) then follows immediately from (3.69) and the fact that $h_e \leq h_T$. □

We remark that (3.72) holds also when p_Γ is sufficiently smooth. In this case, we can approximate this data by a Taylor polynomial approximation and obtain (3.72) with further higher order terms appearing on the right-hand side.

Next, we provide the upper bounds for the estimator terms in (3.36). Our general strategy consists of mimicking the proofs of the results in [140, Section 6] under further assumptions on the data. We have the following lemma.

Lemma 3.19. *Suppose that $\boldsymbol{f}$ and $\boldsymbol{m}_\Gamma$ are piecewise polynomials. There exist constants $c_6, c_7 > 0$, independent of λ and h, such that for all $T \in \mathcal{T}_h$ and $e \in \mathcal{E}_h(\Gamma_p)$,*

$$h_T\|\boldsymbol{f} + \mathbf{div}\,(2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}_h) - \phi_h\boldsymbol{I})\|_{0,T} \leq c_6\left(\|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,T} + \|\phi - \phi_h\|_{0,T}\right), \tag{3.77}$$

$$h_e^{1/2}\|\boldsymbol{m}_\Gamma - (2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}_h) - \phi_h\boldsymbol{I})\boldsymbol{n}\|_{0,e} \leq c_7\left(\|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,T} + \|\phi - \phi_h\|_{0,T}\right). \tag{3.78}$$

Furthermore, there exists a constant $c_8 > 0$, also independent of λ and h, such that for all $e \in \mathcal{E}_h(\Omega)$,

$$h_e^{1/2}\|[\![(2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}_h) - \phi_h\boldsymbol{I})\boldsymbol{n}]\!]\|_{0,e} \leq c_8 \sum_{T \subseteq \omega_e}\left(\|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,T} + \|\phi - \phi_h\|_{0,T}\right). \tag{3.79}$$

Proof. We prove (3.77) and (3.79) using similar arguments as in the proof of Lemma 3.18. We define $\boldsymbol{\chi}_T := \boldsymbol{f} + \mathbf{div}\,(2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}_h) - \phi_h\boldsymbol{I})$ and $\boldsymbol{\chi}_e := [\![(2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}_h) - \phi_h\boldsymbol{I})\boldsymbol{n}]\!]$. Then, applying (3.61) to $\|\boldsymbol{\chi}_T\|_{0,T}$, using that $\boldsymbol{f} = -\mathbf{div}\,(2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}) - \phi\boldsymbol{I})$ in Ω (cf. Lemma 3.3), integrating by parts, and finally using the

inverse estimate (3.64), we obtain

$$\begin{aligned}\|\boldsymbol{\chi}_T\|_{0,T}^2 \leq \gamma_1 \|\Phi_T^{1/2}\boldsymbol{\chi}_T\|_{0,T}^2 &= \gamma_1 \int_T \Phi_T \boldsymbol{\chi}_T^2 \\ &= \gamma_1 \int_T \Phi_T \boldsymbol{\chi}_T \cdot (\boldsymbol{f} + \mathbf{div}\,(2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}_h) - \phi_h \boldsymbol{I})) \\ &= \gamma_1 \int_T \Phi_T \boldsymbol{\chi}_T \cdot \mathbf{div}\,(2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}_h - \boldsymbol{u}) - (\phi_h - \phi)\boldsymbol{I}) \\ &= -\gamma_1 \int_T \nabla(\Phi_T \boldsymbol{\chi}_T) : (2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}_h - \boldsymbol{u}) - (\phi_h - \phi)\boldsymbol{I}) \\ &\leq C h_T^{-1} \|\Phi_T \boldsymbol{\chi}_T\|_{0,T}\, \|2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}_h - \boldsymbol{u}) - (\phi_h - \phi)\boldsymbol{I}\|_{0,T}\,.\end{aligned}$$

By (3.61), $\|\Phi_T \boldsymbol{\chi}_T\|_{0,T} \leq \|\boldsymbol{\chi}_T\|_{0,T}$, thus $h_T \|\boldsymbol{\chi}_T\|_{0,T} \leq \widetilde{C}\,(\|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,T} + \|\phi - \phi_h\|_{0,T})$ providing (3.77).

Next, denoting by $\boldsymbol{\mathcal{L}}$ the vector operator defined componentwise by the extension $\mathcal{L} : \mathcal{C}(e) \to \mathcal{C}(T)$ introduced in Lemma 3.14, using inequality (3.62), and integrating by parts, we find

$$\begin{aligned}\|\boldsymbol{\chi}_e\|_{0,e}^2 \leq \gamma_2 \|\Phi_e^{1/2}\boldsymbol{\chi}_e\|_{0,e}^2 &= \int_e \Phi_e \boldsymbol{\mathcal{L}}(\boldsymbol{\chi}_e) \cdot \boldsymbol{\chi}_e \\ &= \int_e \Phi_e \boldsymbol{\mathcal{L}}(\boldsymbol{\chi}_e) \cdot (\boldsymbol{\chi}_e + [\![(2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}) - \phi\boldsymbol{I})\boldsymbol{n}]\!]) \\ &= \sum_{T \subseteq \omega_e} \left(\int_T \nabla(\Phi_e \boldsymbol{\mathcal{L}}(\boldsymbol{\chi}_e)) : (2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}_h - \boldsymbol{u}) - (\phi_h - \phi)\boldsymbol{I}) + \int_T (\Phi_e \boldsymbol{\mathcal{L}}(\boldsymbol{\chi}_e)) \cdot \boldsymbol{\chi}_T \right) \\ &\leq \sum_{T \subseteq \omega_e} h_T^{-1} \left(\|2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}_h - \boldsymbol{u}) - (\phi_h - \phi)\boldsymbol{I}\|_{0,T} + h_T \|\boldsymbol{\chi}_T\|_{0,T} \right) \|\Phi_e \boldsymbol{\mathcal{L}}(\boldsymbol{\chi}_e)\|_{0,T} \\ &\leq \widehat{C} h_e^{1/2} \sum_{T \subseteq \omega_e} h_T^{-1} \left(\|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,T} + \|\phi - \phi_h\|_{0,T} \right) \|\Phi_e \boldsymbol{\mathcal{L}}(\boldsymbol{\chi}_e)\|_{0,T}.\end{aligned} \tag{3.80}$$

Similar to the steps in the proof of (3.71) we note that $\|\Phi_e \boldsymbol{\mathcal{L}}(\boldsymbol{\chi}_e)\|_{0,T} \leq \gamma_4 h_e^{1/2} \|\boldsymbol{\chi}_e\|$. Combined with (3.80) this implies

$$h_e^{1/2} \|\boldsymbol{\chi}_e\|_{0,e} \leq \overline{C} \sum_{T \subseteq \omega_e} \left(\|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,T} + \|\phi - \phi_h\|_{0,T} \right),$$

since $h_e \leq h_T$ for all $T \subseteq \omega_e$. The result (3.79) follows.

Finally, proceeding as in the proof of (3.72), it is not difficult to see that the proof of (3.79) is similar to that of (3.78). □

Note again that, in the above lemma, if the data is sufficiently smooth instead of piecewise polynomial, then higher order terms arising from suitable polynomial approximations will appear on the corresponding right-hand sides.

The efficiency estimate (3.60) now follows directly from Lemmas 3.13, 3.16, 3.17, 3.18 and 3.19.

3.4.3 Extension of the estimator to three dimensions

We briefly discuss the *a posteriori* error estimator in three dimensions.

Given a sufficiently smooth vector field $\boldsymbol{\tau}$, we let $\mathbf{curl}\,\boldsymbol{\tau} := \nabla \times \boldsymbol{\tau}$. Furthermore, we take a tetrahedralization $\mathcal{T}_h$ of $\overline{\Omega}$ and consider the same notation as in the introduction of Section 3.4 (replacing the

word "edge" by "face"). Given a face $e \in \mathcal{E}_h(\Omega)$, $\boldsymbol{\tau} \in \mathbf{L}^2(\Omega)$ and $\boldsymbol{\xi} \in [\mathrm{L}^2(\Omega)]^{3\times 3}$, such that $\boldsymbol{\tau} \in [\mathcal{C}(T)]^3$ and $\boldsymbol{\xi} \in [\mathcal{C}(T)]^{3\times 3}$ for all $T \in \mathcal{T}_h$, we let $[\![\boldsymbol{\tau} \times \boldsymbol{n}]\!]$ and $[\![\boldsymbol{\xi}\boldsymbol{n}]\!]$ be the corresponding jumps across e, namely, $[\![\boldsymbol{\tau} \times \boldsymbol{n}]\!] := \{(\boldsymbol{\tau}|_T)|_e - (\boldsymbol{\tau}|_{T'})|_e\} \times \boldsymbol{n}$ and $[\![\boldsymbol{\xi}\boldsymbol{n}]\!] := \{(\boldsymbol{\xi}|_T)|_e - (\boldsymbol{\xi}|_{T'})|_e\}\boldsymbol{n}$, respectively, where T and T' are the elements of $\mathcal{T}_h$ sharing a face e.

The local error indicator $\Theta_{f,T}$ now reads

$$\begin{aligned}\Theta_{f,T}^2 := h_T^2 \left\| \nabla p_h - \rho \boldsymbol{g} + \frac{\eta}{\kappa}\boldsymbol{\sigma}_h \right\|_{0,T}^2 + h_T^2 \left\| \mathrm{curl}\left(\frac{\eta}{\kappa}\boldsymbol{\sigma}_h - \rho \boldsymbol{g}\right)\right\|_{0,T}^2 \\ + \sum_{e\in\mathcal{E}(T)\cap\mathcal{E}_h(\Gamma_p)} \left(h_e \|p_\Gamma - p_h\|_{0,e}^2 + h_e \left\| \left(\frac{\eta}{\kappa}\boldsymbol{\sigma}_h - \rho \boldsymbol{g}\right) \times \boldsymbol{n} + \nabla p_\Gamma \times \boldsymbol{n} \right\|_{0,e}^2 \right) \\ + \sum_{e\in\mathcal{E}(T)\cap\mathcal{E}_h(\Omega)} h_e \left\| \left[\!\left[\left(\frac{\eta}{\kappa}\boldsymbol{\sigma}_h - \rho \boldsymbol{g}\right) \times \boldsymbol{n} \right]\!\right] \right\|_{0,e}^2,\end{aligned}$$

while the error indicators $\Theta_{s,T}$ and $\Theta_{sf,T}$ are defined as for the two-dimensional case in (3.36) and (3.38), respectively. We then set the global indicator as in (3.35).

All the results for the reliability estimate in Section 3.4.1 hold also in the three-dimensional case, except the upper bound for $\|\mathcal{F}_2\|_{\mathbf{Z}'}$ in Section 3.4.1. To bound this term, we require the following three results.

We require the 3D analogue of (3.58). This is an immediate consequence of the identity

$$\langle \mathrm{curl}\, \boldsymbol{\varphi} \cdot \boldsymbol{n}, \boldsymbol{\chi}\rangle_{\Gamma_p} = -\langle \nabla \boldsymbol{\chi} \times \boldsymbol{n}, \boldsymbol{\varphi}\rangle_\Gamma \quad \forall\, \boldsymbol{\varphi}, \boldsymbol{\chi} \in \mathbf{H}^1(\Omega).$$

Its proof, like in the 2D case, follows from [68, Lemma 3.5].

We require also the following integration by parts formula:

$$\int_T \mathrm{curl}\, \boldsymbol{\tau} \cdot \boldsymbol{\chi} - \int_T \boldsymbol{\tau} \cdot \mathrm{curl}\, \boldsymbol{\chi} = \langle \boldsymbol{\tau} \times \boldsymbol{n}, \boldsymbol{\chi}\rangle_{\partial T}$$

for all $\boldsymbol{\tau} \in \mathbf{H}(\mathrm{curl}\,;\Omega) := \{\boldsymbol{\tau} \in \mathbf{L}^2(\Omega) :\ \mathrm{curl}\, \boldsymbol{\tau} \in \mathbf{L}^2(\Omega)\}$ and $\boldsymbol{\chi} \in \mathbf{H}^1(\Omega)$. Above, $\langle\cdot,\cdot\rangle_{\partial T}$ stands for the duality pairing between $\mathbf{H}^{-1/2}(\partial T)$ and $\mathbf{H}^{1/2}(\partial T)$.

Finally, the stable Helmholtz decomposition in Lemma 3.10 is also valid in this case (see [74, Theorem 3.2]), where $\mathbf{curl}\, \varphi$ in (3.55) is replaced by $\mathrm{curl}\, \boldsymbol{\varphi}$ ($\boldsymbol{\varphi} \in \mathbf{H}^1_{\Gamma_{\boldsymbol{u}}}(\Omega)$). A proof for the upper bound for $\|\mathcal{F}_2\|_{\mathbf{Z}'}$, the proof of the reliability of Θ, as well as the efficiency estimate, proceed now as in the two-dimensional case.

3.5 Numerical examples

We present several tests illustrating the performance of the Galerkin scheme (3.23), verifying the reliability and efficiency of the a posteriori error estimator Θ, and confirming the locking-free estimates. All simulations were implemented using the *FEniCS library* [2]. As a direct solver we used the Multifrontal Massively Parallel Solver MUMPS [116]. In all our examples we use the finite element spaces (3.30)-(3.33).

In what follows, we denote by N the total number of degrees of freedom. The global error and the effectivity index associated to the global estimator Θ are denoted, respectively, by

$$\mathrm{e}(\boldsymbol{u},\phi,\boldsymbol{\sigma},p) := \Big(\mathrm{e}(\boldsymbol{u})^2 + \mathrm{e}(\phi)^2 + \mathrm{e}(\boldsymbol{\sigma})^2 + \mathrm{e}(p)^2\Big)^{1/2} \quad \text{and} \quad \texttt{eff}(\Theta) := \mathrm{e}(\boldsymbol{u},\phi,\boldsymbol{\sigma},p)/\Theta,$$

where

$$\mathrm{e}(\boldsymbol{u}) := \|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}, \quad \mathrm{e}(\phi) := \|\phi-\phi_h\|_{0,\Omega}, \quad \mathrm{e}(\boldsymbol{\sigma}) := \|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathrm{div},\Omega}, \quad \mathrm{e}(p) := \|p-p_h\|_{0,\Omega}.$$

Moreover, using the fact that $cN^{-1/d} \leq h \leq CN^{-1/d}$, the experimental rate of convergence of any of the above quantities will be computed as

$$\texttt{rate} := -d\left[\log(\mathrm{e}/\mathrm{e}')/\log(N/N')\right],$$

where N and N' denote the total degrees of freedom associated to two consecutive triangulations with errors e and e′.

The examples to be considered in this section are described next. Example 1 is used to explore the performance of the two-dimensional Galerkin scheme (3.23) and the *a posteriori* error estimator Θ under a quasi-uniform refinement, especially in the presence of volumetric locking. Furthermore, the two and three-dimensional simulations in Examples 2, 3 and 4 demonstrate the behavior of the adaptive algorithm associated to Θ, which reads as follows:

1. Start with a coarse mesh $\mathcal{T}_h$ of $\overline{\Omega}$.
2. Solve the discrete problem (3.23) on the current mesh.
3. Compute Θ_T for each $T \in \mathcal{T}_h$.
4. Check the stopping criterion and decide whether to finish or go to the next step.
5. Use Plaza and Carey's algorithm [115] to refine each $T' \in \mathcal{T}_h$ satisfying:

$$\Theta_{T'} \geq C_{\mathrm{per}} \max\{\Theta_T : T \in \mathcal{T}_h\} \quad \text{for some} \quad C_{\mathrm{per}} \in]0,1[.$$

6. Define the resulting mesh as the current mesh $\mathcal{T}_h$, and go to step 2.

Note that the above procedure is the usual adaptive refinement strategy from [139], except that the classical blue-green refinement has been replaced by step 5.

3.5.1 Example 1: Accuracy assessment

This first example is aimed at evaluating the accuracy of the method, as well as the properties of the *a posteriori* error estimator through the effectivity index $\texttt{eff}(\Theta)$, under a quasi-uniform refinement strategy. To that end, we consider the domain $\Omega :=]0,3/2[\times]0,1[$ and split its boundary into $\Gamma_{\boldsymbol{u}} := \Big\{(x_1,x_2)^T \in \mathbb{R}^2 : \ x_1 = 0 \ \text{ or } \ x_2 = 1\Big\}$ and $\Gamma_p := \Big\{(x_1,x_2)^T \in \mathbb{R}^2 : \ x_1 = 3/2 \ \text{ or } \ x_2 = 0\Big\}$. We choose the data $\boldsymbol{f}$, ℓ, p_Γ and $\boldsymbol{m}_\Gamma$ such that the solution of problem (3.11) is given by $\boldsymbol{u} := (u_1,u_2)^T$, where $u_1(x_1,x_2) := 0.1\left(\sin(\pi x_1)\cos(\pi x_2) + \frac{x_1^2}{2\lambda}\right)$ and $u_2(x_1,x_2) := 0.1\left(-\cos(\pi x_1)\sin(\pi x_2) + \frac{x_2^2}{2\lambda}\right)$,

and $p(x_1,x_2) := \pi\sin(\pi x_1)\sin(\pi x_2)$, and ϕ and $\boldsymbol{\sigma}$ defined as in (3.11b) and (3.11c), respectively, with $\boldsymbol{g} := (0,1)^T$.

In Table 3.1 we present the convergence history obtained for this example under the following non-dimensional model parameters: $\eta = \alpha = \rho = 1$, $c_0 = 10^{-3}$, $\kappa(x_1,x_2) := 1 + \sin^2(\pi x_1)\cos^2(\pi x_2)$, $E = 100$, $\boldsymbol{g} = (0,0,-1)^T$ and three cases for the Poisson ratio, $\nu = 0.35$, $\nu = 0.4$ and $\nu = 0.4999$. From Table 3.1 we conclude that there are almost no differences between the corresponding errors when varying ν. This confirms that the estimates given by Lemma 3.4 are independent of $\lambda := 2E/[(1-2\nu)(1+\nu)]$, i.e., our conforming scheme (3.23) is locking-free. Moreover, for each value of ν, the effectivity index $\texttt{eff}(\Theta)$ remains bounded, thus verifying the reliability and efficiency of the *a posteriori* error estimator Θ.

It is worth mentioning that it is desirable to have $\texttt{eff}(\Theta) \to 1$ as $N \to \infty$. For the four-field poroelasticity equations, we claim that $\texttt{eff}(\Theta)$ is affected by the values of η/κ in (3.11c). To show this, we use the same model parameters as before, fix $\nu = 0.4$, and consider the cases of $\eta/\kappa = 10^4$, $\eta/\kappa = 10^0$ and $\eta/\kappa = 10^{-4}$. The decay of the corresponding total errors with respect to the total number of degrees of freedom, as well as the effectivity indexes, using a quasi-uniform refinement strategy are depicted in Figure 3.1. From these results, we conclude that the method is not robust with respect to the ratio η/κ. Moreover, in two cases the effectivity index is far from 1 and for all cases the effectivity index differs from each other, but is still bounded. This behavior is not surprising since our *a posteriori* and *a priori* error estimates may depend on η/κ. Despite this, we proceed as in [127] to modify $\mathrm{e}(\boldsymbol{u},\phi,\boldsymbol{\sigma},p)$ in such a way that $\texttt{eff}(\Theta)$ is closer to 1. For this, we first introduce the estimator terms Θ_i $(i = 1,\ldots,10)$ given by $\Theta_i^2 := \sum_{T\in\mathcal{T}_h} \widehat{\Theta}_i^2$, where

$$\widehat{\Theta}_1^2 := \left\| \left(c_0 + \frac{\alpha^2}{\lambda} \right) p_h - \frac{\alpha}{\lambda}\phi_h + \operatorname{div}\boldsymbol{\sigma}_h - \ell \right\|_{0,T}^2, \qquad \widehat{\Theta}_2^2 := h_T^2 \|\boldsymbol{f} + \mathbf{div}\,(2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}_h) - \phi_h\boldsymbol{I})\|_{0,T}^2,$$

$$\widehat{\Theta}_3^2 := h_T^2 \left\| \operatorname{rot}\left(\frac{\eta}{\kappa}\boldsymbol{\sigma}_h - \rho\boldsymbol{g} \right) \right\|_{0,T}^2, \qquad \widehat{\Theta}_4^2 := \left\| \frac{1}{\lambda}(\phi_h - \alpha p_h) + \operatorname{div}\boldsymbol{u}_h \right\|_{0,T}^2,$$

$$\widehat{\Theta}_5^2 := \sum_{e\in\mathcal{E}(T)\cap\mathcal{E}_h(\Gamma_p)} h_e \left\| \left(\frac{\eta}{\kappa}\boldsymbol{\sigma}_h - \rho\boldsymbol{g} \right)\cdot\boldsymbol{s} + \frac{dp_\Gamma}{d\boldsymbol{s}} \right\|_{0,e}^2, \qquad \widehat{\Theta}_6^2 := \sum_{e\in\mathcal{E}(T)\cap\mathcal{E}_h(\Gamma_p)} h_e \|p_\Gamma - p_h\|_{0,e}^2,$$

$$\widehat{\Theta}_7^2 := \sum_{e\in\mathcal{E}(T)\cap\mathcal{E}_h(\Gamma_p)} h_e \|\boldsymbol{m}_\Gamma - (2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}_h) - \phi_h\boldsymbol{I})\boldsymbol{n}\|_{0,e}^2, \qquad \widehat{\Theta}_8^2 := \sum_{e\in\mathcal{E}(T)\cap\mathcal{E}_h(\Omega)} h_e \left\| \left[\!\left[\left(\frac{\eta}{\kappa}\boldsymbol{\sigma}_h, -\rho\boldsymbol{g} \right)\cdot\boldsymbol{s} \right]\!\right] \right\|_{0,e}^2,$$

$$\widehat{\Theta}_9^2 := \sum_{e\in\mathcal{E}(T)\cap\mathcal{E}_h(\Omega)} h_e \|[\![(2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}_h) - \phi_h\boldsymbol{I})\boldsymbol{n}]\!]\|_{0,e}^2, \qquad \widehat{\Theta}_{10}^2 := h_T^2 \left\| \nabla p_h - \rho\boldsymbol{g} + \frac{\eta}{\kappa}\boldsymbol{\sigma}_h \right\|_{0,T}^2.$$

The history of convergence of these estimator terms for the three values of η/κ are shown in Figure 3.2. Although $\Theta_1 > \Theta_i$ for all $i = 2,\ldots,10$ when $\kappa/\eta = 10^{-4}$, the results for $\kappa/\eta = 10^0$ and $\kappa/\eta = 10^4$ allow us to conjecture that the global estimator Θ focuses on refining where the divergence of $2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}-\boldsymbol{u}_h) - (\phi-\phi_h)\boldsymbol{I}$ (associated to Θ_2) is large. Inspired by [127], this situation leads us to consider, under further regularity of the solution, the modified total error and effectivity index given by

$$\widehat{\mathrm{e}}(\boldsymbol{u},\phi,\boldsymbol{\sigma},p) := \left(\mathrm{e}(\boldsymbol{u},\phi,\boldsymbol{\sigma},p)^2 + \sum_{T\in\mathcal{T}_h} h_T^2 \|\mathbf{div}\,(2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}-\boldsymbol{u}_h) - (\phi-\phi_h)\boldsymbol{I})\|_{0,T}^2 \right)^{1/2}$$

and

$$\widehat{\texttt{eff}}(\Theta) := \widehat{\texttt{e}}(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p)/\Theta,$$

respectively. The left panel of Figure 3.3 illustrates the updated history of convergence, whereas the associated effectivity indexes are shown on the right panel. It can be concluded that, in general, $\widehat{\texttt{eff}}(\Theta)$ is much closer to 1 than $\texttt{eff}(\Theta)$.

$\nu = 0.35$											
	$\texttt{e}(\boldsymbol{u})$		$\texttt{e}(\phi)$		$\texttt{e}(\boldsymbol{\sigma})$		$\texttt{e}(p)$		$\texttt{e}(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p)$		
N	error	rate	error	rate	error	rate	error	rate	error	rate	$\texttt{eff}(\Theta)$
220	6.16e-02	–	4.08e-01	–	1.13e+01	–	2.96e-01	–	1.14e+01	–	0.390
762	1.65e-02	2.12	1.09e-01	2.13	3.21e+00	2.03	8.03e-02	2.10	3.21e+00	2.03	0.410
3146	3.87e-03	2.05	2.44e-02	2.11	7.85e-01	1.99	1.93e-02	2.01	7.86e-01	1.99	0.438
11664	1.03e-03	2.02	6.41e-03	2.04	2.11e-01	2.00	5.02e-03	2.05	2.11e-01	2.00	0.449
46975	2.48e-04	2.05	1.49e-03	2.09	5.23e-02	2.00	1.22e-03	2.03	5.24e-02	2.00	0.457
186597	6.23e-05	2.00	3.72e-04	2.02	1.31e-02	2.00	3.07e-04	2.00	1.31e-02	2.00	0.455
744791	1.56e-05	2.00	9.28e-05	2.01	3.32e-03	1.99	7.69e-05	2.00	3.32e-03	1.99	0.459
$\nu = 0.4$											
	$\texttt{e}(\boldsymbol{u})$		$\texttt{e}(\phi)$		$\texttt{e}(\boldsymbol{\sigma})$		$\texttt{e}(p)$		$\texttt{e}(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p)$		
N	error	rate	error	rate	error	rate	error	rate	error	rate	$\texttt{eff}(\Theta)$
220	6.17e-02	–	4.48e-01	–	1.13e+01	–	2.96e-01	–	1.14e+01	–	0.400
762	1.66e-02	2.12	1.13e-01	2.22	3.21e+00	2.03	8.03e-02	2.10	3.21e+00	2.03	0.420
3146	3.88e-03	2.05	2.50e-02	2.13	7.85e-01	1.99	1.93e-02	2.01	7.86e-01	1.99	0.448
11664	1.03e-03	2.02	6.49e-03	2.06	2.11e-01	2.00	5.02e-03	2.05	2.11e-01	2.00	0.460
46975	2.48e-04	2.05	1.50e-03	2.11	5.23e-02	2.00	1.22e-03	2.03	5.24e-02	2.00	0.467
186597	6.23e-05	2.00	3.72e-04	2.02	1.31e-02	2.00	3.07e-04	2.00	1.31e-02	2.00	0.465
744791	1.56e-05	2.00	9.24e-05	2.01	3.32e-03	1.99	7.69e-05	2.00	3.32e-03	1.99	0.470
$\nu = 0.4999$											
	$\texttt{e}(\boldsymbol{u})$		$\texttt{e}(\phi)$		$\texttt{e}(\boldsymbol{\sigma})$		$\texttt{e}(p)$		$\texttt{e}(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p)$		
N	error	rate	error	rate	error	rate	error	rate	error	rate	$\texttt{eff}(\Theta)$
220	6.19e-02	–	8.81e-01	–	1.13e+01	–	2.96e-01	–	1.14e+01	–	0.415
762	1.66e-02	2.12	1.51e-01	2.84	3.21e+00	2.03	8.04e-02	2.10	3.22e+00	2.04	0.437
3146	3.88e-03	2.05	2.95e-02	2.30	7.85e-01	1.99	1.93e-02	2.01	7.86e-01	1.99	0.468
11664	1.03e-03	2.02	7.16e-03	2.16	2.11e-01	2.00	5.02e-03	2.05	2.11e-01	2.00	0.480
46975	2.48e-04	2.05	1.57e-03	2.18	5.23e-02	2.00	1.22e-03	2.03	5.24e-02	2.00	0.487
186597	6.24e-05	2.00	3.84e-04	2.05	1.31e-02	2.00	3.08e-04	2.00	1.31e-02	2.00	0.485
744791	1.56e-05	2.00	9.46e-05	2.02	3.32e-03	1.99	7.69e-05	2.00	3.32e-03	1.99	0.490

Table 3.1: Example 1: Convergence history of the errors under a quasi-uniform refinement strategy and different values of the Poisson ratio ν. (table produced by the author)

3.5.2 Example 2: Domain with corner singularity

In this example we set the model parameters (in non-dimensional form) as follows: $c_0 = \eta = 0.01$, $E = 100$, $\alpha = 1$ and $\nu = 0.35$. Furthermore, we neglect gravity effects and consider the inverted L-shaped domain $\Omega :=]-1, 1[\times]-1, 1[\setminus[0, 1] \times [-1, 0]$, with boundary parts $\Gamma_p :=]-1, 1[\times\{1\}$ and $\Gamma_{\boldsymbol{u}} := \Gamma \setminus \overline{\Gamma}_p$. The manufactured solution in polar coordinates is given by $\boldsymbol{u} := (u_1, u_2)^T$, where $u_1(r, \theta) := r^{2/3} \sin(2\theta/3)$ and $u_2(r, \theta) := r^{2/3} \cos(2\theta/3)$, and $p(r, \theta) := 1$, $\phi(r, \theta) := \alpha$ and $\boldsymbol{\sigma}(r, \theta) := \mathbf{0}$, with corresponding data. Note that $\Gamma_{\boldsymbol{u}}$ does not satisfy the geometrical assumption made in Lemma 3.10, which means that further regularity of the solution on a bigger convex domain needed by the

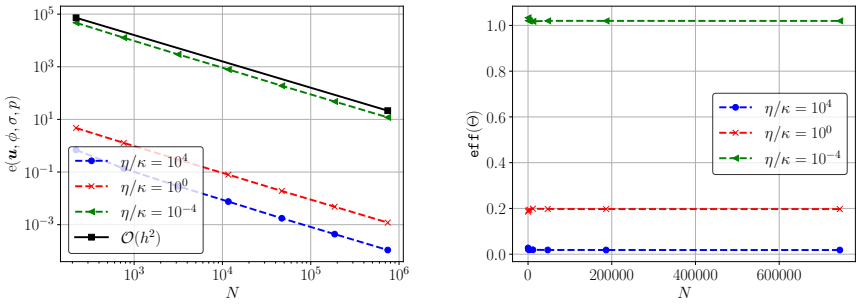


Figure 3.1: Example 1: Log-log plots of N vs $\mathrm{e}(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p)$ (left) and $\mathrm{eff}(\Theta)$ (right) for a quasi-uniform refinement strategy and different values of the ratio η/κ. (figure produced by the author)

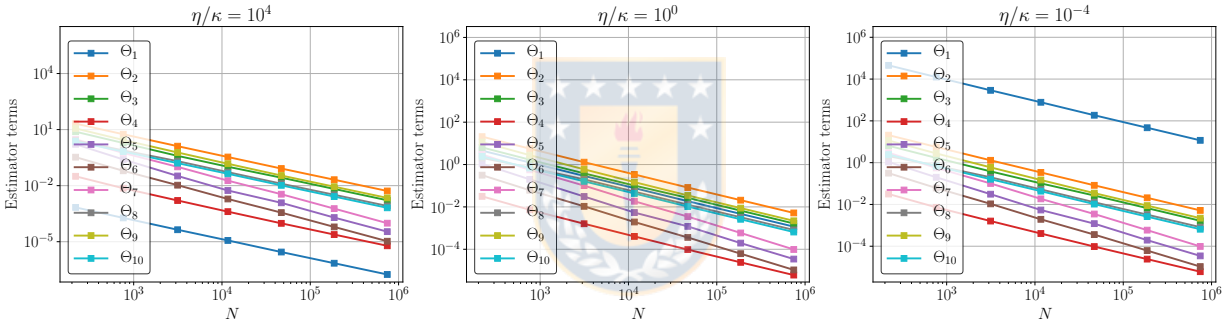


Figure 3.2: Example 1: Log-log plots of N vs Θ_i $(i = 1, \ldots, 10)$ for a quasi-uniform refinement strategy and different values of the ratio η/κ. (figure produced by the author)

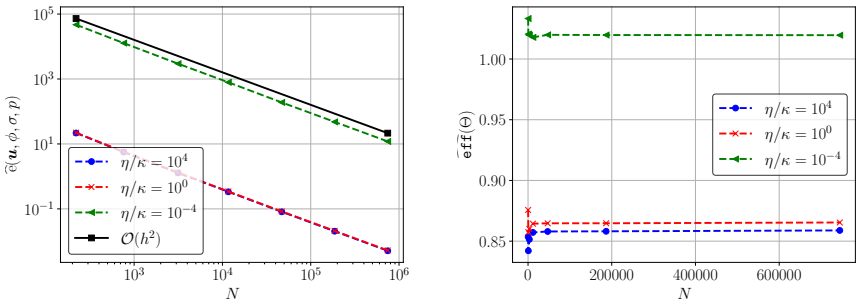


Figure 3.3: Example 1: Log-log plots of N vs $\widehat{\mathrm{e}}(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p)$ (left) and $\widehat{\mathrm{eff}}(\Theta)$ (right) for a quasi-uniform refinement strategy and different values of the ratio η/κ. (figure produced by the author)

Helmholtz decomposition (cf. (3.55)) cannot be guarantied theoretically (see [74] for more details). We omit this fact for the sake of convenience. Furthermore, we note that a negative power of the radius r appears when taking partial derivatives of the components of the displacements; this implies a singularity at the origin. It is well-known that in this case a convergence of $\mathcal{O}(h^{2/3-\delta})$ (with some $\delta > 0$) is expected from Theorem 3.6.

In Figure 3.4 we report the history of convergence of the total error for quasi-uniform and adaptive refinement strategies. It is clear that the errors using the adaptive refinement are considerably smaller than when using quasi-uniform refinement. Moreover, the adaptive procedure reduces the magnitude of $\mathrm{e}(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p)$ with optimal convergence of $\mathcal{O}(h^2)$. Some adapted meshes obtained with $C_{\mathrm{per}} = 0.2$ are depicted in Figure 3.5, where it is evident that the *a posteriori* error estimator Θ detects the singularity.

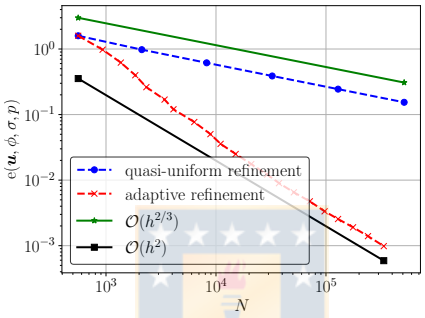


Figure 3.4: Example 2: Log-log plot of $\mathrm{e}(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p)$ vs N for both refinement strategies ($C_{\mathrm{per}} = 0.2$). (figure produced by the author)

3.5.3 Example 3: Three-dimensional L-shaped domain

We next consider a three-dimensional L-shaped domain as shown in the left panel of Figure 3.6. For this example we consider the following non-dimensional model parameters: $c_0 = 0.01$, $\eta = \alpha = \rho = 1$, $E = 10$, $\kappa = 0.05$ and $\nu = 0.4999$. Furthermore, the manufactured exact solution is defined as follows: $\boldsymbol{u} := (u_1, u_2, u_3)^T$, where $u_1(x_1, x_2, x_3) := 0.1\left(4(x_2^3 - 6x_3^5 + 15x_3^2) + \frac{x_1^2}{\lambda}\right)$, $u_2(x_1, x_2, x_3) := 0.1\left(2(x_2 - 10)x_3 + \frac{x_2^2}{\lambda}\right)$ and $u_3(x_1, x_2, x_3) := 0.1\left(x_3^2 + \frac{x_3^2}{\lambda}\right)$, $p(x_1, x_2, x_3) := x_1 x_3^4 - 30x_2^3 + x_3^2 + \frac{0.1(1.2 - x_3)}{[(1.05-x_1)^2 + (1.05-x_3)^2]}$, and ϕ and $\boldsymbol{\sigma}$ are defined as in (3.11b) and (3.11c), respectively, with $\boldsymbol{g} := (0, 0, -1)^T$. We notice that the partial derivatives of p exhibit singularities along the line $\left\{(x_1, x_2, x_3)^T \in \mathbb{R}^3 : \ x_1 = x_3 = 1.05\right\}$ so that high gradients of p are likely to occur near the re-entrant edge of the domain.

The right panel of Figure 3.6 illustrates the decay of the total error with respect to N for quasi-uniform and adaptive refinement strategies. A suboptimal rate of convergence is observed using quasi-uniform refinement. In contrast, the adaptive algorithm restores the optimal rate of convergence (i.e., $\mathcal{O}(h^2)$) and reduces the magnitude of $\mathrm{e}(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p)$ by marking the mesh elements near the re-entrant edge, as shown in Figure 3.7.

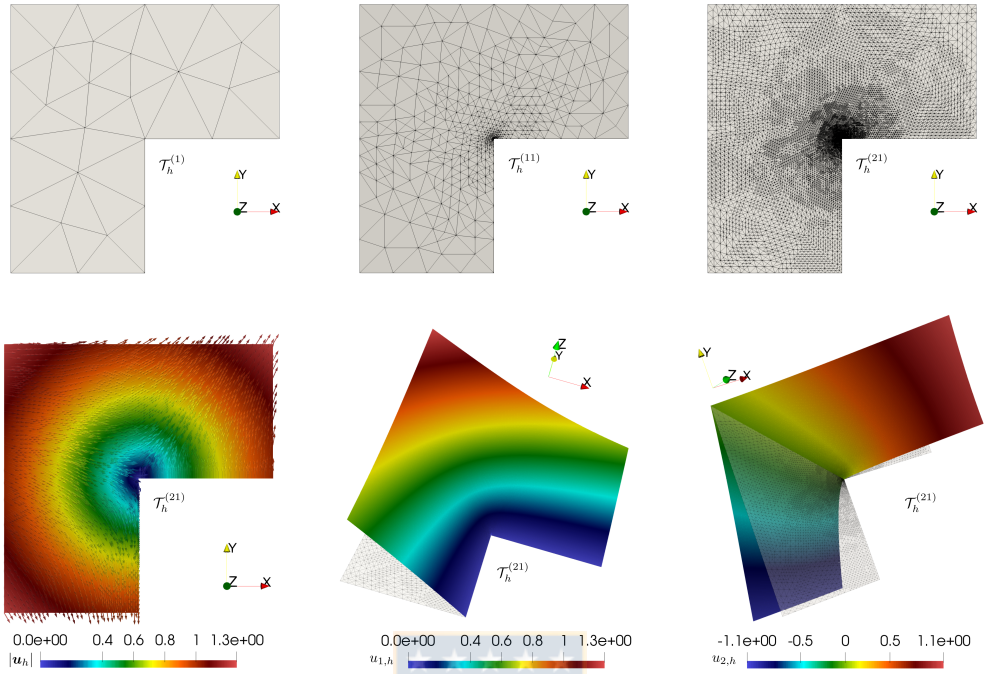


Figure 3.5: Example 2: Initial mesh and two adapted meshes obtained with the adaptive algorithm and $C_{\text{per}} = 0.2$ (first row), and approximate displacement magnitude, and approximate displacement components, denoted by $u_{1,h}$ and $u_{2,h}$, obtained at the $21st$ refinement step (second row). (figure produced by the author)

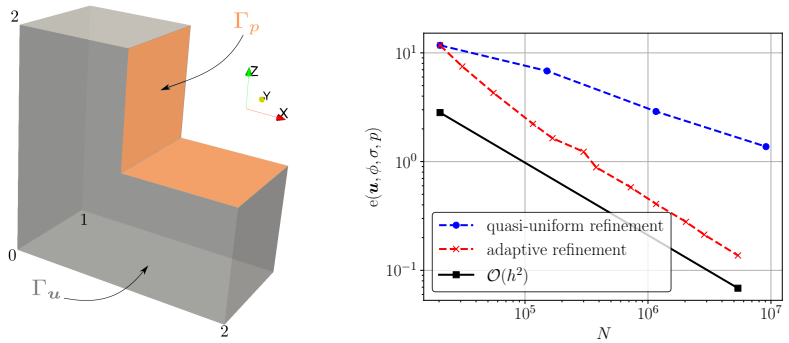


Figure 3.6: Example 3: Domain configuration (left) and log-log plot of $\mathrm{e}(\boldsymbol{u}, \phi, \boldsymbol{\sigma}, p)$ vs N for both refinement strategies (right). The adaptive algorithm was carried out with $C_{\text{per}} = 0.5$. (figure produced by the author)

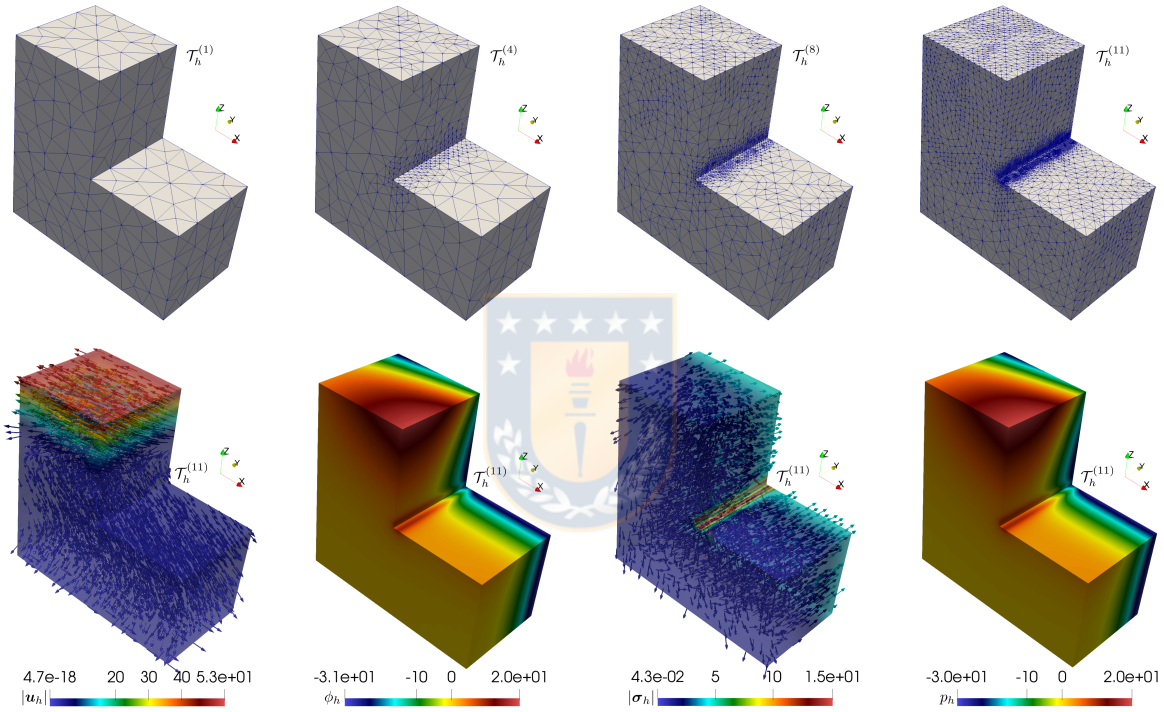


Figure 3.7: Example 3: Initial mesh and three adapted meshes obtained with the adaptive algorithm and $C_{\text{per}} = 0.5$ (fist row), and approximate displacement magnitude, approximate fluid flux and approximate fluid pressure obtained at the $11th$ refinement step (second row). (figure produced by the author)

3.5.4 Example 4: Simple-poroelastic brain model

In our final example we present a 3D computation illustrating the cerebrospinal fluid-tissue interaction in the human brain. For this, we use the Colin 27 mesh [70] as our initial mesh, see Figure 3.8. We neglect effects due to gravity.

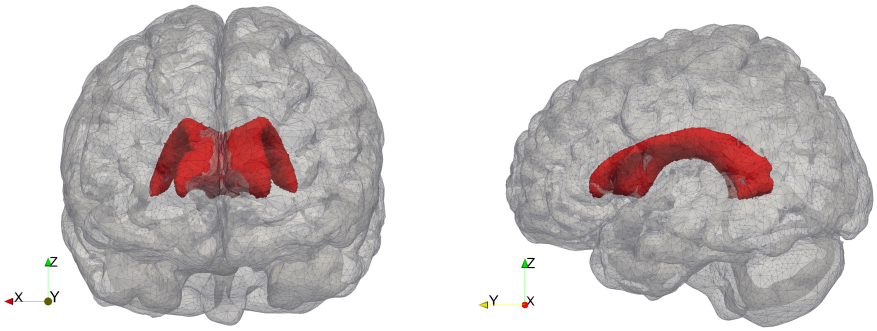
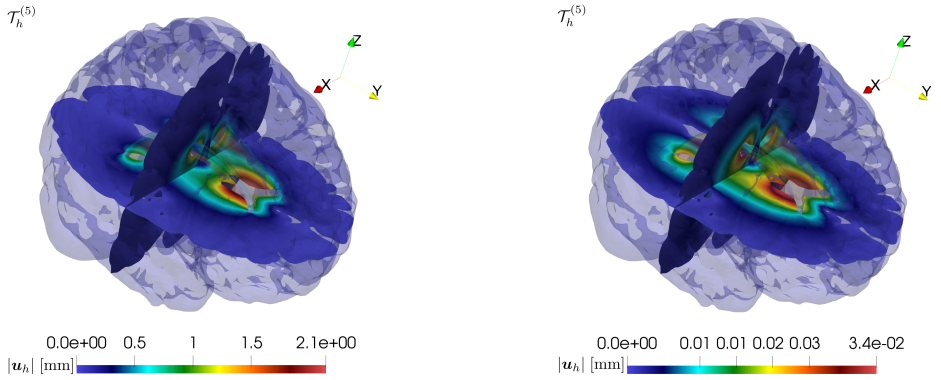


Figure 3.8: Left, posterior and right, lateral views of the initial mesh (with 99605 elements) used in Example 4. The inner ventricular boundary is shown in red. (figure produced by the author)

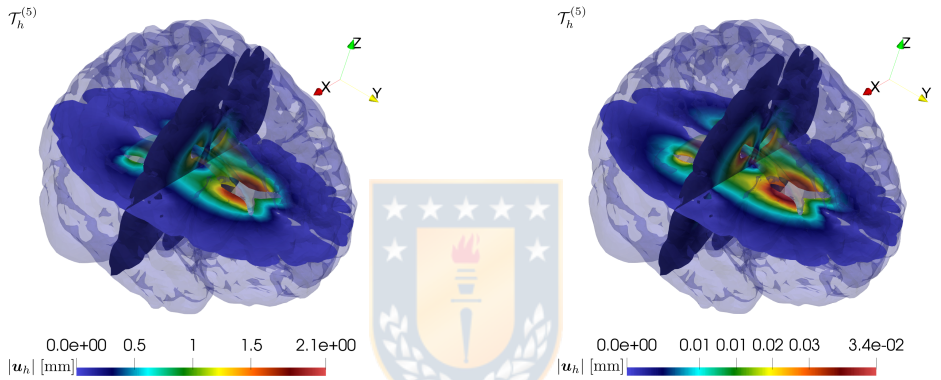
The material properties in our simulations are: $E = 1500\,[\text{Pa}]$, $\alpha = 0.25$, $c_0 = 3 \cdot 10^{-4}$ and $\eta = 100\,[\text{Pa}\cdot\text{s}]$. These are inspired by the numerical example of [94, Section 6]. We also consider three cases for the permeability, $\kappa = 3.75\,[\text{mm}^2]$, $\kappa = 1.57 \cdot 10^{-1}\,[\text{mm}^2]$ and $\kappa = 1.57 \cdot 10^{-3}\,[\text{mm}^2]$, and set $\Gamma_{\boldsymbol{u}}$ and Γ_p as the skull (outer boundary) and the ventricles (inner boundary) of the brain, respectively. Note that $\Gamma_{\boldsymbol{u}}$ does not satisfy the geometrical assumption made in the three-dimensional Helmholtz decomposition (see Lemma 3.10 for details in the two-dimensional case). We simply omit this fact and continue by imposing the following boundary conditions:

$$p = 799.92\ [\text{Pa}] \quad \text{and} \quad (2\mu\boldsymbol{\varepsilon}(\boldsymbol{u}) - \phi\boldsymbol{I})\boldsymbol{n} = -199.98\boldsymbol{n} \quad \text{on} \quad \Gamma_p,$$
$$\boldsymbol{u} = \boldsymbol{0} \quad \text{and} \quad \boldsymbol{\sigma}\cdot\boldsymbol{n} = 0 \quad \text{on} \quad \Gamma_{\boldsymbol{u}}.$$

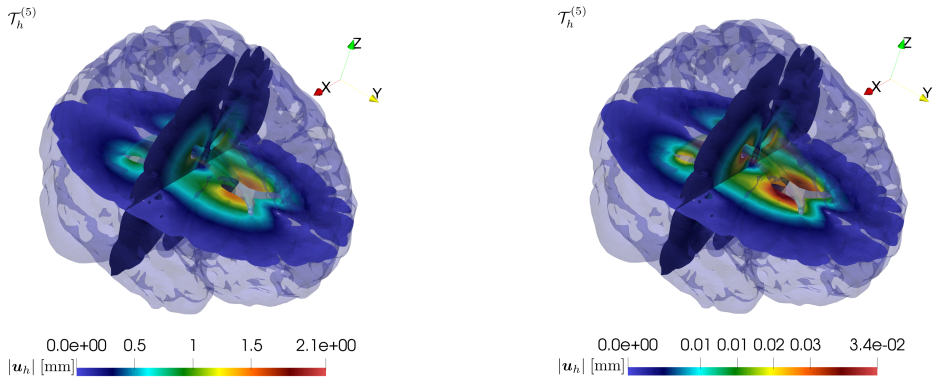
In Figure 3.9 we observe that there is little displacement when the brain behaves like an elastic material ($\nu = 0.4999$). Lowering the Poisson ratio to $\nu = 0.34$, the material is able to relax resulting in more displacement. In the first column we furthermore observe that increasing the permeability results in more displacement. This is due to a higher filtration rate of the fluid. As expected, in the elastic limit there is little effect on the displacement when increasing the permeability. In Figure 3.10 we observe compressibility effects due to high filtration when permeability is large, both for high ($\nu = 0.4999$) and low ($\nu = 0.34$) Poisson ratios. Finally, the $5th$ adapted mesh for the case $\nu = 0.4999$ and $\kappa = 1.57 \cdot 10^{-3}$ $[\text{mm}^2]$ is depicted in Figure 3.11, from which it is concluded that the adaptive algorithm refines near the ventricles. It is here where the pressures and displacement are highest.



(a) $\nu = 0.34$ and $\kappa = 1.57 \cdot 10^{-3}$ [mm^2]. (b) $\nu = 0.4999$ and $\kappa = 1.57 \cdot 10^{-3}$ [mm^2].

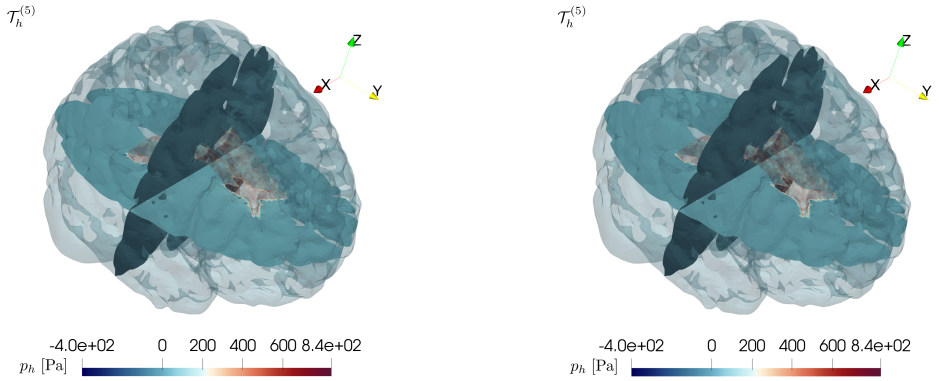


(c) $\nu = 0.34$ and $\kappa = 1.57 \cdot 10^{-1}$ [mm^2]. (d) $\nu = 0.4999$ and $\kappa = 1.57 \cdot 10^{-1}$ [mm^2].

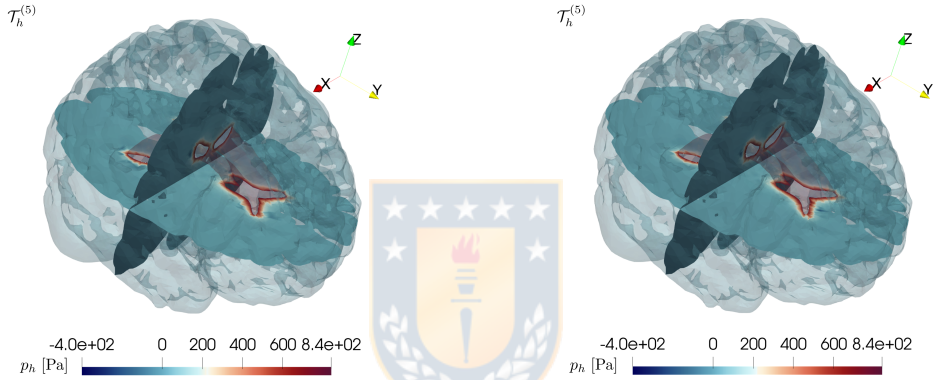


(e) $\nu = 0.34$ and $\kappa = 3.75$ [mm^2]. (f) $\nu = 0.4999$ and $\kappa = 3.75$ [mm^2].

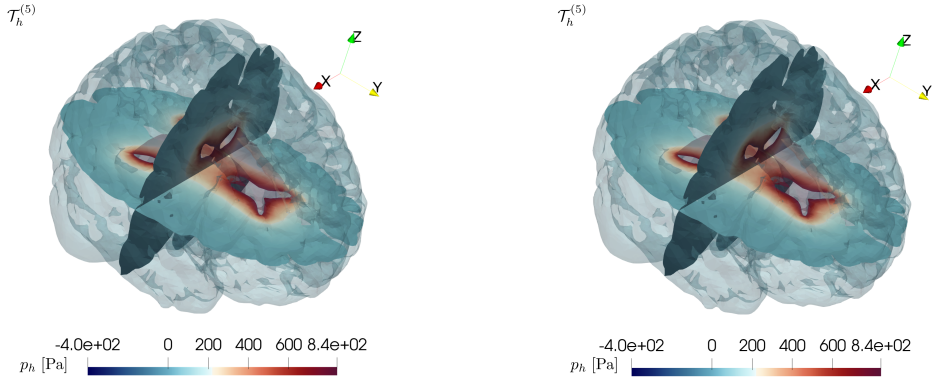
Figure 3.9: Example 4: Approximate displacement magnitude for different values of ν and κ obtained at the $5th$ refinement step ($C_{\text{per}} = 0.3$) with: (a) $N = 4969116$ and 270243 elements, (b) $N = 5290281$ and 288805 elements, (c) $N = 3290456$ and 175830 elements, (d) $N = 3216013$ and 171634 elements, (e) $N = 3865851$ and 209323 elements; and (f) $N = 3369212$ and 180800 elements. (figure produced by the author)



(a) $\nu = 0.34$ and $\kappa = 1.57 \cdot 10^{-3}\,[\mathrm{mm}^2]$. (b) $\nu = 0.4999$ and $\kappa = 1.57 \cdot 10^{-3}\,[\mathrm{mm}^2]$.



(c) $\nu = 0.34$ and $\kappa = 1.57 \cdot 10^{-1}\,[\mathrm{mm}^2]$. (d) $\nu = 0.4999$ and $\kappa = 1.57 \cdot 10^{-1}\,[\mathrm{mm}^2]$.



(e) $\nu = 0.34$ and $\kappa = 3.75\,[\mathrm{mm}^2]$. (f) $\nu = 0.4999$ and $\kappa = 3.75\,[\mathrm{mm}^2]$.

Figure 3.10: Example 4: Approximate fluid pressure for different values of ν and κ obtained at the $5th$ refinement step ($C_{\mathrm{per}} = 0.3$) with: (a) $N = 4969116$ and 270243 elements, (b) $N = 5290281$ and 288805 elements, (c) $N = 3290456$ and 175830 elements, (d) $N = 3216013$ and 171634 elements, (e) $N = 3865851$ and 209323 elements; and (f) $N = 3369212$ and 180800 elements. (figure produced by the author)

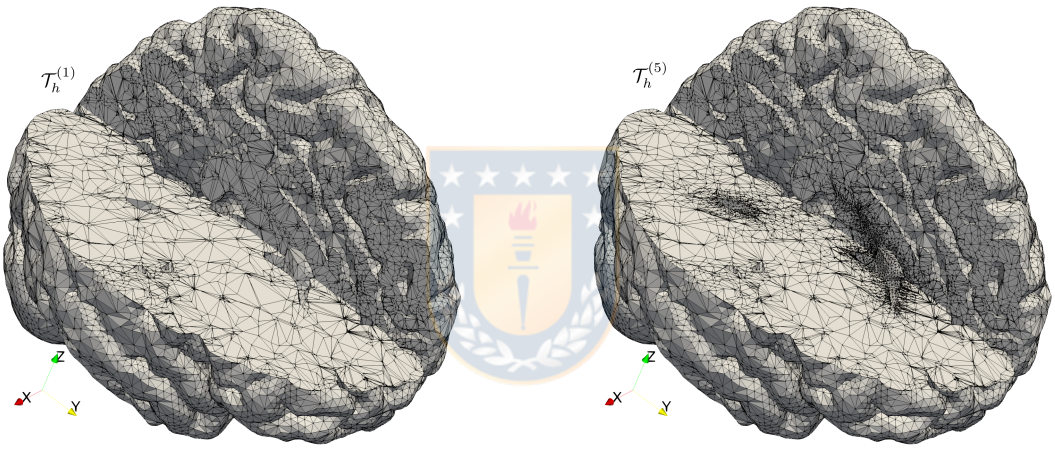


Figure 3.11: Example 4: Initial mesh (left) and the $5th$ adapted mesh obtained with $\nu = 0.4999$ and $\kappa = 1.57 \cdot 10^{-3}$ [mm^2] (right). These meshes have 99605 and 288805 elements, respectively. (figure produced by the author)

Discussion and future work

In this thesis we considered problems of physical interest subjected to curved domains and singularities or high gradients of the solution, and developed new high order mixed finite element methods that can be employed without the risk of losing accuracy in the aforementioned situations. Let us discuss some aspects of the methods proposed and analyzed in this dissertation that we think the reader should take into account.

The analysis we presented in Chapter 1 depends on certain hypotheses on the *transferring paths* which, in practice, we restricted to two situations where numerical evidence suggested that the method can be used. First, we may consider *fitted* methods resulting from a piecewise linear interpolation of the boundary Γ and conclude that Assumptions **D** of Section 1.2.4 hold for h small enough. Moreover, since in this case the *transferring paths* related to the vertices of a boundary edge e can be chosen so that they are perpendicular to e, the equivalence norm involving these paths holds without requiring the assumptions in Lemma 1.6.

The other case we considered is when the domain is immersed in a backgroundmesh and D_h is the union of all elements inside Ω. This technique is very convenient for complex geometries since Γ_h is away from Γ. The first concern here, however, is that Assumption **(D.2)** is difficult to ensure in practice. The second issue concerns the numerical implementation of *transferring paths* satisfying the hypotheses in Lemma 1.6. In our numerical simulations we considered a two-dimensional algorithm to compute these paths for which no problems were detected, but we recognize further research on this computational aspect is needed to see if we can expected the hypotheses in Lemma 1.6 to hold for general geometries in two and three dimensions.

In Chapter 2 we extended the analysis of purely diffusive problems to the incompressible Stokes equations in mixed form. This extension by itself is non-trivial because the introduction of the pseudostress tensor as an additional unknown requires the trace of this tensor to satisfy a zero-mean condition over D_h. The assumptions on the *transferring paths* are similar to those required by the previous work. Furthermore, a residual based *a posteriori* error estimator was developed and both reliability and quasi-efficiency estimates for the estimator were shown. We restricted this estimator to the case where the curved boundary is interpolated by a piecewise linear function. However, the *a priori* error estimates holds for computational meshes that are not necessarily adjusted to the curved boundary. In this way, we are interested in extending our *a posteriori* error estimator to handle those meshes as well.

The theory we developed in Chapters 1 and 2 can be adapted to three dimensions. In fact, by the equivalence norm result in Appendix B and the notation introduced in Section 2.5.4, the extension

of the analyses for both purely diffusive and Stokes problems to three dimensions is straightforward. The numerical implementation of this case is much harder and will be considered in future work. In the context of *unfitted* methods, we would also like to study non-linear problems, such as the Stokes problem for quasi-Newtonian fluids and Navier–Stokes equations, as well as more general boundary conditions.

On the other hand, in Chapter 3 we developed an error analysis for a conforming and locking-free finite element approximation of a four-field formulation of the stationary Biot's model. An important limitation of the results we presented is that all the analysis breaks down for the limit case $c_0 = 0$. In fact, as was pointed out in [92, Section 2], the stability constant in Theorem 3.2 blows up as c_0 becomes very small. However, numerical simulations in [92] suggested that for $c_0 = 0$ the analysis might be derived using a different approach. This is somehow what we might see in two-phase models of partially melted materials that degenerate in regions where the porosity vanishes, for which some mixed and HDG methods based on weighted Sobolev spaces have shown to handle the degenerancies in the porosity [7, 90]. In this direction, it would be interesting to derive a new locking-free method for Biot's model following the ideas in [7, 90]. En esta dirección, sería interesante derivar un nuevo método libre de bloqueo para el modelo de Biot siguiendo las ideas en [7, 90].

As another extension of Chapter 3, we are interested in developing robust numerical methods with respect to the ratio between the viscosity of the pore fluid and the permeability of the porous solid in Biot's consolidation model. Furthermore, since the numerical results obtained for this model are very promising, especially in the context of our fourth example in Section 3.5, we would also like to extend our *a posteriori* error analysis to the more sophisticated model of multiple network poroelasticity [94], which can be used, for example, to study hydrocephalus [135], cerebral oedema [136], and risk factors associated with early stages of Alzheimer's disease [85].

Discusión y trabajos futuros

En esta tesis consideramos problemas de interés físico sujetos a dominios curvos y singularidades o gradientes altos de la solución, y desarrollamos nuevos métodos de elementos finitos mixtos de alto orden que pueden emplearse sin el riesgo de perder precisión en las situaciones antes mencionadas. Discutamos algunos aspectos de los métodos propuestos y analizados en esta tesis que creemos que el lector debe tener en cuenta.

El análisis que presentamos en el Capítulo 1 depende de ciertas hipótesis sobre *los caminos de transferencia* que, en la práctica, restringimos a dos situaciones en las que la evidencia numérica sugiere que el método puede ser utilizado. Primero, podemos considerar los métodos *fitted* que resultan de una interpolación lineal a trozos de la frontera Γ y concluir que las Hipótesis **D** de la Sección 1.2.4 se cumplen para h lo suficientemente pequeño. Además, dado que en este caso los *caminos de transferencia* asociados con los vértices de una cara de frontera e se pueden elegir de tal forma que sean perpendiculares a e, la equivalencia de normas que involucra estos caminos se cumple sin requerir los supuestos en Lemma 1.6.

El otro caso que consideramos es cuando el dominio está inmerso en una malla de fondo y D_h es la unión de todos los elementos contenidos en Ω. Esta técnica es muy conveniente para geometrías complejas ya que Γ_h está lejos de Γ. Sin embargo, la primera preocupación aquí es que la Hipótesis **(D.2)** es difícil de garantizar en la práctica. El segundo problema se refiere a la implementación numérica de *caminos de transferencia* que satisfagan las hipótesis en el Lemma 1.6. En nuestras simulaciones numéricas, consideramos un algoritmo bidimensional para calcular estos caminos para las cuales no se detectaron problemas, pero reconocemos que se necesita más investigación sobre este aspecto computacional para ver si podemos esperar que las hipótesis en el Lemma 1.6 se cimplan para geometrías generales en dos y tres dimensiones.

En el Capítulo 2 ampliamos el análisis de problemas puramente difusivos a las ecuaciones de Stokes incompresibles en forma mixta. Esta extensión en sí misma no es trivial porque la introducción del tensor de pseudo-esfuerzo como una incógnita adicional requiere que la traza de este tensor satisfaga una condición de media cero sobre D_h. Los supuestos sobre los *caminos de transferencia* son similares a los requeridos por el trabajo anterior. Además, se desarrolló un estimador de error *a posteriori* de tipo residual y se mostraron estimaciones de confiabilidad y casi eficiencia para el estimador. Restringimos este estimador al caso en el que la frontera curva se interpola mediante una función lineal a trozos. Sin embargo, las estimaciones de error *a priori* se cumplen para mallas computacionales que no están necesariamente ajustadas a la frontera curva. De esta manera, estamos interesados en extender nuestro estimador de error *a posteriori* para manejar esas mallas también.

La teoría que desarrollamos en los Capítulos 1 y 2 se puede adaptar a tres dimensiones. En efecto, gracias al resultado de la norma de equivalencia en el Apéndice B y la notación introducida en la Sección 2.5.4, la extensión a tres dimensiones de los análisis para problemas puramente difusivos y de Stokes es inmediata. La implementación numérica de este caso es mucho más difícil y será considerada en trabajos futuros. En el contexto de métodos *unfitted*, también nos gustaría estudiar problemas no lineales, como el problema de Stokes para fluidos casi newtonianos y las ecuaciones de Navier-Stokes, así como también condiciones de frontera más generales.

Por otro lado, en el Capítulo 3 desarrollamos un análisis de error para una aproximación de elementos finitos conformes y libre bloqueo de una formulación de cuatro campos del modelo de Biot estacionario. Una limitación importante de los resultados que presentamos es que todo el análisis deja de cumplirse para el caso límite $c_0 = 0$. En efecto, como se señaló en [92, Sección 2], la constante de estabilidad en el Teorema 3.2 explota cuando c_0 se vuelve muy pequeña. Sin embargo, las simulaciones numéricas en [92] sugirieron que para $c_0 = 0$ el análisis podría derivarse usando un enfoque diferente. Esto es de alguna manera lo que podríamos ver en los modelos de dos fases de materiales parcialmente fundidos que se degeneran en regiones donde la porosidad es cero, para lo cual algunos métodos mixtos y HDG basados en espacios Sobolev con pesos han demostrado manejar las degeneraciones en la porosidad [7, 90].

Como otra extensión del Capítulo 3, estamos interesados en desarrollar métodos numéricos robustos con respecto al radio entre la viscosidad del fluido de los poros y la permeabilidad del sólido poroso en el modelo de consolidación de Biot. Además, dado que los resultados numéricos obtenidos para este modelo son muy prometedores, especialmente en el contexto de nuestro cuarto ejemplo en la Sección 3.5, también nos gustaría extender nuestro análisis de error *a posteriori* al modelo más sofisticado de poroelasticidad de redes múltiples [94], que se puede utilizar, por ejemplo, para estudiar hidrocefalia [135], edema cerebral [136], y factores de riesgo asociados con las primeras etapas de la enfermedad de Alzheimer [85].

Appendices

APPENDIX A

Estimates for $\widetilde{C}^e_{ext}$

In this section we provide an estimate of the extrapolation constant (1.15). To that end, we use the norm equivalence given by Lemma 1.6.

The following result extends the estimation in [57, Lemma A.1] to the case when the norm $\|\cdot\|_{0,\widetilde{K}^e_{ext}}$ is considered.

Lemma A.1. *Let e be any edge in $\mathcal{E}_h^{\partial}$. Let $\mathcal{L}$ be the line segment with endpoints given by the center of the biggest ball contained in K^e, and the point of the set $\widetilde{K}_{ext}$ where the polynomial p achieves its maximum. Suppose that Assumption* **(A.1)** *in Section 1.2.4 holds. Assume further that $\mathcal{L}$ is contained in the interior of the closure of the set $K^e \cup \widetilde{K}^e_{ext}$, denoted by B^e. Then, for any $p \in \mathrm{P}_l(B^e)$ we have*

$$\|p\|_{0,\widetilde{K}^e_{ext}} \leq C(\widetilde{r}_e)^{1/2}(l+1)^2\eta_e^l\|p\|_{0,K^e},$$

where $\widetilde{r}_e := \widetilde{H}_e/h_e^{\perp}$ and $\eta_e := 1+2\gamma_{K^e}\widetilde{r}_e+2\left(\gamma_{K^e}\widetilde{r}_e(1+\gamma_{K^e}\widetilde{r}_e)\right)^{1/2}$. Here the constant C solely depends on the shape-regularity constant γ_{K^e}.

Proof. We begin by noting that $\mathcal{L}$ can be subdivided into

$$I^e_{int} := \{\mathbf{x} \in \mathcal{L} : \ \mathbf{x} \cap K^e \neq \varnothing\} \quad \text{and} \quad I^e_{ext} := \{\mathbf{x} \in \mathcal{L} : \ \mathbf{x} \cap \widetilde{K}^e_{ext} \neq \varnothing\},$$

from which

$$\|p\|^2_{0,\widetilde{K}^e_{ext}} \leq |\widetilde{K}^e_{ext}| \max_{\mathbf{x}\in\widetilde{K}^e_{ext}} |p(\mathbf{x})| \leq |\widetilde{K}^e_{ext}|\|p\|^2_{\mathrm{L}^\infty(I^e_{ext})} \leq Ch_{K^e}\widetilde{H}_e\|p\|^2_{\mathrm{L}^\infty(I^e_{ext})},$$

since $|\widetilde{K}^e_{ext}| \leq Ch_{K^e}\widetilde{H}_e$.

On the other hand, the estimate $\|p\|_{\mathrm{L}^\infty(I^e_{ext})} \leq \eta_e^l\|p\|_{\mathrm{L}^\infty(I^e_{int})}$ holds by mimicking similar steps as in the proof of [57, Lemma A.1]. In fact, since $h_e^{\perp} \leq h_{K^e}$ and $h_{K^e} \leq \gamma_{K^e}\rho_{K^e}$, we find

$$\frac{|I^e_{ext}|}{|I^e_{int}|} \leq \frac{|I^e_{ext}|}{\rho_{K^e}} \leq \frac{\widetilde{H}_e}{\rho_{K^e}} \leq \gamma_{K^e}\frac{\widetilde{H}_e}{h_{K^e}} \leq \gamma_{K^e}\widetilde{r}_e,$$

where ρ_{K^e} is the radius of the biggest ball contained in K^e. From this, the estimate on $\|p\|_{\mathrm{L}^\infty(I^e_{ext})}$ follows from [56, Lemma 4.3]. Furthermore, we obtain using standard scaling arguments,

$$\|p\|_{\mathrm{L}^\infty(I^e_{int})} \leq \|p\|_{\mathrm{L}^\infty(K^e)} \leq C\left(h_{K^e}\right)^{-1}(l+1)^2\|p\|_{0,K^e}.$$

The conclusion then follows by using that $h_{K^e}^{-1} \leq (h_e^{\perp})^{-1} \leq \widetilde{r}_e/\widetilde{H}_e$. □

The previous result, combined with the estimates in Lemma 1.6, implies

$$\widetilde{C}^e_{ext} \leq C(C^e_1)^{-1}C^e_2(l+1)^2\eta^l_e.$$

APPENDIX B

Extension of the analysis to three dimensions

Our goal in this section is to comment on the main consideration to extend the analyses of Chapters 1 and 2 to three dimensions.

B.1 Norm equivalence

In this section we extend the proof of Lemma 3.4 in Chapter 1 to show the equivalence between the $\mathbb{L}^2(\widetilde{T}^e_{ext})$-norm and the norm in (1.14) (see also (2.28)) in three dimensions. To that end, let $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$ be the vertices of a boundary face e. A point $\mathbf{x} \in e$, can be parameterized as

$$\mathbf{x}(\theta_1,\theta_2) = \mathbf{p}_1 + \theta_1(\mathbf{p}_2-\mathbf{p}_1) + \theta_2(\mathbf{p}_3-\mathbf{p}_1), \quad \theta_1,\theta_2 \geq 0, \quad \theta_1+\theta_2 \leq 1.$$

Then, the tangent vector $\widehat{\mathbf{m}}$ corresponding to the transferring segment associated to $\mathbf{x}$ can be written as

$$\widehat{\mathbf{m}}(\theta_1,\theta_2) = \mathbf{m}^{\mathbf{p}_1} + \theta_1\left(\mathbf{m}^{\mathbf{p}_2}-\mathbf{m}^{\mathbf{p}_1}\right) + \theta_2\left(\mathbf{m}^{\mathbf{p}_3}-\mathbf{m}^{\mathbf{p}_1}\right).$$

By setting $\alpha(\theta_1,\theta_2) := |\widehat{\mathbf{m}}(\theta_1,\theta_2)|$ if $\widehat{\mathbf{m}}(\theta_1,\theta_2) \neq \mathbf{0}$ and $\alpha(\theta_1,\theta_2) = 1$, otherwise, we define the normalized vector $\mathbf{m}(\theta_1,\theta_2) = \widehat{\mathbf{m}}(\theta_1,\theta_2)/\alpha(\theta_1,\theta_2)$. For $\mathbf{y} \in \widetilde{T}^e_{ext}$, we write

$$\mathbf{y}(s,\theta_1,\theta_2) = \mathbf{x}(\theta_1,\theta_2) + s\mathbf{m}(\theta_1,\theta_2), \quad s \in [0,\ell(\theta_1,\theta_2)], \quad \theta_1,\theta_2 \geq 0, \quad \theta_1+\theta_2 \leq 1,$$

where $\ell(\theta_1,\theta_2)$ stands for the length of the transferring path associated to $\mathbf{x}(\theta_1,\theta_2)$. Then, the Jacobian of the mapping $(s,\theta_1,\theta_2) \mapsto \mathbf{y}(s,\theta_1,\theta_2)$ is given by

$$\begin{aligned}|\mathbf{J}(s,\theta_1,\theta_2)| = \Big|&2|e|\mathbf{m}(\theta_1,\theta_2)\cdot\mathbf{n}_e + \frac{s}{\alpha(\theta_1,\theta_2)}\mathbf{m}(\theta_1,\theta_2)\cdot[(\mathbf{p}_2-\mathbf{p}_1)\times(\mathbf{m}^{\mathbf{p}_3}-\mathbf{m}^{\mathbf{p}_1})\\ &-(\mathbf{p}_3-\mathbf{p}_1)\times(\mathbf{m}^{\mathbf{p}_2}-\mathbf{m}^{\mathbf{p}_1})] + \left(\frac{2s}{\alpha(\theta_1,\theta_2)}\right)^2\mathbf{m}(\theta_1,\theta_2)\cdot(\mathbf{m}^{\mathbf{p}_2}-\mathbf{m}^{\mathbf{p}_1})\times(\mathbf{m}^{\mathbf{p}_3}-\mathbf{m}^{\mathbf{p}_1})\Big|\end{aligned}$$

and the $\mathbb{L}^2(\widetilde{T}^e_{ext})$-norm of a function $\mathbf{v}$ can be written as

$$\|\mathbf{v}\|^2_{0,\widetilde{T}^e_{ext}} = \int_{\widetilde{T}^e_{ext}} |\mathbf{v}(\mathbf{y})|^2\,d\mathbf{y} = \int_0^1\int_0^{1-\theta_1}\int_0^{\ell(\theta_1,\theta_2)} |\mathbf{v}(\mathbf{y}(s,\theta_1,\theta_2))|^2|\mathbf{J}(s,\theta_1,\theta_2)|\,ds d\theta_2 d\theta_1.$$

We have then the following result.

Lemma B.1. *Let* $\mathbf{p} \in \mathbb{L}^2(\widetilde{T}^e_{ext})$ *and consider the following conditions:*

i) $\mathbf{m}^{\mathbf{p}_1} \cdot \mathbf{m}^{\mathbf{p}_2} \geq 0$, $\mathbf{m}^{\mathbf{p}_2} \cdot \mathbf{m}^{\mathbf{p}_3} \geq 0$, *and* $\mathbf{m}^{\mathbf{p}_1} \cdot \mathbf{m}^{\mathbf{p}_3} \geq 0$,

ii) there exists a constant $\delta_e > 0$, *independent of* h, *such that* $\mathbf{m}(\theta_1, \theta_2) \cdot \mathbf{n}_e \geq \delta_e > 0$ *for all* $\theta_1, \theta_2 \geq 0$, *satisfying* $\theta_1 + \theta_2 \leq 1$; *and*

iii) $\mathbf{m}^{\mathbf{p}_1} \cdot [(\mathbf{p}_2 - \mathbf{p}_1) \times \mathbf{m}^{\mathbf{p}_3}] \geq 0$, $\mathbf{m}^{\mathbf{p}_2} \cdot [(\mathbf{p}_2 - \mathbf{p}_1) \times \mathbf{m}^{\mathbf{p}_3}] \geq 0$, $\mathbf{m}^{\mathbf{p}_2} \cdot [\mathbf{m}^{\mathbf{p}_1} \times (\mathbf{p}_2 - \mathbf{p}_1)] \geq 0$, $\mathbf{m}^{\mathbf{p}_1} \cdot [\mathbf{m}^{\mathbf{p}_2} \times (\mathbf{p}_3 - \mathbf{p}_1)] \geq 0$, $\mathbf{m}^{\mathbf{p}_3} \cdot [\mathbf{m}^{\mathbf{p}_2} \times (\mathbf{p}_3 - \mathbf{p}_1)] \geq 0$, $\mathbf{m}^{\mathbf{p}_3} \cdot [(\mathbf{p}_3 - \mathbf{p}_1) \times \mathbf{m}^{\mathbf{p}_1}] \geq 0$, *and* $\mathbf{m}^{\mathbf{p}_1} \cdot [\mathbf{m}^{\mathbf{p}_2} \times \mathbf{m}^{\mathbf{p}_3}] \geq 0$.

If i) is satisfied, there exists $C^e_1 > 0$ *independent of* h *such that* $\|\mathbf{v}\|_{0,\widetilde{T}^e_{ext}} \leq C^e_1 |||\mathbf{v}|||_e$. *Moreover, if ii) and iii) holds, then* $|||\mathbf{v}|||_e \leq C^e_2 \|\mathbf{v}\|_{0,\widetilde{T}^e_{ext}}$, *where* $C^e_2 > 0$ *is also independent of* h.

Proof. Note that

$$[\alpha(\theta_1,\theta_2)]^2 = (1-\theta_1-\theta_2)^2+\theta_1^2+\theta_2^2+2(1-\theta_1-\theta_2)\theta_1\mathbf{m}^{\mathbf{p}_1}\cdot\mathbf{m}^{\mathbf{p}_2}+2\theta_1\theta_2\mathbf{m}^{\mathbf{p}_2}\cdot\mathbf{m}^{\mathbf{p}_3}+2(1-\theta_1-\theta_2)\theta_2\mathbf{m}^{\mathbf{p}_1}\cdot\mathbf{m}^{\mathbf{p}_3}$$

and, by i), we have

$$[\alpha(\theta_1,\theta_2)]^2 \geq (1-\theta_1-\theta_2)^2 + \theta_1^2 + \theta_2^2 \geq \frac{1}{3}.$$

Now, we observe that $\ell(\theta_1,\theta_2) \leq \widetilde{H}_e \leq \widetilde{r}_e h_{T^e}$. Moreover, by regularity of the mesh, there exists γ such that $h_{T^e} \leq \gamma h_e$, and hence

$$|\mathbf{J}(s,\theta_1,\theta_2)| \leq 2|e| + 4h_e\frac{\ell(\theta_1,\theta_2)}{\alpha(\theta_1,\theta_2)} + 16\left(\frac{\ell(\theta_1,\theta_2)}{\alpha(\theta_1,\theta_2)}\right)^2 \leq 2|e|\left(1 + C\left(\sqrt{3}\widetilde{r}_e\gamma^2 + 24\left(\widetilde{r}_e\gamma\right)^2\right)\right) \lesssim 2|e|.$$

Combining this expression and (B.1), we deduce

$$\|\mathbf{v}\|^2_{0,\widetilde{T}^e_{ext}} \lesssim 2|e|\int_0^1\int_0^{1-\theta_1}\int_0^{\ell(\theta_1,\theta_2)} |\mathbf{v}(\mathbf{y}(s,\theta_1,\theta_2))|^2\, ds d\theta_2 d\theta_1 = |||\mathbf{v}|||^2_e,$$

which implies the first assertion of the lemma with C^e_1 the constant hidden in the symbol $\lesssim$. Finally, by algebraic manipulations and assumptions ii) and iii), it is possible to obtain $|\mathbf{J}(s,\theta_1,\theta_2)| \geq 2|e|\delta_e > 0$ and the second assertion follows. □

APPENDIX C

Additional experiments

This section addresses aspects of the method in Chapter 2 that are not covered by our theory, but we consider they are important to take into account. More precisely, we study the condition number, sparsity properties and dependence on the polynomial degree.

C.1 Density and condition number of the matrix

We first notice that (2.16) can be expressed as a linear system

$$\begin{bmatrix} (\mathbf{A}+\mathbf{D}) & \mathbf{B}^T \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{\boldsymbol{\sigma}} \\ \mathbf{X}_{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{G} \\ \mathbf{F} \end{bmatrix},$$

where $\mathbf{X}_{\boldsymbol{\sigma}}$ (resp. $\mathbf{X}_{\mathbf{u}}$) are the coefficients of $\boldsymbol{\sigma}_h$ (resp. $\mathbf{u}_h$) expanded with respect to its corresponding finite element space, whereas $\mathbf{A}$, $\mathbf{D}$ and $\mathbf{B}$ are the matrices related respectively to the bilinear forms a_h, d_h and b_h. We shall denote by $\mathbf{M_D}$ the above matrix and write $\mathbf{M_0}$ when $\mathbf{D}=\mathbf{0}$. Notice that $\mathbf{M_0}$ corresponds to the standard block symmetric matrix of mixed formulations for polyhedral domains. In other words, $\mathbf{M_D}$ can be seen as a perturbation of $\mathbf{M_0}$ due to the transferring technique of our data. In this experiments we want to quantify how this perturbation affects some properties of the linear system.

We consider Ω to be a circle of center at origin and radius 0.75, meshed by following the two procedures indicated along Section 2.6 in such a way that $d(\Gamma,\Gamma_h) \lesssim h$ and $d(\Gamma,\Gamma_h) \lesssim h^2$. In Table C.1 we compare the number of nonzero entries of $\mathbf{M_0}$ vs $\mathbf{M_D}$ for $k=0,1,2,3,4$ and four different meshes. We observe no significant differences between $\mathbf{M_0}$ and $\mathbf{M_D}$ in this aspect.

Now, we compare the condition number of matrices $\mathbf{M_0}$ and $\mathbf{M_D}$ denoted by $\boldsymbol{\kappa_0}$ and $\boldsymbol{\kappa_D}$, respectively. In Figure C.1 we display the ratio $\boldsymbol{\kappa_0}/\boldsymbol{\kappa_D}$ for the meshes considered in the experiment of Table C.1. We observe that when $d(\Gamma,\Gamma_h) \lesssim h^2$, there is no significant differences between $\boldsymbol{\kappa_0}$ and $\boldsymbol{\kappa_D}$. On the other hand, when the distance between Γ and Γ_h is of order h, in lowest order case, $\boldsymbol{\kappa_D}$ behaves as $\boldsymbol{\kappa_0}$. However, when $k \geq 1$, $\boldsymbol{\kappa_D}$ is much larger than $\boldsymbol{\kappa_0}$. We think this bad behavior of the condition number of $\mathbf{M_D}$ might be related to the fact that Assumption **(A2)** is not necessarily satisfied in the latter case, because $\widetilde{C}^e_{ext}$ depends on the polynomial degree as shown in Appendix A.

	N	$\mathbf{M_0}$	$\mathbf{M_D}$	N	$\mathbf{M_0}$	$\mathbf{M_D}$
$k=0$	72	2.3856%	2.3856%	78	2.2812%	2.2812%
	360	0.5006%	0.5007%	374	0.4951%	0.4951%
$k=1$	72	1.5430%	1.5430%	78	1.4645%	1.4645%
	360	0.3184%	0.3186%	374	0.3132%	0.3132%
$k=2$	72	1.2978%	1.2978%	78	1.2200%	1.2200%
	360	0.2661%	0.2661%	374	0.2594%	0.2594%
$k=3$	72	1.1995%	1.1995%	78	1.1236%	1.1236%
	360	0.2449%	0.2449%	374	0.2380%	0.2380%
$k=4$	72	1.1448%	1.1448%	78	1.0695%	1.0695%
	360	0.2330%	0.2330%	374	0.2259%	0.2259%

Table C.1: Percentage of nonzero entries in the matrices $\mathbf{M_0}$ and $\mathbf{M_D}$. Columns 2-4 corresponds to the case $d(\Gamma, \Gamma_h) \lesssim h$, whereas columns 5-7 are the results for $d(\Gamma, \Gamma_h) \lesssim h^2$. N is the number of triangles of the mesh. (table produced by the author)

A rigorous analysis of the condition number and preconditioning techniques will be subject of future work.

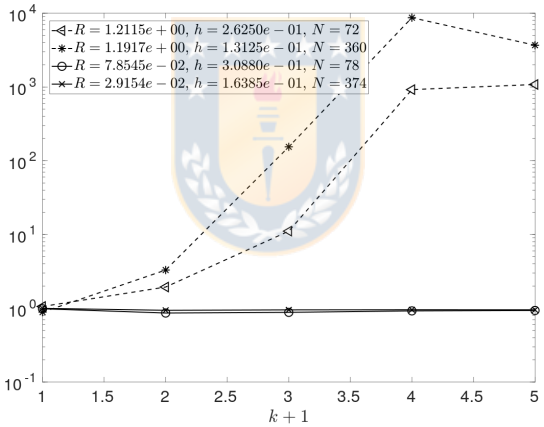


Figure C.1: Semi-log plot of $(k+1)$ vs $\boldsymbol{\kappa_0}/\boldsymbol{\kappa_D}$ for $k = 0, 1, 2, 3, 4$. Dashed lines corresponds to the case $d(\Gamma, \Gamma_h) \lesssim h$, whereas solid lines are the results for $d(\Gamma, \Gamma_h) \lesssim h^2$. (figure produced by the author)

C.2 k-dependence of the method

For a fixed mesh we explore the performance of the method for different polynomial degrees. We consider the same setting of Appendix C.1. In Figure C.2 we observe that the log of the errors linearly decreases with the polynomial degree. The results for the case $k = 6$ are affected by rounded-off errors.

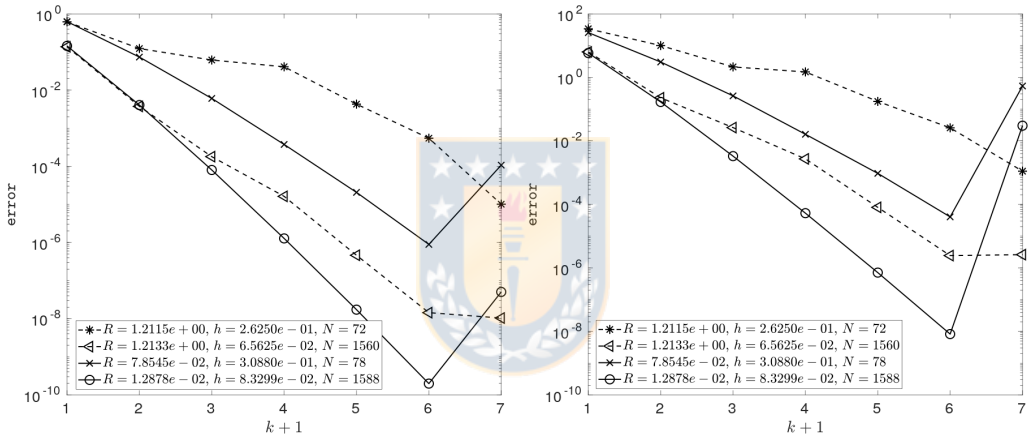


Figure C.2: Left, log-log plot of $(k+1)$ vs $\mathrm{e}(\mathbf{u})$ and right, vs $\mathrm{e}(\boldsymbol{\sigma})$, for $k = 0, 1, \ldots, 6$. Dashed lines corresponds to the case $d(\Gamma, \Gamma_h) \lesssim h$, whereas solid lines are the results for $d(\Gamma, \Gamma_h) \lesssim h^2$. (figure produced by the author)

References

[1] E. Ahmed, F. A. Radu, and J. M. Nordbotten, *Adaptive poromechanics computations based on a posteriori error estimates for fully mixed formulations of Biot's consolidation model*, Comput. Methods Appl. Mech. Engrg., 347 (2019), pp. 264–294.

[2] M. S. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. E. Rognes, and G. N. Wells, *The fenics project version 1.5*, Archive of Numerical Software, 3 (2015), pp. 9–23.

[3] M. Alvarez, G. N. Gatica, and R. Ruiz-Baier, *A posteriori error analysis for a viscous flow-transport problem*, ESAIM Math. Model. Numer. Anal., 50 (2016), pp. 1789–1816.

[4] V. Anaya, Z. de Wijn, D. Mora, and R. Ruiz-Baier, *Mixed displacement–rotation–pressure formulations for linear elasticity*, Computer Methods in Applied Mechanics and Engineering, 344 (2019), pp. 71 – 94.

[5] R. Araya, T. P. Barrios, G. N. Gatica, and N. Heuer, *A posteriori error estimates for a mixed-fem formulation of a non-linear elliptic problem*, Computer methods in applied mechanics and engineering, 191 (2002), pp. 2317–2336.

[6] T. Arbogast, M. A. Hesse, and A. L. Taicher, *Mixed methods for two-phase darcy–stokes mixtures of partially melted materials with regions of zero porosity*, SIAM Journal on Scientific Computing, 39 (2017), pp. B375–B402.

[7] T. Arbogast, M. A. Hesse, and A. L. Taicher, *Mixed methods for two-phase Darcy-Stokes mixtures of partially melted materials with regions of zero porosity*, SIAM J. Sci. Comput., 39 (2017), pp. B375–B402.

[8] M. G. Armentano, C. Padra, and M. Scheble, *An hp finite element adaptive scheme to solve the Poisson problem on curved domains*, Comput. Appl. Math., 34 (2015), pp. 705–727.

[9] D. N. Arnold, *An interior penalty finite element method with discontinuous elements*, SIAM J. Numer. Anal., 19 (1982), pp. 742–760.

[10] D. N. Arnold, F. Brezzi, and J. Douglas, Jr., *PEERS: a new mixed finite element for plane elasticity*, Japan J. Appl. Math., 1 (1984), pp. 347–367.

[11] D. N. Arnold, F. Brezzi, and M. Fortin, *A stable finite element for the Stokes equations*, Calcolo, 21 (1984), pp. 337–344 (1985).

[12] D. N. Arnold, J. Douglas, Jr., and C. P. Gupta, *A family of higher order mixed finite element methods for plane elasticity*, Numer. Math., 45 (1984), pp. 1–22.

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

[14] I. Babuska, B. A. Szabo, and I. N. Katz, *The p-version of the finite element method*, SIAM journal on numerical analysis, 18 (1981), pp. 515–545.

[15] I. Babuška and M. R. Dorr, *Error estimates for the combined h and p versions of the finite element method*, Numer. Math., 37 (1981), pp. 257–277.

[16] I. Babuška and M. Suri, *Locking effects in the finite element approximation of elasticity problems*, Numer. Math., 62 (1992), pp. 439–463.

[17] T. P. Barrios, G. N. Gatica, M. González, and N. Heuer, *A residual based a posteriori error estimator for an augmented mixed finite element method in linear elasticity*, M2AN Math. Model. Numer. Anal., 40 (2006), pp. 843–869 (2007).

[18] S. I. Barry and G. N. Mercer, *Exact Solutions for Two-Dimensional Time-Dependent Flow and Deformation Within a Poroelastic Medium*, Journal of Applied Mechanics, 66 (1999), pp. 536–540.

[19] P. J. Basser, *Interstitial pressure, volume, and flow during infusion into brain tissue*, Microvascular Research, 44 (1992), pp. 143 – 165.

[20] A. Berger, R. Scott, and G. Strang, *Approximate boundary conditions in the finite element method*, in Symposia Mathematica, Vol. X (Convegno di Analisi Numerica, INDAM, Rome, 1972), 1972, pp. 295–313.

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

[22] F. Bertrand, S. Münzenmaier, and G. Starke, *First-order system least squares on curved boundaries: higher-order Raviart-Thomas elements*, SIAM J. Numer. Anal., 52 (2014), pp. 3165–3180.

[23] F. Bertrand and G. Starke, *Parametric Raviart-Thomas elements for mixed methods on domains with curved surfaces*, SIAM J. Numer. Anal., 54 (2016), pp. 3648–3667.

[24] S. Beuchler, V. Pillwein, and S. Zaglmayr, *Sparsity optimized high order finite element functions for $H(\mathrm{div})$ on simplices*, Numer. Math., 122 (2012), pp. 197–225.

[25] M. A. Biot, *General theory of three-dimensional consolidation*, Journal of applied physics, 12 (1941), pp. 155–164.

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

[27] J. J. Blair, *Bounds for the change in the solutions of second order elliptic PDE's when the boundary is perturbed*, SIAM J. Appl. Math., 24 (1973), pp. 277–285.

[28] J. J. Blair, *Higher order approximations to the boundary conditions for the finite element method*, Math. Comput., 30 (1976), pp. 250–262.

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

[30] D. Boffi, F. Brezzi, L. F. Demkowicz, R. G. Durán, R. S. Falk, and M. Fortin, *Mixed finite elements, compatibility conditions, and applications*, vol. 1939 of Lecture Notes in Mathematics, Springer-Verlag, Berlin; Fondazione C.I.M.E., Florence, 2008.

[31] D. Boffi, F. Brezzi, and M. Fortin, *Mixed finite element methods and applications*, vol. 44 of Springer Series in Computational Mathematics, Springer, Heidelberg, 2013.

[32] S. P. A. Bordas, E. Burman, M. G. Larson, and M. A. Olshanskii, eds., *Geometrically unfitted finite element methods and applications*, vol. 121 of Lecture Notes in Computational Science and Engineering, Springer, Cham, 2017. Held January 6–8, 2016.

[33] J. H. Bramble, T. Dupont, and V. Thomée, *Projection methods for Dirichlet's problem in approximating polygonal domains with boundary-value corrections*, Math. Comp., 26 (1972), pp. 869–879.

[34] J. H. Bramble, T. Dupont, and V. Thomée, *Projection methods for Dirichlet's problem in approximating polygonal domains with boundary-value corrections*, Math. Comp., 26 (1972), pp. 869–879.

[35] J. H. Bramble and J. T. King, *A robust finite element method for nonhomogeneous Dirichlet problems in domains with curved boundaries*, Math. Comp., 63 (1994), pp. 1–17.

[36] J. H. Bramble and J. T. King, *A finite element method for interface problems in domains with smooth boundaries and interfaces*, Adv. Comput. Math., 6 (1996), pp. 109–138 (1997).

[37] S. C. Brenner and L. R. Scott, *The mathematical theory of finite element methods*, vol. 15 of Texts in Applied Mathematics, Springer, New York, third ed., 2008.

[38] H. Brezis, *Functional analysis, Sobolev spaces and partial differential equations*, Universitext, Springer, New York, 2011.

[39] F. Brezzi, J. Douglas, Jr., and L. D. Marini, *Recent results on mixed finite element methods for second order elliptic problems*, in Vistas in applied mathematics, Transl. Ser. Math. Engrg., Optimization Software, New York, 1986, pp. 25–43.

[40] F. Brezzi and R. S. Falk, *Stability of higher-order Hood-Taylor methods*, SIAM J. Numer. Anal., 28 (1991), pp. 581–590.

[41] F. Brezzi and M. Fortin, *Mixed and hybrid finite element methods*, vol. 15 of Springer Series in Computational Mathematics, Springer-Verlag, New York, 1991.

[42] E. Burman, S. Claus, P. Hansbo, M. G. Larson, and A. Massing, *CutFEM: discretizing geometry and partial differential equations*, Internat. J. Numer. Methods Engrg., 104 (2015), pp. 472–501.

[43] E. Burman and P. Hansbo, *Fictitious domain finite element methods using cut elements: I. A stabilized Lagrange multiplier method*, Comput. Methods Appl. Mech. Engrg., 199 (2010), pp. 2680–2686.

[44] E. Burman and P. Hansbo, *Fictitious domain finite element methods using cut elements: II. A stabilized Nitsche method*, Appl. Numer. Math., 62 (2012), pp. 328–341.

[45] Z. Cai, B. Lee, and P. Wang, *Least-squares methods for incompressible Newtonian fluid flow: linear stationary problems*, SIAM J. Numer. Anal., 42 (2004), pp. 843–859.

[46] Z. Cai, C. Tong, P. S. Vassilevski, and C. Wang, *Mixed finite element methods for incompressible flow: stationary Stokes equations*, Numer. Methods Partial Differential Equations, 26 (2010), pp. 957–978.

[47] Z. Cai and Y. Wang, *Pseudostress-velocity formulation for incompressible Navier-Stokes equations*, Internat. J. Numer. Methods Fluids, 63 (2010), pp. 341–356.

[48] C. Carstensen, *A posteriori error estimate for the mixed finite element method*, Math. Comp., 66 (1997), pp. 465–476.

[49] C. Carstensen and G. Dolzmann, *A posteriori error estimates for mixed FEM in elasticity*, Numer. Math., 81 (1998), pp. 187–209.

[50] S. Caucao, D. Mora, and R. Oyarzúa, *A priori and a posteriori error analysis of a pseudostress-based mixed formulation of the Stokes problem with varying density*, IMA J. Numer. Anal., 36 (2016), pp. 947–983.

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

[52] J. Cheung, M. Perego, P. Bochev, and M. Gunzburger, *Optimally accurate higher-order finite element methods for polytopial approximations of domains with smooth boundaries*, Math. Comp., 88 (2019), pp. 2187–2219.

[53] P. G. Ciarlet, *The finite element method for elliptic problems*, vol. 40 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002. Reprint of the 1978 original [North-Holland, Amsterdam; MR0520174 (58 #25001)].

[54] P. Clément, *Approximation by finite element functions using local regularization*, Rev. Française Automat. Informat. Recherche Opérationnelle Sér., 9 (1975), pp. 77–84.

[55] B. Cockburn, J. Gopalakrishnan, and F.-J. Sayas, *A projection-based error analysis of HDG methods*, Math. Comp., 79 (2010), pp. 1351–1367.

[56] B. Cockburn, D. Gupta, and F. Reitich, *Boundary-conforming discontinuous Galerkin methods via extensions from subdomains*, J. Sci. Comput., 42 (2010), pp. 144–184.

[57] B. Cockburn, W. Qiu, and M. Solano, *A priori error analysis for HDG methods using extensions from subdomains to achieve boundary conformity*, Math. Comp., 83 (2014), pp. 665–699.

[58] B. Cockburn, F.-J. Sayas, and M. Solano, *Coupling at a distance HDG and BEM*, SIAM J. Sci. Comput., 34 (2012), pp. A28–A47.

[59] B. Cockburn and M. Solano, *Solving Dirichlet boundary-value problems on curved domains by extensions from subdomains*, SIAM J. Sci. Comput., 34 (2012), pp. A497–A519.

[60] B. Cockburn and M. Solano, *Solving convection-diffusion problems on curved domains by extensions from subdomains*, J. Sci. Comput., 59 (2014), pp. 512–543.

[61] B. Cockburn and W. Zhang, *A posteriori error estimates for HDG methods*, J. Sci. Comput., 51 (2012), pp. 582–607.

[62] B. Cockburn and W. Zhang, *An a posteriori error estimate for the variable-degree Raviart-Thomas method*, Math. Comp., 83 (2014), pp. 1063–1082.

[63] E. Colmenares, G. N. Gatica, and R. Oyarzúa, *Analysis of an augmented mixed-primal formulation for the stationary Boussinesq problem*, Numer. Methods Partial Differential Equations, 32 (2016), pp. 445–478.

[64] T. A. Davis, *Algorithm 832: UMFPACK V4.3—an unsymmetric-pattern multifrontal method*, ACM Trans. Math. Software, 30 (2004), pp. 196–199.

[65] M. Dawson, R. Sevilla, and K. Morgan, *The application of a high-order discontinuous Galerkin time-domain method for the computation of electromagnetic resonant modes*, Appl. Math. Model., 55 (2018), pp. 94–108.

[66] S. Deng and W. Cai, *Analysis and application of an orthogonal nodal basis on triangles for discontinuous spectral element methods*, Applied Numerical Analysis & Computational Mathematics, 2 (2005), pp. 326–345.

[67] D. A. Di Pietro and A. Ern, *Mathematical aspects of discontinuous Galerkin methods*, vol. 69 of Mathématiques & Applications (Berlin) [Mathematics & Applications], Springer, Heidelberg, 2012.

[68] C. Domínguez, G. N. Gatica, and S. Meddahi, *A posteriori error analysis of a fully-mixed finite element method for a two-dimensional fluid-solid interaction problem*, J. Comput. Math., 33 (2015), pp. 606–641.

[69] A. Ern and S. Meunier, *A posteriori error analysis of Euler-Galerkin approximations to coupled elliptic-parabolic problems*, M2AN Math. Model. Numer. Anal., 43 (2009), pp. 353–375.

[70] Q. Fang, *Mesh-based monte carlo method using fast ray-tracing in plücker coordinates*, Biomed. Opt. Express, 1 (2010), pp. 165–175.

[71] Z. E. A. Fellah, N. Sebaa, M. Fellah, E. Ogam, F. G. Mitri, C. Depollier, and W. Lauriks, *Application of the biot model to ultrasound in bone: Inverse problem*, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 55 (2008), pp. 1516–1523.

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

[73] G. N. Gatica, *A simple introduction to the mixed finite element method*, SpringerBriefs in Mathematics, Springer, Cham, 2014. Theory and applications.

[74] G. N. Gatica, *A note on stable helmholtz decompositions in 3d*, Applicable Analysis, 0 (2018), pp. 1–12.

[75] G. N. Gatica, L. F. Gatica, and A. Márquez, *Analysis of a pseudostress-based mixed finite element method for the Brinkman model of porous media flow*, Numer. Math., 126 (2014), pp. 635–677.

[76] G. N. Gatica, G. C. Hsiao, S. Meddahi, and F.-J. Sayas, *On the dual-mixed formulation for an exterior Stokes problem*, ZAMM Z. Angew. Math. Mech., 93 (2013), pp. 437–445.

[77] G. N. Gatica, G. C. Hsiao, S. Meddahi, and F. J. Sayas, *New developments on the coupling of mixed-FEM and BEM for the three-dimensional exterior Stokes problem*, Int. J. Numer. Anal. Model., 13 (2016), pp. 457–492.

[78] G. N. Gatica, A. Márquez, and M. A. Sánchez, *Analysis of a velocity–pressure–pseudostress formulation for the stationary Stokes equations*, Computer Methods in Applied Mechanics and Engineering, 199 (2010), pp. 1064–1079.

[79] G. N. Gatica and S. Meddahi, *On the coupling of MIXED-FEM and BEM for an exterior Helmholtz problem in the plane*, Numer. Math., 100 (2005), pp. 663–695.

[80] G. N. Gatica, R. Oyarzúa, and F.-J. Sayas, *A residual-based a posteriori error estimator for a fully-mixed formulation of the Stokes-Darcy coupled problem*, Comput. Methods Appl. Mech. Engrg., 200 (2011), pp. 1877–1891.

[81] G. N. Gatica, R. Oyarzúa, and F.-J. Sayas, *A twofold saddle point approach for the coupling of fluid flow with nonlinear porous media flow*, IMA J. Numer. Anal., 32 (2012), pp. 845–887.

[82] G. N. Gatica, R. Ruiz-Baier, and G. Tierra, *A mixed finite element method for Darcy's equations with pressure dependent porosity*, Math. Comp., 85 (2016), pp. 1–33.

[83] G. N. Gatica and E. P. Stephan, *A mixed-FEM formulation for nonlinear incompressible elasticity in the plane*, Numer. Methods Partial Differential Equations, 18 (2002), pp. 105–128.

[84] V. Girault and P.-A. Raviart, *Finite element methods for Navier-Stokes equations*, vol. 5 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1986. Theory and algorithms.

[85] L. Guo, J. C. Vardakis, T. Lassila, M. Mitolo, N. Ravikumar, D. Chou, M. Lange, A. Sarrami-Foroushani, B. J. Tully, Z. A. Taylor, et al., *Subject-specific multi-poroelastic model for exploring the risk factors associated with the early stages of alzheimer's disease*, Interface focus, 8 (2017), p. 20170019.

[86] A. Hansbo and P. Hansbo, *An unfitted finite element method, based on Nitsche's method, for elliptic interface problems*, Comput. Methods Appl. Mech. Engrg., 191 (2002), pp. 5537–5552.

[87] J. S. Hesthaven and T. Warburton, *Nodal high-order methods on unstructured grids. I. Time-domain solution of Maxwell's equations*, J. Comput. Phys., 181 (2002), pp. 186–221.

[88] R. Hiptmair, *Finite elements in computational electromagnetism*, Acta Numer., 11 (2002), pp. 237–339.

[89] G. Jayaraman, *Water transport in the arterial wall–a theoretical study*, Journal of Biomechanics, 16 (1983), pp. 833 – 840.

[90] S. Kang, T. Bui-Thanh, and T. Arbogast, *A hybridized discontinuous Galerkin method for a linear degenerate elliptic equation arising from two-phase mixtures*, Comput. Methods Appl. Mech. Engrg., 350 (2019), pp. 315–336.

[91] J.-M. Kim and R. R. Parizek, *Three-dimensional finite element modelling for consolidation due to groundwater withdrawal in a desaturating anisotropic aquifer system*, International Journal for Numerical and Analytical Methods in Geomechanics, 23 (1999), pp. 549–571.

[92] S. Kumar, R. Oyarzúa, R. Ruiz-Baier, and R. Sandilya, *Conservative discontinuous finite volume and mixed schemes for a new four-field formulation in poroelasticity*, ESAIM: Mathematical Modelling and Numerical Analysis, (2019), https://doi.org/10.1051/m2an/2019063.

[93] J. J. Lee, K.-A. Mardal, and R. Winther, *Parameter-robust discretization and preconditioning of Biot's consolidation model*, SIAM J. Sci. Comput., 39 (2017), pp. A1–A24.

[94] J. J. Lee, E. Piersanti, K.-A. Mardal, and M. E. Rognes, *A mixed finite element method for nearly incompressible multiple-network poroelasticity*, SIAM J. Sci. Comput., 41 (2019), pp. A722–A747.

[95] C. Lehrenfeld and A. Reusken, *High order unfitted finite element methods for interface problems and PDEs on surfaces*, in Transport processes at fluidic interfaces, Adv. Math. Fluid Mech., Birkhäuser/Springer, Cham, 2017, pp. 33–63.

[96] M. Lenoir, *Optimal isoparametric finite elements and error estimates for domains involving curved boundaries*, SIAM J. Numer. Anal., 23 (1986), pp. 562–580.

[97] R. J. LeVeque and Z. Li, *Immersed interface methods for Stokes flow with elastic boundaries or surface tension*, SIAM J. Sci. Comput., 18 (1997), pp. 709–735.

[98] X. Li, H. von Holst, and S. Kleiven, *Influences of brain tissue poroelastic constants on intracranial pressure (icp) during constant-rate infusion*, Computer Methods in Biomechanics and Biomedical Engineering, 16 (2013), pp. 1330–1343.

[99] A. Main and G. Scovazzi, *The shifted boundary method for embedded domain computations. Part I: Poisson and Stokes problems*, J. Comput. Phys., 372 (2018), pp. 972–995.

[100] A. Main and G. Scovazzi, *The shifted boundary method for embedded domain computations. Part II: Linear advection-diffusion and incompressible Navier-Stokes equations*, J. Comput. Phys., 372 (2018), pp. 996–1026.

[101] W. F. Mitchell, *How high a degree is high enough for high order finite elements?*, Procedia Computer Science, 51 (2015), pp. 246–255.

[102] Y. Mori, *Convergence proof of the velocity field for a Stokes flow immersed boundary method*, Comm. Pure Appl. Math., 61 (2008), pp. 1213–1263.

[103] B. Müller, S. Krämer-Eis, F. Kummer, and M. Oberlack, *A high-order discontinuous Galerkin method for compressible flows with immersed boundaries*, Internat. J. Numer. Methods Engrg., 110 (2017), pp. 3–30.

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

[105] J. Nitsche, *Über ein variationsprinzip zur lösung von dirichlet-problemen bei verwendung von teilräumen, die keinen randbedingungen unterworfen sind*, in Abhandlungen aus dem mathematischen Seminar der Universität Hamburg, vol. 36, Springer, 1971, pp. 9–15.

[106] R. Oyarzúa, S. Rhebergen, M. Solano, and P. Zúñiga, *Error analysis of a conforming and locking-free four-field formulation for the stationary biot's model*. Preprint 2019-31, Centro de Investigación en Ingeniería Matemática (CI^{2}MA), Universidad de Concepción, Chile, Preprint available at `https://ci2ma.udec.cl/publicaciones/prepublicaciones/prepublicacion.php?id=381`.

[107] R. Oyarzúa and R. Ruiz-Baier, *Locking-free finite element methods for poroelasticity*, SIAM J. Numer. Anal., 54 (2016), pp. 2951–2973.

[108] R. Oyarzúa, M. Solano, and P. Zúñiga, *A priori and a posteriori error analyses of a high order unfitted mixed-FEM for Stokes flow*. Preprint 2019-15, Centro de Investigación en Ingeniería Matemática (CI^{2}MA), Universidad de Concepción, Chile, Preprint available at `https://ci2ma.udec.cl/publicaciones/prepublicaciones/prepublicacion.php?id=365`.

[109] R. Oyarzúa, M. Solano, and P. Zúñiga, *A High Order Mixed-FEM for Diffusion Problems on Curved Domains*, J. Sci. Comput., 79 (2019), pp. 49–78.

[110] R. Oyarzúa and P. Zúñiga, *Analysis of a conforming finite element method for the Boussinesq problem with temperature-dependent parameters*, J. Comput. Appl. Math., 323 (2017), pp. 71–94.

[111] C. S. Peskin, *Flow patterns around heart valves: A numerical method*, Journal of Computational Physics, 10 (1972), pp. 252 – 271.

[112] C. S. Peskin, *Flow patterns around heart valves: a numerical method*, Journal of computational physics, 10 (1972), pp. 252–271.

[113] P. J. Phillips and M. F. Wheeler, *A coupling of mixed and discontinuous Galerkin finite-element methods for poroelasticity*, Comput. Geosci., 12 (2008), pp. 417–435.

[114] P. J. Phillips and M. F. Wheeler, *Overcoming the problem of locking in linear elasticity and poroelasticity: an heuristic approach*, Computational Geosciences, 13 (2009), pp. 5–12.

[115] A. Plaza and G. F. Carey, *Local refinement of simplicial grids based on the skeleton*, Appl. Numer. Math., 32 (2000), pp. 195–218.

[116] P.R. Amestoy and I.S. Duff and J.-Y. L'Excellent, *Multifrontal parallel distributed symmetric and unsymmetric solvers*, Computer Methods in Applied Mechanics and Engineering, 184 (2000), pp. 501 – 520.

[117] W. Qiu, M. Solano, and P. Vega, *A high order HDG method for curved-interface problems via approximations from straight triangulations*, J. Sci. Comput., 69 (2016), pp. 1384–1407.

[118] P.-A. Raviart and J. M. Thomas, *A mixed finite element method for 2nd order elliptic problems*, in Mathematical aspects of finite element methods (Proc. Conf., Consiglio Naz. delle Ricerche (C.N.R.), Rome, 1975), 1977, pp. 292–315. Lecture Notes in Math., Vol. 606.

[119] R. Riedlbeck, D. A. Di Pietro, A. Ern, S. Granet, and K. Kazymyrenko, *Stress and flux reconstruction in Biot's poro-elasticity problem with application to a posteriori error analysis*, Comput. Math. Appl., 73 (2017), pp. 1593–1610.

[120] B. Rivière, J. Tan, and T. Thompson, *Error analysis of primal discontinuous Galerkin methods for a mixed formulation of the Biot equations*, Comput. Math. Appl., 73 (2017), pp. 666–683.

[121] J. E. Roberts and J.-M. Thomas, *Mixed and hybrid methods*, in Handbook of numerical analysis, Vol. II, Handb. Numer. Anal., II, North-Holland, Amsterdam, 1991, pp. 523–639.

[122] T. Sánchez-Vizuet and M. E. Solano, *A Hybridizable Discontinuous Galerkin solver for the Grad–Shafranov equation*, Comput. Phys. Commun., 235 (2019), pp. 120–132.

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

[124] R. E. Showalter, *Diffusion in deformable media*, in Resource recovery, confinement, and remediation of environmental hazards (Minneapolis, MN, 2000), vol. 131 of IMA Vol. Math. Appl., Springer, New York, 2002, pp. 115–129.

[125] M. Solano and F. Vargas, *An unfitted HDG method for Oseen equations*, Preprint 2019-8, Centro de Investigación en Ingeniería Matemática (CI^{2}MA), Universidad de Concepción, Chile, Preprint available at `https://ci2ma.udec.cl/publicaciones/prepublicaciones/prepublicacion.php?id=358`.

[126] M. Solano and F. Vargas, *A high order HDG method for Stokes flow in curved domains*, J. Sci. Comput., 79 (2019), pp. 1505–1533.

[127] M. Solano, P. Vega, and R. Araya, *Analysis of an adaptive HDG method for the Brinkman problem*, IMA Journal of Numerical Analysis, (2019), https://doi.org/10.1093/imanum/dry031.

[128] E. Stein, R. de Borst, and T. J. R. Hughes, eds., *Encyclopedia of computational mechanics. Vol. 1*, John Wiley & Sons, Ltd., Chichester, 2004. Fundamentals.

[129] E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton Mathematical Series, No. 30, Princeton University Press, Princeton, N.J., 1970.

[130] R. Stenberg, *A family of mixed finite elements for the elasticity problem*, Numer. Math., 53 (1988), pp. 513–538.

[131] G. Strang, *Variational crimes in the finite element method*, in The mathematical foundations of the finite element method with applications to partial differential equations (Proc. Sympos., Univ. Maryland, Baltimore, Md., 1972), 1972, pp. 689–710.

[132] G. Strang and A. E. Berger, *The change in solution due to change in domain*, in Partial differential equations (Proc. Sympos. Pure Math., Vol. XXIII, Univ. California, Berkeley, Calif., 1971), 1973, pp. 199–205.

[133] K. Terzaghi, *Principle of soil mechanics*, Eng. News Record, A Series of Articles, (1925).

[134] V. Thomée, *Polygonal domain approximation in Dirichlet's problem*, J. Inst. Math. Appl., 11 (1973), pp. 33–44.

[135] B. Tully and Y. Ventikos, *Cerebral water transport using multiple-network poroelastic theory: application to normal pressure hydrocephalus*, Journal of Fluid Mechanics, 667 (2011), p. 188–215.

[136] J. C. Vardakis, D. Chou, B. J. Tully, C. C. Hung, T. H. Lee, P.-H. Tsui, and Y. Ventikos, *Investigating cerebral oedema using poroelasticity*, Medical Engineering & Physics, 38 (2016), pp. 48 – 57. Micro and Nano Flows 2014 (MNF2014) - Biomedical Stream.

[137] R. Verfürth, *A posteriori error estimators for the Stokes equations*, Numer. Math., 55 (1989), pp. 309–325.

[138] R. Verfürth, *A posteriori error estimation and adaptive mesh-refinement techniques*, in Proceedings of the Fifth International Congress on Computational and Applied Mathematics (Leuven, 1992), vol. 50, 1994, pp. 67–83.

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

[140] R. Verfürth, *A review of a posteriori error estimation techniques for elasticity problems*, Comput. Methods Appl. Mech. Engrg., 176 (1999), pp. 419–440. New advances in computational methods (Cachan, 1997).

[141] R. Verfürth, *A posteriori error estimation techniques for finite element methods*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, 2013.

[142] M. Wangen, S. Gasda, and T. Bjørnarå, *Geomechanical consequences of large-scale fluid storage in the utsira formation in the north sea*, Energy Procedia, 97 (2016), pp. 486 – 493. European Geosciences Union General Assembly 2016, EGU Division Energy, Resources & the Environment (ERE).

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

[144] J. A. White and R. I. Borja, *Stabilized low-order finite elements for coupled solid-deformation/fluid-diffusion and their application to fault zone transients*, Computer Methods in Applied Mechanics and Engineering, 197 (2008), pp. 4353 – 4366.

[145] S.-Y. Yi, *A coupling of nonconforming and mixed finite element methods for Biot's consolidation model*, Numer. Methods Partial Differential Equations, 29 (2013), pp. 1749–1777.

[146] S.-Y. Yi, *Convergence analysis of a new mixed finite element method for Biot's consolidation model*, Numer. Methods Partial Differential Equations, 30 (2014), pp. 1189–1210.

[147] S.-Y. Yi, *A study of two modes of locking in poroelasticity*, SIAM J. Numer. Anal., 55 (2017), pp. 1915–1936.

[148] J. Young, B. Rivière, C. S. Cox, Jr., and K. Uray, *A mathematical model of intestinal oedema formation*, Math. Med. Biol., 31 (2014), pp. 1–15.