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

# MÉTODOS DE ELEMENTOS FINITOS MIXTOS PARA PROBLEMAS ACOPLADOS EN MECÁNICA DE FLUIDOS

## (MIXED FINITE ELEMENT METHODS FOR COUPLED PROBLEMS IN FLUID MECHANICS)

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

Mario Andrés Álvarez Guadamúz
concepción-chile

2016

Profesor Guía: Gabriel N. Gatica Pérez
CI$^2$MA y Departamento de Ingeniería Matemática
Universidad de Concepción, Chile

Cotutor: Ricardo Ruiz
Mathematical Institute
Oxford University, United Kingdom

**Mixed Finite Element Methods**
**For Coupled Problems in Fluid Mechanics**

Mario Andrés Álvarez Guadamúz

**Directores de Tesis:** Gabriel N. Gatica, Universidad de Concepción, Chile.
Ricardo Ruiz, Oxford University, United Kingdom.

**Director de Programa:** Raimund Bürger, Universidad de Concepción, Chile.

## Comisión evaluadora

Prof.

Prof.

Prof.

Prof.

## Comisión examinadora

Firma: _____
Prof. Raimund Bürger, Universidad de Concepción, Chile.

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

Firma: _____
Prof. Nelson Moraga, Universidad de La Serena, Chile.

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

Firma: _____
Prof. Ricardo Ruiz, Oxford University, United Kingdom.

Calificación: _____

Concepción, 16 de Diciembre de 2016

# Abstract

The main goal of this dissertation is to develop and to apply diverse mathematical and numerical techniques, based on mixed finite element methods and fixed-point strategies, in order to establish the solvability of coupled problems arising in the context of fluid mechanics and that are frequently found in transport processes through viscous flows in porous media.

We firstly derive an augmented mixed–primal finite element method for a viscous flow–transport problem. The model consists in the coupling of a scalar nonlinear convection–diffusion equation with the Stokes problem (written in terms of Cauchy pseudostress) where the viscosity depends on the distribution of the solution to the transport equation, which in turn, exhibits a diffusivity term depending on the gradient norm of that solution. One of the main difficulties to obtain numerical approximations of such a coupling is due to the nonlinearities present in the field equations. A novel result in this work is the derivation of a variational formulation for this system by introducing an augmented mixed approach for the fluid flow coupled with a primal method for the transport equation. The solvability of such formulation is stated by combining fixed point arguments and suitable estimates arising from the connection between certain regularity assumptions, and the Sobolev embedding and Rellich-Kondrachov compactness theorems. As a consequence, the existence of solution of the continuous and discrete schemes is obtained by applying the classical Schauder and Brouwer fixed point theorems, respectively. For the corresponding Galerkin scheme we make use of Raviart-Thomas spaces of order $k$ for the Cauchy pseudostress computation, and continuous piecewise polynomials of degree $\leq k+1$ for both velocity and the scalar field. Afterwards, under the assumption of sufficiently small data and employing suitable Strang-type estimates, we provide optimal *a priori* error bounds. Secondly, we derive two different efficient and reliable residual-based *a posteriori* error estimators for the associated Galerkin scheme. Our analysis of reliability of the proposed estimators is mainly based on the use of suitable ellipticity and inf-sup conditions together with a Helmholtz decomposition, and the local approximation properties of the Clément interpolant and Raviart-Thomas operator. On the other hand, the main tools employed to show their efficiency include suitable inverse inequalities and the localization technique based on triangle-bubble and edge-bubble functions.

Next, the aforementioned methodology is extended to establish the *a priori* and *a posteriori* error analysis of a mixed–primal finite element method for a sedimentation-consolidation system, where the flow patterns are governed by the Brinkman equations with variable viscosity and the diffusion coefficient in the transport equation depends only on the concentration. As a consequence, we derive optimal *a priori* error estimates for the associated Galerkin scheme, together with two efficient and reliable residual-based *a posteriori* error estimators, which constitutes one of the main contributions of this work. Here the Cauchy pseudostress is approximated by Raviart-Thomas spaces of order $k$, whereas the velocity and concentration are computed by using continuous piecewise polynomials of degree $\leq k+1$.

Finally, we close by providing a vorticity-based fully-mixed finite element method to numerically approximate the flow patterns of a viscous fluid within a highly permeable medium, described by Brinkman equations (written in terms of vorticity, velocity and pressure), and its interaction with classical porous media flow governed by Darcy's law. The two domains are separated by an essentially fixed interface, across which the flow passes from viscous to a non-viscous regime. In turn, suitable boundary and transmission conditions on velocities, pressures and the vorticity are considered. In this work we introduce a mixed variational formulation which leads to the incorporation of a Lagrange multiplier enforcing the pressure continuity across the interface. Then, the classical Babuška-Brezzi theory is applied in such a way that the continuous inf-sup conditions of the main bilinear form are obtained by employing the so-called $T$-coercivity approach, which allows us to guarantee the well-posedness of the continuous formulation. Next, we adapt the aforementioned arguments to the discrete case and show that, under suitable assumptions on the discrete subspaces involved, the corresponding Galerkin scheme is well-posed. Here we use a finite element family for which the **curl** of the subspace approximating the vorticity must be contained in the space where the discrete velocity of the fluid lives, and hence Raviart-Thomas and Nédélec finite elements for velocities and vorticity, respectively, become feasible choices. In turn, the pressures and the Lagrange multiplier are approximated, respectively, by discontinuous and continuous piecewise polynomials. Later, we modify the mixed formulation by incorporating a residual-type term arising from the Brinkman momentum equation, and show that the resulting augmented scheme yields a strongly elliptic global bilinear form, without requiring the aforementioned constraint on the discrete space approximating the vorticity. Whereby any finite element subspace of continuous product space can be applied.

For all situations described above we provide several numerical experiments illustrating the satisfactory performance of the proposed methods, and confirming the theoretical results such as the orders of convergence as well as the reliability and efficiency of the proposed *a posteriori* error estimators. These examples also serve to gain insight about the behavior of the underlying physical phenomena of interest.

# Resumen

El objetivo principal de esta disertación es desarrollar y aplicar diversas técnicas matemáticas y numéricas, basadas en métodos de elementos finitos mixtos y extrategias de punto fijo, con el propósito de establecer la solubilidad de problemas acoplados que surgen dentro del contexto de la mecánica de fluidos y que se encuentran con frecuencia en procesos de transporte a través de flujos viscosos en medios porosos.

En primer lugar, se deriva un método de elementos finitos mixto–primal aumentado para un problema acoplado de flujo viscoso con transporte. El modelo consiste en el acoplamiento de un problema de convección-difusión escalar no lineal con el problema de Stokes (escrito en términos del pseudo-esfuerzo de Cauchy) donde la viscosidad depende de la distribución de la solución a la ecuación de transporte, la cual a su vez, exhibe un término de difusión que depende de la norma del gradiente de esa solución. Una de las principales dificultades para obtener aproximaciones numéricas de un acoplamiento de este tipo se debe a las no linealidades presentes en las ecuaciones de campo. Un resultado novedoso en este trabajo es la derivación de una formulación variacional para este acoplamiento mediante la introducción de un enfoque mixto aumentado para el flujo del fluido acoplado con un método primal para el problema de transporte. La solubilidad de dicha formulación se establece combinando argumentos de punto fijo y estimaciones adecuadas que surgen a partir de la conexión entre ciertos supuestos de regularidad y los teoremas de inclusión y compacidad de Sobolev y Rellich-Kondrachov, respectivamente. Como consecuencia de ello, la existencia de solución de los esquemas continuo y discreto se concluye aplicando los teoremas clásicos de punto fijo de Schauder y Brouwer, respectivamente. Para el esquema de Galerkin correspondiente se utilizan espacios de Raviart-Thomas de orden $k$ para el cálculo del pseudo-esfuerzo de Cauchy, y polinomios continuos a trozos de grado $\leq k+1$ para la velocidad y el campo escalar. Luego, bajo datos suficientemente pequeños y aplicando desigualdades adecuadas de tipo Strang, se proporcionan cotas de error *a priori* óptimas. En segundo lugar, se desarrolla un análisis de error *a posteriori* para el esquema de Galerkin asociado. Aquí se derivan dos estimadores de error *a posteriori* residuales para ese esquema, los cuales son confiables y eficientes. El análisis de confiabilidad de los estimadores propuestos se basa principalmente en el uso de elipticidades y condiciones inf-sup adecuadas junto con una descomposición de Helmholtz, y las propiedades de aproximación locales del operador de interpolación de Clément y del operador de Raviart-Thomas. Por otra parte, las principales herramientas empleadas para mostrar sus eficiencias incluyen desigualdades inversas adecuadas y la técnica de localización que se basa en funciones burbujas sobre triángulos y lados.

Luego, la metodología anterior es extendida para establecer el análisis de error *a priori* y *a posteriori* de un método de elementos finitos mixto–primal para un sistema de sedimentación-consolidación, donde los patrones de flujos se rigen por las ecuaciones de Brinkman con viscosidad variable y el coeficiente de difusión en la ecuación de transporte depende solamente de la concentración. En consecuencia,

se derivan estimaciones óptimas de error *a priori* para el esquema de Galerkin asociado, junto con dos estimadores de error *a posteriori* residuales, confiables y eficientes, lo cual constituye una de las principales contribuciones de este trabajo. Aquí el pseudo-esfuerzo de Cauchy es aproximado con espacios de Raviart-Thomas de orden $k$, mientras que la velocidad y la concentración son calculadas usando polinomios continuos a trozos de grado $\leq k+1$.

Finalmente, se cierra esta tesis con el desarrollo de un método de elementos finitos completamente mixto basado en vorticidad para aproximar los patrones de flujo de un fluido viscoso dentro de un medio altamente permeable, descrito por las ecuaciones de Brinkman (escritas en términos de la vorticidad, velocidad y la presión del fluido), y su interacción con un flujo en un medio poroso gobernado por la ley de Darcy. Los dos dominios están separados por una interfaz esencialmente fija, a través de la cual el flujo pasa de viscoso a un régimen no viscoso. A su vez, se consideran condiciones de frontera y de transmisión adecuadas sobre las velocidades, las presiones y la vorticidad. En este trabajo se introduce una formulación variacional mixta, la cual conduce a la incorporación de un multiplicador de Lagrange, el cual impone la continuidad de la presión a través de la interfaz. Despúes, la teoría clásica de Babuška-Brezzi se aplica de tal forma que las condiciones inf-sup continuas de la forma bilineal principal se obtienen empleando la técnica conocida como $T$-coercividad, lo cual nos permite garantizar el buen planteamiento de la formulación continua. Luego, los argumentos mencionados anteriormente son adaptados al caso discreto para mostrar que, bajo supuestos adecuados sobre los subespacios discretos involucrados, el esquema de Galerkin correspondiente está bien puesto. Aquí se utiliza una familia de elementos finitos donde el **curl** del subespacio que aproxima a la vorticidad debe estar contenido en el espacio donde vive la velocidad discreta del fluido, y por lo tanto una elección factible consiste en elementos finitos de Raviart-Thomas y de Nédélec para las velocidades y la vorticidad, respectivamente. A su vez, las presiones y el multiplicador de Lagrange son aproximados, respectivamente, por polinomios continuos y discontinuos a trozos. Posteriormente, la formulación mixta es modificada incorporando un término de tipo residual que surge de la ecuación de momentum de Brinkman, y se muestra que el esquema aumentado resultante, el cual conduce a una forma bilineal global fuertemente elíptica, no requiere la restricción mencionada anteriormente sobre el espacio que aproxima la vorticidad, con lo cual cualquier subespacio de elementos finitos del espacio producto continuo puede ser aplicado.

Para todas las situaciones descritas anteriormente se proporcionan varios experimentos numéricos que ilustran el desempeño satisfactorio de los métodos propuestos, y que confirman los resultados teóricos de convergencia así como la confiabilidad y la eficiencia de los estimadores de error *a posteriori* derivados. Estos ejemplos también sirven para obtener información sobre el comportamiento de los fenómenos físicos subyacentes de interés.

# Agradecimientos

En primer lugar, agradezco infinitamente a Dios quien me ha acompañado y me ha proporcionado la salud y la sabiduría necesarias a lo largo de todos estos años. En segundo lugar, agradezco profundamente a mi madre María F. Guadamúz y a mis hermanos; José, Haide y Luis, por estar siempre pendientes de mi y apoyarme constantemente.

A mi tutor de tesis, el profesor Gabriel N. Gatica, por permitirme crecer en el área de Análisis Numérico e instruirme y guiarme a lo largo del desarrollo de mis estudios doctorales y el de esta tesis. Su apoyo constante, sus observaciones, y sus valiosos consejos técnicos y de edición fueron fundamentales para culminar con éxito esta tesis. A su vez, quiero expresar mi gratitud ya que además de ser un extraordinario expositor en el salón de clases y un excelente guía de tesis, ha sido un gran ser humano y un apreciado amigo a lo largo de todos estos años.

A mi co-tutor de tesis, el profesor Ricardo Ruiz, quien ha sido un pilar importante en mi formación como investigador en Análisis Numérico. Su excelente disposición y asesoramiento constante me permitieron aprender y madurar aspectos claves en la implementación y la simulación de diversos fenómenos en el campo de la mecánica de fluidos. También quiero expresar mi agradecimiento ya que a parte de ser un excelente guía en este proyecto, ha sido un apreciado amigo y un gran anfitrión durante mis estadías de investigación en Lausanne, Suiza y en Oxford, Inglaterra.

A mis estimados compañeros y amigos de doctorado, Filander A. Sequeira y Eligio A. Colmenares. Ha sido un privilegio contar con su amistad y un placer compartir con ustedes este recorrido de doctorado durante todos estos años. Realmente fue muy grato poder convivir y seguir creciendo como persona en la compañía de ustedes, ya que gracias a ello, la lejanía de mis seres queridos se me hizo más llevadera.

A mis amigos chilenos, Felipe Lepe, Joaquín Fernandez, y a todos los miembros que conforman el Departamento de Ingeniería Matemática (DIM) y el Centro de Investigación en Ingeniería Matemática ($\text{CI}^2\text{MA}$). Muchas gracias por su hospitalidad y amistad durante mi estancia en Concepción. Además, agradezco profundamente al Director del programa de Doctorado en Ciencias Aplicadas con mención en Ingeniería Matemática, el profesor Raimund Bürger, por el apoyo constante y su excelente gestión en todos aquellos trámites que tuvieron lugar durante mis estudios doctorales.

Un agradecimiento especial a los compañeros y amigos de la Sección de Matemática de la Universidad de Costa Rica (UCR), Sede de Occidente. En particular a Carlos Ml. Ulate y a Carlos Márquez por impulsarme en el camino de los estudios doctorales. También agradezco mucho a los profesores de la escuela de Matemática de la UCR, William Ugalde y José Rosales, quienes de alguna u otra forma tomaron parte de su tiempo para orientarme y apoyarme hacia los estudios de postgrado fuera de las fronteras de Costa Rica.

A mis compatriotas y amigos, Carlos Solórzano, Evaristo Alvarado y Jesús Rodríguez. Muchas gracias por su apoyo y por haber confiado en mi.

Finalmente, pero no por eso menos importante, agradezco profundamente a la Red Doctoral REDOC.CTA de la Universidad de Concepción (UdeC), al CI$^2$MA y al proyecto Basal en conjunto con el Centro de Modelamiento Matemático (CMM) de la Universidad de Chile, a la Comisión Nacional de Ciencia y Tecnología (CONICYT), y a la Universidad de Costa Rica (UCR), por haber financiado mis estudios doctorales. Su apoyo fue trascendental para que yo pudiese dedicarme de lleno a mis estudios y así culminar con éxito esta tesis doctoral.

Mario Andrés Álvarez Guadamúz

# Contents

**Abstract** iii

**Resumen** v

**Agradecimientos** vii

**Contents** ix

**List of Tables** xii

**List of Figures** xiii

**Introduction** 1

**Introducción** 7

**1 An augmented mixed–primal finite element method for a coupled flow–transport problem** 13

1.1 Introduction . . . 13

1.2 The model problem . . . 14

1.3 The continuous formulation . . . 16

1.3.1 The augmented mixed–primal formulation . . . 16

1.3.2 A fixed point strategy . . . 18

1.3.3 Well-posedness of the uncoupled problems . . . 19

1.3.4 Solvability analysis of the fixed point equation . . . 23

1.4 The Galerkin scheme . . . 27

1.5 A priori error analysis . . . 30

1.6 Numerical results . . . 34

**2 A posteriori error analysis for a viscous flow–transport problem** 41

2.1 Introduction . . . 41
2.2 A coupled viscous flow–transport problem . . . 42
2.2.1 The three-field formulation . . . 42
2.2.2 The augmented mixed–primal formulation . . . 42
2.2.3 The augmented mixed–primal finite element method . . . 43
2.3 A residual-based a posteriori error estimator . . . 44
2.3.1 The local error indicator . . . 44
2.3.2 Reliability . . . 45
2.3.3 Efficiency . . . 55
2.4 A second a posteriori error estimator . . . 60
2.5 Numerical tests . . . 62

**3 A mixed–primal finite element approximation of a sedimentation–consolidation system 70**
3.1 Introduction . . . 70
3.2 The model problem . . . 72
3.2.1 The sedimentation-consolidation system . . . 72
3.3 The variational formulation . . . 73
3.3.1 An augmented mixed–primal approach . . . 73
3.3.2 Fixed point strategy . . . 75
3.3.3 Well-posedness of the uncoupled problems . . . 76
3.3.4 Solvability of the fixed point equation . . . 79
3.4 The Galerkin scheme . . . 82
3.5 A priori error analysis . . . 85
3.6 Numerical tests . . . 90

**4 A posteriori error analysis for a sedimentation-consolidation system 97**
4.1 Introduction . . . 97
4.2 The sedimentation-consolidation system . . . 98
4.2.1 The governing equations . . . 98
4.2.2 The augmented mixed–primal formulation . . . 99
4.2.3 The augmented mixed–primal finite element method . . . 99
4.3 A residual-based a posteriori error estimator . . . 100
4.3.1 The local error indicator . . . 100

4.3.2 Reliability . . . 101
4.3.3 Efficiency . . . 109
4.4 A second a posteriori error estimator . . . 113
4.5 Residual-based a posteriori error estimators: The 3D case . . . 115
4.6 Numerical tests . . . 117

**5 A vorticity-based fully-mixed formulation for the 3D Brinkman-Darcy problem 126**
5.1 Introduction . . . 126
5.2 The coupled problem and its mixed formulation . . . 128
5.3 Solvability analysis of the mixed formulation . . . 131
5.4 The mixed finite element method . . . 137
5.4.1 Preliminaries and main results . . . 137
5.4.2 Specific finite element subspaces . . . 140
5.5 An augmented mixed formulation . . . 147
5.6 Numerical results . . . 150
5.6.1 Accuracy of the mixed and augmented formulations on two embedded cubes . . . 150
5.6.2 Flow into a cracked porous medium . . . 151
5.6.3 Perpendicular infiltration through a porous capsule . . . 151
5.6.4 Flow simulations imposing Dirichlet conditions for the velocity . . . 152

**Conclusions and future works 155**

**Conclusiones y trabajos futuros 158**

**References 161**

# List of Tables


1.1 Example 1: Convergence history and Newton iteration count for the mixed–primal $\mathbf{RT}_k - \mathbf{P}_{k+1} - \mathrm{P}_{k+1}$ approximations of the coupled problem, $k = 0, 1$. Here $N_h$ stands for the number of degrees of freedom associated to each triangulation $\mathcal{T}_h$. . . . . 35

1.2 Example 2: Outputs of interest (Nusselt number, maximum value of the normalized horizontal velocity on the mid-plane attained at $(0.5, x_2^{\infty}, 0.5)$, and maximum value of the normalized vertical velocity and its position $(x_1^{\infty}, 0.5, 0.5)$ on the central horizontal plane, respectively) for different values of the Rayleigh number, and comparison with respect to values from [47, 58]. . . . . 38

2.1 Test 1: convergence history, average Newton iteration count, Picard steps to reach the desired tolerance, effectivity and quasi-effectivity indexes for the mixed–primal $\mathrm{RT}_k - P_{k+1} - P_{k+1}$ approximations of concentration, Cauchy stress, and velocity, with $k = 0, 1$. 63

2.2 Test 2: convergence history, Picard iteration count, effectivity and quasi-effectivity indexes for the mixed–primal approximation of the coupled problem under quasi-uniform, and adaptive refinement according to the indicators introduced in Sections 2.3,2.4. . . 69

3.1 Example 1: Convergence history and Newton iteration count for the mixed–primal $\mathbf{RT}_k - \mathbf{P}_{k+1} - \mathrm{P}_{k+1}$ approximations of the coupled problem, $k = 0, 1$. Here $N_h$ stands for the number of degrees of freedom associated to each triangulation $\mathcal{T}_h$. . . . . 91

3.2 Example 2: Model constants employed in the simulation of steady sedimentation of PMMA into glycerol/water within a contracted cylinder. . . . . 92

4.1 Test 1: convergence history, Picard iteration count, error `e` and quasi-error `m`, effectivity and quasi-effectivity indexes for the approximation of the coupled Brinkman-transport problem, under quasi-uniform, and adaptive refinement according to the indicators introduced in Sections 4.3 and 4.4. . . . . 119

5.1 Example 1: Error history associated to fully mixed and augmented $\mathbb{RT}_0 - \mathbb{ND}_1 - \mathbb{RT}_0 - \mathbf{P}_0 - \mathbf{P}_0 - \mathbf{P}_1$ discretizations (top and middle rows), and augmented $\mathbb{RT}_0 - \mathbb{ND}_2 - \mathbb{RT}_0 - \mathbf{P}_0 - \mathbf{P}_0 - \mathbf{P}_1$ FE family (bottom row) for problem (5.5)-(5.6) on a 3D domain. . . . . 152

# List of Figures


1.1 Example 1: Computed errors $e(\phi), e(\boldsymbol{\sigma}), e(\boldsymbol{u})$ associated to the mixed–primal approximation versus the number of degrees of freedom $N_h$ for $\mathbf{RT}_0-\mathbf{P}_1-\mathrm{P}_1$ and $\mathbf{RT}_1-\mathbf{P}_2-\mathrm{P}_2$ finite elements (left and right, respectively). See values in Table 1.1. . . . . 34

1.2 Example 1: $\mathbf{RT}_0-\mathbf{P}_1-\mathrm{P}_1$ approximation of stress components $\boldsymbol{\sigma}_h$ (top panels), velocity components $\boldsymbol{u}_h$ (with vector directions, bottom left and center, respectively), and scalar field $\phi_h$ (bottom right) solving (1.15). The mesh has 37249 vertices and 74496 triangular elements. . . . . 36

1.3 Example 2: Computed temperature iso-surfaces (top left) and velocity streamlines and vectors colored by magnitude (top center and right, respectively) and principal components of the Cauchy stress (center and bottom rows) for the thermal cavity test. . . . 37

1.4 Example 2: Temperature profiles (solid blue, left axis) and velocity components (dashed green, right axis) at $x_3 = 0.5$, and comparison with respect to benchmark solutions. . 38

1.5 Example 3: Geometry of the clarifier-thickener unit (left panel) and tetrahedral mesh $\mathcal{T}_h$ with 64135 vertices and 370597 elements (right panel). . . . . 39

1.6 Example 3: Simulation of a clarifier-thickener unit. From left-top: Approximated concentration profile, opposite of the trace of the Cauchy stress tensor (which corresponds to the suggested approximation of the pressure field), and velocity components. . . . . 40

2.1 Extension of $\Omega$ to a convex domain $B$ for the Helmholtz decomposition. . . . . 52

2.2 Test 2: approximate solutions obtained with the lowest order method, after six steps of adaptive mesh refinement following the second indicator $\widetilde{\boldsymbol{\theta}}$. Concentration, velocity components, and stress components are depicted. . . . . 64

2.3 Test 2: log-log plot of the total errors vs. degrees of freedom associated to uniform and adaptive mesh refinements using the two proposed indicators. . . . . 65

2.4 Test 2: from left to right, four snapshots of successively refined meshes according to the indicators $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}}$ (top and bottom panels, respectively). . . . . 66

2.5 Test 3: approximate solutions obtained with the lowest order method, after six steps of adaptive mesh refinement following the second indicator $\widetilde{\boldsymbol{\theta}}$. Concentration, velocity components, and stress components are depicted. . . . . 67

2.6 Test 3: from left to right, four snapshots of successively refined meshes according to the indicator $\widetilde{\boldsymbol{\theta}}$. . . . . 68

3.1 Example 1: Computed errors associated to the mixed–primal approximation versus the number of degrees of freedom $N_h$ for $\mathbf{RT}_0 - \mathbf{P}_1 - \mathrm{P}_1$ and $\mathbf{RT}_1 - \mathbf{P}_2 - \mathrm{P}_2$ finite elements (left and right, respectively). Values are detailed in Table 3.1. . . . . 92

3.2 Example 1: $\mathbf{RT}_0 - \mathbf{P}_1 - \mathrm{P}_1$ approximation of pseudo-stress components (top panels), velocity components with vector directions (bottom left and center, respectively), and concentration profile (bottom right), solutions to (3.15). The finest mesh has 204847 vertices and 409692 triangular elements. . . . . 93

3.3 Example 2: Typical dimensions and boundary setting for a two-dimensional computational domain representing the batch sedimentation within a cylinder with a contraction (left), and magnitude of the velocity field shown on the rotationally extruded domain (right). . . . . 94

3.4 Example 2: principal principal components of the Cauchy pseudo-stress (top rows), velocity components $\boldsymbol{u}_h$ with vector directions (bottom left and center), and computed concentration $\phi_h$ (bottom right) for the test of batch sedimentation in a cylinder with a contraction. . . . . 95

3.5 Example 3: sketch of the computational domain $\Omega = [0, 6] \times [0, 4] \times [0, 0.8]$, a coarse mesh, and boundary setting, with $\partial\Omega = \Gamma \cup \Gamma_{\mathrm{in}} \cup \Gamma_{\mathrm{out}}$. . . . . 95

3.6 Example 3: approximate solutions to the so-called Coanda effect using an augmented mixed formulation. Trace of the Cauchy pseudo-stress tensor (left), velocity vectors and streamlines (middle), and concentration profile (right). . . . . 96

4.1 Test 1: approximate solutions obtained with the lowest order method, after six steps of adaptive mesh refinement following the second indicator $\widetilde{\boldsymbol{\theta}}$. Concentration, velocity components, and stress components are depicted. . . . . 120

4.2 Test 1: from left to right, four snapshots of successively refined meshes according to the indicators $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}}$ (top and bottom panels, respectively). . . . . 121

4.3 Test 1: approximate postprocessed pressure and adapted mesh near $(a_1, a_2)$, after six refinement steps following the second estimator $\widetilde{\boldsymbol{\theta}}$. . . . . 121

4.4 Test 2: approximate solutions at 3 (left), 6 (middle) and 12 (right) pseudo-time steps. A lowest order method and mesh adaptive refinement guided by (4.59) were used. . . . 122

4.5 Test 2: adapted meshes at 2, 4, and 6 steps, generated following the second estimator (4.59). . . . . 123

4.6 Test 3: sketch of the clipped domain and different boundaries in a clarifier-thickener device (top left panel), and snapshots of the approximate concentration, postprocessed pressure, and velocity components and streamlines computed with the proposed lowest order mixed-primal method. . . . . 124

4.7 Test 3: zoom on the produced meshes after the first three steps of adaptive refinement using the first estimator as defined in (4.64). . . . . 125

5.1 Sketch of the domains occupied by the incompressible fluid and by the porous medium ($\Omega_{\mathrm{B}}$ and $\Omega_{\mathrm{D}}$, respectively), interface $\Sigma$, and corresponding boundaries. . . . . 129

5.2 Example 1: Two-domain geometry and mesh (top left), approximated Darcy velocity streamlines (top middle), approximated Darcy pressure isosurfaces (top right), zoom of approximated Brinkman vorticity vectors (bottom left), zoom of approximated Brinkman velocity streamlines (bottom middle), and isosurfaces of the computed Brinkman pressure (bottom right). . . . . 150

5.3 Example 2: Two-domain geometry and boundaries (top left), approximated Brinkman vorticity magnitude (top right), approximated velocity magnitude and vectors (bottom left), and computed pressure profiles (bottom right) for the Brinkman-Darcy coupling, using a fully-mixed scheme. . . . . 153

5.4 Example 3: From left to right: Two-domain geometry and boundaries, approximated Brinkman vorticity magnitude and vectors, approximated velocity magnitude and vectors, and computed pressure profile for the Brinkman-Darcy coupling, using an augmented finite element formulation. . . . . 153

5.5 Example 4: Two-domain geometry and boundaries (top left), approximated Brinkman vorticity and Darcy permeability field (top right), approximated velocity (bottom left), and computed pressure profile (bottom right) for the Brinkman-Darcy coupling imposing Dirichlet boundary conditions for the velocity. . . . . 154

# Introduction

## Motivation

The modelling of phenomena arising from processes classically described by using continuum mechanics, as well as the design of suitable numerical methods to approximate the solution of the corresponding systems of partial differential equations, remains a very active field of research in the scientific community of Numerical Analysis. A notable example is the flow and transport of constituents within incompressible fluids. Many of these phenomena involve linear and nonlinear couplings, for which numerical approximations of the velocity field, pressure, temperature, stress, or pseudostress are usually required. In such scenarios, mixed finite element methods are very suitable since, besides the original unknowns, they yield direct computation of several quantities of physical interest. In particular, the need of obtaining accurate approximations of additional fields has motivated the successful derivation of a wide range of formulations in the framework of Stokes and Navier-Stokes equations, including for instance, stress-velocity formulations (see [34, 53, 62, 75] and the references therein). They feature the clear advantage (with respect to classical velocity-pressure formulations) that these auxiliary fields of interest are computed directly, without resorting to any kind of post-process of the velocity by numerical differentiation, which usually yields an important loss in accuracy. Other physical quantities such as the vorticity and gradient of velocity (or temperature), can also be directly approximated by a mixed finite element approach (see e.g. [11, 43, 65, 73]).

On the other hand, fully mixed finite element methods, mixed–primal approximations, and suitable augmented versions allow us to derive new numerical schemes leading to the numerical solution of a wide range of problems arising in fluid mechanics (see e.g. [11, 12, 35, 43, 44, 72, 73, 75]). In particular, one advantage of augmented methods is given by the fact that, besides to greatly simplifying the analysis, they lead to numerical approximations of the unknowns of interest using standard finite element spaces.

This thesis is motivated by the design and analysis of numerical schemes based on finite elements to solve coupled problems that have a relevant physical interest in fluid mechanics and engineering applications including: wastewater treatment, distillation processes, pulp-and-paper industries, cooling systems for electronic devices, oil extraction processes, and contamination of the groundwater, among others. More precisely, we are interested in studying transport problems through viscous flows in porous media. Many of these phenomena can be governed by nonlinear systems of partial differential equations consisting in the coupling of the Stokes, Brinkman, Navier-Stokes or Darcy equations with a convection–diffusion equation. Solving these problems numerically, both accurately and efficiently, is usually difficult due to the combination of strong nonlinearities with the coupling between the field equations.

## Goals of the thesis

According to the above discussion, this thesis contributes to the development of new mixed finite element methods to simulate several problems in fluid mechanics, specially oriented to the transport of particles of low concentration through viscous flows in porous media. In turn, throughout the mathematical and numerical analysis in this work, we aim to establish a general theoretical framework to guarantee the solvability and stability of a wide class of coupled problems.

More precisely, we are interested in:

- Providing suitable variational formulations to establish existence and uniqueness of the continuous solution for the set of governing equations, using fixed-point strategies and classical results for variational problems.
- Deriving the corresponding Galerkin scheme and employing appropiate finite element spaces, in order to respect the mathematical and physical structure of the underlying problem.
- Analyzing the solvability of the Galerkin scheme and establishing the corresponding convergence results.
- Deriving *a posteriori* error estimators to establish adaptive methods allowing to improve the accuracy of the numerical approximations, mainly under presence of singularities or high gradients of the solution.
- Validating the theoretical results through extensive testing and illustrative numerical simulations, covering both academic and application-oriented examples.

## Model problems

First let us stress that modelling aspects are not part of the main objectives of this work, and therefore the governing continuum-based models will be assumed as given, and no further effort will be made to derive the particular set of equations.

We firstly are interested in studying the following system of partial differential equations

$$\begin{aligned} \boldsymbol{\sigma} &= \mu(\phi)\,\nabla \boldsymbol{u} - p\,\mathbb{I}\,, \quad -\mathbf{div}\,\boldsymbol{\sigma} = \boldsymbol{f}\phi\,, \quad \mathbf{div}\,\boldsymbol{u} = 0\,, \\ \widetilde{\boldsymbol{\sigma}} &= \vartheta(|\nabla\phi|)\,\nabla\phi - \phi\boldsymbol{u} - \gamma(\phi)\boldsymbol{k}\,, \quad -\operatorname{div}\widetilde{\boldsymbol{\sigma}} = g\,, \end{aligned} \tag{1}$$

which describes the stationary state of the transport of species $\phi$ (*cf.* second row in (1)) in an immiscible fluid whose dynamic is governed by the Stokes equations (*cf.* first row of (1)) in a given domain $\Omega$. In this problem the sought quantities are: the Cauchy fluid stress $\boldsymbol{\sigma}$, the velocity of fluid $\boldsymbol{u}$, the pressure $p$, and the local concentration of species $\phi$. Here the viscosity $\mu$, the diffusion coefficient $\vartheta$, and the flux function $\gamma$, depend nonlinearly on $\phi$. In turn, $\boldsymbol{f}$ and $g$ are given functions. In addition, suitable mixed boundary conditions on $\boldsymbol{u}$, $\boldsymbol{\sigma}$, $\widetilde{\boldsymbol{\sigma}}$, and $\phi$ complement the system (1). Depending on the nature of $\phi$ (which may represent the concentration of some chemical component, or temperature) this problem can be relevant to a number of practical engineering applications including; natural and thermal convection, aluminum production, chemical distillation processes, formation of fog, granular flows, motion of biomenbranes, among others.

Next, we address an extention of the problem given in (1), described by the following system of equations

$$\begin{aligned} \boldsymbol{\sigma} &= \mu(\phi)\,\nabla \boldsymbol{u} - p\,\mathbb{I}, \quad \boldsymbol{K}^{-1}\boldsymbol{u} - \mathbf{div}\,\boldsymbol{\sigma} = \boldsymbol{f}\phi\,, \quad \mathbf{div}\,\boldsymbol{u} = 0\,, \\ \widetilde{\boldsymbol{\sigma}} &= \vartheta(\phi)\,\nabla\phi - \phi\boldsymbol{u} - f_{\rm bk}(\phi)\boldsymbol{k}\,, \quad \beta\,\phi - \operatorname{div}\widetilde{\boldsymbol{\sigma}} = g\,. \end{aligned} \tag{2}$$

where, in contrast with (1), the Brinkman term $\boldsymbol{K}^{-1}\boldsymbol{u}$ is incorporated into the equations modeling the fluid. In turn, the diffusion coefficient $\vartheta$ depends on $\phi$ instead of $|\nabla\phi|$, and the mass term $\beta\,\phi$ is added to the second equation of the second row in (1). In this case, the model described by (2) has a relevant physical interest, since it corresponds to a specific configuration in the modelling of sedimentation-consolidation processes (see [31, 32, 96]). The numerical approximation of the macroscopic description in sedimentation processes is needed in a wide variety of physical phenomena and industrial processes including wasterwater treatment, mineral processing, clot formation within the blood, fluidized beds, and many others.

It is important to remark that the problems given by (1) and (2) are strongly nonlinearly coupled, representing a further difficulty. On the other hand, the solvability of the sedimentation-consolidation problem has been previously discussed in [29], for the case of large fluid viscosity, using the technique of parabolic regularization. Nevertheless, as we will see more precisely in Section 3.2.1 (see eq. (3.3)), we still remain in the framework of non-degeneracy of the diffusion coefficient. On the other hand, it is worth mentioning that models of sedimentation-consolidation share some structural similarities with Boussinesq- and Oldroyd- type models, for which several mixed formulations have been proposed (see [44, 45, 53, 54, 55, 90]). However, up to our knowledge, mixed formulations specifically tailored for the study of sedimentation processes are not yet available from the literature.

Finally, we focus on the mathematical and numerical analysis for the following Brinkman-Darcy coupled problem in $\mathrm{R}^3$

$$\left.\begin{aligned} \alpha\,\boldsymbol{u}_{\rm B} + \nu\,\mathbf{curl}\,\boldsymbol{\omega}_{\rm B} + \nabla p_{\rm B} &= \boldsymbol{f}_{\rm B} \\ \boldsymbol{\omega}_{\rm B} - \mathbf{curl}\boldsymbol{u}_{\rm B} &= \mathbf{0} \\ \operatorname{div}\boldsymbol{u}_{\rm B} &= 0 \end{aligned}\right\} \quad \text{en} \quad \Omega_{\rm B}\,, \tag{3}$$

and

$$\left.\begin{aligned} \mu\boldsymbol{u}_{\rm D} + \nabla p_{\rm D} &= \boldsymbol{f}_{\rm D} \\ \operatorname{div}\boldsymbol{u}_{\rm D} &= 0 \end{aligned}\right\} \quad \text{en} \quad \Omega_{\rm D}\,, \tag{4}$$

which is formulated in terms of the velocity $\boldsymbol{u}_{\rm B}$, the vorticity $\boldsymbol{\omega}_{\rm B}$, and the pressure $p_{\rm B}$, for the Brinkman model. In this case, the above model describes the flow patterns of a viscous fluid within a highly permeable medium $\Omega_{\rm B}$, given by Brinkman equations (*cf.* (3)), and its interaction with non-viscous flow within classical porous media $\Omega_{\rm D}$, governed by Darcy's equations (*cf.* (4)). In this model, $\nu > 0$ represents the viscosity of the fluid, $\mu > 0$ depends on this viscosity and on the permeability of the porous media, which is assumed to be homogeneous, and $\alpha > 0$ is a parameter related to the relaxation time. Here $\boldsymbol{f}_{\rm B}$ and $\boldsymbol{f}_{\rm D}$ are given functions. In addition, for the coupled problem (3)-(4) suitable transmission conditions (on the interface that split the domains $\Omega_{\rm B}$ and $\Omega_{\rm D}$) for the velocities $\boldsymbol{u}_{\rm B}$, $\boldsymbol{u}_{\rm D}$, and the pressures $p_{\rm B}$ and $p_{\rm D}$, are introduced. Finally, suitable boundary conditions on $\boldsymbol{u}_{\rm B}$, $\boldsymbol{u}_{\rm D}$, and $\boldsymbol{\omega}_{\rm B}$ complement the model problem (3)-(4). Such a model is often encountered in the modeling petroleum reservoir, perfusion of physiological fluids into soft tissues, and other phenomena related with filtration processes. Up to our knowledge, the coupling of Brinkman and Darcy flows has been only addressed in terms of the primal unknowns of velocity and pressure (see [25, 52, 85]). Vorticity-based formulations for the Stokes-Darcy coupling were introduced in [19] and later studied in [18, 51].

In particular, the contributions [19] and [18] include a slightly different formulation for a model where the fluid boundary coincides with the interface between both domains. On the other hand, in [18], higher regularity of the fluid pressure is required for the analysis, whereas at discrete level a family of nonconforming discretizations, consisting in Nédélec elements for vorticity, piecewise constant elements for velocity, and Crouzeix-Raviart elements for the pressure, is employed.

## Outline of the thesis

The present work is organized as follows. In **Chapter 1** we begin by analyzing the nonlinear coupling given in (1). An augmented variational approach for the fluid flow coupled with a primal formulation for the transport model is proposed. The resulting Galerkin scheme yields an augmented mixed–primal finite element method employing Raviart-Thomas spaces of order $k$ for the Cauchy stress, and continuous piecewise polynomials of degree $\leq k+1$ for the velocity and also for the scalar field. Here the classical Schauder and Brouwer fixed point theorems are utilized to establish existence of solution of the continuous and discrete formulations, respectively. In turn, suitable estimates arising from the connection between a regularity assumption and the Sobolev embedding and Rellich-Kondrachov compactness theorems, are also employed in the continuous analysis. Then, sufficiently small data allow us to prove uniqueness and to derive optimal a priori error estimates. The contents of this chapter originally appeared in the following paper:

> [4] M. Álvarez, G.N. Gatica and R. Ruiz-Baier, *An augmented mixed–primal finite element method for a coupled flow–transport problem.* **ESAIM: Mathematical Modelling and Numerical Analysis**. Vol. 49, 3, pp. 1399–1427, (2015).

In **Chapter 2** we develop an *a posteriori* error analysis for the augmented mixed–primal finite element method of the coupled fow–transport problem studied in [4]. More precisely, we derive two efficient and reliable residual-based *a posteriori* error estimators for that scheme: For the first estimator, and under suitable assumptions on the domain, we apply a Helmholtz decomposition and exploit local approximation properties of the Clément interpolant and Raviart-Thomas operator to show its reliability. On the other hand, its efficiency follows from inverse inequalities and the localization arguments based on triangle-bubble and edge-bubble functions. Next, an alternative error estimator is proposed, whose reliability can be proved without resorting to Helmholtz decompositions. The contents of this chapter are part of the following paper:

> [7] M. Álvarez, G.N. Gatica and R. Ruiz-Baier, *A posteriori error analysis for a viscous flow–transport problem.* **ESAIM: Mathematical Modelling and Numerical Analysis**. Vol. 50, 6, 1789–1816, (2016).

The **Chapter 3** is devoted to the mathematical and numerical analysis of the strongly coupled flow and transport system given by (2). The model focuses on the steady-state regime of a solid-liquid suspension immersed in a viscous fluid within a permeable medium, and the governing equations consist in the Brinkman problem with variable viscosity, written in terms of Cauchy pseudostresses and bulk velocity of the mixture; coupled with a nonlinear advection–diffusion equation describing the transport of the solids volume fraction. Here, we extend the approach and the fixed point strategy introduced

in [4] to the present context. The contents of this chapter gave rise to the following paper:

> [5] M. Álvarez, G.N. Gatica and R. Ruiz-Baier, *A mixed–primal finite element approximation of a sedimentation-consolidation system.* **M3AS: Mathematical Models and Methods in Applied Sciences**. vol. 26, 5, pp. 867–900, (2016).

In the **Chapter 4** we develop the *a posteriori* error analysis of an augmented mixed–primal finite element method for the 2D and 3D versions of the sedimentation-consolidation system studied in [5]. In this chapter, we derive two efficient and reliable residual-based *a posteriori* error estimators for the corresponding Galerkin scheme. The derivation of the 2D version of these estimators follows very closely the analysis developed in [7]. For the first estimator we make use of suitable ellipticity and inf-sup conditions together with a Helmholtz decomposition and the local approximation properties of the Clément interpolant and Raviart-Thomas operator to show its reliability. Then, its efficiency is derived from inverse inequalities and localization arguments based on triangle-bubble and edge-bubble functions. Afterwards, we deduce a second reliable and efficient *a posteriori* error estimator, where the Helmholtz decomposition is not employed in the corresponding proof of reliability. Next, we employ the recent results from [64] to extend the analysis above to the 3D version of these two estimators. The contents of this chapter have inspired the following paper, submitted for publication:

> [6] M. Álvarez, G.N. Gatica and R. Ruiz-Baier, *A posteriori error analysis for a sedimentation–consolidation system.* Preprint 2016-26, Centro de Investigación en Ingeniería Matemática (CI$^2$MA). Universidad de Concepción, Chile, (2016).

Finally, in **Chapter 5**, we propose and analyze a vorticity-based fully-mixed finite element method to numerically approximate the flow patterns governed by the Brinkman-Darcy equations (3)-(4). Here for sake of the analysis, the tangential component of the vorticity is supposed to vanish on the whole boundary of the fluid, whereas null normal components of both velocities are assumed on the respective boundaries, except on the interface where suitable transmission conditions are considered. In this way, the derivation of the corresponding mixed variational formulation leads to incorporate a Lagrange multiplier enforcing the pressure continuity across the interface, whereas mass balance results from essential boundary conditions on each domain. As a consequence, a typical saddle-point operator equation is obtained, and hence the classical Babuška-Brezzi theory is applied to show the well-posedness of the continuous and discrete schemes. In particular, we remark that the continuous and discrete inf-sup conditions of the main bilinear form are proved by using suitably chosen injective operators to get lower bounds of the corresponding suprema, which constitutes a previously known technique, introduced in [69, 70, 71], and recently denominated $T$-coercivity (see [22, 39]). In turn, and consistently with the above, the stability of the Galerkin scheme requires that the **curl** of the finite element subspace approximating the vorticity be contained in the space where the discrete velocity of the fluid lives, which yields Raviart-Thomas and Nédélec finite element subspaces as feasible choices. Then we show that the aforementioned constraint can be avoided by augmenting the mixed formulation with a residual-type term arising from the Brinkman momentum equation. The contents of this chapter originally appeared in the following paper:

[8] M. Álvarez, G.N. Gatica and R. Ruiz-Baier, *Analysis of a vorticity-based fully-mixed formulation for the 3D Brinkman-Darcy problem.* **Computer Methods in Applied Mechanics and Engineering**. vol. 307, pp. 68–95, (2016).

Throughout the five chapters of this thesis, our theoretical results such as orders of convergence and the reliability and efficiency of the corresponding *a posteriori* error estimators, are illustrated *via* several numerical tests, highlighting also the good performance of the discrete scheme proposed and the associated adaptive algorithms. The set of numerical experiments includes examples in 2D and 3D, where we provided even examples not fully covered by the theory. The computational implementations were obtained employing the freely available finite element libraries; FreeFem++, Lifev, and FEniCS, the open source mesh generator Gmsh, and the illustrator Paraview.

## Preliminary Notations

Let us denote by $\Omega \subseteq \mathrm{R}^n$, $n = 2, 3$ a given bounded domain with polyhedral boundary $\Gamma = \bar{\Gamma}_{\mathrm{D}} \cup \bar{\Gamma}_{\mathrm{N}}$, with $\Gamma_{\mathrm{D}} \cap \Gamma_{\mathrm{N}} = \emptyset$ and $|\Gamma_{\mathrm{D}}|, |\Gamma_{\mathrm{N}}| > 0$, and denote by $\boldsymbol{\nu}$ the outward unit normal vector on $\Gamma$. Standard notation will be adopted for Lebesgue spaces $\mathrm{L}^p(\Omega)$ and Sobolev spaces $\mathrm{H}^s(\Omega)$ with norm $\|\cdot\|_{s,\Omega}$ and seminorm $|\cdot|_{s,\Omega}$. In particular, $\mathrm{H}^{1/2}(\Gamma)$ stands for the space of traces of functions of $\mathrm{H}^1(\Omega)$ and $\mathrm{H}^{-1/2}(\Gamma)$ denotes its dual. By $\mathbf{M}$ and $\mathbb{M}$ we will denote the corresponding vectorial and tensorial counterparts of the generic scalar functional space M, and $\|\cdot\|$, with no subscripts, will stand for the natural norm of either an element or an operator in any product functional space. In turn, for any vector field $\boldsymbol{v} = (v_i)_{i=1,n}$ we set the gradient and divergence operators as

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

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

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

Furthermore, we recall that

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

equipped with the usual norm

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

is a standard Hilbert space in the realm of mixed problems. Finally, in what follows $\mathbb{I}$ stands for the identity tensor in $\mathbb{R} := \mathrm{R}^{n \times n}$, and $|\cdot|$ denotes the Euclidean norm in $\mathbf{R} := \mathrm{R}^n$.

# Introducción

## Motivación

El modelamiento de fenómenos provenientes de procesos clásicos descritos a través de la mecánica de medios continuos, así como el diseño de métodos numéricos apropiados para aproximar la solución de los sistemas de ecuaciones diferenciales parciales correspondientes, sigue siendo un campo de investigación que mantiene muy activa a la comunidad científica del área de Análisis Numérico. Un ejemplo notable corresponde al flujo y transporte de constituyentes dentro de fluidos incompresibles. Muchos de estos fenómenos involucran acoplamientos lineales y no lineales, para los cuales por lo general se requieren aproximaciones numéricas del campo de velocidad, presión, temperatura, esfuerzos, o pseudo-esfuerzos. En este tipo de escenarios, los métodos de elementos finitos mixtos resultan muy apropiados ya que, además de las incógnitas originales, ellos proporcionan el cálculo directo de varias otras cantidades de interés físico. En particular, la necesidad de obtener aproximaciones precisas de campos adicionales ha motivado la derivación con éxito de una amplia gama de formulaciones en el marco de las ecuaciones de Stokes y Navier-Stokes, incluyendo por ejemplo formulaciones para el esfuerzo y la velocidad (ver [34, 53, 62, 75]). Estas se caracterizan por la ventaja (con respecto a las formulaciones clásicas para la velocidad y la presión) de que los campos auxiliares de interés se calculan directamente, sin recurrir a ningún tipo de post-proceso del campo de velocidades por diferenciación numérica, que por lo general produce una pérdida importante en la precisión. Otras cantidades físicas tales como la vorticidad y el gradiente de la velocidad (o de la temperatura), también se pueden aproximar de manera directa a través de un enfoque de elementos finitos mixtos (ver por ejemplo [11, 43, 65, 73]).

Por otra parte, métodos de elementos finitos completamente mixtos, aproximaciones de tipo mixto-primal, y versiones aumentadas adecuadas, permiten derivar nuevos esquemas numéricos que conducen a la solución numérica de una gran cantidad de problemas que surgen en mecánica de fluidos (ver por ejemplo [11, 12, 35, 43, 44, 72, 73, 75]). En particular, una de las ventajas de los métodos aumentados corresponde al hecho de que, además de simplificar enormemente el análisis, ellos conducen a aproximaciones numéricas de las incógnitas de intéres utilizando espacios de elementos finitos habituales.

Esta tesis está motivada por el diseño y el análisis de esquemas numéricos basados en un enfoque de elementos finitos para resolver problemas acoplados que tienen un gran interés físico en mecánica de fluidos y en aplicaciones de ingeniería, las cuales incluyen: tratamiento de aguas residuales, procesos de destilación, industrias de pulpa y papel, sistemas de enfriamiento para dispositivos electrónicos, procesos de extracción de petróleo, y contaminación de aguas subterráneas, entre otras. Más precisamente, interesa estudiar problemas de transporte a través de flujos viscosos en medios porosos. Muchos de estos fenómenos pueden regirse por sistemas no lineales de ecuaciones diferenciales parciales los cuales consisten en el acoplamiento de las ecuaciones de Stokes, Brinkman, Navier-Stokes, o Darcy, con una ecuación de convección–difusión. Resolver numericamente estos problemas, de manera precisa y efi-

ciente, es usualmente díficil debido a la combinación de fuertes no linealidades con los acoplamientos de las ecuaciones de campo.

## Objetivos de la tesis

De acuerdo a la discusión anterior, esta tesis contribuye con el desarrollo de nuevos métodos de elementos finitos mixtos para simular varios problemas en mecánica de fluidos, especialmente orientados al transporte de particulas de baja concentración a través de un flujo viscoso en medios porosos. A su vez, durante el desarrollo matemático y numérico de esta tesis, se busca establecer un marco teórico general para garantizar la solubilidad y la estabilidad de una clase amplia de problemas acoplados.

Más precisamente, interesa:

- Proporcionar formulaciones variacionales adecuadas para establecer existencia y unicidad de la solución continua del conjunto de ecuaciones gobernantes, utilizando estrategias de punto fijo y resultados clásicos para problemas variacionales.
- Derivar el esquema de Galerkin correspondiente y emplear espacios de elementos finitos apropiados con la finalidad de respetar la estructura física y matemática del problema subyacente.
- Analizar la solubilidad del esquema de Galerkin y establecer los resultados de convergencia correspondientes.
- Derivar estimadores de error *a posteriori* para establecer métodos adaptativos que permitan mejorar la precisión de las aproximaciones numéricas, principalmente bajo la presencia de singularidades o altos gradientes de la solución.
- Validar los resultados teóricos a través de varios ensayos y simulaciones numéricas ilustrativas, que incluyan ejemplos académicos y orientados a aplicaciones.

## Problemas Modelo

En primer lugar se enfatiza en que los aspectos de modelamiento no son parte de los objetivos principales de este trabajo, con lo cual los modelos que rigen (los cuales están basados en el continuo) se asumirán como dados y no se hará esfuerzo adicional para derivar el conjunto particular de ecuaciones.

En primer instancia interesa estudiar el siguiente sistema de ecuaciones diferenciales parciales

$$\begin{aligned} \boldsymbol{\sigma} &= \mu(\phi)\,\nabla \boldsymbol{u} - p\,\mathbb{I}\,, \quad -\mathbf{div}\,\boldsymbol{\sigma} = \boldsymbol{f}\phi\,, \quad \mathbf{div}\,\boldsymbol{u} = 0\,, \\ \widetilde{\boldsymbol{\sigma}} &= \vartheta(|\nabla\phi|)\,\nabla\phi - \phi\boldsymbol{u} - \gamma(\phi)\boldsymbol{k}\,, \quad -\operatorname{div}\widetilde{\boldsymbol{\sigma}} = g\,, \end{aligned} \tag{5}$$

el cual describe el estado estacionario del transporte de especies $\phi$ (segunda fila en (5)) en un fluido inmiscible cuya dinámica se rige por las ecuaciones de Stokes (primera fila en (5)) en una región dada del espacio $\Omega$. En este problema las incógnitas son: el tensor de esfuerzos de Cauchy $\boldsymbol{\sigma}$, la velocidad del fluido $\boldsymbol{u}$, la presion $p$, y la concentración local de especies $\phi$. Aquí la viscocidad $\mu$, el coeficiente de difusión $\vartheta$, y la función de flujo $\gamma$, dependen no linealmente de $\phi$. A su vez, $\boldsymbol{f}$ y $g$ son funciones dadas. Adicionalmente, al sistema (5) se incorporan condiciones de contorno mixtas sobre $\boldsymbol{u}$, $\boldsymbol{\sigma}$, $\widetilde{\boldsymbol{\sigma}}$,

y $\phi$, las cuales permiten cerrar el sistema de manera apropiada. Dependiendo de la naturaleza de $\phi$ (el cual puede representar la concentración de algún componente químico, ó la temperatura) este problema puede ser relevante en varias aplicaciones prácticas en ingeniería que incluyen; convección térmica y natural, producción de aluminio, procesos de destilación química, formación de niebla, flujos granulares, y movimiento de membranas, entre otros.

Luego, se aborda una extensión del problema dado en (5), descrita por el siguiente sistema de ecuaciones

$$\begin{aligned} \boldsymbol{\sigma} &= \mu(\phi)\,\nabla \boldsymbol{u} - p\,\mathbb{I}\,, \quad \boldsymbol{K}^{-1}\boldsymbol{u} - \mathbf{div}\,\boldsymbol{\sigma} = \boldsymbol{f}\phi\,, \quad \mathbf{div}\,\boldsymbol{u} = 0\,, \\ \widetilde{\boldsymbol{\sigma}} &= \vartheta(\phi)\,\nabla\phi - \phi\boldsymbol{u} - f_{\mathrm{bk}}(\phi)\boldsymbol{k}\,, \quad \beta\,\phi - \mathrm{div}\,\widetilde{\boldsymbol{\sigma}} = g\,. \end{aligned} \tag{6}$$

el cual, a diferencia de (5), incorpora el término de Brinkman $\boldsymbol{K}^{-1}\boldsymbol{u}$ en las ecuaciones que modelan el fluido. A su vez, el coeficiente de difusión $\vartheta$ ahora depende de $\phi$ en lugar de $|\nabla\phi|$, y se agrega el término de masa $\beta\,\phi$ a la segunda ecuación de la segunda fila en (5). En este caso, el modelo descrito por (6) es de gran interés físico, debido a que corresponde a una configuración específica en el modelamiento de procesos de sedimentación-consolidación (ver [31, 32, 96]). La aproximación numérica de la descripción macroscópica en los procesos de sedimentación se necesita en una amplia variedad de fenómenos físicos y procesos industriales que incluyen el tratamiento de aguas residuales, procesamiento de minerales, la formación de coágulos en la sangre, lechos fluidizados, y muchos otros.

Es importante notar que los problemas dados por (5) y (6) involucran acoplamientos fuertemente no lineales, donde la no linealidad representa una dificultad adicional. Por otra parte, la solubilidad del problema de sedimentación-consolidación se ha discutido previamente en [29], para el caso de un fluido altamente viscoso, utilizando la técnica de regularización parabólica. Sin embargo, como se verá precisamente en la Sección 3.2.1 (ver ecuación (3.3)), aquí se permanecerá en el marco no degenerativo del coeficiente de difusión. Por otra parte, cabe mencionar que los modelos de sedimentación-consolidación comparten algunas similitudes estructurales con los modelos del tipo Oldroyd y Boussinesq, para los cuales varias formulaciones mixtas han sido propuestas (ver [44, 45, 53, 54, 55, 90]). No obstante, hasta donde se sabe, formulaciones mixtas específicamente diseñadas para el estudio de procesos de sedimentación no se encuentran todavía disponibles en la literatura.

Finalmente, la atención se centra en el análisis matemático y numérico para el siguente problema acoplado Brinkman-Darcy en $\mathrm{R}^3$

$$\left.\begin{aligned} \alpha\,\boldsymbol{u}_{\mathrm{B}} + \nu\,\mathbf{curl}\,\boldsymbol{\omega}_{\mathrm{B}} + \nabla p_{\mathrm{B}} &= \boldsymbol{f}_{\mathrm{B}} \\ \boldsymbol{\omega}_{\mathrm{B}} - \mathbf{curl}\boldsymbol{u}_{\mathrm{B}} &= \mathbf{0} \\ \mathrm{div}\,\boldsymbol{u}_{\mathrm{B}} &= 0 \end{aligned}\right\} \quad \text{en} \quad \Omega_{\mathrm{B}}\,, \tag{7}$$

y

$$\left.\begin{aligned} \mu\boldsymbol{u}_{\mathrm{D}} + \nabla p_{\mathrm{D}} &= \boldsymbol{f}_{\mathrm{D}} \\ \mathrm{div}\,\boldsymbol{u}_{\mathrm{D}} &= 0 \end{aligned}\right\} \quad \text{en} \quad \Omega_{\mathrm{D}}\,, \tag{8}$$

el cual se formula para el modelo de Brinkman, en términos de la velocidad $\boldsymbol{u}_{\mathrm{B}}$, la vorticidad $\boldsymbol{\omega}_{\mathrm{B}}$, y la presión $p_{\mathrm{B}}$. En este caso, el modelo anterior describe los patrones de flujo de un fluido viscoso dentro de un medio altamente permeable $\Omega_{\mathrm{B}}$, modelado por las ecuaciones de Brinkman (*cf.* (7)), y su interacción con un flujo no viscoso dentro de un medio poroso $\Omega_{\mathrm{D}}$, gobernado por las ecuaciones de Darcy (*cf.* (8)). En este modelo, $\nu > 0$ representa la viscocidad del fluido, $\mu > 0$ depende de la viscocidad y la permeabilidad del medio poroso, el cual se asume homogéneo, y $\alpha > 0$ es un parámetro relacionado con el tiempo de relajación. Aquí $\boldsymbol{f}_{\mathrm{B}}$ y $\boldsymbol{f}_{\mathrm{D}}$ son funciones dadas. Adicionalmente al problema acoplado

descrito por (7)-(8), se imponen condiciones de transmisión adecuadas (en la interfaz que separa los dominios $\Omega_{\mathrm{B}}$ y $\Omega_{\mathrm{D}}$) para las velocidades $\boldsymbol{u}_{\mathrm{B}}$, $\boldsymbol{u}_{\mathrm{D}}$, y las presiones $p_{\mathrm{B}}$ y $p_{\mathrm{D}}$. Finalmente el sistema (7)-(8) se cierra incorporando condiciones de frontera apropiadas sobre $\boldsymbol{u}_{\mathrm{B}}$, $\boldsymbol{u}_{\mathrm{D}}$, y $\boldsymbol{\omega}_{\mathrm{B}}$. Modelos como el descrito anteriormente, se encuentran con frecuencia en el modelamiento de yacimientos petrolíferos, perfusión de fluidos fisiológicos en tejidos blandos, y otros fenómenos relacionados con procesos de filtración. Hasta donde se sabe, el acoplamiento de flujos Brinkman y Darcy solamente ha sido abordado en términos de las incógnitas primales de la velocidad y la presión (ver [25, 52, 85]). Formulaciones basadas en vorticidad para el acoplamiento Stokes-Darcy fueron introducidas en [19] y luego estudiadas en [18, 51]. En particular, las contribuciones [19] y [18] incluyen una formulación ligeramente diferente para un modelo en el cual la frontera del fluido coincide con la interfaz entre ambos dominios. Por otra parte, el análisis en [18] requiere mayor regularidad de la presión del fluido, miestras que a nivel discreto se utiliza una familia de discretizaciones no conformes que consiste en elementos de Nédélec para la vorticidad, elementos constantes a trozos para la velocidad, y elementos Crouzeix-Raviart para la presión.

## Organización de la tesis

El presente trabajo está organizado como sigue. En el **Capítulo 1** se estudia el acoplamiento no lineal dado por (5). Se propone un enfoque variacional aumentado para el flujo del fluido acoplado con una formulación primal para el modelo de transporte. El esquema de Galerkin resultante conduce a un método de elementos finitos mixto-primal aumentado para el cual se emplean espacios de Raviart-Thomas de orden $k$ para el tensor de Cauchy, y polinomios continuos a trozos de grado $\leq k+1$ para la velocidad del fluido y también para el campo escalar. Aquí los teoremas clásicos de punto de fijo de Schauder y Brouwer son utilizados para establecer la existencia de solución de las formulaciones continuas y discretas, respectivamente. A su vez, estimaciones adecuadas que surgen a partir de la conexión entre un supuesto de regularidad y los teoremas de inclusión de Sobolev y de compacidad de Rellich-Kondrachov, también son empleados en el análisis continuo. Luego, datos suficientemente pequeños permiten probar unicidad y derivar estimaciones de error *a priori* óptimas. Este capítulo está constituido por la siguiente publicación:

> [4] M. Álvarez, G.N. Gatica and R. Ruiz-Baier, *An augmented mixed–primal finite element method for a coupled flow–transport problem.* **ESAIM: Mathematical Modelling and Numerical Analysis**. vol. 49, 3, pp. 1399–1427, (2015).

En el **Capítulo 2** se desarrolla un análisis de error *a posteriori* para el método de elementos finitos mixto-primal del problema acoplado de flujo con transporte estudiado en [4]. En este trabajo se derivan dos estimadores de error *a posteriori* residuales para ese esquema, los cuales resultan ser confiables y eficientes. Para el primer estimador, bajo supuestos adecuados sobre el dominio, se aplica una descomposición de Helmholtz y se explotan las propiedades de aproximación locales del interpolador de Clément y del operador de Raviart-Thomas para mostrar su confiabilidad. Por otra parte, su eficiencia se sigue a partir de desigualdades inversas y argumentos de localización basados en funciones burbujas sobre triángulos y lados. Luego, se propone un estimador alternativo cuya confiabilidad se logra probar sin recurrir a descomposiciones de Helmholtz. Este capítulo está constituido por el siguiente artículo:

[7] M. Álvarez, G.N. Gatica and R. Ruiz-Baier, *A posteriori error analysis for a viscous flow–transport problem.* **ESAIM: Mathematical Modelling and Numerical Analysis**. Vol. 50, 6, 1789–1816, (2016).

El **Capítulo 3** está dedicado al análisis matemático y numérico del sistema de flujo y transporte fuertemente acoplado dado por (2). El modelo se enfoca en el regimen estacionario de una suspension sólido-liquido inmersa en un fluido viscoso dentro de un medio permeable, y las ecuaciones gobernantes corresponden al problema de Brinkman con viscosidad variable, escrito en términos del pseudo esfuerzo de Cauchy y la velocidad de la mezcla; acoplado con una ecuación de advección-difusión no lineal que describe el transporte de la fracción de volumen de sólidos. Aquí se extiende el enfoque y la estrategía de punto fijo que se introdujó en [4] al presente contexto. Este capítulo dió origen a la siguiente publicación:

[5] M. Álvarez, G.N. Gatica and R. Ruiz-Baier, *A mixed–primal finite element approximation of a sedimentation-consolidation system.* **M3AS: Mathematical Models and Methods in Applied Sciences**. vol. 26, 5, pp. 867–900, (2016).

En el **Capítulo 4** se desarrolla el análisis de error *a posteriori* del método de elementos finitos mixto-primal aumentado para las versiones 2D y 3D del sistema de sedimentación-consolidación estudiado en [5]. En este capítulo se derivan dos estimadores de error *a posteriori* residuales para el esquema de Galerkin correspondiente, los cuales son confiables y eficientes. La derivación de la version 2D de estos dos estimadores sigue muy de cerca el análisis desarrollado en [7]. Para el primer estimador se hace uso de elipticidades y condiciones inf-sup adecuadas, junto con una descomposición de Helmholtz y las propiedades de aproximación locales del operador de interpolación de Clément y el operador de Raviart-Thomas, para mostrar su confiabilidad. Luego, su eficiencia se deriva a partir de desigualdades inversas y argumentos de localización basados en funciones burbujas sobre triángulos y lados. Posteriormente, se deduce un segundo estimador de error *a posteriori*, donde la descomposición de Helmholtz no se emplea en la prueba de confiabilidad correspondiente. Seguidamente, se emplean los resultados recientes en [64] para extender el análisis anterior a la versión 3D de estos dos estimadores. Este capítulo ha inspirado el siguiente artículo, el cual se encuentra sometido a publicación:

[6] M. Álvarez, G.N. Gatica and R. Ruiz-Baier, *A posteriori error analysis for a sedimentation-consolidation system.* Preprint 2016-26, Centro de Investigación en Ingeniería Matemática (CI$^2$MA). Universidad de Concepción, Chile, (2016).

Finalmente, en el **Capítulo 5** se propone y se analiza un método de elementos finitos completamente mixto basado en vorticidad para aproximar numéricamente los patrones de flujo gobernados por las ecuaciones Brinkman-Darcy (3)-(4). Aquí, para efectos del análisis, se asume que la componente tangencial de la vorticidad se anula sobre toda la frontera del dominio Brinkman, mientras que las componentes normales de ambas velocidades se suponen nulas sobre las fronteras respectivas, excepto sobre la interfaz donde se consideran condiciones de transmisión adecuadas. De esta manera, la derivación de la formulación variacional mixta correspondiente conduce a la incorporación de un multiplicador de Lagrange, el cual impone la continuidad de la presión a través de la interfaz, mientras que los balances de masas resultan a partir de las condiciones de fronteras esenciales sobre cada dominio. En consecuencia, se obtiene una ecuación típica de operadores de punto silla, y por lo tanto

la teoría clásica de Babuška-Brezzi se aplica para mostrar que el esquema continuo y discreto están bien puestos. En particular, las condiciones inf-sup continuas y discretas de la forma bilinear principal se demuestran usando adecuadamente operadores inyectivos para obtener las cotas inferiores de los supremos correspondientes, lo cual constituye una técnica previamente conocida, e introducida en [69, 70, 71], y recientemente denominada $T-$coercividad (ver [22, 39]). A su vez, y consistentemente con lo anterior, la estabilidad del esquema de Galerkin requiere que el **curl** del subespacio de elementos finitos que aproxima a la vorticidad este contenido en el espacio donde vive la velocidad discreta del fluido, lo cual conduce a los subespacios de elementos finitos de Raviart-Thomas y Nédélec como elecciones factibles. Luego, se prueba que las restricciones mencionadas anteriormente se pueden evitar aumentando la formulación mixta con un término de tipo residual que surge a partir de la ecuación de momentum de Brinkman. Este capítulo está constituido por la siguiente publicación:

> [8] M. Álvarez, G.N. Gatica and R. Ruiz-Baier, *Analysis of a vorticity-based fully-mixed formulation for the 3D Brinkman-Darcy problem.* **Computer Methods in Applied Mechanics and Engineering**. vol. 307, pp. 68–95, (2016).

A lo largo de los cinco capítulos que conforman esta tesis, los resultados teóricos como órdenes de convergencia y la confiabilidad y eficiencia de los estimadores de error *a posteriori* correspondientes, son ilustrados a través de varios tests numéricos, que destacan también el buen desempeño de los esquemas discretos propuestos y los algoritmos adaptativos asociados. El conjunto de experimentos numéricos incluye ejemplos en 2D y 3D, donde incluso algunos de ellos no están cubiertos por la teoría. A su vez, las implementaciones computacionales se obtuvieron empleando las librerías de elementos finitos de acceso libre; FreeFem++, Lifev, y FEniCS, el generador de mallas de código abierto Gmsh, y el ilustrador Paraview.

# CHAPTER 1

## An augmented mixed–primal finite element method for a coupled flow–transport problem

### 1.1 Introduction

We are interested in studying mixed finite element approximations to simulate the transport of a species density in an immiscible fluid. Depending on the nature of the species, this problem can be relevant to a number of practical engineering applications including natural and thermal convection, aluminum production, chemical distillation processes, formation of fog, impedance tomography, motion of bio-membranes, semiconductors, granular flows, and so on. In this chapter we pay particular attention to the steady state regime in the phenomenon of sedimentation-consolidation of particles (see e.g. [31, 32, 96]), where the sought physical quantities in the model include the velocity of the flow and the local solids concentration. On the other hand, it is well known that other variables, such as the principal components of the fluid or solids stress tensors, are of great interest in this context (see e.g. [33, Chapter 3]). In a more general sense, the need of obtaining accurate approximations of additional fields has motivated the successful derivation of a wide range of formulations in the framework of Stokes and Navier-Stokes equations, including for instance, stress-velocity formulations (see [34, 53, 62, 75] and the references therein). They feature the clear advantage (with respect to classical velocity-pressure formulations) that these auxiliary quantities of interest are computed directly, without resorting to any kind of post-process of the velocity field by numerical differentiation, which usually yields an important loss in accuracy. The attached difficulties are that the finite dimensional spaces involved in the resulting discrete formulation must be properly selected in order to satisfy the corresponding inf-sup condition [26], and that approximation of stresses may become quite expensive if adequate finite elements are not used.

Now, concerning the problem we are interested in here, we realize that, in order to be able to analyze the solvability of a mixed formulation for the fluid flow coupled with a primal method for the transport model, we require $\mathrm{H}^1(\Omega)$ smoothness for the components of both the fluid velocity and its discrete approximation. However, since the usual mixed approach is only able to guarantee that they live in $\mathrm{L}^2(\Omega)$, in this chapter we follow [57] (see also [56], [62]) and propose an augmented mixed scheme in which the stress stays in its original space $\mathbb{H}(\mathbf{div};\Omega)$, but the velocity components lie now in the smaller space $\mathrm{H}^1(\Omega)$. In other words, the original problem is reformulated as an augmented variational approach for the fluid flow coupled with a primal formulation for the transport model, which constitutes one of the key ideas of this work. According to the above, we will approximate each row of the fluid

Cauchy stress tensor with Raviart-Thomas elements of order $k$, whereas the velocity and scalar field (which will represent a solids concentration, or temperature, depending on the specific application) will be approximated with continuous piecewise polynomials of degree $\leq k+1$. The existence of solution of the continuous and associated Galerkin schemes is established by a combination of a suitable regularity assumption, fixed point arguments, the well-know Lax-Milgram theorem, and a classical result on bijective monotone operators. In addition, the Sobolev embedding and Rellich-Kondrachov compactness theorems are also essential in the continuous analysis. Furthermore, the assumption of sufficiently small data allows us to conclude uniqueness of solution and to derive optimal *a priori* error estimates. The extension of the above described general approach to the more realistic case of steady sedimentation-consolidation systems is under development in [4]. In addition, the incorporation of a similar augmented formulation, and the consequent application of basically the same fixed point theorems employed here, will appear in the forthcoming work [44] where a mixed-primal formulation for the stationary Boussinesq problem is introduced and analyzed.

### Outline

We have organized the contents of this chapter as follows. The remainder of this section introduces some standard notations and functional spaces. In Section 1.2 we first describe the boundary value problem of interest and then slightly simplify it by eliminating the pressure unknown in the fluid. Next, in Section 1.3, we introduce and analyze the continuous formulation, which is defined by an augmented mixed approach for the fluid flow coupled with a primal method for the transport equation. The necessity of augmentation is clearly justified, and the solvability analysis is based on a fixed point strategy that makes use of the Lax-Milgram and Schauder theorems together with a well-known result on monotone operators. We prove existence of solution and for sufficiently small data we derive uniqueness. The associated Galerkin scheme is introduced in Section 1.4 by employing Raviart-Thomas elements for the stress, and continuous piecewise polynomial approximations for the velocity and concentration. Here the solvability is established by applying now the Brouwer fixed point theorem and analogous arguments to those employed in Section 1.3. In Section 1.5 we assume again sufficiently small data and, applying a suitable Strang-type estimate for nonlinear problems, provide optimal *a priori* error estimates. Finally, in Section 1.6 we present numerical examples illustrating the good performance of the mixed-primal method and confirming the theoretical rates of convergence.

## 1.2 The model problem

The following system of partial differential equations describes the stationary state of the transport of species $\phi$ in an immiscible fluid occupying the domain $\Omega$:

$$\begin{aligned} \boldsymbol{\sigma} &= \mu(\phi)\,\nabla\boldsymbol{u} - p\,\mathbb{I}\,, \quad -\mathbf{div}\,\boldsymbol{\sigma} = \boldsymbol{f}\phi\,, \quad \operatorname{div}\boldsymbol{u} = 0\,, \\ \widetilde{\boldsymbol{\sigma}} &= \vartheta(|\nabla\phi|)\,\nabla\phi - \phi\boldsymbol{u} - \gamma(\phi)\boldsymbol{k}\,, \quad -\operatorname{div}\widetilde{\boldsymbol{\sigma}} = g\,, \end{aligned} \tag{1.1}$$

where the sought quantities are the Cauchy fluid stress $\boldsymbol{\sigma}$, the local volume-average velocity of the fluid $\boldsymbol{u}$, the pressure $p$, and the local concentration of species $\phi$. For sake of clarity in the presentation, we will restrict ourselves to a specific physical scenario corresponding to the process of sedimentation-consolidation of a mixture (see e.g. [32]). There, it is assumed that a suspension of solid particles within a viscous fluid, undergoes settling due to gravity. The hypotheses of slow sedimentation velocity,

short relaxation time, constant density, and negligible expansion viscosities, allow to derive (1.1) from the classical principles of mass and momentum conservation in mixture theory. In this model, the kinematic effective viscosity, $\mu$; the diffusion coefficient, $\vartheta$; and the one-dimensional flux function describing hindered settling, $\gamma$; depend nonlinearly on $\phi$, and $\boldsymbol{k}$ is a vector pointing in the direction of gravity. In addition, $\vartheta$ is assumed of class $C^1$ and we suppose that there exist positive constants $\mu_1$, $\mu_2$, $\gamma_1$, $\gamma_2$, $\vartheta_1$, and $\vartheta_2$, such that

$$\mu_1 \leq \mu(s) \leq \mu_2 \quad \text{and} \quad \gamma_1 \leq \gamma(s) \leq \gamma_2 \qquad \forall\, s \in \mathrm{R}\,, \tag{1.2}$$

$$\vartheta_1 \leq \vartheta(s) \leq \vartheta_2 \quad \text{and} \quad \vartheta_1 \leq \vartheta(s) + s\,\vartheta'(s) \leq \vartheta_2 \qquad \forall\, s \geq 0\,. \tag{1.3}$$

Note that (1.2) and the first assumption in (1.3) guarantee, in particular, that the corresponding Nemytsky operators, say $U$ for $\mu$, defined by $U(\phi)(x) := \mu(\phi(x))\ \forall\, \phi \in \mathrm{L}^2(\Omega)$, $\forall\, x \in \Omega$ a.e., and analogously for $\vartheta$, $\gamma$, $\mu^{-1}$, $\vartheta^{-1}$, and $\gamma^{-1}$, are all well defined and continuous from $\mathrm{L}^2(\Omega)$ into $\mathrm{L}^2(\Omega)$. Furthermore, it is easy to show (see, e.g. [76, Theorem 3.8]) that the assumptions in (1.3) imply Lipschitz-continuity and strong monotonicity of the nonlinear operator induced by $\vartheta$. We will go back to this fact later on in Section 1.3.

Some examples of concentration-dependent coefficients typically found in the literature are (see [31, 33, 96, 97])

$$\mu(\phi) = \mu_\infty \left(1 - \frac{\phi}{\phi_m}\right)^{-n_\mu}, \quad \gamma(\phi) = g_\infty \phi (1-\phi)^{n_\gamma}, \quad \vartheta(\phi) = \vartheta_\infty \left(1 - \frac{\phi}{\phi_m}\right)^{-n_\vartheta},$$

where $\mu_\infty, \phi_m, \gamma_\infty, n_\mu, n_\gamma, n_\vartheta, \vartheta_\infty$ are positive model parameters. These functions violate assumptions (1.2)-(1.3) required in the subsequent analysis. We will therefore consider regularized concentration-dependent coefficients. In turn, some examples of nonlinear functions $\vartheta$ that indeed satisfy (1.3) are the following:

$$\vartheta(s) := 2 + \frac{1}{1+s} \quad \text{and} \quad \vartheta(s) := \alpha_0 + \alpha_1 (1+s^2)^{(\beta-2)/2}\,,$$

where $\alpha_0$, $\alpha_1 > 0$ and $\beta \in (0, 2)$. The first example is basically academic but the second one corresponds to a particular case of the well-known Carreau law in fluid mechanics. We also stress that the structure of (1.1) may also serve as prototype model for generalized Boussinesq models and natural convection equations describing the interaction of a fluid driven by gravity and thermal changes. In such a context, $\phi$ can be viewed as the adimensional temperature of the fluid, and typical examples for the variable coefficients (now interpreted as temperature-dependent viscosity and thermal diffusivity) are

$$\mu(\phi) = \mu_\infty \exp(-\phi), \quad \vartheta(\phi) = \vartheta_\infty \exp(\phi), \quad \gamma = 0,$$

(see e.g. [45, 55, 90]).

The driving force of the mixture also depends on the local fluctuations of the concentration, so the right hand side of the second equation in (1.1) is linear with respect to $\phi$, and $\boldsymbol{f} \in \mathbf{L}^\infty(\Omega)$ and $g \in \mathrm{L}^2(\Omega)$ are given functions. Finally, given $\boldsymbol{u}_{\mathrm{D}} \in \mathbf{H}^{1/2}(\Gamma_{\mathrm{D}})$, the following mixed boundary conditions complement (1.1):

$$\boldsymbol{u} = \boldsymbol{u}_{\mathrm{D}} \quad \text{on} \quad \Gamma_{\mathrm{D}}\,, \quad \boldsymbol{\sigma}\boldsymbol{\nu} = \mathbf{0} \quad \text{on} \quad \Gamma_{\mathrm{N}}\,, \quad \phi = 0 \quad \text{on} \quad \Gamma_{\mathrm{D}}\,, \quad \text{and} \quad \widetilde{\boldsymbol{\sigma}} \cdot \boldsymbol{\nu} = 0 \quad \text{on} \quad \Gamma_{\mathrm{N}}\,. \tag{1.4}$$

On the other hand, it is easy to see that the first and third equations in (1.1) are equivalent to

$$\boldsymbol{\sigma} = \mu(\phi)\, \nabla \boldsymbol{u} - p\, \mathbb{I} \quad \text{and} \quad p + \frac{1}{n} \mathrm{tr}(\boldsymbol{\sigma}) = 0 \quad \text{in} \quad \Omega\,,$$

which permits us to eliminate the pressure $p$ from the first equation. Consequently, we arrive at the following coupled system:

$$\begin{gathered}\frac{1}{\mu(\phi)}\,\boldsymbol{\sigma}^{\mathrm{d}} = \nabla\boldsymbol{u} \quad \text{in} \quad \Omega\,, \quad -\mathbf{div}\,\boldsymbol{\sigma} = \boldsymbol{f}\phi \quad \text{in} \quad \Omega\,,\\ \widetilde{\boldsymbol{\sigma}} = \vartheta(|\nabla\phi|)\,\nabla\phi - \phi\boldsymbol{u} - \gamma(\phi)\boldsymbol{k} \quad \text{in} \quad \Omega\,, \quad -\mathrm{div}\widetilde{\boldsymbol{\sigma}} = g \quad \text{in} \quad \Omega\,,\\ \boldsymbol{u} = \boldsymbol{u}_{\mathrm{D}} \quad \text{on} \quad \Gamma_{\mathrm{D}}\,, \quad \boldsymbol{\sigma}\boldsymbol{\nu} = \mathbf{0} \quad \text{on} \quad \Gamma_{\mathrm{N}}\,,\\ \phi = 0 \quad \text{on} \quad \Gamma_{\mathrm{D}}\,, \quad \text{and} \quad \widetilde{\boldsymbol{\sigma}}\cdot\boldsymbol{\nu} = 0 \quad \text{on} \quad \Gamma_{\mathrm{N}}\,.\end{gathered} \tag{1.5}$$

We remark here that the incompressibility constraint is implicitly present in the first equation of (1.5), that is in the constitutive equation relating $\boldsymbol{\sigma}$ and $\boldsymbol{u}$.

## 1.3 The continuous formulation

### 1.3.1 The augmented mixed–primal formulation

We first observe that the homogeneous Neumann boundary condition for $\boldsymbol{\sigma}$ on $\Gamma_{\mathrm{N}}$ (*cf.* second relation of (1.4)) suggests to introduce the space

$$\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega) := \Big\{\boldsymbol{\tau}\in\mathbb{H}(\mathbf{div};\Omega): \quad \boldsymbol{\tau}\boldsymbol{\nu}=\mathbf{0} \quad \text{on} \quad \Gamma_{\mathrm{N}}\Big\}\,.$$

Hence, multiplying the first equation of (1.5) by $\boldsymbol{\tau}\in\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)$, integrating by parts, and using the Dirichlet boundary condition for $\boldsymbol{u}$ (*cf.* third row of (1.5)), we obtain

$$\int_\Omega \frac{1}{\mu(\phi)}\,\boldsymbol{\sigma}^{\mathrm{d}}:\boldsymbol{\tau}^{\mathrm{d}} + \int_\Omega \boldsymbol{u}\cdot\mathbf{div}\boldsymbol{\tau} = \langle\boldsymbol{\tau}\boldsymbol{\nu},\boldsymbol{u}_{\mathrm{D}}\rangle_{\Gamma_{\mathrm{D}}} \qquad \forall\,\boldsymbol{\tau}\in\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)\,,$$

where $\langle\cdot,\cdot\rangle_{\Gamma_{\mathrm{D}}}$ is the duality pairing between $\mathbf{H}^{-1/2}(\Gamma_{\mathrm{D}})$ and $\mathbf{H}^{1/2}(\Gamma_{\mathrm{D}})$. In addition, the equilibrium equation is then rewritten as

$$\int_\Omega \boldsymbol{v}\cdot\mathbf{div}\boldsymbol{\sigma} = -\int_\Omega \boldsymbol{f}\phi\cdot\boldsymbol{v} \qquad \forall\,\boldsymbol{v}\in\mathbf{L}^2(\Omega)\,.$$

On the other hand, the Dirichlet boundary condition for $\phi$ (*cf.* fourth row of (1.5)) motivates the introduction of the space

$$\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) := \Big\{\psi\in\mathrm{H}^1(\Omega): \quad \psi=0 \quad \text{on} \quad \Gamma_{\mathrm{D}}\Big\}\,,$$

for which, thanks to the generalized Poincaré inequality, there exists $c_p>0$, depending only on $\Omega$ and $\Gamma_{\mathrm{D}}$, such that

$$\|\psi\|_{1,\Omega} \le c_p\,|\psi|_{1,\Omega} \qquad \forall\,\psi\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,. \tag{1.6}$$

Therefore , given $\phi\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$, we arrive at the following mixed formulation for the flow: Find $(\boldsymbol{\sigma},\boldsymbol{u})\in\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)\times\mathbf{L}^2(\Omega)$ such that

$$\begin{aligned} a_\phi(\boldsymbol{\sigma},\boldsymbol{\tau}) + b(\boldsymbol{\tau},\boldsymbol{u}) &= \langle\boldsymbol{\tau}\boldsymbol{\nu},\boldsymbol{u}_{\mathrm{D}}\rangle_{\Gamma_{\mathrm{D}}} \qquad \forall\,\boldsymbol{\tau}\in\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega),\\ b(\boldsymbol{\sigma},\boldsymbol{v}) &= -\int_\Omega \boldsymbol{f}\phi\cdot\boldsymbol{v} \qquad \forall\,\boldsymbol{v}\in\mathbf{L}^2(\Omega),\end{aligned} \tag{1.7}$$

where $a_\phi : \mathbb{H}_{\rm N}(\mathbf{div};\Omega)\times\mathbb{H}_{\rm N}(\mathbf{div};\Omega)\to\mathbb{R}$ and $b : \mathbb{H}_{\rm N}(\mathbf{div};\Omega)\times\mathbf{L}^2(\Omega)\to\mathbb{R}$ are bounded bilinear forms defined as

$$a_\phi(\boldsymbol{\zeta},\boldsymbol{\tau}) := \int_\Omega \frac{1}{\mu(\phi)}\,\boldsymbol{\zeta}^{\rm d} : \boldsymbol{\tau}^{\rm d}, \qquad b(\boldsymbol{\tau},\boldsymbol{v}) := \int_\Omega \boldsymbol{v}\cdot\mathbf{div}\boldsymbol{\tau},$$

for $\boldsymbol{\zeta},\boldsymbol{\tau}\in\mathbb{H}_{\rm N}(\mathbf{div};\Omega)$ and $\boldsymbol{v}\in\mathbf{L}^2(\Omega)$.

In turn, given $\boldsymbol{u}\in\mathbf{L}^2(\Omega)$, and using the homogeneous Neumann boundary condition for $\widetilde{\boldsymbol{\sigma}}$ (*cf.* fourth row of (1.5)), we deduce that the primal formulation for the concentration equation becomes: Find $\phi\in\mathrm{H}^1_{\Gamma_{\rm D}}(\Omega)$ such that

$$A_{\boldsymbol{u}}(\phi,\psi) \,=\, \int_\Omega \gamma(\phi)\,\boldsymbol{k}\cdot\nabla\psi \,+\, \int_\Omega g\psi \qquad \forall\,\psi\in\mathrm{H}^1_{\Gamma_{\rm D}}(\Omega)\,, \tag{1.8}$$

where

$$A_{\boldsymbol{u}}(\phi,\psi) \,:=\, \int_\Omega \vartheta(|\nabla\phi|)\,\nabla\phi\cdot\nabla\psi \,-\, \int_\Omega \phi\,\boldsymbol{u}\cdot\nabla\psi \qquad \forall\,\phi,\,\psi\in\mathrm{H}^1_{\Gamma_{\rm D}}(\Omega)\,. \tag{1.9}$$

At this point we observe that the assumption on $\mu$ given in (1.2) and the well known Babuška-Brezzi theory suffice to show that (1.7) is well-posed (see, e.g. [72, Theorem 2.1] for details). However, in order to deal with the analysis of (1.8), and particularly to estimate the second term defining $A_{\boldsymbol{u}}$, we would require $\boldsymbol{u}\in\mathbf{H}^1(\Omega)$. In fact, we now recall from [90, eq. (2.20)] that Hölder's inequality and standard Sobolev embeddings estimates (*cf.* [1, Theorem 4.12], [92, Theorem 1.3.4]) yield the existence of a positive constant $c(\Omega)$, depending only on $\Omega$, such that

$$\left|\int_\Omega \varphi\,\boldsymbol{v}\cdot\nabla\psi\right| \,\leq\, c(\Omega)\,\|\varphi\|_{1,\Omega}\,\|\boldsymbol{v}\|_{1,\Omega}\,|\psi|_{1,\Omega} \qquad \forall\,\varphi,\,\psi\in\mathrm{H}^1(\Omega)\,, \quad \forall\,\boldsymbol{v}\in\mathbf{H}^1(\Omega)\,. \tag{1.10}$$

Furthermore, while the exact solution of (1.7) actually satisfies $\nabla\boldsymbol{u} = \dfrac{1}{\mu(\phi)}\,\boldsymbol{\sigma}^{\rm d}$ in $\mathcal{D}'(\Omega)$, which implies that $\boldsymbol{u}$ does belong to $\mathbf{H}^1(\Omega)$, the foregoing distributional identity does not necessarily extend to the discrete setting of (1.7), and hence the aforementioned difficulty would appear again when trying to analyze the Galerkin scheme associated to (1.8). Therefore, in order to circumvent this inconvenience, we proceed similarly as in [57, Section 3] and incorporate into (1.7) the following redundant Galerkin terms

$$\begin{aligned}
\kappa_1\int_\Omega\left(\nabla\boldsymbol{u}-\frac{1}{\mu(\phi)}\,\boldsymbol{\sigma}^{\rm d}\right):\nabla\boldsymbol{v} \;&=\; 0 && \forall\,\boldsymbol{v}\in\mathbf{H}^1(\Omega)\,,\\
\kappa_2\int_\Omega \mathbf{div}\boldsymbol{\sigma}\cdot\mathbf{div}\boldsymbol{\tau} \;&=\; -\kappa_2\int_\Omega \boldsymbol{f}\phi\cdot\mathbf{div}\boldsymbol{\tau} && \forall\,\boldsymbol{\tau}\in\mathbb{H}_{\rm N}(\mathbf{div};\Omega)\,,\\
\kappa_3\int_{\Gamma_{\rm D}}\boldsymbol{u}\cdot\boldsymbol{v} \;&=\; \kappa_3\int_{\Gamma_{\rm D}}\boldsymbol{u}_{\rm D}\cdot\boldsymbol{v} && \forall\,\boldsymbol{v}\in\mathbf{H}^1(\Omega)\,,
\end{aligned} \tag{1.11}$$

where $(\kappa_1,\kappa_2,\kappa_3)$ is a vector of positive parameters to be specified later. Notice that the first and third equations in (1.11) implicitly require the velocity $\boldsymbol{u}$ to live in $\mathbf{H}^1(\Omega)$. In this way, instead of (1.7), we consider from now on the following augmented mixed formulation: Find $(\boldsymbol{\sigma},\boldsymbol{u})\in\mathbb{H}_{\rm N}(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega)$ such that

$$B_\phi((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau},\boldsymbol{v})) \,=\, F_\phi(\boldsymbol{\tau},\boldsymbol{v}) \quad \forall\,(\boldsymbol{\tau},\boldsymbol{v})\in\mathbb{H}_{\rm N}(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega)\,, \tag{1.12}$$

where

$$\begin{aligned}
B_\phi((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau},\boldsymbol{v})) \,:=\, a_\phi(\boldsymbol{\sigma},\boldsymbol{\tau}) \,+\, b(\boldsymbol{\tau},\boldsymbol{u}) \,-\, b(\boldsymbol{\sigma},\boldsymbol{v}) \,+\, \kappa_1\int_\Omega\left(\nabla\boldsymbol{u}\,-\,\frac{1}{\mu(\phi)}\,\boldsymbol{\sigma}^{\rm d}\right):\nabla\boldsymbol{v}\\
+\quad \kappa_2\int_\Omega \mathbf{div}\boldsymbol{\sigma}\cdot\mathbf{div}\boldsymbol{\tau} \,+\, \kappa_3\int_{\Gamma_{\rm D}}\boldsymbol{u}\cdot\boldsymbol{v}\,,
\end{aligned} \tag{1.13}$$

and

$$F_\phi(\boldsymbol{\tau},\boldsymbol{v}) := \langle \boldsymbol{\tau}\boldsymbol{\nu}, \boldsymbol{u}_{\mathrm{D}}\rangle_{\Gamma_{\mathrm{D}}} + \int_\Omega \boldsymbol{f}\phi\cdot\boldsymbol{v} - \kappa_2\int_\Omega \boldsymbol{f}\phi\cdot\mathbf{div}\boldsymbol{\tau} + \kappa_3\int_{\Gamma_{\mathrm{D}}}\boldsymbol{u}_{\mathrm{D}}\cdot\boldsymbol{v}\,. \tag{1.14}$$

We remark in advance that the well-posedness of (1.12) is proved below in Section 1.3.3. In particular, we emphasize that the positiveness of the parameter $\kappa_2$ in (1.13) is crucial for the ellipticity of the bilinear form $B_\phi$ in the product space $\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega)$ (see below (1.23) - (1.24) in the proof of Lemma 1.4), which enables to choose arbitrary finite element subspaces to define the associated discrete formulation. Otherwise, one would need inf-sup conditions for the bilinear form $b$, which, involving the spaces $\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)$ and $\mathbf{H}^1(\Omega)$, and suitable subspaces of them, do not seem to be easily verifiable. In turn, the term multiplying $\kappa_2$ in (1.13) just adds a minor complexity to the Galerkin scheme since the corresponding matrix becomes block-diagonal. On the other hand, since the unique solution of (1.7) is obviously a solution of (1.12) as well, we will conclude that both continuous problems share the same unique solution.

In this way, the augmented mixed-primal formulation of our original coupled problem (1.5) reduces to (1.12) and (1.8), that is: Find $(\boldsymbol{\sigma},\boldsymbol{u},\phi)\in\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega)\times\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ such that

$$\begin{aligned} B_\phi((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau},\boldsymbol{v})) &= F_\phi(\boldsymbol{\tau},\boldsymbol{v}) \quad \forall(\boldsymbol{\tau},\boldsymbol{v})\in\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega)\,,\\ A_{\boldsymbol{u}}(\phi,\psi) &= \int_\Omega \gamma(\phi)\,\boldsymbol{k}\cdot\nabla\psi + \int_\Omega g\psi \qquad \forall\,\psi\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,. \end{aligned} \tag{1.15}$$

### 1.3.2 A fixed point strategy

Having proposed the alternative formulation (1.12), whose continuous and discrete solutions have second components living in $\mathbf{H}^1(\Omega)$, we are able now to take a second look at (1.8). More precisely, given $\phi\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ and the corresponding solution $(\boldsymbol{\sigma},\boldsymbol{u})\in\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega)$ of (1.12), we can set, instead of (1.8), the modified primal formulation: Find $\widetilde{\phi}\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ such that

$$A_{\boldsymbol{u}}(\widetilde{\phi},\widetilde{\psi}) = G_\phi(\widetilde{\psi}) \qquad \forall\,\widetilde{\psi}\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,, \tag{1.16}$$

where

$$G_\phi(\widetilde{\psi}) := \int_\Omega \gamma(\phi)\,\boldsymbol{k}\cdot\nabla\widetilde{\psi} + \int_\Omega g\,\widetilde{\psi} \qquad \forall\,\widetilde{\psi}\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,. \tag{1.17}$$

The well-posedness of this nonlinear problem will also be addressed in Section 1.3.3. Alternatively, one could also deal, instead of (1.16), with the linear problem: Find $\widetilde{\phi}\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ such that

$$A_{\phi,\boldsymbol{u}}(\widetilde{\phi},\widetilde{\psi}) = G_\phi(\widetilde{\psi}) \qquad \forall\,\widetilde{\psi}\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,,$$

where

$$A_{\phi,\boldsymbol{u}}(\widetilde{\phi},\widetilde{\psi}) := \int_\Omega \vartheta(|\nabla\phi|)\,\nabla\widetilde{\phi}\cdot\nabla\widetilde{\psi} - \int_\Omega \widetilde{\phi}\,\boldsymbol{u}\cdot\nabla\widetilde{\psi} \qquad \forall\,\widetilde{\phi},\,\widetilde{\psi}\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,.$$

Nevertheless, and for easiness of the analysis, throughout the rest of the chapter we stay with (1.16).

Hence, the description of problems (1.12) and (1.16) suggests a fixed point strategy to analyze (1.15). Indeed, let $\mathbf{S}:\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\longrightarrow\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega)$ be the operator defined by

$$\mathbf{S}(\phi) = (\mathbf{S}_1(\phi),\mathbf{S}_2(\phi)) := (\boldsymbol{\sigma},\boldsymbol{u}) \in \mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega) \qquad \forall\,\phi\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,,$$

where $(\boldsymbol{\sigma},\boldsymbol{u})$ is the unique solution of (1.12) with the given $\phi$. In turn, let $\widetilde{\mathbf{S}} : \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) \times \mathbf{H}^1(\Omega) \longrightarrow \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ be the operator defined by

$$\widetilde{\mathbf{S}}(\phi,\boldsymbol{u}) := \widetilde{\phi} \qquad \forall\,(\phi,\boldsymbol{u}) \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) \times \mathbf{H}^1(\Omega)\,,$$

where $\widetilde{\phi}$ is the unique solution of (1.16) with the given $(\phi,\boldsymbol{u})$. Then, we define the operator $\mathbf{T} : \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) \longrightarrow \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ by

$$\mathbf{T}(\phi) := \widetilde{\mathbf{S}}(\phi,\mathbf{S}_2(\phi)) \qquad \forall\,\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,,$$

and realize that solving (1.15) is equivalent to seeking a fixed point of $\mathbf{T}$, that is: Find $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ such that

$$\mathbf{T}(\phi) = \phi\,. \tag{1.18}$$

### 1.3.3 Well-posedness of the uncoupled problems

In this section we show that the uncoupled problems (1.12) and (1.16) are in fact well-posed. We begin by recalling (see, e.g. [26]) that $\mathbb{H}(\mathbf{div};\Omega) = \mathbb{H}_0(\mathbf{div};\Omega) \oplus \mathbb{R}\,\mathbb{I}$, where

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

More precisely, for each $\boldsymbol{\zeta} \in \mathbb{H}(\mathbf{div};\Omega)$ there exist unique $\boldsymbol{\zeta}_0 := \boldsymbol{\zeta} - \left\{ \dfrac{1}{n|\Omega|} \displaystyle\int_\Omega \mathrm{tr}(\boldsymbol{\zeta}) \right\} \mathbb{I} \in \mathbb{H}_0(\mathbf{div};\Omega)$ and $d := \dfrac{1}{n|\Omega|} \displaystyle\int_\Omega \mathrm{tr}(\boldsymbol{\zeta}) \in \mathbb{R}$, such that $\boldsymbol{\zeta} = \boldsymbol{\zeta}_0 + d\mathbb{I}$. The following three lemmas from [26, 57, 62], which concern the above decomposition and an equivalence of norm, will be employed to show the well-posedness of (1.12) for a given $\phi$.

**Lemma 1.1.** *There exists $c_1 = c_1(\Omega) > 0$ such that*

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

*with $\boldsymbol{\tau}_0 \in \mathbb{H}_0(\mathbf{div};\Omega)$ and $c \in \mathbb{R}$.*

*Proof.* See [26, Proposition 3.1]. □

**Lemma 1.2.** *There exists $c_2 = c_2(\Omega,\Gamma_{\mathrm{N}}) > 0$ such that*

$$c_2 \,\|\boldsymbol{\tau}\|^2_{\mathbf{div};\Omega} \leq \|\boldsymbol{\tau}_0\|^2_{\mathbf{div};\Omega} \qquad \forall\,\boldsymbol{\tau} = \boldsymbol{\tau}_0 + c\mathbb{I} \in \mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega),$$

*with $\boldsymbol{\tau}_0 \in \mathbb{H}_0(\mathbf{div};\Omega)$ and $c \in \mathbb{R}$.*

*Proof.* See [62, Lemma 2.2]. □

**Lemma 1.3.** *There exists $c_3 = c_3(\Omega,\Gamma_{\mathrm{D}}) > 0$ such that*

$$|\boldsymbol{v}|^2_{1,\Omega} + \|\boldsymbol{v}\|^2_{0,\Gamma_{\mathrm{D}}} \geq c_3 \,\|\boldsymbol{v}\|^2_{1,\Omega} \qquad \forall\,\boldsymbol{v} \in \mathbf{H}^1(\Omega)\,.$$

*Proof.* It corresponds to a slight modification of the proof of [57, Lemma 3.3]. □

Furthermore, for sake of the subsequent analysis we will also require Lipschitz continuity assumptions for $\gamma$ and $\mu$. More precisely, we assume that there exist positive constants $L_\gamma$ and $L_\mu$ such that

$$|\gamma(s)-\gamma(t)| \leq L_\gamma\,|s-t| \qquad \forall\, s,\, t \in \mathrm{R}\,, \tag{1.19}$$

and

$$|\mu(s)-\mu(t)| \leq L_\mu\,|s-t| \qquad \forall\, s,\, t \in \mathrm{R}\,. \tag{1.20}$$

We now begin the solvability analysis of the uncoupled problems with the following result.

**Lemma 1.4.** *Assume that* $\kappa_1 \in \left(0, \frac{2\delta\mu_1}{\mu_2}\right)$ *with* $\delta \in (0, 2\mu_1)$, *and that* $0 < \kappa_2,\, \kappa_3$. *Then, for each* $\phi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)$ *the problem* (1.12) *has a unique solution* $\mathbf{S}(\phi) := (\boldsymbol{\sigma}, \boldsymbol{u}) \in H := \mathbb{H}_\mathrm{N}(\mathbf{div};\Omega) \times \mathbf{H}^1(\Omega)$. *Moreover, there exists* $C_S > 0$, *independent of* $\phi$, *such that*

$$\|\mathbf{S}(\phi)\|_H \,=\, \|(\boldsymbol{\sigma},\boldsymbol{u})\|_H \,\leq\, C_S \left\{ \|\boldsymbol{u}_\mathrm{D}\|_{1/2,\Gamma_\mathrm{D}} \,+\, \|\boldsymbol{f}\|_{\infty,\Omega}\, \|\phi\|_{0,\Omega} \right\} \qquad \forall\, \phi \,\in\, \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)\,. \tag{1.21}$$

*Proof.* We first observe from (1.13) that, given $\phi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)$, $B_\phi$ is clearly a bilinear form. Next, applying the Cauchy-Schwarz inequality, the lower bound for $\mu$ (*cf.* (1.2)), and the trace theorem (with constant $c_0$), we also obtain from (1.13) that

$$\begin{aligned} |B_\phi((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau},\boldsymbol{v}))| \,&\leq\, \frac{1}{\mu_1}\,\|\boldsymbol{\sigma}^\mathtt{d}\|_{0,\Omega}\,\|\boldsymbol{\tau}^\mathtt{d}\|_{0,\Omega} \,+\, \|\boldsymbol{u}\|_{0,\Omega}\,\|\mathbf{div}\boldsymbol{\tau}\|_{0,\Omega} \,+\, \|\boldsymbol{v}\|_{0,\Omega}\,\|\mathbf{div}\boldsymbol{\sigma}\|_{0,\Omega} \\ +\quad \kappa_1\,|\boldsymbol{u}|_{1,\Omega}\,|\boldsymbol{v}|_{1,\Omega} \,&+\, \frac{\kappa_1}{\mu_1}\,\|\boldsymbol{\sigma}^\mathtt{d}\|_{0,\Omega}\,|\boldsymbol{v}|_{1,\Omega} \,+\, \kappa_2\,\|\mathbf{div}\boldsymbol{\sigma}\|_{0,\Omega}\,\|\mathbf{div}\boldsymbol{\tau}\|_{0,\Omega} \,+\, c_0^2\,\kappa_3\,\|\boldsymbol{u}\|_{1,\Omega}\,\|\boldsymbol{v}\|_{1,\Omega}\,. \end{aligned}$$

It follows that there exists a positive constant, denoted $\|B\|$ and depending on $\mu_1$, $\kappa_1$, $\kappa_2$, $\kappa_3$, and $c_0$, such that

$$|B_\phi((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau},\boldsymbol{v}))| \,\leq\, \|B\|\,\|(\boldsymbol{\sigma},\boldsymbol{u})\|_H\,\|(\boldsymbol{\tau},\boldsymbol{v})\|_H \qquad \forall\,(\boldsymbol{\sigma},\boldsymbol{u}),\,(\boldsymbol{\tau},\boldsymbol{v}) \,\in\, H\,, \tag{1.22}$$

and hence $B_\phi$ is bounded independently of $\phi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)$.

In turn, we now aim to show that $B_\phi$ is $H$-elliptic. In fact, given $(\boldsymbol{\tau},\boldsymbol{v}) \in H$, we have again from (1.13) that

$$B_\phi((\boldsymbol{\tau},\boldsymbol{v}),(\boldsymbol{\tau},\boldsymbol{v})) = \int_\Omega \frac{1}{\mu(\phi)}\,\boldsymbol{\tau}^\mathtt{d} : \boldsymbol{\tau}^\mathtt{d} + \kappa_1\,|\boldsymbol{v}|^2_{1,\Omega} - \kappa_1 \int_\Omega \frac{1}{\mu(\phi)}\,\boldsymbol{\tau}^\mathtt{d} : \nabla\boldsymbol{v} + \kappa_2\|\mathbf{div}\boldsymbol{\tau}\|^2_{0,\Omega} + \kappa_3\,\|\boldsymbol{v}\|^2_{0,\Gamma_\mathrm{D}}\,,$$

which, using the bounds for $\mu$ (*cf.* (1.2)), the Young inequality, and Lemmas 1.1, 1.2, and 1.3, and taking $\delta$, $\kappa_1$, $\kappa_2$, and $\kappa_3$ as stated in the hypotheses, yields

$$\begin{aligned} B_\phi((\boldsymbol{\tau},\boldsymbol{v}),(\boldsymbol{\tau},\boldsymbol{v})) \,&\geq\, \left(\frac{1}{\mu_2} - \frac{\kappa_1}{2\delta\mu_1}\right)\,\|\boldsymbol{\tau}^\mathtt{d}\|^2_{0,\Omega} \,+\, \kappa_2\|\mathbf{div}\boldsymbol{\tau}\|^2_{0,\Omega} \,+\, \kappa_1\left(1-\frac{\delta}{2\mu_1}\right)\,|\boldsymbol{v}|^2_{1,\Omega} \,+\, \kappa_3\,\|\boldsymbol{v}\|^2_{0,\Gamma_\mathrm{D}} \\ &\geq\, c_1\alpha_1\,\|\boldsymbol{\tau}_0\|^2_{0,\Omega} \,+\, \frac{\kappa_2}{2}\,\|\mathbf{div}\boldsymbol{\tau}\|^2_{0,\Omega} \,+\, \kappa_1\left(1-\frac{\delta}{2\mu_1}\right)\,|\boldsymbol{v}|^2_{1,\Omega} \,+\, \kappa_3\,\|\boldsymbol{v}\|^2_{0,\Gamma_\mathrm{D}} \\ &\geq\, \alpha_2\,\|\boldsymbol{\tau}_0\|^2_{\mathbf{div};\Omega} \,+\, \alpha_3\left\{|\boldsymbol{v}|^2_{1,\Omega} \,+\, \|\boldsymbol{v}\|^2_{0,\Gamma_\mathrm{D}}\right\} \\ &\geq\, c_2\alpha_2\,\|\boldsymbol{\tau}\|^2_{\mathbf{div};\Omega} \,+\, c_3\alpha_3\,\|\boldsymbol{v}\|^2_{1,\Omega}\,, \end{aligned} \tag{1.23}$$

where $\alpha_1 := \min\left\{\left(\frac{1}{\mu_2} - \frac{\kappa_1}{2\delta\mu_1}\right), \frac{\kappa_2}{2}\right\}$, $\alpha_2 := \min\left\{c_1\alpha_1, \frac{\kappa_2}{2}\right\}$, and $\alpha_3 := \min\left\{\kappa_1\left(1-\frac{\delta}{2\mu_1}\right), \kappa_3\right\}$. In this way, defining $\alpha := \min\{c_2\alpha_2, c_3\alpha_3\}$, which depends on $\mu_1$, $\mu_2$, $\delta$, $\kappa_1$, $\kappa_2$, $\kappa_3$, $c_1$, $c_2$, and $c_3$, we conclude that

$$B_\phi((\boldsymbol{\tau},\boldsymbol{v}),(\boldsymbol{\tau},\boldsymbol{v})) \,\geq\, \alpha\,\|(\boldsymbol{\tau},\boldsymbol{v})\|^2_H \qquad \forall\,(\boldsymbol{\tau},\boldsymbol{v}) \,\in\, H\,, \tag{1.24}$$

thus confirming the $H$-ellipticity of $B_\phi$ independently of $\phi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)$ as well. In particular, choosing the feasible values $\delta = \mu_1$ and $\kappa_1 = \frac{\mu_1^2}{\mu_2}$, and then taking $\kappa_2 = 2\left(\frac{1}{\mu_2} - \frac{\kappa_1}{2\delta\mu_1}\right)$ and $\kappa_3 = \kappa_1\left(1 - \frac{\delta}{2\mu_1}\right)$, we find that $\kappa_2 = \frac{1}{\mu_2}$, $\kappa_3 = \frac{\mu_1^2}{2\mu_2}$, and $\alpha = \frac{1}{2\mu_2}\min\{c_1c_2, c_2, c_3\mu_1^2\}$.

Next, given $\phi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)$, we look at the functional $F_\phi$ (*cf.* (1.14)), which is certainly linear. Then, using the Cauchy-Schwarz inequality and the trace estimates in $\mathbb{H}(\mathbf{div};\Omega)$ and $\mathbf{H}^1(\Omega)$, with constants 1 and $c_0$, respectively, we deduce that for each $(\boldsymbol{\tau},\boldsymbol{v}) \in H$ there holds

$$|F_\phi(\boldsymbol{\tau},\boldsymbol{v})| \leq \|\boldsymbol{\tau}\|_{\mathbf{div};\Omega}\,\|\boldsymbol{u}_\mathrm{D}\|_{1/2,\Gamma_\mathrm{D}} + \|\boldsymbol{f}\|_{\infty,\Omega}\,\|\phi\|_{0,\Omega}\left\{\|\boldsymbol{v}\|_{0,\Omega} + \kappa_2\,\|\mathbf{div}\boldsymbol{\tau}\|_{0,\Omega}\right\}$$
$$+ \quad c_0\,\kappa_3\,\|\boldsymbol{u}_\mathrm{D}\|_{1/2,\Gamma_\mathrm{D}}\,\|\boldsymbol{v}\|_{1,\Omega}\,,$$

which provides the existence of a positive constant, denoted $\|F\|$ and depending on $\kappa_2$, $\kappa_3$, and $c_0$, such that

$$|F_\phi(\boldsymbol{\tau},\boldsymbol{v})| \leq \|F\|\left\{\|\boldsymbol{u}_\mathrm{D}\|_{1/2,\Gamma_\mathrm{D}} + \|\boldsymbol{f}\|_{\infty,\Omega}\,\|\phi\|_{0,\Omega}\right\}\|(\boldsymbol{\tau},\boldsymbol{v})\| \qquad \forall\,(\boldsymbol{\tau},\boldsymbol{v}) \in H\,. \tag{1.25}$$

The foregoing inequality shows the boundedness of $F_\phi$ with

$$\|F_\phi\| \leq \|F\|\left\{\|\boldsymbol{u}_\mathrm{D}\|_{1/2,\Gamma_\mathrm{D}} + \|\boldsymbol{f}\|_{\infty,\Omega}\,\|\phi\|_{0,\Omega}\right\}. \tag{1.26}$$

Finally, a straightforward application of the Lax-Milgram Lemma (see, e.g. [63, Theorem 1.1]), proves that, for each $\phi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)$, problem (1.12) has a unique solution $\mathbf{S}(\phi) := (\boldsymbol{\sigma},\boldsymbol{u}) \in H$. Moreover, the corresponding continuous dependence result together with the estimates (1.24) and (1.25) give

$$\|\mathbf{S}(\phi)\|_H = \|(\boldsymbol{\sigma},\boldsymbol{u})\|_H \leq \frac{1}{\alpha}\,\|F_\phi\|_{H'} \leq C_S\left\{\|\boldsymbol{u}_\mathrm{D}\|_{1/2,\Gamma_\mathrm{D}} + \|\boldsymbol{f}\|_{\infty,\Omega}\,\|\phi\|_{0,\Omega}\right\},$$

with $C_S := \frac{\|F\|}{\alpha}$, thus completing the proof. □

Throughout the rest of the chapter we suppose further regularity for the problem defining the operator $\mathbf{S}$. More precisely, we assume that $\boldsymbol{u}_\mathrm{D} \in \mathbf{H}^{1/2+\varepsilon}(\Gamma_\mathrm{D})$ for some $\varepsilon \in (0,1)$ (when $n = 2$) or $\varepsilon \in (\frac{1}{2},1)$ (when $n = 3$), and that for each $\psi \in \mathrm{H}^1_{\Gamma_\mathrm{D}}(\Omega)$ with $\|\psi\|_{1,\Omega} \leq r$, $r > 0$ given, there hold $(\boldsymbol{\zeta},\boldsymbol{w}) := \mathbf{S}(\psi) \in \mathbb{H}_\mathrm{N}(\mathbf{div};\Omega) \cap \mathbb{H}^\varepsilon(\Omega) \times \mathbf{H}^{1+\varepsilon}(\Omega)$ and

$$\|\boldsymbol{\zeta}\|_{\varepsilon,\Omega} + \|\boldsymbol{w}\|_{1+\varepsilon,\Omega} \leq \widetilde{C}_S(r)\left\{\|\boldsymbol{u}_\mathrm{D}\|_{1/2+\varepsilon,\Gamma_\mathrm{D}} + \|\boldsymbol{f}\|_{\infty,\Omega}\,\|\psi\|_{0,\Omega}\right\}, \tag{1.27}$$

with a positive constant $\widetilde{C}_S(r)$ independent of the given $\psi$ but depending on the upper bound $r$ of its $\mathrm{H}^1$-norm. We remark that the reason of the stipulated ranges for $\varepsilon$ will be clarified in the forthcoming analysis (see below proof of Lemma 1.11). In turn, while the actual verification of (1.27) is beyond the goals of the present work, we observe that the fact that (1.7) and (1.12) share the same solution implies that this issue reduces finally to the regularity of the Stokes problem with variable viscosity $\mu$ depending on $\phi$ (see, e.g. [10] for analogous regularity results). To this respect, we would like to mention that the equilibrium equation $-\mathbf{div}\,\boldsymbol{\zeta} = \boldsymbol{f}\,\psi$ in $\Omega$, obviously controls $\mathbf{div}\,\boldsymbol{\zeta}$, whereas the constitutive equation $\frac{1}{\mu(\psi)}\,\boldsymbol{\zeta}^\mathtt{d} = \nabla\boldsymbol{w}$ in $\Omega$, may serve to control the curl of $\boldsymbol{\zeta}$. Certainly, the Lipschitz-continuity of $\mu$ (*cf.* (1.20)) and the upper bound of $\|\psi\|_{1,\Omega}$ are essential here. In addition, the Dirichlet boundary condition on $\boldsymbol{w}$ should be used under the form of its tangential derivative, and the eventual presence of corners in $\Gamma$ should not be a problem.

According to the above, the present assumption is indeed quite reasonable at least for $n = 2$ since just $\varepsilon \in (0,1)$ is required in this case. However, due to the hypothesis $\varepsilon \in (\frac{1}{2},1)$ when $n = 3$, we conjecture that mixed boundary conditions for $\boldsymbol{u}$ and $\boldsymbol{\sigma}$ will have to be excluded of the corresponding 3D problem and that either Dirichlet or Neumann boundary conditions only will be allowed. Finally, and while the estimate (1.27) will be employed only to bound $\|\boldsymbol{\zeta}\|_{\varepsilon,\Omega}$, we have stated it including the term $\|\boldsymbol{w}\|_{1+\varepsilon,\Omega}$ since, because of the constitutive equation, the regularities of $\boldsymbol{\zeta}$ and $\boldsymbol{w}$ will most likely be connected.

We now establish the unique solvability of the nonlinear problem (1.16).

**Lemma 1.5.** *Let $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ and $\boldsymbol{u} \in \mathbf{H}^1(\Omega)$ such that $\|\boldsymbol{u}\|_{1,\Omega} < \frac{\vartheta_1}{c_p\, c(\Omega)}$ (cf. (1.3), (1.6), (1.10)). Then, there exists a unique $\widetilde{\phi} := \widetilde{\mathbf{S}}(\phi,\boldsymbol{u}) \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ solution of (1.16), and there holds*

$$\|\widetilde{\mathbf{S}}(\phi,\boldsymbol{u})\|_{1,\Omega} \,=\, \|\widetilde{\phi}\|_{1,\Omega} \,\leq\, \frac{c_p^2}{\big(\vartheta_1 - c_p c(\Omega)\|\boldsymbol{u}\|_{1,\Omega}\big)} \Big\{ \gamma_2\, |\Omega|^{1/2}\, |\boldsymbol{k}| \,+\, \|g\|_{0,\Omega} \Big\}\,. \tag{1.28}$$

*Proof.* We begin by recalling from [76, Theorem 3.8] that the nonlinear operator induced by the first term defining $A_{\boldsymbol{u}}$ (*cf.* (1.9)) is strongly monotone and Lipschitz-continuous with constants $\vartheta_1$ and $\widetilde{\vartheta}_2 := \max\big\{\vartheta_2, 2\vartheta_2 - \vartheta_1\big\}$ (*cf.* (1.3)), respectively. It follows, using also the Cauchy-Schwarz inequality, (1.10), and (1.6), that for all $\widetilde{\varphi},\, \widetilde{\psi} \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ there holds

$$\begin{aligned}
&A_{\boldsymbol{u}}(\widetilde{\varphi},\widetilde{\varphi}-\widetilde{\psi}) \,-\, A_{\boldsymbol{u}}(\widetilde{\psi},\widetilde{\varphi}-\widetilde{\psi}) \\
&\qquad = \quad \int_\Omega \Big\{ \vartheta(|\nabla\widetilde{\varphi}|)\nabla\widetilde{\varphi} - \vartheta(|\nabla\widetilde{\psi}|)\nabla\widetilde{\psi} \Big\} \cdot \nabla(\widetilde{\varphi}-\widetilde{\psi}) - \int_\Omega (\widetilde{\varphi}-\widetilde{\psi})\boldsymbol{u}\cdot\nabla(\widetilde{\varphi}-\widetilde{\psi}) \\
&\qquad \geq \quad \vartheta_1\, |\widetilde{\varphi}-\widetilde{\psi}|^2_{1,\Omega} \,-\, c(\Omega)\, \|\widetilde{\varphi}-\widetilde{\psi}\|_{1,\Omega}\, \|\boldsymbol{u}\|_{1,\Omega}\, |\widetilde{\varphi}-\widetilde{\psi}|_{1,\Omega} \\
&\qquad \geq \quad \Big\{ \vartheta_1 \,-\, c_p\, c(\Omega)\, \|\boldsymbol{u}\|_{1,\Omega} \Big\}\, |\widetilde{\varphi}-\widetilde{\psi}|^2_{1,\Omega} \\
&\qquad \geq \quad c_p^{-2} \Big\{ \vartheta_1 \,-\, c_p\, c(\Omega)\, \|\boldsymbol{u}\|_{1,\Omega} \Big\}\, \|\widetilde{\varphi}-\widetilde{\psi}\|^2_{1,\Omega}\,,
\end{aligned}$$

which shows that $A_{\boldsymbol{u}}$ is strongly monotone with constant $\widetilde{\alpha}_{\boldsymbol{u}} := c_p^{-2} \Big\{ \vartheta_1 \,-\, c_p\, c(\Omega)\, \|\boldsymbol{u}\|_{1,\Omega} \Big\}$. In turn, proceeding similarly, we find that for all $\widetilde{\varphi},\, \widetilde{\psi},\, \widetilde{\rho} \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ there holds

$$\begin{aligned}
\big|A_{\boldsymbol{u}}(\widetilde{\varphi},\widetilde{\rho}) \,-\, A_{\boldsymbol{u}}(\widetilde{\psi},\widetilde{\rho})\big| \,&=\, \left| \int_\Omega \Big\{ \vartheta(|\nabla\widetilde{\varphi}|)\nabla\widetilde{\varphi} - \vartheta(|\nabla\widetilde{\psi}|)\nabla\widetilde{\psi} \Big\} \cdot \nabla\widetilde{\rho} \,-\, \int_\Omega (\widetilde{\varphi}-\widetilde{\psi})\boldsymbol{u}\cdot\nabla\widetilde{\rho} \right| \\
&\leq \quad \widetilde{\vartheta}_2\, |\widetilde{\varphi}-\widetilde{\psi}|_{1,\Omega}\, |\widetilde{\rho}|_{1,\Omega} \,+\, c(\Omega)\, \|\widetilde{\varphi}-\widetilde{\psi}\|_{1,\Omega}\, \|\boldsymbol{u}\|_{1,\Omega}\, |\widetilde{\rho}|_{1,\Omega} \\
&\leq \quad \Big\{ \widetilde{\vartheta}_2 \,+\, c(\Omega)\, \|\boldsymbol{u}\|_{1,\Omega} \Big\}\, \|\widetilde{\varphi}-\widetilde{\psi}\|_{1,\Omega}\, \|\widetilde{\rho}\|_{1,\Omega}\,,
\end{aligned}$$

which proves that $A_{\boldsymbol{u}}$ is Lipschitz-continuous with constant $\widetilde{L}_{\boldsymbol{u}} := \widetilde{\vartheta}_2 \,+\, c(\Omega)\, \|\boldsymbol{u}\|_{1,\Omega}$. Therefore, a direct application of a classical result on the bijectivity of monotone operators (see, e.g. [88, Theorem 3.3.23]) implies the existence of a unique solution $\widetilde{\phi} := \widetilde{\mathbf{S}}(\phi,\boldsymbol{u}) \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ of (1.16). Moreover, applying the strong monotonicity of $A_{\boldsymbol{u}}$ to $\widetilde{\varphi} = \widetilde{\phi}$ and $\widetilde{\psi} = 0$, and noting from (1.9) that $A_{\boldsymbol{u}}(0,\cdot) = 0$, we deduce that

$$\widetilde{\alpha}_{\boldsymbol{u}}\, \|\widetilde{\phi}\|^2_{1,\Omega} \,\leq\, A_{\boldsymbol{u}}(\widetilde{\phi},\widetilde{\phi}) \,=\, G_\phi(\widetilde{\phi})\,,$$

which gives $\widetilde{\alpha}_{\boldsymbol{u}}\, \|\widetilde{\phi}\|_{1,\Omega} \,\leq\, \|G_\phi\|$. Finally, using the Cauchy-Schwarz inequality and the upper bound of $\gamma$ (*cf.* (1.2)), it follows from (1.17) that $\|G_\phi\| \,\leq\, \gamma_2\, |\Omega|^{1/2}\, |\boldsymbol{k}| \,+\, \|g\|_{0,\Omega}$, which yields (1.28) and finishes the proof. □

A simple corollary of the above lemma, which removes the dependence on $\boldsymbol{u}$ of the strong monotonicity constant of $A_{\boldsymbol{u}}$ and of the estimate (1.28), is given as follows.

**Lemma 1.6.** *Let $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ and $\boldsymbol{u} \in \mathbf{H}^1(\Omega)$ such that $\|\boldsymbol{u}\|_{1,\Omega} < \frac{\vartheta_1}{2\,c_p\,c(\Omega)}$ (cf. (1.3), (1.6), (1.10)). Then, there exists a unique $\widetilde{\phi} := \widetilde{\mathbf{S}}(\phi,\boldsymbol{u}) \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ solution of (1.16), and there holds*

$$\|\widetilde{\mathbf{S}}(\phi,\boldsymbol{u})\|_{1,\Omega} \,=\, \|\widetilde{\phi}\|_{1,\Omega} \,\leq\, \frac{2\,c_p^2}{\vartheta_1}\left\{\gamma_2\,|\Omega|^{1/2}\,|\boldsymbol{k}| \,+\, \|g\|_{0,\Omega}\right\}. \tag{1.29}$$

*Proof.* It follows directly from the proof of Lemma 1.5. Note in particular that the strong monotonicity of $A_{\boldsymbol{u}}$ holds with the constant $\widetilde{\alpha} := \frac{\vartheta_1}{2\,c_p^2}$. Further details are omitted. □

We end this section by remarking that the restriction on $\|\boldsymbol{u}\|_{1,\Omega}$ in Lemma 1.6 could also have been taken as $\|\boldsymbol{u}\|_{1,\Omega} < \delta\frac{\vartheta_1}{c_p\,c(\Omega)}$ with any $\delta \in (0,1)$. However, we have chosen $\delta = \frac{1}{2}$ for simplicity and because it yields a joint maximization of the constant $\widetilde{\alpha}$ and the upper bound for $\|\boldsymbol{u}\|_{1,\Omega}$.

### 1.3.4 Solvability analysis of the fixed point equation

Having established in the previous section the well-posedness of the uncoupled problems (1.12) and (1.16), which confirms that the operators $\mathbf{S}$, $\widetilde{\mathbf{S}}$, and $\mathbf{T}$ (*cf.* Section 1.3.2) are well defined, we now address the solvability analysis of the fixed point equation (1.18). For this purpose, in what follows we verify the hypotheses of the Schauder fixed point theorem, which is stated as follows (see, e.g. [41, Theorem 9.12-1(b)]).

**Theorem 1.1.** *Let $W$ be a closed and convex subset of a Banach space $X$ and let $T : W \to W$ be a continuous mapping such that $\overline{T(W)}$ is compact. Then $T$ has at least one fixed point.*

We begin the analysis with the following result.

**Lemma 1.7.** *Given $r > 0$, we let $W$ be the closed and convex subset of $\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ defined by*

$$W \,:=\, \left\{\phi \,\in\, \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) : \quad \|\phi\|_{1,\Omega} \,\leq\, r\right\},$$

*and assume that the data satisfy*

$$\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} \,+\, r\,\|\boldsymbol{f}\|_{\infty,\Omega} \,<\, \frac{\vartheta_1}{2\,C_S\,c_p\,c(\Omega)} \qquad \textit{and} \qquad \gamma_2\,|\Omega|^{1/2}\,|\boldsymbol{k}| \,+\, \|g\|_{0,\Omega} \,\leq\, \frac{\vartheta_1\,r}{2\,c_p^2}. \tag{1.30}$$

*Then $\mathbf{T}(W) \,\subseteq\, W$.*

*Proof.* Given $\phi \in W$, we get from (1.21) (*cf.* Lemma 1.4) that

$$\|\mathbf{S}(\phi)\|_H \,=\, \|(\boldsymbol{\sigma},\boldsymbol{u})\|_H \,\leq\, C_S\left\{\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} \,+\, r\,\|\boldsymbol{f}\|_{\infty,\Omega}\,\right\},$$

and hence, thanks to the first restriction in (1.30), we observe that $\boldsymbol{u} = \mathbf{S}_2(\phi)$ satisfies the hypotheses of Lemma 1.6. Moreover, the corresponding estimate (1.29) gives

$$\|\mathbf{T}(\phi)\|_{1,\Omega} \,=\, \|\widetilde{\mathbf{S}}(\phi,\boldsymbol{u})\|_{1,\Omega} \,\leq\, \frac{2\,c_p^2}{\vartheta_1}\left\{\gamma_2\,|\Omega|^{1/2}\,|\boldsymbol{k}| \,+\, \|g\|_{0,\Omega}\right\},$$

which, due to the second inequality in (1.30), proves that $\mathbf{T}(\phi) \,\in\, W$, thus finishing the proof. □

Next, we aim to prove the continuity and compactness properties of $\mathbf{T}$, which basically will be direct consequences of the following two lemmas providing the continuity of $\mathbf{S}$ and $\widetilde{\mathbf{S}}$, respectively. We remark in advance that a combination of the Cauchy-Schwarz and Hölder inequalities with the further regularity assumption specified by (1.27) plays a key role in the proof of the first result.

**Lemma 1.8.** *There exists a positive constant $C$, depending on $\mu_1$, $\kappa_1$, $\kappa_2$, $L_\mu$, $\alpha$, and $\varepsilon$ (*cf. (1.2), (1.11), (1.20), (1.24), (1.27)*), such that*

$$\|\mathbf{S}(\phi)-\mathbf{S}(\psi)\|_H \leq C\left\{\|\boldsymbol{f}\|_{\infty,\Omega}\,\|\phi-\psi\|_{0,\Omega}+\|\mathbf{S}_1(\psi)\|_{\varepsilon,\Omega}\,\|\phi-\psi\|_{\mathrm{L}^{n/\varepsilon}(\Omega)}\right\} \quad \forall\,\phi,\,\psi\in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,. \tag{1.31}$$

*Proof.* Given $\phi,\,\psi\in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$, we let $(\boldsymbol{\sigma},\boldsymbol{u})=\mathbf{S}(\phi)$ and $(\boldsymbol{\zeta},\boldsymbol{w})=\mathbf{S}(\psi)$, that is

$$B_\phi((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau},\boldsymbol{v}))=F_\phi(\boldsymbol{\tau},\boldsymbol{v}) \quad\text{and}\quad B_\psi((\boldsymbol{\zeta},\boldsymbol{w}),(\boldsymbol{\tau},\boldsymbol{v}))=F_\psi(\boldsymbol{\tau},\boldsymbol{v}) \qquad \forall\,(\boldsymbol{\tau},\boldsymbol{v})\in H\,.$$

It follows, using the ellipticity of $B_\phi$ (*cf.* (1.24)) and then subtracting and adding the expression $F_\psi((\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta},\boldsymbol{w}))=B_\psi((\boldsymbol{\zeta},\boldsymbol{w}),(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta},\boldsymbol{w}))$, that

$$\begin{aligned}\alpha\,\|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta},\boldsymbol{w})\|^2_H &\leq B_\phi((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta},\boldsymbol{w}))-B_\phi((\boldsymbol{\zeta},\boldsymbol{w}),(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta},\boldsymbol{w}))\\ &= (F_\phi-F_\psi)\big((\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta},\boldsymbol{w})\big)+(B_\psi-B_\phi)\big((\boldsymbol{\zeta},\boldsymbol{w}),(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta},\boldsymbol{w})\big)\,.\end{aligned} \tag{1.32}$$

Then, according to the definition of $F_\phi$ (*cf.* (1.14)), and applying the Cauchy-Schwarz inequality, we deduce that

$$\begin{aligned}\Big|(F_\phi-F_\psi)\big((\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta},\boldsymbol{w})\big)\Big| &= \left|\int_\Omega \boldsymbol{f}(\phi-\psi)\cdot(\boldsymbol{u}-\boldsymbol{w})-\kappa_2\int_\Omega \boldsymbol{f}(\phi-\psi)\cdot\mathbf{div}(\boldsymbol{\sigma}-\boldsymbol{\zeta})\right|\\ &\leq \|\boldsymbol{f}\|_{\infty,\Omega}\,\|\phi-\psi\|_{0,\Omega}\left\{\|\boldsymbol{u}-\boldsymbol{w}\|_{0,\Omega}+\kappa_2\,\|\mathbf{div}(\boldsymbol{\sigma}-\boldsymbol{\zeta})\|_{0,\Omega}\right\}\\ &\leq \big(1+\kappa_2^2\big)^{1/2}\,\|\boldsymbol{f}\|_{\infty,\Omega}\,\|\phi-\psi\|_{0,\Omega}\,\|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta},\boldsymbol{w})\|_H\,.\end{aligned} \tag{1.33}$$

In turn, it follows easily from (1.13) that

$$(B_\psi-B_\phi)\big((\boldsymbol{\zeta},\boldsymbol{w}),(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta},\boldsymbol{w})\big)=\int_\Omega\left\{\frac{\mu(\phi)-\mu(\psi)}{\mu(\phi)\,\mu(\psi)}\right\}\boldsymbol{\zeta}^{\mathtt{d}}:\Big\{(\boldsymbol{\sigma}-\boldsymbol{\zeta})^{\mathtt{d}}-\kappa_1\,\nabla(\boldsymbol{u}-\boldsymbol{w})\Big\}\,,$$

from which, thanks to the lower bound of $\mu$ (*cf.* (1.2)) and its Lipschitz-continuity assumption (1.20), and applying Cauchy-Schwarz and Hölder inequalities, we find that

$$\begin{aligned}\Big|(B_\psi-B_\phi)\big((\boldsymbol{\zeta},\boldsymbol{w}),(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta},\boldsymbol{w})\big)\Big| &\leq \frac{L_\mu}{\mu_1^2}\int_\Omega\big|(\phi-\psi)\,\boldsymbol{\zeta}^{\mathtt{d}}\big|\,\big|(\boldsymbol{\sigma}-\boldsymbol{\zeta})^{\mathtt{d}}-\kappa_1\,\nabla(\boldsymbol{u}-\boldsymbol{w})\big|\\ &\leq \frac{L_\mu}{\mu_1^2}\,\big\|(\phi-\psi)\,\boldsymbol{\zeta}\big\|_{0,\Omega}\,\big\|(\boldsymbol{\sigma}-\boldsymbol{\zeta})^{\mathtt{d}}-\kappa_1\,\nabla(\boldsymbol{u}-\boldsymbol{w})\big\|_{0,\Omega}\\ &\leq \frac{L_\mu}{\mu_1^2}\,\|\boldsymbol{\zeta}\|_{\mathbb{L}^{2p}(\Omega)}\,\|\phi-\psi\|_{\mathrm{L}^{2q}(\Omega)}\left\{\|\boldsymbol{\sigma}-\boldsymbol{\zeta}\|_{0,\Omega}+\kappa_1\,|\boldsymbol{u}-\boldsymbol{w}|_{1,\Omega}\right\}\\ &\leq \frac{L_\mu\,(1+\kappa_1^2)^{1/2}}{\mu_1^2}\,\|\boldsymbol{\zeta}\|_{\mathbb{L}^{2p}(\Omega)}\,\|\phi-\psi\|_{\mathrm{L}^{2q}(\Omega)}\,\|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta},\boldsymbol{w})\|_H\,,\end{aligned} \tag{1.34}$$

where $p,\,q\in[1,+\infty)$ are such that $\frac{1}{p}+\frac{1}{q}=1$. Next, bearing in mind the further regularity $\varepsilon$ assumed in (1.27), we notice that the Sobolev embedding Theorem (*cf.* [1, Theorem 4.12], [92, Theorem 1.3.4]) establishes the continuous injection $i_\varepsilon:\mathrm{H}^\varepsilon(\Omega)\longrightarrow\mathrm{L}^{\varepsilon^*}(\Omega)$ with boundedness constant $C_\varepsilon$, where

$$\varepsilon^*:=\begin{cases}\frac{2}{1-\varepsilon} & \text{if } n=2\,,\\ \frac{6}{3-2\varepsilon} & \text{if } n=3\,.\end{cases}$$

Thus, choosing $p$ such that $2p = \varepsilon^*$, we deduce that $\boldsymbol{\zeta} := \mathbf{S}_1(\psi)$ does in fact belong to $\mathbb{L}^{2p}(\Omega)$, and hence, thanks to the aforementioned continuity, there holds

$$\|\boldsymbol{\zeta}\|_{\mathbb{L}^{2p}(\Omega)} \leq C_\varepsilon \, \|\boldsymbol{\zeta}\|_{\varepsilon,\Omega} \, , \tag{1.35}$$

which, when needed, can be bounded by (1.27), yielding for each $\psi$ with $\|\psi\|_{1,\Omega} \leq r$

$$\|\boldsymbol{\zeta}\|_{\mathbb{L}^{2p}(\Omega)} \leq C_\varepsilon \, \widetilde{C}_S(r) \left\{ \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma_{\mathrm{D}}} + \|\boldsymbol{f}\|_{\infty,\Omega} \, \|\psi\|_{0,\Omega} \right\} .$$

In addition, according to the above choice of $p$, that is $p = \varepsilon^*/2$, we readily find that

$$2q := \frac{2p}{p-1} = \begin{cases} \frac{2}{\varepsilon} & \text{if } n = 2 \, , \\ \\ \frac{3}{\varepsilon} & \text{if } n = 3 \, . \end{cases} = \frac{n}{\varepsilon} \, . \tag{1.36}$$

In this way, inequalities (1.32), (1.33), (1.34), and (1.35), together with the identity (1.36), imply (1.31) and complete the proof. □

**Lemma 1.9.** *Let $\widetilde{\alpha} := \frac{\vartheta_1}{2\, c_p^2}$ be the strong monotonicity constant provided in the proof of Lemma 1.6. Then, there exists a positive constant $\widetilde{C}$, depending on $\widetilde{\alpha}$, $c(\Omega)$, and $L_\gamma$ (*cf.* (1.10), (1.19)), such that for all $(\phi, \boldsymbol{u})$, $(\varphi, \boldsymbol{w}) \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) \times \mathbf{H}^1(\Omega)$, with $\|\boldsymbol{u}\|_{1,\Omega}$, $\|\boldsymbol{w}\|_{1,\Omega} < \frac{\vartheta_1}{2\, c_p \, c(\Omega)}$, there holds*

$$\|\widetilde{\mathbf{S}}(\phi, \boldsymbol{u}) - \widetilde{\mathbf{S}}(\varphi, \boldsymbol{w})\|_{1,\Omega} \leq \widetilde{C} \left\{ |\boldsymbol{k}| \, \|\phi - \varphi\|_{0,\Omega} + \|\widetilde{\mathbf{S}}(\varphi, \boldsymbol{w})\|_{1,\Omega} \, \|\boldsymbol{u} - \boldsymbol{w}\|_{1,\Omega} \right\} . \tag{1.37}$$

*Proof.* Given $(\phi, \boldsymbol{u})$, $(\varphi, \boldsymbol{w})$ as stated, we let $\widetilde{\phi} := \widetilde{\mathbf{S}}(\phi, \boldsymbol{u})$ and $\widetilde{\varphi} := \widetilde{\mathbf{S}}(\varphi, \boldsymbol{w})$, that is (*cf.* (1.16))

$$A_{\boldsymbol{u}}(\widetilde{\phi}, \widetilde{\psi}) = G_\phi(\widetilde{\psi}) \quad \text{and} \quad A_{\boldsymbol{w}}(\widetilde{\varphi}, \widetilde{\psi}) = G_\varphi(\widetilde{\psi}) \qquad \forall \, \widetilde{\psi} \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) \, .$$

It follows, according to the strong monotonicity of $A_{\boldsymbol{u}}$ with constant $\widetilde{\alpha}$, and then subtracting and adding $G_\varphi(\widetilde{\phi} - \widetilde{\psi}) = A_{\boldsymbol{w}}(\widetilde{\varphi}, \widetilde{\phi} - \widetilde{\psi})$, that

$$\begin{aligned} \widetilde{\alpha} \, \|\widetilde{\phi} - \widetilde{\varphi}\|^2_{1,\Omega} &\leq A_{\boldsymbol{u}}(\widetilde{\phi}, \widetilde{\phi} - \widetilde{\varphi}) - A_{\boldsymbol{u}}(\widetilde{\varphi}, \widetilde{\phi} - \widetilde{\varphi}) \\ &= G_\phi(\widetilde{\phi} - \widetilde{\varphi}) - G_\varphi(\widetilde{\phi} - \widetilde{\varphi}) + A_{\boldsymbol{w}}(\widetilde{\varphi}, \widetilde{\phi} - \widetilde{\psi}) - A_{\boldsymbol{u}}(\widetilde{\varphi}, \widetilde{\phi} - \widetilde{\varphi}) \\ &= \int_\Omega \big(\gamma(\phi) - \gamma(\psi)\big) \, \boldsymbol{k} \cdot \nabla(\widetilde{\phi} - \widetilde{\psi}) + \int_\Omega \widetilde{\varphi} \, (\boldsymbol{u} - \boldsymbol{w}) \cdot \nabla(\widetilde{\phi} - \widetilde{\psi}) \, , \end{aligned}$$

where the last equality has employed the definitions given by (1.9) and (1.17). Then, applying the Lipschitz-continuity of $\gamma$ (*cf.* (1.19)), the Cauchy-Schwarz inequality, and the estimate (1.10), we deduce from the foregoing equation that

$$\widetilde{\alpha} \, \|\widetilde{\phi} - \widetilde{\varphi}\|^2_{1,\Omega} \leq \left\{ L_\gamma \, |\boldsymbol{k}| \, \|\phi - \psi\|_{0,\Omega} + c(\Omega) \, \|\widetilde{\varphi}\|_{1,\Omega} \, \|\boldsymbol{u} - \boldsymbol{w}\|_{1,\Omega} \right\} |\widetilde{\phi} - \widetilde{\psi}|_{1,\Omega} \, ,$$

which gives (1.37) and finishes the proof. □

The following result is a straightforward corollary of Lemmas 1.8 and 1.9.

**Lemma 1.10.** *Given $r > 0$, we let $W := \Big\{ \phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) : \quad \|\phi\|_{1,\Omega} \leq r \Big\}$, and assume that*

$$\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + r \, \|\boldsymbol{f}\|_{\infty,\Omega} < \frac{\vartheta_1}{2\, C_S \, c_p \, c(\Omega)} \qquad \textit{and} \qquad \gamma_2 \, |\Omega|^{1/2} \, |\boldsymbol{k}| + \|g\|_{0,\Omega} \leq \frac{\vartheta_1 \, r}{2\, c_p^2} \, .$$

*Then, with the constants $C$ and $\widetilde{C}$ from Lemmas 1.8 and 1.9, for all $\phi$, $\varphi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ there holds*

$$\begin{aligned}\|\mathbf{T}(\phi) - \mathbf{T}(\varphi)\|_{1,\Omega} \leq & \left\{\widetilde{C}\,|\boldsymbol{k}| + C\,\widetilde{C}\,\|\mathbf{T}(\varphi)\|_{1,\Omega}\,\|\boldsymbol{f}\|_{\infty,\Omega}\right\}\|\phi-\varphi\|_{0,\Omega} \\ & + \quad C\,\widetilde{C}\,\|\mathbf{T}(\varphi)\|_{1,\Omega}\,\|\mathbf{S}_1(\varphi)\|_{\varepsilon,\Omega}\,\|\phi-\varphi\|_{\mathrm{L}^{n/\varepsilon}(\Omega)}\,.\end{aligned} \tag{1.38}$$

*Proof.* It suffices to recall from Section 1.3.2 that $\mathbf{T}(\phi) = \widetilde{\mathbf{S}}(\phi, \mathbf{S}_2(\phi)) \quad \forall\, \phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$, and then apply Lemmas 1.7, 1.8, and 1.9. □

The announced properties of $\mathbf{T}$ are proved now.

**Lemma 1.11.** *Given $r > 0$, we let $W := \left\{\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) : \quad \|\phi\|_{1,\Omega} \leq r\right\}$, and assume that*

$$\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + r\,\|\boldsymbol{f}\|_{\infty,\Omega} < \frac{\vartheta_1}{2\,C_S\,c_p\,c(\Omega)} \qquad \text{and} \qquad \gamma_2\,|\Omega|^{1/2}\,|\boldsymbol{k}| + \|g\|_{0,\Omega} \leq \frac{\vartheta_1\,r}{2\,c_p^2}\,.$$

*Then, $\mathbf{T} : W \longrightarrow W$ is continuous and $\overline{\mathbf{T}(W)}$ is compact.*

*Proof.* We first observe, thanks now to the Rellich-Kondrachov compactness Theorem (*cf.* [1, Theorem 6.3], [92, Theorem 1.3.5]), that the injection $i : \mathrm{H}^1(\Omega) \longrightarrow \mathrm{L}^s(\Omega)$ is compact, and hence continuous, for each $s \geq 1$ (when $n = 2$), and for each $s \in [1, 6)$ (when $n = 3$). Then, according to the assumptions on the further regularity $\varepsilon$ (*cf.* (1.27)), namely $\varepsilon \in (0, 1)$ in $\mathrm{R}^2$ and $\varepsilon \in (\frac{1}{2}, 1)$ in $\mathrm{R}^3$, we realize that $n/\varepsilon$ belongs to the indicated ranges for $s$, and therefore $\mathrm{H}^1(\Omega)$ is compactly, and hence continuously, embedded in $\mathrm{L}^{n/\varepsilon}(\Omega)$, which together with (1.38) imply the continuity of $\mathbf{T}$. In turn, let $\{\phi_k\}_{k\in\mathrm{N}}$ be a sequence of $W$, which is clearly bounded. It follows that there exist a subsequence $\{\phi_k^{(1)}\}_{k\in\mathrm{N}} \subseteq \{\phi_k\}_{k\in\mathrm{N}}$ and $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ such that $\phi_k^{(1)} \xrightarrow{w} \phi$. In this way, since the injections $i_1 : \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) \longrightarrow \mathrm{L}^2(\Omega)$ and $\widetilde{i}_1 : \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) \longrightarrow \mathrm{L}^{n/\varepsilon}(\Omega)$ are compact, we deduce that $\phi_k^{(1)} \longrightarrow \phi$ in $\mathrm{L}^2(\Omega)$ and $\phi_k^{(1)} \longrightarrow \phi$ in $\mathrm{L}^{n/\varepsilon}(\Omega)$, which, combined with (1.38), implies that $\mathbf{T}(\phi_k^{(1)}) \longrightarrow \mathbf{T}(\phi)$ in $\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$. This proves the compactness of $\overline{\mathbf{T}(W)}$ and finishes the proof. □

Finally, the main result of this section is given as follows.

**Theorem 1.2.** *Given $r > 0$, we let $W := \left\{\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) : \quad \|\phi\|_{1,\Omega} \leq r\right\}$, and assume that*

$$\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + r\,\|\boldsymbol{f}\|_{\infty,\Omega} < \frac{\vartheta_1}{2\,C_S\,c_p\,c(\Omega)} \qquad \text{and} \qquad \gamma_2\,|\Omega|^{1/2}\,|\boldsymbol{k}| + \|g\|_{0,\Omega} \leq \frac{\vartheta_1\,r}{2\,c_p^2}\,. \tag{1.39}$$

*Then the augmented mixed-primal problem (1.15) has at least one solution $(\boldsymbol{\sigma}, \boldsymbol{u}, \phi) \in \mathbb{H}_{\mathrm{N}}(\mathbf{div}; \Omega) \times \mathbf{H}^1(\Omega) \times \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ with $\phi \in W$, and there holds*

$$\|\phi\|_{1,\Omega} \leq \frac{2\,c_p^2}{\vartheta_1\,r}\left\{\gamma_2\,|\Omega|^{1/2}\,|\boldsymbol{k}| + \|g\|_{0,\Omega}\right\} \tag{1.40}$$

*and*

$$\|(\boldsymbol{\sigma}, \boldsymbol{u})\|_H \leq C_S\left\{\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + \|\boldsymbol{f}\|_{\infty,\Omega}\,\|\phi\|_{1,\Omega}\right\}. \tag{1.41}$$

*Moreover, if the data $\boldsymbol{k}$, $\boldsymbol{f}$, and $\boldsymbol{u}_{\mathrm{D}}$ are sufficiently small so that, with the constants $C$, $\widetilde{C}$, and $\widetilde{C}_S(r)$ from Lemmas 1.8 and 1.9, and estimate (1.27), and denoting by $\widetilde{C}_\varepsilon$ the boundedness constant of the continuous injection of $\mathrm{H}^1(\Omega)$ into $\mathrm{L}^{n/\varepsilon}(\Omega)$, there holds*

$$\widetilde{C}\,|\boldsymbol{k}| + C\,\widetilde{C}\,r\left\{\left(1 + r\,\widetilde{C}_\varepsilon\,\widetilde{C}_S(r)\right)\|\boldsymbol{f}\|_{\infty,\Omega} + \widetilde{C}_\varepsilon\,\widetilde{C}_S(r)\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma_{\mathrm{D}}}\right\} < 1\,, \tag{1.42}$$

*then the solution $\phi$ is unique in $W$.*

*Proof.* According to the equivalence between (1.15) and the fixed point equation (1.18), and thanks to the previous Lemmas 1.7 and 1.11, the existence of solution is just a straightforward application of the Schauder fixed point theorem (*cf.* Theorem 1.1). In turn, the estimates (1.40) and (1.41) follow from (1.21) (*cf.* Lemma 1.4) and (1.29) (*cf.* Lemma 1.6). Furthermore, given another solution $\varphi \in W$ of (1.18), the estimates $\|\mathbf{T}(\varphi)\|_{1,\Omega} = \|\varphi\|_{1,\Omega} \leq r$, $\|\mathbf{S}_1(\varphi)\|_{\varepsilon,\Omega} \leq \widetilde{C}_S(r)\left\{\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma_{\mathrm{D}}} + \|\boldsymbol{f}\|_{\infty,\Omega}\,\|\varphi\|_{1,\Omega}\right\}$ (*cf.* (1.27)), and $\|\psi\|_{\mathrm{L}^{n/\varepsilon}(\Omega)} \leq \widetilde{C}_\varepsilon\,\|\psi\|_{1,\Omega} \quad \forall\,\psi \in \mathrm{H}^1(\Omega)$, confirm (1.42) as a sufficient condition for concluding, together with (1.38), that $\phi = \varphi$. □

It is important to highlight here that the uniqueness of $\phi$ certainly implies, according to Lemma 1.4, the uniqueness of the solution $\mathbf{S}(\phi) := (\boldsymbol{\sigma}, \boldsymbol{u}) \in H$ of problem (1.12), and hence the foregoing theorem actually guarantees that, under the asumption (1.39) on the data, there exists a unique solution $(\boldsymbol{\sigma}, \boldsymbol{u}, \phi) \in \mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega) \times \mathbf{H}^1(\Omega) \times \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ of problem (1.15) such that $\phi \in W$.

## 1.4 The Galerkin scheme

In this section we introduce and analyze the Galerkin scheme of the augmented mixed-primal problem (1.15). To this end, we now let $\mathcal{T}_h$ be a regular triangulation of $\Omega$ by triangles $K$ (resp. tetrahedra $K$ in $\mathrm{R}^3$) of diameter $h_K$, and define the meshsize $h := \max\left\{h_K : \ K \in \mathcal{T}_h\right\}$. In addition, given an integer $k \geq 0$, for each $K \in \mathcal{T}_h$ we let $\mathrm{P}_k(K)$ be the space of polynomial functions on $K$ of degree $\leq k$, and define the corresponding local Raviart-Thomas space of order $k$ as

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

where, according to the notations described in Section 1.1, $\mathbf{P}_k(K) = [\mathrm{P}_k(K)]^n$, and $\boldsymbol{x}$ is the generic vector in $\mathrm{R}^n$. Then, we introduce the finite element subspaces approximating the unknowns $\boldsymbol{\sigma}$, $\boldsymbol{u}$, and $\phi$, respectively, as the global Raviart-Thomas space of order $k$, and the corresponding Lagrange spaces given by the continuous piecewise polynomials of degree $\leq k+1$, that is

$$\mathbb{H}_h^{\boldsymbol{\sigma}} := \left\{\boldsymbol{\tau}_h \in \mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega) : \quad \mathbf{c}^{\mathtt{t}}\,\boldsymbol{\tau}_h|_K \in \mathbf{RT}_k(K) \quad \forall\,\mathbf{c} \in \mathrm{R}^n\,, \quad \forall\,K \in \mathcal{T}_h\right\}, \tag{1.43}$$

$$\mathbf{H}_h^{\boldsymbol{u}} := \left\{\boldsymbol{v}_h \in \mathbf{C}(\Omega) : \quad \boldsymbol{v}_h|_K \in \mathbf{P}_{k+1}(K) \quad \forall\,K \in \mathcal{T}_h\right\}, \tag{1.44}$$

$$\mathrm{H}_h^{\phi} := \left\{\psi_h \in C(\Omega) \cap \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) : \quad \psi_h|_K \in \mathrm{P}_{k+1}(K) \quad \forall\,K \in \mathcal{T}_h\right\}. \tag{1.45}$$

In this way, the underlying Galerkin's scheme, given by the discrete counterpart of (1.15), reads: Find $(\boldsymbol{\sigma}_h, \boldsymbol{u}_h, \phi_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}} \times \mathrm{H}_h^{\phi}$ such that

$$\begin{aligned} B_{\phi_h}((\boldsymbol{\sigma}_h, \boldsymbol{u}_h), (\boldsymbol{\tau}_h, \boldsymbol{v}_h)) &= F_{\phi_h}(\boldsymbol{\tau}_h, \boldsymbol{v}_h) \quad \forall (\boldsymbol{\tau}_h, \boldsymbol{v}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}\,, \\ A_{\boldsymbol{u}_h}(\phi_h, \psi_h) &= \int_\Omega \gamma(\phi_h)\,\boldsymbol{k} \cdot \nabla\psi_h + \int_\Omega g\psi_h \qquad \forall\,\psi_h \in \mathrm{H}_h^{\phi}\,. \end{aligned} \tag{1.46}$$

Throughout the rest of this section we adopt the discrete analogue of the fixed point strategy introduced in Section 1.3.2. Hence, we now let $\mathbf{S}_h : \mathrm{H}_h^{\phi} \longrightarrow \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}$ be the operator defined by

$$\mathbf{S}_h(\phi_h) = (\mathbf{S}_{1,h}(\phi_h), \mathbf{S}_{2,h}(\phi_h)) := (\boldsymbol{\sigma}_h, \boldsymbol{u}_h) \qquad \forall\,\phi_h \in \mathrm{H}_h^{\phi}\,,$$

where $(\boldsymbol{\sigma}_h, \boldsymbol{u}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}$ is the unique solution of

$$B_{\phi_h}((\boldsymbol{\sigma}_h, \boldsymbol{u}_h), (\boldsymbol{\tau}_h, \boldsymbol{v}_h)) \,=\, F_{\phi_h}(\boldsymbol{\tau}_h, \boldsymbol{v}_h) \quad \forall\, (\boldsymbol{\tau}_h, \boldsymbol{v}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}\,, \tag{1.47}$$

with $B_{\phi_h}$ and $F_{\phi_h}$ being defined by (1.13) and (1.14), respectively, with $\phi = \phi_h$. In addition, we let $\widetilde{\mathbf{S}}_h : \mathrm{H}_h^{\phi} \times \mathbf{H}_h^{\boldsymbol{u}} \longrightarrow \mathrm{H}_h^{\phi}$ be the operator defined by

$$\widetilde{\mathbf{S}}_h(\phi_h, \boldsymbol{u}_h) \,:=\, \widetilde{\phi}_h \qquad \forall\, (\phi_h, \boldsymbol{u}_h) \,\in\, \mathrm{H}_h^{\phi} \times \mathbf{H}_h^{\boldsymbol{u}}\,,$$

where $\widetilde{\phi}_h \in \mathrm{H}_h^{\phi}$ is the unique solution of

$$A_{\boldsymbol{u}_h}(\widetilde{\phi}_h, \widetilde{\psi}_h) \,=\, G_{\phi_h}(\widetilde{\psi}_h) \qquad \forall\, \widetilde{\psi}_h \in \mathrm{H}_h^{\phi}\,, \tag{1.48}$$

with $A_{\boldsymbol{u}_h}$ and $G_{\phi_h}$ being defined by (1.9) and (1.17), respectively, with $\boldsymbol{u} = \boldsymbol{u}_h$ and $\phi = \phi_h$. Finally, we define the operator $\mathbf{T}_h : \mathrm{H}_h^{\phi} \longrightarrow \mathrm{H}_h^{\phi}$ by

$$\mathbf{T}_h(\phi_h) \,:=\, \widetilde{\mathbf{S}}_h(\phi_h, \mathbf{S}_{2,h}(\phi_h)) \qquad \forall\, \phi_h \,\in\, \mathrm{H}_h^{\phi}\,,$$

and realize that (1.46) can be rewritten, equivalently, as: Find $\phi_h \in \mathrm{H}_h^{\phi}$ such that

$$\mathbf{T}_h(\phi_h) \,=\, \phi_h\,. \tag{1.49}$$

Certainly, all the above makes sense if we guarantee that the discrete problems (1.47) and (1.48) are well-posed. Indeed, it is easy to see that the respective proofs are almost verbatim of the continuous analogues provided in Section 1.3.3, and hence we simply state the corresponding results as follows.

**Lemma 1.12.** *Assume that* $\kappa_1 \in \left(0, \frac{2\delta\mu_1}{\mu_2}\right)$ *with* $\delta \in (0, 2\mu_1)$, *and that* $0 < \kappa_2, \kappa_3$. *Then, for each* $\phi_h \in \mathrm{H}_h^{\phi}$ *the problem* (1.47) *has a unique solution* $\mathbf{S}_h(\phi_h) := (\boldsymbol{\sigma}_h, \boldsymbol{u}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}$. *Moreover, with the same constant* $C_S > 0$ *from Lemma* 1.4, *there holds*

$$\|\mathbf{S}_h(\phi_h)\|_H \,=\, \|(\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\|_H \,\leq\, C_S \left\{ \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} \,+\, \|\boldsymbol{f}\|_{\infty,\Omega}\, \|\phi_h\|_{0,\Omega} \right\} \qquad \forall\, \phi_h \,\in\, \mathrm{H}_h^{\phi}\,.$$

*Proof.* It suffices to see that for each $\phi_h \in \mathrm{H}_h^{\phi}$, $B_{\phi_h}$ is elliptic on $\mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}$ with the same constant $\alpha$ from Lemma 1.4 (*cf.* (1.24)), and that $\|F_{\phi_h}\|_{\left(\mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}\right)'}$ is bounded as in (1.26) with $\phi_h$ in place of $\phi$. The rest of the proof is a direct application of the Lax-Milgram lemma. □

**Lemma 1.13.** *Let* $\phi_h \in \mathrm{H}_h^{\phi}$ *and* $\boldsymbol{u}_h \in \mathbf{H}_h^{\boldsymbol{u}}$ *such that* $\|\boldsymbol{u}_h\|_{1,\Omega} < \frac{\vartheta_1}{2\, c_p\, c(\Omega)}$ (cf. (1.3), (1.6), (1.10)). *Then, there exists a unique* $\widetilde{\phi}_h := \widetilde{\mathbf{S}}_h(\phi_h, \boldsymbol{u}_h) \in \mathrm{H}_h^{\phi}$ *solution of* (1.48), *and there holds*

$$\|\widetilde{\mathbf{S}}_h(\phi_h, \boldsymbol{u}_h)\|_{1,\Omega} \,=\, \|\widetilde{\phi}_h\|_{1,\Omega} \,\leq\, \frac{2\, c_p^2}{\vartheta_1} \left\{ \gamma_2\, |\Omega|^{1/2}\, |\boldsymbol{k}| \,+\, \|g\|_{0,\Omega} \right\}.$$

*Proof.* It basically follows by observing that, under the assumption on $\|\boldsymbol{u}_h\|_{1,\Omega}$, $A_{\boldsymbol{u}_h}$ becomes Lipschitz-continuous and strongly monotone on $\mathrm{H}_h^{\phi} \times \mathrm{H}_h^{\phi}$ with the constants $\widetilde{L}_{\boldsymbol{u}_h} := \widetilde{\vartheta}_2 \,+\, c(\Omega)\, \|\boldsymbol{u}_h\|_{1,\Omega}$ and $\widetilde{\alpha} := \frac{\vartheta_1}{2\, c_p^2}$ given in the proofs of Lemmas 1.5) and 1.6, respectively, and then applying again [88, Theorem 3.3.23]. In addition, the fact that $\|G_{\phi}\|$ is bounded independently of $\phi$ (*cf.* proof of Lemma 1.5), confirms the same upper bound for $\|G_{\phi_h}\|_{\left(\mathrm{H}_h^{\phi}\right)'}$. □

We now aim to show the solvability of (1.46) by analyzing the equivalent fixed point equation (1.49). To this end, in what follows we verify the hypotheses of the Brouwer fixed point theorem, which is given as follows (see, e.g. [41, Theorem 9.9-2]).

**Theorem 1.3.** *Let $W$ be a compact and convex subset of a finite dimensional Banach space $X$ and let $T : W \to W$ be a continuous mapping. Then $T$ has at least one fixed point.*

We begin with the discrete version of Lemma 1.7.

**Lemma 1.14.** *Given $r > 0$, we let $W_h := \Big\{ \phi_h \in \mathrm{H}^{\phi}_h : \quad \|\phi_h\|_{1,\Omega} \leq r \Big\}$, and assume that*

$$\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + r\,\|\boldsymbol{f}\|_{\infty,\Omega} < \frac{\vartheta_1}{2\,C_S\,c_p\,c(\Omega)} \qquad \text{and} \qquad \gamma_2\,|\Omega|^{1/2}\,|\boldsymbol{k}| + \|g\|_{0,\Omega} \leq \frac{\vartheta_1\,r}{2\,c_p^2}\,.$$

*Then $\mathbf{T}_h(W_h) \subseteq W_h$.*

*Proof.* It is a straightforward consequence of Lemmas 1.12 and 1.13. □

The discrete analogue of Lemma 1.8 is provided next. We notice in advance that, instead of the regularity assumption employed in the proof of that result, which actually is not needed nor could be applied in the present discrete case, we simply utilize a $\mathrm{L}^4 - \mathrm{L}^4 - \mathrm{L}^2$ argument.

**Lemma 1.15.** *There exists a positive constant $C$, depending on $\mu_1$, $\kappa_1$, $\kappa_2$, $L_\mu$, and $\alpha$ (*cf. (1.2), (1.11), (1.20), (1.24)*), such that*

$$\|\mathbf{S}_h(\phi_h) - \mathbf{S}_h(\psi_h)\|_H \leq C \Big\{ \|\boldsymbol{f}\|_{\infty,\Omega}\,\|\phi_h - \psi_h\|_{0,\Omega} + \|\mathbf{S}_{1,h}(\psi_h)\|_{\mathbb{L}^4(\Omega)}\,\|\phi_h - \psi_h\|_{\mathrm{L}^4(\Omega)} \Big\}\,, \tag{1.50}$$

*for all $\phi_h$, $\psi_h \in \mathrm{H}^{\phi}_h$.*

*Proof.* It proceeds exactly as in the proof of Lemma 1.8, except for the derivation of the discrete analogue of (1.34), where, instead of choosing the values of $p$ and $q$ determined by the regularity parameter $\delta$, it suffices to take $p = q = 2$, thus obtaining

$$\begin{aligned} &\Big| (B_{\psi_h} - B_{\phi_h})\big((\boldsymbol{\zeta}_h, \boldsymbol{w}_h), (\boldsymbol{\sigma}_h, \boldsymbol{u}_h) - (\boldsymbol{\zeta}_h, \boldsymbol{w}_h)\big) \Big| \\ &\qquad \leq \frac{L_\mu\,(1 + \kappa_1^2)^{1/2}}{\mu_1^2}\,\|\boldsymbol{\zeta}_h\|_{\mathbb{L}^4(\Omega)}\,\|\phi_h - \psi_h\|_{\mathrm{L}^4(\Omega)}\,\|(\boldsymbol{\sigma}_h, \boldsymbol{u}_h) - (\boldsymbol{\zeta}_h, \boldsymbol{w}_h)\|_H\,, \end{aligned}$$

for all $\phi_h$, $\psi_h \in \mathrm{H}^{\phi}_h$, with $(\boldsymbol{\sigma}_h, \boldsymbol{u}_h) := \mathbf{S}_h(\phi_h)$ and $(\boldsymbol{\zeta}_h, \boldsymbol{w}_h) := \mathbf{S}_h(\psi_h)$. Thus, the fact that the elements of $\mathbb{H}^{\boldsymbol{\sigma}}_h$ are piecewise polynomials insures that $\|\boldsymbol{\zeta}_h\|_{\mathbb{L}^4(\Omega)} < +\infty$ for each $\boldsymbol{\zeta}_h \in \mathbb{H}^{\boldsymbol{\sigma}}_h$. Further details are omitted. □

Now, utilizing Lemma 1.15 and the discrete analogue of Lemma 1.9 (which for sake of space saving is not specified here), we can prove the discrete version of Lemma 1.10.

**Lemma 1.16.** *Given $r > 0$, we let $W_h := \Big\{ \phi_h \in \mathrm{H}^{\phi}_h : \quad \|\phi_h\|_{1,\Omega} \leq r \Big\}$, and assume that*

$$\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + r\,\|\boldsymbol{f}\|_{\infty,\Omega} < \frac{\vartheta_1}{2\,C_S\,c_p\,c(\Omega)} \qquad \text{and} \qquad \gamma_2\,|\Omega|^{1/2}\,|\boldsymbol{k}| + \|g\|_{0,\Omega} \leq \frac{\vartheta_1\,r}{2\,c_p^2}\,.$$

*Then, with the constants $C$ and $\widetilde{C}$ from Lemmas 1.15 and 1.9, for all $\phi_h$, $\varphi_h \in \mathrm{H}_h^\phi$ there holds*

$$\begin{aligned}\|\mathbf{T}_h(\phi_h) - \mathbf{T}_h(\varphi_h)\|_{1,\Omega} \leq & \left\{\widetilde{C}\,|\boldsymbol{k}| + C\,\widetilde{C}\,\|\mathbf{T}_h(\varphi_h)\|_{1,\Omega}\,\|\boldsymbol{f}\|_{\infty,\Omega}\right\}\|\phi_h - \varphi_h\|_{0,\Omega} \\ & + \quad C\,\widetilde{C}\,\|\mathbf{T}_h(\varphi_h)\|_{1,\Omega}\,\|\mathbf{S}_{1,h}(\varphi_h)\|_{\mathrm{L}^4(\Omega)}\,\|\phi_h - \varphi_h\|_{\mathrm{L}^4(\Omega)}\,.\end{aligned} \tag{1.51}$$

Consequently, since the foregoing lemma and the continuous injection of $\mathrm{H}^1(\Omega)$ into $\mathrm{L}^4(\Omega)$ confirm the continuity of $\mathbf{T}_h$, we conclude, thanks to the Brouwer fixed point theorem (*cf.* Theorem 1.3) and Lemmas 1.14 and 1.16, the main result of this section.

**Theorem 1.4.** *Given $r > 0$, we let $W_h := \left\{\phi_h \in \mathrm{H}_h^\phi : \quad \|\phi_h\|_{1,\Omega} \leq r\right\}$, and assume that*

$$\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + r\,\|\boldsymbol{f}\|_{\infty,\Omega} < \frac{\vartheta_1}{2\,C_S\,c_p\,c(\Omega)} \qquad \text{and} \qquad \gamma_2\,|\Omega|^{1/2}\,|\boldsymbol{k}| + \|g\|_{0,\Omega} \leq \frac{\vartheta_1\,r}{2\,c_p^2}\,.$$

*Then the Galerkin scheme (1.46) has at least one solution $(\boldsymbol{\sigma}_h, \boldsymbol{u}_h, \phi_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}} \times \mathrm{H}_h^\phi$ with $\phi_h \in W_h$, and there holds*

$$\|\phi_h\|_{1,\Omega} \leq \frac{2\,c_p^2}{\vartheta_1\,r}\left\{\gamma_2\,|\Omega|^{1/2}\,|\boldsymbol{k}| + \|g\|_{0,\Omega}\right\}$$

*and*

$$\|(\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\|_H \leq C_S\left\{\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + \|\boldsymbol{f}\|_{\infty,\Omega}\,\|\phi_h\|_{1,\Omega}\right\}.$$

## 1.5 A priori error analysis

Given $(\boldsymbol{\sigma}, \boldsymbol{u}, \phi) \in \mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega) \times \mathbf{H}^1(\Omega) \times \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ with $\phi \in W$, and $(\boldsymbol{\sigma}_h, \boldsymbol{u}_h, \phi_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}} \times \mathrm{H}_h^\phi$ with $\phi_h \in W_h$, solutions of (1.15) and (1.46), respectively, we now aim to derive a corresponding *a priori* error estimate. For this purpose, we now recall from (1.15) and (1.46), that the above means that

$$\begin{aligned} B_\phi((\boldsymbol{\sigma}, \boldsymbol{u}), (\boldsymbol{\tau}, \boldsymbol{v})) &= F_\phi(\boldsymbol{\tau}, \boldsymbol{v}) \quad \forall (\boldsymbol{\tau}, \boldsymbol{v}) \in \mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega) \times \mathbf{H}^1(\Omega)\,, \\ B_{\phi_h}((\boldsymbol{\sigma}_h, \boldsymbol{u}_h), (\boldsymbol{\tau}_h, \boldsymbol{v}_h)) &= F_{\phi_h}(\boldsymbol{\tau}_h, \boldsymbol{v}_h) \quad \forall (\boldsymbol{\tau}_h, \boldsymbol{v}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}\,, \end{aligned} \tag{1.52}$$

and

$$\begin{aligned} A_{\boldsymbol{u}}(\phi, \psi) &= G_\phi(\psi) \qquad \forall\, \psi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,, \\ A_{\boldsymbol{u}_h}(\phi_h, \psi_h) &= G_{\phi_h}(\psi_h) \qquad \forall\, \psi_h \in \mathrm{H}_h^\phi\,. \end{aligned} \tag{1.53}$$

Next, we recall from [68] a Strang-type lemma, which will be utilized in our subsequent analysis.

**Lemma 1.17.** *Let $H$ be a Hilbert space, $F \in H'$, and $\mathbf{A} : H \to H'$ a nonlinear operator. In addition, let $\{H_n\}_{n\in N}$ be a sequence of finite dimensional subspaces of $H$, and for each $n \in N$ consider a nonlinear operator $\mathbf{A}_n : H_n \to H_n'$ and a functional $F_n \in H_n'$. Assume that the family $\{\mathbf{A}\} \cup \{\mathbf{A}_n\}_{n\in N}$ is uniformly Lipschitz continuous and strongly monotone with constants $\Lambda_{\mathtt{LC}}$ and $\Lambda_{\mathtt{SM}}$, respectively. In turn, let $u \in H$ and $u_n \in H_n$ such that*

$$[\mathbf{A}(u), v] = [F, v] \qquad \forall\, v \in H \quad \text{and} \quad [\mathbf{A}_n(u_n), v_n] = [F_n, v_n] \qquad \forall\, v_n \in H_n\,,$$

*where* $[\cdot,\cdot]$ *denotes the duality pairings of both* $H' \times H$ *and* $H_n' \times H_n$. *Then for each* $n \in N$ *there holds*

$$\|u-u_n\|_H \leq \Lambda_{\mathtt{ST}} \left\{ \sup_{\substack{w_n \in H_n \\ w_n \neq \mathbf{0}}} \frac{\big| [F,w_n] - [F_n,w_n] \big|}{\|w_n\|_H} + \inf_{\substack{v_n \in H_n \\ v_n \neq \mathbf{0}}} \left( \|u-v_n\|_H + \sup_{\substack{w_n \in H_n \\ w_n \neq \mathbf{0}}} \frac{\big| [\mathbf{A}(v_n),w_n] - [\mathbf{A}_n(v_n),w_n] \big|}{\|w_n\|_H} \right) \right\},$$

*with* $\Lambda_{\mathtt{ST}} := \Lambda_{\mathtt{SM}}^{-1} \max\big\{1, \Lambda_{\mathtt{SM}} + \Lambda_{\mathtt{LC}}\big\}$.

*Proof.* It is a particular case of [68, Theorem 6.4]. □

We begin our analysis by denoting as usual

$$\mathrm{dist}\big(\phi, \mathrm{H}_h^{\phi}\big) := \inf_{\varphi_h \in \mathrm{H}_h^{\phi}} \|\phi - \varphi_h\|_{1,\Omega},$$

and

$$\mathrm{dist}\big((\boldsymbol{\sigma},\boldsymbol{u}), \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}\big) := \inf_{(\boldsymbol{\tau}_h,\boldsymbol{v}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}} \|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\tau}_h,\boldsymbol{v}_h)\|_H .$$

Then, we have the following result concerning $\|\phi - \phi_h\|_{1,\Omega}$.

**Lemma 1.18.** *Let* $\widetilde{C}_{\mathtt{ST}} := \widetilde{\alpha}^{-1} \max\big\{1, \widetilde{\alpha} + \widetilde{L}\big\}$, *with* $\widetilde{\alpha} := \frac{\vartheta_1}{2\,c_p^2}$ *and* $\widetilde{L} := \widetilde{\vartheta}_2 + \frac{\vartheta_1}{2\,c_p}$. *Then there holds*

$$\|\phi - \phi_h\|_{1,\Omega} \leq \widetilde{C}_{\mathtt{ST}} \left\{ L_\gamma\,|\boldsymbol{k}|\,\|\phi-\phi_h\|_{0,\Omega} + c(\Omega)\,\|\phi\|_{1,\Omega}\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega} + \Big(1 + c(\Omega)\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}\Big)\,\mathrm{dist}(\phi, \mathrm{H}_h^{\phi}) \right\}. \tag{1.54}$$

*Proof.* We first observe from Lemmas 1.5, 1.6, and 1.13, that the nonlinear operators $A_{\boldsymbol{u}}$ and $A_{\boldsymbol{u}_h}$ are both strongly monotone and Lipschitz-continuous on their corresponding spaces with constants $\widetilde{\alpha}$ and $\widetilde{L}$, respectively. Then, by applying the abstract Lemma 1.17 to the context (1.53), we find that

$$\|\phi - \phi_h\|_{1,\Omega} \leq \widetilde{C}_{\mathtt{ST}} \left\{ \sup_{\substack{\psi_h \in \mathrm{H}_h^{\phi} \\ \psi_h \neq \mathbf{0}}} \frac{\big| G_\phi(\psi_h) - G_{\phi_h}(\psi_h) \big|}{\|\psi_h\|_{1,\Omega}} + \inf_{\substack{\varphi_h \in \mathrm{H}_h^{\phi} \\ \varphi_h \neq \mathbf{0}}} \left( \|\phi - \varphi_h\|_{1,\Omega} + \sup_{\substack{\psi_h \in \mathrm{H}_h^{\phi} \\ \psi_h \neq \mathbf{0}}} \frac{\big| A_{\boldsymbol{u}}(\varphi_h,\psi_h) - A_{\boldsymbol{u}_h}(\varphi_h,\psi_h) \big|}{\|\psi_h\|_{1,\Omega}} \right) \right\}. \tag{1.55}$$

Next, we proceed similarly as in the proof of Lemma 1.9 to estimate each term in the foregoing equation involving a supremum. In fact, according to the definition of $G_\phi$ (*cf.* (1.17)), and applying the same arguments from that proof, we readily see that

$$\sup_{\substack{\psi_h \in \mathrm{H}_h^{\phi} \\ \psi_h \neq \mathbf{0}}} \frac{\big| G_\phi(\psi_h) - G_{\phi_h}(\psi_h) \big|}{\|\psi_h\|_{1,\Omega}} \leq L_\gamma\,|\boldsymbol{k}|\,\|\phi-\phi_h\|_{0,\Omega}. \tag{1.56}$$

In turn, it is clear from the definition of $A_{\boldsymbol{u}}$ (*cf.* (1.9)) and the estimate (1.10) that for each $\varphi_h \in \mathrm{H}_h^\phi$ there holds

$$\begin{aligned}\sup_{\substack{\psi_h \in \mathrm{H}_h^\phi \\ \psi_h \neq \mathbf{0}}} \frac{\big| A_{\boldsymbol{u}}(\varphi_h,\psi_h) - A_{\boldsymbol{u}_h}(\varphi_h,\psi_h)\big|}{\|\psi_h\|_{1,\Omega}} &\leq c(\Omega)\,\|\varphi_h\|_{1,\Omega}\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega} \\ &\leq c(\Omega)\,\|\phi-\varphi_h\|_{1,\Omega}\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega} + c(\Omega)\,\|\phi\|_{1,\Omega}\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}\,.\end{aligned} \tag{1.57}$$

In this way, replacing (1.56) and (1.57) back into (1.55), we arrive at (1.54) and end the proof. □

The following lemma provides a preliminary estimate for the error $\|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H$.

**Lemma 1.19.** *Let $C_{\mathtt{ST}} := \alpha^{-1}\max\{1,\alpha+\|B\|\}$, where $\|B\|$ and $\alpha$ are the boundedness and ellipticity constants, respectively, of the bilinear forms $B_\phi$ (*cf.* (1.22), (1.24)). Then there holds*

$$\begin{aligned}\|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H \leq C_{\mathtt{ST}}\Big\{&\big(1+2\,\|B\|\big)\,\mathrm{dist}\big((\boldsymbol{\sigma},\boldsymbol{u}),\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}}\big) \\ &+ \big(1+\kappa_2^2\big)^{1/2}\,\|\boldsymbol{f}\|_{\infty,\Omega}\,\|\phi-\phi_h\|_{0,\Omega} + \frac{L_\mu\,(1+\kappa_1^2)^{1/2}}{\mu_1^2}\,C_\varepsilon\,\|\boldsymbol{\sigma}\|_{\varepsilon,\Omega}\,\|\phi-\phi_h\|_{\mathrm{L}^{n/\varepsilon}(\Omega)}\Big\}.\end{aligned} \tag{1.58}$$

*Proof.* By applying the abstract Lemma 1.17 to the context (1.52), we obtain

$$\begin{aligned}&\|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H \\ &\leq C_{\mathtt{ST}}\Bigg\{\sup_{\substack{(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\in\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}} \\ (\boldsymbol{\tau}_h,\boldsymbol{v}_h)\neq\mathbf{0}}} \frac{\big|F_\phi(\boldsymbol{\tau}_h,\boldsymbol{v}_h)-F_{\phi_h}(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\big|}{\|(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\|_H} \\ &+ \inf_{\substack{(\boldsymbol{\zeta}_h,\boldsymbol{w}_h)\in\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}} \\ (\boldsymbol{\zeta}_h,\boldsymbol{w}_h)\neq\mathbf{0}}} \Bigg(\|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta}_h,\boldsymbol{w}_h)\|_H \\ &+ \sup_{\substack{(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\in\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}} \\ (\boldsymbol{\tau}_h,\boldsymbol{v}_h)\neq\mathbf{0}}} \frac{\big|B_\phi((\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h))-B_{\phi_h}((\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h))\big|}{\|(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\|_H}\Bigg)\Bigg\}.\end{aligned} \tag{1.59}$$

Then, proceeding analogously as in the proof of Lemma 1.8 (*cf.* (1.33)), we first deduce that

$$\sup_{\substack{(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\in\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}} \\ (\boldsymbol{\tau}_h,\boldsymbol{v}_h)\neq\mathbf{0}}} \frac{\big|F_\phi(\boldsymbol{\tau}_h,\boldsymbol{v}_h)-F_{\phi_h}(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\big|}{\|(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\|_H} \leq \big(1+\kappa_2^2\big)^{1/2}\,\|\boldsymbol{f}\|_{\infty,\Omega}\,\|\phi-\phi_h\|_{0,\Omega}\,. \tag{1.60}$$

In turn, in order to estimate the supremum in (1.59), we add and substract suitable terms to write

$$\begin{aligned}B_\phi((\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h)) - B_{\phi_h}((\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h)) &= B_\phi((\boldsymbol{\zeta}_h,\boldsymbol{w}_h)-(\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau}_h,\boldsymbol{v}_h)) \\ &+ \big(B_\phi - B_{\phi_h}\big)((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau}_h,\boldsymbol{v}_h)) + B_{\phi_h}((\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h))\,,\end{aligned}$$

whence, applying the boundedness (1.22) to the first and third terms on the right hand side of the foregoing equation, and proceeding analogously as for the derivation of (1.34) and (1.35) with the second one, we find that

$$\begin{aligned}&\sup_{\substack{(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\in\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}} \\ (\boldsymbol{\tau}_h,\boldsymbol{v}_h)\neq\mathbf{0}}} \frac{\big|B_\phi((\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h))-B_{\phi_h}((\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h))\big|}{\|(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\|_H} \\ &\leq 2\,\|B\|\,\|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta}_h,\boldsymbol{w}_h)\|_H + \frac{L_\mu\,(1+\kappa_1^2)^{1/2}}{\mu_1^2}\,C_\varepsilon\,\|\boldsymbol{\sigma}\|_{\varepsilon,\Omega}\,\|\phi-\phi_h\|_{\mathrm{L}^{n/\varepsilon}(\Omega)}\,.\end{aligned} \tag{1.61}$$

Finally, by replacing (1.60) and (1.61) into (1.59), we arrive at (1.58), which ends the proof. □

We now combine the inequalities provided by Lemmas 1.18 and 1.19 to derive the Céa estimate for the total error $\|\phi-\phi_h\|_{1,\Omega} + \|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H$. To this end, and in order to simplify the subsequent writing, we introduce the following constants

$$C_1 := \widetilde{C}_{\mathtt{ST}}\, L_\gamma\,, \quad C_2 := \widetilde{C}_{\mathtt{ST}}\, c(\Omega)\, r\, C_{\mathtt{ST}}\, (1+\kappa_2^2)^{1/2}\,, \quad C_3 := \widetilde{C}_{\mathtt{ST}}\, c(\Omega)\, r\, C_{\mathtt{ST}}\, \frac{L_\mu\,(1+\kappa_1^2)^{1/2}}{\mu_1^2}\, C_\varepsilon\, \widetilde{C}_S(r)\, \widetilde{C}_\varepsilon\,.$$

Hence, by replacing the bound for $\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}$ given by (1.58) into the second term on the right hand side of (1.54), recalling that $\|\phi\|_{1,\Omega} \leq r$, employing from (1.27) that

$$\|\boldsymbol{\sigma}\|_{\varepsilon,\Omega} \leq \widetilde{C}_S(r)\Big\{\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma_{\mathrm{D}}} + \|\boldsymbol{f}\|_{\infty,\Omega}\,\|\phi\|_{0,\Omega}\Big\}\,,$$

using again that $\widetilde{C}_\varepsilon$ is the boundedness constant of the continuous injection of $\mathrm{H}^1(\Omega)$ into $\mathrm{L}^{n/\varepsilon}(\Omega)$, and performing several algebraic manipulations, we can assert that

$$\begin{aligned}\|\phi-\phi_h\|_{1,\Omega} \leq\ & \Big\{C_1\,|\boldsymbol{k}| + \big(C_2 + r\,C_3\big)\,\|\boldsymbol{f}\|_{\infty,\Omega} + C_3\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma_{\mathrm{D}}}\Big\}\,\|\phi-\phi_h\|_{1,\Omega}\\ &+ \widetilde{C}_{\mathtt{ST}}\, c(\Omega)\, r\, C_{\mathtt{ST}}\,\big(1+2\|B\|\big)\,\mathrm{dist}\big((\boldsymbol{\sigma},\boldsymbol{u}),\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}}\big)\\ &+ \widetilde{C}_{\mathtt{ST}}\,\Big(1 + c(\Omega)\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}\Big)\,\mathrm{dist}(\phi,\mathrm{H}_h^{\phi})\,.\end{aligned} \tag{1.62}$$

Note here that $\|\boldsymbol{u}\|_{1,\Omega}$ and $\|\boldsymbol{u}_h\|_{1,\Omega}$ are estimated according to (1.21), and hence the expression in (1.62) multiplying $\mathrm{dist}(\phi,\mathrm{H}_h^{\phi})$ is already controlled by constants, parameters, and data only. As a consequence of the foregoing discussion, we can establish the following result providing the requested Céa estimate.

**Theorem 1.5.** *Assume that the data $\boldsymbol{k}$, $\boldsymbol{f}$, and $\boldsymbol{u}_{\mathrm{D}}$ are sufficiently small so that*

$$C_1\,|\boldsymbol{k}| + \big(C_2 + r\,C_3\big)\,\|\boldsymbol{f}\|_{\infty,\Omega} + C_3\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma_{\mathrm{D}}} < \frac{1}{2}\,. \tag{1.63}$$

*Then, there exist positive constants $C_4$ and $C_5$, depending only on parameters, data, and other constants, all them independent of $h$, such that*

$$\|\phi-\phi_h\|_{1,\Omega} + \|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H \leq C_4\,\mathrm{dist}\big((\boldsymbol{\sigma},\boldsymbol{u}),\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}}\big) + C_5\,\mathrm{dist}(\phi,\mathrm{H}_h^{\phi})\,. \tag{1.64}$$

*Proof.* The estimate for $\|\phi-\phi_h\|_{1,\Omega}$ follows straightforwardly from (1.62) and (1.63), and then, the replacement of it back into (1.58), using also that $\|\phi-\phi_h\|_{\mathrm{L}^{n/\varepsilon}(\Omega)} \leq \widetilde{C}_\varepsilon\,\|\phi-\phi_h\|_{1,\Omega}$, completes the proof. □

We end this section with the corresponding rates of convergence of our Galerkin scheme (1.46).

**Theorem 1.6.** *In addition to the hypotheses of Theorems 1.2, 1.4, and 1.5, assume that there exists $s>0$ such that $\boldsymbol{\sigma}\in\mathbb{H}^s(\Omega)$, $\mathbf{div}\boldsymbol{\sigma}\in\mathbf{H}^s(\Omega)$, $\boldsymbol{u}\in\mathbf{H}^{1+s}(\Omega)$, and $\phi\in\mathrm{H}^{1+s}(\Omega)$. Then, there exists $\widehat{C}>0$, independent of $h$, such that, with the finite element subspaces defined by (1.43), (1.44), and (1.45), there holds*

$$\begin{aligned}&\|\phi-\phi_h\|_{1,\Omega} + \|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H\\ &\qquad\leq\ \widehat{C}\,h^{\min\{s,k+1\}}\Big\{\|\boldsymbol{\sigma}\|_{s,\Omega} + \|\mathbf{div}\boldsymbol{\sigma}\|_{s,\Omega} + \|\boldsymbol{u}\|_{1+s,\Omega} + \|\phi\|_{1+s,\Omega}\Big\}\,.\end{aligned} \tag{1.65}$$

*Proof.* It follows directly from the Céa estimate (1.64) and the approximation properties of $\mathbb{H}_h^{\boldsymbol{\sigma}}$, $\mathbf{H}_h^{\boldsymbol{u}}$, and $\mathrm{H}_h^{\phi}$ (*cf.* [26, 40, 63]). □

## 1.6 Numerical results

We illustrate the performance of our mixed–primal finite element method with some numerical tests. We first study the accuracy of the approximations by manufacturing an exact solution of the nonlinear problem (1.1) defined on $\Omega = (0,1)^2$. We introduce the coefficients $\mu(\phi) = (1-c\phi)^{-2}$, $\gamma(\phi) = c\phi(1-c\phi)^2$, $\vartheta(|\nabla\phi|) = m_1 + m_2(1+|\nabla\phi|^2)^{m_3/2-1}$, and the source terms on the right hand sides are adjusted in such a way that the exact solutions are given by the smooth functions

$$\phi(x_1,x_2) = b - b\exp(-x_1(x_1-1)x_2(x_2-1)), \quad \boldsymbol{u}(x_1,x_2) = \begin{pmatrix} \sin(2\pi x_1)\cos(2\pi x_2) \\ -\cos(2\pi x_1)\sin(2\pi x_2) \end{pmatrix},$$

$$\boldsymbol{\sigma}(x_1,x_2) = 2\pi \begin{pmatrix} \dfrac{\cos(2\pi x_1)\cos(2\pi x_2)}{(1-bc+bce^{-x_1(x_1-1)x_2(x_2-1)})^2} & \dfrac{-\sin(2\pi x_1)\sin(2\pi x_2)}{(1-bc+bce^{-x_1(x_1-1)x_2(x_2-1)})^2} \\ \dfrac{\sin(2\pi x_1)\sin(2\pi x_2)}{(1-bc+bce^{-x_1(x_1-1)x_2(x_2-1)})^2} & \dfrac{-\cos(2\pi x_1)\cos(2\pi x_2)}{(1-bc+bce^{-x_1(x_1-1)x_2(x_2-1)})^2} \end{pmatrix} - (x_1^2-x_2^2)\mathbb{I},$$

for $(x_1,x_2) \in \overline{\Omega}$. We take $b = 15, c = m_1 = m_2 = 1/2, m_3 = 3/2$ and set $\Gamma_{\mathrm{D}} = \partial\Omega$, where $\phi$ vanishes and $\boldsymbol{u}_{\mathrm{D}}$ is imposed accordingly to the exact solution. The mean value of $\mathrm{tr}(\boldsymbol{\sigma})_h$ over $\Omega$ is fixed via a penalization strategy. As defined above, the scalar field $\phi$ is bounded in $\Omega$ and so the coefficients are also bounded. In particular we have $\mu_1 = 0.99$ and $\mu_2 = 3.35$. Therefore, and as suggested by Lemma 1.4, the stabilization constants are chosen as $\kappa_1 = \mu_1^2/\mu_2 = 0.2976$, $\kappa_2 = 1/\mu_2 = 0.2985$, and $\kappa_3 = \kappa_1/2 = 0.1488$.

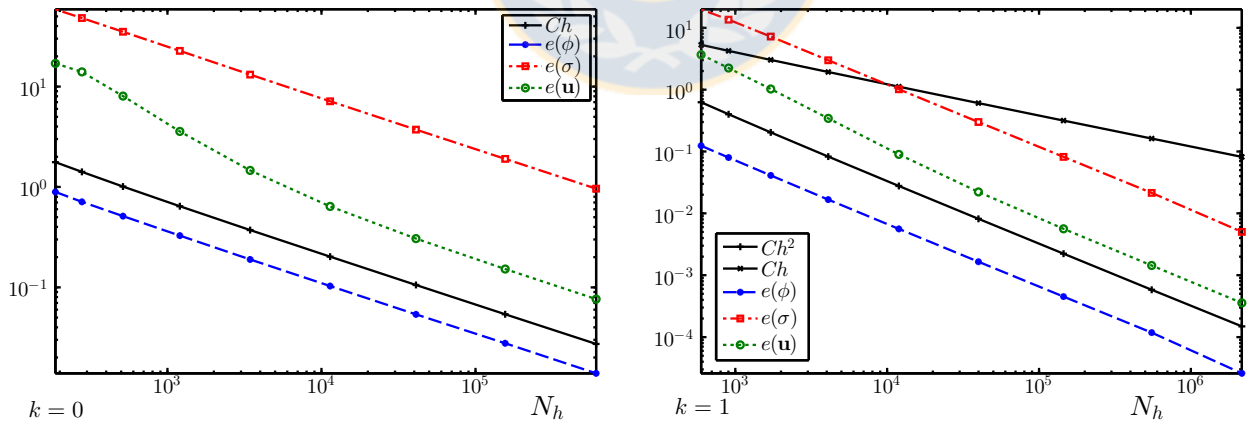


Figure 1.1: Example 1: Computed errors $e(\phi), e(\boldsymbol{\sigma}), e(\boldsymbol{u})$ associated to the mixed–primal approximation versus the number of degrees of freedom $N_h$ for $\mathbf{RT}_0 - \mathbf{P}_1 - \mathrm{P}_1$ and $\mathbf{RT}_1 - \mathbf{P}_2 - \mathrm{P}_2$ finite elements (left and right, respectively). See values in Table 1.1.

The domain is partitioned into quasi-uniform meshes with $2^n + 3$, $n = 0, 1, \ldots, 8$ vertices on each side of the domain. The convergence of the approximate solutions is assessed by computing errors in

| $N_h$ | $h$ | $e(\boldsymbol{\sigma})$ | $r(\boldsymbol{\sigma})$ | $e(\boldsymbol{u})$ | $r(\boldsymbol{u})$ | $e(\phi)$ | $r(\phi)$ | iter |
|---|---|---|---|---|---|---|---|---|
| Augmented $\mathbf{RT}_0 - \mathbf{P}_1 - \mathrm{P}_1$ scheme | | | | | | | | |
| 187 | 0.353553 | 58.80212 | – | 16.97841 | – | 0.891473 | – | 8 |
| 278 | 0.282843 | 48.21425 | 0.852938 | 13.99512 | 1.014962 | 0.711188 | 0.970463 | 7 |
| 514 | 0.202031 | 35.19082 | 0.935794 | 8.041585 | 1.424675 | 0.512189 | 0.975540 | 7 |
| 1202 | 0.128565 | 22.67913 | 0.972039 | 3.573343 | 1.579459 | 0.327347 | 0.990462 | 7 |
| 3442 | 0.074432 | 13.16677 | 0.994888 | 1.461483 | 1.563582 | 0.189813 | 0.997142 | 6 |
| 11378 | 0.040406 | 7.138732 | 1.002043 | 0.639297 | 1.235346 | 0.103089 | 0.999241 | 6 |
| 41074 | 0.021107 | 3.722753 | 1.002661 | 0.305779 | 1.113577 | 0.053859 | 0.999801 | 6 |
| 155762 | 0.010795 | 1.904552 | 1.002240 | 0.152283 | 1.034021 | 0.027705 | 0.999948 | 6 |
| 606322 | 0.005460 | 0.961174 | 1.001041 | 0.076408 | 1.010863 | 0.013933 | 0.999987 | 6 |
| Augmented $\mathbf{RT}_1 - \mathbf{P}_2 - \mathrm{P}_2$ scheme | | | | | | | | |
| 595 | 0.353553 | 19.88141 | – | 3.675443 | – | 0.123752 | – | 7 |
| 903 | 0.282843 | 13.55213 | 1.717465 | 2.237812 | 2.223581 | 0.079988 | 1.955574 | 6 |
| 1711 | 0.202031 | 7.213065 | 1.874291 | 1.026756 | 2.215637 | 0.041028 | 1.984189 | 6 |
| 4095 | 0.128565 | 2.989083 | 1.949025 | 0.343355 | 2.223416 | 0.016689 | 1.990120 | 6 |
| 11935 | 0.074432 | 1.012340 | 1.981522 | 0.089977 | 2.150313 | 0.005607 | 1.995567 | 6 |
| 39903 | 0.040406 | 0.299392 | 1.994287 | 0.022247 | 2.187332 | 0.001654 | 1.998442 | 6 |
| 144991 | 0.021107 | 0.081778 | 1.998531 | 0.005629 | 2.116371 | 0.000451 | 1.999545 | 6 |
| 551775 | 0.010795 | 0.021401 | 1.999468 | 0.001439 | 2.034801 | 0.000118 | 1.999836 | 6 |
| 2164783 | 0.005460 | 0.005014 | 2.006076 | 0.000357 | 2.013878 | 0.000026 | 1.999935 | 6 |

Table 1.1: Example 1: Convergence history and Newton iteration count for the mixed–primal $\mathbf{RT}_k - \mathbf{P}_{k+1} - \mathrm{P}_{k+1}$ approximations of the coupled problem, $k = 0, 1$. Here $N_h$ stands for the number of degrees of freedom associated to each triangulation $\mathcal{T}_h$.

the respective norms and experimental rates, that we define as usual

$$e(\boldsymbol{\sigma}) := \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{\mathbf{div},\Omega}, \quad e(\boldsymbol{u}) := \|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,\Omega}, \quad e(\phi) := \|\phi - \phi_h\|_{1,\Omega},$$
$$r(\boldsymbol{\sigma}) := \frac{\log(e(\boldsymbol{\sigma})/\hat{e}(\boldsymbol{\sigma}))}{\log(h/\hat{h})}, \quad r(\boldsymbol{u}) := \frac{\log(e(\boldsymbol{u})/\hat{e}(\boldsymbol{u}))}{\log(h/\hat{h})}, \quad r(\phi) := \frac{\log(e(\phi)/\hat{e}(\phi))}{\log(h/\hat{h})},$$

where $e$ and $\hat{e}$ denote errors computed on two consecutive meshes of sizes $h$ and $\hat{h}$, respectively. Notice that these errors are computed between the finite element approximation and the corresponding interpolate of the exact solution. Values and plots of errors and corresponding rates associated to $\mathbf{RT}_k - \mathbf{P}_{k+1} - \mathrm{P}_{k+1}$ approximations with $k = 0$ and $k = 1$ are summarized in Table 1.1 and Figure 1.1, respectively, where we observe convergence rates of $O(h^{k+1})$ for stresses, velocities and the scalar field in the relevant norms. These findings are in agreement with the theoretical error bounds of Section 1.5 (*cf.* (1.65)). A Newton-Raphson algorithm with a tolerance of 1E-08 has been applied to the resolution of the nonlinear problem (1.46), and at each iteration the linear systems resulting from the linearization were solved by means of the multifrontal massively parallel solver (MUMPS [9]). We mention that an average number of 7 Newton steps were required to reach the desired tolerance. All remaining examples were carried out using $k = 0$, i.e., lowest-order Raviart-Thomas finite element approximations for the

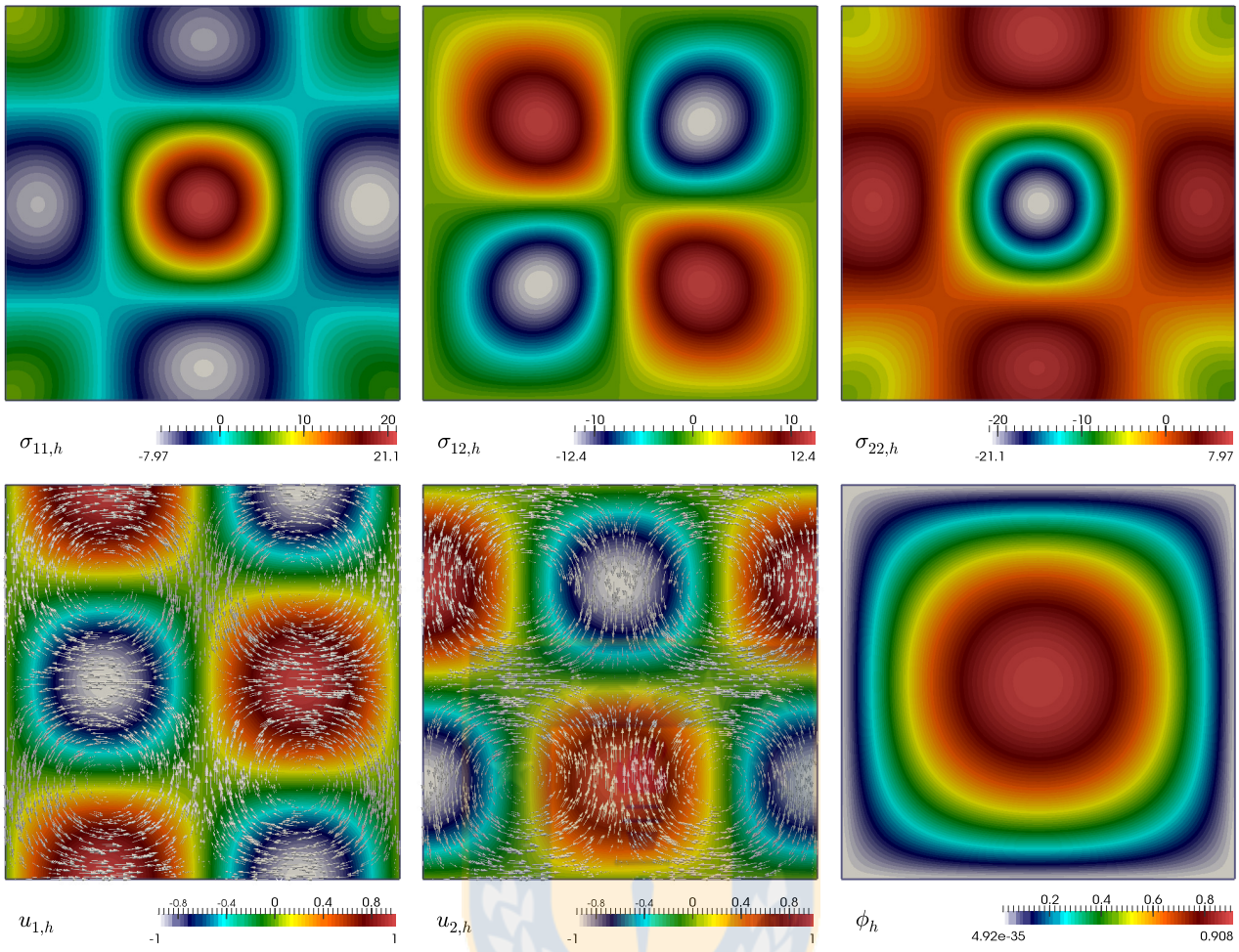


Figure 1.2: Example 1: $\mathbf{RT}_0 - \mathbf{P}_1 - \mathrm{P}_1$ approximation of stress components $\boldsymbol{\sigma}_h$ (top panels), velocity components $\boldsymbol{u}_h$ (with vector directions, bottom left and center, respectively), and scalar field $\phi_h$ (bottom right) solving (1.15). The mesh has 37249 vertices and 74496 triangular elements.

rows of the Cauchy stress tensor, and piecewise linear approximations of velocity components and the scalar field $\phi$. The augmented mixed–primal approximations computed on a mesh of 37249 vertices and 74496 elements are depicted in Figure 1.2.

In our second example we assess the capability of a 3D implementation by carrying out the benchmark test of thermal convection on the cube $\Omega = (0,1)^3$ (see e.g. [58, 84]). The relevant equations, here written in terms of stresses $\boldsymbol{\sigma}$, velocities $\boldsymbol{u}$, and *temperature* $\phi$ correspond to the Boussinesq approximation and can be readily recovered from (1.5) by setting $g = 0$, $\boldsymbol{f}\phi = \frac{1}{\rho}(0, \phi - 1, 0)^{\mathtt{t}}$, $\mu(\phi) = \mathrm{Re}^{-1} = (\mathrm{Ra}/\mathrm{Pr})^{-1/2}$, $\vartheta(\phi) = (\mathrm{Re}\,\mathrm{Pr})^{-1}$, $\gamma(\phi) = 0$, where $\mathrm{Pr} = 0.71$, $\mathrm{Ra} =$1E05, and $\rho = 0.1$ are the Prandtl (ratio between the viscous and thermal diffusions), Rayleigh (only parameter remaining after nondimensionalization of the Boussinesq approximation), and overheat ratio coefficients, respectively. Notice that this problem is linear, except for the convection term. Even if the problem setting does not coincide exactly with the case analyzed previously, our goal is to illustrate the applicability of the present coupling strategy in diverse scenarios. In fact, if we redefine $\boldsymbol{f} := \frac{1}{\rho}(0, 1, 0)^{\mathtt{t}}$, then the functional (1.14) will eventually contain two additional terms independent of $\boldsymbol{f}$, and all the subsequent

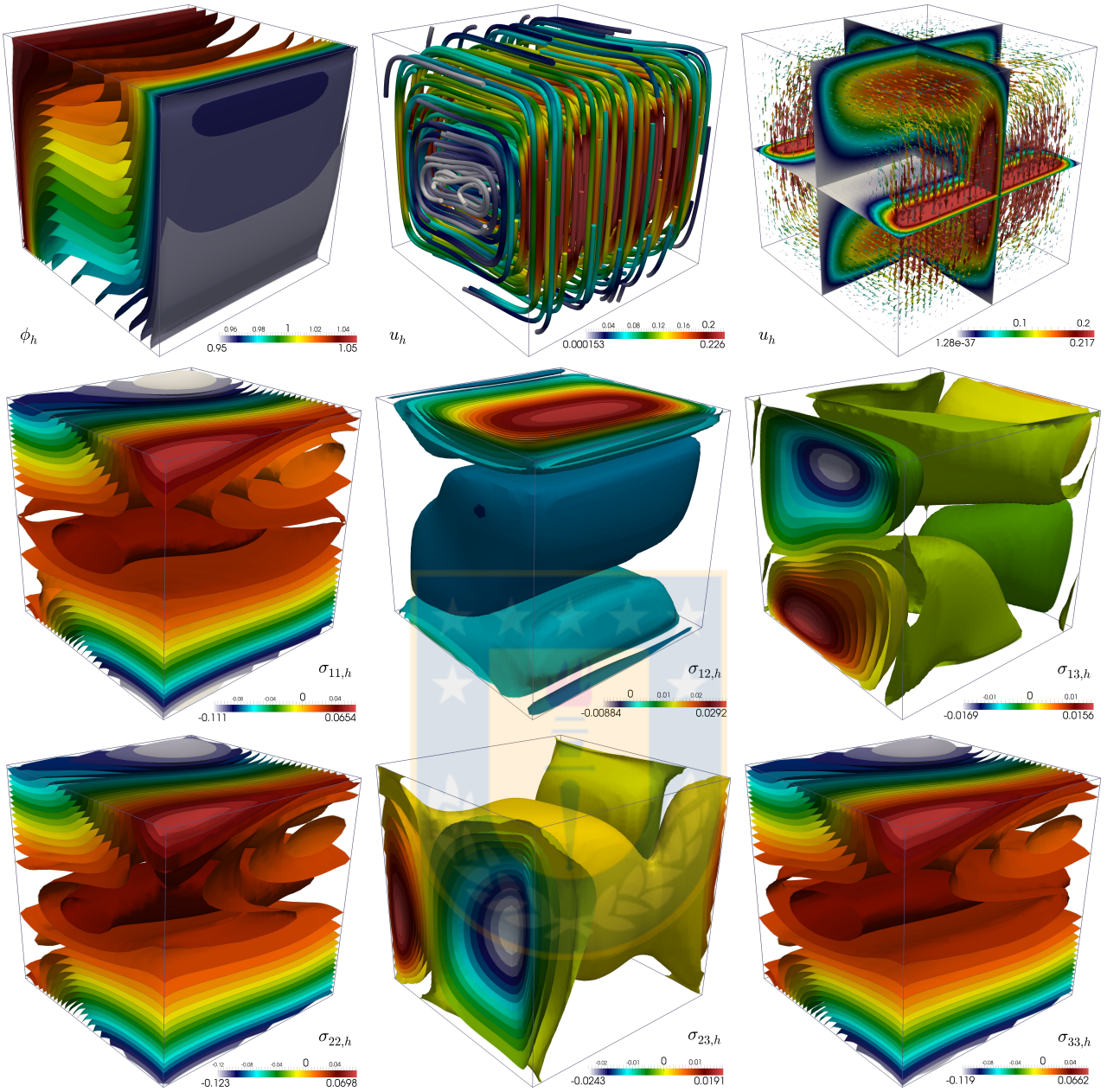


Figure 1.3: Example 2: Computed temperature iso-surfaces (top left) and velocity streamlines and vectors colored by magnitude (top center and right, respectively) and principal components of the Cauchy stress (center and bottom rows) for the thermal cavity test.

continuous and discrete analysis would remain unchanged after replacing $\boldsymbol{f}\phi$ by $\boldsymbol{f}\phi - \boldsymbol{f}$.

The stabilization constants are chosen as $\kappa_1 = \mu$, $\kappa_2 = 1/\mu$, and $\kappa_3 = \mu/2$. As boundary data we impose $\boldsymbol{u}_{\mathrm{D}} = \boldsymbol{0}$ on the whole $\partial\Omega$, whereas we put $\phi = (2-\rho)/2$ at $x_1 = 0$ and $\phi = (2+\rho)/2$ at $x_1 = 1$. On the remainder of $\partial\Omega$ we impose zero-flux conditions for $\phi$, that is $\widetilde{\boldsymbol{\sigma}} \cdot \boldsymbol{\nu} = 0$. The domain is discretized on a mesh $\mathcal{T}_h$ of 46656 vertices and 271950 tetrahedra, and we represent the field quantities of interest in Figure 1.3. From these plots we can observe a satisfactory qualitative agreement with respect to published data (see e.g. [47, 58, 84]).

Moreover, Figure 1.4 reports on the mid-plane ($x_3 = 0.5$) profiles and a comparison with respect to values described in [58], including the average Nusselt number associated to a plane $S$ (at fixed $x_1$) and computed as $\mathrm{Nu} = \int_S \mathrm{Pr}\,\mathrm{Re}\,u_1\phi - \partial_1\phi$. Our findings, after an average of 9 Newton iterations to reach a tolerance of 1E-08, satisfactorily match the benchmark data in terms of maximum and minimum velocities and temperature profiles at the symmetry lines $x_1 = 0.5$ and $x_2 = 0.5$. More quantitative comparisons are also presented in Table 1.2, where we have collected some outputs of interest for different values of the Rayleigh number. For larger Rayleigh numbers, an homotopy (or continuation) method was carried out on the Rayleigh number in order to ensure convergence of the algorithm.

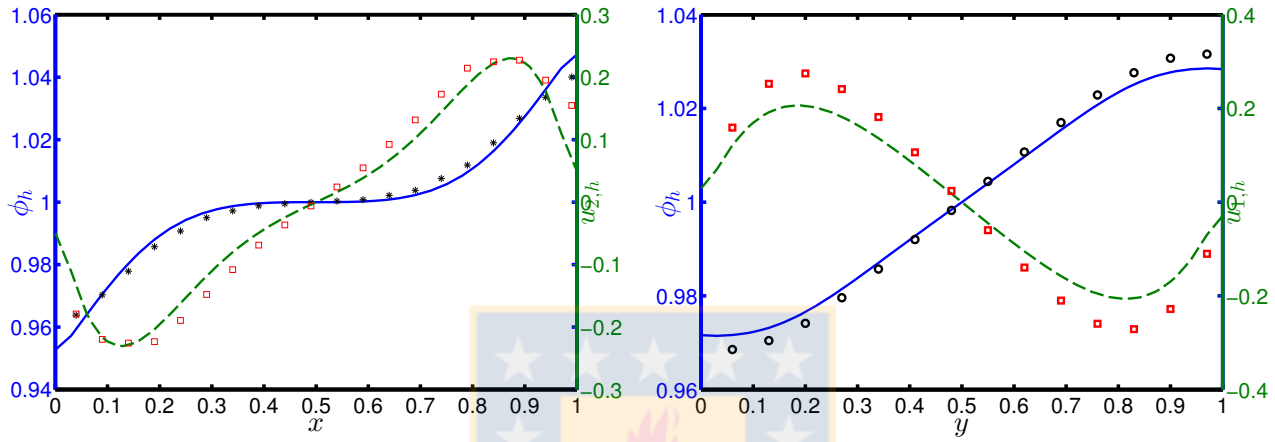


Figure 1.4: Example 2: Temperature profiles (solid blue, left axis) and velocity components (dashed green, right axis) at $x_3 = 0.5$, and comparison with respect to benchmark solutions.

| | Ra | Nu | $\max(\lvert\hat{u}_{1,h}\rvert)$ | $\max(\lvert\hat{u}_{2,h}\rvert)$ | $x_1^\infty$ | $x_2^\infty$ |
|---|---|---|---|---|---|---|
| Computed | $10^3$ | 1.134 | 0.129 | 0.131 | 0.176 | 0.845 |
| [47] | $10^3$ | 1.117 | 0.136 | 0.138 | 0.178 | 0.813 |
| [58] | $10^3$ | – | 0.132 | 0.131 | 0.200 | 0.833 |
| Computed | $10^4$ | 2.030 | 0.195 | 0.229 | 0.121 | 0.819 |
| [47] | $10^4$ | 2.054 | 0.192 | 0.234 | 0.119 | 0.823 |
| [58] | $10^4$ | 2.100 | 0.201 | 0.225 | 0.117 | 0.817 |
| Computed | $10^5$ | 4.321 | 0.145 | 0.244 | 0.064 | 0.843 |
| [47] | $10^5$ | 4.337 | 0.153 | 0.261 | 0.066 | 0.855 |
| [58] | $10^5$ | 4.361 | 0.147 | 0.247 | 0.065 | 0.855 |

Table 1.2: Example 2: Outputs of interest (Nusselt number, maximum value of the normalized horizontal velocity on the mid-plane attained at $(0.5, x_2^\infty, 0.5)$, and maximum value of the normalized vertical velocity and its position $(x_1^\infty, 0.5, 0.5)$ on the central horizontal plane, respectively) for different values of the Rayleigh number, and comparison with respect to values from [47, 58].

Our last example focuses on the simulation of the steady state of a clarifying-thickening process. The basin, the different boundaries of the geometry, and the generated volumetric mesh consisting of 64135 vertices and 370597 tetrahedra are sketched in Figure 1.5. The size of the mesh and the finite

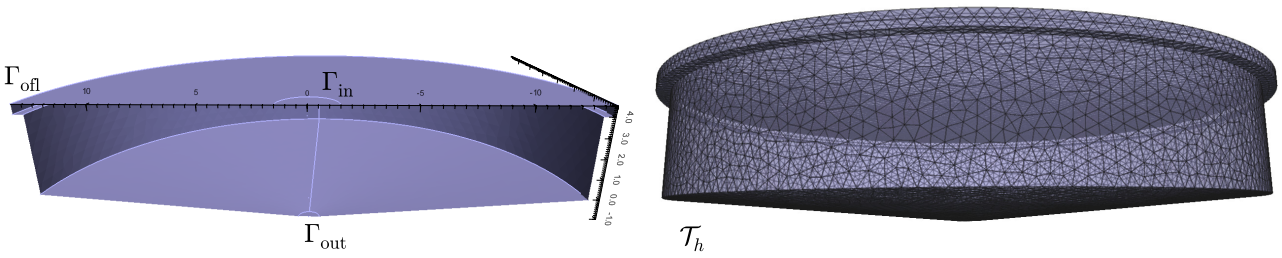


Figure 1.5: Example 3: Geometry of the clarifier-thickener unit (left panel) and tetrahedral mesh $\mathcal{T}_h$ with 64135 vertices and 370597 elements (right panel).

element choice (row-wise Raviart-Thomas approximations for stresses and piecewise linear elements for velocity components and concentration) implies that at each Newton step we solve for a total of 2515211 degrees of freedom. The nonlinear functions of the concentration are taken as in [32]: $\mu(\phi) = (1 - \phi/\phi_{\max})^{-2.5}$, $\gamma(\phi) = u_\infty(1 + \phi(1 - \phi/\phi_{\max}))^2$, $\vartheta(\phi) = \frac{\gamma(\phi)\sigma_0\alpha(\phi/\phi_c)^{\alpha-1}}{\phi\phi_c G\Delta\rho} + u_\infty$ and the source terms are $\boldsymbol{f} = (0, 0, -G)^{\mathtt{t}}$, $g = 0$. The physical values assumed by the concentration (it remains bounded between 0 and $\phi_{\max}$) imply that the viscosity, hindered flux, and compressibility coefficients satisfy (1.2)-(1.3) with $\mu_1 = 1, \mu_2 = 2.7, \gamma_1 = u_\infty, \gamma_2 = 1.15u_\infty, \vartheta_1 = 4.28, \vartheta_2 = 29.74$. However, notice that $\vartheta$ depends explicitly on $\phi$ and not on the concentration gradient, which was not addressed in the solvability analysis of the model problem. While one could try to analyze this case by using some classical results on pseudomonotone operators (see, e.g. [41, Section 9.3], [88, Section 3.3]), in the forthcoming work [5] we have chosen to extend the present approach to this modified model since in this way we are able to derive not only the existence of continuous and discrete solutions but also the corresponding *a priori* error analysis.

Boundary conditions are set as follows: Concentration and velocities are fixed on the inlet disc $\Gamma_{\text{in}}$ according to $\phi = \phi_{\text{in}}$ and $\boldsymbol{u} = \boldsymbol{u}_{\text{in}} = (0, 0, -u_{3,\text{in}})^{\mathtt{t}}$. At the outlet disk $\Gamma_{\text{out}}$ we prescribe $\boldsymbol{u} = \boldsymbol{u}_{\text{out}} = (0, 0, -u_{3,\text{out}})^{\mathtt{t}}$, at the overflow annulus we do not constraint the velocity field, and on the remainder of $\partial\Omega$ we put no slip boundary data for the velocity and zero-flux conditions for the concentration. Model parameters are set as $u_{3,\text{in}}$ =1.29E-02, $u_{3,\text{out}}$ =2.54E-03, $\Delta\rho = 1562$, $\phi_{\max} = 0.9$, $\phi_c = 0.1$, $u_\infty$ =2.2E-03, $G = 9.81$, $\phi_{\text{in}} = 0.08$, $\alpha = 5$, and $\sigma_0$ =5E-02. We mimic the behavior of a transient simulation by adding a mass term $\eta\phi$ to the concentration equation, with $\eta$ =1E-03. Such a modification does not entail a major change in the analysis: it suffices to replace the part of the flux $\phi\boldsymbol{u}$ by $\phi(\boldsymbol{u} + \eta)$.

According to the bounds of the viscosity, the stabilization parameters were set as $\kappa_1 = \kappa_2 = 0.4784$, and $\kappa_3 = 0.2392$. We mention that 8 Newton iterations were needed to achieve a tolerance of 1E-07 for the energy norm of the incremental approximations. The numerical results are depicted in Figure 1.6 (we show half of the tank for visualization purposes), including concentration profile, velocity vectors, pressure approximation (computed in terms of the trace of the Cauchy stress), and velocity components. We can observe that the material is removed from the unit at the boundary $\Gamma_{\text{out}}$ with concentration $\phi \approx 0.24$, which agrees with the results in e.g. [31, Example 3].

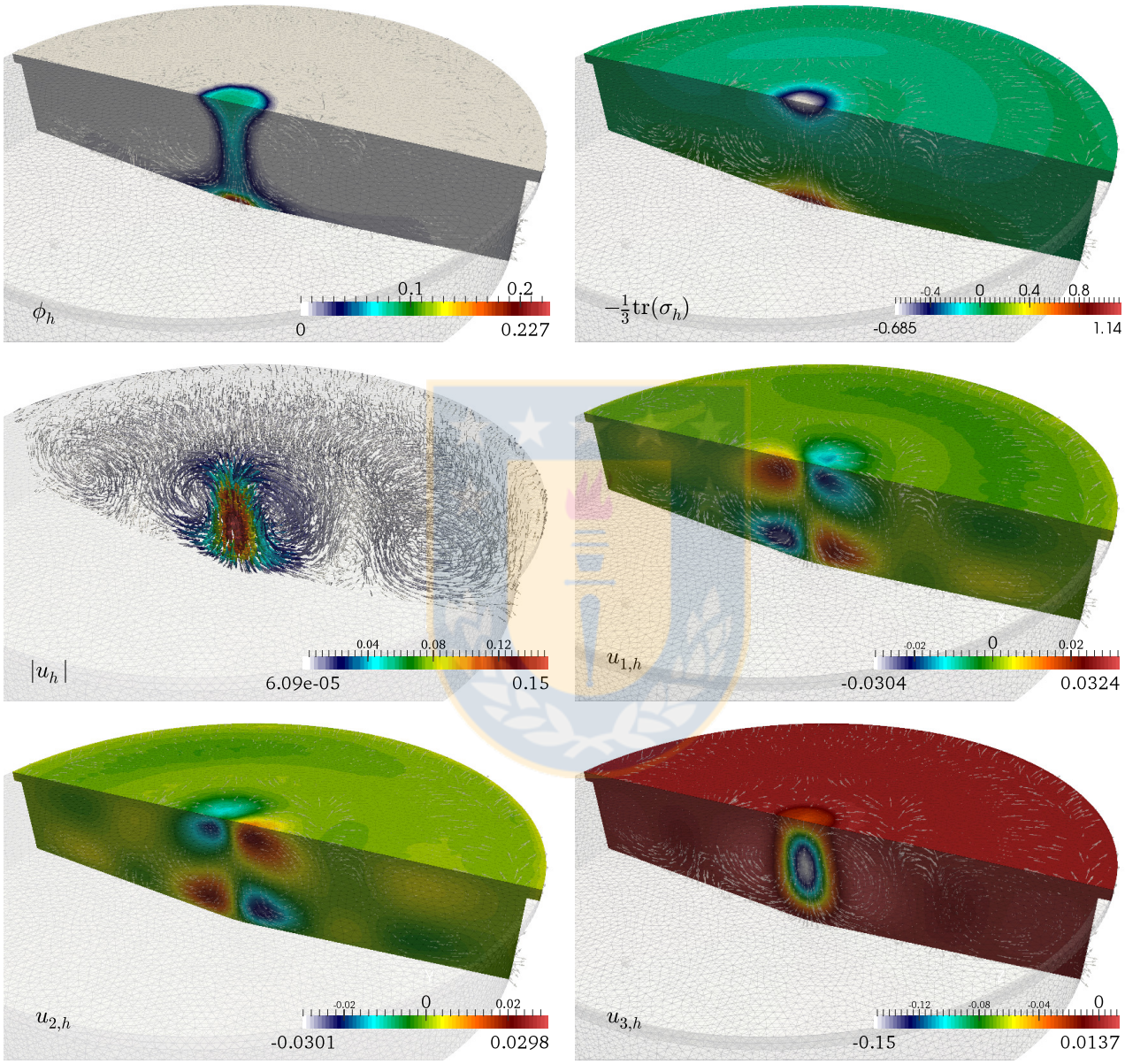


Figure 1.6: Example 3: Simulation of a clarifier-thickener unit. From left-top: Approximated concentration profile, opposite of the trace of the Cauchy stress tensor (which corresponds to the suggested approximation of the pressure field), and velocity components.

# CHAPTER 2

## A posteriori error analysis for a viscous flow–transport problem

### 2.1 Introduction

**Chapter 1** was concerned with the solvability of a three-field flow-transport problem given by the coupling of a scalar nonlinear convection-diffusion problem with the Stokes equations where the viscosity depends on the distribution of the solution to the transport problem. There, an augmented mixed–primal variational formulation was proposed, where the Cauchy stresses are sought in $\mathbb{H}(\mathbf{div};\Omega)$, the velocity is in $\mathbf{H}^1(\Omega)$, and the solution to the transport problem has $\mathrm{H}^1(\Omega)$ regularity. The associated numerical scheme employed Raviart-Thomas spaces of order $k$ for the Cauchy stress, whereas the velocity and a coupled scalar field (e.g. concentration as in [5], or temperature) were approximated with continuous piecewise polynomials of degree $\leq k+1$. Optimal *a priori* error estimates were also derived.

Our goal in this chapter is to propose reliable and efficient residual-based *a posteriori* error estimators for the coupled flow–transport problem studied in [4]. Estimators of this kind are typically used to guide adaptive mesh refinement in order to guarantee an adequate convergence behavior of the Galerkin approximations, even under the eventual presence of singularities. The global estimator $\boldsymbol{\theta}$ depends on local estimators $\theta_T$ defined on each element $T$ of a given mesh $\mathcal{T}_h$. Then, $\boldsymbol{\theta}$ is said to be efficient (resp. reliable) if there exists a constant $C_{\mathtt{eff}} > 0$ (resp. $C_{\mathtt{rel}} > 0$), independent of meshsizes, such that

$$C_{\mathtt{eff}}\,\boldsymbol{\theta} \,+\, \text{h.o.t.} \leq \|error\| \leq C_{\mathtt{rel}}\,\boldsymbol{\theta} \,+\, \text{h.o.t.}\,,$$

where h.o.t. is a generic expression denoting one or several terms of higher order. A number of *a posteriori* error estimators specifically targeted for non-viscous (e.g., Darcy) flow coupled with transport problems area available in the literature (see, e.g. [17, 48, 82, 99, 104]). However, only a couple of contributions deal with *a posteriori* error analysis for coupled viscous flow-transport problems. In particular, we mention the reactive flow equations studied in [24] and the adaptive finite element method for heat transfer in incompressible fluid flow proposed in [83], which is based on dual weighted residual error estimation.

In contrast, here we apply a Helmholtz decomposition, local approximation properties of the Clément interpolant and Raviart-Thomas operator, and known estimates from [15, 60, 66, 72, 73], to prove the reliability of a residual-based estimator. Then, inverse inequalities, the localization technique based on triangle-bubble and edge-bubble functions imply the efficiency of the estimator. An alternative (also reliable and efficient) residual-based *a posteriori* error estimator is proposed, where

the Helmholtz decomposition is not employed in the corresponding proof of reliability. The remainder of this chapter is structured as follows. In Section 2.2, we first recall from [4] the model problem and a corresponding augmented mixed-primal formulation as well as the associated Galerkin scheme. In Section 2.3, we derive a reliable and efficient residual-based *a posteriori* error estimator for our Galerkin scheme. A second estimator is introduced and studied in Section 2.4. Finally, in Section 2.5 we provide some numerical results confirming the reliability and efficiency of the estimators, and illustrating the good performance of the associated adaptive algorithm for the augmented mixed-primal finite element method.

## 2.2 A coupled viscous flow–transport problem

### 2.2.1 The three-field formulation

The following system of partial differential equations describes the stationary state of the transport of species $\phi$ in an immiscible fluid occupying the domain $\Omega$ (*cf.* [4]):

$$\begin{aligned} \frac{1}{\mu(\phi)}\,\boldsymbol{\sigma}^{\mathrm{d}} &= \nabla \boldsymbol{u} \quad \text{in} \quad \Omega\,, \qquad -\mathbf{div}\boldsymbol{\sigma} = \boldsymbol{f}\phi \quad \text{in} \quad \Omega\,, \\ \widetilde{\boldsymbol{\sigma}} &= \vartheta(|\nabla\phi|)\,\nabla\phi - \phi\boldsymbol{u} - \gamma(\phi)\boldsymbol{k} \quad \text{in} \quad \Omega\,, \qquad -\mathrm{div}\widetilde{\boldsymbol{\sigma}} = g \quad \text{in} \quad \Omega\,, \end{aligned} \tag{2.1}$$

$$\begin{aligned} \boldsymbol{u} &= \boldsymbol{u}_{\mathrm{D}} \quad \text{on} \quad \Gamma_{\mathrm{D}}\,, \qquad \boldsymbol{\sigma}\boldsymbol{\nu} = \mathbf{0} \quad \text{on} \quad \Gamma_{\mathrm{N}}\,, \\ \phi &= \phi_{\mathrm{D}} \quad \text{on} \quad \Gamma_{\mathrm{D}}\,, \quad \text{and} \quad \widetilde{\boldsymbol{\sigma}}\cdot\boldsymbol{\nu} = 0 \quad \text{on} \quad \Gamma_{\mathrm{N}}\,. \end{aligned} \tag{2.2}$$

where the sought quantities are the Cauchy fluid stress $\boldsymbol{\sigma}$, the local volume-average velocity of the fluid $\boldsymbol{u}$, and the local concentration of species $\phi$. In this model, the kinematic effective viscosity, $\mu$; the diffusion coefficient, $\vartheta$; and the one-directional flux function describing hindered settling, $\gamma$; depend nonlinearly on $\phi$. In turn, $\boldsymbol{k}$ is a vector pointing in the direction of gravity and $\boldsymbol{f} \in \mathbf{L}^\infty(\Omega)$, $\boldsymbol{u}_{\mathrm{D}} \in \mathbf{H}^{1/2}(\Gamma_{\mathrm{D}})$, $g \in \mathrm{L}^2(\Omega)$ are given functions. For sake of the subsequent analysis, the Dirichlet datum for the concentration will be assumed homogeneous, that is $\phi_{\mathrm{D}} = 0$, $\vartheta$ is assumed of class $C^1$, and we suppose that there exist positive constants $\mu_1$, $\mu_2$, $\gamma_1$, $\gamma_2$, $\vartheta_1$, $\vartheta_2$, $L_\mu$ and $L_\gamma$, such that

$$\mu_1 \leq \mu(s) \leq \mu_2 \quad \text{and} \quad \gamma_1 \leq \gamma(s) \leq \gamma_2 \qquad \forall\, s \in \mathrm{R}\,, \tag{2.3}$$

$$\vartheta_1 \leq \vartheta(s) \leq \vartheta_2 \quad \text{and} \quad \vartheta_1 \leq \vartheta(s) + s\,\vartheta'(s) \leq \vartheta_2 \qquad \forall\, s \geq 0, \tag{2.4}$$

$$|\mu(s) - \mu(t)| \leq L_\mu\,|s-t| \qquad \forall\, s,\, t \in \mathrm{R}\,, \tag{2.5}$$

$$|\gamma(s) - \gamma(t)| \leq L_\gamma\,|s-t| \qquad \forall\, s,\, t \in \mathrm{R}\,. \tag{2.6}$$

### 2.2.2 The augmented mixed–primal formulation

The homogeneous Neumann boundary condition for $\boldsymbol{\sigma}$ on $\Gamma_{\mathrm{N}}$ and the Dirichlet datum for $\phi$ (*cf.* second and third relations of (2.2), respectively) suggest the introduction of the spaces

$$\begin{aligned} \mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega) &:= \Big\{ \boldsymbol{\tau} \in \mathbb{H}(\mathbf{div};\Omega) : \quad \boldsymbol{\tau}\boldsymbol{\nu} = \mathbf{0} \quad \text{on} \quad \Gamma_{\mathrm{N}} \Big\}, \\ \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) &:= \Big\{ \psi \in \mathrm{H}^1(\Omega) : \quad \psi = 0 \quad \text{on} \quad \Gamma_{\mathrm{D}} \Big\}. \end{aligned}$$

Also, due to the generalized Poincaré inequality, there exists $c_p > 0$, depending only on $\Omega$ and $\Gamma_{\rm D}$, such that

$$\|\psi\|_{1,\Omega} \leq c_p\, |\psi|_{1,\Omega} \qquad \forall\, \psi \in \mathrm{H}^1_{\Gamma_{\rm D}}(\Omega)\,. \tag{2.7}$$

An augmented mixed-primal formulation for our original coupled problem (2.1) reads as follows: Find $(\boldsymbol{\sigma}, \boldsymbol{u}, \phi) \in \mathbb{H}_{\rm N}(\mathbf{div};\Omega) \times \mathbf{H}^1(\Omega) \times \mathrm{H}^1_{\Gamma_{\rm D}}(\Omega)$ such that

$$\begin{aligned} B_\phi((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau},\boldsymbol{v})) &= F_\phi(\boldsymbol{\tau},\boldsymbol{v}) \quad \forall (\boldsymbol{\tau},\boldsymbol{v}) \in \mathbb{H}_{\rm N}(\mathbf{div};\Omega) \times \mathbf{H}^1(\Omega)\,, \\ A_{\boldsymbol{u}}(\phi,\psi) &= G_\phi(\psi) \qquad \forall\, \psi \in \mathrm{H}^1_{\Gamma_{\rm D}}(\Omega) \end{aligned} \tag{2.8}$$

where

$$\begin{aligned} B_\phi((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau},\boldsymbol{v})) :=& \int_\Omega \frac{1}{\mu(\phi)}\, \boldsymbol{\sigma}^{\rm d} : \boldsymbol{\tau}^{\rm d} + \int_\Omega \boldsymbol{u}\cdot\mathbf{div}\boldsymbol{\tau} - \int_\Omega \boldsymbol{v}\cdot\mathbf{div}\boldsymbol{\sigma} \\ &+ \kappa_1 \int_\Omega \left(\nabla \boldsymbol{u} - \frac{1}{\mu(\phi)}\,\boldsymbol{\sigma}^{\rm d}\right) : \nabla \boldsymbol{v} + \kappa_2 \int_\Omega \mathbf{div}\boldsymbol{\sigma}\cdot\mathbf{div}\boldsymbol{\tau} + \kappa_3 \int_{\Gamma_{\rm D}} \boldsymbol{u}\cdot\boldsymbol{v}\,, \end{aligned} \tag{2.9}$$

$$F_\phi(\boldsymbol{\tau},\boldsymbol{v}) := \langle \boldsymbol{\tau}\boldsymbol{\nu}, \boldsymbol{u}_{\rm D}\rangle_{\Gamma_{\rm D}} + \int_\Omega \boldsymbol{f}\phi\cdot\boldsymbol{v} - \kappa_2 \int_\Omega \boldsymbol{f}\phi\cdot\mathbf{div}\boldsymbol{\tau} + \kappa_3 \int_{\Gamma_{\rm D}} \boldsymbol{u}_{\rm D}\cdot\boldsymbol{v}\,, \tag{2.10}$$

$$A_{\boldsymbol{u}}(\phi,\psi) := \int_\Omega \vartheta(|\nabla\phi|)\,\nabla\phi\cdot\nabla\psi - \int_\Omega \phi\,\boldsymbol{u}\cdot\nabla\psi \qquad \forall\, \phi,\, \psi \in \mathrm{H}^1_{\Gamma_{\rm D}}(\Omega), \tag{2.11}$$

$$G_\phi(\psi) := \int_\Omega \gamma(\phi)\,\boldsymbol{k}\cdot\nabla\psi + \int_\Omega g\psi \qquad \forall\, \psi \in \mathrm{H}^1_{\Gamma_{\rm D}}(\Omega)\,,$$

where $\kappa_i$, $i \in \{1,2,3\}$, are the stabilization parameters specified in [4, Lemma 4.1]. Further details yielding the weak formulation (2.8) can be found in [4, Section 3.1], whereas its solvability follows from the fixed point strategy developed in [4, Theorem 3.13].

### 2.2.3 The augmented mixed–primal finite element method

We denote by $\mathcal{T}_h$ a regular partition of $\Omega$ into triangles $T$ (resp. tetrahedra $T$ in $\mathrm{R}^3$) of diameter $h_T$, and meshsize $h := \max\left\{h_T : T \in \mathcal{T}_h\right\}$. In addition, given an integer $k \geq 0$, the space $\mathrm{P}_k(T)$ contains polynomial functions on $T$ of degree $\leq k$, and we define the corresponding local Raviart-Thomas space of order $k$ as $\mathrm{RT}_k(T) := \mathbf{P}_k(T) \oplus \mathrm{P}_k(T)\,\boldsymbol{x}$, where, according to the notations described in Section 2.2, $\mathbf{P}_k(T) = [\mathrm{P}_k(T)]^n$, and $\boldsymbol{x} \in \mathrm{R}^n$. Then, the Galerkin scheme associated to (2.8) is as follows: Find $(\boldsymbol{\sigma}_h, \boldsymbol{u}_h, \phi_h) \in \mathbb{H}^{\boldsymbol{\sigma}}_h \times \mathbf{H}^{\boldsymbol{u}}_h \times \mathrm{H}^{\phi}_h$ such that

$$\begin{aligned} B_{\phi_h}((\boldsymbol{\sigma}_h,\boldsymbol{u}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h)) &= F_{\phi_h}(\boldsymbol{\tau}_h,\boldsymbol{v}_h) \quad \forall (\boldsymbol{\tau}_h,\boldsymbol{v}_h) \in \mathbb{H}^{\boldsymbol{\sigma}}_h \times \mathbf{H}^{\boldsymbol{u}}_h\,, \\ A_{\boldsymbol{u}_h}(\phi_h,\psi_h) &= \int_\Omega \gamma(\phi_h)\,\boldsymbol{k}\cdot\nabla\psi_h + \int_\Omega g\psi_h \qquad \forall\, \psi_h \in \mathrm{H}^{\phi}_h\,, \end{aligned} \tag{2.12}$$

where the involved finite element spaces are defined as

$$\begin{aligned} \mathbb{H}^{\boldsymbol{\sigma}}_h &:= \left\{\boldsymbol{\tau}_h \in \mathbb{H}_{\rm N}(\mathbf{div};\Omega) : \quad \mathbf{c}^{\rm t}\,\boldsymbol{\tau}_h|_T \in \mathrm{RT}_k(T) \quad \forall\, \mathbf{c} \in \mathrm{R}^n\,, \quad \forall\, T \in \mathcal{T}_h\right\}, \\ \mathbf{H}^{\boldsymbol{u}}_h &:= \left\{\boldsymbol{v}_h \in \mathbf{C}(\overline{\Omega}) : \quad \boldsymbol{v}_h|_T \in \mathbf{P}_{k+1}(T) \quad \forall\, T \in \mathcal{T}_h\right\}, \\ \mathrm{H}^{\phi}_h &:= \left\{\psi_h \in C(\overline{\Omega}) \cap \mathrm{H}^1_{\Gamma_{\rm D}}(\Omega) : \quad \psi_h|_T \in \mathrm{P}_{k+1}(T) \quad \forall\, T \in \mathcal{T}_h\right\}. \end{aligned}$$

The solvability analysis and *a priori* error bounds for (2.12) are established in [4, Section 5].

## 2.3 A residual-based a posteriori error estimator

In this section we derive a reliable and efficient residual-based *a posteriori* error estimator for the Galerkin scheme (2.12). The analysis will be restricted to the two-dimensional case, with the discrete spaces introduced in Section 2.2.3. However, we point out that a straightforward extension of our analysis to 3D does also apply since the key estimate given by the stability of the corresponding Helmholtz decomposition (see (2.48) below for our 2D case) follows from the technique suggested in [67, Lemma 4.3] and the results provided in [10, Theorems 2.17 and 3.12, and Corollary 3.16].

Now, given a suitable chosen $r > 0$ (see [4] for details), we define the balls

$$W := \{\phi \in \mathrm{H}_{\Gamma_{\mathrm{D}}} : \quad \|\phi\|_{1,\Omega} \leq r\} \quad \text{and} \quad W_h := \{\phi_h \in \mathrm{H}_h^{\phi} : \quad \|\phi_h\|_{1,\Omega} \leq r\}, \tag{2.13}$$

and throughout the rest of the chapter we let $(\boldsymbol{\sigma}, \boldsymbol{u}, \phi) \in \mathbb{H}_{\mathrm{N}}(\mathbf{div}; \Omega) \times \mathbf{H}^1(\Omega) \times \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ with $\phi \in W$ and $(\boldsymbol{\sigma}_h, \boldsymbol{u}_h, \phi_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}} \times \mathrm{H}_h^{\phi}$ with $\phi_h \in W_h$ be the solutions of the continuous and discrete formulations (2.8) and (2.12), respectively. In addition, we set

$$H := \mathbb{H}_{\mathrm{N}}(\mathbf{div}, \Omega) \times \mathbf{H}^1(\Omega), \qquad \|(\boldsymbol{\tau}, \boldsymbol{v})\|_H := \|\boldsymbol{\tau}\|_{\mathbf{div};\Omega} + \|\boldsymbol{v}\|_{1,\Omega} \quad \forall\, (\boldsymbol{\tau}, \boldsymbol{v}) \in H,$$

and recall from [4, Theorems 3.13 and 4.7] that the following *a priori* estimates hold

$$\begin{aligned} \|(\boldsymbol{\sigma}, \boldsymbol{u})\|_H &\leq C_S \left\{ \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + \|\phi\|_{1,\Omega} \|\boldsymbol{f}\|_{\infty,\Omega} \right\}, \\ \|(\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\|_H &\leq C_S \left\{ \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + \|\phi_h\|_{1,\Omega} \|\boldsymbol{f}\|_{\infty,\Omega} \right\}, \end{aligned} \tag{2.14}$$

where $C_S$ is a positive constant independent of $\phi$ and $\phi_h$.

### 2.3.1 The local error indicator

Given $T \in \mathcal{T}_h$, we let $\mathcal{E}_h(T)$ be the set of its edges, and let $\mathcal{E}_h$ be the set of all edges of the triangulation $\mathcal{T}_h$. Then we write $\mathcal{E}_h = \mathcal{E}_h(\Omega) \cup \mathcal{E}_h(\Gamma_{\mathrm{D}}) \cup \mathcal{E}_h(\Gamma_{\mathrm{N}})$, where $\mathcal{E}_h(\Omega) := \{e \in \mathcal{E}_h : e \subseteq \Omega\}$, $\mathcal{E}_h(\Gamma_{\mathrm{D}}) := \{e \in \mathcal{E}_h : e \subseteq \Gamma_{\mathrm{D}}\}$ and $\mathcal{E}_h(\Gamma_{\mathrm{N}}) := \{e \in \mathcal{E}_h : e \subseteq \Gamma_{\mathrm{N}}\}$. Also, for each edge $e \in \mathcal{E}_h$ we fix a unit normal vector $\boldsymbol{\nu}_e := (\nu_1, \nu_2)^{\mathtt{t}}$, and let $\boldsymbol{s}_e := (-\nu_2, \nu_1)^{\mathtt{t}}$ be the corresponding fixed unit tangential vector along $e$. Then, given $e \in \mathcal{E}_h(\Omega)$ and $\tau \in \mathbf{L}^2(\Omega)$ such that $\tau|_T \in [C(T)]^2$ on each $T \in \mathcal{T}_h$, we let $[\![\tau \cdot \boldsymbol{\nu}_e]\!]$ be the corresponding jump across $e$, that is, $[\![\tau \cdot \boldsymbol{\nu}_e]\!] := (\tau|_T - \tau|_{T'})|_e \cdot \boldsymbol{\nu}_e$, where $T$ and $T'$ are the triangles of $\mathcal{T}_h$ having $e$ as a common edge. Similarly, given $\boldsymbol{\tau} \in \mathbb{L}^2(\Omega)$ such that $\boldsymbol{\tau}|_T \in [C(T)]^{2\times 2}$ on each $T \in \mathcal{T}_h$, we let $[\![\boldsymbol{\tau}\boldsymbol{s}_e]\!]$ be the corresponding jump across $e$, that is, $[\![\boldsymbol{\tau}\boldsymbol{s}_e]\!] := (\boldsymbol{\tau}|_T - \boldsymbol{\tau}|_{T'})|_e\, \boldsymbol{s}_e$. If no confusion arises, we will simple write $\boldsymbol{s}$ and $\boldsymbol{\nu}$ instead of $\boldsymbol{s}_e$ and $\boldsymbol{\nu}_e$, respectively. The curl operator applied to scalar, vector and tensor valued fields $v$, $\boldsymbol{\varphi} := (\varphi_1, \varphi_2)$ and $\boldsymbol{\tau} := (\tau_{ij})$, respectively, will be denoted as

$$\mathbf{curl}(v) := \begin{pmatrix} \frac{\partial v}{\partial x_2} \\ -\frac{\partial v}{\partial x_1} \end{pmatrix}, \quad \underline{\mathbf{curl}}(\boldsymbol{\varphi}) := \begin{pmatrix} \mathbf{curl}(\varphi_1)^{\mathtt{t}} \\ \mathbf{curl}(\varphi_2)^{\mathtt{t}} \end{pmatrix}, \quad \text{and} \quad \mathbf{curl}(\boldsymbol{\tau}) := \begin{pmatrix} \frac{\partial \tau_{12}}{\partial x_1} - \frac{\partial \tau_{11}}{\partial x_2} \\ \frac{\partial \tau_{22}}{\partial x_1} - \frac{\partial \tau_{21}}{\partial x_2} \end{pmatrix}.$$

Then, we let $\widetilde{\boldsymbol{\sigma}}_h := \vartheta(|\nabla\phi_h|)\nabla\phi_h - \phi_h\boldsymbol{u}_h - \gamma(\phi_h)\boldsymbol{k}$ and define for each $T \in \mathcal{T}_h$ a local error indicator $\theta_T$ as follows

$$\begin{aligned}\theta_T^2 :=& \|\boldsymbol{f}\phi_h + \mathbf{div}\boldsymbol{\sigma}_h\|_{0,T}^2 + \left\|\nabla\boldsymbol{u}_h - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\|_{0,T}^2 + h_T^2\,\|g + \mathrm{div}\widetilde{\boldsymbol{\sigma}}_h\|_{0,T}^2 \\ &+ h_T^2\left\|\mathbf{curl}\left\{\frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\}\right\|_{0,T}^2 + \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Omega)} h_e\left\|\left[\!\left[\frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\,\boldsymbol{s}\right]\!\right]\right\|_{0,e}^2 \\ &+ \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Omega)} h_e\,\|[\![\widetilde{\boldsymbol{\sigma}}_h\cdot\boldsymbol{\nu}_e]\!]\|_{0,e}^2 + \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{N}})} h_e\,\|\widetilde{\boldsymbol{\sigma}}_h\cdot\boldsymbol{\nu}\|_{0,e}^2 \\ &+ \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{D}})} \|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|_{0,e}^2 + \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{D}})} h_e\left\|\frac{d\boldsymbol{u}_{\mathrm{D}}}{d\boldsymbol{s}} - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\,\boldsymbol{s}\right\|_{0,e}^2 .\end{aligned} \tag{2.15}$$

The residual character of each term defining $\theta_T^2$ is clear, and hence, proceeding as usual, a *global* residual error estimator can be defined as

$$\boldsymbol{\theta} := \left\{\sum_{T\in\mathcal{T}_h}\theta_T^2\right\}^{1/2} . \tag{2.16}$$

Note that the last term defining $\theta_T^2$ requires that $\left.\frac{d\boldsymbol{u}_{\mathrm{D}}}{d\boldsymbol{s}}\right|_e \in \mathbf{L}^2(e)$ for each $e \in \mathcal{E}_h(\Gamma_{\mathrm{D}})$. This is ensured below by assuming that $\boldsymbol{u}_{\mathrm{D}} \in \mathbf{H}_0^1(\Gamma_{\mathrm{D}})$.

### 2.3.2 Reliability

The following theorem constitutes the main result of this section

**Theorem 2.1.** *Assume that $\Omega$ is a connected domain and that $\Gamma_{\mathrm{N}}$ is the boundary of a convex part of $\Omega$, that is $\Omega$ can be extended to a convex domain $B$ such that $\overline{\Omega} \subseteq B$ and $\Gamma_{\mathrm{N}} \subseteq \partial B$ (see Figure 2.1 below). In addition, assume that $\boldsymbol{u}_{\mathrm{D}} \in \mathbf{H}_0^1(\Gamma_{\mathrm{D}})$ and that for some $\varepsilon \in (0,1)$ (when $n = 2$) or some $\varepsilon \in (1/2, 1)$ (when $n = 3$) there holds*

$$C_3\,|\boldsymbol{k}| + C_6\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma_{\mathrm{D}}} + C_7\,\|\boldsymbol{f}\|_{\infty,\Omega} < \frac{1}{2}, \tag{2.17}$$

*where $C_3$, $C_6$ and $C_7$ are the constants given below in (2.34). Then, there exists a constant $C_{\mathtt{rel}} > 0$, which depends only on parameters, $\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma_{\mathrm{D}}}$, $\|\boldsymbol{f}\|_{\infty,\Omega}$, and other constants, all them independent of $h$, such that*

$$\|\phi - \phi_h\|_{1,\Omega} + \|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H \leq C_{\mathtt{rel}}\,\boldsymbol{\theta}\,. \tag{2.18}$$

We begin the proof of (2.18) with the upper bounds derived in the following two subsections.

#### 2.3.2.1 A preliminary estimate for $\|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H$

In order to simplify the subsequent writing, we first introduce the following constants

$$C_0 := \frac{1}{\alpha}, \quad C_1 := 2\,C_0\,C_\varepsilon\,\widetilde{C}_\varepsilon\,\widetilde{C}_S(r)\frac{L_\mu(1+\kappa_1^2)^{1/2}}{\mu_1^2}, \quad C_2 := C_0\,(1+\kappa_2^2)^{1/2} + r\,C_1, \tag{2.19}$$

where $\widetilde{C}_S(r)$ and $C_\varepsilon$, $\widetilde{C}_\varepsilon$ are defined in [4, *cf.* (3.22) ] and [4, Lemma 3.9 and Theorem 3.13], respectively.

**Lemma 2.1.** *Let* $\boldsymbol{\theta}_0^2 := \sum_{T\in\mathcal{T}_h} \theta_{0,T}^2$, *where for each* $T \in \mathcal{T}_h$ *we set*

$$\theta_{0,T}^2 := \|\boldsymbol{f}\phi_h + \mathbf{div}\boldsymbol{\sigma}_h\|_{0,T}^2 + \left\|\nabla\boldsymbol{u}_h - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\|_{0,T}^2 + \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{D}})} \|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|_{0,e}^2 . \tag{2.20}$$

*Then there exists* $\bar{C} > 0$, *depending on* $C_0$, $\kappa_1$, $\kappa_3$, *and the trace operator in* $\mathbf{H}^1(\Omega)$, *such that*

$$\begin{aligned} &\|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H \\ &\quad\leq \bar{C}\left\{\boldsymbol{\theta}_0 + \|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}\right\} + \left\{C_1\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma_{\mathrm{D}}} + C_2\,\|\boldsymbol{f}\|_{\infty,\Omega}\right\}\|\phi-\phi_h\|_{1,\Omega}\,, \end{aligned} \tag{2.21}$$

*where* $C_1$ *and* $C_2$ *are given by* (2.19), *and* $E_h \in \mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'$, *defined for each* $\boldsymbol{\zeta} \in \mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)$ *by*

$$E_h(\boldsymbol{\zeta}) := \langle \boldsymbol{\zeta}\,\boldsymbol{\nu}, \boldsymbol{u}_{\mathrm{D}}\rangle_{\Gamma_{\mathrm{D}}} - \int_\Omega \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}} : \boldsymbol{\zeta} - \int_\Omega \boldsymbol{u}_h\cdot\mathbf{div}\boldsymbol{\zeta} - \kappa_2\int_\Omega (\boldsymbol{f}\phi_h + \mathbf{div}\boldsymbol{\sigma}_h)\cdot\mathbf{div}\boldsymbol{\zeta}\,, \tag{2.22}$$

*satisfies*

$$E_h(\boldsymbol{\zeta}_h) = 0 \qquad \forall\,\boldsymbol{\zeta}_h \in \mathbb{H}_h^{\boldsymbol{\sigma}}\,. \tag{2.23}$$

*Proof.* We first deduce from the $H$-ellipticity of $B_\phi$ (see [4, Lemma 3.4]) that there holds the global inf-sup condition

$$\alpha\,\|(\boldsymbol{\tau},\boldsymbol{v})\|_H \leq \sup_{\substack{(\boldsymbol{\zeta},\boldsymbol{w})\in H\\(\boldsymbol{\zeta},\boldsymbol{w})\neq\mathbf{0}}} \frac{B_\phi((\boldsymbol{\tau},\boldsymbol{v}),(\boldsymbol{\zeta},\boldsymbol{w}))}{\|(\boldsymbol{\zeta},\boldsymbol{w})\|_H} \qquad \forall\,(\boldsymbol{\tau},\boldsymbol{v})\in H, \tag{2.24}$$

where $\alpha$ is the constant of ellipticity, which depends only on $\mu_1,\mu_2,\Omega,\Gamma_{\mathrm{N}}$ and $\Gamma_{\mathrm{D}}$ (see [4, Lemma 3.4]). Then, applying (2.24) to the error $(\boldsymbol{\tau},\boldsymbol{v}) := (\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,\boldsymbol{u}-\boldsymbol{u}_h)$, we find that

$$\alpha\,\|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H \leq \sup_{\substack{(\boldsymbol{\zeta},\boldsymbol{w})\in H\\(\boldsymbol{\zeta},\boldsymbol{w})\neq\mathbf{0}}} \frac{F_\phi(\boldsymbol{\zeta},\boldsymbol{w}) - B_\phi((\boldsymbol{\sigma}_h,\boldsymbol{u}_h),(\boldsymbol{\zeta},\boldsymbol{w}))}{\|(\boldsymbol{\zeta},\boldsymbol{w})\|_H}. \tag{2.25}$$

Next, using the definitions of $B_\phi$ (*cf.* (2.9)) and $F_\phi$ (*cf.* (2.10)), and adding and subtracting suitable terms, we can write

$$\begin{aligned} F_\phi(\boldsymbol{\zeta},\boldsymbol{w}) - B_\phi((\boldsymbol{\sigma}_h,\boldsymbol{u}_h),(\boldsymbol{\zeta},\boldsymbol{w})) &= F_{\phi_h}(\boldsymbol{\zeta},\boldsymbol{w}) - B_{\phi_h}((\boldsymbol{\sigma}_h,\boldsymbol{u}_h),(\boldsymbol{\zeta},\boldsymbol{w}))\\ + B_{\phi_h}((\boldsymbol{\sigma}_h,\boldsymbol{u}_h),(\boldsymbol{\zeta},\boldsymbol{w})) - B_\phi((\boldsymbol{\sigma}_h,\boldsymbol{u}_h),(\boldsymbol{\zeta},\boldsymbol{w})) &+ F_\phi(\boldsymbol{\zeta},\boldsymbol{w}) - F_{\phi_h}(\boldsymbol{\zeta},\boldsymbol{w})\,. \end{aligned} \tag{2.26}$$

In this way, employing the estimate for $|B_{\phi_h}(\,\cdot\,,(\boldsymbol{\zeta},\boldsymbol{w})) - B_\phi(\,\cdot\,,(\boldsymbol{\zeta},\boldsymbol{w}))|$ (see [4, eq. (3.29)]) and $|F_\phi(\boldsymbol{\zeta},\boldsymbol{w}) - F_{\phi_h}(\boldsymbol{\zeta},\boldsymbol{w})|$ (see [4, eq. (3.28)]), we deduce from (2.25) and (2.26) that

$$\begin{aligned} \|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H &\leq C_0\,\|F_{\phi_h}(\,\cdot\,) - B_{\phi_h}((\boldsymbol{\sigma}_h,\boldsymbol{u}_h),\,\cdot\,)\|_{H'}\\ &+ \left\{C_1\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma_{\mathrm{D}}} + C_2\,\|\boldsymbol{f}\|_{\infty,\Omega}\right\}\|\phi-\phi_h\|_{1,\Omega}, \end{aligned} \tag{2.27}$$

where, bearing in mind (2.22), there holds

$$F_{\phi_h}(\boldsymbol{\zeta},\boldsymbol{w}) - B_{\phi_h}((\boldsymbol{\sigma}_h,\boldsymbol{u}_h),(\boldsymbol{\zeta},\boldsymbol{w})) = E_h(\boldsymbol{\zeta}) + \widehat{E}_h(\boldsymbol{w}) \qquad \forall\,(\boldsymbol{\zeta},\boldsymbol{w})\in H\,, \tag{2.28}$$

with $\widehat{E}_h \in \mathbf{H}^1(\Omega)'$ defined for each $\boldsymbol{w}\in\mathbf{H}^1(\Omega)$ by

$$\widehat{E}_h(\boldsymbol{w}) := = \int_\Omega (\boldsymbol{f}\phi_h + \mathbf{div}\boldsymbol{\sigma}_h)\cdot\boldsymbol{w} + \kappa_1\int_\Omega \left(\nabla\boldsymbol{u}_h - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right) : \nabla\boldsymbol{w} + \kappa_3\int_{\Gamma_{\mathrm{D}}} (\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h)\cdot\boldsymbol{w}\,.$$

Then, applying the Cauchy-Schwarz inequality we readily deduce the existence of a constant $\widehat{c} > 0$, depending on $\kappa_1$, $\kappa_3$, and the trace operator in $\mathbf{H}^1(\Omega)$, such that

$$\|\widehat{E}_h\|_{\mathbf{H}^1(\Omega)'} \leq \widehat{c}\,\boldsymbol{\theta}_0\,,$$

which, together with (2.27) and (2.28), imply the main inequality (2.21). Moreover, using the fact that

$$F_{\phi_h}(\boldsymbol{\zeta}_h, \boldsymbol{w}_h) - B_{\phi_h}((\boldsymbol{\sigma}_h, \boldsymbol{u}_h), (\boldsymbol{\zeta}_h, \boldsymbol{w}_h)) = 0 \qquad \forall\, (\boldsymbol{\zeta}_h, \boldsymbol{w}_h) \in H_h\,,$$

and taking in particular $\boldsymbol{w}_h = \mathbf{0}$, we deduce (2.23), which completes the proof. □

Observe, according to (2.23), that for each $\boldsymbol{\zeta} \in \mathbb{H}_{\mathrm{N}}(\mathbf{div}, \Omega)$ there holds

$$E_h(\boldsymbol{\zeta}) = E_h(\boldsymbol{\zeta} - \boldsymbol{\zeta}_h) \qquad \forall\, \boldsymbol{\zeta}_h \in \mathbb{H}_h^{\boldsymbol{\sigma}}\,,$$

and hence the upper bound of $\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}$ to be derived below (see Section 2.3.2.4) will employ the foregoing expression with a suitable choice of $\boldsymbol{\zeta}_h \in \mathbb{H}_h^{\boldsymbol{\sigma}}$.

#### 2.3.2.2 A preliminary estimate for $\|\phi - \phi_h\|_{1,\Omega}$

We begin with the following results.

**Lemma 2.2.** *The nonlinear operator $\mathbf{A}_{\boldsymbol{u}} : \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) \to [\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)]'$ induced by $A_{\boldsymbol{u}}$ (cf. (2.11)), that is*

$$[\mathbf{A}_{\boldsymbol{u}}(\phi), \psi] := \int_\Omega \vartheta(|\nabla\phi|)\,\nabla\phi\cdot\nabla\psi - \int_\Omega \phi\,\boldsymbol{u}\cdot\nabla\psi \qquad \forall\,\psi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,, \tag{2.29}$$

*where $[\cdot,\cdot]$ is the duality pairing between $\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ and $[\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)]'$, is Gâteaux differentiable in $\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$.*

*Proof.* We begin by observing, thanks to simple computations and the $C^1$-regularity of $\vartheta$, that for all $\widehat{\phi}, \psi, \varphi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$, with $\nabla\widehat{\phi} \neq \mathbf{0}$ there holds

$$\begin{aligned}\lim_{\epsilon\to 0}\frac{[\mathbf{A}_{\boldsymbol{u}}(\widehat{\phi}+\epsilon\,\psi) - \mathbf{A}_{\boldsymbol{u}}(\widehat{\phi}), \varphi]}{\epsilon} &= \int_\Omega \vartheta'(|\nabla\widehat{\phi}|)\frac{\left(\nabla\widehat{\phi}\cdot\nabla\psi\right)}{|\nabla\widehat{\phi}|}\nabla\widehat{\phi}\cdot\nabla\varphi \\ &+ \int_\Omega \vartheta(|\nabla\widehat{\phi}|)\nabla\psi\cdot\nabla\varphi - \int_\Omega \psi\boldsymbol{u}\cdot\nabla\varphi,\end{aligned} \tag{2.30}$$

whereas for $\nabla\widehat{\phi} = \mathbf{0}$, we find that

$$\lim_{\epsilon\to 0}\frac{[\mathbf{A}_{\boldsymbol{u}}(\widehat{\phi}+\epsilon\,\psi) - \mathbf{A}_{\boldsymbol{u}}(\widehat{\phi}), \varphi]}{\epsilon} = \int_\Omega \vartheta(0)\nabla\psi\cdot\nabla\varphi - \int_\Omega \psi\boldsymbol{u}\cdot\nabla\varphi. \tag{2.31}$$

In this way, the identities (2.30) and (2.31) show that $\mathbf{A}_{\boldsymbol{u}}$ is Gâteaux differentiable at $\widehat{\phi}$. Moreover, $\mathcal{D}\mathbf{A}_{\boldsymbol{u}}(\widehat{\phi})$ is the bounded linear operator of $\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ into $[\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)]'$ that can be identified with the bilinear form $\mathcal{D}\mathbf{A}_{\boldsymbol{u}}(\widehat{\phi}) : \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\times\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) \to \mathrm{R}$ defined by

$$\mathcal{D}\mathbf{A}_{\boldsymbol{u}}(\widehat{\phi})(\psi,\varphi) := \lim_{\epsilon\to 0}\frac{[\mathbf{A}_{\boldsymbol{u}}(\widehat{\phi}+\epsilon\,\psi) - \mathbf{A}_{\boldsymbol{u}}(\widehat{\phi}), \varphi]}{\epsilon} \qquad \forall\,\psi,\varphi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega). \tag{2.32}$$

□

**Lemma 2.3.** *Let $c_p$ and $c(\Omega)$ be the constants given by (2.7) and [4, eq. (3.5)], respectively, and let $\boldsymbol{u} \in \mathbf{H}^1(\Omega)$ be such that*

$$\|\boldsymbol{u}\|_{1,\Omega} < \frac{\vartheta_1}{2\, c_p\, c(\Omega)}.$$

*Then, the family of Gâteaux derivates $\{\mathcal{D}\mathbf{A}_{\boldsymbol{u}}(\widehat{\phi})\}_{\widehat{\phi}\in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)}$ is uniformly bounded and uniformly elliptic on $\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\times \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$. More precisely, there exist positive constants $\widetilde{\lambda}, \widetilde{\alpha}$, depending only on $\vartheta_1, \vartheta_2$ (cf. (2.4)), $c(\Omega)$, and $c_p$, such that for all $\widehat{\phi}, \varphi, \psi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$, there holds*

$$|\mathcal{D}\mathbf{A}_{\boldsymbol{u}}(\widehat{\phi})(\psi,\varphi)| \,\leq\, \widetilde{\lambda}\,\|\psi\|_{1,\Omega}\,\|\varphi\|_{1,\Omega} \qquad \text{and} \qquad \mathcal{D}\mathbf{A}_{\boldsymbol{u}}(\widehat{\phi})(\psi,\psi) \,\geq\, \widetilde{\alpha}\,\|\psi\|^2_{1,\Omega}.$$

*Proof.* It proceeds similarly to the proof of [60, Lemma 5.1]. □

As a consequence of the ellipticity of the family $\{\mathcal{D}\mathbf{A}_{\boldsymbol{u}}(\widehat{\phi})\}_{\widehat{\phi}\in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)}$, we obtain the following global inf-sup condition

$$\widetilde{\alpha}\,\|\psi\|_{1,\Omega} \,\leq\, \sup_{\substack{\varphi\in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\\ \varphi\neq 0}} \frac{\mathcal{D}\mathbf{A}_{\boldsymbol{u}}(\widehat{\phi})(\psi,\varphi)}{\|\varphi\|_{1,\Omega}} \qquad \forall\, \psi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega). \tag{2.33}$$

Next, similarly as before, we simplify the subsequent writing by introducing the following constants

$$\widetilde{C} \,:=\, \frac{1}{\widetilde{\alpha}}, \quad C_3 := \widetilde{C}\, L_\gamma, \quad C_4 := r\, c(\Omega)\, \bar{C}, \quad C_5 := r\, c(\Omega)\, \widetilde{C}, \quad C_6 := C_1\, C_5 \qquad C_7 := C_2\, C_5\,, \tag{2.34}$$

where $\bar{C}$ is the constant provided by Lemma 2.1.

**Lemma 2.4.** *Assume that*

$$C_3\,|\boldsymbol{k}| \,+\, C_6\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma_{\mathrm{D}}} \,+\, C_7\,\|\boldsymbol{f}\|_{\infty,\Omega} \,<\, \frac{1}{2}. \tag{2.35}$$

*Then, there exists $\widehat{C} > 0$, depending on $\widetilde{C}$ and $C_4$ (cf. (2.34)), such that*

$$\|\phi - \phi_h\|_{1,\Omega} \,\leq\, \widehat{C}\,\Big\{\boldsymbol{\theta}_0 \,+\, \|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'} \,+\, \|\widetilde{E}_h\|_{\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'}\Big\}\,, \tag{2.36}$$

*where $\boldsymbol{\theta}_0$ and $E_h$ are given in the statement of Lemma 2.1 and (2.22), respectively, and $\widetilde{E}_h \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'$, defined for each $\varphi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ by*

$$\widetilde{E}_h(\varphi) \,:=\, \int_\Omega g\,\varphi \,-\, \int_\Omega \Big\{\vartheta(|\nabla\phi_h|)\,\nabla\phi_h \,-\, \phi_h\,\boldsymbol{u}_h \,-\, \gamma(\phi_h)\,\boldsymbol{k}\Big\}\cdot\nabla\varphi\,, \tag{2.37}$$

*satisfies*

$$\widetilde{E}_h(\varphi_h) \,=\, 0 \qquad \forall\, \varphi_h \in \mathrm{H}^\phi_h\,. \tag{2.38}$$

*Proof.* Since $\phi$ and $\phi_h$ belong to $\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$, a straightforward application of the mean value theorem yields the existence of a convex combination of $\phi$ and $\phi_h$, say $\widehat{\phi}_h \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$, such that

$$\mathcal{D}\mathbf{A}_{\boldsymbol{u}}(\widehat{\phi}_h)(\phi-\phi_h,\varphi) \,=\, [\mathbf{A}_{\boldsymbol{u}}(\phi) - \mathbf{A}_{\boldsymbol{u}}(\phi_h),\varphi] \qquad \forall\, \varphi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega).$$

Next, applying (2.33) to the Galerkin error $\psi := \phi - \phi_h$, we find that

$$\widetilde{\alpha}\,\|\phi-\phi_h\|_{1,\Omega} \,\leq\, \sup_{\substack{\varphi\in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\\ \varphi\neq 0}} \frac{[\mathbf{A}_{\boldsymbol{u}}(\phi)-\mathbf{A}_{\boldsymbol{u}}(\phi_h),\varphi]}{\|\varphi\|_{1,\Omega}}. \tag{2.39}$$

Now, using the fact that $[\mathbf{A}_{\boldsymbol{u}}(\phi),\varphi] = [G_\phi,\varphi]$, the definition of $\mathbf{A}_{\boldsymbol{u}}$ (*cf.* (2.29)), and adding and substracting suitable terms, it follows that

$$[\mathbf{A}_{\boldsymbol{u}}(\phi)-\mathbf{A}_{\boldsymbol{u}}(\phi_h),\varphi] \,=\, [G_{\phi_h}-\mathbf{A}_{\boldsymbol{u}_h}(\phi_h),\varphi] + [G_\phi - G_{\phi_h},\varphi] + [\mathbf{A}_{\boldsymbol{u}_h}(\phi_h)-\mathbf{A}_{\boldsymbol{u}}(\phi_h),\varphi]. \tag{2.40}$$

In this way, applying the estimate for $|[G_\phi - G_{\phi_h},\varphi]|$ (see [4, eq. (5.5)]) and $|[\mathbf{A}_{\boldsymbol{u}}(\phi_h)-\mathbf{A}_{\boldsymbol{u}_h}(\phi_h),\varphi]|$ (see [4, eq. (5.6)]), we deduce from (2.39) and (2.40) that

$$\|\phi-\phi_h\|_{1,\Omega} \,\leq\, \widetilde{C}\,\|G_{\phi_h}-\mathbf{A}_{\boldsymbol{u}_h}(\phi_h)\|_{[\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)]'} \,+\, \widetilde{C}\,L_\gamma\,|\boldsymbol{k}|\,\|\phi-\phi_h\|_{1,\Omega} \,+\, r\,c(\Omega)\,\widetilde{C}\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}. \tag{2.41}$$

Then, bounding $\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}$ by the error estimate provided by (2.21) (*cf.* Lemma 2.1), and then employing (2.35), we arrive at

$$\|\phi-\phi_h\|_{1,\Omega} \,\leq\, 2\,\widetilde{C}\left\{\|G_{\phi_h}-\mathbf{A}_{\boldsymbol{u}_h}(\phi_h)\|_{[\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)]'} \,+\, C_4\left(\boldsymbol{\theta}_0 \,+\, \|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}\right)\right\},$$

where, bearing in mind (2.37), there holds

$$[G_{\phi_h}-\mathbf{A}_{\boldsymbol{u}_h}(\phi_h),\varphi] \,=\, \widetilde{E}_h(\varphi) \qquad \forall\,\varphi\in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega).$$

Finally, using the fact that $[G_{\phi_h}-\mathbf{A}_{\boldsymbol{u}_h}(\phi_h),\varphi_h] = 0 \quad \forall\varphi_h\in \mathrm{H}^\phi_h$, we obtain (2.38) and the proof concludes. □

At this point we remark, similarly as we did at the end of Section 2.3.2.1, and thanks now to (2.38), that for each $\varphi\in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ there holds

$$\widetilde{E}_h(\varphi) \,=\, \widetilde{E}_h(\varphi-\varphi_h) \qquad \forall\,\varphi_h\in \mathrm{H}^\phi_h,$$

and therefore $\|\widetilde{E}_h\|_{\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'}$ will be estimated below (see Section 2.3.2.4) by employing the foregoing expression with a suitable choice of $\varphi_h\in \mathrm{H}^\phi_h$.

#### 2.3.2.3 A preliminary estimate for the total error

We now combine the inequalities provided by Lemmas 2.1 and 2.4 to derive a first estimate for the total error $\|\phi-\phi_h\|_{1,\Omega}+\|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H$. To this end, we now introduce the constants

$$C(\boldsymbol{u}_{\mathrm{D}},\boldsymbol{f}) := \widehat{C}\left\{C_1\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma_{\mathrm{D}}} \,+\, C_2\,\|\boldsymbol{f}\|_{\infty,\Omega} \,+\, 1\right\} \quad\text{and}\quad c(\boldsymbol{u}_{\mathrm{D}},\boldsymbol{f}) := \bar{C} \,+\, C(\boldsymbol{u}_{\mathrm{D}},\boldsymbol{f}),$$

where $\bar{C}$ and $\widehat{C}$ are provided by Lemmas 2.1 and 2.4, respectively, and $C_1$ and $C_2$ are given by (2.19).

**Theorem 2.2.** *Assume that*

$$C_3\,|\boldsymbol{k}| \,+\, C_6\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma_{\mathrm{D}}} \,+\, C_7\,\|\boldsymbol{f}\|_{\infty,\Omega} \,<\, \frac{1}{2}.$$

*Then there holds*

$$\|\phi-\phi_h\|_{1,\Omega}+\|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H \leq C(\boldsymbol{u}_{\mathrm{D}},\boldsymbol{f})\,\|\widetilde{E}_h\|_{\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'} + c(\boldsymbol{u}_{\mathrm{D}},\boldsymbol{f})\left\{\boldsymbol{\theta}_0+\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}\right\}. \tag{2.42}$$

*Proof.* It suffices to replace the upper bound for $\|\phi - \phi_h\|_{1,\Omega}$ given by (2.36) into the second term on the right hand side of (2.21), and then add the resulting estimate to the right hand side of (2.36). We omit further details. □

It is clear from (2.42) that, in order to obtain an explicit estimate for the total error, it only remains to derive suitable upper bounds for $\|\widetilde{E}_h\|_{\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'}$ and $\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}$. This is precisely the purpose of the next subsection.

#### 2.3.2.4 Upper bounds for $\|\widetilde{E}_h\|_{\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'}$ and $\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}$

In what follows we make use of the Clément interpolation operator $\mathcal{I}_h : \mathrm{H}^1(\Omega) \to X_h$ (*cf.* [42]), where $X_h$ is given by

$$X_h := \{v_h \in C(\overline{\Omega}) : v_h|_T \in \mathrm{P}_1(T) \quad \forall T \in \mathcal{T}_h\}.$$

The following Lemma establishes the local approximation properties of $\mathcal{I}_h$.

**Lemma 2.5.** *There exist constants $c_1, c_2 > 0$, independent of $h$, such that for all $v \in \mathrm{H}^1(\Omega)$ there hold*

$$\|v - \mathcal{I}_h(v)\|_{0,T} \leq c_1 h_T \|v\|_{1,\Delta(T)} \quad \forall T \in \mathcal{T}_h, \tag{2.43}$$

*and*

$$\|v - \mathcal{I}_h(v)\|_{0,e} \leq c_2 h_e^{1/2} \|v\|_{1,\Delta(e)} \quad \forall e \in \mathcal{E}_h, \tag{2.44}$$

*where $\Delta(T)$ and $\Delta(e)$ are the union of all elements intersecting with $T$ and $e$, respectively.*

*Proof.* See [42]. □

We now recall from Subsection 2.3.1 that we have defined there

$$\widetilde{\boldsymbol{\sigma}}_h := \vartheta(|\nabla\phi_h|)\nabla\phi_h - \phi_h \boldsymbol{u}_h - \gamma(\phi_h)\boldsymbol{k}\,. \tag{2.45}$$

Then, the following lemma provides an upper bound for $\|\widetilde{E}_h\|_{\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'}$.

**Lemma 2.6.** *Let $\widetilde{\boldsymbol{\eta}}^2 := \sum_{T\in\mathcal{T}_h} \widetilde{\eta}_T^2$, where for each $T \in \mathcal{T}_h$ we set*

$$\widetilde{\eta}_T^2 := h_T^2 \|g + \mathrm{div}\widetilde{\boldsymbol{\sigma}}_h\|_{0,T}^2 + \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Omega)} h_e \|[\![\widetilde{\boldsymbol{\sigma}}_h \cdot \boldsymbol{\nu}_e]\!]\|_{0,e}^2 + \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{N}})} h_e \|\widetilde{\boldsymbol{\sigma}}_h \cdot \boldsymbol{\nu}\|_{0,e}^2\,.$$

*Then there exists $c > 0$, independent of $h$, such that*

$$\|\widetilde{E}_h\|_{\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'} \leq c\,\widetilde{\boldsymbol{\eta}}\,. \tag{2.46}$$

*Proof.* Given $\varphi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ we let $\varphi_h := \mathcal{I}_h(\varphi) \in \mathrm{H}^{\phi}_h$, and observe, according to (2.37), (2.38), and (2.45), that

$$\widetilde{E}_h(\varphi) = \widetilde{E}_h(\varphi - \varphi_h) = \sum_{T\in\mathcal{T}_h} \int_T g\,(\varphi - \varphi_h) - \sum_{T\in\mathcal{T}_h} \int_T \widetilde{\boldsymbol{\sigma}}_h \cdot \nabla(\varphi - \varphi_h)\,.$$

Next, integrating by parts on each $T \in \mathcal{T}_h$ in the last term on the right hand side of the foregoing equation, we find that

$$\widetilde{E}_h(\varphi) \,=\, \sum_{T\in\mathcal{T}_h}\int_T (g \,+\, \mathrm{div}\widetilde{\boldsymbol{\sigma}}_h)\,(\varphi-\varphi_h) \,-\, \sum_{e\in\mathcal{E}_h(\Omega)}\int_e (\varphi-\varphi_h)\,[\![\widetilde{\boldsymbol{\sigma}}_h\cdot\boldsymbol{\nu}_e]\!] \,-\, \sum_{e\in\mathcal{E}_h(\Gamma_{\mathrm{N}})}\int_e (\varphi-\varphi_h)\,\widetilde{\boldsymbol{\sigma}}_h\cdot\boldsymbol{\nu}\,,$$

from which, applying Cauchy-Schwarz inequality, employing the approximation properties of the Clément operator given by (2.43) and (2.44), and performing some algebraic rearrangements, we readily conclude that

$$|\widetilde{E}_h(\varphi)| \,\leq\, c\,\widetilde{\boldsymbol{\eta}}\,\|\varphi\|_{1,\Omega}\,,$$

which yields (2.46) and finishes the proof. □

We now aim to provide an upper bound for $\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}$ (*cf.* (2.22)), which, being less straightforward than Lemma 2.6, requires several preliminary results and estimates. We begin by introducing the space

$$\mathbf{H}^1_{\Gamma_{\mathrm{N}}}(\Omega) \,:=\, \Big\{\boldsymbol{\varphi}\in\mathbf{H}^1(\Omega):\quad \boldsymbol{\varphi}\,=\,\mathbf{0}\quad\text{on}\quad\Gamma_{\mathrm{N}}\Big\}\,,$$

and establishing a suitable Helmholtz decomposition of our space $\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)$.

**Lemma 2.7.** *Assume that $\Omega$ is a connected domain and that $\Gamma_{\mathrm{N}}$ is contained in the boundary of a convex part of $\Omega$, that is there exists a convex domain $B$ such that $\overline{\Omega}\subseteq B$ and $\Gamma_{\mathrm{N}}\subseteq\partial B$ (see Figure 2.1). Then, for each $\boldsymbol{\zeta}\in\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)$, there exist $\boldsymbol{\tau}\in\mathbb{H}^1(\Omega)$ and $\boldsymbol{\chi}\in\mathbf{H}^1_{\Gamma_{\mathrm{N}}}(\Omega)$ such that*

$$\boldsymbol{\zeta} \,=\, \boldsymbol{\tau} \,+\, \underline{\mathbf{curl}}(\boldsymbol{\chi}) \quad \textit{in} \quad \Omega\,, \tag{2.47}$$

*and*

$$\|\boldsymbol{\tau}\|_{1,\Omega} \,+\, \|\boldsymbol{\chi}\|_{1,\Omega} \,\leq\, C\,\|\boldsymbol{\zeta}\|_{\mathbf{div},\Omega}\,, \tag{2.48}$$

*with a positive constant $C$ independent of $\boldsymbol{\zeta}$.*

*Proof.* Given $\boldsymbol{\zeta}\in\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)$, we let $\boldsymbol{z}\in\mathbf{H}^2(B)$ be the unique weak solution of the boundary value problem:

$$\Delta\boldsymbol{z} \,=\, \begin{cases} \mathbf{div}\boldsymbol{\zeta} & \text{in} \quad \Omega \\ \\ \dfrac{-1}{|\,B\backslash\overline{\Omega}\,|}\displaystyle\int_\Omega \mathbf{div}\boldsymbol{\zeta} & \text{in} \quad B\backslash\overline{\Omega} \end{cases}\,, \qquad \nabla\boldsymbol{z}\,\boldsymbol{\nu} \,=\, \mathbf{0} \quad \text{on} \quad \partial B, \qquad \int_\Omega \boldsymbol{z} \,=\, \mathbf{0}. \tag{2.49}$$

Thanks to the elliptic regularity result of (2.49) we have that $\boldsymbol{z}\in\mathbf{H}^2(B)$ and

$$\|\boldsymbol{z}\|_{2,B} \,\leq\, c\,\|\mathbf{div}\boldsymbol{\zeta}\|_{0,\Omega}\,, \tag{2.50}$$

where $c>0$ is independent of $\boldsymbol{z}$. In addition, it is clear that $\boldsymbol{\tau}:=(\nabla\boldsymbol{z})|_\Omega\in\mathbb{H}^1(\Omega)$, $\mathbf{div}\boldsymbol{\tau}\,=\,\Delta\boldsymbol{z}\,=\,\mathbf{div}\boldsymbol{\zeta}$ in $\Omega$, $\boldsymbol{\tau}\,\boldsymbol{\nu}=\mathbf{0}$ on $\partial B$ (which certainly yields $\boldsymbol{\tau}\,\boldsymbol{\nu}=\mathbf{0}$ on $\Gamma_{\mathrm{N}}$), and

$$\|\boldsymbol{\tau}\|_{1,\Omega} \,\leq\, \|\boldsymbol{z}\|_{2,\Omega} \,\leq\, \|\boldsymbol{z}\|_{2,B} \,\leq\, c\,\|\mathbf{div}\boldsymbol{\zeta}\|_{0,\Omega}\,. \tag{2.51}$$

On the other hand, since $\mathbf{div}(\boldsymbol{\zeta}-\boldsymbol{\tau}) \,=\, \mathbf{0}$ in $\Omega$, and $\Omega$ is connected, there exists $\boldsymbol{\chi}\in\mathbf{H}^1(\Omega)$ such that

$$\boldsymbol{\zeta}\,-\,\boldsymbol{\tau} \,=\, \underline{\mathbf{curl}}(\boldsymbol{\chi}) \qquad \text{in} \quad \Omega\,. \tag{2.52}$$

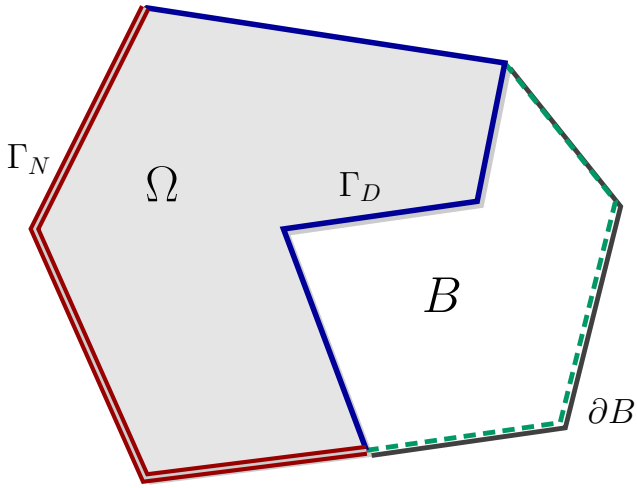


Figure 2.1: Extension of $\Omega$ to a convex domain $B$ for the Helmholtz decomposition.

In turn, noting that $\mathbf{0} = (\boldsymbol{\zeta} - \boldsymbol{\tau})\,\boldsymbol{\nu} = \underline{\mathbf{curl}}(\boldsymbol{\chi})\,\boldsymbol{\nu} = \frac{d\boldsymbol{\chi}}{ds}$ on $\Gamma_{\rm N}$, we deduce that $\boldsymbol{\chi}$ is constant on $\Gamma_{\rm N}$, and therefore $\boldsymbol{\chi}$ can be chosen so that $\boldsymbol{\chi} \in \mathbf{H}^1_{\Gamma_{\rm N}}(\Omega)$, which, together with (2.52), gives (2.47). In addition, from the equivalence between $\|\boldsymbol{\chi}\|_{1,\Omega}$ and $|\boldsymbol{\chi}|_{1,\Omega} = \|\underline{\mathbf{curl}}(\boldsymbol{\chi})\|_{0,\Omega}$ (which is a consequence of the generalized Poincaré inequality), and employing (2.51) and (2.47), we deduce that there exists a constant $\widetilde{c} > 0$ such that

$$\|\boldsymbol{\chi}\|_{1,\Omega} \leq \widetilde{c}\|\boldsymbol{\zeta}\|_{\mathbf{div},\Omega}\,. \tag{2.53}$$

Finally, it is clear that (2.51) and (2.53) yield (2.48), which is the stability estimate for (2.47). □

We remark here, as already announced at the beginning of this section, that Lemma 2.7 also holds in the 3D case. The corresponding proof follows by combining the extension technique introduced in [67, Lemma 4.3] with the approach suggested in the foregoing proof (*cf.* auxiliary problem (2.49)), and the results stated in [10, Theorems 2.17 and 3.12, and Corollary 3.16].

We now consider the finite element subspace of $\mathbf{H}_{\Gamma_{\rm N}}(\Omega)$ given by

$$\mathbf{X}_{h,N} := \Big\{\boldsymbol{\varphi}_h \in \mathbf{C}(\overline{\Omega}) : \quad \boldsymbol{\varphi}_h|_T \in \mathbf{P}_1(T) \quad \forall\, T \in \mathcal{T}_h, \quad \boldsymbol{\varphi}_h = \mathbf{0} \text{ on } \Gamma_{\rm N}\Big\}\,,$$

and introduce, analogously as before, the Clément interpolation operator $\boldsymbol{\mathcal{I}}_{h,{\rm N}} : \mathbf{H}_{\Gamma_{\rm N}}(\Omega) \to \mathbf{X}_{h,N}$. In addition, we let $\Pi_h : \mathbb{H}^1(\Omega) \to \mathbb{H}^{\boldsymbol{\sigma}}_h$ be the Raviart-Thomas interpolation operator (see [26],[95]), which, given $\bar{\boldsymbol{\tau}} \in \mathbb{H}^1(\Omega)$, is characterized by the identities:

$$\int_e \Pi_h(\bar{\boldsymbol{\tau}})\,\boldsymbol{\nu}\cdot\boldsymbol{p} = \int_e \bar{\boldsymbol{\tau}}\,\boldsymbol{\nu}\cdot\boldsymbol{p}, \quad \forall \text{ edge } e \in \mathcal{T}_h, \quad \forall \boldsymbol{p} \in \mathbf{P}_k(e), \quad \text{when } k \geq 0, \tag{2.54}$$

and

$$\int_T \Pi_h(\bar{\boldsymbol{\tau}}) : \boldsymbol{\rho} = \int_T \bar{\boldsymbol{\tau}} : \boldsymbol{\rho}, \quad \forall\, T \in \mathcal{T}_h, \quad \forall \boldsymbol{\rho} \in \mathbb{P}_{k-1}(T), \quad \text{when } k \geq 1, \tag{2.55}$$

where $\mathbf{P}_k(e) := [P_k(e)]^2$ and $\mathbb{P}_{k-1}(T) := [P_{k-1}(T)]^{2\times 2}$. It is easy to show, using (2.54) and (2.55), that (see, e.g. [63, Lemma 3.7], [92, eq. (3.4.23)])

$$\mathbf{div}(\Pi_h(\bar{\boldsymbol{\tau}})) = \mathcal{P}_h(\mathbf{div}\bar{\boldsymbol{\tau}})\,, \tag{2.56}$$

where $\mathcal{P}_h : \mathbf{L}^2(\Omega) \to \mathbf{Q}_h$ is the $\mathbf{L}^2(\Omega)$-orthogonal projector and

$$\mathbf{Q}_h := \big\{ \boldsymbol{v} \in \mathbf{L}^2(\Omega) : \quad \boldsymbol{v}|_T \in \mathbf{P}_k(T) \quad \forall\, T \in \mathcal{T}_h \big\}.$$

Note that $\mathcal{P}_h$ can also be identified with $(P_h, P_h)$, where $P_h$ is the orthogonal projector from $\mathrm{L}^2(\Omega)$ into $\mathrm{Q}_h$ with $\mathrm{Q}_h := \{ v \in \mathrm{L}^2(\Omega) : \quad v|_T \in P_k(T) \quad \forall\, T \in \mathcal{T}_h \}$. Furthermore, the following approximation properties hold (*cf.* [26, 40, 63, 95]):

$$\|v - P_h(v)\|_{0,T} \,\leq\, c\, h_T^m\, |v|_{m,T} \quad \forall\, T \in \mathcal{T}_h\,, \tag{2.57}$$

for each $v \in H^m(\Omega)$, with $0 \leq m \leq k+1$,

$$\|\bar{\boldsymbol{\tau}} - \Pi_h(\bar{\boldsymbol{\tau}})\|_{0,T} \,\leq\, c\, h_T^m\, |\bar{\boldsymbol{\tau}}|_{m,T} \quad \forall\, T \in \mathcal{T}_h\,, \tag{2.58}$$

for each $\bar{\boldsymbol{\tau}} \in \mathbb{H}^m(\Omega)$, with $1 \leq m \leq k+1$,

$$\|\mathbf{div}(\bar{\boldsymbol{\tau}} - \Pi_h(\bar{\boldsymbol{\tau}}))\|_{0,T} \,\leq\, c\, h_T^m\, |\mathbf{div}\bar{\boldsymbol{\tau}}|_{m,T} \quad \forall\, T \in \mathcal{T}_h\,, \tag{2.59}$$

for each $\bar{\boldsymbol{\tau}} \in \mathbb{H}^1(\Omega)$ such that $\mathbf{div}\bar{\boldsymbol{\tau}} \in \mathbf{H}^m(\Omega)$, with $0 \leq m \leq k+1$, and

$$\|\bar{\boldsymbol{\tau}}\boldsymbol{\nu} - \Pi_h(\bar{\boldsymbol{\tau}})\,\boldsymbol{\nu}\|_{0,e} \,\leq\, c\, h_e^{1/2}\, \|\bar{\boldsymbol{\tau}}\|_{1,T_e} \quad \forall\, \mathrm{edge}\, e \in \mathcal{T}_h\,, \tag{2.60}$$

for each $\bar{\boldsymbol{\tau}} \in \mathbb{H}^1(\Omega)$, where $T_e \in \mathcal{T}_h$ contains $e$ on its boundary.

Then, given $\boldsymbol{\zeta} \in \mathbb{H}_{\mathrm{N}}(\mathbf{div}, \Omega)$ and its Helmholtz decomposition (2.47), we define $\boldsymbol{\chi}_h := \boldsymbol{\mathcal{I}}_{h,\mathrm{N}}(\boldsymbol{\chi})$, and set

$$\boldsymbol{\zeta}_h := \Pi_h(\boldsymbol{\tau}) + \underline{\mathbf{curl}}(\boldsymbol{\chi}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \tag{2.61}$$

as its associated discrete Helmholtz decomposition. It follows that

$$\boldsymbol{\zeta} - \boldsymbol{\zeta}_h \;=\; \boldsymbol{\tau} - \Pi_h(\boldsymbol{\tau}) + \underline{\mathbf{curl}}(\boldsymbol{\chi} - \boldsymbol{\chi}_h)\,,$$

from which, using (2.56) and the fact that $\mathbf{div}\,\boldsymbol{\tau} = \Delta \boldsymbol{z} = \mathbf{div}\,\boldsymbol{\zeta}$ in $\Omega$, yields

$$\mathbf{div}(\boldsymbol{\zeta} - \boldsymbol{\zeta}_h) \;=\; \mathbf{div}(\boldsymbol{\tau} - \Pi_h(\boldsymbol{\tau})) \;=\; (\mathbf{I} - \mathcal{P}_h)(\mathbf{div}\boldsymbol{\zeta})\,. \tag{2.62}$$

Hence, according to (2.22) and (2.23), and noting from (2.62) that

$$\int_\Omega \boldsymbol{u}_h \cdot \mathbf{div}(\boldsymbol{\tau} - \Pi_h(\boldsymbol{\tau})) \;=\; \int_\Omega \boldsymbol{u}_h \cdot (\mathbf{I} - \mathcal{P}_h)(\mathbf{div}\,\boldsymbol{\zeta}) \;=\; 0\,,$$

we find that

$$E_h(\boldsymbol{\zeta}) \;=\; E_h(\boldsymbol{\zeta} - \boldsymbol{\zeta}_h) \;=\; E_{h,1}(\boldsymbol{\tau}) + E_{h,2}(\boldsymbol{\chi})\,, \tag{2.63}$$

where

$$\begin{aligned} E_{h,1}(\boldsymbol{\tau}) \;:=\; & \langle (\boldsymbol{\tau} - \Pi_h(\boldsymbol{\tau}))\,\boldsymbol{\nu}, \boldsymbol{u}_D \rangle_{\Gamma_{\mathrm{D}}} \;-\; \int_\Omega \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} : \big(\boldsymbol{\tau} - \Pi_h(\boldsymbol{\tau})\big) \\ & +\quad \kappa_2 \int_\Omega (\boldsymbol{f}\phi_h \,+\, \mathbf{div}\boldsymbol{\sigma}_h) \cdot (\mathbf{I} - \mathcal{P}_h)(\mathbf{div}\,\boldsymbol{\tau}))\,, \end{aligned} \tag{2.64}$$

and

$$E_{h,2}(\boldsymbol{\chi}) \;:=\; \langle \underline{\mathbf{curl}}(\boldsymbol{\chi} - \boldsymbol{\chi}_h)\,\boldsymbol{\nu}, \boldsymbol{u}_{\mathrm{D}} \rangle_{\Gamma_{\mathrm{D}}} \;-\; \int_\Omega \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} : \underline{\mathbf{curl}}(\boldsymbol{\chi} - \boldsymbol{\chi}_h)\,. \tag{2.65}$$

It is now evident from (2.63) that, in order to estimate $\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}$, it only remains to bound $|E_{h,1}(\boldsymbol{\tau})|$ and $|E_{h,2}(\boldsymbol{\chi})|$ in terms of a multiple of $\|\boldsymbol{\zeta}\|_{\mathbf{div},\Omega}$, which is indeed the purpose of the following two lemmas.

**Lemma 2.8.** *Let* $\boldsymbol{\theta}_1^2 := \sum_{T\in\mathcal{T}_h} \theta_{1,T}^2$, *where for each* $T \in \mathcal{T}_h$ *we set*

$$\theta_{1,T}^2 := h_T^2 \left\| \nabla \boldsymbol{u}_h - \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\|_{0,T}^2 + \| \boldsymbol{f}\phi_h + \mathbf{div}\boldsymbol{\sigma}_h \|_{0,T}^2 + \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{D}})} h_e \, \| \boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h \|_{0,e}^2 \, .$$

*Then there exists* $c > 0$, *independent of* $h$, *such that*

$$|E_{h,1}(\boldsymbol{\tau})| \leq c\, \boldsymbol{\theta}_1 \, \|\boldsymbol{\zeta}\|_{\mathbf{div},\Omega} \, . \tag{2.66}$$

*Proof.* The analysis for the first two terms defining $E_{h,1}$ (*cf.* (2.64)) follows as in the proof of [72, Lemma 4.4], after replacing $\Gamma$ by $\Gamma_{\mathrm{D}}$, and then employing the characterization (2.54) - (2.55), the Cauchy-Schwarz inequality, the approximation properties (2.58) and (2.60), and the stability estimate (2.48). In turn, for the corresponding third term it suffices to see that

$$\begin{aligned} &\left| \int_\Omega (\boldsymbol{f}\phi_h + \mathbf{div}\boldsymbol{\sigma}_h) \cdot (\mathbf{I} - \mathcal{P}_h)(\mathbf{div}\,\boldsymbol{\tau})) \right| \\ &\qquad \leq \quad \| \boldsymbol{f}\phi_h + \mathbf{div}\boldsymbol{\sigma}_h \|_{0,\Omega} \, \| \mathbf{div}\,\boldsymbol{\tau} \|_{0,\Omega} \leq \| \boldsymbol{f}\phi_h + \mathbf{div}\boldsymbol{\sigma}_h \|_{0,\Omega} \, \|\boldsymbol{\zeta}\|_{\mathbf{div},\Omega} \, , \end{aligned}$$

which concludes the proof. □

**Lemma 2.9.** *Assume that* $\boldsymbol{u}_{\mathrm{D}} \in \mathbf{H}_0^1(\Gamma_{\mathrm{D}})$, *and let* $\boldsymbol{\theta}_2^2 := \sum_{T\in\mathcal{T}_h} \theta_{2,T}^2$, *where for each* $T \in \mathcal{T}_h$ *we set*

$$\begin{aligned} \theta_{2,T}^2 := h_T^2 \left\| \mathbf{curl} \left\{ \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\} \right\|_{0,T}^2 &+ \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Omega)} h_e \left\| \left[\!\!\left[ \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \boldsymbol{s} \right]\!\!\right] \right\|_{0,e}^2 \\ &+ \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{D}})} h_e \left\| \frac{d\boldsymbol{u}_{\mathrm{D}}}{d\boldsymbol{s}} - \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \, \boldsymbol{s} \right\|_{0,e}^2 . \end{aligned}$$

*Then there exists* $c > 0$, *independent of* $h$, *such that*

$$|E_{h,2}(\boldsymbol{\chi})| \leq c\, \boldsymbol{\theta}_2 \, \|\boldsymbol{\zeta}\|_{\mathbf{div},\Omega} \, . \tag{2.67}$$

*Proof.* We proceed similarly as in the proof of [72, Lemma 4.3]. In fact, using that $\underline{\mathbf{curl}}(\boldsymbol{\chi} - \boldsymbol{\chi}_h)\,\boldsymbol{\nu} = \frac{d}{d\boldsymbol{s}}(\boldsymbol{\chi} - \boldsymbol{\chi}_h)$, noting that $\dfrac{d\boldsymbol{u}_D}{d\boldsymbol{s}} \in \mathbf{L}^2(\Gamma_{\mathrm{D}})$, and then integrating by parts on $\Gamma_{\mathrm{D}}$, we find that

$$\langle \underline{\mathbf{curl}}(\boldsymbol{\chi} - \boldsymbol{\chi}_h)\,\boldsymbol{\nu}, \boldsymbol{u}_{\mathrm{D}} \rangle_{\Gamma_{\mathrm{D}}} = -\langle \boldsymbol{\chi} - \boldsymbol{\chi}_h, \frac{d\boldsymbol{u}_{\mathrm{D}}}{d\boldsymbol{s}} \rangle_{\Gamma_{\mathrm{D}}} = - \sum_{e\in\mathcal{E}_h(\Gamma_{\mathrm{D}})} \int_e (\boldsymbol{\chi} - \boldsymbol{\chi}_h) \frac{d\boldsymbol{u}_{\mathrm{D}}}{d\boldsymbol{s}} .$$

On the other hand, integrating by parts on each $T \in \mathcal{T}_h$, we obtain that

$$\begin{aligned} \int_\Omega \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} : \underline{\mathbf{curl}}(\boldsymbol{\chi} - \boldsymbol{\chi}_h) &= \sum_{T\in\mathcal{T}_h} \left\{ \int_T \mathbf{curl} \left\{ \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\} \cdot (\boldsymbol{\chi} - \boldsymbol{\chi}_h) - \int_{\partial T} \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \, \boldsymbol{s} \cdot (\boldsymbol{\chi} - \boldsymbol{\chi}_h) \right\} \\ &= \sum_{T\in\mathcal{T}_h} \int_T \mathbf{curl} \left\{ \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\} \cdot (\boldsymbol{\chi} - \boldsymbol{\chi}_h) - \sum_{e\in\mathcal{E}_h(\Omega)} \int_e \left[\!\!\left[ \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \boldsymbol{s} \right]\!\!\right] \cdot (\boldsymbol{\chi} - \boldsymbol{\chi}_h) \\ &\quad - \sum_{e\in\mathcal{E}_h(\Gamma_{\mathrm{D}})} \int_e \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \, \boldsymbol{s} \cdot (\boldsymbol{\chi} - \boldsymbol{\chi}_h) - \sum_{e\in\mathcal{E}_h(\Gamma_{\mathrm{N}})} \int_e \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \, \boldsymbol{s} \cdot (\boldsymbol{\chi} - \boldsymbol{\chi}_h) . \end{aligned}$$

Then, replacing the above expressions on the right hand side of (2.65), and using the fact that $\boldsymbol{\chi}|_{\Gamma_{\mathrm{N}}} = \boldsymbol{\chi}_h|_{\Gamma_{\mathrm{N}}} = \mathbf{0}$, we deduce that

$$E_{h,2}(\boldsymbol{\chi}) = \sum_{e\in\mathcal{E}_h(\Omega)} \int_e \left[\!\!\left[ \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \boldsymbol{s} \right]\!\!\right] \cdot (\boldsymbol{\chi} - \boldsymbol{\chi}_h) - \sum_{T\in\mathcal{T}_h} \int_T \mathbf{curl} \left\{ \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\} \cdot (\boldsymbol{\chi} - \boldsymbol{\chi}_h)$$
$$- \sum_{e\in\mathcal{E}_h(\Gamma_{\mathrm{D}})} \int_e \left\{ \frac{d\boldsymbol{u}_{\mathrm{D}}}{d\boldsymbol{s}} - \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\} \cdot (\boldsymbol{\chi} - \boldsymbol{\chi}_h) \,.$$

Next, since $\boldsymbol{\chi}_h := \boldsymbol{\mathcal{I}}_{h,\mathrm{N}}(\boldsymbol{\chi})$, the approximation properties of $\boldsymbol{\mathcal{I}}_{h,\mathrm{N}}$ (*cf.* Lemma 2.5) yield

$$\|\boldsymbol{\chi} - \boldsymbol{\chi}_h\|_{0,T} \leq c_1 \, h_T \, \|\boldsymbol{\chi}\|_{1,\Delta(T)} \qquad \forall\, T \in \mathcal{T}_h, \tag{2.68}$$

and

$$\|\boldsymbol{\chi} - \boldsymbol{\chi}_h\|_{0,e} \leq c_2 \, h_e^{1/2} \, \|\boldsymbol{\chi}\|_{1,\Delta(e)} \qquad \forall\, e \in \mathcal{E}_h. \tag{2.69}$$

In this way, applying the Cauchy-Schwarz inequality to each term in the above expression for $E_{h,2}(\boldsymbol{\chi})$, and making use of (2.68), (2.69), and (2.48), together with the fact that the number of triangles in $\Delta(T)$ and $\Delta(e)$ are bounded, the proof is finished. □

As a consequence of Lemmas 2.8 and 2.9 we conclude the following upper bound for $\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}$.

**Lemma 2.10.** *There exists $c > 0$, independent of $h$, such that*

$$\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'} \leq c \left\{ \boldsymbol{\theta}_1 + \boldsymbol{\theta}_2 \right\}.$$

*Proof.* It follows straightforwardly from (2.63) and the upper bounds (2.66) and (2.67). □

We now observe that the terms $h_T^2 \|\nabla \boldsymbol{u}_h - \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}}\|_{0,T}^2$ and $h_e \|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|_{0,e}^2$, which appear in the definition of $\theta_{1,T}^2$ (*cf.* Lemma 2.8), are dominated by $\|\nabla \boldsymbol{u}_h - \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}}\|_{0,T}^2$ and $\|\boldsymbol{u}_D - \boldsymbol{u}_h\|_{0,e}^2$, respectively, which form part of $\theta_{0,T}^2$ (*cf.* (2.20)). In this way, the reliability estimate (2.18) (*cf.* Theorem 2.1) is a direct consequence of Theorem 2.2, the definition of $\boldsymbol{\theta}_0$ (*cf.* Lemma 2.1), and Lemmas 2.6, 2.8, 2.9, and 2.10.

We end this section by remarking that the assumption (2.17) on the data $\boldsymbol{k}$, $\boldsymbol{u}_{\mathrm{D}}$, and $\boldsymbol{f}$, which, as shown throughout the foregoing analysis, is a key estimate to derive (2.18), is, unfortunately, unverifiable in practice. In fact, while the data are certainly known in advance, the constants $C_3$, $C_6$, and $C_7$ involved in that condition (*cf.* (2.34)), which in turn are expressed in terms of the previous constants $C_1$ and $C_2$ (*cf.* (2.19)), depend all on boundedness and regularity constants of operators, as well as on parameters, some of which are not explicitly calculable, and hence it is not possible to check whether (2.17) is indeed satisfied or not. This is, however, a quite common fact arising in the analysis of many nonlinear problems, and only in very particular cases (usually related to simple geometries of the domain) it could eventually be circumvented.

### 2.3.3 Efficiency

The following theorem is the main result of this section.

**Theorem 2.3.** *There exists a constant $\overline{C}_{\mathtt{eff}} > 0$, which depends only on parameters, $|\boldsymbol{k}|$, $\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}}$, $\|\boldsymbol{f}\|_{\infty,\Omega}$, and other constants, all them independent of $h$, such that*

$$\overline{C}_{\mathtt{eff}}\,\boldsymbol{\theta} \leq \|\phi-\phi_h\|_{1,\Omega} + \|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega} + \|\mathbf{div}(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_{0,\Omega} + \left\|\frac{1}{\mu(\phi)}\boldsymbol{\sigma}^{\mathtt{d}} - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\|_{0,\Omega} + \text{h.o.t.} \quad (2.70)$$

*where* h.o.t. *stands for one or several terms of higher order. Moreover, under the assumption that $\boldsymbol{\sigma} \in \mathbb{L}^4(\Omega)$, there exists a constant $C_{\mathtt{eff}} > 0$, which depends only on parameters, $|\boldsymbol{k}|$, $\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}}$, $\|\boldsymbol{f}\|_{\infty,\Omega}$, $\|\boldsymbol{\sigma}\|_{\mathbb{L}^4(\Omega)}$, and other constants, all them independent of $h$, such that*

$$C_{\mathtt{eff}}\,\boldsymbol{\theta} \leq \|\phi-\phi_h\|_{1,\Omega} + \|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H + \text{h.o.t.} \quad (2.71)$$

Throughout this and the following sections we assume for simplicity that the nonlinear functions $\mu$, $\vartheta$, and $\gamma$ are such that $\dfrac{1}{\mu(\phi_h)}$, $\vartheta(|\nabla\phi_h|)$, $\gamma(\phi_h)$, and hence $\widetilde{\boldsymbol{\sigma}}_h$ as well, are all piecewise polynomials. The same is assumed for the data $\boldsymbol{u}_{\mathrm{D}}$ and $g$. Otherwise, and if $\mu^{-1}$, $\vartheta$, $\gamma$, $\boldsymbol{u}_{\mathrm{D}}$, and $g$ are sufficiently smooth, higher order terms given by the errors arising from suitable polynomial approximations of these expressions and functions would appear in (2.70) and (2.71) (*cf.* Theorem 2.3), which explains the eventual h.o.t. in these inequalities. In this regard, we remark that (2.70) constitutes what we call a *quasi-efficiency* estimate for the global residual error estimator $\boldsymbol{\theta}$ (*cf.* (2.16)). Indeed, the *quasi-efficiency* concept refers here to the fact that the expression appearing on the right hand side of (2.70) is not exactly the error, but part of it plus the nonlinear term given by $\|\frac{1}{\mu(\phi)}\boldsymbol{\sigma}^{\mathtt{d}} - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\|_{0,\Omega}$. However, assuming additionally that $\boldsymbol{\sigma} \in \mathbb{L}^4(\Omega)$, we show at the end of this section that the latter can be bounded by $\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{0,\Omega} + \|\phi-\phi_h\|_{1,\Omega}$, thus yielding the efficiency estimate given by (2.71).

In order to prove (2.70) and (2.71), in what follows we derive suitable upper bounds for the ten terms defining the local error indicator $\theta_T^2$ (*cf.* (2.15)). We first notice, using that $\boldsymbol{f}\phi = -\mathbf{div}\boldsymbol{\sigma}$ in $\Omega$, that there holds

$$\begin{aligned}\|\boldsymbol{f}\phi_h + \mathbf{div}\boldsymbol{\sigma}_h\|_{0,T}^2 &\leq 2\,\|\boldsymbol{f}(\phi-\phi_h)\|_{0,T}^2 + 2\,\|\mathbf{div}(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_{0,T}^2 \\ &\leq 2\,\|\boldsymbol{f}\|_{\infty,\Omega}^2\,\|\phi-\phi_h\|_{0,T}^2 + 2\,\|\mathbf{div}(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_{0,T}^2\,.\end{aligned} \quad (2.72)$$

In addition, since $\nabla\boldsymbol{u} = \frac{1}{\mu(\phi)}\boldsymbol{\sigma}^{\mathtt{d}}$ in $\Omega$, we find that

$$\left\|\nabla\boldsymbol{u}_h - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\|_{0,T}^2 \leq 2\,\|\nabla\boldsymbol{u}-\nabla\boldsymbol{u}_h\|_{0,T}^2 + 2\left\|\frac{1}{\mu(\phi)}\boldsymbol{\sigma}^{\mathtt{d}} - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\|_{0,T}^2. \quad (2.73)$$

Furthermore, employing that $\boldsymbol{u} = \boldsymbol{u}_{\mathrm{D}}$ on $\Gamma_{\mathrm{D}}$ and applying the trace theorem, we obtain that

$$\sum_{e\in\mathcal{E}_h(\Gamma_{\mathrm{D}})} \|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|_{0,e}^2 = \|\boldsymbol{u}-\boldsymbol{u}_h\|_{0,\Gamma_{\mathrm{D}}}^2 \leq c_0^2\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}^2\,, \quad (2.74)$$

where $c_0$ is the norm of the trace operator in $\mathbf{H}^1(\Omega)$.

The upper bounds of the remaining seven terms, which depend on the mesh parameters $h_T$ and $h_e$, will be derived next. We proceed as in [37, 38] (see also [61]), and apply results ultimately based on inverse inequalities (see [40]) and the localization technique introduced in [103], which is based on triangle-bubble and edge-bubble functions. To this end, we now introduce further notations and preliminary results. In fact, given $T \in \mathcal{T}_h$ and $e \in \mathcal{E}_h(T)$, we let $\psi_T$ and $\psi_e$ be the usual triangle-bubble and edge-bubble functions, respectively (see [103, eqs. (1.4) and (1.6)]), which satisfy:

i) $\psi_T \in P_3(T)$, $\mathrm{supp}(\psi_T) \subseteq T$, $\psi_T = 0$ on $\partial T$, and $0 \leq \psi_T \leq 1$ in $T$.

ii) $\psi_e|_T \in P_2(T)$, $\mathrm{supp}(\psi_e) \subseteq \omega_e := \cup\{T' \in \mathcal{T}_h : e \in \mathcal{E}_h(T')\}$, $\psi_e = 0$ on $\partial T \backslash \{e\}$, and $0 \leq \psi_e \leq 1$ in $\omega_e$.

We also recall from [102] that, given $k \in \mathbb{N} \cup \{0\}$, there exists a linear operator $L : C(e) \to C(T)$ that satisfies $L(p) \in P_k(T)$ and $L(p)|_e = p \quad \forall p \in P_k(e)$. A corresponding vectorial version of $L$, that is the component-wise application of $L$, is denoted by $\mathbf{L}$. Additional properties of $\psi_T, \psi_e$ and $L$ are collected in the following Lemma.

**Lemma 2.11.** *Given $k \in \mathbb{N} \cup \{0\}$, there exist positive constants $c_1, c_2, c_3$, and $c_4$, depending only on $k$ and the shape regularity of the triangulations (minimum angle condition), such that for each $T \in \mathcal{T}_h$ and $e \in \mathcal{E}_h(T)$, there hold*

$$\begin{aligned} \|\psi_T\, q\|_{0,T}^2 \;&\leq\; \|q\|_{0,T}^2 \;\leq\; c_1\, \|\psi_T^{1/2}\, q\|_{0,T}^2 \qquad \forall\, q \in P_k(T), \\ \|\psi_e\, L(p)\|_{0,T}^2 \;&\leq\; \|p\|_{0,e}^2 \;\leq\; c_2\, \|\psi_e^{1/2}\, p\|_{0,e}^2 \qquad \forall\, p \in P_k(e), \\ c_3\, h_e\, \|p\|_{0,e}^2 \;&\leq\; \|\psi_e^{1/2}\, L(p)\|_{0,T}^2 \;\leq\; c_4\, h_e\, \|p\|_{0,e}^2 \qquad \forall\, p \in P_k(e). \end{aligned} \tag{2.75}$$

*Proof.* See [102, Lemma 4.1]. □

The following inverse estimate is also needed.

**Lemma 2.12.** *Let $l, m \in \mathbb{N} \cup \{0\}$ such that $l \leq m$. Then, there exists $c > 0$, depending only on $k, l, m$ and the shape regularity of the triangulations, such that for each $T \in \mathcal{T}_h$ there holds*

$$|q|_{m,T} \leq c\, h_T^{l-m}\, |q|_{l,T} \qquad \forall\, q \in P_k(T). \tag{2.76}$$

*Proof.* See [40, Theorem 3.2.6]. □

The following Lemma is required for the terms involving the **curl** operator and the tangential jumps across the edges of $\mathcal{T}_h$. It proofs, which makes use of Lemmas 2.11 and 2.12, can be found in [37].

**Lemma 2.13.** *Let $\boldsymbol{\rho}_h \in \mathbb{L}^2(\Omega)$ be a piecewise polynomial of degree $k \geq 0$ on each $T \in \mathcal{T}_h$. In addition, let $\boldsymbol{\rho} \in \mathbb{L}^2(\Omega)$ be such that $\mathbf{curl}(\boldsymbol{\rho}) = 0$ on each $T \in \mathcal{T}_h$. Then, there exist $c, \widetilde{c} > 0$, independent of $h$, such that*

$$\|\mathbf{curl}(\boldsymbol{\rho}_h)\|_{0,T} \;\leq\; c\, h_T^{-1}\, \|\boldsymbol{\rho} - \boldsymbol{\rho}_h\|_{0,T} \qquad \forall\, T \in \mathcal{T}_h$$

*and*

$$\|[\![\boldsymbol{\rho}_h \boldsymbol{s}_e]\!]\|_{0,e} \;\leq\; \widetilde{c}\, h_e^{-1/2}\, \|\boldsymbol{\rho} - \boldsymbol{\rho}_h\|_{0,\omega_e} \qquad \forall\, e \in \mathcal{E}_h\,.$$

*Proof.* For the first estimate we refer to [37, Lemma 4.3], whereas the second one follows from a slight modification of the proof of [37, Lemma 4.4]. Further details are omitted. □

We now apply Lemma 2.13 to obtain upper bounds for two other terms defining $\theta_T^2$.

**Lemma 2.14.** *There exist $\widetilde{c}_1, \widetilde{c}_2 > 0$, independent of $h$ such that*

$$h_T^2 \left\| \mathbf{curl}\left\{ \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\} \right\|_{0,T}^2 \leq \widetilde{c}_1 \left\| \frac{1}{\mu(\phi)} \boldsymbol{\sigma}^{\mathtt{d}} - \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\|_{0,T}^2 \qquad \forall\, T \in \mathcal{T}_h,$$

$$h_e \left\| \left[\!\!\left[ \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}}\, \boldsymbol{s} \right]\!\!\right] \right\|_{0,e}^2 \leq \widetilde{c}_2 \left\| \frac{1}{\mu(\phi)} \boldsymbol{\sigma}^{\mathtt{d}} - \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\|_{0,\omega_e}^2 \qquad \forall\, e \in \mathcal{E}_h(\Omega).$$

*Proof.* It suffices to apply Lemma 2.13 to $\boldsymbol{\rho}_h := \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}}$ and $\boldsymbol{\rho} := \frac{1}{\mu(\phi)} \boldsymbol{\sigma}^{\mathtt{d}} = \nabla \boldsymbol{u}$. □

**Lemma 2.15.** *There exists $\widetilde{c}_3 > 0$, independent of $h$, such that*

$$h_e \left\| \frac{d\boldsymbol{u}_{\mathrm{D}}}{d\boldsymbol{s}} - \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}}\, \boldsymbol{s} \right\|_{0,e}^2 \leq \widetilde{c}_3 \left\| \frac{1}{\mu(\phi)} \boldsymbol{\sigma}^{\mathtt{d}} - \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\|_{0,T_e}^2 \qquad \forall\, e \in \mathcal{E}_h(\Gamma_{\mathrm{D}}). \tag{2.77}$$

*Proof.* We proceed similarly as in the proof of [72, Lemma 4.15], by replacing $\mathbf{g}$, $\Gamma$, and $\frac{1}{\mu} \boldsymbol{\sigma}_h^{\mathtt{d}}$ by $\boldsymbol{u}_D$, $\Gamma_{\mathrm{D}}$, and $\frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}}$, respectively. □

Finally, it only remains to provide upper bounds for the three terms completing the definition of the local error indicator $\theta_T^2$ (*cf.* (2.15)). This requires, however, the preliminary result given by the following *a priori* estimate for the error $\|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{0,T}^2$.

**Lemma 2.16.** *There exists $C > 0$, depending on $\vartheta_1$, $\vartheta_2$, $L_\gamma$ (*cf.* (2.4), (2.6)), and $|\boldsymbol{k}|$, such that*

$$\|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{0,T}^2 \leq C \left\{ \|\phi - \phi_h\|_{1,T}^2 + \|\boldsymbol{u}(\phi - \phi_h)\|_{0,T}^2 + \|\phi_h(\boldsymbol{u} - \boldsymbol{u}_h)\|_{0,T}^2 \right\}. \tag{2.78}$$

*Proof.* According to the definitions of $\widetilde{\boldsymbol{\sigma}}$ (*cf.* (1.5)) and $\widetilde{\boldsymbol{\sigma}}_h$ (*cf.* Section 2.3.1), and applying the triangle inequality, we obtain that

$$\begin{aligned} \|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{0,T}^2 \leq 2 \Big\{ & \|\vartheta(|\nabla\phi|)\nabla\phi - \vartheta(|\nabla\phi_h|)\nabla\phi_h\|_{0,T}^2 + 2\,\|\boldsymbol{k}(\gamma(\phi) - \gamma(\phi_h))\|_{0,T}^2 \\ & + \; 4\,\|\boldsymbol{u}(\phi - \phi_h)\|_{0,T}^2 + 4\,\|\phi_h(\boldsymbol{u} - \boldsymbol{u}_h)\|_{0,T}^2 \Big\}. \end{aligned} \tag{2.79}$$

We now recall from [76, Theorem 3.8] that the nonlinear operator induced by the first term defining $\mathbf{A}_{\boldsymbol{u}}$ (*cf.* (2.29)) is Lipschitz-continuous with constant $L := \max\{\vartheta_2, 2\vartheta_2 - \vartheta_1\}$. In this way, applying the aforementioned Lipschitz continuity, but restricted to each triangle $T \in \mathcal{T}_h$ instead of $\Omega$, and using the Lipschitz continuity assumption for $\gamma$ (*cf.* (2.6)), we deduce from (2.79) that

$$\begin{aligned} \|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{0,T}^2 \leq 2 \Big\{ & L^2\,\|\nabla\phi - \nabla\phi_h\|_{0,T}^2 + 2L_\gamma^2\,|\boldsymbol{k}|^2\,\|\phi - \phi_h\|_{0,T}^2 \\ & + \; 4\,\|\boldsymbol{u}(\phi - \phi_h)\|_{0,T}^2 + 4\,\|\phi_h(\boldsymbol{u} - \boldsymbol{u}_h)\|_{0,T}^2 \Big\}, \end{aligned} \tag{2.80}$$

which readily yields (2.78) and ends the proof. □

Having proved the previous result we now establish the efficiency estimates given by the following three lemmas.

**Lemma 2.17.** *There exists $\widetilde{c}_4 > 0$, which depends only on $L$, $L_\gamma$, $|\boldsymbol{k}|$, and other constants, all them independent of $h$, such that*

$$h_T^2 \, \|g \, + \, \mathrm{div}\widetilde{\boldsymbol{\sigma}}_h\|_{0,T}^2 \, \leq \, \widetilde{c}_4 \left\{ \|\phi - \phi_h\|_{1,T}^2 \, + \, \|\boldsymbol{u}(\phi - \phi_h)\|_{0,T}^2 \, + \, \|\phi_h(\boldsymbol{u} - \boldsymbol{u}_h)\|_{0,T}^2 \right\}. \tag{2.81}$$

*Proof.* We proceed as in the proof of [15, Lemma 4.4]. In fact, given $T \in \mathcal{T}_h$ we first observe, using that $\mathrm{div}\widetilde{\boldsymbol{\sigma}} = -g$ in $\Omega$, and integrating by parts, that

$$\|g + \mathrm{div}\widetilde{\boldsymbol{\sigma}}_h\|_{0,T}^2 \, \leq \, c_1 \, \|\psi_T^{1/2} \, (g + \mathrm{div}\widetilde{\boldsymbol{\sigma}}_h)\|_{0,T}^2 \, = \, -c_1 \int_T \nabla\big(\psi_T \, (g + \mathrm{div}\widetilde{\boldsymbol{\sigma}}_h)\big) \cdot (\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h) \, .$$

Next, the Cauchy-Schwarz inequality, the inverse estimate (2.76), the fact that $0 \leq \psi_T \leq 1$, and the triangle inequality imply that

$$\|g + \mathrm{div}\widetilde{\boldsymbol{\sigma}}_h\|_{0,T}^2 \, \leq \, c_1 \, |\psi_T \, (g + \mathrm{div}\widetilde{\boldsymbol{\sigma}}_h)|_{1,T} \, \|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{0,T} \, \leq \, C \, h_T^{-1} \, \|g + \mathrm{div}\widetilde{\boldsymbol{\sigma}}_h\|_{0,T} \, \|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{0,T} \, ,$$

which gives

$$\|g + \mathrm{div}\widetilde{\boldsymbol{\sigma}}_h\|_{0,T} \, \leq \, C \, h_T^{-1} \, \|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{0,T} \, .$$

The foregoing inequality and (2.78) (*cf.* Lemma 2.16) imply (2.81) and complete the proof. □

**Lemma 2.18.** *There exists $\widetilde{c}_5 > 0$, which depends only on $L$, $L_\gamma$, $|\boldsymbol{k}|$, and other constants, all them independent of $h$, such that for each $e \in \mathcal{E}_h(\Omega)$ there holds*

$$h_e \, \|[\![\widetilde{\boldsymbol{\sigma}}_h \cdot \boldsymbol{\nu}_e]\!]\|_{0,e}^2 \, \leq \, \widetilde{c}_5 \sum_{T \subseteq \omega_e} \left\{ \|\phi - \phi_h\|_{1,T}^2 \, + \, \|\boldsymbol{u}(\phi - \phi_h)\|_{0,T}^2 \, + \, \|\phi_h(\boldsymbol{u} - \boldsymbol{u}_h)\|_{0,T}^2 \right\}, \tag{2.82}$$

*where $\omega_e$ is the union of the two triangles in $\mathcal{T}_h$ having $e$ as an edge.*

*Proof.* Proceeding analogously as in the proof of [15, Lemma 4.5], we find that

$$h_e \, \|[\![\widetilde{\boldsymbol{\sigma}}_h \cdot \boldsymbol{\nu}_e]\!]\|_{0,e}^2 \, \leq \, c \sum_{T \subseteq \omega_e} \left\{ h_T^2 \, \|g \, + \, \mathrm{div}\widetilde{\boldsymbol{\sigma}}_h\|_{0,T}^2 \, + \, \|\widetilde{\boldsymbol{\sigma}}_h \, - \, \widetilde{\boldsymbol{\sigma}}\|_{0,T}^2 \right\},$$

which, together with (2.78) and (2.81) (*cf.* Lemmas 2.16 and 2.17), yields (2.82) and ends the proof.

□

**Lemma 2.19.** *There exists $\widetilde{c}_6 > 0$, which depends only on $L$, $L_\gamma$, $|\boldsymbol{k}|$, and other constants, all them independent of $h$, such that for each $e \in \mathcal{E}_h(\Gamma_{\mathrm{N}})$ there holds*

$$h_e \, \|\widetilde{\boldsymbol{\sigma}}_h \cdot \boldsymbol{\nu}\|_{0,e}^2 \, \leq \, \widetilde{c}_6 \left\{ \|\phi - \phi_h\|_{1,T}^2 \, + \, \|\boldsymbol{u}(\phi - \phi_h)\|_{0,T}^2 \, + \, \|\phi_h(\boldsymbol{u} - \boldsymbol{u}_h)\|_{0,T}^2 \right\}, \tag{2.83}$$

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

*Proof.* Following a similar reasoning to the proof of [15, Lemma 4.6], we find that

$$h_e \, \|\widetilde{\boldsymbol{\sigma}}_h \cdot \boldsymbol{\nu}\|_{0,e}^2 \, \leq \, c \left\{ h_T^2 \, \|g \, + \, \mathrm{div}\widetilde{\boldsymbol{\sigma}}_h\|_{0,T}^2 \, + \, \|\widetilde{\boldsymbol{\sigma}}_h \, - \, \widetilde{\boldsymbol{\sigma}}\|_{0,T}^2 \right\},$$

which, thanks again to (2.78) and (2.81), provides (2.83) and ends the proof. □

In order to complete the global efficiency given by (2.70), we now need to estimate the terms $\|\boldsymbol{u}(\phi-\phi_h)\|^2_{0,T}$ and $\|\phi_h(\boldsymbol{u}-\boldsymbol{u}_h)\|^2_{0,T}$ appearing in the upper bounds provided by the last three lemmas. In fact, applying Cauchy-Schwarz's inequality, the compactness (and hence continuity) of the injections $\mathrm{i}:\mathrm{H}^1(\Omega)\to\mathrm{L}^4(\Omega)$ and $\mathbf{i}:\mathbf{H}^1(\Omega)\to\mathbf{L}^4(\Omega)$ (*cf.* [1, Theorem 6.3], [92, Theorem 1.3.5]), and the *a priori* bound for $\|\boldsymbol{u}\|_{1,\Omega}$ given by (2.14), we find that

$$\begin{aligned}\sum_{T\in\mathcal{T}_h}\|\boldsymbol{u}(\phi-\phi_h)\|^2_{0,T} &\leq \sum_{T\in\mathcal{T}_h}\|\boldsymbol{u}\|^2_{\mathbf{L}^4(T)}\,\|\phi-\phi_h\|^2_{\mathrm{L}^4(T)}\\ &\leq \|\boldsymbol{u}\|^2_{\mathbf{L}^4(\Omega)}\,\|\phi-\phi_h\|^2_{\mathrm{L}^4(\Omega)} \leq C\,\|\phi-\phi_h\|^2_{1,\Omega}\,,\end{aligned}\tag{2.84}$$

where $C$ is a positive constant, independent of $h$, that depends only on $\|i\|$, $\|\mathbf{i}\|$, $\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}}$, $\|\boldsymbol{f}\|_{\infty,\Omega}$, and $r$ (*cf.* (2.13)). Similar arguments allow to establish the existence of another constant $C>0$, also independent of $h$, and depending now on $\|i\|$, $\|\mathbf{i}\|$, and $r$, such that

$$\sum_{T\in\mathcal{T}_h}\|\phi_h(\boldsymbol{u}-\boldsymbol{u}_h)\|^2_{0,T} \leq C\,\|\boldsymbol{u}-\boldsymbol{u}_h\|^2_{1,\Omega}\,.\tag{2.85}$$

Consequently, it is not difficult to see that (2.70) follows straightforwardly from (2.72), (2.73), (2.74), Lemmas 2.14, 2.15, 2.17, 2.18, and 2.19, and the final estimates given by (2.84) and (2.85). Furthermore, adding and subtracting a suitable term, using the lower bound (*cf.* (2.3)) and the Lipschitz continuity (*cf.* (2.5)) of $\mu$, and applying the boundedness of $\boldsymbol{\tau}\to\boldsymbol{\tau}^{\mathtt{d}}$, we find that

$$\left\|\frac{1}{\mu(\phi)}\boldsymbol{\sigma}^{\mathtt{d}}-\frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\|_{0,\Omega}\leq\frac{1}{\mu_1}\,\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{0,\Omega}+\frac{L_\mu}{\mu_1^2}\,\|(\phi-\phi_h)\boldsymbol{\sigma}\|_{0,\Omega}\,,\tag{2.86}$$

from which, assuming now that $\boldsymbol{\sigma}\in\mathbb{L}^4(\Omega)$, and estimating $\|(\phi-\phi_h)\boldsymbol{\sigma}\|_{0,\Omega}$ almost verbatim as we derived (2.84) and (2.85), we arrive at (2.71), thus concluding the proof of Theorem 2.3.

## 2.4 A second a posteriori error estimator

In this section we introduce and analyze another *a posteriori* error estimator for our augmented mixed-primal finite element scheme (2.12), which is not based on the Helmholtz decomposition. More precisely, this second estimator arises simply from a different way of bounding $\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}$ in the preliminary estimate for the total error given by (2.42) (*cf.* Theorem 2.2). Then, with the same notations and discrete spaces introduced in Sections 2.2 and 2.3, we now set for each $T\in\mathcal{T}_h$ the local error indicator

$$\begin{aligned}\widetilde{\theta}^2_T :=\ &\|\boldsymbol{f}\phi_h+\mathbf{div}\boldsymbol{\sigma}_h\|^2_{0,T}+\left\|\nabla\boldsymbol{u}_h-\frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\|^2_{0,T}+h_T^2\,\|g+\mathrm{div}\widetilde{\boldsymbol{\sigma}}_h\|^2_{0,T}\\ &+\sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Omega)}h_e\,\|[\![\widetilde{\boldsymbol{\sigma}}_h\cdot\boldsymbol{\nu}_e]\!]\|^2_{0,e}+\sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{N}})}h_e\,\|\widetilde{\boldsymbol{\sigma}}_h\cdot\boldsymbol{\nu}\|^2_{0,e}+\sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{D}})}\|\boldsymbol{u}_{\mathrm{D}}-\boldsymbol{u}_h\|^2_{0,e}\,,\end{aligned}$$

and define the following global residual error estimator

$$\widetilde{\boldsymbol{\theta}}^{\,2}:=\sum_{T\in\mathcal{T}_h}\widetilde{\theta}^2_T+\|\boldsymbol{u}_{\mathrm{D}}-\boldsymbol{u}_h\|^2_{1/2,\Gamma_{\mathrm{D}}}\,.\tag{2.87}$$

In what follows we establish *quasi-local* reliability and efficiency for the estimator $\widetilde{\boldsymbol{\theta}}$. The name *quasi-local* refers here to the fact that the last term defining $\widetilde{\boldsymbol{\theta}}$ can not be decomposed into local

quantities associated to each triangle $T \in \mathcal{T}_h$ (unless it is either conveniently bounded or previously modified, as we will see below).

**Theorem 2.4.** *Assume that* $\boldsymbol{u}_{\mathrm{D}} \in \mathbf{H}^1_0(\Gamma_{\mathrm{D}})$ *and*

$$C_3\,|\boldsymbol{k}| \,+\, C_6\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma_{\mathrm{D}}} \,+\, C_7\,\|\boldsymbol{f}\|_{\infty,\Omega} \,<\, \frac{1}{2},$$

*where* $C_3$, $C_6$ *and* $C_7$ *are the constants given in* (2.34). *Then, there exists a constant* $\widetilde{C}_{\mathtt{rel}} > 0$, *which depends only on* $\|\boldsymbol{u}_D\|_{1/2+\varepsilon,\Gamma_{\mathrm{D}}}$, $\|\boldsymbol{f}\|_{\infty,\Omega}$ *and other constants, all them independent of* $h$, *such that*

$$\|\phi - \phi_h\|^2_{1,\Omega} \,+\, \|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|^2_H \,\leq\, \widetilde{C}_{\mathtt{rel}}\,\widetilde{\boldsymbol{\theta}}^2\,. \tag{2.88}$$

*Proof.* As mentioned at the beginning of this section, the proof reduces basically to derive another upper bound for $\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}$. Indeed, Integrating by parts the third term defining $E_h$ (*cf.* (2.22)), and then using the homogeneous Neumann boundary condition on $\Gamma_{\mathrm{N}}$, we find that for each $\boldsymbol{\zeta} \in \mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)$ there holds

$$E_h(\boldsymbol{\zeta}) \,=\, \langle \boldsymbol{\zeta}\,\boldsymbol{\nu}, \boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\rangle_{\Gamma_{\mathrm{D}}} \,+\, \int_\Omega \Big(\nabla \boldsymbol{u}_h \,-\, \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}^{\mathtt{d}}_h\Big) : \boldsymbol{\zeta} \,-\, \kappa_2 \int_\Omega (\boldsymbol{f}\phi_h \,+\, \mathbf{div}\boldsymbol{\sigma}_h)\cdot \mathbf{div}\boldsymbol{\zeta}\,,$$

from which, applying the Cauchy-Schwarz inequality, we readily deduce that

$$\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'} \,\leq\, C\,\left\{\|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|_{1/2,\Gamma_{\mathrm{D}}} \,+\, \Big\|\nabla \boldsymbol{u}_h \,-\, \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}^{\mathtt{d}}_h\Big\|_{0,\Omega} \,+\, \|\boldsymbol{f}\phi_h + \mathbf{div}\boldsymbol{\sigma}_h\|_{0,\Omega}\right\}\,, \tag{2.89}$$

where $C$ is a positive constant independent of $h$. In this way, replacing (2.89) back into (2.42) (*cf.* Theorem 2.2), and employing again the upper bound for $\|\widetilde{E}_h\|_{\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'}$ (*cf.* Lemma 2.6), and the definition of $\boldsymbol{\theta}_0$ (*cf.* Lemma 2.1), we obtain (2.88) and finish the proof. □

**Theorem 2.5.** *There exists a constant* $C^*_{\mathtt{eff}} > 0$, *which depends only on parameters,* $|\boldsymbol{k}|$, $\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}}$, $\|\boldsymbol{f}\|_{\infty,\Omega}$, *and other constants, all them independent of* $h$, *such that*

$$C^*_{\mathtt{eff}}\,\widetilde{\boldsymbol{\theta}}^2 \,\leq\, \|\phi - \phi_h\|^2_{1,\Omega} + \|\boldsymbol{u} - \boldsymbol{u}_h\|^2_{1,\Omega} + \|\mathbf{div}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)\|^2_{0,\Omega} + \Big\|\frac{1}{\mu(\phi)}\boldsymbol{\sigma}^{\mathtt{d}} - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}^{\mathtt{d}}_h\Big\|^2_{0,\Omega} + \text{h.o.t.} \tag{2.90}$$

*where* h.o.t. *stands for one or several terms of higher order. Moreover, assuming* $\boldsymbol{\sigma} \in \mathbb{L}^4(\Omega)$, *there exists a constant* $\widetilde{C}_{\mathtt{eff}} > 0$, *which depends only on parameters,* $|\boldsymbol{k}|$, $\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}}$, $\|\boldsymbol{f}\|_{\infty,\Omega}$, $\|\boldsymbol{\sigma}\|_{\mathbb{L}^4(\Omega)}$, *and other constants, all them independent of* $h$, *such that*

$$\widetilde{C}_{\mathtt{eff}}\,\widetilde{\boldsymbol{\theta}}^2 \,\leq\, \|\phi - \phi_h\|^2_{1,\Omega} \,+\, \|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|^2_H \,+\, \text{h.o.t.} \tag{2.91}$$

*Proof.* We simply observe, thanks to the trace theorem in $\mathbf{H}^1(\Omega)$, that there exists $c > 0$, depending on $\Gamma_{\mathrm{D}}$ and $\Omega$, such that

$$\|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|^2_{1/2,\Gamma_{\mathrm{D}}} \,\leq\, c\,\|\boldsymbol{u} - \boldsymbol{u}_h\|^2_{1,\Omega}\,. \tag{2.92}$$

The rest of the arguments are contained in the proof of Theorem 2.3 (*cf.* Section 2.3.3), and hence we omit further details. □

At this point we remark that the eventual use of $\widetilde{\boldsymbol{\theta}}$ (*cf.* (2.87)) in an adaptive algorithm solving (2.12) would be discouraged by the non-local character of the expression $\|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|^2_{1/2,\Gamma_{\mathrm{D}}}$. In order to circumvent this situation, we now apply an interpolation argument and replace this term by a suitable upper bound, which yields a reliable and fully local *a posteriori* error estimate.

**Theorem 2.6.** *Assume that* $\boldsymbol{u}_{\mathrm{D}} \in \mathbf{H}^1_0(\Gamma_{\mathrm{D}})$ *and that*

$$C_3\,|\boldsymbol{k}| \,+\, C_6\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma_{\mathrm{D}}} \,+\, C_7\,\|\boldsymbol{f}\|_{\infty,\Omega} \,<\, \frac{1}{2},$$

*where* $C_3$, $C_6$ *and* $C_7$ *are given in* (2.34). *In turn, let* $\widehat{\boldsymbol{\theta}}^2 := \sum_{T\in\mathcal{T}_h} \widehat{\theta}^2_T$, *where for each* $T \in \mathcal{T}_h$ *we set*

$$\widehat{\theta}^2_T \,:=\, \|\boldsymbol{f}\phi_h \,+\, \mathbf{div}\boldsymbol{\sigma}_h\|^2_{0,T} \,+\, \left\|\nabla\boldsymbol{u}_h \,-\, \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}^{\mathtt{d}}_h\right\|^2_{0,T} \,+\, h^2_T\,\|g \,+\, \mathrm{div}\widetilde{\boldsymbol{\sigma}}_h\|^2_{0,T}$$
$$+ \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Omega)} h_e\,\|[\![\widetilde{\boldsymbol{\sigma}}_h\cdot\boldsymbol{\nu}_e]\!]\|^2_{0,e} \,+\, \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{N}})} h_e\,\|\widetilde{\boldsymbol{\sigma}}_h\cdot\boldsymbol{\nu}\|^2_{0,e} \,+\, \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{D}})} \|\boldsymbol{u}_{\mathrm{D}}-\boldsymbol{u}_h\|^2_{1,e}\,.$$

*Then, there exists a constant* $\widehat{C}_{\mathtt{rel}} > 0$, *which depends only on parameters,* $\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma_{\mathrm{D}}}$, $\|\boldsymbol{f}\|_{\infty,\Omega}$, *and other constants, all them independent of* $h$, *such that*

$$\|\phi-\phi_h\|^2_{1,\Omega} \,+\, \|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|^2_H \,\leq\, \widehat{C}_{\mathtt{rel}}\,\widehat{\boldsymbol{\theta}}^2\,. \tag{2.93}$$

*Proof.* It reduces to bound $\|\boldsymbol{u}_{\mathrm{D}}-\boldsymbol{u}_h\|_{1/2,\Gamma_{\mathrm{D}}}$. In fact, since $H^{1/2}(\Gamma_{\mathrm{D}})$ is the interpolation space with index $1/2$ between $H^1(\Gamma_{\mathrm{D}})$ and $L^2(\Gamma_{\mathrm{D}})$, there exists a constant $c_{\mathrm{D}} > 0$, depending on $\Gamma_{\mathrm{D}}$, such that

$$\begin{aligned}\|\boldsymbol{u}_{\mathrm{D}}-\boldsymbol{u}_h\|^2_{1/2,\Gamma_{\mathrm{D}}} &\leq c_{\mathrm{D}}\,\|\boldsymbol{u}_{\mathrm{D}}-\boldsymbol{u}_h\|_{0,\Gamma_{\mathrm{D}}}\,\|\boldsymbol{u}_{\mathrm{D}}-\boldsymbol{u}_h\|_{1,\Gamma_{\mathrm{D}}}\\ &\leq c_{\mathrm{D}}\,\|\boldsymbol{u}_D-\boldsymbol{u}_h\|^2_{1,\Gamma_{\mathrm{D}}} \,=\, c_{\mathrm{D}}\sum_{e\in\mathcal{E}_h(\Gamma_{\mathrm{D}})}\|\boldsymbol{u}_{\mathrm{D}}-\boldsymbol{u}_h\|^2_{1,e}\,,\end{aligned} \tag{2.94}$$

which, together with (2.88), implies (2.93) and finishes the proof. □

The same remark stated at the end of Section 2.3.2 concerning the assumption (2.17) (which is also required in Theorem 2.6) is valid here.

## 2.5 Numerical tests

This section serves to illustrate the properties of the estimators introduced in Sections 2.3-2.4. The domain of each example to be considered below is discretized into a series of nested uniform triangulations, where errors and experimental convergence rates will be computed as usual

$$\begin{aligned}&e(\boldsymbol{\sigma}) := \|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathbf{div},\Omega}, \quad e(\boldsymbol{u}) := \|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}, \quad e(\phi) := \|\phi-\phi_h\|_{1,\Omega},\\ &r(\boldsymbol{\sigma}) := \frac{\log(e(\boldsymbol{\sigma})/\hat{e}(\boldsymbol{\sigma}))}{\log(h/\hat{h})}, \quad r(\boldsymbol{u}) := \frac{\log(e(\boldsymbol{u})/\hat{e}(\boldsymbol{u}))}{\log(h/\hat{h})}, \quad r(\phi) := \frac{\log(e(\phi)/\hat{e}(\phi))}{\log(h/\hat{h})},\end{aligned}$$

with $e$ and $\hat{e}$ denoting errors associated to two consecutive meshes of sizes $h$ and $\hat{h}$, respectively. In addition, the total error, the modified error suggested by (2.70) and (2.90), and the effectivity and quasi-effectivity indexes associated to a given global estimator $\boldsymbol{\eta}$ are defined, respectively, as

$$\mathtt{e} = \left\{[e(\boldsymbol{\sigma})]^2+[e(\boldsymbol{u})]^2+[e(\phi)]^2\right\}^{1/2}, \qquad \mathtt{eff}(\boldsymbol{\eta}) = \frac{\mathtt{e}}{\boldsymbol{\eta}},$$
$$\mathtt{m} = \left\{[e(\boldsymbol{u})]^2+[e(\phi)]^2+\|\mathbf{div}\boldsymbol{\sigma}-\mathbf{div}\boldsymbol{\sigma}_h\|^2_{0,\Omega}+\left\|\frac{\boldsymbol{\sigma}^{\mathtt{d}}}{\mu(\phi)}-\frac{\boldsymbol{\sigma}^{\mathtt{d}}_h}{\mu(\phi_h)}\right\|^2_{0,\Omega}\right\}^{1/2}, \qquad \mathtt{qeff}(\boldsymbol{\eta}) = \frac{\mathtt{m}}{\boldsymbol{\eta}}.$$

| $h$ | $e(\boldsymbol{\sigma})$ | $r(\boldsymbol{\sigma})$ | $e(\boldsymbol{u})$ | $r(\boldsymbol{u})$ | $e(\phi)$ | $r(\phi)$ | $i_N$ | $i_P$ | $\mathtt{eff}(\boldsymbol{\theta})$ | $\mathtt{qeff}(\boldsymbol{\theta})$ | $\mathtt{eff}(\widetilde{\boldsymbol{\theta}})$ | $\mathtt{qeff}(\widetilde{\boldsymbol{\theta}})$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Augmented $\mathrm{RT}_0 - P_1 - P_1$ scheme | | | | | | | | | | | | |
| 0.7071 | 99.1853 | – | 10.1168 | – | 1.5980 | – | 4 | 12 | 1.8570 | 1.8603 | 1.8740 | 1.8773 |
| 0.4714 | 83.1416 | 0.4351 | 8.6706 | 0.3804 | 1.1558 | 0.7990 | 3 | 15 | 1.5604 | 1.5627 | 1.5950 | 1.5973 |
| 0.2828 | 56.1085 | 0.7698 | 6.1721 | 0.6653 | 0.7191 | 0.9288 | 5 | 17 | 1.2142 | 1.2158 | 1.2474 | 1.2490 |
| 0.1571 | 31.7872 | 0.9667 | 2.9676 | 1.2458 | 0.4035 | 0.9828 | 4 | 16 | 1.0694 | 1.0707 | 1.1021 | 1.1034 |
| 0.0831 | 16.7731 | 1.0051 | 1.3190 | 1.2749 | 0.2136 | 0.9998 | 4 | 16 | 1.0088 | 1.0101 | 1.0409 | 1.0421 |
| 0.0428 | 8.5927 | 1.0083 | 0.6226 | 1.1316 | 0.1100 | 1.0009 | 4 | 16 | 0.9861 | 0.9873 | 1.0180 | 1.0193 |
| 0.0217 | 4.3466 | 1.0053 | 0.3071 | 1.0422 | 0.0558 | 1.0003 | 5 | 16 | 0.9777 | 0.9789 | 1.0097 | 1.0110 |
| Augmented $\mathrm{RT}_1 - P_2 - P_2$ scheme | | | | | | | | | | | | |
| 0.7071 | 51.6128 | – | 7.1534 | – | 0.4563 | – | 5 | 13 | 1.0353 | 1.0370 | 1.1487 | 1.1506 |
| 0.4714 | 29.6423 | 1.3677 | 3.2044 | 1.9805 | 0.2157 | 1.8475 | 4 | 14 | 0.9638 | 0.9650 | 1.0549 | 1.0562 |
| 0.2828 | 15.2131 | 1.3058 | 1.2003 | 1.9222 | 0.0799 | 1.9439 | 4 | 16 | 0.9573 | 0.9580 | 1.0198 | 1.0205 |
| 0.1571 | 4.9972 | 1.8940 | 0.3289 | 2.2022 | 0.0249 | 1.9837 | 5 | 15 | 0.9525 | 0.9532 | 1.0088 | 1.0095 |
| 0.0831 | 1.4251 | 1.9726 | 0.0868 | 2.0947 | 0.0070 | 1.9944 | 5 | 16 | 0.9515 | 0.9522 | 1.0045 | 1.0052 |
| 0.0428 | 0.3799 | 1.9928 | 0.0225 | 2.0320 | 0.0018 | 1.9984 | 4 | 16 | 0.9488 | 0.9495 | 1.0003 | 1.0011 |
| 0.0217 | 0.0980 | 1.9983 | 0.0057 | 2.0070 | 0.0004 | 1.9996 | 5 | 16 | 0.9396 | 0.9403 | 0.9895 | 0.9902 |

Table 2.1: Test 1: convergence history, average Newton iteration count, Picard steps to reach the desired tolerance, effectivity and quasi-effectivity indexes for the mixed–primal $\mathrm{RT}_k - P_{k+1} - P_{k+1}$ approximations of concentration, Cauchy stress, and velocity, with $k = 0, 1$.

According to the coupling structure of the scheme (2.12), the linearization of the coupled problem can follow a Newton method solving the nonlinear transport problem, nested within a Picard iteration to establish the coupling with the Stokes problem. This procedure requires the computation of the Gâteaux derivative (2.32). When the residuals from Newton-Raphson and Picard iterations reach the tolerances $\varepsilon_{\mathrm{N}} = 1\text{e-}8$ and $\varepsilon_{\mathrm{P}} = 1\text{e-}7$, respectively, the algorithms are terminated. The unsymmetric multi-frontal direct solver for sparse matrices (UMFPACK) is used to solve the linear systems appearing at each linearization step.

In a first example, the following exact solutions to system (2.1) are considered

$$\phi(x_1, x_2) = b - b\exp(-x_1(x_1-1)x_2(x_2-1)), \quad \boldsymbol{u}(x_1, x_2) = \begin{pmatrix} \sin(2\pi x_1)\cos(2\pi x_2) \\ -\cos(2\pi x_1)\sin(2\pi x_2) \end{pmatrix},$$

$$\boldsymbol{\sigma}(x_1, x_2) = \mu(\phi)\,\nabla\boldsymbol{u} - \mu(\phi)\frac{\partial u_1}{\partial x_1}\mathbb{I},$$

defined on the unit square $\Omega = (0,1)^2$ and satisfying the first and third conditions of (2.2) on the whole boundary $\Gamma_{\mathrm{D}} = \partial\Omega$. The data $\boldsymbol{u}_{\mathrm{D}}, \boldsymbol{f}, g$ are constructed with these manufactured exact solutions, and the involved coefficients in the equations (and in the solutions) are $\boldsymbol{k} = (0,-1)^T$, $\mu(\phi) = (1-c\phi)^{-2}$, $\gamma(\phi) = c\phi(1-c\phi)^2$, $\vartheta(|\nabla\phi|) = m_1 + m_2(1+|\nabla\phi|^2)^{m_3/2-1}$, with $b = 15, c = m_1 = m_2 = 1/2, m_3 = 3/2$. These values imply $\mu_1 = 0.99$, $\mu_2 = 3.35$, and consequently the stabilization parameters adopt the values $\kappa_1 = \mu_1^2/\mu_2 = 0.2976$, $\kappa_2 = 1/\mu_2 = 0.2985$, and $\kappa_3 = \kappa_1/2 = 0.1488$.

The manufactured solutions on the considered (convex) domain are smooth, and the *a posteriori*

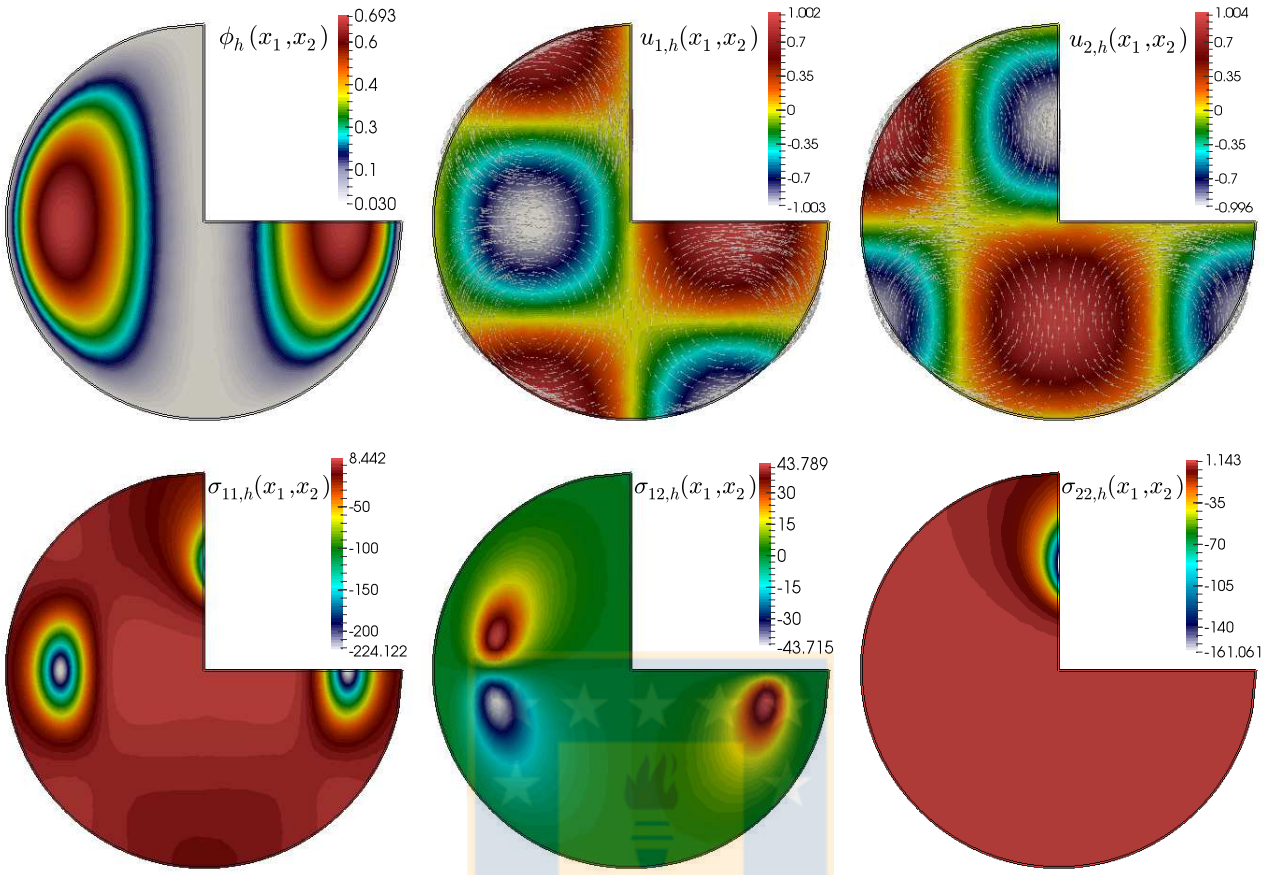


Figure 2.2: Test 2: approximate solutions obtained with the lowest order method, after six steps of adaptive mesh refinement following the second indicator $\widetilde{\boldsymbol{\theta}}$. Concentration, velocity components, and stress components are depicted.

error indicators show effectivity (and quasi-effectivity) indexes close to one in all studied cases. This behavior can be observed in Table 2.1, where errors in different norms indicate optimal convergence rates for the two lowest order methods ($k = 0, 1$). We also show the average number of Newton steps to achieve the tolerance $\varepsilon_{\mathrm{N}}$ and the total Picard iteration count at each refinement level. The subsequent examples will be restricted to the lowest order method $k = 0$.

Our second test focuses on the case where, under uniform mesh refinement, the convergence rates are affected by the loss of regularity of the exact solutions. The problem setting is as follows: the domain is taken as the non-convex *pacman*-shaped domain $\Omega = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1\} \setminus (0, 1)^2$, where an exact solution to (2.1) is given by the same velocity as in the previous test, while concentration and Cauchy stress now read

$$\begin{aligned} \phi(x_1, x_2) &= b - b \exp(x_1^2(x_1^2 + x_2^2 - 1)) \,, \\ \boldsymbol{\sigma}(x_1, x_2) &= \mu(\phi)\, \nabla \boldsymbol{u} - \left[ \mu(\phi) \frac{\partial u_2}{\partial x_2} + \frac{x_2}{((x_1 - a_1)^2 + (x_2 - a_2)^2)^2} \right] \mathbb{I} \,. \end{aligned} \tag{2.95}$$

Now the boundary is indeed split into $\Gamma_{\mathrm{N}} = (0, 1) \times \{0\}$ (the horizontal segment of $\partial\Omega$) and $\Gamma_{\mathrm{D}} = \partial\Omega \backslash \Gamma_{\mathrm{N}}$ (the arch and vertical borders of the domain), and the only difference with respect to (2.1) is that a non-homogeneous concentration flux $\widetilde{\boldsymbol{\sigma}} \cdot \boldsymbol{\nu} = j$ is imposed on $\Gamma_{\mathrm{N}}$, where $j$ is manufactured according

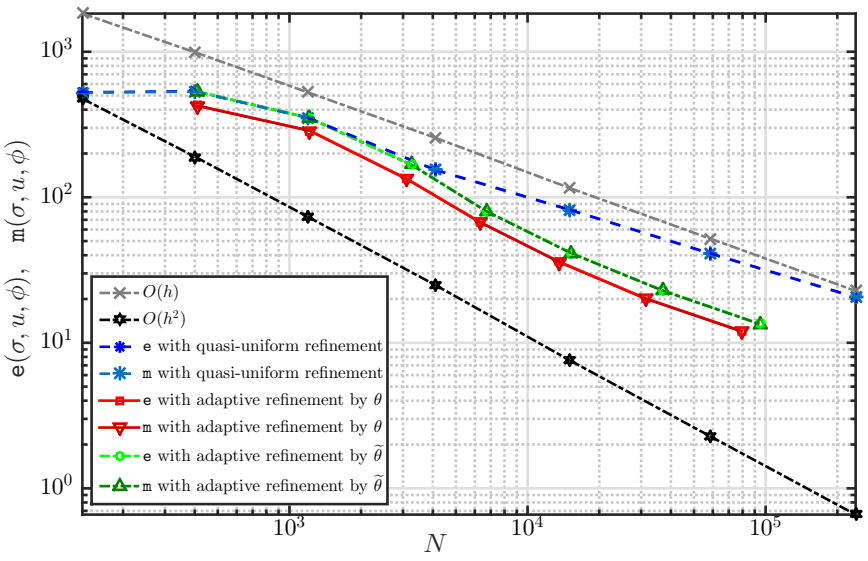


Figure 2.3: Test 2: log-log plot of the total errors vs. degrees of freedom associated to uniform and adaptive mesh refinements using the two proposed indicators.

to (2.95). In this case, the relevant term in the *a posteriori* error estimators will be evidently replaced by

$$\sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\rm N})} h_e \,\|\widetilde{\boldsymbol{\sigma}}_h\cdot\boldsymbol{\nu}-j\|^2_{0,e},$$

whose estimation from below and above follows in a straightforward manner. For this example, the individual and total convergence rates are determined by the expression

$$r(\cdot) := -2\log(e(\cdot)/\hat{e}(\cdot))[\log(N/\hat{N})]^{-1},$$

where $N$ and $\hat{N}$ denote the total number of degrees of freedom associated to each triangulation. Alternatively to the first test, here the Picard tolerance is set to $\varepsilon_{\rm P}$ = 1e-6, and no inner Newton linearization will be employed for the transport problem.

The viscosity, hindered settling and diffusivity functions $\mu,\gamma,\vartheta$ are taken as in the first example with the parameters $a_1 = 0.1$, $a_2 = 0.5$, $b = 3$, $c = 4/3$. Notice that the isotropic part of the stress in (2.95) exhibits a singularity just outside the domain, at $(a_1,a_2)$. With the chosen parameters, the concentration has a high gradient near $\Gamma_{\rm D}$, and the viscosity vanishes whenever the concentration attains its maximum value. Therefore, and according to (2.95), high gradients are also expected in the stress approximation; and optimal convergence, especially in that field, is no longer evidenced under uniform mesh refinement (see first rows of Table 2.2). On the other hand, if an adaptive mesh refinement step (guided by the proposed residual error indicators) is applied, optimal convergence can be restored, as shown in the last two blocks of Table 2.2. This feature is also seen in Figure 2.3, where we plot the total errors $\mathtt{e},\mathtt{m}$ versus the degrees of freedom associated to each triangulation. Total errors under adaptive refinement exhibit a superconvergence whereas uniform refinement yields suboptimal rates. From the figure we also observe that the curves for $\mathtt{e}$ and $\mathtt{m}$ coincide for each algorithm.

Once the local and global error indicators are computed, the adaptation procedure uses the automatic `adaptmesh` tool (see particulars in e.g. [80]) to construct the next triangulation. The algorithm

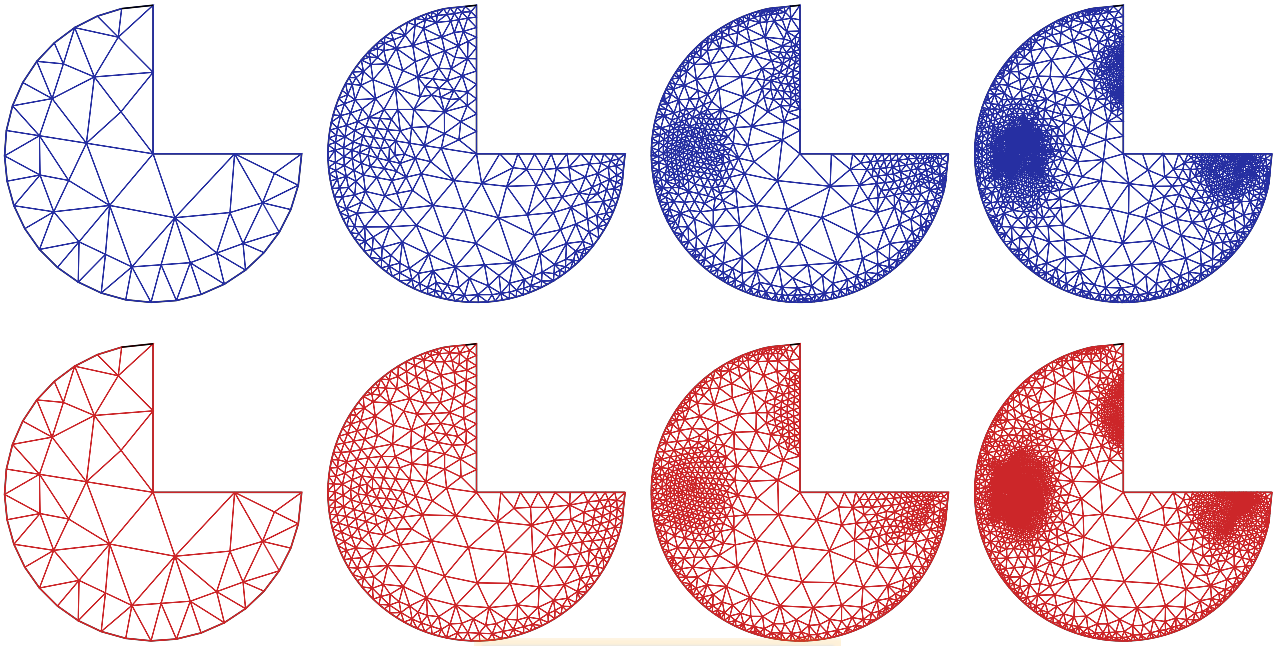

Figure 2.4: Test 2: from left to right, four snapshots of successively refined meshes according to the indicators $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}}$ (top and bottom panels, respectively).

is based on an equi-distribution of the discrete *a posteriori* error indicators, where the diameter of each triangle in $\mathcal{T}_{h_{i+1}}$, which is contained in a generic element $T \in \mathcal{T}_{h_i}$ in the new step of the algorithm, is proportional to $h_T$ times the ratio $\hat{\boldsymbol{\zeta}}_T/\boldsymbol{\zeta}_T$, where $\hat{\boldsymbol{\zeta}}$ denotes the mean value of an estimator over $\mathcal{T}_h$. Approximate solutions obtained after six adaptation steps are depicted in Figure 2.2, whereas a few adaptive meshes generated using the two proposed indicators are collected in Figure 2.4. At least for this particular configuration, the second *a posteriori* error estimator produces smaller errors but the convergence rates coincide with the ones obtained with the first indicator.

In our last example the assumptions on the diffusivity $\vartheta$ will not hold anymore: we allow $\vartheta$ to be constant and very small for any concentration below a so-called gel point $\phi \in [0, \phi_c]$. This extension (whose limit case translates into a loss of ellipticity in the concentration equation) implies that for low volume fractions, one may expect shock-like fronts to develop (see e.g. the monograph [33] and the recent review [20]). Adaptive mesh refinement would then be highly appreciated in this particular case; not only to restitute optimal convergence orders, but also to resolve concentration profiles accurately without the need of refining the grid everywhere (even more important if higher-order schemes are used, or transient models are studied). The problem configuration corresponds to the so-called Boycott effect (*cf.* [23]), where the sedimentation-consolidation of small particles within an enclosure is enhanced by tilting the vessel (from the gravity direction), thus allowing the formation of recirculation zones carrying low concentration fluid along the underside of the inclined wall (see also [30]). The diffusivity function will be set to

$$\vartheta(\phi) = \begin{cases} \varepsilon & \text{for } \phi \leq \phi_c, \\ \vartheta_0 \dfrac{\alpha}{\phi_c} \left( \dfrac{\phi}{\phi_c} \right)^{\alpha-1} & \text{otherwise,} \end{cases}$$

with $\alpha = 5, \vartheta_0 = 0.055, \varepsilon =$1e-6. As computational domain we consider an inclined rectangle of height

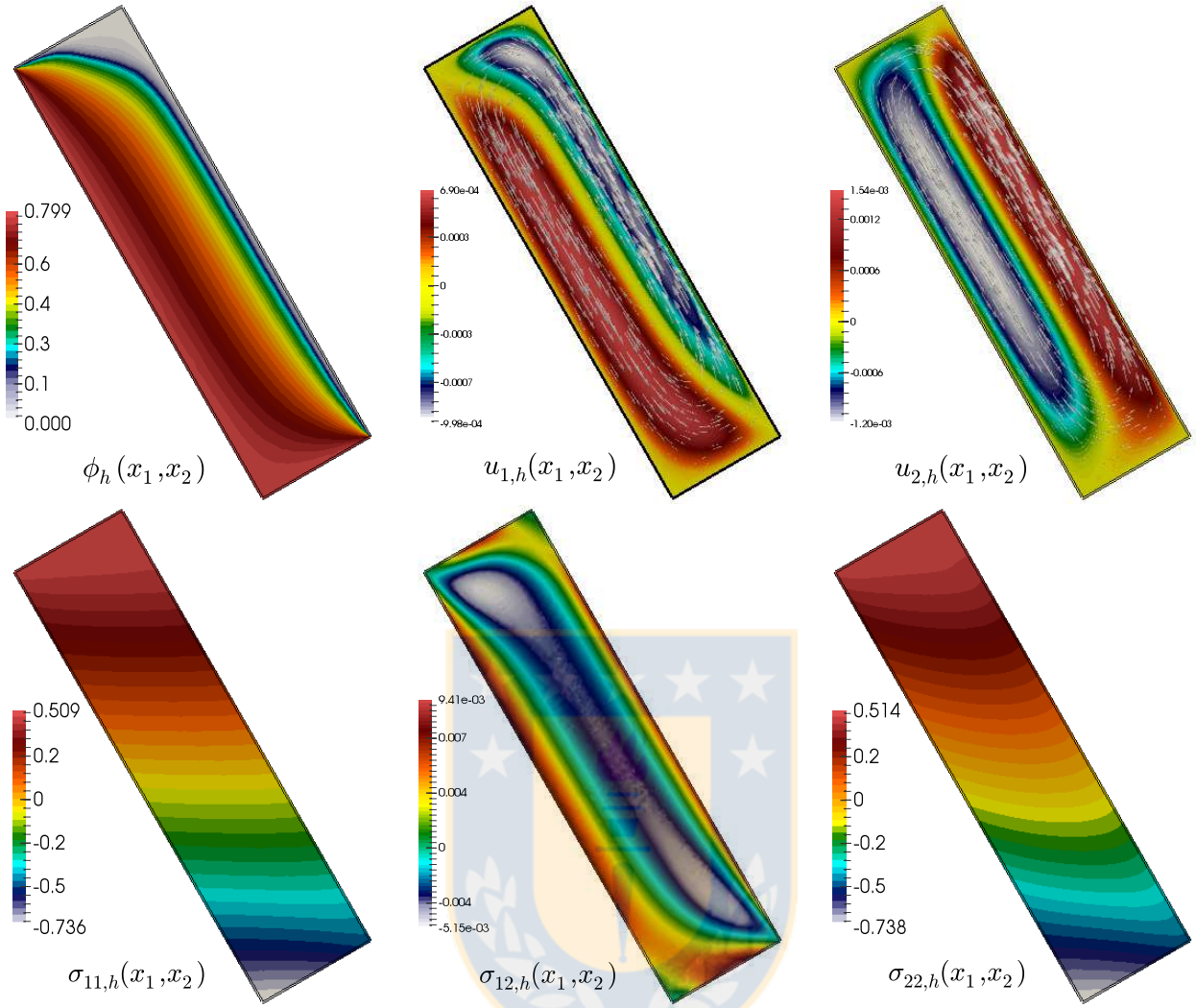


Figure 2.5: Test 3: approximate solutions obtained with the lowest order method, after six steps of adaptive mesh refinement following the second indicator $\widetilde{\boldsymbol{\theta}}$. Concentration, velocity components, and stress components are depicted.

1.5 m and width of 6 m forming an angle of $2\pi/3$ with the positive $x_1-$axis, which we initially discretize into a coarse mesh of 102 triangular elements. Viscosity is set as in the previous examples and the remaining coefficients are $\boldsymbol{f} = \Delta\rho\boldsymbol{k}$, $g = 0$, $\boldsymbol{u}_{\mathrm{D}} = \boldsymbol{0}$, $\phi_{\mathrm{D}} = \{0.8$ on the bottom and overside inclined wall; 0.0001 elsewhere in $\partial\Omega\}$, $\gamma(\phi) = \{\gamma_0\phi(1-\phi)^2$ if $\phi \geq \phi_{\mathrm{c}}$; 0 otherwise$\}$, $\phi_{\mathrm{c}} = 0.07$, $c = 2/3$, $\gamma_0 = 4.4\text{e-}3$, $\Delta\rho = 700$, The stabilization constants will depend on $\mu_1 = 1$ and $\mu_2 = 4.75$.

The numerical solutions are collected in Figure 2.5, where velocity shows a main circulation zone at the center of the domain, directing the flow towards the bottom along the lower inclined side and moving upwards on the opposite side. In addition, high concentration zones are located at the bottom of the vessel, while clear fluid forms at the top. These flow patterns are in accordance with the observations in [28, 30]. Four intermediate steps of mesh adaptation guided by the second *a posteriori* error estimator $\widetilde{\boldsymbol{\theta}}$ are displayed in Figure 2.6. We can see the capturing of the high concentration gradient and velocity boundary layer near the upper inclined side of the domain.

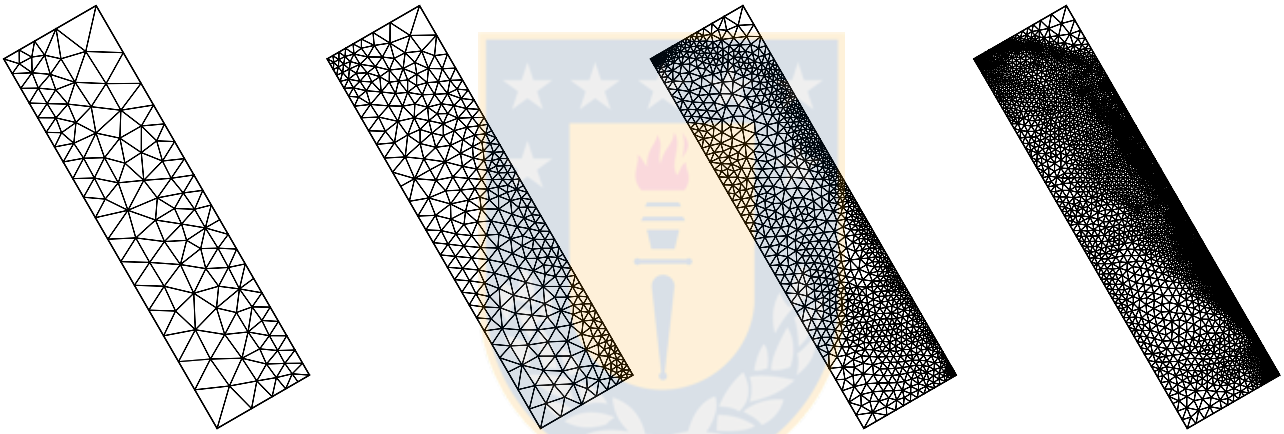

Figure 2.6: Test 3: from left to right, four snapshots of successively refined meshes according to the indicator $\widetilde{\boldsymbol{\theta}}$.

| $N$ | $e(\boldsymbol{\sigma})$ | $r(\boldsymbol{\sigma})$ | $e(\boldsymbol{u})$ | $r(\boldsymbol{u})$ | $e(\phi)$ | $r(\phi)$ | $i_P$ | $\texttt{eff}(\boldsymbol{\theta})$ | $\texttt{qeff}(\boldsymbol{\theta})$ | $\texttt{eff}(\widetilde{\boldsymbol{\theta}})$ | $\texttt{qeff}(\widetilde{\boldsymbol{\theta}})$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Augmented $\mathrm{RT}_0 - P_1 - P_1$ scheme with quasi-uniform refinement | | | | | | | | | | | |
| 135 | 497.2285 | – | 172.3239 | – | 1.6796 | – | 25 | 1.0930 | 1.0855 | 1.0736 | 1.0662 |
| 403 | 501.6796 | -0.0174 | 189.3379 | -0.1835 | 1.5883 | 0.1090 | 29 | 0.9688 | 0.9610 | 0.9654 | 0.9576 |
| 1191 | 325.2386 | 0.8264 | 136.1004 | 0.6295 | 0.7731 | 1.3728 | 28 | 1.0512 | 1.0490 | 1.0413 | 1.0391 |
| 4090 | 150.5401 | 1.2842 | 35.1355 | 2.2575 | 0.4022 | 1.0896 | 29 | 1.0112 | 1.0078 | 1.0096 | 1.0062 |
| 15074 | 81.0276 | 0.9336 | 12.4395 | 1.5649 | 0.1990 | 1.0606 | 28 | 1.0022 | 0.9988 | 1.0031 | 0.9997 |
| 58289 | 41.1328 | 1.0175 | 3.0515 | 2.1089 | 0.1012 | 1.0146 | 31 | 1.0004 | 0.9968 | 1.0031 | 0.9996 |
| 238705 | 20.5693 | 1.0063 | 0.8447 | 1.8650 | 0.0510 | 0.9956 | 29 | 0.9997 | 0.9962 | 1.0032 | 0.9997 |
| Augmented $\mathrm{RT}_0 - P_1 - P1$ scheme with adaptive refinement according to $\boldsymbol{\theta}$ | | | | | | | | | | | |
| 409 | 482.0538 | – | 229.0604 | – | 1.1557 | – | 21 | 1.1461 | 1.1431 | – | – |
| 1215 | 325.3517 | 0.7222 | 151.3681 | 0.7610 | 0.5798 | 1.2670 | 20 | 1.0059 | 1.0040 | – | – |
| 3108 | 160.3139 | 1.5071 | 47.6470 | 2.4614 | 0.4017 | 0.7817 | 19 | 1.0032 | 1.0007 | – | – |
| 6346 | 82.9351 | 1.8466 | 13.4484 | 3.5441 | 0.3298 | 0.5527 | 18 | 1.0060 | 1.0030 | – | – |
| 13629 | 44.6689 | 1.6201 | 4.6110 | 2.8026 | 0.2538 | 0.6851 | 21 | 0.9984 | 0.9949 | – | – |
| 31278 | 25.1967 | 1.3780 | 1.8416 | 2.2089 | 0.1908 | 0.6873 | 20 | 0.9941 | 0.9893 | – | – |
| 79064 | 15.0459 | 1.1118 | 0.8192 | 1.7468 | 0.1332 | 0.7739 | 19 | 0.9903 | 0.9849 | – | – |
| Augmented $\mathrm{RT}_0 - P_1 - P_1$ scheme with adaptive refinement according to $\widetilde{\boldsymbol{\theta}}$ | | | | | | | | | | | |
| 409 | 482.0538 | – | 229.0604 | – | 1.1557 | – | 21 | – | – | 1.1219 | 1.1190 |
| 1206 | 318.9239 | 0.7641 | 148.2081 | 0.8052 | 0.5510 | 1.3700 | 19 | – | – | 0.9979 | 0.9962 |
| 3247 | 160.2352 | 1.3899 | 49.7563 | 2.2041 | 0.3399 | 0.9758 | 20 | – | – | 0.9986 | 0.9965 |
| 6703 | 79.3292 | 1.9399 | 12.8689 | 3.7315 | 0.3024 | 0.3221 | 21 | – | – | 1.0038 | 1.0011 |
| 15393 | 41.1173 | 1.5928 | 3.7409 | 2.9945 | 0.2553 | 0.4104 | 20 | – | – | 1.0073 | 1.0039 |
| 36869 | 22.9177 | 1.3296 | 1.2198 | 2.5490 | 0.1777 | 0.8247 | 20 | – | – | 1.0055 | 1.0008 |
| 94817 | 13.4813 | 1.1231 | 0.5302 | 1.7635 | 0.1289 | 0.6790 | 19 | – | – | 1.0058 | 1.0006 |

Table 2.2: Test 2: convergence history, Picard iteration count, effectivity and quasi-effectivity indexes for the mixed–primal approximation of the coupled problem under quasi-uniform, and adaptive refinement according to the indicators introduced in Sections 2.3,2.4.

# CHAPTER 3

## A mixed–primal finite element approximation of a sedimentation–consolidation system

### 3.1 Introduction

The interaction of solid-liquid suspensions is often encountered in a wide variety of natural and engineering applications, including fluidized beds, clot formation within the blood, solid-liquid separation and purification in wastewater treatment, macroscopic biofilm characterization, and many others. In sedimentation-consolidation processes, suspended solid particles settle down due to gravity acceleration and can be subsequently removed from the fluid. Here we are interested in a continuum-based framework where the viscous fluid is incompressible and the flow patterns are laminar so the mass and momentum balances are governed by the Brinkman equations with variable viscosity, and the mass balance of the solid phase (here allowed to sediment into the fluid phase) is described by a nonlinear advection-diffusion equation. A number of difficulties are associated to the understanding and prediction of the behavior of such a problem, including highly nonlinear (and typically degenerate) advection and diffusion terms, strong interaction of velocity and solids volume fraction via the Cauchy stress tensor and the forcing term, nonlinear structure of the overall coupled flow and transport problem, saddle-point structure of the flow problem, non-homogeneous and mixed boundary conditions, and so on. These complications are usually reflected, not only in the solvability analysis of the governing equations, but also in the construction of appropriate schemes for their numerical approximation, and in the derivation of stability results and error bounds.

The solvability of the sedimentation-consolidation problem has been previously discussed in [29] for the case of large fluid viscosity, using the technique of parabolic regularization. Moreover, a modified formulation based on Stokes flow has been recently studied in [4], where the solution of the transport equation required an explicit dependence of the effective diffusivity on the gradient of the concentration. With a similar restriction (the viscosity depending on the concentration and on the velocity gradient), the existence of solutions to a model of chemically reacting non-Newtonian fluid has been established in [27]. In contrast, here these hypotheses have been modified, enlarging the applicability of the present results, in particular to classical models of sedimentation of suspensions. More specifically, we assume both the viscosity and the diffusivity to depend only on the scalar value of the concentration. Nevertheless, we still remain in the framework of non-degeneracy of the diffusion term. On the other hand, it is worth mentioning that models of sedimentation-consolidation share some structural similarities with Boussinesq- and Oldroyd- type models, for which several mixed formulations

have been proposed [44, 45, 53, 54, 55, 90]. In particular, the mixed finite element method for the Boussinesq problem developed in [54] is based on the introduction of the gradient of velocity as an auxiliary unknown, and the utilization of refined meshes near the singular corners and suitable finite element subspaces. In turn, the approach from [44] first introduces the same nonlinear pseudostress tensor linking the pseudostress and the convective term that has been employed before in [36] for the Navier-Stokes problem, and then augment the resulting mixed formulation of the stationary Boussinesq problem with Galerkin type terms arising from the constitutive and equilibrium equations, and the Dirichlet boundary condition. In this way, the Banach and Brouwer fixed point theorems, together with the Lax-Milgram lemma and the Babuška-Brezzi theory are applied to conclude the well-posedness of the continuous and discrete formulations. Nevertheless, up to our knowledge, mixed formulations specifically tailored for the study of sedimentation processes are not yet available from the literature.

According to the above bibliographic discussion, our present purpose is to examine mixed finite element approximations of the model problem, where also the Cauchy stress enters in the formulation as an additional unknown. Given the arrangement of the equations and the implicit smoothness requirements of the fluid velocity and its discrete approximation, we realize as in [4] that applying an augmentation strategy to the Brinkman problem simplifies the treatment of both the continuous and Galerkin schemes. More precisely, we propose an augmented variational formulation where stresses are sought in $\mathbb{H}(\mathbf{div};\Omega)$, the velocity is in $\mathbf{H}^1(\Omega)$, and the solids volume fraction has $\mathrm{H}^1(\Omega)$ regularity. Consequently, the rows of the Cauchy stress tensor will be approximated with Raviart-Thomas elements of order $k$, whereas the velocity and solids concentration will be discretized with continuous piecewise polynomials of degree $\leq k+1$. The solvability analysis of the continuous formulation is based on a strategy combining classical fixed-point arguments, suitable regularity assumptions on the decoupled problems, the Lax-Milgram lemma, and the Sobolev embedding and Rellich-Kondrachov compactness theorems. In addition, and provided that the data are sufficiently small, we also establish uniqueness of weak solution. On the other hand, well-posedness of the discrete problem relies on the Brouwer fixed point theorem and analogous arguments to those employed in the continuous analysis. Finally, applying a suitable Strang-type lemma valid for linear problems to the fluid flow equations, and explicitly deriving our own Strang-type estimate for the transport equations, we are also able to derive the corresponding Céa estimate, and to provide optimal *a priori* error bounds for the Galerkin solution.

The rest of the chapter is organized as follows. Section 3.2 compiles some preliminary notation and outlines the boundary value problem of interest, which is rewritten by eliminating the pressure unknown from the system. In Section 3.3 we introduce the corresponding variational formulation following an augmented mixed approach for the Brinkman equations, coupled with a primal method for the transport problem. The associated Galerkin scheme is introduced in Section 3.4, followed by the development of its solvability analysis. In Section 3.5 we proceed with the study of accuracy of the augmented formulation, establishing optimal error bounds; and we close in Section 3.6 with some numerical examples illustrating the good performance of the mixed-primal method and confirming the predicted rates of convergence.

## 3.2 The model problem

### 3.2.1 The sedimentation-consolidation system

We consider the steady state of the sedimentation-consolidation process consisting on the transport and suspension of a solid phase into an immiscible fluid contained in a vessel $\Omega$. The flow patterns are influenced by gravity and by the local fluctuations of the solids volume fraction. The process is governed by the following system of partial differential equations:

$$\begin{aligned} \boldsymbol{\sigma} &= \mu(\phi)\,\nabla \boldsymbol{u} - p\,\mathbb{I}\,, \quad \boldsymbol{K}^{-1}\boldsymbol{u} - \mathbf{div}\,\boldsymbol{\sigma} = \boldsymbol{f}\,\phi\,, \quad \operatorname{div}\boldsymbol{u} = 0\,, \\ \widetilde{\boldsymbol{\sigma}} &= \vartheta(\phi)\,\nabla\phi - \phi\boldsymbol{u} - \mathit{f}_{\rm bk}(\phi)\boldsymbol{k}\,, \quad \beta\,\phi - \operatorname{div}\widetilde{\boldsymbol{\sigma}} = g\,. \end{aligned} \tag{3.1}$$

The sought quantities are the Cauchy fluid pseudo-stress $\boldsymbol{\sigma}$, the average velocity of the mixture $\boldsymbol{u}$, the fluid pressure $p$, and the volumetric fraction of the solids (in short, concentration) $\phi$. In this model we also assume that the vessel initially contains an array of fixed-concentration particles (see the discussion in [14]). In this context, the parameter $\beta$ is a positive constant representing the porosity of the medium, and the permeability tensor $\boldsymbol{K} \in \mathbb{C}(\bar{\Omega}) := [C(\bar{\Omega})]^{n\times n}$ and its inverse are symmetric and uniformly positive definite, which means that there exists $\alpha_{\boldsymbol{K}} > 0$ such that

$$\boldsymbol{v}^{\mathtt{t}}\boldsymbol{K}^{-1}(\boldsymbol{x})\boldsymbol{v} \geq \alpha_{\boldsymbol{K}}|\boldsymbol{v}|^2 \quad \forall\, \boldsymbol{v} \in \mathrm{R}^n,\ \forall\, \boldsymbol{x} \in \Omega. \tag{3.2}$$

Here $\boldsymbol{k}$ is a constant vector pointing in the direction of gravity, and we assume that the kinematic effective viscosity, $\mu$; the one-directional Kynch batch flux density function describing hindered settling, $\mathit{f}_{\rm bk}$; and the diffusion or sediment compressibility, $\vartheta$; are nonlinear scalar functions of the concentration $\phi$. In particular, we can take

$$\mu(\phi) := \mu_\infty\left(1-\frac{\phi}{\phi_m}\right)^{-\gamma_\mu}, \quad \mathit{f}_{\rm bk}(\phi) := f_\infty\left[1+\phi\left(1-\frac{\phi}{\phi_m}\right)^{\gamma_f}\right], \quad \vartheta(\phi) := \vartheta_\infty\left[\phi+\left(1-\frac{\phi}{\phi_m}\right)^{-\gamma_\vartheta}\right],$$

where $\mu_\infty, \phi_m, f_\infty, \gamma_\mu, \gamma_f, \gamma_\vartheta, \vartheta_\infty$ are positive model parameters. Notice that $\mathit{f}_{\rm bk}$ and $\vartheta$ are regularized versions of the Kynch flux and compressibility functions typically employed in sedimentation models (see e.g. [28, 29, 31]). Nevertheless, the subsequent analysis allows for arbitrary concentration-dependent functions, as long as the following properties are satisfied: there exist positive constants $\mu_1$, $\mu_2$, $\gamma_1$, $\gamma_2$, $\vartheta_1$, and $\vartheta_2$, such that

$$\mu_1 \leq \mu(s) \leq \mu_2\,, \quad \vartheta_1 \leq \vartheta(s) \leq \vartheta_2\,, \quad \text{and} \quad \gamma_1 \leq \mathit{f}_{\rm bk}(s) \leq \gamma_2 \qquad \forall\, s \in \mathrm{R}\,. \tag{3.3}$$

Note that (3.3) guarantees, in particular, that the corresponding Nemytsky operators, say $U$ for $\mu$, defined by $U(\phi)(x) := \mu(\phi(x))\ \forall\,\phi \in \mathrm{L}^2(\Omega),\ \forall\, x \in \Omega$ a.e., and analogously for $\vartheta$, $\mathit{f}_{\rm bk}$, $\mu^{-1}$, $\vartheta^{-1}$, and $\mathit{f}_{\rm bk}^{-1}$, are all well defined and continuous from $\mathrm{L}^2(\Omega)$ into $\mathrm{L}^2(\Omega)$.

The driving force of the mixture also depends on the local fluctuations of the concentration, so the right hand side of the second equation in (3.1) is linear with respect to $\phi$, and $\boldsymbol{f} \in \mathbf{L}^\infty(\Omega)$ and $g \in \mathrm{L}^2(\Omega)$ are given functions. Finally, given $\boldsymbol{u}_{\rm D} \in \mathbf{H}^{1/2}(\Gamma_{\rm D})$, the following mixed boundary conditions complement (3.1):

$$\boldsymbol{u} = \boldsymbol{u}_{\rm D} \quad \text{on} \quad \Gamma_{\rm D}\,, \quad \boldsymbol{\sigma}\boldsymbol{\nu} = \boldsymbol{0} \quad \text{on} \quad \Gamma_{\rm N}\,, \quad \phi = 0 \quad \text{on} \quad \Gamma_{\rm D}\,, \quad \text{and} \quad \widetilde{\boldsymbol{\sigma}}\cdot\boldsymbol{\nu} = 0 \quad \text{on} \quad \Gamma_{\rm N}\,,$$

where we remark that the homogeneous datum for $\boldsymbol{\sigma}$ represents a pseudo-traction boundary condition, since we are employing $\nabla\boldsymbol{u}$ instead of the symmetrized gradient $\mathbf{e}(\boldsymbol{u}) := \frac{1}{2}(\nabla\boldsymbol{u}+\nabla\boldsymbol{u}^{\mathtt{t}})$ in the definition of the stress.

On the other hand, it is easy to see that the first and third equations in (3.1) are equivalent to

$$\boldsymbol{\sigma} = \mu(\phi)\,\nabla\boldsymbol{u} - p\,\mathbb{I} \quad \text{and} \quad p + \frac{1}{n}\,\mathrm{tr}(\boldsymbol{\sigma}) = 0 \quad \text{in} \quad \Omega\,,$$

which permits us to eliminate the pressure $p$ from the first equation. Consequently, we arrive at the following coupled system:

$$\begin{gathered}
\frac{1}{\mu(\phi)}\,\boldsymbol{\sigma}^{\mathrm{d}} = \nabla\boldsymbol{u} \quad \text{in} \quad \Omega\,, \quad \boldsymbol{K}^{-1}\boldsymbol{u} - \mathbf{div}\,\boldsymbol{\sigma} = \boldsymbol{f}\phi \quad \text{in} \quad \Omega\,, \\
\widetilde{\boldsymbol{\sigma}} = \vartheta(\phi)\,\nabla\phi - \phi\boldsymbol{u} - \mathit{f}_{\mathrm{bk}}(\phi)\boldsymbol{k} \quad \text{in} \quad \Omega\,, \quad \beta\,\phi - \mathrm{div}\widetilde{\boldsymbol{\sigma}} = g \quad \text{in} \quad \Omega\,, \\
\boldsymbol{u} = \boldsymbol{u}_{\mathrm{D}} \quad \text{on} \quad \Gamma_{\mathrm{D}}\,, \quad \boldsymbol{\sigma}\boldsymbol{\nu} = \mathbf{0} \quad \text{on} \quad \Gamma_{\mathrm{N}}\,, \\
\phi = 0 \quad \text{on} \quad \Gamma_{\mathrm{D}}\,, \quad \text{and} \quad \widetilde{\boldsymbol{\sigma}}\cdot\boldsymbol{\nu} = 0 \quad \text{on} \quad \Gamma_{\mathrm{N}}\,.
\end{gathered} \tag{3.4}$$

We stress that the incompressibility constraint is implicitly present in the constitutive equation $(3.4)_1$ relating $\boldsymbol{\sigma}$ and $\boldsymbol{u}$. Systems like (3.1) are well established and have been extensively validated to describe sediment-flow patterns in permeable media (see [81, 89] and the references therein). Furthermore, if we wanted to deal with a traction boundary condition on $\Gamma_{\mathrm{N}}$, then $(3.1)_1$ and $(3.4)_1$ would be replaced simply by $\boldsymbol{\sigma} = \mu(\phi)\,\mathbf{e}(\boldsymbol{u}) - p\,\mathbb{I}$ and $\dfrac{1}{\mu(\phi)}\,\boldsymbol{\sigma}^{\mathrm{d}} = \nabla\boldsymbol{u} - \gamma$, respectively, where $\gamma := \frac{1}{2}(\nabla\boldsymbol{u} - \nabla\boldsymbol{u}^{\mathtt{t}})$ is the additional unknown given by the vorticity. In this case, however, the rest of the corresponding analysis would be very close to the one to be developed in what follows.

## 3.3 The variational formulation

In this section we proceed similarly as in [4] to derive a suitable variational formulation of (3.4) and analyze its corresponding solvability by using a fixed-point strategy.

### 3.3.1 An augmented mixed–primal approach

Notice that the homogeneous boundary condition for $\boldsymbol{\sigma}$ on $\Gamma_{\mathrm{N}}$ (*cf.* third row of (3.4)) suggests the introduction of the functional space

$$\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega) := \Big\{\boldsymbol{\tau} \in \mathbb{H}(\mathbf{div};\Omega) : \quad \boldsymbol{\tau}\boldsymbol{\nu} = \mathbf{0} \quad \text{on} \quad \Gamma_{\mathrm{N}}\Big\}\,.$$

Multiplying the first equation of (3.4) by $\boldsymbol{\tau} \in \mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)$, integrating by parts, and using the Dirichlet boundary condition for $\boldsymbol{u}$ (*cf.* third row of (3.4)), we obtain

$$\int_\Omega \frac{1}{\mu(\phi)}\,\boldsymbol{\sigma}^{\mathrm{d}} : \boldsymbol{\tau}^{\mathrm{d}} + \int_\Omega \boldsymbol{u}\cdot\mathbf{div}\,\boldsymbol{\tau} = \langle\boldsymbol{\tau}\boldsymbol{\nu}, \boldsymbol{u}_{\mathrm{D}}\rangle_{\Gamma_{\mathrm{D}}} \qquad \forall\,\boldsymbol{\tau} \in \mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)\,,$$

where $\langle\cdot,\cdot\rangle_{\Gamma_{\mathrm{D}}}$ is the duality pairing between $\mathbf{H}^{-1/2}(\Gamma_{\mathrm{D}})$ and $\mathbf{H}^{1/2}(\Gamma_{\mathrm{D}})$. Moreover, the momentum balance is then rewritten as

$$-\int_\Omega \boldsymbol{K}^{-1}\boldsymbol{u}\cdot\boldsymbol{v} + \int_\Omega \boldsymbol{v}\cdot\mathbf{div}\boldsymbol{\sigma} = -\int_\Omega \boldsymbol{f}\phi\cdot\boldsymbol{v} \qquad \forall\,\boldsymbol{v} \in \mathbf{L}^2(\Omega)\,.$$

On the other hand, the Dirichlet boundary condition for $\phi$ (*cf.* fourth row of (3.4)) motivates the introduction of the space

$$\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) := \Big\{\psi \in \mathrm{H}^1(\Omega) : \quad \psi = 0 \quad \text{on} \quad \Gamma_{\mathrm{D}}\Big\}\,,$$

for which, thanks to the generalized Poincaré inequality, there exists $c_p > 0$, depending only on $\Omega$ and $\Gamma_{\rm D}$, such that

$$\|\psi\|_{1,\Omega} \leq c_p\, |\psi|_{1,\Omega} \qquad \forall\, \psi \in \mathrm{H}^1_{\Gamma_{\rm D}}(\Omega)\,. \tag{3.5}$$

Therefore, given $\phi \in \mathrm{H}^1_{\Gamma_{\rm D}}(\Omega)$, we arrive at the following mixed formulation for the Brinkman flow: Find $(\boldsymbol{\sigma}, \boldsymbol{u}) \in \mathbb{H}_{\rm N}(\mathbf{div};\Omega) \times \mathbf{L}^2(\Omega)$ such that

$$\begin{array}{rcll} \mathbf{a}_\phi(\boldsymbol{\sigma},\boldsymbol{\tau}) + \mathbf{b}(\boldsymbol{\tau},\boldsymbol{u}) & = & \langle \boldsymbol{\tau}\boldsymbol{\nu}, \boldsymbol{u}_{\rm D}\rangle_{\Gamma_{\rm D}} & \forall\, \boldsymbol{\tau} \in \mathbb{H}_{\rm N}(\mathbf{div};\Omega), \\ \mathbf{b}(\boldsymbol{\sigma},\boldsymbol{v}) - \mathbf{c}(\boldsymbol{u},\boldsymbol{v}) & = & -\displaystyle\int_\Omega \boldsymbol{f}\phi\cdot\boldsymbol{v} & \forall\, \boldsymbol{v} \in \mathbf{L}^2(\Omega), \end{array} \tag{3.6}$$

where $\mathbf{a}_\phi : \mathbb{H}_{\rm N}(\mathbf{div};\Omega) \times \mathbb{H}_{\rm N}(\mathbf{div};\Omega) \to \mathbb{R}$, $\mathbf{b} : \mathbb{H}_{\rm N}(\mathbf{div};\Omega) \times \mathbf{L}^2(\Omega) \to \mathbb{R}$ and $\mathbf{c} : \mathbf{L}^2(\Omega) \times \mathbf{L}^2(\Omega) \to \mathbb{R}$ are bounded bilinear forms defined as

$$\mathbf{a}_\phi(\boldsymbol{\zeta},\boldsymbol{\tau}) := \int_\Omega \frac{1}{\mu(\phi)}\, \boldsymbol{\zeta}^{\rm d} : \boldsymbol{\tau}^{\rm d}, \qquad \mathbf{b}(\boldsymbol{\tau},\boldsymbol{v}) := \int_\Omega \boldsymbol{v}\cdot\mathbf{div}\boldsymbol{\tau}, \quad \mathbf{c}(\boldsymbol{u},\boldsymbol{v}) := \int_\Omega \boldsymbol{K}^{-1}\boldsymbol{u}\cdot\boldsymbol{v} \tag{3.7}$$

for $\boldsymbol{\zeta}, \boldsymbol{\tau} \in \mathbb{H}_{\rm N}(\mathbf{div};\Omega)$ and $\boldsymbol{u}$, $\boldsymbol{v} \in \mathbf{L}^2(\Omega)$.

In turn, given $\boldsymbol{u} \in \mathbf{L}^2(\Omega)$, and using the homogeneous Neumann boundary condition for $\widetilde{\boldsymbol{\sigma}}$ (*cf.* fourth row of (3.4)), we deduce that the primal formulation for the concentration equation becomes: Find $\phi \in \mathrm{H}^1_{\Gamma_{\rm D}}(\Omega)$ such that

$$A_{\boldsymbol{u}}(\phi,\psi) \,=\, \int_\Omega f_{\rm bk}(\phi)\, \boldsymbol{k}\cdot\nabla\psi \,+\, \int_\Omega g\psi \qquad \forall\, \psi \in \mathrm{H}^1_{\Gamma_{\rm D}}(\Omega)\,, \tag{3.8}$$

where

$$A_{\boldsymbol{u}}(\phi,\psi) \,:=\, \int_\Omega \vartheta(\phi)\, \nabla\phi\cdot\nabla\psi \,-\, \int_\Omega \phi\, \boldsymbol{u}\cdot\nabla\psi \,+\, \int_\Omega \beta\, \phi\psi \qquad \forall\, \phi,\, \psi \in \mathrm{H}^1_{\Gamma_{\rm D}}(\Omega)\,. \tag{3.9}$$

We remark at this point that the well-posedness of (3.6) is a straightforward consequence of the assumption on $\mu$ given in (3.3) and the well known Babuška-Brezzi theory (see, e.g. [72, Theorem 2.1] and [21, Proposition 4.3.1] for details). However, in order to deal with the analysis of (3.8) - (3.9), and particularly to estimate the second term defining $A_{\boldsymbol{u}}$, we would require $\boldsymbol{u} \in \mathbf{H}^1(\Omega)$. In fact, we know from the Rellich-Kondrachov compactness Theorem (*cf.* [1, Theorem 6.3] , [92, Theorem 1.3.5]), that the injection $i_c : \mathrm{H}^1(\Omega) \longrightarrow \mathrm{L}^4(\Omega)$ is compact, and hence continuous, which, after applying Hölder's inequality, yields the existence of a positive constant $c(\Omega) = \|i_c\|^2$, depending only on $\Omega$, such that

$$\left|\int_\Omega \phi\, \boldsymbol{v}\cdot\nabla\psi\right| \leq c(\Omega)\, \|\phi\|_{1,\Omega}\, \|\boldsymbol{v}\|_{1,\Omega}\, |\psi|_{1,\Omega} \qquad \forall\, \phi,\, \psi \in \mathrm{H}^1(\Omega)\,, \quad \forall\, \boldsymbol{v} \in \mathbf{H}^1(\Omega)\,. \tag{3.10}$$

Furthermore, we now observe, as we did in [4], that while the exact solution of (3.6) actually satisfies $\nabla\boldsymbol{u} = \dfrac{1}{\mu(\phi)}\,\boldsymbol{\sigma}^{\rm d}$ in $\mathcal{D}'(\Omega)$, which implies that $\boldsymbol{u}$ does belong to $\mathbf{H}^1(\Omega)$, the foregoing distributional identity does not necessarily extend to the discrete setting of (3.6), and hence the aforementioned difficulty would appear again when trying to analyze the Galerkin scheme associated to (3.8). In order to overcome this inconvenience, we proceed similarly as in [4, Section 3.1] (see also [57, Section 3]) and incorporate into (3.6) the following residual Galerkin type terms

$$\begin{array}{rcll} \kappa_1 \displaystyle\int_\Omega \left(\nabla\boldsymbol{u} - \frac{1}{\mu(\phi)}\,\boldsymbol{\sigma}^{\rm d}\right) : \nabla\boldsymbol{v} & = & 0 & \forall\, \boldsymbol{v} \in \mathbf{H}^1(\Omega)\,, \\ -\kappa_2 \displaystyle\int_\Omega \boldsymbol{K}^{-1}\boldsymbol{u}\cdot\mathbf{div}\boldsymbol{\tau} + \kappa_2 \int_\Omega \mathbf{div}\boldsymbol{\sigma}\cdot\mathbf{div}\boldsymbol{\tau} & = & -\kappa_2 \displaystyle\int_\Omega \boldsymbol{f}\phi\cdot\mathbf{div}\boldsymbol{\tau} & \forall\, \boldsymbol{\tau} \in \mathbb{H}_{\rm N}(\mathbf{div};\Omega)\,, \end{array} \tag{3.11}$$

where $(\kappa_1,\kappa_2)$ is a vector of positive parameters to be specified later. In this way, instead of (3.6), we consider from now on the following augmented mixed formulation: Find $(\boldsymbol{\sigma},\boldsymbol{u})\in\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega)$ such that

$$B_\phi((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau},\boldsymbol{v}))\,=\,F_\phi(\boldsymbol{\tau},\boldsymbol{v})\quad\forall\,(\boldsymbol{\tau},\boldsymbol{v})\in\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega)\,, \tag{3.12}$$

where

$$\begin{aligned}B_\phi((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau},\boldsymbol{v}))\,&:=\,\mathbf{a}_\phi(\boldsymbol{\sigma},\boldsymbol{\tau})\,+\,\mathbf{b}(\boldsymbol{\tau},\boldsymbol{u})\,-\,\mathbf{b}(\boldsymbol{\sigma},\boldsymbol{v})\,+\,\mathbf{c}(\boldsymbol{u},\boldsymbol{v})\\&+\kappa_1\int_\Omega\Big(\nabla\boldsymbol{u}-\frac{1}{\mu(\phi)}\,\boldsymbol{\sigma}^{\mathtt{d}}\Big):\nabla\boldsymbol{v}\,-\,\kappa_2\int_\Omega\boldsymbol{K}^{-1}\boldsymbol{u}\cdot\mathbf{div}\boldsymbol{\tau}\,+\,\kappa_2\int_\Omega\mathbf{div}\boldsymbol{\sigma}\cdot\mathbf{div}\boldsymbol{\tau}\,,\end{aligned} \tag{3.13}$$

and

$$F_\phi(\boldsymbol{\tau},\boldsymbol{v})\,:=\,\langle\boldsymbol{\tau}\boldsymbol{\nu},\boldsymbol{u}_{\mathrm{D}}\rangle_{\Gamma_{\mathrm{D}}}\,+\,\int_\Omega\boldsymbol{f}\phi\cdot\boldsymbol{v}\,-\,\kappa_2\int_\Omega\boldsymbol{f}\phi\cdot\mathbf{div}\boldsymbol{\tau}\,. \tag{3.14}$$

We remark in advance that the well-posedness of (3.12) is proved below in Section 3.3.3. To this respect, it is important to highlight that, differently from [4], here we do not need to add any stabilization term on the Dirichlet boundary, as we did in [4, eq. (3.6)], since the required $\mathbf{H}^1(\Omega)$-norm is obtained thanks to the first equation in (3.11) and the presence now of the positive definite bilinear form $\mathbf{c}$ (*cf.* (3.7)) in the definition of $B_\phi$. Furthermore, since the unique solution of (3.6) is obviously a solution of (3.12) as well, we will conclude that both continuous problems share the same unique solution.

Summarizing the foregoing discussion, we find that the augmented mixed-primal formulation of the initial coupled problem (3.4) reduces to (3.12) and (3.8), that is: Find $(\boldsymbol{\sigma},\boldsymbol{u},\phi)\in\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega)\times\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ such that

$$\begin{aligned}B_\phi((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau},\boldsymbol{v}))\,&=\,F_\phi(\boldsymbol{\tau},\boldsymbol{v})&&\forall(\boldsymbol{\tau},\boldsymbol{v})\in\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega)\,,\\A_{\boldsymbol{u}}(\phi,\psi)\,&=\,\int_\Omega f_{\mathrm{bk}}(\phi)\,\boldsymbol{k}\cdot\nabla\psi\,+\,\int_\Omega g\psi&&\forall\,\psi\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,.\end{aligned} \tag{3.15}$$

### 3.3.2 Fixed point strategy

We begin by noticing that the alternative formulation (3.12) will certainly require continuous and discrete solutions with second components living in $\mathbf{H}^1(\Omega)$. Now, given $\phi\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ and the corresponding solution $(\boldsymbol{\sigma},\boldsymbol{u})\in\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega)$ of (3.12), we can set, instead of (3.8), the modified primal formulation: Find $\widetilde{\phi}\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ such that

$$A_{\phi,\boldsymbol{u}}(\widetilde{\phi},\widetilde{\psi})\,=\,G_\phi(\widetilde{\psi})\qquad\forall\,\widetilde{\psi}\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,, \tag{3.16}$$

where

$$A_{\phi,\boldsymbol{u}}(\widetilde{\phi},\widetilde{\psi})\,:=\,\int_\Omega\vartheta(\phi)\,\nabla\widetilde{\phi}\cdot\nabla\widetilde{\psi}\,-\,\int_\Omega\widetilde{\phi}\,\boldsymbol{u}\cdot\nabla\widetilde{\psi}\,+\,\int_\Omega\beta\,\widetilde{\phi}\widetilde{\psi}\qquad\forall\,\widetilde{\phi},\,\widetilde{\psi}\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,, \tag{3.17}$$

and

$$G_\phi(\widetilde{\psi})\,:=\,\int_\Omega f_{\mathrm{bk}}(\phi)\,\boldsymbol{k}\cdot\nabla\widetilde{\psi}\,+\,\int_\Omega g\,\widetilde{\psi}\qquad\forall\,\widetilde{\psi}\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,. \tag{3.18}$$

The well-posedness of (3.16) will also be addressed in Section 3.3.3.

In turn, the description of problems (3.12) and (3.16) naturally suggests a fixed point strategy to analyze (3.15). Indeed, let $\mathbf{S}:\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\longrightarrow\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega)$ be the operator defined by

$$\mathbf{S}(\phi)\,=\,(\mathbf{S}_1(\phi),\mathbf{S}_2(\phi))\,:=\,(\boldsymbol{\sigma},\boldsymbol{u})\,\in\,\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega)\qquad\forall\,\phi\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,,$$

where $(\boldsymbol{\sigma},\boldsymbol{u})$ is the unique solution of (3.12) with the given $\phi$. In turn, let $\widetilde{\mathbf{S}} : \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\times\mathbf{H}^1(\Omega) \longrightarrow \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ be the operator defined by

$$\widetilde{\mathbf{S}}(\phi,\boldsymbol{u}) := \widetilde{\phi} \qquad \forall\,(\phi,\boldsymbol{u}) \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\times\mathbf{H}^1(\Omega)\,,$$

where $\widetilde{\phi}$ is the unique solution of (3.16) with the given $(\phi,\boldsymbol{u})$. Then, we define the operator $\mathbf{T} : \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) \longrightarrow \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ by

$$\mathbf{T}(\phi) := \widetilde{\mathbf{S}}(\phi,\mathbf{S}_2(\phi)) \qquad \forall\,\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,,$$

and realize that solving (3.15) is equivalent to seeking a fixed point of $\mathbf{T}$, that is: Find $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ such that

$$\mathbf{T}(\phi) = \phi\,. \tag{3.19}$$

We find it important to remark here that, due to the dependence on $\phi$ (instead of $|\nabla\phi|$ as in [4]) of the diffusivity $\vartheta$, our nonlinear operator $A_{\boldsymbol{u}}$ (*cf.* (3.9)) does not become strongly monotone (as it was the case for the corresponding nonlinear operator in [4, eq. (3.4) and Lemma 3.5]), and hence we realize that for easing the present analysis we need to stay with the linear problem (3.16) instead of the nonlinear one suggested by (3.8) and (3.9).

### 3.3.3 Well-posedness of the uncoupled problems

In this section we show that the uncoupled problems (3.12) and (3.16) are in fact well-posed. We begin by recalling (see, e.g. [26]) that $\mathbb{H}(\mathbf{div};\Omega) = \mathbb{H}_0(\mathbf{div};\Omega)\oplus\mathrm{R}\,\mathbb{I}$, where

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

More precisely, for each $\boldsymbol{\zeta}\in\mathbb{H}(\mathbf{div};\Omega)$ there exist unique $\boldsymbol{\zeta}_0 := \boldsymbol{\zeta} - \left\{\dfrac{1}{n|\Omega|}\displaystyle\int_\Omega \mathrm{tr}(\boldsymbol{\zeta})\right\}\mathbb{I} \in \mathbb{H}_0(\mathbf{div};\Omega)$ and $d := \dfrac{1}{n|\Omega|}\displaystyle\int_\Omega \mathrm{tr}(\boldsymbol{\zeta}) \in \mathrm{R}$, such that $\boldsymbol{\zeta} = \boldsymbol{\zeta}_0 + d\mathbb{I}$. As for the analysis in [4], the following two Lemmas concerning the above decomposition will be instrumental in showing the well-posedness of (3.12) for a given $\phi$.

**Lemma 3.1** (Proposition 3.1 in [26]). *There exists $c_1 = c_1(\Omega) > 0$ such that*

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

*with $\boldsymbol{\tau}_0\in\mathbb{H}_0(\mathbf{div};\Omega)$ and $c\in\mathrm{R}$.*

**Lemma 3.2** (Lemma 2.2 in [62]). *There exists $c_2 = c_2(\Omega,\Gamma_{\mathrm{N}}) > 0$ such that*

$$c_2\,\|\boldsymbol{\tau}\|^2_{\mathbf{div};\Omega} \le \|\boldsymbol{\tau}_0\|^2_{\mathbf{div};\Omega} \qquad \forall\,\boldsymbol{\tau} = \boldsymbol{\tau}_0 + c\mathbb{I}\in\mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega),$$

*with $\boldsymbol{\tau}_0\in\mathbb{H}_0(\mathbf{div};\Omega)$ and $c\in\mathrm{R}$.*

We now begin the solvability analysis of the uncoupled problems with the following result.

**Lemma 3.3.** *Assume that $\kappa_1 \in \left(0,\frac{2\delta\mu_1}{\mu_2}\right)$ and $\kappa_2 \in \left(0,\frac{2\widetilde{\delta}\,\alpha_{\boldsymbol{K}}}{n\|\boldsymbol{K}^{-1}\|_\infty}\right)$, with $\delta\in(0,2\mu_1)$ and $\widetilde{\delta}\in\left(0,\frac{2}{n\|\boldsymbol{K}^{-1}\|_\infty}\right)$. Then, for each $\phi\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$, problem (3.12) has a unique solution $\mathbf{S}(\phi) := (\boldsymbol{\sigma},\boldsymbol{u})\in H := \mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)\times\mathbf{H}^1(\Omega)$. Moreover, there exists $C_S>0$, independent of $\phi$, such that*

$$\|\mathbf{S}(\phi)\|_H = \|(\boldsymbol{\sigma},\boldsymbol{u})\|_H \le C_S\left\{\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + \|\phi\|_{0,\Omega}\,\|\boldsymbol{f}\|_{\infty,\Omega}\right\} \qquad \forall\,\phi\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,. \tag{3.20}$$

*Proof.* We first observe from (3.13) that, given $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$, $B_\phi$ is clearly a bilinear form. Next, applying Cauchy-Schwarz inequality and the lower bound for $\mu$ (*cf.* (3.3)), we find from (3.13) that there exists a positive constant $\|B\|$, depending on $\mu_1$, $\kappa_1$, $\kappa_2$, $n$, and $\|\boldsymbol{K}^{-1}\|_\infty$, such that

$$|B_\phi((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau},\boldsymbol{v}))| \,\leq\, \|B\|\,\|(\boldsymbol{\sigma},\boldsymbol{u})\|_H\,\|(\boldsymbol{\tau},\boldsymbol{v})\|_H \qquad \forall\,(\boldsymbol{\sigma},\boldsymbol{u}),\,(\boldsymbol{\tau},\boldsymbol{v}) \,\in\, H\,, \tag{3.21}$$

which confirms the boundedness of $B_\phi$ independently of $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$. Next, we show that $B_\phi$ is $H$-elliptic. In fact, given $(\boldsymbol{\tau},\boldsymbol{v}) \in H$, we have again from (3.13) that

$$\begin{aligned} B_\phi((\boldsymbol{\tau},\boldsymbol{v}),(\boldsymbol{\tau},\boldsymbol{v})) \,=\, & \int_\Omega \frac{1}{\mu(\phi)}\,|\boldsymbol{\tau}^{\mathtt{d}}|^2 \,+\, \kappa_1\,|\boldsymbol{v}|^2_{1,\Omega} \,-\, \kappa_1 \int_\Omega \frac{1}{\mu(\phi)}\,\boldsymbol{\tau}^{\mathtt{d}} : \nabla \boldsymbol{v} \\ & +\quad \kappa_2 \|\mathbf{div}\,\boldsymbol{\tau}\|^2_{0,\Omega} \,+\, \int_\Omega \boldsymbol{K}^{-1}\boldsymbol{v}\cdot\boldsymbol{v} \,-\, \kappa_2 \int_\Omega \boldsymbol{K}^{-1}\boldsymbol{v}\cdot\mathbf{div}\,\boldsymbol{\tau}\,, \end{aligned}$$

which, using the lower and upper bounds for $\mu$ (*cf.* (3.3)), the Cauchy-Schwarz and Young inequalities, and the estimate (3.2), yields for any $\delta,\,\widetilde{\delta} > 0$,

$$\begin{aligned} B_\phi((\boldsymbol{\tau},\boldsymbol{v}),(\boldsymbol{\tau},\boldsymbol{v})) \,\geq\, & \left(\frac{1}{\mu_2} - \frac{\kappa_1}{2\delta\mu_1}\right)\,\|\boldsymbol{\tau}^{\mathtt{d}}\|^2_{0,\Omega} \,+\, \kappa_2\left(1 - \frac{n\,\widetilde{\delta}}{2}\,\|\boldsymbol{K}^{-1}\|_\infty\right)\|\mathbf{div}\,\boldsymbol{\tau}\|^2_{0,\Omega} \\ & +\quad \kappa_1\left(1 - \frac{\delta}{2\mu_1}\right)\,|\boldsymbol{v}|^2_{1,\Omega} \,+\, \left(\alpha_{\boldsymbol{K}} \,-\, \frac{n\,\kappa_2}{2\widetilde{\delta}}\,\|\boldsymbol{K}^{-1}\|_\infty\right)\|\boldsymbol{v}\|^2_{0,\Omega}\,. \end{aligned} \tag{3.22}$$

Then, assuming the indicated hypotheses on $\delta$, $\kappa_1$, $\widetilde{\delta}$, and $\kappa_2$, we can introduce the positive constants

$$\begin{aligned} \alpha_0(\Omega) \;&:=\; \min\left\{\left(\frac{1}{\mu_2} - \frac{\kappa_1}{2\delta\mu_1}\right),\, \frac{\kappa_2}{2}\left(1 - \frac{n\,\widetilde{\delta}}{2}\,\|\boldsymbol{K}^{-1}\|_\infty\right)\right\}\,, \\ \alpha_1(\Omega) \;&:=\; c_2\,\min\left\{c_1\,\alpha_0(\Omega)\,,\, \frac{\kappa_2}{2}\left(1 - \frac{n\,\widetilde{\delta}}{2}\,\|\boldsymbol{K}^{-1}\|_\infty\right)\right\}\,, \\ \alpha_2(\Omega) \;&:=\; \min\left\{\kappa_1\left(1 - \frac{\delta}{2\mu_1}\right)\,,\, \left(\alpha_{\boldsymbol{K}} \,-\, \frac{n\,\kappa_2}{2\widetilde{\delta}}\,\|\boldsymbol{K}^{-1}\|_\infty\right)\right\}\,, \end{aligned}$$

which, according to Lemmas 3.1 and 3.2, and defining $\alpha(\Omega) \,:=\, \min\Big\{\alpha_1(\Omega),\alpha_2(\Omega)\Big\}$, implies from (3.22) that

$$B_\phi((\boldsymbol{\tau},\boldsymbol{v}),(\boldsymbol{\tau},\boldsymbol{v})) \,\geq\, \alpha(\Omega)\,\|(\boldsymbol{\tau},\boldsymbol{v})\|^2 \qquad \forall\,(\boldsymbol{\tau},\boldsymbol{v}) \,\in\, \mathbf{H}\,, \tag{3.23}$$

thus confirming the $H$-ellipticity of $B_\phi$ independently of $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ as well. Next, given $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$, it is easy to see from (3.14) that there exists a positive constant $\|F\|$, depending only on $\kappa_2$, such that

$$\|F_\phi\| \,\leq\, \|F\|\,\Big\{\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} \,+\, \|\phi\|_{0,\Omega}\,\|\boldsymbol{f}\|_{\infty,\Omega}\Big\}\,. \tag{3.24}$$

Finally, a straightforward application of the Lax-Milgram Lemma (see, e.g. [63, Theorem 1.1]), proves that, for each $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$, problem (3.12) has a unique solution $\mathbf{S}(\phi) \,:=\, (\boldsymbol{\sigma},\boldsymbol{u}) \,\in\, H$. Moreover, the corresponding continuous dependence result together with the estimates (3.23) and (3.24) yield (3.20) with $C_S \,:=\, \dfrac{\|F\|}{\alpha(\Omega)}$, which completes the proof. □

We now establish the unique solvability of the linear problem (3.16).

**Lemma 3.4.** *Let $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ and $\boldsymbol{u} \in \mathbf{H}^1(\Omega)$ such that $\|\boldsymbol{u}\|_{1,\Omega} < \frac{\vartheta_1}{2\,c_p\,c(\Omega)}$ (cf. (3.3), (3.5), (3.10)). Then, there exists a unique $\widetilde{\phi} := \widetilde{\mathbf{S}}(\phi,\boldsymbol{u}) \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ solution of (3.16). Moreover, there exists $C_{\widetilde{S}} > 0$, independent of $(\phi,\boldsymbol{u})$, such that*

$$\|\widetilde{\mathbf{S}}(\phi,\boldsymbol{u})\|_{1,\Omega} \,=\, \|\widetilde{\phi}\|_{1,\Omega} \,\leq\, C_{\widetilde{S}}\,\Big\{\gamma_2\,|\Omega|^{1/2}\,|\boldsymbol{k}| \,+\, \|g\|_{0,\Omega}\Big\}\,. \tag{3.25}$$

*Proof.* First all, notice that $A_{\phi,\boldsymbol{u}}$ (*cf.* (3.17)) is clearly a bilinear form. In turn, according to (3.3) and (3.10), it readily follows from (3.17) that there exists a positive constant $\|A\|$, depending on $\vartheta_2$, $c(\Omega)$, and the bound for $\|\boldsymbol{u}\|_{1,\Omega}$ assumed here, that

$$|A_{\phi,\boldsymbol{u}}(\widetilde{\phi},\widetilde{\psi})| \,\leq\, \|A\|\,\|\widetilde{\phi}\|_{1,\Omega}\,\|\widetilde{\psi}\|_{1,\Omega} \quad \forall\,\widetilde{\phi},\,\widetilde{\psi} \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,,$$

which proves that $A_{\phi,\boldsymbol{u}}$ is bounded independently of $\phi$ and $\boldsymbol{u}$. Next, applying the estimate (3.10) and the inequality (3.5), we find that for each $\widetilde{\phi} \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ there holds

$$\begin{aligned} A_{\phi,\boldsymbol{u}}(\widetilde{\phi},\widetilde{\phi}) \,&=\, \int_\Omega \vartheta(\phi)\,|\nabla\widetilde{\phi}|^2 \,-\, \int_\Omega \widetilde{\phi}\boldsymbol{u}\cdot\nabla\widetilde{\phi} \,+\, \beta\,\|\widetilde{\phi}\|^2_{0,\Omega} \\ &\geq\, \big(\vartheta_1 \,-\, c_p\,c(\Omega)\,\|\boldsymbol{u}\|_{1,\Omega}\big)\,|\widetilde{\phi}|^2_{1,\Omega} \,\geq\, \frac{\vartheta_1}{2}\,|\widetilde{\phi}|^2_{1,\Omega} \,\geq\, \frac{\vartheta_1}{2\,c_p^2}\,\|\widetilde{\phi}\|^2_{1,\Omega}\,, \end{aligned} \tag{3.26}$$

which proves that $A_{\phi,\boldsymbol{u}}$ is $\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$-elliptic with constant $\widetilde{\alpha} := \frac{\vartheta_1}{2\,c_p^2}$, independently of $\phi$ and $\boldsymbol{u}$ as well. On the other hand, applying Cauchy-Schwarz inequality and the upper bound for $f_{\mathrm{bk}}$ given in (3.3), we easily deduce that

$$|G_\phi(\widetilde{\psi})| \,\leq\, \Big\{\gamma_2\,|\Omega|^{1/2}\,|\boldsymbol{k}| \,+\, \|g\|_{0,\Omega}\Big\}\,\|\widetilde{\psi}\|_{1,\Omega} \qquad \forall\,\widetilde{\psi} \,\in\, \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,,$$

which says that $G_\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'$ and $\|G_\phi\| \,\leq\, \Big\{\gamma_2\,|\Omega|^{1/2}\,|\boldsymbol{k}| \,+\, \|g\|_{0,\Omega}\Big\}$. Consequently, a direct application of the Lax-Milgram Lemma implies the existence of a unique solution $\widetilde{\phi} := \widetilde{\mathbf{S}}(\phi,\boldsymbol{u}) \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ of (3.16), and the corresponding continuous dependence result becomes (1.29) with $C_{\widetilde{S}} = \dfrac{1}{\widetilde{\alpha}} = \dfrac{2\,c_p^2}{\vartheta_1}$. □

At this point we remark that the restriction on $\|\boldsymbol{u}\|_{1,\Omega}$ in Lemma 3.4 could also have been taken as $\|\boldsymbol{u}\|_{1,\Omega} < \varepsilon\frac{\vartheta_1}{c_p\,c(\Omega)}$ with any $\varepsilon \in (0,1)$. However, we have chosen $\varepsilon = \frac{1}{2}$ for simplicity and because it yields a joint maximization of the ellipticity constant of $A_{\phi,\boldsymbol{u}}$ and the upper bound for $\|\boldsymbol{u}\|_{1,\Omega}$. In addition, when dropping the term $\beta\,\|\widetilde{\phi}\|^2_{0,\Omega}$ in (3.26) we have first assumed that $\beta$ is small and then utilized the Poincaré inequality (3.5). In turn, when $\beta$ is sufficiently large, say $\beta \geq \vartheta_1$, then the aforementioned expression is kept along the whole derivation of (3.26), so that in this case the Poincaré inequality (3.5) does not need to be applied.

We end this section by introducing suitable regularity hypotheses on the operators $\mathbf{S}$ and $\widetilde{\mathbf{S}}$, which will be employed later on. In fact, for the remainder of this chapter we follow [4, eq. (3.22)], and suppose that $\boldsymbol{u}_{\mathrm{D}} \in \mathrm{H}^{1/2+\delta}(\Gamma_{\mathrm{D}})$ and $g \in \mathrm{H}^{\delta}(\Omega)$, for some $\delta \in (0,1)$ (when $n=2$) or $\delta \in (1/2,1)$ (when $n=3$). Then, we assume that for each $\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ and $(\varphi,\boldsymbol{w}) \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\times\mathbf{H}^1(\Omega)$, with $\|\phi\|_{1,\Omega} \leq r$ and $\|\varphi\|_{1,\Omega}+\|\boldsymbol{w}\|_{1,\Omega} \leq r$, $r>0$ given, there holds, respectively, $\mathbf{S}(\phi) \in \mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega)\cap\mathbb{H}^{\delta}(\Omega)\times\mathbf{H}^{1+\delta}(\Omega)$ and $\widetilde{\mathbf{S}}(\varphi,\boldsymbol{w}) \in \mathrm{H}^{1+\delta}_{\Gamma_{\mathrm{D}}}(\Omega)$, with

$$\|\mathbf{S}_1(\phi)\|_{\delta,\Omega} \,+\, \|\mathbf{S}_2(\phi)\|_{1+\delta,\Omega} \,\leq\, \widehat{C}_S(r)\,\Big\{\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} \,+\, \|\phi\|_{0,\Omega}\,\|\boldsymbol{f}\|_{\infty,\Omega}\Big\}\,, \tag{3.27}$$

and

$$\|\widetilde{\mathbf{S}}(\varphi,\boldsymbol{w})\|_{1+\delta,\Omega} \leq \widehat{C}_{\widetilde{S}}(r)\left\{\gamma_2\,|\Omega|^{1/2}\,|\boldsymbol{k}| + \|g\|_{\delta,\Omega}\right\}, \tag{3.28}$$

where $\widehat{C}_S(r)$ and $\widehat{C}_{\widetilde{S}}(r)$ are positive constants independent of $\phi$ and $(\varphi,\boldsymbol{w})$, respectively, but depending on the upper bound $r$ of their norms. The reason of the stipulated ranges for $\delta$ will be clarified in the forthcoming analysis (see below proofs of Lemmas 1.8 and 1.9). More precisely, we remark in advance that the regularity estimate (3.27) is needed in the proof of Lemma 1.8 to bound an expression of the form $\|\mathbf{S}_1(\phi)\|_{\mathbb{L}^{2p}(\Omega)}$ in terms of $\|\mathbf{S}_1(\phi)\|_{\delta,\Omega}$, and hence of the data at the right hand side of (3.27) (further details are available in the proof of [4, Lemma 3.9]). In turn, (3.28) is employed in the proof of Lemma 1.9 to bound an expression of the form $\|\nabla\widetilde{\mathbf{S}}(\varphi,\boldsymbol{w})\|_{\mathbb{L}^{2p}(\Omega)}$ in terms of $\|\widetilde{\mathbf{S}}(\varphi,\boldsymbol{w})\|_{1+\delta,\Omega}$, and hence of the data at the right hand side of (3.28) (see (3.35) below for details).

Though the actual verification of (3.27) - (3.28) is beyond the goals of this chapter, we remark that some insights confirming the feasibility of the assumed regularity for the nonlinear problem defining $\mathbf{S}$ were already provided in [4, remarks below eq. (3.22)]. In turn, the assumed regularity of the linear problem defining $\widetilde{\mathbf{S}}$ is quite standard in the realm of elliptic boundary value problems, and we just refer the interested reader to [46] or [78].

### 3.3.4 Solvability of the fixed point equation

We begin by emphasizing that the well-posedness of the uncoupled problems (3.12) and (3.16) confirms that the operators $\mathbf{S}$, $\widetilde{\mathbf{S}}$ and $\mathbf{T}$ (*cf.* Section 3.3.2) are well defined, and hence now we can address the solvability analysis of the fixed point equation (3.19). To this end, we will verify below the hypotheses of the Schauder fixed point theorem (see, e.g. [41, Theorem 9.12-1(b)]), for which we require Lipschitz continuity of the nonlinear functions $f_{\rm bk}$, $\vartheta$ and $\mu$. More precisely, we assume that there exist positive constants $L_\mu$, $L_\vartheta$, and $L_f$, such that for each $s$, $t\in\mathrm{R}$ there hold

$$|\mu(s)-\mu(t)| \leq L_\mu\,|s-t|\,, \quad |\vartheta(s)-\vartheta(t)| \leq L_\vartheta\,|s-t|\,, \quad \text{and} \quad |f_{\rm bk}(s)-f_{\rm bk}(t)| \leq L_f\,|s-t|\,. \tag{3.29}$$

We begin the analysis with the following straightforward consequence of Lemmas 3.3 and 3.4.

**Lemma 3.5.** *Given $r>0$, we let $W := \left\{\phi\in \mathrm{H}^1_{\Gamma_{\rm D}}(\Omega): \quad \|\phi\|_{1,\Omega}\leq r\right\}$, and assume that*

$$C_S\left\{\|\boldsymbol{u}_{\rm D}\|_{1/2,\Gamma_{\rm D}} + r\,\|\boldsymbol{f}\|_{\infty,\Omega}\right\} < \frac{\vartheta_1}{2\,c_p\,c(\Omega)} \qquad \text{and} \qquad C_{\widetilde{S}}\left\{\gamma_2\,|\Omega|^{1/2}\,|\boldsymbol{k}| + \|g\|_{0,\Omega}\right\} \leq r\,, \tag{3.30}$$

*where $C_S$ and $C_{\widetilde{S}}$ are the constants specified in Lemmas 3.3 and 3.4, respectively. Then $\mathbf{T}(W)\subseteq W$.*

*Proof.* It corresponds to a slight modification of the proof of [4, Lemma 3.8]. □

Next, similarly as in [4], the continuity and compactness of $\mathbf{T}$ will essentially be direct consequences of the following two Lemmas providing the continuity of $\mathbf{S}$ and $\widetilde{\mathbf{S}}$, respectively.

**Lemma 3.6.** *There exists a positive constant $C$, depending on $\mu_1$, $\kappa_1$, $\kappa_2$, $L_\mu$, $\alpha(\Omega)$, and $\delta$ (*cf.* (3.3), (3.11), (3.29), (3.23), (3.27)), such that*

$$\|\mathbf{S}(\phi)-\mathbf{S}(\varphi)\|_H \leq C\left\{\|\boldsymbol{f}\|_{\infty,\Omega}\,\|\phi-\varphi\|_{0,\Omega} + \|\mathbf{S}_1(\varphi)\|_{\delta,\Omega}\,\|\phi-\varphi\|_{\mathrm{L}^{n/\delta}(\Omega)}\right\} \quad \forall\,\phi,\,\varphi\in \mathrm{H}^1_{\Gamma_{\rm D}}(\Omega)$$

*Proof.* Even though the present bilinear form $B_\phi$ (*cf.* (3.13)) and the corresponding one from [4] differ in a couple of linear terms, the present proof is almost verbatim as [4, Lemma 3.9]), particularly concerning the utilization of the Lipschitz-continuity of $\mu$ (*cf.* (3.29)), the regularity estimate (3.27), and the Sobolev embedding Theorem (*cf.* [1, Theorem 4.12], [92, Theorem 1.3.4]), and hence further details are omitted. □

On the contrary to the foregoing lemma, and due to the fact already mentioned that the diffusivity $\vartheta$ depends now on the scalar value of the concentration instead of the magnitude of its gradient (as it is in [4]), the proof of the Lipschitz-continuity of the operator $\widetilde{\mathbf{S}}$, being more involved, differs substantially from the one given for the analogue result of [4, Lemma 3.10]. In particular, as a consequence of the aforementioned dependences, the regularity assumption (3.28), which was not required for the proof of [4, Lemma 3.10], will definitely be employed next.

**Lemma 3.7.** *Let $C_{\widetilde{S}}$ be the constant provided by Lemma 3.4. Then, there exists a positive constant $\widetilde{C}$, depending on $C_{\widetilde{S}}$, $c(\Omega)$, $L_f$, $L_\vartheta$ and $\delta$ (*cf.* (3.10), (3.29), (3.28)), such that for all $(\phi,\boldsymbol{u})$, $(\varphi,\boldsymbol{w}) \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\times\mathbf{H}^1(\Omega)$, with $\|\boldsymbol{u}\|_{1,\Omega}$, $\|\boldsymbol{w}\|_{1,\Omega} < \frac{\vartheta_1}{2\,c_p\,c(\Omega)}$, there holds*

$$\begin{aligned}&\|\widetilde{\mathbf{S}}(\phi,\boldsymbol{u})-\widetilde{\mathbf{S}}(\varphi,\boldsymbol{w})\|_{1,\Omega}\\ &\quad\leq\ \widetilde{C}\Big\{|\boldsymbol{k}|\,\|\phi-\varphi\|_{0,\Omega}\,+\,\|\widetilde{\mathbf{S}}(\varphi,\boldsymbol{w})\|_{1,\Omega}\|\boldsymbol{u}-\boldsymbol{w}\|_{1,\Omega}\,+\,\|\widetilde{\mathbf{S}}(\varphi,\boldsymbol{w})\|_{1+\delta,\Omega}\,\|\phi-\varphi\|_{\mathrm{L}^{n/\delta}(\Omega)}\Big\}\,.\end{aligned}\tag{3.31}$$

*Proof.* Given $(\phi,\boldsymbol{u})$, $(\varphi,\boldsymbol{w})$ as stated, we let $\widetilde{\phi}:=\widetilde{\mathbf{S}}(\phi,\boldsymbol{u})$ and $\widetilde{\varphi}:=\widetilde{\mathbf{S}}(\varphi,\boldsymbol{w})$, that is (*cf.* (3.16))

$$A_{\phi,\boldsymbol{u}}(\widetilde{\phi},\widetilde{\psi})\,=\,G_\phi(\widetilde{\psi})\quad\text{and}\quad A_{\varphi,\boldsymbol{w}}(\widetilde{\varphi},\widetilde{\psi})\,=\,G_\varphi(\widetilde{\psi})\qquad\forall\,\widetilde{\psi}\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,.$$

It follows, according to the ellipticity of $A_{\phi,\boldsymbol{u}}$ with constant $\widetilde{\alpha}$, and then subtracting and adding $G_\varphi(\widetilde{\phi}-\widetilde{\varphi})\,=\,A_{\varphi,\boldsymbol{w}}(\widetilde{\varphi},\widetilde{\phi}-\widetilde{\varphi})$, that

$$\begin{aligned}\widetilde{\alpha}\,\|\widetilde{\phi}-\widetilde{\varphi}\|^2_{1,\Omega}\,&\leq\,A_{\phi,\boldsymbol{u}}(\widetilde{\phi},\widetilde{\phi}-\widetilde{\varphi})\,-\,A_{\phi,\boldsymbol{u}}(\widetilde{\varphi},\widetilde{\phi}-\widetilde{\varphi})\\ &=\quad G_\phi(\widetilde{\phi}-\widetilde{\varphi})-G_\varphi(\widetilde{\phi}-\widetilde{\varphi})\,+\,A_{\varphi,\boldsymbol{w}}(\widetilde{\varphi},\widetilde{\phi}-\widetilde{\varphi})-A_{\phi,\boldsymbol{u}}(\widetilde{\varphi},\widetilde{\phi}-\widetilde{\varphi})\\ &=\quad\int_\Omega\big(f_{\mathrm{bk}}(\phi)-f_{\mathrm{bk}}(\varphi)\big)\,\boldsymbol{k}\cdot\nabla(\widetilde{\phi}-\widetilde{\varphi})\,+\,\int_\Omega\widetilde{\varphi}\,(\boldsymbol{u}-\boldsymbol{w})\cdot\nabla(\widetilde{\phi}-\widetilde{\varphi})\\ &+\quad\int_\Omega\Big(\vartheta(\varphi)-\vartheta(\phi)\Big)\nabla\widetilde{\varphi}\cdot\nabla(\widetilde{\phi}-\widetilde{\varphi})\,,\end{aligned}\tag{3.32}$$

where the last equality has employed the definitions given by (3.17) and (3.18). Then applying Cauchy-Schwarz's inequality, the Lipschitz-continuity assumption (3.29) on the last term in (3.32), and then Hölder's inequality, we obtain

$$\begin{aligned}\widetilde{\alpha}\,\|\widetilde{\phi}-\widetilde{\varphi}\|^2_{1,\Omega}\,\leq\,&\Big\{L_f\,|\boldsymbol{k}|\,\|\phi\,-\,\varphi\|_{0,\Omega}\,+\,c(\Omega)\,\|\widetilde{\varphi}\|_{1,\Omega}\,\|\boldsymbol{u}-\boldsymbol{w}\|_{1,\Omega}\Big\}\,|\widetilde{\phi}-\widetilde{\varphi}|_{1,\Omega}\\ &+\quad L_\vartheta\,\|\phi-\varphi\|_{\mathrm{L}^{2q}(\Omega)}\,\|\nabla\widetilde{\varphi}\|_{\mathrm{L}^{2p}(\Omega)}\,|\widetilde{\phi}-\widetilde{\varphi}|_{1,\Omega}\,,\end{aligned}\tag{3.33}$$

where $p,q\in[1,+\infty)$ are such that $1/p+1/q=1$. Next, given the further regularity $\delta$ assumed in (3.28), we recall that the Sobolev embedding Theorem (*cf.* [1, Theorem 4.12], [92, Theorem 1.3.4])

establishes the continuous injection $i_\delta : \mathrm{H}^\delta(\Omega) \longrightarrow \mathrm{L}^{\delta^*}(\Omega)$ with boundedness constant $C_\delta$, where

$$\delta^* := \begin{cases} \dfrac{2}{1-\delta} & \text{if} \quad n = 2\,, \\ \\ \dfrac{6}{3-2\delta} & \text{if} \quad n = 3\,. \end{cases} \tag{3.34}$$

Thus, choosing $p$ such that $2p = \delta^*$, we find that

$$\|\nabla\widetilde{\varphi}\|_{\mathrm{L}^{2p}(\Omega)} \,=\, \|\nabla\widetilde{\mathbf{S}}(\varphi,\boldsymbol{w})\|_{\mathrm{L}^{2p}(\Omega)} \,\leq\, C_\delta\,\|\nabla\widetilde{\mathbf{S}}(\varphi,\boldsymbol{w})\|_{\delta,\Omega} \,\leq\, C_\delta\,\|\widetilde{\mathbf{S}}(\varphi,\boldsymbol{w})\|_{1+\delta,\Omega}\,. \tag{3.35}$$

In turn, according to the above choice of $p$, that is $p = \delta^*/2$, it readily follows that

$$2q \,:=\, \frac{2p}{p-1} \,=\, \begin{cases} \dfrac{2}{\delta} & \text{if} \quad n = 2\,, \\ \\ \dfrac{3}{\delta} & \text{if} \quad n = 3 \end{cases} \,=\, \frac{n}{\delta}. \tag{3.36}$$

In this way, inequalities (3.32), (3.33), and (3.35) together with identity (3.36) imply (3.31), which finishes the proof. □

The following result, which is the analogue of [4, Lemma 3.11], is a straightforward corollary of Lemmas 3.5, 3.6, and 3.7.

**Lemma 3.8.** *Given $r > 0$, we let $W := \Big\{\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) : \ \|\phi\|_{1,\Omega} \leq r\Big\}$, and assume (3.30) (cf. Lemma 3.5). Then, with the constants $C$ and $\widetilde{C}$ from Lemmas 3.6 and 3.7, for all $\phi$, $\varphi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ there holds*

$$\begin{aligned} \|\mathbf{T}(\phi) - \mathbf{T}(\varphi)\|_{1,\Omega} \,&\leq\, \widetilde{C}\Big\{|\boldsymbol{k}| \,+\, C\,\|\mathbf{T}(\varphi)\|_{1,\Omega}\,\|\boldsymbol{f}\|_{\infty,\Omega}\Big\}\,\|\phi - \varphi\|_{0,\Omega} \\ &+\quad \widetilde{C}\Big\{C\,\|\mathbf{T}(\varphi)\|_{1,\Omega}\,\|\mathbf{S}_1(\varphi)\|_{\delta,\Omega} \,+\, \|\mathbf{T}(\varphi)\|_{1+\delta,\Omega}\Big\}\,\|\phi - \varphi\|_{\mathrm{L}^{n/\delta}(\Omega)}\,. \end{aligned} \tag{3.37}$$

*Proof.* It suffices to recall from Section 3.3.2 that $\mathbf{T}(\phi) \,=\, \widetilde{\mathbf{S}}(\phi, \mathbf{S}_2(\phi)) \quad \forall\, \phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$, and then apply Lemmas 3.5, 3.6, and 3.7. □

The announced properties of $\mathbf{T}$ are proved now.

**Lemma 3.9.** *Given $r > 0$, we let $W := \Big\{\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) : \ \|\phi\|_{1,\Omega} \leq r\Big\}$, and assume (3.30) (cf. Lemma 3.5). Then, $\mathbf{T} : W \longrightarrow W$ is continuous and $\overline{\mathbf{T}(W)}$ is compact.*

*Proof.* It follows almost verbatim as the proof of [4, Lemma 3.12]. Indeed, it is basically consequence of the Rellich-Kondrachov compactness Theorem (*cf.* [1, Theorem 6.3], [92, Theorem 1.3.5]), the specified range of the constant $\delta$ involved in the further regularity assumptions given by (3.27) and (3.28), and the well-known fact that every bounded sequence in a Hilbert space has a weakly convergent subsequence. We omit the rest of details. □

Finally, the main result of this section is given as follows, where the proof can be obtained very much as in [4, Theorem 3.13].

**Theorem 3.1.** *Assume that the hypotheses of the Lemmas 3.5 – 3.9 are met. Then the augmented mixed-primal problem (3.15) has at least one solution* $(\boldsymbol{\sigma}, \boldsymbol{u}, \phi) \in \mathbb{H}_{\mathrm{N}}(\mathbf{div}; \Omega) \times \mathbf{H}^1(\Omega) \times \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ *with* $\phi \in W$, *and there holds*

$$\|\phi\|_{1,\Omega} \, \leq \, C_{\widetilde{S}} \left\{ \gamma_2 \, |\Omega|^{1/2} \, |\boldsymbol{k}| \, + \, \|g\|_{\delta,\Omega} \right\} , \tag{3.38}$$

*and*

$$\|(\boldsymbol{\sigma}, \boldsymbol{u})\|_H \, \leq \, C_S \left\{ \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} \, + \, r \, \|\boldsymbol{f}\|_{\infty,\Omega} \right\} , \tag{3.39}$$

*where* $C_S$ *and* $C_{\widetilde{S}}$ *are the constants specified in Lemmas 3.3 and 3.4, respectively. Moreover, if the data* $\boldsymbol{k}$, $\boldsymbol{f}$, $g$ *and* $\boldsymbol{u}_{\mathrm{D}}$ *are sufficiently small so that, with the constants* $C$, $\widetilde{C}$, $\widehat{C}_S(r)$, *and* $\widehat{C}_{\widetilde{S}}(r)$ *from Lemmas 3.6 and 3.7, and estimates (3.27) and (3.28), and denoting by* $\widetilde{C}_\delta$ *the boundedness constant of the continuous injection of* $\mathrm{H}^1(\Omega)$ *into* $\mathrm{L}^{n/\delta}(\Omega)$, *there holds*

$$\begin{aligned} \widetilde{C} \, (1 + \widehat{C}_{\widetilde{S}}(r) \, \widetilde{C}_\delta \, C \, \gamma_2 \, |\Omega|^{1/2}) \, |\boldsymbol{k}| \, + \, \widehat{C}_{\widetilde{S}}(r) \, \widetilde{C} \, \widetilde{C}_\delta \, \|g\|_{\delta,\Omega} \, + \, \widetilde{C} \, \widetilde{C}_\delta \, C \, \widehat{C}_S(r) \, \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} \\ + \quad r \, \widetilde{C} \, C \left( 1 \, + \, r \, \widetilde{C}_\delta \, \widehat{C}_S(r) \right) \|\boldsymbol{f}\|_{\infty,\Omega} \, < \, 1 \, , \end{aligned} \tag{3.40}$$

*then the solution* $\phi$ *is unique in* $W$.

*Proof.* According to the equivalence between (3.15) and the fixed point equation (3.19), and thanks to the previous Lemmas 3.5 and 3.9, the existence of solution is just a straightforward application of the Schauder fixed point theorem (*cf.* [41, Theorem 9.12-1(b)]). In turn, the estimates (3.38) and (3.39) follow from (3.20) (*cf.* Lemma 3.3) and (3.25) (*cf.* Lemma 3.4). Furthermore, given another solution $\varphi \in W$ of (3.19), the estimates $\|\mathbf{T}(\varphi)\|_{1,\Omega} \, = \, \|\varphi\|_{1,\Omega} \, \leq \, r$,

$$\|\mathbf{S}_1(\varphi)\|_{\delta,\Omega} \, \leq \, \widehat{C}_S(r) \left\{ \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} \, + \, \|\varphi\|_{0,\Omega} \, \|\boldsymbol{f}\|_{\infty,\Omega} \right\} \qquad (\text{cf. (3.27)}) \, ,$$

$$\|\widetilde{\varphi}\|_{1+\delta,\Omega} \, \leq \, \widehat{C}_{\widetilde{S}}(r) \left\{ \gamma_2 \, |\Omega|^{1/2} \, |\boldsymbol{k}| \, + \, \|g\|_{\delta,\Omega} \right\} \qquad (\text{cf. (3.28)}) \, ,$$

and

$$\|\psi\|_{\mathrm{L}^{n/\delta}(\Omega)} \, \leq \, \widetilde{C}_\delta \, \|\psi\|_{1,\Omega} \quad \forall \, \psi \in \mathrm{H}^1(\Omega) \, , \tag{3.41}$$

confirm (3.40) as a sufficient condition for concluding, together with (3.37), that $\phi = \varphi$.

In other words, (3.40) constitutes the condition arising from (3.37) – once (3.41), and the *a priori* and regularity estimates for $\|\mathbf{T}(\varphi)\|_{1,\Omega}$, $\|\mathbf{S}_1(\varphi)\|_{\delta,\Omega}$ and $\|\mathbf{T}(\varphi)\|_{1+\delta,\Omega}$, respectively, are employed – that makes the operator $\mathbf{T}$ to become a contraction, thus yielding the existence of a unique fixed point of $\mathbf{T}$ in $W$. □

## 3.4 The Galerkin scheme

Let $\mathcal{T}_h$ be a regular triangulation of $\Omega$ by triangles $K$ (resp. tetrahedra $K$ in $\mathrm{R}^3$) of diameter $h_K$, and define the meshsize $h := \max \left\{ h_K : \, K \in \mathcal{T}_h \right\}$. In addition, given an integer $k \geq 0$, for each $K \in \mathcal{T}_h$ we let $\mathrm{P}_k(K)$ be the space of polynomial functions on $K$ of degree $\leq k$, and define the corresponding local Raviart-Thomas space of order $k$ as

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

where, according to the notations described in Section 3.1, $\mathbf{P}_k(K) \, = \, [\mathrm{P}_k(K)]^n$, and $\boldsymbol{x}$ is the generic vector in $\mathrm{R}^n$. Then, we introduce the finite element subspaces approximating the unknowns $\boldsymbol{\sigma}$, $\boldsymbol{u}$, and

$\phi$, respectively, as the global Raviart-Thomas space of order $k$, and the corresponding Lagrange spaces given by the continuous piecewise polynomials of degree $\leq k+1$, that is

$$\mathbb{H}_h^{\boldsymbol{\sigma}} := \left\{ \boldsymbol{\tau}_h \in \mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega) : \quad \mathbf{c}^{\mathtt{t}}\,\boldsymbol{\tau}_h|_K \in \mathbf{RT}_k(K) \quad \forall\, \mathbf{c} \in \mathrm{R}^n\,, \quad \forall\, K \in \mathcal{T}_h \right\}, \tag{3.42}$$

$$\mathbf{H}_h^{\boldsymbol{u}} := \left\{ \boldsymbol{v}_h \in \mathbf{C}(\Omega) : \quad \boldsymbol{v}_h|_K \in \mathbf{P}_{k+1}(K) \quad \forall\, K \in \mathcal{T}_h \right\}, \tag{3.43}$$

$$\mathrm{H}_h^{\phi} := \left\{ \psi_h \in C(\Omega) \cap \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) : \quad \psi_h|_K \in \mathrm{P}_{k+1}(K) \quad \forall\, K \in \mathcal{T}_h \right\}. \tag{3.44}$$

In this way, the underlying Galerkin scheme, given by the discrete counterpart of (3.15), reads: Find $(\boldsymbol{\sigma}_h, \boldsymbol{u}_h, \phi_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}} \times \mathrm{H}_h^{\phi}$ such that

$$\begin{aligned} B_{\phi_h}((\boldsymbol{\sigma}_h, \boldsymbol{u}_h), (\boldsymbol{\tau}_h, \boldsymbol{v}_h)) &= F_{\phi_h}(\boldsymbol{\tau}_h, \boldsymbol{v}_h) \quad \forall (\boldsymbol{\tau}_h, \boldsymbol{v}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}\,, \\ A_{\boldsymbol{u}_h}(\phi_h, \psi_h) &= \int_\Omega f_{\mathrm{bk}}(\phi_h)\, \boldsymbol{k} \cdot \nabla \psi_h + \int_\Omega g \psi_h \qquad \forall\, \psi_h \in \mathrm{H}_h^{\phi}\,. \end{aligned} \tag{3.45}$$

Throughout the rest of this section we adopt the discrete analogue of the fixed point strategy introduced in Section 3.3.2. Hence, we now let $\mathbf{S}_h : \mathrm{H}_h^{\phi} \longrightarrow \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}$ be the operator defined by

$$\mathbf{S}_h(\phi_h) = (\mathbf{S}_{1,h}(\phi_h), \mathbf{S}_{2,h}(\phi_h)) := (\boldsymbol{\sigma}_h, \boldsymbol{u}_h) \qquad \forall\, \phi_h \in \mathrm{H}_h^{\phi}\,,$$

where $(\boldsymbol{\sigma}_h, \boldsymbol{u}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}$ is the unique solution of

$$B_{\phi_h}((\boldsymbol{\sigma}_h, \boldsymbol{u}_h), (\boldsymbol{\tau}_h, \boldsymbol{v}_h)) = F_{\phi_h}(\boldsymbol{\tau}_h, \boldsymbol{v}_h) \quad \forall\, (\boldsymbol{\tau}_h, \boldsymbol{v}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}\,, \tag{3.46}$$

with $B_{\phi_h}$ and $F_{\phi_h}$ being defined by (3.13) and (3.14), respectively, with $\phi = \phi_h$. In addition, we let $\widetilde{\mathbf{S}}_h : \mathrm{H}_h^{\phi} \times \mathbf{H}_h^{\boldsymbol{u}} \longrightarrow \mathrm{H}_h^{\phi}$ be the operator defined by

$$\widetilde{\mathbf{S}}_h(\phi_h, \boldsymbol{u}_h) := \widetilde{\phi}_h \qquad \forall\, (\phi_h, \boldsymbol{u}_h) \in \mathrm{H}_h^{\phi} \times \mathbf{H}_h^{\boldsymbol{u}}\,,$$

where $\widetilde{\phi}_h \in \mathrm{H}_h^{\phi}$ is the unique solution of

$$A_{\phi_h, \boldsymbol{u}_h}(\widetilde{\phi}_h, \widetilde{\psi}_h) = G_{\phi_h}(\widetilde{\psi}_h) \qquad \forall\, \widetilde{\psi}_h \in \mathrm{H}_h^{\phi}\,, \tag{3.47}$$

with $A_{\phi_h, \boldsymbol{u}_h}$ and $G_{\phi_h}$ being defined by (3.17) and (3.18), respectively, with $\boldsymbol{u} = \boldsymbol{u}_h$ and $\phi = \phi_h$. Finally, we define the operator $\mathbf{T}_h : \mathrm{H}_h^{\phi} \longrightarrow \mathrm{H}_h^{\phi}$ by

$$\mathbf{T}_h(\phi_h) := \widetilde{\mathbf{S}}_h(\phi_h, \mathbf{S}_{2,h}(\phi_h)) \qquad \forall\, \phi_h \in \mathrm{H}_h^{\phi}\,,$$

and realize that (3.45) can be rewritten, equivalently, as: Find $\phi_h \in \mathrm{H}_h^{\phi}$ such that

$$\mathbf{T}_h(\phi_h) = \phi_h\,. \tag{3.48}$$

Certainly, all the above makes sense if we guarantee that the discrete problems (3.46) and (3.47) are well-posed. Indeed, it is easy to see that the respective proofs are almost verbatim of the continuous analogues provided in Section 3.3.3, and hence we simply state the corresponding results as follows.

**Lemma 3.10.** *Assume that* $\kappa_1 \in \left(0, \frac{2\delta\mu_1}{\mu_2}\right)$ *and* $\kappa_2 \in \left(0, \frac{2\widetilde{\delta}\,\alpha_{\boldsymbol{K}}}{n\|\boldsymbol{K}^{-1}\|_\infty}\right)$, *with* $\delta \in (0, 2\mu_1)$ *and* $\widetilde{\delta} \in \left(0, \frac{2}{n\|\boldsymbol{K}^{-1}\|_\infty}\right)$. *Then, for each* $\phi_h \in \mathrm{H}_h^{\phi}$ *the problem (3.46) has a unique solution* $\mathbf{S}_h(\phi_h) := (\boldsymbol{\sigma}_h, \boldsymbol{u}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}$. *Moreover, with the same constant* $C_S > 0$ *from Lemma 3.3, there holds*

$$\|\mathbf{S}_h(\phi_h)\|_H = \|(\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\|_H \leq C_S \left\{ \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + \|\phi_h\|_{0,\Omega}\, \|\boldsymbol{f}\|_{\infty,\Omega} \right\} \qquad \forall\, \phi_h \in \mathrm{H}_h^{\phi}\,.$$

**Lemma 3.11.** *Let $\phi_h \in \mathrm{H}_h^\phi$ and $\boldsymbol{u}_h \in \mathbf{H}_h^{\boldsymbol{u}}$ such that $\|\boldsymbol{u}_h\|_{1,\Omega} < \frac{\vartheta_1}{2\,c_p\,c(\Omega)}$ (cf. (3.3), (3.5), (3.10)). Then, there exists a unique $\widetilde{\phi}_h := \widetilde{\mathbf{S}}_h(\phi_h, \boldsymbol{u}_h) \in \mathrm{H}_h^\phi$ solution of (3.47). Moreover, with the same constant $C_{\widetilde{S}} > 0$ from Lemma 3.4, there holds*

$$\|\widetilde{\mathbf{S}}_h(\phi_h, \boldsymbol{u}_h)\|_{1,\Omega} = \|\widetilde{\phi}_h\|_{1,\Omega} \leq C_{\widetilde{S}} \left\{ \gamma_2 \, |\Omega|^{1/2} \, |\boldsymbol{k}| + \|g\|_{0,\Omega} \right\}.$$

We now aim to show the solvability of (3.45) by analyzing the equivalent fixed point equation (3.48). To this end, in what follows we verify the hypotheses of the Brouwer fixed point theorem (*cf.* [41, Theorem 9.9-2]). We begin with the discrete version of Lemma 3.5.

**Lemma 3.12.** *Given $r > 0$, we let $W_h := \left\{ \phi_h \in \mathrm{H}_h^\phi : \quad \|\phi_h\|_{1,\Omega} \leq r \right\}$, and assume (3.30) (cf. Lemma 3.5). Then $\mathbf{T}_h(W_h) \subseteq W_h$.*

*Proof.* It is a straightforward consequence of Lemmas 3.10 and 3.11. □

The discrete analogues of Lemmas 3.6 and 3.7 are provided next. We notice in advance that, instead of the regularity assumptions employed in the proof of those results, which actually are not needed nor could be applied in the present discrete case, we simply utilize a $\mathrm{L}^4 - \mathrm{L}^4 - \mathrm{L}^2$ argument.

**Lemma 3.13.** *There exist a positive constant $C$, depending on $\mu_1$, $\kappa_1$, $\kappa_2$, $L_\mu$, and $\alpha(\Omega)$ (cf. (3.3), (3.11), (3.29), (3.23)), such that*

$$\|\mathbf{S}_h(\phi_h) - \mathbf{S}_h(\varphi_h)\|_H \leq C \left\{ \|\boldsymbol{f}\|_{\infty,\Omega} \, \|\phi_h - \varphi_h\|_{0,\Omega} + \|\mathbf{S}_{1,h}(\varphi_h)\|_{\mathbb{L}^4(\Omega)} \, \|\phi_h - \varphi_h\|_{\mathrm{L}^4(\Omega)} \right\},$$

*for all $\phi_h$, $\varphi_h \in \mathrm{H}_h^\phi$.*

*Proof.* It proceeds exactly as in the proof of Lemma 3.6 (see [4, Lemma 3.9]), except for the derivation of the discrete analogue of [4, eq. (3.29), Lemma 3.9], where, instead of choosing the values of $p$ and $q$ determined by the regularity parameter $\delta$, it suffices to take $p = q = 2$, thus obtaining

$$\begin{aligned} &\left| (B_{\varphi_h} - B_{\phi_h})\big((\boldsymbol{\zeta}_h, \boldsymbol{w}_h), (\sigma_h, \boldsymbol{u}_h) - (\boldsymbol{\zeta}_h, \boldsymbol{w}_h)\big) \right| \\ &\leq \frac{L_\mu (1 + \kappa_1^2)^{1/2}}{\mu_1^2} \, \|\boldsymbol{\zeta}_h\|_{\mathbb{L}^4(\Omega)} \, \|\phi_h - \varphi_h\|_{\mathrm{L}^4(\Omega)} \, \|(\boldsymbol{\sigma}_h, \boldsymbol{u}_h) - (\boldsymbol{\zeta}_h, \boldsymbol{w}_h)\|_H, \end{aligned}$$

for all $\phi_h$, $\varphi_h \in \mathrm{H}_h^\phi$, with $(\sigma_h, \boldsymbol{u}_h) := \mathbf{S}_h(\phi_h)$ and $(\boldsymbol{\zeta}_h, \boldsymbol{w}_h) := \mathbf{S}_h(\varphi)$. Thus, the fact that the elements of $\mathbb{H}_h^{\boldsymbol{\sigma}}$ are piecewise polynomials insures that $\|\boldsymbol{\zeta}_h\|_{\mathbb{L}^4(\Omega)} < +\infty$ for each $\boldsymbol{\zeta}_h \in \mathbb{H}_h^{\boldsymbol{\sigma}}$. Further details are omitted. □

**Lemma 3.14.** *Let $C_{\widetilde{S}}$ be the constant provided by Lemma 3.4. Then, there exists a positive constant $\widetilde{C}$, depending on $C_{\widetilde{S}}$, $c(\Omega)$, $L_f$, and $L_\vartheta$ (cf. (3.10), (3.29)), such that for all $(\phi_h, \boldsymbol{u}_h)$, $(\varphi_h, \boldsymbol{w}_h) \in \mathrm{H}_h^\phi \times \mathbf{H}_h^{\boldsymbol{u}}$, with $\|\boldsymbol{u}_h\|_{1,\Omega}$, $\|\boldsymbol{w}_h\|_{1,\Omega} < \frac{\vartheta_1}{2\,c_p\,c(\Omega)}$, there holds*

$$\begin{aligned} \|\widetilde{\mathbf{S}}_h(\phi_h, \boldsymbol{u}_h) - \widetilde{\mathbf{S}}_h(\varphi_h, \boldsymbol{w}_h)\|_{1,\Omega} &\leq \widetilde{C} \Big\{ |\boldsymbol{k}| \, \|\phi_h - \varphi_h\|_{0,\Omega} \\ + \quad \|\widetilde{\mathbf{S}}_h(\varphi_h, \boldsymbol{w}_h)\|_{1,\Omega} \, \|\boldsymbol{u}_h - \boldsymbol{w}_h\|_{1,\Omega} &+ \|\nabla \widetilde{\mathbf{S}}_h(\varphi_h, \boldsymbol{w}_h)\|_{\mathrm{L}^4(\Omega)} \, \|\phi_h - \varphi_h\|_{\mathrm{L}^4(\Omega)} \Big\}. \end{aligned} \tag{3.49}$$

*Proof.* Given $(\phi_h, \boldsymbol{u}_h)$ and $(\varphi_h, \boldsymbol{w}_h)$ as stated, we first let $\widetilde{\phi}_h := \widetilde{\mathbf{S}}_h(\phi_h, \boldsymbol{u}_h)$ and $\widetilde{\varphi}_h := \widetilde{\mathbf{S}}_h(\varphi_h, \boldsymbol{w}_h)$. Next, we proceed analogously as in the proof of Lemma 3.7, except for the derivation of the discrete analogue of the third term in (3.33), where, employing the same argument of the previous Lemma 3.13, it suffices to take $p = q = 2$, thus obtaining

$$\widetilde{\alpha}\,\|\widetilde{\phi}_h - \widetilde{\varphi}_h\|^2_{1,\Omega} \leq \Big\{ L_f\,|\boldsymbol{k}|\,\|\phi_h - \varphi_h\|_{0,\Omega} + c(\Omega)\,\|\widetilde{\varphi}_h\|_{1,\Omega}\,\|\boldsymbol{u}_h - \boldsymbol{w}_h\|_{1,\Omega} \Big\} |\widetilde{\phi}_h - \widetilde{\varphi}_h|_{1,\Omega}$$
$$+ \quad L_\vartheta\,\|\phi_h - \varphi_h\|_{\mathrm{L}^4(\Omega)}\,\|\nabla\widetilde{\varphi}_h\|_{\mathrm{L}^4(\Omega)}\,|\widetilde{\phi}_h - \widetilde{\varphi}_h|_{1,\Omega}\,.$$

Then, since the elements of $\mathrm{H}^\phi_h$ are piecewise polynomials it follows that $\|\nabla\widetilde{\varphi}_h\|_{\mathrm{L}^4(\Omega)} < +\infty$, and hence the foregoing equation yields (3.49). Further details are omitted. □

Now, utilizing Lemmas 3.13 and 3.14, we can prove the discrete version of Lemma 3.8.

**Lemma 3.15.** *Suppose that the assumptions in Lemma 3.12 are satisfied. Then, with the constants $C$ and $\widetilde{C}$ from Lemmas 3.13 and 3.14, for all $\phi_h$, $\varphi_h \in \mathrm{H}^\phi_h$ there holds*

$$\begin{aligned} \|\mathbf{T}_h(\phi_h) - \mathbf{T}_h(\varphi_h)\|_{1,\Omega} &\leq \widetilde{C}\,\Big\{ |\boldsymbol{k}| + C\,\|\mathbf{T}_h(\varphi_h)\|_{1,\Omega}\,\|\boldsymbol{f}\|_{\infty,\Omega} \Big\}\,\|\phi_h - \varphi_h\|_{0,\Omega} \\ &+ \widetilde{C}\,\Big\{ C\,\|\mathbf{T}_h(\varphi_h)\|_{1,\Omega}\,\|\mathbf{S}_{1,h}(\varphi_h)\|_{\mathbb{L}^4(\Omega)} + \|\nabla\mathbf{T}_h(\varphi_h)\|_{\mathbb{L}^4(\Omega)} \Big\}\,\|\phi_h - \varphi_h\|_{\mathrm{L}^4(\Omega)}\,. \end{aligned} \tag{3.50}$$

Consequently, since the foregoing lemma and the continuous injection of $\mathrm{H}^1(\Omega)$ into $\mathrm{L}^4(\Omega)$ confirm the continuity of $\mathbf{T}_h$, we conclude, thanks to the Brouwer fixed point theorem (*cf.* [41] [Theorem 9.9-2]) and Lemmas 3.12 and 3.15, the main result of this section.

**Theorem 3.2.** *Under the assumptions of Lemma 3.12, the Galerkin scheme (3.45) has at least one solution $(\boldsymbol{\sigma}_h, \boldsymbol{u}_h, \phi_h) \in \mathbb{H}^{\boldsymbol{\sigma}}_h \times \mathbf{H}^{\boldsymbol{u}}_h \times \mathrm{H}^\phi_h$ with $\phi_h \in W_h$, and there holds*

$$\|(\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\|_H \leq C_S \Big\{ \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + |\boldsymbol{k}|\,\|\phi_h\|_{1,\Omega} \Big\}\,,$$

*and*

$$\|\phi_h\|_{1,\Omega} \leq C_{\widetilde{S}} \Big\{ \gamma_2\,|\Omega|^{1/2}\,|\boldsymbol{k}| + \|g\|_{0,\Omega} \Big\}\,,$$

*where $C_S$ and $C_{\widetilde{S}}$ are the constants provided by Lemmas 3.3 and 3.4, respectively.*

We end this section by remarking that the lack of suitable estimates for $\|\mathbf{S}_{1,h}(\varphi_h)\|_{\mathbb{L}^4(\Omega)}$ and $\|\nabla\mathbf{T}_h(\varphi_h)\|_{\mathbb{L}^4(\Omega)}$ stops us of trying to use (3.50) to derive a contraction estimate for $\mathbf{T}_h$. This is the reason why in the foregoing Theorem 3.2 we are able only to guarantee existence, but not uniqueness, of a discrete solution.

## 3.5 A priori error analysis

Given $(\boldsymbol{\sigma}, \boldsymbol{u}, \phi) \in \mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega) \times \mathbf{H}^1(\Omega) \times \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ with $\phi \in W$, and $(\boldsymbol{\sigma}_h, \boldsymbol{u}_h, \phi_h) \in \mathbb{H}^{\boldsymbol{\sigma}}_h \times \mathbf{H}^{\boldsymbol{u}}_h \times \mathrm{H}^\phi_h$ with $\phi_h \in W_h$, solutions of (3.15) and (3.45), respectively, we now aim to derive a corresponding *a priori* error estimate. For this purpose, we first observe from (3.15) and (3.45) that the above problems can be rewritten as two pairs of corresponding continuous and discrete formulations, namely

$$\begin{aligned} B_\phi((\boldsymbol{\sigma}, \boldsymbol{u}), (\boldsymbol{\tau}, \boldsymbol{v})) &= F_\phi(\boldsymbol{\tau}, \boldsymbol{v}) \qquad \forall (\boldsymbol{\tau}, \boldsymbol{v}) \in \mathbb{H}_{\mathrm{N}}(\mathbf{div};\Omega) \times \mathbf{H}^1(\Omega)\,, \\ B_{\phi_h}((\boldsymbol{\sigma}_h, \boldsymbol{u}_h), (\boldsymbol{\tau}_h, \boldsymbol{v}_h)) &= F_{\phi_h}(\boldsymbol{\tau}_h, \boldsymbol{v}_h) \qquad \forall (\boldsymbol{\tau}_h, \boldsymbol{v}_h) \in \mathbb{H}^{\boldsymbol{\sigma}}_h \times \mathbf{H}^{\boldsymbol{u}}_h\,, \end{aligned} \tag{3.51}$$

and

$$\begin{aligned} A_{\boldsymbol{u}}(\phi,\psi) &= G_\phi(\psi) \qquad \forall\, \psi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\,, \\ A_{\boldsymbol{u}_h}(\phi_h,\psi_h) &= G_{\phi_h}(\psi_h) \qquad \forall\, \psi_h \in \mathrm{H}^\phi_h\,. \end{aligned} \tag{3.52}$$

Then, as suggested by the structure of the foregoing systems, in what follows we apply a suitable Strang-type lemma valid for linear problems to (3.51), and then derive our own Strang-type estimate for (3.52). The reason of the latter is that the present form $A_{\boldsymbol{u}}$ is not strongly monotone as it was in [4] where $\vartheta$ depended on $|\nabla\phi|$ instead of just $\phi$, and hence it does not fit the corresponding Strang-type estimates for nonlinear problems (see, e.g. [4, Lemma 5.1]).

We begin our analysis by recalling from [40] the first Strang Lemma for linear problems.

**Lemma 3.16.** *Let $H$ be a Hilbert space, $F \in H'$, and $A : H \times H \to \mathrm{R}$ a bounded and elliptic bilinear form. In addition, let $\{H_n\}_{n\in\mathrm{N}}$ be a sequence of finite dimensional subspaces of $H$, and for each $n \in \mathrm{N}$ consider a functional $F_n \in H_n'$ and a bounded bilinear form $A_n : H_n \times H_n \to \mathrm{R}$. Assume that the family $\{A\} \cup \{A_h\}_{n\in\mathrm{N}}$ is uniformly bounded and uniformly elliptic with constants $\Lambda_B$ and $\Lambda_E$, respectively. In turn, let $u \in H$ and $u_n \in H_n$ such that*

$$A(u,v) = F(v) \qquad \forall\, v \in H \quad \text{and} \quad A_n(u_n,v_n) = F_n(v_n) \qquad \forall\, v_n \in H_n\,.$$

*Then for each $n \in \mathrm{N}$ there holds*

$$\|u-u_n\|_H \leq C_{\mathtt{ST}} \left\{ \sup_{\substack{w_n\in H_n \\ w_n\neq \mathbf{0}}} \frac{\big|\, F(w_n) - F_n(w_n)\,\big|}{\|w_n\|_H} + \inf_{\substack{v_n\in H_n \\ v_n\neq \mathbf{0}}} \left( \|u-v_n\|_H + \sup_{\substack{w_n\in H_n \\ w_n\neq \mathbf{0}}} \frac{\big|\, A(v_n,w_n) - A_n(v_n,w_n)\,\big|}{\|w_n\|_H} \right) \right\},$$

*with* $C_{\mathtt{ST}} := \Lambda_E^{-1} \max\big\{1, \Lambda_E + \Lambda_B\big\}$.

*Proof.* See [40, Lemma 4.1.1]. □

We now denote as usual

$$\mathrm{dist}\big(\phi, \mathrm{H}^\phi_h\big) := \inf_{\varphi_h\in \mathrm{H}^\phi_h} \|\phi - \varphi_h\|_{1,\Omega}\,,$$

and

$$\mathrm{dist}\big((\boldsymbol{\sigma},\boldsymbol{u}), \mathbb{H}^{\boldsymbol{\sigma}}_h \times \mathbf{H}^{\boldsymbol{u}}_h\big) := \inf_{(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\in \mathbb{H}^{\boldsymbol{\sigma}}_h \times \mathbf{H}^{\boldsymbol{u}}_h} \|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\tau}_h,\boldsymbol{v}_h)\|_H\,.$$

The following Lemma provides a preliminary estimate for the error $\|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H$.

**Lemma 3.17.** *Let $C_{\mathtt{ST}} := \alpha^{-1}(\Omega) \max\big\{1, \alpha(\Omega) + \|B\|\big\}$, where $\|B\|$ and $\alpha(\Omega)$ are the boundedness and ellipticity constants, respectively, of the bilinear forms $B_\phi$ (cf. (3.21), (3.23)). Then there holds*

$$\begin{aligned} \|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H \leq C_{\mathtt{ST}} \Big\{ & \big(1 + 2\|B\|\big)\, \mathrm{dist}\big((\boldsymbol{\sigma},\boldsymbol{u}), \mathbb{H}^{\boldsymbol{\sigma}}_h \times \mathbf{H}^{\boldsymbol{u}}_h\big) \\ + \quad & (1+\kappa_2^2)^{1/2}\, \|\boldsymbol{f}\|_{\infty,\Omega}\, \|\phi-\phi_h\|_{0,\Omega} + \frac{L_\mu\,(1+\kappa_1^2)^{1/2}}{\mu_1^2}\, C_\delta\, \|\boldsymbol{\sigma}\|_{\delta,\Omega}\, \|\phi-\phi_h\|_{\mathrm{L}^{n/\delta}(\Omega)} \Big\}. \end{aligned} \tag{3.53}$$

*Proof.* By applying Lemma 3.16 to the context (3.51), we obtain

$$\begin{aligned}
&\|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H \\
&\quad\leq\ C_{\mathrm{ST}}\left\{\sup_{\substack{(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\in\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}}\\(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\neq\mathbf{0}}}\frac{\left|\,F_\phi(\boldsymbol{\tau}_h,\boldsymbol{v}_h)-F_{\phi_h}(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\,\right|}{\|(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\|_H}\right. \\
&\quad+\ \inf_{\substack{(\boldsymbol{\zeta}_h,\boldsymbol{w}_h)\in\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}}\\(\boldsymbol{\zeta}_h,\boldsymbol{w}_h)\neq\mathbf{0}}}\left(\|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta}_h,\boldsymbol{w}_h)\|_H\right. \\
&\quad\left.\left.+\ \sup_{\substack{(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\in\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}}\\(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\neq\mathbf{0}}}\frac{\left|\,B_\phi((\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h))-B_{\phi_h}((\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h))\,\right|}{\|(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\|_H}\right)\right\}.
\end{aligned}\tag{3.54}$$

Then, proceeding analogously as in the proof of [4, Lemma 3.9], we easily deduce that

$$\sup_{\substack{(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\in\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}}\\(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\neq\mathbf{0}}}\frac{\left|\,F_\phi(\boldsymbol{\tau}_h,\boldsymbol{v}_h)-F_{\phi_h}(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\,\right|}{\|(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\|_H}\ \leq\ \left(1+\kappa_2^2\right)^{1/2}\|\boldsymbol{f}\|_{\infty,\Omega}\,\|\phi-\phi_h\|_{0,\Omega}\,.\tag{3.55}$$

In turn, in order to estimate the supremum in (3.54), we add and subtract suitable terms to write

$$\begin{aligned}
&B_\phi((\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h))-B_{\phi_h}((\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h))\ =\ B_\phi((\boldsymbol{\zeta}_h,\boldsymbol{w}_h)-(\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau}_h,\boldsymbol{v}_h)) \\
&+\ (B_\phi-B_{\phi_h})((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau}_h,\boldsymbol{v}_h))\ +\ B_{\phi_h}((\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h)),
\end{aligned}$$

whence, applying the boundedness (3.21) to the first and third terms on the right hand side of the foregoing equation, and proceeding analogously as for the derivation of [4, eqs. (3.29), (3.30)] with the second one, we find that

$$\begin{aligned}
&\sup_{\substack{(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\in\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}}\\(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\neq\mathbf{0}}}\frac{\left|\,B_\phi((\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h))-B_{\phi_h}((\boldsymbol{\zeta}_h,\boldsymbol{w}_h),(\boldsymbol{\tau}_h,\boldsymbol{v}_h))\,\right|}{\|(\boldsymbol{\tau}_h,\boldsymbol{v}_h)\|_H} \\
&\qquad\leq\ 2\|B\|\,\|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\zeta}_h,\boldsymbol{w}_h)\|_H\ +\ \frac{L_\mu\,(1+\kappa_1^2)^{1/2}}{\mu_1^2}\,C_\delta\,\|\boldsymbol{\sigma}\|_{\delta,\Omega}\,\|\phi-\phi_h\|_{\mathrm{L}^{n/\delta}(\Omega)}\,.
\end{aligned}\tag{3.56}$$

Finally, by replacing the inequalities (3.55) and (3.56) into (3.54), we get (3.53), which ends the proof. □

Next, we have the following result concerning $\|\phi-\phi_h\|_{1,\Omega}$. To this end, and in order to simplify the subsequent writing, we introduce the following constants, independent of the data $\boldsymbol{k}$, $g$, $\boldsymbol{u}_{\mathrm{D}}$, and $\boldsymbol{f}$,

$$K_1\ :=\ C_{\widetilde{S}}\left\{L_f\ +\ L_\vartheta\,C_\delta\,\widetilde{C}_\delta\,\widehat{C}_{\widetilde{S}}(r)\,\gamma_2\,|\Omega|^{1/2}\right\},\qquad K_2\ :=\ C_{\widetilde{S}}\,L_\vartheta\,C_\delta\,\widetilde{C}_\delta\,\widehat{C}_{\widetilde{S}}(r)\,,$$

$$K_3\ :=\ 1+C_{\widetilde{S}}\,(\vartheta_2+\beta)\,,\qquad K_4\ :=\ 3\,C_{\widetilde{S}}\,c(\Omega)\,C_S\,,\qquad\text{and}\qquad K_5\ :=\ C_{\widetilde{S}}\,c(\Omega)\,r\,,$$

where $C_\delta$ is the boundedness constant of the continuous injection $i_\delta:\mathrm{H}^\delta(\Omega)\longrightarrow\mathrm{L}^{\delta^*}(\Omega)$, with $\delta^*$ given by (3.34), and $\widetilde{C}_\delta$ is the boundedness constant of the compact injection $i:\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\longrightarrow\mathrm{L}^{n/\delta}(\Omega)$.

**Lemma 3.18.** *Assume that the data $\boldsymbol{k}$ and $g$ satisfy*

$$K_1\,|\boldsymbol{k}|\ +\ K_2\,\|g\|_{\delta,\Omega}\ \leq\ \frac{1}{2}\,.\tag{3.57}$$

*Then, there holds*

$$\|\phi-\phi_h\|_{1,\Omega} \leq \widetilde{K}_3(\boldsymbol{u}_{\mathrm{D}},\boldsymbol{f})\,\mathrm{dist}(\phi,\mathrm{H}_h^{\phi}) + \widetilde{K}_5\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}\,, \tag{3.58}$$

*where*

$$\widetilde{K}_3(\boldsymbol{u}_{\mathrm{D}},\boldsymbol{f}) := 2\left\{K_3 + K_4\left(\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + r\,\|\boldsymbol{f}\|_{\infty,\Omega}\right)\right\} \quad \textit{and} \quad \widetilde{K}_5 := 2\,K_5\,. \tag{3.59}$$

*Proof.* We first observe by triangle inequality that

$$\|\phi-\phi_h\|_{1,\Omega} \leq \|\phi-\varphi_h\|_{1,\Omega} + \|\phi_h-\varphi_h\|_{1,\Omega} \qquad \forall\,\varphi_h\in\mathrm{H}_h^{\phi}\,. \tag{3.60}$$

Then employing the ellipticity of the bilinear form $A_{\phi_h,\boldsymbol{u}_h}$ with constant $\widetilde{\alpha}$, using that (*cf.* (3.52)) $A_{\phi_h,\boldsymbol{u}_h}(\phi_h,\phi_h-\varphi_h) = A_{\boldsymbol{u}_h}(\phi_h,\phi_h-\varphi_h) = G_{\phi_h}(\phi_h-\varphi_h)$, and then adding and subtracting the expression (*cf.* (3.52)) $G_{\phi}(\phi_h-\varphi_h) = A_{\boldsymbol{u}}(\phi,\phi_h-\varphi_h) = A_{\phi,\boldsymbol{u}}(\phi,\phi_h-\varphi_h)$, we deduce that

$$\begin{aligned}\widetilde{\alpha}\,\|\phi_h-\varphi_h\|_{1,\Omega}^2 &\leq A_{\phi_h,\boldsymbol{u}_h}(\phi_h-\varphi_h,\phi_h-\varphi_h)\\ &\leq |G_{\phi_h}(\phi_h-\varphi_h) - G_{\phi}(\phi_h-\varphi_h)| + |A_{\phi,\boldsymbol{u}}(\phi,\phi_h-\varphi_h) - A_{\phi_h,\boldsymbol{u}_h}(\varphi_h,\phi_h-\varphi_h)|\,.\end{aligned} \tag{3.61}$$

Next, according to the definition of $G_{\phi}$ (*cf.* (3.18)), and applying Cauchy-Schwarz inequality, we get

$$|G_{\phi_h}(\phi_h-\varphi_h) - G_{\phi}(\phi_h-\varphi_h)| \leq L_f\,|\boldsymbol{k}|\,\|\phi-\phi_h\|_{0,\Omega}\,|\phi_h-\varphi_h|_{1,\Omega}\,. \tag{3.62}$$

In turn, adding and subtracting $\vartheta(\phi_h)$ and $\boldsymbol{u}$ within appropriate expressions of $A_{\phi,\boldsymbol{u}}(\phi,\phi_h-\varphi_h)$ and $A_{\phi_h,\boldsymbol{u}_h}(\varphi_h,\phi_h-\varphi_h)$, respectively, and then applying Hölder's inequality, the upper bound of $\vartheta$ (*cf.* (3.3)), and (3.10), we find that

$$\begin{aligned}&|A_{\phi,\boldsymbol{u}}(\phi,\phi_h-\varphi_h) - A_{\phi_h,\boldsymbol{u}_h}(\varphi_h,\phi_h-\varphi_h)|\\ &\leq L_{\vartheta}\,\|\phi-\phi_h\|_{\mathrm{L}^{2q}(\Omega)}\,\|\nabla\phi\|_{\mathrm{L}^{2p}(\Omega)}\,|\phi_h-\varphi_h|_{1,\Omega} + \vartheta_2\,|\phi-\varphi_h|_{1,\Omega}\,|\phi_h-\varphi_h|_{1,\Omega}\\ &+ c(\Omega)\,\|\varphi_h\|_{1,\Omega}\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}\,|\phi_h-\varphi_h|_{1,\Omega} + c(\Omega)\,\|\phi-\varphi_h\|_{1,\Omega}\,\|\boldsymbol{u}\|_{1,\Omega}\,|\phi_h-\varphi_h|_{1,\Omega}\\ &+ \beta\,\|\phi-\varphi_h\|_{0,\Omega}\,\|\phi_h-\varphi_h\|_{0,\Omega}\,,\end{aligned} \tag{3.63}$$

where $p,q\in[1,+\infty)$ are such that $1/p+1/q=1$. In this way, using the Sobolev embedding Theorem (*cf.* [1, Theorem 4.12], [92, Theorem 1.3.4]), the regularity estimate (3.28), and applying the same arguments used for the derivation of (3.33) (*cf.* proof of Lemma 3.7), in particular the fact that $\mathrm{H}^1(\Omega)$ is compactly, and hence continuously embedded in $\mathrm{L}^{n/\delta}(\Omega)$ with boundedness constant $\widetilde{C}_{\delta}$, it follows from (3.63) that

$$\begin{aligned}&|A_{\phi,\boldsymbol{u}}(\phi,\phi_h-\varphi_h) - A_{\phi_h,\boldsymbol{u}_h}(\varphi_h,\phi_h-\varphi_h)|\\ &\leq L_{\vartheta}\,C_{\delta}\,\widetilde{C}_{\delta}\,\widehat{C}_{\widetilde{S}}(r)\left\{\gamma_2|\Omega|^{1/2}|\boldsymbol{k}| + \|g\|_{\delta,\Omega}\right\}\|\phi-\phi_h\|_{1,\Omega}\,\|\phi_h-\varphi_h\|_{1,\Omega}\\ &+ \vartheta_2\,\|\phi-\varphi_h\|_{1,\Omega}\,\|\phi_h-\varphi_h\|_{1,\Omega} + c(\Omega)\,\|\varphi_h\|_{1,\Omega}\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}\,\|\phi_h-\varphi_h\|_{1,\Omega}\\ &+ c(\Omega)\,\|\phi-\varphi_h\|_{1,\Omega}\,\|\boldsymbol{u}\|_{1,\Omega}\,\|\phi_h-\varphi_h\|_{1,\Omega} + \beta\,\|\phi-\varphi_h\|_{1,\Omega}\,\|\phi_h-\varphi_h\|_{1,\Omega}\,.\end{aligned} \tag{3.64}$$

Thus, by replacing (3.62) and (3.64) into (3.61), and then the resulting estimate into (3.60), employing the constants defined previously to the statement of the present lemma, using that both $\|\boldsymbol{u}\|_{1,\Omega}$ and

$\|\boldsymbol{u}_h\|_{1,\Omega}$ are bounded by $C_S \left\{ \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + r \|\boldsymbol{f}\|_{\infty,\Omega} \right\}$ (*cf.* Lemmas 3.3 and 3.10), and recalling from the proof of Lemma 3.4 that $\widetilde{\alpha} = C_{\widetilde{S}}^{-1}$, we find, after several algebraic manipulations, that

$$\begin{aligned} \|\phi - \phi_h\|_{1,\Omega} \, \leq \, & \left\{ K_1 \, |\boldsymbol{k}| \, + \, K_2 \, \|g\|_{\delta,\Omega} \right\} \|\phi - \phi_h\|_{1,\Omega} \\ & + \left\{ K_3 + K_4 \left( \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + r \, \|\boldsymbol{f}\|_{\infty,\Omega} \right) \right\} \|\phi - \varphi_h\|_{1,\Omega} \, + \, K_5 \, \|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,\Omega} \qquad \forall \, \varphi_h \in \mathrm{H}_h^{\phi} \, , \end{aligned}$$

which, according to the assumption (3.57) and the notation (3.59), and taking the infimum on $\varphi_h \in \mathrm{H}_h^{\phi}$, yields (3.58) and completes the proof. □

We now combine the inequalities provided by Lemmas 3.17 and 3.18 to derive the Céa estimate for the total error $\|(\boldsymbol{\sigma}, \boldsymbol{u}) - (\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\| \, + \, \|\phi - \phi_h\|_{1,\Omega}$. To this end, we now introduce the constants

$$K_6 \, := \, C_{\mathtt{ST}} \, \frac{L_\mu \, (1 + \kappa_1^2)^{1/2}}{\mu_1^2} \, C_S \, \widetilde{C}_\delta \, \widehat{C}_S(r) \, ,$$

and

$$K_7 \, := \, C_{\mathtt{ST}} \, (1 + \kappa_2^2)^{1/2} \, + \, \frac{L_\mu \, (1 + \kappa_1^2)^{1/2}}{\mu_1^2} \, C_S \, \widetilde{C}_\delta \, \widehat{C}_S(r) \, r \, .$$

Then, employing from (3.27) that $\|\boldsymbol{\sigma}\|_{\delta,\Omega} \, \leq \, \widehat{C}_S(r) \left\{ \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} \, + \, \|\phi\|_{0,\Omega} \, \|\boldsymbol{f}\|_{\infty,\Omega} \right\}$, recalling that $\|\phi\|_{1,\Omega} \, \leq \, r$, and using that $\widetilde{C}_\delta$ is the boundedness constant of the continuous injection of $\mathrm{H}^1(\Omega)$ into $\mathrm{L}^{n/\delta}(\Omega)$, we can assert from (3.53) that

$$\begin{aligned} \|(\boldsymbol{\sigma}, \boldsymbol{u}) - (\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\|_H \, \leq \, & C_{\mathtt{ST}} \, \big(1 + 2 \, \|B\|\big) \, \mathrm{dist}\big((\boldsymbol{\sigma}, \boldsymbol{u}), \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}\big) \\ & + \quad \left\{ K_6 \, \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} \, + \, K_7 \, \|\boldsymbol{f}\|_{\infty,\Omega} \right\} \|\phi - \phi_h\|_{1,\Omega} \, , \end{aligned}$$

which, employing the estimate for $\|\phi - \phi_h\|_{1,\Omega}$ given by (3.58), implies

$$\begin{aligned} \|(\boldsymbol{\sigma}, \boldsymbol{u}) - (\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\|_H \, \leq \, & C_{\mathtt{ST}} \, \big(1 + 2 \, \|B\|\big) \, \mathrm{dist}\big((\boldsymbol{\sigma}, \boldsymbol{u}), \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}\big) \, + \, \widetilde{K}_6(\boldsymbol{u}_{\mathrm{D}}, \boldsymbol{f}) \, \mathrm{dist}\big(\phi, \mathrm{H}_h^{\phi}\big) \\ & + \quad \widetilde{K}_5 \left\{ K_6 \, \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} \, + \, K_7 \, \|\boldsymbol{f}\|_{\infty,\Omega} \right\} \|\boldsymbol{u} - \boldsymbol{u}_h\|_{1,\Omega} \, , \end{aligned}$$

where

$$\widetilde{K}_6(\boldsymbol{u}_{\mathrm{D}}, \boldsymbol{f}) \, := \, \widetilde{K}_3(\boldsymbol{u}_{\mathrm{D}}, \boldsymbol{f}) \left\{ K_6 \, \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} \, + \, K_7 \, \|\boldsymbol{f}\|_{\infty,\Omega} \right\} .$$

In this way, assuming now that $\boldsymbol{u}_{\mathrm{D}}$ and $\boldsymbol{f}$ satisfy

$$\widetilde{K}_5 \left\{ K_6 \, \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} \, + \, K_7 \, \|\boldsymbol{f}\|_{\infty,\Omega} \right\} \, \leq \, \frac{1}{2} \, ,$$

we conclude from the foregoing equations that

$$\|(\boldsymbol{\sigma}, \boldsymbol{u}) - (\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\|_H \, \leq \, 2 \, C_{\mathtt{ST}} \, \big(1 + 2 \, \|B\|\big) \, \mathrm{dist}\big((\boldsymbol{\sigma}, \boldsymbol{u}), \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}\big) \, + \, 2 \, \widetilde{K}_6(\boldsymbol{u}_{\mathrm{D}}, \boldsymbol{f}) \, \mathrm{dist}\big(\phi, \mathrm{H}_h^{\phi}\big) \, . \quad (3.65)$$

Consequently, we can establish the following result providing the complete Céa estimate.

**Theorem 3.3.** *Assume that the data* $\boldsymbol{k}$, $g$, $\boldsymbol{u}_{\mathrm{D}}$, *and* $\boldsymbol{f}$ *are sufficiently small so that*

$$K_1\,|\boldsymbol{k}| \,+\, K_2\,\|g\|_{\delta,\Omega} \,\leq\, \frac{1}{2} \qquad \textit{and} \qquad \widetilde{K}_5\left\{K_6\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} \,+\, K_7\,\|\boldsymbol{f}\|_{\infty,\Omega}\right\} \,\leq\, \frac{1}{2}\,.$$

*Then, there exists a positive constant* $C$, *independent of* $h$, *but depending on data, parameters, and other constants, such that*

$$\begin{aligned} &\|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H \,+\, \|\phi-\phi_h\|_{1,\Omega} \\ &\qquad\leq\quad C\left\{\mathrm{dist}\big((\boldsymbol{\sigma},\boldsymbol{u}),\mathbb{H}_h^{\boldsymbol{\sigma}}\times\mathbf{H}_h^{\boldsymbol{u}}\big) \,+\, \mathrm{dist}(\phi,\mathrm{H}_h^{\phi})\right\}. \end{aligned} \tag{3.66}$$

*Proof.* It follows straightforwardly from (3.65) and (3.58). □

We end this section with the corresponding rates of convergence of our Galerkin scheme (3.45).

**Theorem 3.4.** *In addition to the hypotheses of Theorems 3.1, 3.2, and 3.3, assume that there exists* $s>0$ *such that* $\boldsymbol{\sigma}\in\mathbb{H}^s(\Omega)$, $\mathbf{div}\boldsymbol{\sigma}\in\mathbf{H}^s(\Omega)$, $\boldsymbol{u}\in\mathbf{H}^{1+s}(\Omega)$, *and* $\phi\in\mathrm{H}^{1+s}(\Omega)$. *Then, there exists* $\widehat{C}>0$, *independent of* $h$, *such that, with the finite element subspaces defined by (3.42), (3.43), and (3.44), there holds*

$$\begin{aligned} &\|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H \,+\, \|\phi-\phi_h\|_{1,\Omega} \\ &\qquad\leq\quad \widehat{C}\,h^{\min\{s,k+1\}}\left\{\|\boldsymbol{\sigma}\|_{s,\Omega}+\|\mathbf{div}\boldsymbol{\sigma}\|_{s,\Omega}+\|\boldsymbol{u}\|_{1+s,\Omega}+\|\phi\|_{1+s,\Omega}\right\}. \end{aligned}$$

*Proof.* It follows directly from the Céa estimate (3.66) and the approximation properties of $\mathbb{H}_h^{\boldsymbol{\sigma}}$, $\mathbf{H}_h^{\boldsymbol{u}}$, and $\mathrm{H}_h^{\phi}$ (*cf.* [26, 40, 63]). □

## 3.6 Numerical tests

**Example 1.** Our first example aims at testing the accuracy of our augmented finite element formulation. As usual, experimental errors and convergence rates are defined as

$$e(\boldsymbol{\sigma}) := \|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathbf{div},\Omega}\,, \qquad e(\boldsymbol{u}) := \|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}\,,$$

$$e(\phi) := \|\phi-\phi_h\|_{1,\Omega}\,, \qquad r(\cdot) := \log(e(\cdot)/\widehat{e}(\cdot))[\log(h/\widehat{h})]^{-1}\,,$$

where $e$ and $\widehat{e}$ stand for errors computed on two consecutive meshes of sizes $h$ and $\widehat{h}$, respectively. In all examples we consider $\boldsymbol{K}=K\mathbb{I}$, with $K$ constant. The following exact solution to (3.1) defined on the unit disk is manufactured

$$\phi(x_1,x_2)=c-c\exp(1-x_1^2-x_2^2),\ \boldsymbol{u}(x_1,x_2)=\begin{pmatrix}\sin(\pi x_1)\cos(\pi x_2)\\ -\cos(\pi x_1)\sin(\pi x_2)\end{pmatrix},\ \boldsymbol{\sigma}(x_1,x_2)=\mu(\phi)\nabla\boldsymbol{u}-(x_1^2-x_2^2)\mathbb{I}\,,$$

where $K^{-1}=0.01$, $\beta=10$, $\boldsymbol{k}=(0,-1)^{\mathtt{t}}$, $\mu(\phi)=(1-b\phi)^{-2}$, $f_{\mathrm{bk}}(\phi)=b\phi(1-b\phi)^2$, $\vartheta(\phi)=\phi+(1-b\phi)^2$, and the source terms are

$$\boldsymbol{f}(x_1,x_2)=\phi^{-1}(K^{-1}\boldsymbol{u}-\mathbf{div}\boldsymbol{\sigma}),\quad g(x_1,x_2)=\beta\phi-\mathrm{div}(\vartheta(\phi)\nabla\phi)+\boldsymbol{u}\cdot\nabla\phi+f'_{\mathrm{bk}}(\phi)\boldsymbol{k}\cdot\nabla\phi,$$

| $N_h$ | $h$ | $e(\boldsymbol{\sigma})$ | $r(\boldsymbol{\sigma})$ | $e(\boldsymbol{u})$ | $r(\boldsymbol{u})$ | $e(\phi)$ | $r(\phi)$ | iter |
|---|---|---|---|---|---|---|---|---|
| Augmented $\mathbf{RT}_0 - \mathbf{P}_1 - \mathrm{P}_1$ scheme | | | | | | | | |
| 45 | 1.000000 | 37.83630 | – | 5.078982 | – | 0.794267 | – | 7 |
| 150 | 0.752986 | 29.63322 | 0.861350 | 3.864984 | 0.962793 | 0.551649 | 1.284790 | 6 |
| 567 | 0.381608 | 14.53602 | 1.047988 | 1.893370 | 1.049950 | 0.261571 | 1.097927 | 6 |
| 1986 | 0.202981 | 7.685891 | 1.009446 | 0.917330 | 1.147899 | 0.142439 | 0.962789 | 5 |
| 7587 | 0.107277 | 3.855674 | 1.081757 | 0.449265 | 1.119414 | 0.071858 | 1.072957 | 5 |
| 29652 | 0.056293 | 1.929090 | 1.073920 | 0.222210 | 1.091739 | 0.036226 | 1.062147 | 5 |
| 116820 | 0.029796 | 0.967180 | 1.085224 | 0.111677 | 1.081444 | 0.018210 | 1.081158 | 5 |
| 465243 | 0.015539 | 0.480698 | 1.073919 | 0.056004 | 1.060163 | 0.009068 | 1.070848 | 5 |
| 1840545 | 0.008139 | 0.243228 | 1.053393 | 0.028080 | 1.067485 | 0.004619 | 1.043095 | 5 |
| Augmented $\mathbf{RT}_1 - \mathbf{P}_2 - \mathrm{P}_2$ scheme | | | | | | | | |
| 129 | 1.000000 | 32.06255 | – | 3.909169 | – | 0.549477 | – | 7 |
| 447 | 0.752986 | 16.36007 | 1.505829 | 1.686632 | 1.605829 | 0.163634 | 1.919349 | 6 |
| 2043 | 0.381608 | 6.518447 | 1.697863 | 0.374835 | 1.820361 | 0.040489 | 1.899365 | 5 |
| 6835 | 0.202981 | 2.511864 | 1.781622 | 0.097825 | 1.967394 | 0.011124 | 1.841454 | 5 |
| 26243 | 0.107277 | 0.695590 | 1.902700 | 0.026791 | 2.146465 | 0.002491 | 1.857032 | 5 |
| 104867 | 0.056293 | 0.209899 | 1.925679 | 0.006716 | 1.743774 | 0.000774 | 1.864551 | 5 |
| 412611 | 0.029796 | 0.056163 | 1.945132 | 0.001700 | 2.095484 | 0.000182 | 1.928363 | 5 |
| 1643907 | 0.015539 | 0.014448 | 1.960714 | 0.000427 | 2.132144 | 0.000039 | 1.987836 | 5 |

Table 3.1: Example 1: Convergence history and Newton iteration count for the mixed–primal $\mathbf{RT}_k - \mathbf{P}_{k+1} - \mathrm{P}_{k+1}$ approximations of the coupled problem, $k = 0, 1$. Here $N_h$ stands for the number of degrees of freedom associated to each triangulation $\mathcal{T}_h$.

for $(x_1, x_2) \in \overline{\Omega}$. We take $b = 1/2, c = 1/(1-e)$ and set $\Gamma_{\mathrm{D}} = \partial\Omega$, where $\phi$ vanishes and the velocity is imposed accordingly to the exact solution. The mean value of $\mathrm{tr}(\boldsymbol{\sigma}_h)$ over $\Omega$ is fixed via a penalization strategy. As defined above, the concentration is bounded in $\Omega$ and so are the concentration-dependent coefficients as well. In particular we have $\mu_1 = 1, \mu_2 = 4$ and as suggested by Lemma 3.3, the stabilization constants are chosen as $\kappa_1 = \frac{\delta\mu_1}{\mu_2}$ with $\delta = \mu_1$, and $\kappa_2 = 0.025$ for $\widetilde{\delta} = \frac{1}{4|K^{-1}|}$.

A Newton-Raphson algorithm with a fixed tolerance of 1e-6 has been used for the nonlinear problem (3.45). At each iteration the linear systems resulting from the linearization were solved by means of the multifrontal solver MUMPS. Independently of the refinement level, we observe that an average number of 5 steps was required to reach the desired tolerance. Values and plots of errors and corresponding rates associated to $\mathbf{RT}_k - \mathbf{P}_{k+1} - \mathrm{P}_{k+1}$ approximations with $k = 0$ and $k = 1$ are summarized in Table 3.1 and Figure 3.1. The results show optimal asymptotic convergence rates for all fields (of order $k + 1$ for the pseudo-stress, the velocity and the concentration), which agree with the accuracy predicted in Theorem 3.4. We also remark that for both degrees of approximation, the concentration errors are always below the velocity errors, and both are dominated by the errors in the pseudo-stress approximation. The augmented mixed–primal approximations computed on a mesh of 204847 vertices and 409692 elements are depicted in Figure 3.2, where stress, velocity, and concentration profiles are well resolved.

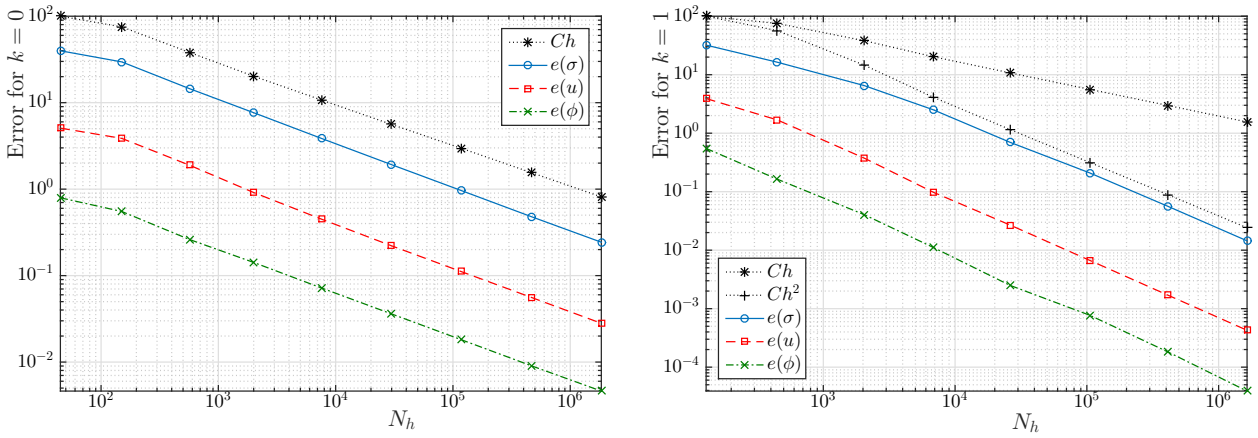


Figure 3.1: Example 1: Computed errors associated to the mixed–primal approximation versus the number of degrees of freedom $N_h$ for $\mathbf{RT}_0 - \mathbf{P}_1 - \mathrm{P}_1$ and $\mathbf{RT}_1 - \mathbf{P}_2 - \mathrm{P}_2$ finite elements (left and right, respectively). Values are detailed in Table 3.1.

For all remaining examples we stick to the case $k = 0$, i.e., row-wise lowest-order Raviart-Thomas finite element approximations for the Cauchy pseudo-stress, and piecewise linear approximations of velocity and concentration.

**Example 2.** Our next example corresponds to a test of batch sedimentation in a cylinder with a contraction (see e.g. [91, 93]). In this case the model parameters and concentration-dependent coefficients assume the values $K = 60$, $\beta = K^{-1}$, $\mu(\phi) = \mu_0\Big(1 - \frac{\phi}{\phi_{\max}}\Big)^{-\eta}$, $f_{\rm bk}(\phi) = C_1 C_2 \frac{(1-\phi_0)}{\mu_\phi}\phi$, $\vartheta(\phi) = C_1\Big(C_3\gamma\phi^2 + C_2\Big)$ and $\boldsymbol{f} = \Delta\rho G\boldsymbol{k}$, with $C_1 = (\rho_1 + \phi_{\max}\Delta\rho)(\rho_1\rho_2)^{-1}$, $C_2 = \frac{2}{9\mu_0}\Delta\rho G a^2$, $C_3 = 0.68355a^2$, $\Delta\rho = \rho_2 - \rho_1$, and $\gamma = 0.88$. Values for the remaining dimensional constants are collected in Table 3.2, and the model parameters yield the following stabilization constants $\mu_1 = \mu_0$, $\mu_2 = 6.5365$, $\kappa_1 = \delta\mu_1/\mu_2$ with $\delta = \mu_1$ and $\kappa_2 = 0.025$ for $\widetilde{\delta} = \frac{1}{4|K^{-1}|}$.

| Quantity | Value |
|---|---|
| Density glycerol/water solution, $\rho_1$ | $1.175\, g/cm^3$ |
| Density PMMA, $\rho_2$ | $1.18\, g/cm^3$ |
| Viscosity glycerol/water solution, $\mu_0$ | $0.184\, g/cm^3$ |
| Initial volume fraction, $\phi_0$ | 0.192 |
| Maximum volume fraction, $\phi_{\max}$ | 0.64 |
| Particle radius, $a$ | $0.0397\, cm$ |
| Viscosity constant, $\eta$ | 1.82 |
| Gravity, $G$ | $980.665\, cm/s^2$ |

Table 3.2: Example 2: Model constants employed in the simulation of steady sedimentation of PMMA into glycerol/water within a contracted cylinder.

On $\partial\Omega$ we impose zero-flux conditions for $\phi$, that is $\widetilde{\boldsymbol{\sigma}}\cdot\boldsymbol{\nu} = 0$. In addition, the following boundary conditions are imposed for the velocity (see sketch in Figure 3.3): $\boldsymbol{u}|_{\Gamma_1} = \boldsymbol{u}|_{\Gamma_3} = \mathbf{0}$, and $u_2|_{\Gamma_2} = 0$

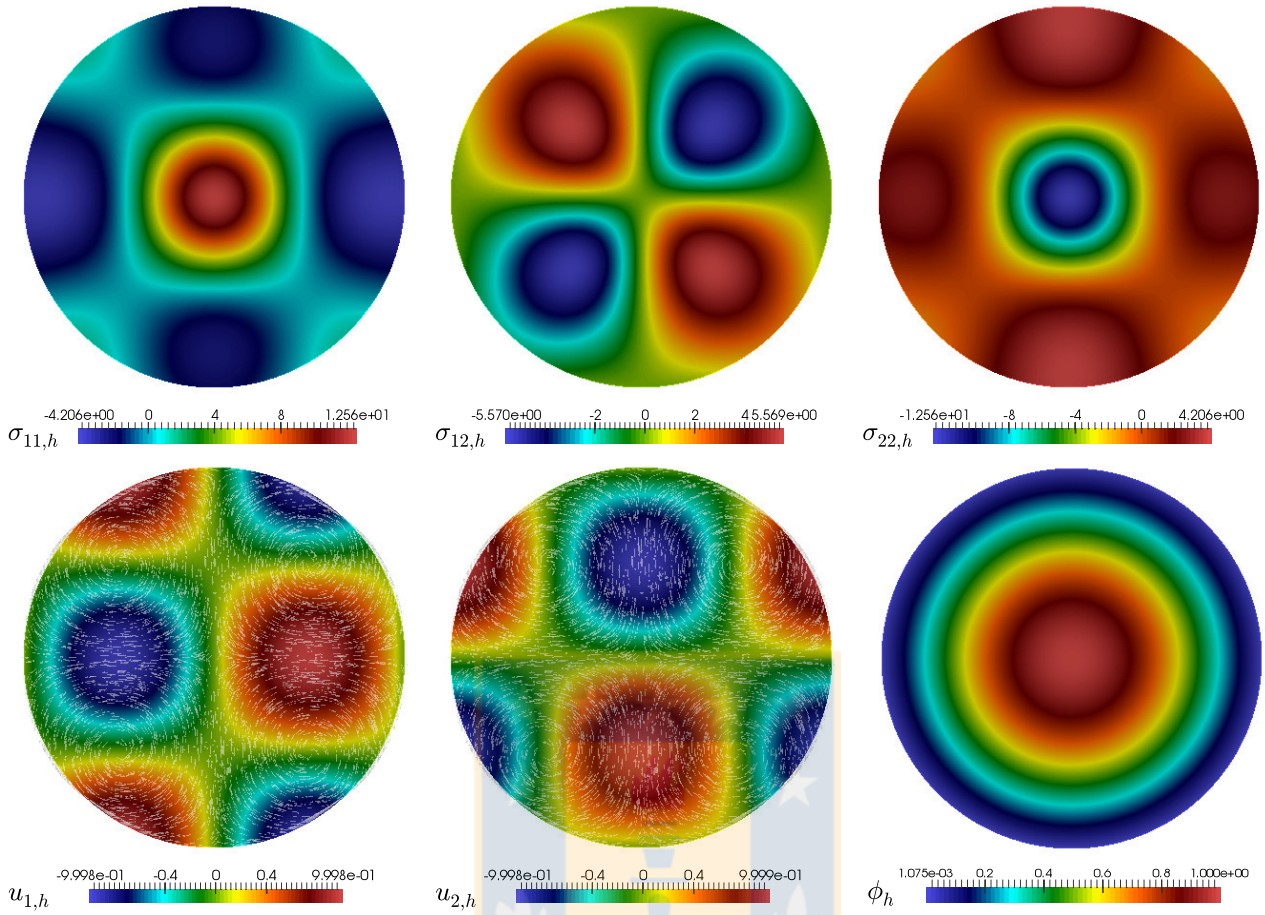


Figure 3.2: Example 1: $\mathbf{RT}_0 - \mathbf{P}_1 - \mathrm{P}_1$ approximation of pseudo-stress components (top panels), velocity components with vector directions (bottom left and center, respectively), and concentration profile (bottom right), solutions to (3.15). The finest mesh has 204847 vertices and 409692 triangular elements.

(representing a symmetry axis). The domain is discretized into 13131 vertices and 26260 triangles, and we represent the obtained field quantities of interest in Figure 3.4. The maximum concentration has been packed at the bottom of the vessel, whereas throughout the rest of the domain is filled with low-concentration material. More interesting phenomena are observed from the velocity plots, where a main recirculation zone is observed at the center of the domain. Moreover, a countercurrent flow is observed along the symmetry axis (clearly identified in the horizontal velocity plot), and these flow patterns are further highlighted in the diagonal components of the pseudo-stress.

**Example 3.** Finally we turn to the simulation of the steady state of flow patterns on a box (see the domain, dimensions and boundary configuration illustrated in Figure 3.5), using a modification to the single phase model described in [100] to reproduce the so-called Coandă effect, which corresponds to the tendency of a fluid jet to be attracted to a nearby surface [101]. In this case the nonlinear concentration-dependent coefficients are $\mu(\phi) = \mu_0(1 - \phi/\phi_{\max})^\eta$, $f_{\mathrm{bk}}(\phi) = u_\infty\phi(1 - \phi/\phi_{\max})^\eta$ and $\vartheta(\phi) = \vartheta_0(\phi^3 + 1)$ where $\eta = 1.82$, $\mu_0 = 0.02$, $\vartheta_0 = 0.0001$, $\beta = 0.01$, $K = 1000$, $g = 0$ and $\boldsymbol{f} = \Delta\rho G\boldsymbol{k}$, with $\Delta\rho = 0.0045$ and $G = 0.98$.

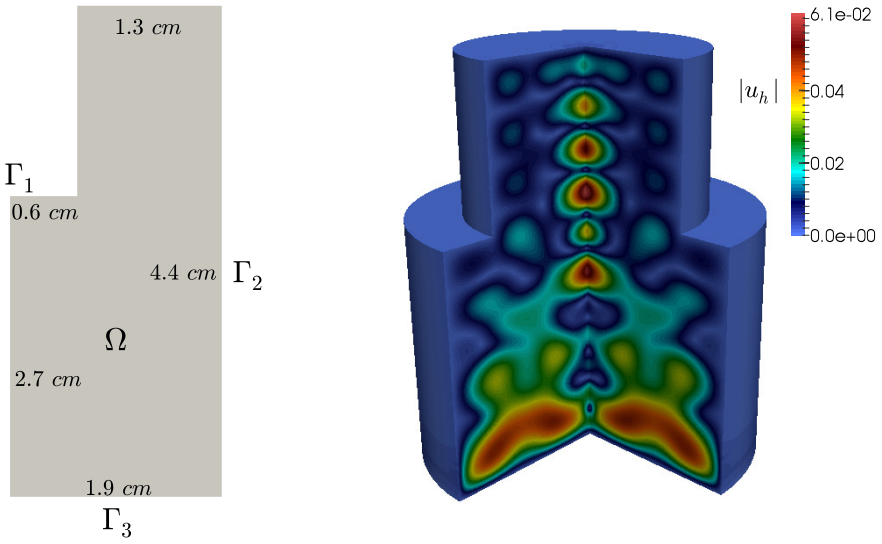


Figure 3.3: Example 2: Typical dimensions and boundary setting for a two-dimensional computational domain representing the batch sedimentation within a cylinder with a contraction (left), and magnitude of the velocity field shown on the rotationally extruded domain (right).

Concentration and velocities are fixed at the inlet surface $\Gamma_{\text{in}}$ (a rectangle of width $0.5\,cm$ and height $0.35\,cm$ located on the top, at $x_1 = 0$) according to $\phi = \phi_{\text{in}}$ and $\boldsymbol{u} = \boldsymbol{u}_{\text{in}} = (u_{1,\text{in}}, 0, 0)^{\mathtt{t}}$. At the outlet $\Gamma_{\text{out}}$ (a rectangle with the same dimensions as the inlet, but located at $x_1 = 6$, on the bottom) we let the material exit the domain with a velocity $\boldsymbol{u} = \boldsymbol{u}_{\text{out}} = (u_{1,\text{out}}, 0, 0)^{\mathtt{t}}$, but the concentration is not prescribed. On the remainder of $\partial\Omega$ we put no slip boundary data for the velocity and zero-flux conditions for the concentration. Other model parameters are set as $u_{1,\text{in}} = u_{1,\text{out}} = 0.01$, $\phi_{\max} = 0.9$, $u_\infty = 0.0022$, and $\phi_{\text{in}} = 0.3$.

According to the bounds of the viscosity, the stabilization parameters were set as $\mu_1 = \mu_0, \mu_2 = \mu_0(1 - \phi_{\text{in}}/\phi_{\max})^{-\eta}$, $\kappa_1 = \delta\frac{\mu_1}{\mu_2}$ with $\delta = \mu_1$ and $\kappa_2 = 0,055$ for $\widetilde{\delta} = \frac{1}{6|K^{-1}|}$. For this problem, 7 Newton iterations were needed to achieve a tolerance of 1e-6 for the energy norm of the incremental approximations. The numerical results are depicted in Figure 3.6 including concentration profiles, velocity vectors and streamlines, and trace of the Cauchy pseudo-stress tensor. As in [100], from the center plot of Figure 3.6 we see a clear attachment of the fluid stream to the side walls, whereas the material with high concentration at the inlet dissolves almost completely at the outlet. This effect corresponds to a relatively high inlet velocity, and since higher concentration material is injected, it penetrates the clear fluid pushing it towards the outlet.

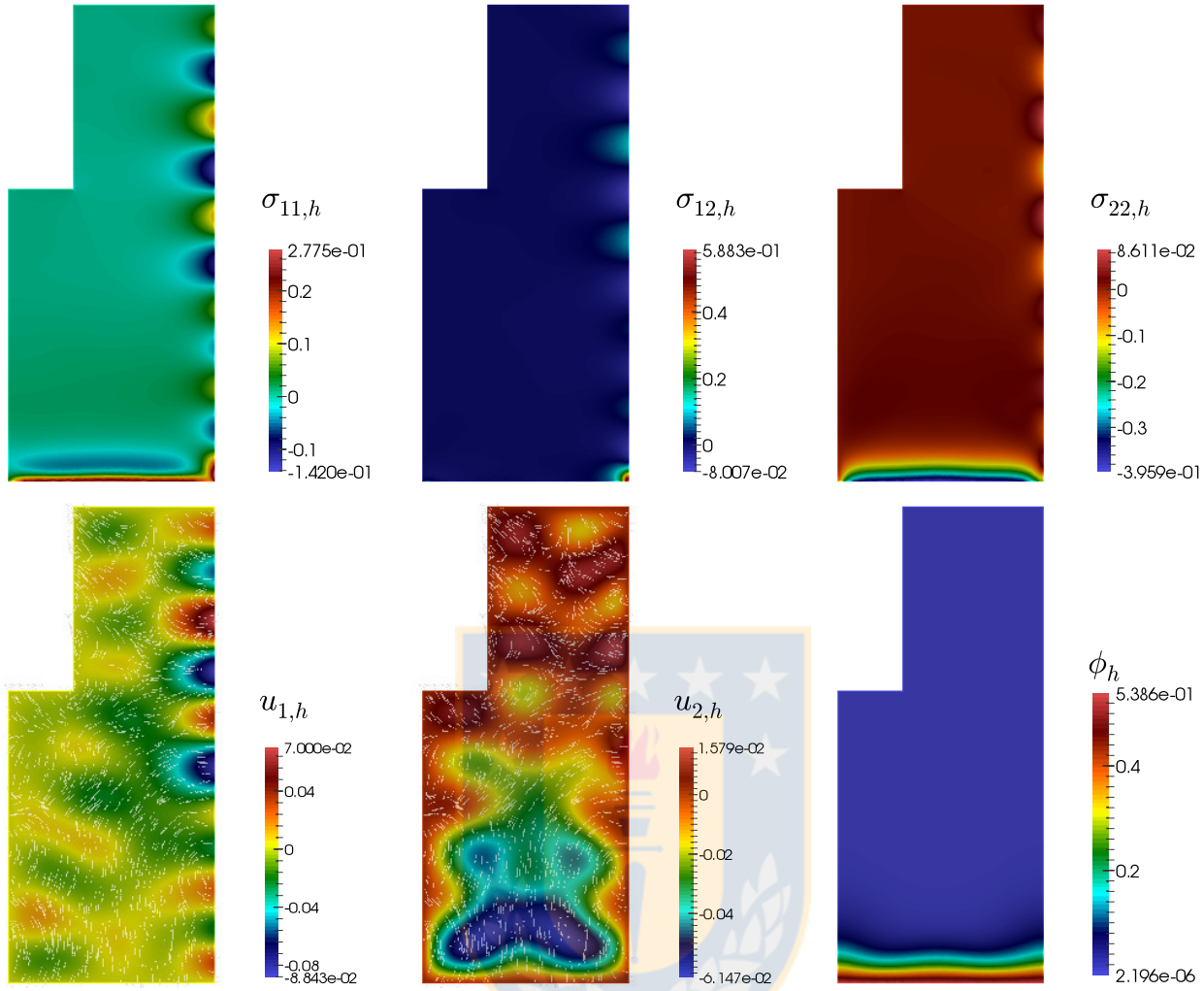


Figure 3.4: Example 2: principal principal components of the Cauchy pseudo-stress (top rows), velocity components $\boldsymbol{u}_h$ with vector directions (bottom left and center), and computed concentration $\phi_h$ (bottom right) for the test of batch sedimentation in a cylinder with a contraction.

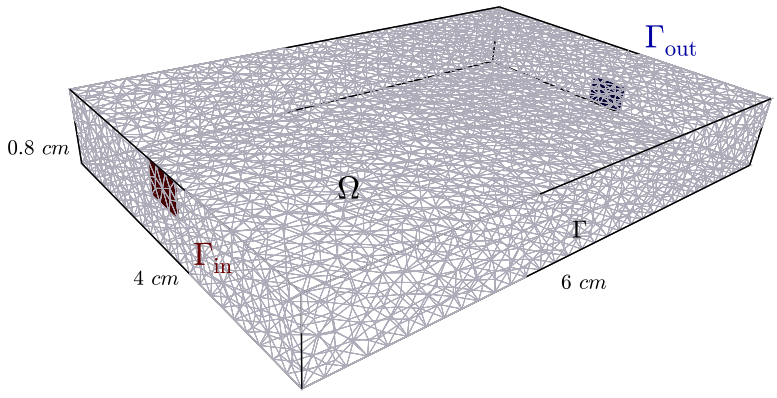


Figure 3.5: Example 3: sketch of the computational domain $\Omega = [0, 6] \times [0, 4] \times [0, 0.8]$, a coarse mesh, and boundary setting, with $\partial\Omega = \Gamma \cup \Gamma_{\text{in}} \cup \Gamma_{\text{out}}$.

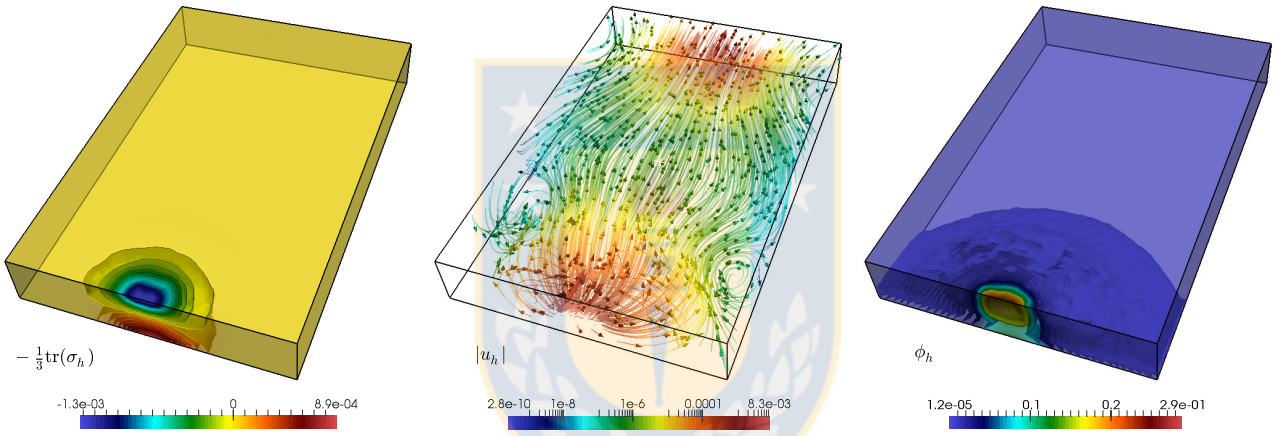


Figure 3.6: Example 3: approximate solutions to the so-called Coanda effect using an augmented mixed formulation. Trace of the Cauchy pseudo-stress tensor (left), velocity vectors and streamlines (middle), and concentration profile (right).

# CHAPTER 4

## A posteriori error analysis for a sedimentation-consolidation system

### 4.1 Introduction

We have recently analyzed in **Chapter 3**, the solvability of a strongly coupled flow and transport system typically encountered in continuum-based models of sedimentation-consolidation processes. More precisely, the steady-state regime of a solid-liquid suspension immersed in a viscous fluid within a permeable medium is considered in [5], and the governing equations consist in the Brinkman problem with variable viscosity coupled with a nonlinear advection – nonlinear diffusion equation describing the transport model. An augmented variational formulation is then proposed there with main unknowns given by the Cauchy pseudo-stress and bulk velocity of the mixture, and the solids volume fraction, which are sought in $\mathbb{H}(\mathbf{div};\Omega)$, $\mathbf{H}^1(\Omega)$, and $\mathrm{H}^1(\Omega)$, respectively. Fixed point arguments, certain regularity assumptions, and some classical results concerning variational problems and Sobolev spaces are combined to establish the solvability of the continuous and discrete coupled formulations. Consequently, the rows of the Cauchy stress tensor were approximated with Raviart-Thomas elements of order $k$, whereas the velocity and solids concentration were discretized with continuous piecewise polynomials of degree $\leq k+1$. Suitable Strang-type estimates are employed to derive optimal *a priori* error estimates for the Galerkin solution.

The purpose of this work is to provide reliable and efficient residual-based *a posteriori* error estimators for the steady sedimentation-consolidation system studied in [5]. Estimators of this kind are frequently employed to guide adaptive mesh refinement in order to guarantee an adequate convergence behavior of the Galerkin approximations, even under the eventual presence of singularities. The global estimator $\boldsymbol{\eta}$ depends on local estimators $\eta_T$ defined on each element $T$ of a given mesh $\mathcal{T}_h$. Then, $\boldsymbol{\eta}$ is said to be efficient (resp. reliable) if there exists a constant $C_{\mathtt{eff}} > 0$ (resp. $C_{\mathtt{rel}} > 0$), independent of meshsizes, such that

$$C_{\mathtt{eff}}\,\boldsymbol{\eta} + \text{h.o.t.} \leq \|error\| \leq C_{\mathtt{rel}}\,\boldsymbol{\eta} + \text{h.o.t.}\,,$$

where h.o.t. is a generic expression denoting one or several terms of higher order. Up to the authors knowledge, a number of *a posteriori* error estimators specifically targeted for non-viscous flow (e.g, Darcy) with transport problems are available in the literature [17, 48, 82, 99, 104]. However, only [24, 83] and [7] are devoted to the *a posteriori* error analysis for coupled viscous flow-transport problems. In particular, we derive in [7], two efficient and reliable residual-based *a posteriori* error estimators for an augmented mixed–primal finite element approximation of a stationary viscous flow and transport problem, which serves as a prototype model for sedimentation-consolidation processes

and other phenomena where the transport of species concentration within a viscous fluid is of interest.

In this chapter, as well as in [7], we make use of ellipticity and inf-sup conditions together with a Helmholtz decomposition, local approximation properties of the Clément interpolant and Raviart-Thomas operator, and known estimates from [15, 60, 66, 72] and [73], to prove the reliability of a residual-based estimator. Then, inverse inequalities, the localization technique based on triangle-bubble and edge-bubble functions imply the efficiency of the estimator. Alternatively, we deduce a second reliable and efficient residual-based *a posteriori* error estimator, where the Helmholtz decomposition is not employed in the corresponding proof of reliability. The rest of this chapter is organized as follows. In Section 4.2, we first recall from [5] the model problem and a corresponding augmented mixed-primal formulation as well as the associated Galerkin scheme. In Section 4.3, we derive a reliable and efficient residual-based *a posteriori* error estimator for our Galerkin scheme. A second estimator is introduced and analyzed in Section 4.4. Next, the analysis and results from Section 4.3 and 4.4 are extend to the three-dimensional case in Section 4.5. Finally, in Section 4.6, our theoretical results are illustrated via some numerical examples, highlighting also the good performance of the scheme and properties of the proposed error indicators.

## 4.2 The sedimentation-consolidation system

### 4.2.1 The governing equations

The following model describes the steady state of the sedimentation-consolidation process consisting on the transport and suspension of a solid phase into an immiscible fluid contained in a vessel $\Omega$ (*cf.* [5]). The flow patterns are influenced by gravity and by the local fluctuations of the solids volume fraction. After elimination of the fluid pressure (*cf.* [5]), the process is governed by the following system of partial differential equations:

$$\begin{aligned} &\frac{1}{\mu(\phi)}\boldsymbol{\sigma}^{\mathtt{d}} = \nabla \boldsymbol{u}\,, \quad \boldsymbol{K}^{-1}\boldsymbol{u} - \mathbf{div}\,\boldsymbol{\sigma} = \boldsymbol{f}\phi\,, \quad \operatorname{div}\boldsymbol{u} = 0 \qquad \text{in} \quad \Omega\,, \\ &\widetilde{\boldsymbol{\sigma}} = \vartheta(\phi)\,\nabla\phi - \phi\boldsymbol{u} - f_{\mathrm{bk}}(\phi)\boldsymbol{k}\,, \quad \beta\,\phi - \operatorname{div}\widetilde{\boldsymbol{\sigma}} = g \qquad \text{in} \quad \Omega\,, \end{aligned} \tag{4.1}$$

along with the following boundary conditions:

$$\begin{aligned} &\boldsymbol{u} = \boldsymbol{u}_{\mathrm{D}} \quad \text{on} \quad \Gamma_{\mathrm{D}}\,, \quad \boldsymbol{\sigma\nu} = \mathbf{0} \quad \text{on} \quad \Gamma_{\mathrm{N}}\,, \\ &\phi = \phi_{\mathrm{D}} \quad \text{on} \quad \Gamma_{\mathrm{D}}\,, \quad \text{and} \quad \widetilde{\boldsymbol{\sigma}}\cdot\boldsymbol{\nu} = 0 \quad \text{on} \quad \Gamma_{\mathrm{N}}\,, \end{aligned} \tag{4.2}$$

where $(\cdot)^{\mathtt{d}}$ denotes the deviatoric operator. The sought quantities are the Cauchy fluid pseudo-stress $\boldsymbol{\sigma}$, the average velocity of the mixture $\boldsymbol{u}$, and the volumetric fraction of the solids (in short, concentration) $\phi$. In this context, the parameter $\beta$ is a positive constant representing the porosity of the medium, and the permeability tensor $\boldsymbol{K} \in \mathbb{C}(\bar{\Omega}) := [C(\bar{\Omega})]^{n\times n}$ and its inverse are symmetric and uniformly positive definite, which means that there exists $\alpha_{\boldsymbol{K}} > 0$ such that

$$\boldsymbol{v}^{\mathtt{t}}\boldsymbol{K}^{-1}(\boldsymbol{x})\boldsymbol{v} \geq \alpha_{\boldsymbol{K}}|\boldsymbol{v}|^2 \quad \forall\, \boldsymbol{v} \in \mathrm{R}^n,\ \forall\, \boldsymbol{x} \in \Omega.$$

Here, we assume that the kinematic effective viscosity, $\mu$; the one-directional Kynch batch flux density function describing hindered settling, $f_{\mathrm{bk}}$; and the diffusion or sediment compressibility, $\vartheta$; are nonlinear scalar functions of the concentration $\phi$. In turn, $\boldsymbol{k}$ is a vector pointing in the direction of gravity and

$\boldsymbol{f} \in \mathbf{L}^\infty(\Omega)$, $\boldsymbol{u}_{\rm D} \in \mathbf{H}^{1/2}(\Gamma_{\rm D})$, $g \in \mathrm{L}^2(\Omega)$ are given functions. For sake of the subsequent analysis, the Dirichlet datum for the concentration will be assumed homogeneous $\phi_{\rm D} = 0$; $\vartheta$ is assumed of class $C^1$; and we suppose that there exist positive constants $\mu_1$, $\mu_2$, $\gamma_1$, $\gamma_2$, $\vartheta_1$, $\vartheta_2$, $L_\mu$, $L_\vartheta$, and $L_f$, such that for each $s$, $t \in \mathrm{R}$ there holds

$$\mu_1 \leq \mu(s) \leq \mu_2\,, \quad \gamma_1 \leq \mathit{f}_{\rm bk}(s) \leq \gamma_2, \quad \vartheta_1 \leq \vartheta(s) \leq \vartheta_2, \tag{4.3}$$

$$|\mu(s) - \mu(t)| \,\leq\, L_\mu\,|s-t|\,, \quad |\vartheta(s) - \vartheta(t)| \,\leq\, L_\vartheta\,|s-t|\,, \quad \text{and} \quad |\mathit{f}_{\rm bk}(s) - \mathit{f}_{\rm bk}(t)| \,\leq\, L_f\,|s-t|\,. \tag{4.4}$$

### 4.2.2 The augmented mixed–primal formulation

The homogeneous Neumann and Dirichlet boundary conditions for $\boldsymbol{\sigma}$ on $\Gamma_{\rm N}$ and $\phi$ on $\Gamma_{\rm D}$ (second and third relations of (4.2), respectively) suggest the introduction of the following functional spaces

$$\mathbb{H}_{\rm N}(\mathbf{div};\Omega) := \Big\{ \boldsymbol{\tau} \in \mathbb{H}(\mathbf{div};\Omega) : \quad \boldsymbol{\tau}\boldsymbol{\nu} = \mathbf{0} \quad \text{on} \quad \Gamma_{\rm N} \Big\},$$
$$\mathrm{H}^1_{\Gamma_{\rm D}}(\Omega) := \Big\{ \psi \in \mathrm{H}^1(\Omega) : \quad \psi = 0 \quad \text{on} \quad \Gamma_{\rm D} \Big\}.$$

Consequently, an augmented mixed-primal formulation for our original coupled problem (4.1) reads as follows: Find $(\boldsymbol{\sigma}, \boldsymbol{u}, \phi) \in \mathbb{H}_{\rm N}(\mathbf{div};\Omega) \times \mathbf{H}^1(\Omega) \times \mathrm{H}^1_{\Gamma_{\rm D}}(\Omega)$ such that

$$\begin{array}{rcll} B_\phi((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau},\boldsymbol{v})) & = & F_\phi(\boldsymbol{\tau},\boldsymbol{v}) & \forall (\boldsymbol{\tau},\boldsymbol{v}) \in \mathbb{H}_{\rm N}(\mathbf{div};\Omega) \times \mathbf{H}^1(\Omega)\,, \\ A_{\boldsymbol{u}}(\phi,\psi) & = & G_\phi(\psi) & \forall\, \psi \in \mathrm{H}^1_{\Gamma_{\rm D}}(\Omega)\,, \end{array} \tag{4.5}$$

where

$$\begin{aligned} B_\phi((\boldsymbol{\sigma},\boldsymbol{u}),(\boldsymbol{\tau},\boldsymbol{v})) \,:=\, & \int_\Omega \frac{1}{\mu(\phi)}\,\boldsymbol{\sigma}^{\rm d} : \boldsymbol{\tau}^{\rm d} \,+\, \int_\Omega \boldsymbol{u}\cdot\mathbf{div}\boldsymbol{\tau} \,-\, \int_\Omega \boldsymbol{v}\cdot\mathbf{div}\boldsymbol{\sigma} \,+\, \int_\Omega \boldsymbol{K}^{-1}\boldsymbol{u}\cdot\boldsymbol{v} \\ & +\, \kappa_1 \int_\Omega \Big( \nabla\boldsymbol{u} - \frac{1}{\mu(\phi)}\,\boldsymbol{\sigma}^{\rm d} \Big) : \nabla\boldsymbol{v} \,-\, \kappa_2 \int_\Omega \boldsymbol{K}^{-1}\boldsymbol{u}\cdot\mathbf{div}\boldsymbol{\tau} \,+\, \kappa_2 \int_\Omega \mathbf{div}\boldsymbol{\sigma}\cdot\mathbf{div}\boldsymbol{\tau}\,, \end{aligned} \tag{4.6}$$

$$\begin{aligned} F_\phi(\boldsymbol{\tau},\boldsymbol{v}) \,:=\, & \langle \boldsymbol{\tau}\boldsymbol{\nu}, \boldsymbol{u}_{\rm D}\rangle_{\Gamma_{\rm D}} \,+\, \int_\Omega \boldsymbol{f}\phi\cdot\boldsymbol{v} \,-\, \kappa_2 \int_\Omega \boldsymbol{f}\phi\cdot\mathbf{div}\boldsymbol{\tau}\,, \\ A_{\boldsymbol{u}}(\phi,\psi) \,:=\, & \int_\Omega \vartheta(\phi)\,\nabla\phi\cdot\nabla\psi \,-\, \int_\Omega \phi\,\boldsymbol{u}\cdot\nabla\psi \,+\, \int_\Omega \beta\,\phi\psi \qquad \forall\,\phi,\,\psi \in \mathrm{H}^1_{\Gamma_{\rm D}}(\Omega)\,, \\ G_\phi(\psi) \,:=\, & \int_\Omega \mathit{f}_{\rm bk}(\phi)\,\boldsymbol{k}\cdot\nabla\psi \,+\, \int_\Omega g\psi \qquad \forall\,\psi \in \mathrm{H}^1_{\Gamma_{\rm D}}(\Omega)\,, \end{aligned} \tag{4.7}$$

and $\kappa_1, \kappa_2$ are positive parameters satisfying $\kappa_1 \in \Big(0, \frac{2\delta\mu_1}{\mu_2}\Big)$ and $\kappa_2 \in \Big(0, \frac{2\widetilde{\delta}\alpha_{\boldsymbol{K}}}{n\|\boldsymbol{K}^{-1}\|_\infty}\Big)$, with $\delta \in (0, 2\mu_1)$ and $\widetilde{\delta} \in \Big(0, \frac{2}{n\|\boldsymbol{K}^{-1}\|_\infty}\Big)$. Further details yielding the weak formulation (4.5), along with its fixed-point based solvability analysis can be found in [5, Section 3].

### 4.2.3 The augmented mixed–primal finite element method

We denote by $\mathcal{T}_h$ a regular partition of $\Omega$ into triangles $T$ (resp. tetrahedra $T$ in $\mathrm{R}^3$) of diameter $h_T$, and meshsize $h := \max\Big\{ h_T : \; T \in \mathcal{T}_h \Big\}$. In addition, given an integer $k \geq 0$, $\mathrm{P}_k(T)$ denotes the space of polynomial functions on $T$ of degree $\leq k$, and we define the corresponding local Raviart-Thomas space of order $k$ as $\mathbf{RT}_k(T) := \mathbf{P}_k(T) \oplus \mathrm{P}_k(T)\,\boldsymbol{x}$, where, according to the notations described in Section

4.1, $\mathbf{P}_k(T) = [\mathrm{P}_k(T)]^n$, and $\boldsymbol{x} \in \mathrm{R}^n$. Then, the Galerkin scheme associated to (4.5), corresponds to: Find $(\boldsymbol{\sigma}_h, \boldsymbol{u}_h, \phi_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}} \times \mathrm{H}_h^{\phi}$ such that

$$\begin{aligned} B_{\phi_h}((\boldsymbol{\sigma}_h, \boldsymbol{u}_h), (\boldsymbol{\tau}_h, \boldsymbol{v}_h)) &= F_{\phi_h}(\boldsymbol{\tau}_h, \boldsymbol{v}_h) \quad \forall (\boldsymbol{\tau}_h, \boldsymbol{v}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}}\,, \\ A_{\boldsymbol{u}_h}(\phi_h, \psi_h) &= \int_\Omega f_{\mathrm{bk}}(\phi_h)\, \boldsymbol{k} \cdot \nabla \psi_h + \int_\Omega g \psi_h \qquad \forall\, \psi_h \in \mathrm{H}_h^{\phi}\,, \end{aligned} \tag{4.8}$$

where the involved finite element spaces are defined as

$$\begin{aligned} \mathbb{H}_h^{\boldsymbol{\sigma}} &:= \left\{ \boldsymbol{\tau}_h \in \mathbb{H}_{\mathrm{N}}(\mathbf{div}; \Omega) : \quad \mathbf{c}^{\mathtt{t}}\, \boldsymbol{\tau}_h|_T \in \mathrm{RT}_k(T) \quad \forall\, \mathbf{c} \in \mathrm{R}^n\,, \quad \forall\, T \in \mathcal{T}_h \right\}, \\ \mathbf{H}_h^{\boldsymbol{u}} &:= \left\{ \boldsymbol{v}_h \in \mathbf{C}(\overline{\Omega}) : \quad \boldsymbol{v}_h|_T \in \mathbf{P}_{k+1}(T) \quad \forall\, T \in \mathcal{T}_h \right\}, \\ \mathrm{H}_h^{\phi} &:= \left\{ \psi_h \in C(\overline{\Omega}) \cap \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) : \quad \psi_h|_T \in \mathrm{P}_{k+1}(T) \quad \forall\, T \in \mathcal{T}_h \right\}. \end{aligned}$$

The solvability analysis and *a priori* error estimates for (4.8) have been derived in [5, Section 5].

## 4.3 A residual-based a posteriori error estimator

In this section we introduce a reliable and efficient residual-based *a posteriori* error estimator for the Galerkin scheme (4.8). In particular, as well as in [7], a Helmholtz decomposition will be employed in the corresponding proof of reliability. Even if this analysis will be restricted to the two-dimensional case using the discrete spaces from Section 4.2.3, an extension to the 3D case will be addressed in detail in Section 4.5, below.

Given a suitably chosen $r > 0$ (see [5] for details), we define the balls

$$W := \{\phi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega) : \quad \|\phi\|_{1,\Omega} \leq r\} \quad \text{and} \quad W_h := \{\phi_h \in \mathrm{H}_h^{\phi} : \quad \|\phi_h\|_{1,\Omega} \leq r\}\,, \tag{4.9}$$

and throughout the rest of the chapter we let $(\boldsymbol{\sigma}, \boldsymbol{u}, \phi) \in \mathbb{H}_{\mathrm{N}}(\mathbf{div}; \Omega) \times \mathbf{H}^1(\Omega) \times \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ with $\phi \in W$ and $(\boldsymbol{\sigma}_h, \boldsymbol{u}_h, \phi_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \times \mathbf{H}_h^{\boldsymbol{u}} \times \mathrm{H}_h^{\phi}$ with $\phi_h \in W_h$ be the solutions of the continuous and discrete formulations (4.5) and (4.8), respectively. In addition, we set

$$H := \mathbb{H}_{\mathrm{N}}(\mathbf{div}, \Omega) \times \mathbf{H}^1(\Omega)\,, \qquad \|(\boldsymbol{\tau}, \boldsymbol{v})\|_H := \|\boldsymbol{\tau}\|_{\mathbf{div};\Omega} + \|\boldsymbol{v}\|_{1,\Omega} \quad \forall\, (\boldsymbol{\tau}, \boldsymbol{v}) \in H\,,$$

and recall from [5, Theorems 3.13 and 4.7] that the following *a priori* estimates hold

$$\begin{aligned} \|(\boldsymbol{\sigma}, \boldsymbol{u})\|_H &\leq C_S \left\{ \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + \|\phi\|_{1,\Omega}\, \|\boldsymbol{f}\|_{\infty,\Omega} \right\}, \\ \|(\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\|_H &\leq C_S \left\{ \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}} + \|\phi_h\|_{1,\Omega}\, \|\boldsymbol{f}\|_{\infty,\Omega} \right\}, \end{aligned}$$

where $C_S$ is a positive constant independent of $\phi$ and $\phi_h$.

### 4.3.1 The local error indicator

Given $T \in \mathcal{T}_h$, we let $\mathcal{E}_h(T)$ be the set of its edges, and let $\mathcal{E}_h$ be the set of all edges of the triangulation $\mathcal{T}_h$. Then we write $\mathcal{E}_h = \mathcal{E}_h(\Omega) \cup \mathcal{E}_h(\Gamma_{\mathrm{D}}) \cup \mathcal{E}_h(\Gamma_{\mathrm{N}})$, where $\mathcal{E}_h(\Omega) := \{e \in \mathcal{E}_h : e \subseteq \Omega\}$, $\mathcal{E}_h(\Gamma_{\mathrm{D}}) := \{e \in \mathcal{E}_h : e \subseteq \Gamma_{\mathrm{D}}\}$ and $\mathcal{E}_h(\Gamma_{\mathrm{N}}) := \{e \in \mathcal{E}_h : e \subseteq \Gamma_{\mathrm{N}}\}$. Also, for each edge $e \in \mathcal{E}_h$ we fix a unit

normal vector $\boldsymbol{\nu}_e := (\nu_1, \nu_2)^{\mathtt{t}}$, and let $\boldsymbol{s}_e := (-\nu_2, \nu_1)^{\mathtt{t}}$ be the corresponding fixed unit tangential vector along $e$. Then, given $e \in \mathcal{E}_h(\Omega)$ and $\boldsymbol{v} \in \mathbf{L}^2(\Omega)$ such that $\boldsymbol{v}|_T \in \mathbf{C}(T)$ on each $T \in \mathcal{T}_h$, we let $[\![\boldsymbol{v} \cdot \boldsymbol{\nu}_e]\!]$ be the corresponding jump across $e$, that is, $[\![\boldsymbol{v} \cdot \boldsymbol{\nu}_e]\!] := (\boldsymbol{v}|_T - \boldsymbol{v}|_{T'})|_e \cdot \boldsymbol{\nu}_e$, where $T$ and $T'$ are the triangles of $\mathcal{T}_h$ having $e$ as a common edge. Similarly, given a tensor field $\boldsymbol{\tau} \in \mathbb{L}^2(\Omega)$ such that $\boldsymbol{\tau}|_T \in \mathbb{C}(T)$ on each $T \in \mathcal{T}_h$, we let $[\![\boldsymbol{\tau} \boldsymbol{s}_e]\!]$ be the corresponding jump across $e$, that is, $[\![\boldsymbol{\tau} \boldsymbol{s}_e]\!] := (\boldsymbol{\tau}|_T - \boldsymbol{\tau}|_{T'})|_e \, \boldsymbol{s}_e$. If no confusion arises, we will simple write $\boldsymbol{s}$ and $\boldsymbol{\nu}$ instead $\boldsymbol{s}_e$ and $\boldsymbol{\nu}_e$, respectively.

Moreover, given scalar, vector, and tensor valued fields $v$, $\boldsymbol{\varphi} := (\varphi_1, \varphi_2)$ and $\boldsymbol{\tau} := (\tau_{ij})$, respectively, we denote

$$\mathbf{curl}(v) := \begin{pmatrix} \frac{\partial v}{\partial x_2} \\ -\frac{\partial v}{\partial x_1} \end{pmatrix}, \quad \underline{\mathbf{curl}}(\boldsymbol{\varphi}) := \begin{pmatrix} \mathbf{curl}(\varphi_1)^{\mathtt{t}} \\ \mathbf{curl}(\varphi_2)^{\mathtt{t}} \end{pmatrix}, \quad \text{and} \quad \mathbf{curl}(\boldsymbol{\tau}) := \begin{pmatrix} \frac{\partial \tau_{12}}{\partial x_1} - \frac{\partial \tau_{11}}{\partial x_2} \\ \frac{\partial \tau_{22}}{\partial x_1} - \frac{\partial \tau_{21}}{\partial x_2} \end{pmatrix}.$$

Then we let $\widetilde{\boldsymbol{\sigma}}_h := \vartheta(\phi_h)\nabla\phi_h - \phi_h \boldsymbol{u}_h - \mathit{f}_{\mathrm{bk}}(\phi_h)\boldsymbol{k}$ and define for each $T \in \mathcal{T}_h$ a local error indicator as follows

$$\begin{aligned}
\theta_T^2 :=& \|\boldsymbol{f}\phi_h - (\boldsymbol{K}^{-1}\boldsymbol{u}_h - \mathbf{div}\boldsymbol{\sigma}_h)\|_{0,T}^2 + \left\|\nabla \boldsymbol{u}_h - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\|_{0,T}^2 + h_T^2 \|g - (\beta\phi_h - \mathrm{div}\widetilde{\boldsymbol{\sigma}}_h)\|_{0,T}^2 \\
&+ h_T^2 \left\|\mathbf{curl}\left\{\frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\}\right\|_{0,T}^2 + \sum_{e \in \mathcal{E}_h(T) \cap \mathcal{E}_h(\Omega)} h_e \left\|\left[\!\left[\frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\, \boldsymbol{s}\right]\!\right]\right\|_{0,e}^2 \\
&+ \sum_{e \in \mathcal{E}_h(T) \cap \mathcal{E}_h(\Omega)} h_e \|[\![\widetilde{\boldsymbol{\sigma}}_h \cdot \boldsymbol{\nu}_e]\!]\|_{0,e}^2 + \sum_{e \in \mathcal{E}_h(T) \cap \mathcal{E}_h(\Gamma_{\mathrm{N}})} h_e \|\widetilde{\boldsymbol{\sigma}}_h \cdot \boldsymbol{\nu}\|_{0,e}^2 \\
&+ \sum_{e \in \mathcal{E}_h(T) \cap \mathcal{E}_h(\Gamma_{\mathrm{D}})} \|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|_{0,e}^2 + \sum_{e \in \mathcal{E}_h(T) \cap \mathcal{E}_h(\Gamma_{\mathrm{D}})} h_e \left\|\frac{d\boldsymbol{u}_{\mathrm{D}}}{d\boldsymbol{s}} - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\, \boldsymbol{s}\right\|_{0,e}^2 .
\end{aligned} \tag{4.10}$$

We remark that the last term defining $\theta_T^2$ requires that $\left.\frac{d\boldsymbol{u}_{\mathrm{D}}}{d\boldsymbol{s}}\right|_e \in \mathbf{L}^2(e)$ for each $e \in \mathcal{E}_h(\Gamma_{\mathrm{D}})$. This is fixed by assuming from now on that $\boldsymbol{u}_{\mathrm{D}} \in \mathbf{H}^1_0(\Gamma_{\mathrm{D}})$. In turn, it is not difficult to see that each term defining $\theta_T^2$ has a residual character, and hence, proceeding as usual, a *global* residual error estimator can be defined as

$$\boldsymbol{\theta} := \left\{\sum_{T \in \mathcal{T}_h} \theta_T^2\right\}^{1/2}. \tag{4.11}$$

### 4.3.2 Reliability

Throughout the rest of the chapter we assume that $\Gamma_{\mathrm{N}}$ is contained in the boundary of a convex extension of $\Omega$, that is, there exists a convex domain $B$ such that $\Omega \subseteq B$ and $\Gamma_{\mathrm{N}} \subseteq \partial B$ (see, e.g. [64, Theorem 3.2 and Figure 3.1]). Furthermore, according to the regularity estimate given in [5, eq. (3.24)], we also suppose from now on that $g \in \mathrm{H}^{\delta}(\Omega)$ for some $\delta \in (0, 1)$. Then the main result of this section is stated as follows.

**Theorem 4.1.** *Assume that $\Omega$ is a connected domain and that $\boldsymbol{u}_{\mathrm{D}}$, $\Gamma_{\mathrm{N}}$, and $g$ are as stated above. In addition, assume that the data $\boldsymbol{k}$, $g$, $\vartheta$, $\boldsymbol{u}_{\mathrm{D}}$, and $\boldsymbol{f}$ are sufficiently small so that there holds*

$$C_4\,|\boldsymbol{k}| + C_5\,\|g\|_{\delta,\Omega} + C_6\,\vartheta_2 + C_7\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} + C_8\,\|\boldsymbol{f}\|_{\infty,\Omega} < \frac{1}{2}, \tag{4.12}$$

*where the involved constants are made precise in (4.20), below. Then, there exists a constant* $C_{\mathtt{rel}} > 0$, *which depends only on the model parameters, on* $\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}}$, $\|\boldsymbol{f}\|_{\infty,\Omega}$, *and possibly other constants, but all independent of* $h$, *such that*

$$\|\phi - \phi_h\|_{1,\Omega} + \|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H \leq C_{\mathtt{rel}}\,\boldsymbol{\theta}. \tag{4.13}$$

A couple of preliminary estimates aiming to prove (4.13) are given in the following two subsections.

#### 4.3.2.1 A preliminary estimate for $\|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H$

In order to simplify the subsequent writing, we introduce in advance the following constants

$$C_0 := \frac{1}{\alpha(\Omega)}, \quad C_1 := 2\,C_0\,C_\delta\,\widetilde{C}_\delta\,\widehat{C}_S(r)\frac{L_\mu(1+\kappa_1^2)^{1/2}}{\mu_1^2}, \quad C_2 := C_0\,(1+\kappa_2^2)^{1/2} + r\,C_1, \tag{4.14}$$

where $\widehat{C}_S(r)$ and $C_\delta$, $\widetilde{C}_\delta$ are defined in [5, eq. (3.23) ] and [5, Lemma 3.6 and Theorem 3.10], respectively.

**Lemma 4.1.** *Let* $\boldsymbol{\theta}_0^2 := \sum_{T\in\mathcal{T}_h}\theta_{0,T}^2$, *where for each* $T\in\mathcal{T}_h$ *we set*

$$\theta_{0,T}^2 := \|\boldsymbol{f}\phi_h - (\boldsymbol{K}^{-1}\boldsymbol{u}_h - \mathbf{div}\boldsymbol{\sigma}_h)\|_{0,T}^2 + \left\|\nabla\boldsymbol{u}_h - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\|_{0,T}^2. \tag{4.15}$$

*Then there exists* $\bar{C} > 0$, *depending on* $C_0$, $\kappa_1$, *such that*

$$\|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H \leq \bar{C}\left\{\boldsymbol{\theta}_0 + \|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}\right\} + \left\{C_1\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} + C_2\,\|\boldsymbol{f}\|_{\infty,\Omega}\right\}\|\phi-\phi_h\|_{1,\Omega}, \tag{4.16}$$

*where* $C_1$ *and* $C_2$ *are given by (4.14), and the functional* $E_h\in\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'$ *is defined by*

$$\begin{aligned} E_h(\boldsymbol{\zeta}) := {} & \langle\boldsymbol{\zeta}\,\boldsymbol{\nu},\boldsymbol{u}_{\mathrm{D}}\rangle_{\Gamma_{\mathrm{D}}} - \int_\Omega \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}} : \boldsymbol{\zeta} - \int_\Omega \boldsymbol{u}_h\cdot\mathbf{div}\boldsymbol{\zeta} \\ & - \kappa_2\int_\Omega(\boldsymbol{f}\phi_h - (\boldsymbol{K}^{-1}\boldsymbol{u}_h - \mathbf{div}\boldsymbol{\sigma}_h))\cdot\mathbf{div}\boldsymbol{\zeta} \qquad \forall\,\boldsymbol{\zeta}\in\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega). \end{aligned} \tag{4.17}$$

*In addition, there holds*

$$E_h(\boldsymbol{\zeta}_h) = 0 \qquad \forall\,\boldsymbol{\zeta}_h\in\mathbb{H}_h^{\boldsymbol{\sigma}}. \tag{4.18}$$

*Proof.* Even though the present bilinear form $B_\phi$ (*cf.* (4.6)) and the corresponding one from [7, eq. (2.9)] differ in a couple of linear terms, the present proof is almost verbatim as [7, Lemma 3.2], particularly concerning the application of the $H$-ellipticity (see [5, Lemma 3.3]) of $B_\phi$ to the error $(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\sigma}_h,\boldsymbol{u}_h)$, and the estimates for $|B_{\phi_h}(\,\cdot\,,(\boldsymbol{\zeta},\boldsymbol{w})) - B_\phi(\,\cdot\,,(\boldsymbol{\zeta},\boldsymbol{w}))|$ and $|F_\phi(\boldsymbol{\zeta},\boldsymbol{w}) - F_{\phi_h}(\boldsymbol{\zeta},\boldsymbol{w})|$, and hence further details are omitted. □

Observe, according to (4.18), that for each $\boldsymbol{\zeta}\in\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)$ we can write

$$E_h(\boldsymbol{\zeta}) = E_h(\boldsymbol{\zeta} - \boldsymbol{\zeta}_h) \qquad \forall\,\boldsymbol{\zeta}_h\in\mathbb{H}_h^{\boldsymbol{\sigma}},$$

and hence the upper bound of $\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}$ to be derived below (see Section 4.3.2.4) will employ the foregoing expression with a suitable choice of $\boldsymbol{\zeta}_h\in\mathbb{H}_h^{\boldsymbol{\sigma}}$.

We end this section with an alternative expression for the functional $E_h$, which will be used later on in Section 4.3.2.4 to obtain a partial estimate for $\boldsymbol{\theta}$, and then in Section 4.4 to derive a second *a posteriori* error estimator. In fact, integrating by parts the expression $\int_\Omega \boldsymbol{u}_h \cdot \mathbf{div}\boldsymbol{\zeta}$, and using the homogeneous Neumann boundary condition on $\Gamma_{\mathrm{N}}$, we find that $E_h$ can be rewritten as

$$\begin{aligned} E_h(\boldsymbol{\zeta}) := \langle \boldsymbol{\zeta}\,\boldsymbol{\nu}, \boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\rangle_{\Gamma_{\mathrm{D}}} + \int_\Omega \Big(\nabla \boldsymbol{u}_h - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\Big) : \boldsymbol{\zeta} \\ - \kappa_2 \int_\Omega (\boldsymbol{f}\phi_h - (\boldsymbol{K}^{-1}\boldsymbol{u}_h - \mathbf{div}\boldsymbol{\sigma}_h)) \cdot \mathbf{div}\boldsymbol{\zeta} \qquad \forall\, \boldsymbol{\zeta} \in \mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)\,. \end{aligned} \tag{4.19}$$

#### 4.3.2.2 A preliminary estimate for $\|\phi - \phi_h\|_{1,\Omega}$

In contrast with [7, Section 3.2.2], in this section we establish an estimate for the error $\|\phi - \phi_h\|_{1,\Omega}$ by employing the ellipticity of the bilinear form $A_{\phi,\boldsymbol{u}}$ [5, eq. (3.13)]. The reason of the latter is due to fact that the Gâteaux derivative of the nonlinear induced operator by the form $A_{\boldsymbol{u}}$ (*cf.* (4.7)) is not elliptic as it was in [7], and hence we can not apply [7, Lemma 3.4] to derive the corresponding preliminary bound. In light of the above, we now set the following constants

$$\begin{aligned} &\widetilde{C} := \frac{1}{\widetilde{\alpha}}, \quad C_3 := L_\vartheta\, C_\delta\, \widetilde{C}_\delta\, \widehat{C}_{\widetilde{S}}(r), \quad C_4 := \widetilde{C}\,(L_f + C_3\,\gamma_2|\Omega|^{1/2}), \quad C_5 := \widetilde{C}\,C_3, \\ &C_6 := 2\,\widetilde{C}, \quad C_7 := r\,c(\Omega)\,\widetilde{C}\,C_1, \quad C_8 := r\,c(\Omega)\,\widetilde{C}\,C_2, \quad C_9 := r\,c(\Omega)\,\bar{C}, \end{aligned} \tag{4.20}$$

where $\widehat{C}_{\widetilde{S}}(r)$, $C_\delta$, $\widetilde{C}_\delta$ and $\bar{C}$ are the constants provided by [5, eq. (3.24)], [5, Lemma 3.7, Theorem 3.10], and Lemma 4.1, respectively.

**Lemma 4.2.** *Assume that the data $\boldsymbol{k}$, $g$, $\vartheta$, $\boldsymbol{u}_{\mathrm{D}}$, and $\boldsymbol{f}$ are sufficiently small so that there holds*

$$C_4\,|\boldsymbol{k}| + C_5\,\|g\|_{\delta,\Omega} + C_6\,\vartheta_2 + C_7\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} + C_8\,\|\boldsymbol{f}\|_{\infty,\Omega} < \frac{1}{2}. \tag{4.21}$$

*Then, there exists $\widehat{C} > 0$, depending on $\widetilde{C}$ and $C_9$ (cf. (4.20)), such that*

$$\|\phi - \phi_h\|_{1,\Omega} \leq \widehat{C}\left\{\|\boldsymbol{\theta}_0 + \|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'} + \|\widetilde{E}_h\|_{\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'}\right\}, \tag{4.22}$$

*where $\boldsymbol{\theta}_0$ and $E_h$ are given in the statement of Lemma 4.1 and (4.17), respectively, and $\widetilde{E}_h \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'$ is defined for each $\psi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ by*

$$\widetilde{E}_h(\psi) := \int_\Omega (g - \beta\,\phi_h)\,\psi - \int_\Omega \Big\{\vartheta(\phi_h)\,\nabla\phi_h - \phi_h\,\boldsymbol{u}_h - f_{\mathrm{bk}}(\phi_h)\,\boldsymbol{k}\Big\} \cdot \nabla\psi\,. \tag{4.23}$$

*In addition, there holds*

$$\widetilde{E}_h(\psi_h) = 0 \qquad \forall\, \psi_h \in \mathrm{H}^\phi_h\,. \tag{4.24}$$

*Proof.* We begin by recalling, from [5, Lemma 3.4], that the bilinear form

$$A_{\phi,\boldsymbol{u}}(\varphi,\psi) := \int_\Omega \vartheta(\phi)\,\nabla\varphi\cdot\nabla\psi - \int_\Omega \varphi\,\boldsymbol{u}\cdot\nabla\psi + \int_\Omega \beta\,\varphi\,\psi \quad \forall\,\varphi,\psi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega), \tag{4.25}$$

is $\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$-elliptic with constant $\widetilde{\alpha} := \frac{\vartheta_1}{2\,c_p^2}$, from which we deduce the following global inf-sup condition

$$\widetilde{\alpha}\,\|\varphi\|_{1,\Omega} \leq \sup_{\substack{\psi\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\\ \psi\neq 0}} \frac{A_{\phi,\boldsymbol{u}}(\varphi,\psi)}{\|\psi\|_{1,\Omega}} \qquad \forall\,\varphi \in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega). \tag{4.26}$$

Next, applying (4.26) to the Galerkin error $\varphi := \phi - \phi_h$, we find that

$$\widetilde{\alpha}\,\|\phi-\phi_h\|_{1,\Omega} \;\leq\; \sup_{\substack{\psi\in \mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)\\ \psi\neq 0}} \frac{A_{\phi,\boldsymbol{u}}(\phi,\psi)\;-\;A_{\phi,\boldsymbol{u}}(\phi_h,\psi)}{\|\psi\|_{1,\Omega}}. \tag{4.27}$$

Now, using the fact that $A_{\phi,\boldsymbol{u}}(\phi,\psi) = A_{\boldsymbol{u}}(\phi,\psi) = G_\phi(\psi)$, and adding and subtracting suitable terms, it follows that

$$A_{\phi,\boldsymbol{u}}(\phi,\psi)-A_{\phi,\boldsymbol{u}}(\phi_h,\psi) \;=\; G_\phi(\psi)-G_{\phi_h}(\psi)+G_{\phi_h}(\psi)-A_{\boldsymbol{u}_h}(\phi_h,\psi)+A_{\boldsymbol{u}_h}(\phi_h,\psi)-A_{\phi,\boldsymbol{u}}(\phi_h,\psi)\,. \tag{4.28}$$

In turn, using the definition of $A_{\boldsymbol{u}_h}$ (*cf.* (4.7)) and $A_{\phi,\boldsymbol{u}}$ (*cf.* (4.25)), we find that

$$\begin{aligned} A_{\boldsymbol{u}_h}(\phi_h,\psi)\;-\;A_{\phi,\boldsymbol{u}}(\phi_h,\psi) \;&=\; \int_\Omega (\vartheta(\phi)-\vartheta(\phi_h))\,\nabla(\phi-\phi_h)\cdot\nabla\psi \\ &+\int_\Omega \phi_h\,(\boldsymbol{u}-\boldsymbol{u}_h)\cdot\nabla\psi \;-\; \int_\Omega (\vartheta(\phi)-\vartheta(\phi_h))\,\nabla\phi\cdot\nabla\psi, \end{aligned} \tag{4.29}$$

from which, employing the upper bound of $\vartheta$ (*cf.* (4.3)), (4.9), and proceeding as in [5, eq. (5.13)-(5.14)] on the third term to the right hand side of (4.29), we arrive at

$$\begin{aligned} |A_{\boldsymbol{u}_h}(\phi_h,\psi)\;-\;A_{\phi,\boldsymbol{u}}(\phi_h,\psi)| \;\leq\; &\left\{\,2\,\vartheta_2\;+\;C_3\left(\gamma_2\,|\Omega|^{1/2}\,|\boldsymbol{k}|\;+\;\|g\|_{\delta,\Omega}\right)\right\}\|\phi-\phi_h\|_{1,\Omega}\,|\psi|_{1,\Omega} \\ &+\quad r\,c(\Omega)\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}\,|\psi|_{1,\Omega}\,. \end{aligned} \tag{4.30}$$

Thus, applying the estimate for $|G_\phi(\psi)-G_{\phi_h}(\psi)|$ (see [5, eq. (5.12)]) and estimate (4.30), we obtain from (4.27) and (4.28) that

$$\begin{aligned} \|\phi-\phi_h\|_{1,\Omega} \;\leq\; &\widetilde{C}\,\|G_{\phi_h}-A_{\boldsymbol{u}_h}(\phi_h,\cdot)\|_{\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'} \\ &+\left\{C_4|\boldsymbol{k}|\;+\;C_5\|g\|_{\delta,\Omega}+C_6\vartheta_2\right\}\|\phi-\phi_h\|_{1,\Omega}\;+\;r\,c(\Omega)\,\widetilde{C}\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}\,. \end{aligned}$$

Then, bounding $\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}$ by the error estimate provided by (4.16) (*cf.* Lemma 4.1), and employing (4.21), we deduce that

$$\|\phi-\phi_h\|_{1,\Omega} \;\leq\; 2\,\widetilde{C}\left\{\|G_{\phi_h}-A_{\boldsymbol{u}_h}(\phi_h,\cdot)\|_{\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'}\;+\;C_9\left(\boldsymbol{\theta}_0\;+\;\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}\right)\right\}, \tag{4.31}$$

where, bearing in mind (4.23), there holds

$$G_{\phi_h}\;-\;A_{\boldsymbol{u}_h}(\phi_h,\cdot)\;=\;\widetilde{E}_h\,,$$

and hence (4.31) yields (4.22). Finally, using the fact that $G_{\phi_h}(\psi_h)-A_{\boldsymbol{u}_h}(\phi_h,\psi_h)=0 \quad \forall\psi_h\in\mathrm{H}^\phi_h$, we obtain (4.24) and the proof concludes. □

We observe here that the upper bound in the assumption (4.21) could have been taken as any constant in $(0,1)$. We have chosen $\frac{1}{2}$ for simplicity and also in order to minimize the resulting constant $\widehat{C}$ in (4.22). Furthermore, it is important to remark, according to (4.24), that for each $\psi\in\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)$ there holds $\widetilde{E}_h(\psi)\;=\;\widetilde{E}_h(\psi-\psi_h) \quad \forall\,\psi_h\in\mathrm{H}^\phi_h$, and therefore $\|\widetilde{E}_h\|_{\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'}$ will be estimated below (see Section 4.3.2.4) by employing the foregoing expression with a suitable choice of $\psi_h\in\mathrm{H}^\phi_h$.

#### 4.3.2.3 A preliminary estimate for the total error

We now combine the inequalities provided by Lemmas 4.1 and 4.2 to derive a first estimate for the total error $\|\phi-\phi_h\|_{1,\Omega}+\|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H$. To this end, we now introduce the constants

$$C(\boldsymbol{u}_{\mathrm{D}},\boldsymbol{f}) := \widehat{C}\left\{C_1\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\varepsilon,\Gamma_{\mathrm{D}}} + C_2\,\|\boldsymbol{f}\|_{\infty,\Omega} + 1\right\} \quad \text{and} \quad c(\boldsymbol{u}_{\mathrm{D}},\boldsymbol{f}) := \bar{C} + C(\boldsymbol{u}_{\mathrm{D}},\boldsymbol{f})\,,$$

where $\bar{C}$ and $\widehat{C}$ are provided by Lemmas 4.1 and 4.2, respectively, and $C_1$ and $C_2$ are given by (4.14).

**Theorem 4.2.** *Assume that*

$$C_4\,|\boldsymbol{k}| + C_5\,\|g\|_{\delta,\Omega} + C_6\,\vartheta_2 + C_7\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} + C_8\,\|\boldsymbol{f}\|_{\infty,\Omega} < \frac{1}{2}.$$

*Then there holds*

$$\|\phi-\phi_h\|_{1,\Omega}+\|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H \leq C(\boldsymbol{u}_{\mathrm{D}},\boldsymbol{f})\,\|\widetilde{E}_h\|_{\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'} + c(\boldsymbol{u}_{\mathrm{D}},\boldsymbol{f})\left\{\boldsymbol{\theta}_0+\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}\right\}. \quad (4.32)$$

*Proof.* The estimate (4.32) is obtained by replacing the upper bound for $\|\phi-\phi_h\|_{1,\Omega}$, given by (4.22), into the second term on the right hand side of (4.16), and then adding the result to the right hand side of (4.22). □

Having established the upper bound (4.32), and in order to obtain an explicit estimate for the total error, we turn to the derivation of suitable upper bounds for $\|\widetilde{E}_h\|_{\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'}$ and $\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}$.

#### 4.3.2.4 Upper bounds for $\|\widetilde{E}_h\|_{\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'}$ and $\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}$

We begin by recalling the Clément interpolation operator $\mathcal{I}_h : \mathrm{H}^1(\Omega) \to X_h$ (*cf.* [42]), where

$$X_h := \{v_h \in C(\overline{\Omega}) : v_h|_T \in \mathrm{P}_1(T) \quad \forall T \in \mathcal{T}_h\}.$$

The following result states the local approximation properties of $\mathcal{I}_h$ (for a proof, see [42]).

**Lemma 4.3.** *There exist constants $c_1, c_2 > 0$, independent of $h$, such that for all $v \in \mathrm{H}^1(\Omega)$ there hold*

$$\|v-\mathcal{I}_h(v)\|_{0,T} \leq c_1 h_T \|v\|_{1,\Delta(T)} \quad \forall\, T \in \mathcal{T}_h,$$

*and*

$$\|v-\mathcal{I}_h(v)\|_{0,e} \leq c_2 h_e^{1/2} \|v\|_{1,\Delta(e)} \quad \forall\, e \in \mathcal{E}_h,$$

*where $\Delta(T)$ and $\Delta(e)$ are the union of all elements intersecting with $T$ and $e$, respectively.*

We now recall the definition of the concentration flux

$$\widetilde{\boldsymbol{\sigma}}_h := \vartheta(\phi_h)\nabla\phi_h - \phi_h\boldsymbol{u}_h - f_{\mathrm{bk}}(\phi_h)\boldsymbol{k}\,. \quad (4.33)$$

Then, the following lemma provides an upper bound for $\|\widetilde{E}_h\|_{\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'}$.

**Lemma 4.4.** *Let* $\widetilde{\boldsymbol{\eta}}^2 := \sum_{T\in\mathcal{T}_h} \widetilde{\eta}_T^2$, *where for each* $T \in \mathcal{T}_h$ *we set*

$$\widetilde{\eta}_T^2 := h_T^2 \|g - (\beta\phi_h - \mathrm{div}\widetilde{\boldsymbol{\sigma}}_h)\|_{0,T}^2 + \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Omega)} h_e \|[\![\widetilde{\boldsymbol{\sigma}}_h\cdot\boldsymbol{\nu}_e]\!]\|_{0,e}^2 + \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{N}})} h_e \|\widetilde{\boldsymbol{\sigma}}_h\cdot\boldsymbol{\nu}\|_{0,e}^2 .$$

*Then there exists* $c>0$, *independent of* $h$, *such that*

$$\|\widetilde{E}_h\|_{\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'} \leq c\,\widetilde{\boldsymbol{\eta}} . \tag{4.34}$$

*Proof.* It corresponds to a slight modification in the proof of [7, Lemma 3.8]. □

Our next goal is to provide an upper bound for $\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}$ (*cf.* (4.17)), which, being less straightforward than Lemma 4.4, requires several preliminary results. To this end, we start by introducing the space

$$\mathbf{H}^1_{\Gamma_{\mathrm{N}}}(\Omega) := \Big\{\boldsymbol{\varphi}\in\mathbf{H}^1(\Omega) : \quad \boldsymbol{\varphi} = \mathbf{0} \quad \text{on} \quad \Gamma_{\mathrm{N}}\Big\},$$

and establishing a suitable Helmholtz decomposition of our space $\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)$.

**Lemma 4.5.** *Assume that* $\Omega$ *is a connected domain and that* $\Gamma_{\mathrm{N}}$ *is contained in the boundary of a convex extension of* $\Omega$. *Then, for each* $\boldsymbol{\zeta}\in\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)$, *there exist* $\boldsymbol{\tau}\in\mathbb{H}^1(\Omega)$ *and* $\boldsymbol{\chi}\in\mathbf{H}^1_{\Gamma_{\mathrm{N}}}(\Omega)$ *such that*

$$\boldsymbol{\zeta} = \boldsymbol{\tau} + \underline{\mathbf{curl}}(\boldsymbol{\chi}) \quad \text{in} \quad \Omega , \tag{4.35}$$

*and*

$$\|\boldsymbol{\tau}\|_{1,\Omega} + \|\boldsymbol{\chi}\|_{1,\Omega} \leq C\,\|\boldsymbol{\zeta}\|_{\mathbf{div},\Omega} , \tag{4.36}$$

*with a positive constant* $C$ *independent of* $\boldsymbol{\zeta}$.

*Proof.* See [7, Lemma 3.9]. □

We continue our analysis by introducing the following finite element subspace of $\mathbf{H}_{\Gamma_{\mathrm{N}}}(\Omega)$

$$\mathbf{X}_{h,N} := \Big\{\boldsymbol{\varphi}_h\in\mathbf{C}(\overline{\Omega}) : \quad \boldsymbol{\varphi}_h|_T\in\mathbf{P}_1(T) \quad \forall\, T\in\mathcal{T}_h, \quad \boldsymbol{\varphi}_h = \mathbf{0} \text{ on } \Gamma_{\mathrm{N}}\Big\},$$

and considering, analogously as before, the Clément interpolation operator $\boldsymbol{\mathcal{I}}_{h,N} : \mathbf{H}_{\Gamma_{\mathrm{N}}}(\Omega)\to\mathbf{X}_{h,N}$. In addition, we let $\Pi_h : \mathbb{H}^1(\Omega)\to\mathbb{H}^{\boldsymbol{\sigma}}_h$ be the Raviart-Thomas interpolation operator (see [26, 95]), which, according to its characterization properties (see e.g. [63, Section 3.4.1]), verifies

$$\mathbf{div}(\Pi_h(\bar{\boldsymbol{\tau}})) = \mathcal{P}_h(\mathbf{div}\bar{\boldsymbol{\tau}}) \qquad \forall\,\bar{\boldsymbol{\tau}}\in\mathbb{H}^1(\Omega), \tag{4.37}$$

where $\mathcal{P}_h : \mathbf{L}^2(\Omega)\to\mathbf{Q}_h$ is the $\mathbf{L}^2(\Omega)$-orthogonal projector and

$$\mathbf{Q}_h := \big\{\boldsymbol{v}\in\mathbf{L}^2(\Omega) : \quad \boldsymbol{v}|_T\in\mathbf{P}_k(T) \quad \forall\, T\in\mathcal{T}_h\big\}.$$

Further approximation properties of $\Pi_h$ are summarized as follows (see [63, Lemmas 3.16 and 3.18]).

**Lemma 4.6.** *There exist $c_3, c_4 > 0$, independent of $h$, such that for all $\bar{\boldsymbol{\tau}} \in \mathbb{H}^1(\Omega)$ there holds*

$$\|\bar{\boldsymbol{\tau}} - \Pi_h(\bar{\boldsymbol{\tau}})\|_{0,T} \leq c_3\, h_T\, \|\bar{\boldsymbol{\tau}}\|_{1,T} \quad \forall\, T \in \mathcal{T}_h\,,$$

*and*

$$\|(\bar{\boldsymbol{\tau}} - \Pi_h(\bar{\boldsymbol{\tau}}))\,\boldsymbol{\nu}\|_{0,e} \leq c_4\, h_e^{1/2}\, \|\bar{\boldsymbol{\tau}}\|_{1,T_e} \quad \forall\, e \in \mathcal{E}_h(\Omega) \cup \mathcal{E}_h(\Gamma_{\rm D})\,,$$

*where $T_e$ is a triangle of $\mathcal{T}_h$ containing the edge $e$ on its boundary.*

Next, given $\boldsymbol{\zeta} \in \mathbb{H}_{\rm N}(\mathbf{div}, \Omega)$ and its Helmholtz decomposition (4.35), we define $\boldsymbol{\chi}_h := \boldsymbol{\mathcal{I}}_{h,N}(\boldsymbol{\chi})$, and set

$$\boldsymbol{\zeta}_h := \Pi_h(\boldsymbol{\tau}) + \underline{\mathbf{curl}}(\boldsymbol{\chi}_h) \in \mathbb{H}_h^{\boldsymbol{\sigma}} \tag{4.38}$$

as its associated discrete Helmholtz decomposition. Then, from (4.35) and (4.38), it follows that

$$\boldsymbol{\zeta} - \boldsymbol{\zeta}_h = \boldsymbol{\tau} - \Pi_h(\boldsymbol{\tau}) + \underline{\mathbf{curl}}(\boldsymbol{\chi} - \boldsymbol{\chi}_h)\,.$$

Therefore, according to (4.17) and (4.18), we deduce that

$$E_h(\boldsymbol{\zeta}) = E_h(\boldsymbol{\zeta} - \boldsymbol{\zeta}_h) = E_h(\boldsymbol{\tau} - \Pi_h(\boldsymbol{\tau})) + E_h(\underline{\mathbf{curl}}(\boldsymbol{\chi} - \boldsymbol{\chi}_h))\,. \tag{4.39}$$

Notice from (4.39) that, in order to estimate $\|E_h\|_{\mathbb{H}_{\rm N}(\mathbf{div},\Omega)'}$, it only remains to bound $|E_h(\boldsymbol{\tau} - \Pi_h(\boldsymbol{\tau}))|$ and $|E_h(\underline{\mathbf{curl}}(\boldsymbol{\chi} - \boldsymbol{\chi}_h))|$ in terms of a multiple of $\|\boldsymbol{\zeta}\|_{\mathbf{div},\Omega}$, which is done in the rest of the present Section 4.3.2.4. To this end, we now recall from [50] the following integration by parts formula on the boundary.

**Lemma 4.7.** *There holds*

$$\langle \underline{\mathbf{curl}}\, \boldsymbol{\chi}\, \boldsymbol{\nu}, \boldsymbol{\phi} \rangle = -\langle \frac{d\boldsymbol{\phi}}{d\boldsymbol{s}}, \boldsymbol{\chi} \rangle \qquad \forall\, \boldsymbol{\chi}, \boldsymbol{\phi} \in \mathbf{H}^1(\Omega). \tag{4.40}$$

*Proof.* It follows from suitable applications of the Green formulae provided in [77, Chapter I, eq. (2.17) and Theorem 2.11]. □

**Lemma 4.8.** *Let $\boldsymbol{\theta}_1^2 := \sum_{T \in \mathcal{T}_h} \theta_{1,T}^2$, where for each $T \in \mathcal{T}_h$ we set*

$$\theta_{1,T}^2 := h_T^2 \left\| \mathbf{curl} \left\{ \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\} \right\|_{0,T}^2 + \sum_{e \in \mathcal{E}_h(T) \cap \mathcal{E}_h(\Omega)} h_e \left\| \left[\!\!\left[ \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \boldsymbol{s} \right]\!\!\right] \right\|_{0,e}^2$$
$$+ \sum_{e \in \mathcal{E}_h(T) \cap \mathcal{E}_h(\Gamma_{\rm D})} h_e \left\| \frac{d\boldsymbol{u}_{\rm D}}{d\boldsymbol{s}} - \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}}\, \boldsymbol{s} \right\|_{0,e}^2 .$$

*Then there exists $c > 0$, independent of $h$, such that*

$$|E_h(\underline{\mathbf{curl}}(\boldsymbol{\chi} - \boldsymbol{\chi}_h)| \leq c\, \boldsymbol{\theta}_1\, \|\boldsymbol{\zeta}\|_{\mathbf{div},\Omega}\,. \tag{4.41}$$

*Proof.* See [7, Lemma 3.11]. □

**Lemma 4.9.** *Let* $\boldsymbol{\theta}_2^2 := \sum_{T\in\mathcal{T}_h} \theta_{2,T}^2$, *where for each* $T \in \mathcal{T}_h$ *we set*

$$\theta_{2,T}^2 := h_T^2 \left\| \nabla \boldsymbol{u}_h - \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\|_{0,T}^2 + \|\boldsymbol{f}\phi_h - (\boldsymbol{K}^{-1}\boldsymbol{u}_h - \mathbf{div}\boldsymbol{\sigma}_h)\|_{0,T}^2 + \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{D}})} h_e \|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|_{0,e}^2 .$$

*Then there exists* $c > 0$, *independent of* $h$, *such that*

$$|E_h(\boldsymbol{\tau} - \Pi_h(\boldsymbol{\tau}))| \leq c\,\boldsymbol{\theta}_2\,\|\boldsymbol{\zeta}\|_{\mathbf{div},\Omega} . \tag{4.42}$$

*Proof.* Using the alternative definition of the functional $E_h$ (*cf.* (4.19)), applying the identity (4.37), and denoting by $\mathcal{I}$ a generic identity operator, we find that

$$\begin{aligned} E_h(\boldsymbol{\tau} - \Pi_h(\boldsymbol{\tau})) \;=\; & \langle (\boldsymbol{\tau} - \Pi_h(\boldsymbol{\tau}))\,\boldsymbol{\nu}, \boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\rangle_{\Gamma_{\mathrm{D}}} + \int_\Omega \left( \nabla \boldsymbol{u}_h - \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \right) : (\boldsymbol{\tau} - \Pi_h(\boldsymbol{\tau})) \\ & - \kappa_2 \int_\Omega (\boldsymbol{f}\phi_h - (\boldsymbol{K}^{-1}\boldsymbol{u}_h - \mathbf{div}\boldsymbol{\sigma}_h)) \cdot (\mathcal{I} - \mathcal{P}_h)(\mathbf{div}\boldsymbol{\tau}). \end{aligned} \tag{4.43}$$

Next, the first two terms on the right hand side of (4.43) are simply bounded by applying the Cauchy-Schwarz in $\mathbf{L}^2(\Gamma_{\mathrm{D}})$ and $\mathbb{L}^2(\Omega)$, and then employing the approximation properties of $\Pi_h$ provided by Lemma 4.6. In turn, for the corresponding third term, it suffices to see, thanks to the Cauchy-Schwarz inequality and the stability estimate (4.36), that

$$\begin{aligned} & \left| \int_\Omega (\boldsymbol{f}\phi_h - (\boldsymbol{K}^{-1}\boldsymbol{u}_h - \mathbf{div}\boldsymbol{\sigma}_h)) \cdot (\mathcal{I} - \mathcal{P}_h)(\mathbf{div}\,\boldsymbol{\tau})) \right| \\ & \qquad \leq \quad \|\boldsymbol{f}\phi_h - (\boldsymbol{K}^{-1}\boldsymbol{u}_h - \mathbf{div}\boldsymbol{\sigma}_h)\|_{0,\Omega}\, \|\mathbf{div}\,\boldsymbol{\tau}\|_{0,\Omega} \leq \|\boldsymbol{f}\phi_h - (\boldsymbol{K}^{-1}\boldsymbol{u}_h - \mathbf{div}\boldsymbol{\sigma}_h)\|_{0,\Omega}\, \|\boldsymbol{\zeta}\|_{\mathbf{div},\Omega} , \end{aligned}$$

which ends the proof. □

By virtue of Lemmas 4.8 and 4.9 we deduce the following upper bound for $\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}$.

**Lemma 4.10.** *There exists* $c > 0$, *independent of* $h$, *such that*

$$\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'} \leq c \left\{ \boldsymbol{\theta}_1 + \boldsymbol{\theta}_2 \right\} .$$

*Proof.* It follows straightforwardly from (4.39) and the upper bounds (4.41) and (4.42). □

At this point we remark that the terms $h_T^2\|\nabla \boldsymbol{u}_h - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\|_{0,T}^2$ and $h_e\|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|_{0,e}^2$, which appear in the definition of $\theta_{2,T}^2$ (*cf.* Lemma 4.9), are dominated by $\|\nabla \boldsymbol{u}_h - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\|_{0,T}^2$ and $\|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|_{0,e}^2$, respectively, which form part of $\theta_{0,T}^2$ (*cf.* (4.15)). Therefore, the reliability estimate (4.13) (*cf.* Theorem 4.1) is a direct consequence of Theorem 4.2, the definition of $\boldsymbol{\theta}_0$ (*cf.* Lemma 4.1), and Lemmas 4.4, 4.8, 4.9, and 4.10.

We close this section by mentioning that the assumption (4.12) on the data $\vartheta$, $\boldsymbol{k}$, $g$, $\boldsymbol{u}_{\mathrm{D}}$, and $\boldsymbol{f}$, which, as shown throughout the foregoing analysis, is a key estimate to derive (4.13), is, unfortunately, unverifiable in practice. In fact, while the data are certainly known in advance, the constants $C_4$, $C_5$, $C_6$, $C_7$, $C_8$ involved in that condition (*cf.* (4.20)), which in turn are expressed in terms of the previous constants $C_1$ and $C_2$ (*cf.* (4.14)), depend all on boundedness and regularity constants of operators, as well as on parameters, some of which are not explicitly calculable, and hence it is not possible to check whether (4.12) is indeed satisfied or not. This is, however, a quite common fact arising in the analysis of many nonlinear problems, and only in very particular cases (usually related to simple geometries of the domain) it could eventually be circumvented.

### 4.3.3 Efficiency

The main result of this section is stated as follows.

**Theorem 4.3.** *Assume that $\nabla\phi \in \mathrm{L}^4(\Omega)$. Then, there exists a constant $\overline{C}_{\mathtt{eff}} > 0$, which depends only on parameters, $\|\boldsymbol{K}^{-1}\|_\infty$, $|\boldsymbol{k}|$, $\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}}$, $\|\boldsymbol{f}\|_{\infty,\Omega}$, $\|\nabla\phi\|_{\mathrm{L}^4(\Omega)}$ and other constants, all them independent of $h$, such that*

$$\overline{C}_{\mathtt{eff}}\,\boldsymbol{\theta} \leq \|\phi-\phi_h\|_{1,\Omega} + \|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega} + \|\mathbf{div}(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_{0,\Omega} + \left\|\frac{1}{\mu(\phi)}\boldsymbol{\sigma}^{\mathtt{d}} - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\|_{0,\Omega} + \text{h.o.t.} \quad (4.44)$$

*where* h.o.t. *stands for one or several terms of higher order. Moreover, under the assumption that $\boldsymbol{\sigma} \in \mathbb{L}^4(\Omega)$, there exists a constant $C_{\mathtt{eff}} > 0$, which depends only on parameters, $\|\boldsymbol{K}^{-1}\|_\infty$, $|\boldsymbol{k}|$, $\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}}$, $\|\boldsymbol{f}\|_{\infty,\Omega}$, $\|\boldsymbol{\sigma}\|_{\mathbb{L}^4(\Omega)}$, $\|\nabla\phi\|_{\mathrm{L}^4(\Omega)}$ and other constants, all them independent of $h$, such that*

$$C_{\mathtt{eff}}\,\boldsymbol{\theta} \leq \|\phi-\phi_h\|_{1,\Omega} + \|(\boldsymbol{\sigma},\boldsymbol{u})-(\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H + \text{h.o.t.} \quad (4.45)$$

In the subsequent analysis, such as in [7], we assume for simplicity that the nonlinear functions $\mu$, $\vartheta$, and $f_{\mathrm{bk}}$ are such that $\dfrac{1}{\mu(\phi_h)}$, $\vartheta(\phi_h)$, $f_{\mathrm{bk}}(\phi_h)$, and hence $\widetilde{\boldsymbol{\sigma}}_h$ as well, are all piecewise polynomials. In addition, we assume that the data $\boldsymbol{u}_{\mathrm{D}}$ and $g$ are piecewise polynomials. Otherwise, and if $\mu^{-1}$, $\vartheta$, $f_{\mathrm{bk}}$, $\boldsymbol{u}_{\mathrm{D}}$, and $g$ are sufficiently smooth, higher order terms given by the errors arising from suitable polynomial approximations of these expressions and functions would appear in (4.44) and (4.45) (*cf.* Theorem 4.3), which explains the eventual h.o.t. in these expressions. In this regard, and similarly as observed in [7], we remark that (4.44) constitutes a *quasi-efficiency* estimate for the global residual error estimator $\boldsymbol{\theta}$ (*cf.* (4.11)). Indeed, the fact that the expression appearing on the right hand side of (4.44) is not exactly the error, but part of it plus the nonlinear term given by $\|\frac{1}{\mu(\phi)}\boldsymbol{\sigma}^{\mathtt{d}} - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\|_{0,\Omega}$, explains the *quasi-efficiency* concept employed here. Nevertheless, we show at the end of this section that, under the assumption that $\boldsymbol{\sigma} \in \mathbb{L}^4(\Omega)$, the latter can be bounded by $\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{0,\Omega} + \|\phi-\phi_h\|_{1,\Omega}$, thus yielding the efficiency estimate given by (4.45).

In order to prove (4.44) and (4.45), in the rest of this section we derive suitable upper bounds for the ten terms defining the local error indicator $\theta_T^2$ (*cf.* (4.10)). We begin by observing, thanks to the fact that $\boldsymbol{f}\phi = \boldsymbol{K}^{-1}\boldsymbol{u} - \mathbf{div}\boldsymbol{\sigma}$ in $\Omega$, that there holds

$$\begin{aligned}\|\boldsymbol{f}\phi_h - (\boldsymbol{K}^{-1}\boldsymbol{u}_h - \mathbf{div}\boldsymbol{\sigma}_h)\|_{0,T}^2 &\leq 2\,\|\boldsymbol{f}\|_{\infty,\Omega}^2\,\|\phi-\phi_h\|_{0,T}^2 \\ &+ 4\,\|\boldsymbol{K}^{-1}\|_\infty^2\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{0,T}^2 + 4\,\|\mathbf{div}(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_{0,T}^2\,.\end{aligned} \quad (4.46)$$

On the other hand, using that $\nabla\boldsymbol{u} = \frac{1}{\mu(\phi)}\boldsymbol{\sigma}^{\mathtt{d}}$ in $\Omega$, $\boldsymbol{u} = \boldsymbol{u}_{\mathrm{D}}$ on $\Gamma_{\mathrm{D}}$, and proceeding as in [7, Section 3.3], we deduce that

$$\left\|\nabla\boldsymbol{u}_h - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\|_{0,T}^2 \leq 2\,\|\nabla\boldsymbol{u}-\nabla\boldsymbol{u}_h\|_{0,T}^2 + 2\left\|\frac{1}{\mu(\phi)}\boldsymbol{\sigma}^{\mathtt{d}} - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\|_{0,T}^2 \quad (4.47)$$

and

$$\sum_{e\in\mathcal{E}_h(\Gamma_{\mathrm{D}})} \|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|_{0,e}^2 \leq c_0^2\,\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}^2\,, \quad (4.48)$$

where $c_0$ is the norm of the trace operator in $\mathbf{H}^1(\Omega)$.

The efficiency estimates for the remaining seven terms given in (4.11), are provided next. To this end, we proceed as in [37] and [38] (see also [61]), and apply the localization technique (see [103]) based on triangle-bubble and edge-bubble functions, together with extension operators, and inverse inequalities. Therefore, we now introduce further notations and preliminary results. In fact, given $T \in \mathcal{T}_h$ and $e \in \mathcal{E}_h(T)$, we let $\psi_T$ and $\psi_e$ be the usual triangle-bubble and edge-bubble functions, respectively (see [103, eqs. (1.4) and (1.6)]), which satisfy:

i) $\psi_T \in P_3(T)$, $\mathrm{supp}(\psi_T) \subseteq T$, $\psi_T = 0$ on $\partial T$, and $0 \leq \psi_T \leq 1$ in $T$.

ii) $\psi_e|_T \in P_2(T)$, $\mathrm{supp}(\psi_e) \subseteq \omega_e := \cup\{T' \in \mathcal{T}_h : e \in \mathcal{E}_h(T')\}$, $\psi_e = 0$ on $\partial T \backslash \{e\}$, and $0 \leq \psi_e \leq 1$ in $\omega_e$.

We also know from [102] that, given $k \in \mathbb{N} \cup \{0\}$, there exists an extension operator $L : C(e) \to C(T)$ that satisfies $L(p) \in P_k(T)$ and $L(p)|_e = p \quad \forall p \in P_k(e)$. A corresponding vectorial version of $L$, that is the component-wise application of $L$, is denoted by $\mathbf{L}$. Additional properties of $\psi_T, \psi_e$ and $L$ are collected in the following Lemma.

**Lemma 4.11.** *Given $k \in \mathbb{N} \cup \{0\}$, there exist positive constants $c_1, c_2, c_3$, and $c_4$, depending only on $k$ and the shape regularity of the triangulations (minimum angle condition), such that for each $T \in \mathcal{T}_h$ and $e \in \mathcal{E}_h(T)$, there hold*

$$\begin{aligned} \|\psi_T\, q\|^2_{0,T} \;\leq\; \|q\|^2_{0,T} \;\leq\; c_1\, \|\psi_T^{1/2}\, q\|^2_{0,T} \qquad \forall\, q \in P_k(T), \\ \|\psi_e\, L(p)\|^2_{0,T} \;\leq\; \|p\|^2_{0,e} \;\leq\; c_2\, \|\psi_e^{1/2}\, p\|^2_{0,e} \qquad \forall\, p \in P_k(e), \\ c_3\, h_e\, \|p\|^2_{0,e} \;\leq\; \|\psi_e^{1/2}\, L(p)\|^2_{0,T} \;\leq\; c_4\, h_e\, \|p\|^2_{0,e} \qquad \forall\, p \in P_k(e). \end{aligned} \tag{4.49}$$

*Proof.* See [102, Lemma 4.1]. □

The following inverse estimate is also required.

**Lemma 4.12.** *Let $l, m \in \mathbb{N} \cup \{0\}$ such that $l \leq m$. Then, there exists $c > 0$, depending only on $k, l, m$ and the shape regularity of the triangulations, such that for each $T \in \mathcal{T}_h$ there holds*

$$|q|_{m,T} \;\leq c\, h_T^{l-m}\, |q|_{l,T} \qquad \forall\, q \in P_k(T). \tag{4.50}$$

*Proof.* See [40, Theorem 3.2.6]. □

In turn, the following lemma, whose proof make use of lemmas 4.11 and 4.12, will be required for the terms involving the **curl** operator and the tangential jumps across the edges of $\mathcal{T}_h$.

**Lemma 4.13.** *Let $\boldsymbol{\rho}_h \in \mathbb{L}^2(\Omega)$ be a piecewise polynomial of degree $k \geq 0$ on each $T \in \mathcal{T}_h$. In addition, let $\boldsymbol{\rho} \in \mathbb{L}^2(\Omega)$ be such that $\mathbf{curl}(\boldsymbol{\rho}) = 0$ on each $T \in \mathcal{T}_h$. Then, there exist $c, \widetilde{c} > 0$, independent of $h$, such that*

$$\|\mathbf{curl}(\boldsymbol{\rho}_h)\|_{0,T} \;\leq\; c\, h_T^{-1}\, \|\boldsymbol{\rho} - \boldsymbol{\rho}_h\|_{0,T} \qquad \forall\, T \in \mathcal{T}_h$$

*and*

$$\|[\![\boldsymbol{\rho}_h \boldsymbol{s}_e]\!]\|_{0,e} \;\leq\; \widetilde{c}\, h_e^{-1/2}\, \|\boldsymbol{\rho} - \boldsymbol{\rho}_h\|_{0,\omega_e} \qquad \forall\, e \in \mathcal{E}_h\,.$$

*Proof.* For the first estimate we refer to [37, Lemma 4.3], whereas the second one follows from a slight modification of the proof of [37, Lemma 4.4]. Further details are omitted. □

We now apply Lemma 4.13 to obtain upper bounds for two other terms defining $\theta_T^2$.

**Lemma 4.14.** *There exist $\widetilde{c}_1, \widetilde{c}_2 > 0$, independent of $h$ such that*

$$h_T^2 \left\| \mathbf{curl}\left\{ \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\} \right\|_{0,T}^2 \leq \widetilde{c}_1 \left\| \frac{1}{\mu(\phi)} \boldsymbol{\sigma}^{\mathtt{d}} - \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\|_{0,T}^2 \qquad \forall\, T \in \mathcal{T}_h,$$

$$h_e \left\| \left[\!\!\left[ \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}}\, \boldsymbol{s} \right]\!\!\right] \right\|_{0,e}^2 \leq \widetilde{c}_2 \left\| \frac{1}{\mu(\phi)} \boldsymbol{\sigma}^{\mathtt{d}} - \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\|_{0,\omega_e}^2 \qquad \forall\, e \in \mathcal{E}_h(\Omega).$$

*Proof.* It suffices to apply Lemma 4.13 to $\boldsymbol{\rho}_h := \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}}$ and $\boldsymbol{\rho} := \frac{1}{\mu(\phi)} \boldsymbol{\sigma}^{\mathtt{d}} = \nabla \boldsymbol{u}$. □

**Lemma 4.15.** *There exists $\widetilde{c}_3 > 0$, independent of $h$, such that*

$$h_e \left\| \frac{d\boldsymbol{u}_{\mathrm{D}}}{d\boldsymbol{s}} - \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}}\, \boldsymbol{s} \right\|_{0,e}^2 \leq \widetilde{c}_3 \left\| \frac{1}{\mu(\phi)} \boldsymbol{\sigma}^{\mathtt{d}} - \frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \right\|_{0,T_e}^2 \qquad \forall\, e \in \mathcal{E}_h(\Gamma_{\mathrm{D}}).$$

*Proof.* We proceed similarly as in the proof of [72, Lemma 4.15], by replacing $\mathbf{g}$, $\Gamma$, and $\frac{1}{\mu} \boldsymbol{\sigma}_h^{\mathtt{d}}$ in [72] by $\boldsymbol{u}_{\mathrm{D}}$, $\Gamma_{\mathrm{D}}$, and $\frac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}}$, respectively. □

We now aim to provide upper bounds for the three terms completing the definition of the local error indicator $\theta_T^2$ (*cf.* (4.10)). This requires, however, the preliminary result given by the following *a priori* estimate for the error $\|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{0,T}^2$.

**Lemma 4.16.** *There exists $C > 0$, depending on $\vartheta_2$, $L_f$ (cf. (1.2), (4.4)), and $|\boldsymbol{k}|$, such that*

$$\|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{0,T}^2 \leq C \left\{ \|\phi - \phi_h\|_{1,T}^2 + \|\boldsymbol{u}(\phi - \phi_h)\|_{0,T}^2 + \|\phi_h(\boldsymbol{u} - \boldsymbol{u}_h)\|_{0,T}^2 + \|(\vartheta(\phi) - \vartheta(\phi_h))\nabla\phi\|_{0,T}^2 \right\}. \quad (4.51)$$

*Proof.* Employing the definitions of $\widetilde{\boldsymbol{\sigma}}$ (*cf.* (4.1)) and $\widetilde{\boldsymbol{\sigma}}_h$ (*cf.* (4.33)), applying the triangle inequality, and using the Lipschitz continuity assumption on $f_{\mathrm{bk}}$ (*cf.* (4.4)), but restricted to each $T \in \mathcal{T}_h$ instead of $\Omega$, we obtain that

$$\begin{aligned} \|\widetilde{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}_h\|_{0,T}^2 \leq 2 \Big\{ & \|\vartheta(\phi)\nabla\phi - \vartheta(\phi_h)\nabla\phi_h\|_{0,T}^2 + 2\, L_f^2\, |\boldsymbol{k}|^2\, \|\phi - \phi_h\|_{0,T}^2 \\ & + 4\, \|\boldsymbol{u}(\phi - \phi_h)\|_{0,T}^2 + 4\, \|\phi_h(\boldsymbol{u} - \boldsymbol{u}_h)\|_{0,T}^2 \Big\}. \end{aligned} \quad (4.52)$$

In turn, applying Cauchy-Schwarz's inequality and the upper bound for $\vartheta$ (*cf.* (4.21)), we deduce that

$$\|\vartheta(\phi)\nabla\phi - \vartheta(\phi_h)\nabla\phi_h\|_{0,T}^2 \leq 2\, \|(\vartheta(\phi) - \vartheta(\phi_h))\nabla\phi\|_{0,T}^2 + 2\, \vartheta_2^2\, \|\nabla\phi - \nabla\phi_h\|_{0,T}^2. \quad (4.53)$$

In this way, (4.52) and (4.53) imply (4.51), which finalizes the proof. □

We consider important to remark here that, due to the dependence on $\phi$ (instead of $|\nabla\phi|$ as in [7]) of the diffusivity $\vartheta$, the first term of our nonlinear operator $A_{\boldsymbol{u}}$ is not necessarily Lipschitz-continuous (as it was the case for the corresponding nonlinear operator in [7, eq. (2.11)]) and hence, in contrast

with [7, Lemma 3.19], now the term $\|(\vartheta(\phi)-\vartheta(\phi_h))\nabla\phi\|^2_{0,T}$ appears in the estimate (4.51) of Lemma 4.16. The treatment of such additional term will be postponed to Lemma 4.20.

We now establish the aforementioned efficiency estimates, split into three separate results.

**Lemma 4.17.** *There exists $\widetilde{c}_4>0$, which depends only on $\vartheta_2$, $L_f$, $\beta$ (cf. (4.3), (4.4), (1.5)), $|\boldsymbol{k}|$, and other constants, all them independent of $h$, such that*

$$h_T^2\,\|g-(\beta\phi_h-\mathrm{div}\widetilde{\boldsymbol{\sigma}}_h)\|^2_{0,T}\leq \widetilde{c}_4\Big\{\|\phi-\phi_h\|^2_{1,T}+\|\boldsymbol{u}(\phi-\phi_h)\|^2_{0,T} +\|\phi_h(\boldsymbol{u}-\boldsymbol{u}_h)\|^2_{0,T}+\|(\vartheta(\phi)-\vartheta(\phi_h))\nabla\phi\|^2_{0,T}+h_T^2\,\|\phi-\phi_h\|^2_{0,T}\Big\}. \tag{4.54}$$

*Proof.* It is an adaptation of the proof of [7, Lemma 3.22]. Indeed, given $T\in\mathcal{T}_h$, using the properties (4.49), the fact that $\beta\phi-\mathrm{div}\widetilde{\boldsymbol{\sigma}}=g$ in $\Omega$, and integrating by parts, we deduce that

$$\begin{aligned}\|g-(\beta\phi_h-\mathrm{div}\widetilde{\boldsymbol{\sigma}}_h)\|^2_{0,T}&\leq c_1\,\|\psi_T^{1/2}\,(g-(\beta\phi_h-\mathrm{div}\widetilde{\boldsymbol{\sigma}}_h))\|^2_{0,T}\\ &=-c_1\int_T\nabla\big(\psi_T\,(g-(\beta\phi_h-\mathrm{div}\widetilde{\boldsymbol{\sigma}}_h)\big)\cdot(\widetilde{\boldsymbol{\sigma}}-\widetilde{\boldsymbol{\sigma}}_h)+c_1\,\beta\int_T\psi_T\,(g-(\beta\phi_h-\mathrm{div}\widetilde{\boldsymbol{\sigma}}_h))\,(\phi-\phi_h).\end{aligned}$$

Then, the Cauchy-Schwarz inequality, the inverse estimate (4.50), the fact that $0\leq\psi_T\leq 1$, and the triangle inequality imply that

$$\begin{aligned}\|g-(\beta\phi_h-\mathrm{div}\widetilde{\boldsymbol{\sigma}}_h)\|^2_{0,T}&\leq c_1\,c\;h_T^{-1}\|g-(\beta\phi_h-\mathrm{div}\widetilde{\boldsymbol{\sigma}}_h)\|_{0,T}\,\|\widetilde{\boldsymbol{\sigma}}-\widetilde{\boldsymbol{\sigma}}_h\|_{0,T}\\ &+\quad c_1\beta\,\|g-(\beta\phi_h-\mathrm{div}\widetilde{\boldsymbol{\sigma}}_h)\|_{0,T}\,\|\phi-\phi_h\|_{0,T}\,,\end{aligned}$$

which gives

$$h_T\|g-(\beta\phi_h-\mathrm{div}\widetilde{\boldsymbol{\sigma}}_h)\|_{0,T}\leq C\Big\{\|\widetilde{\boldsymbol{\sigma}}-\widetilde{\boldsymbol{\sigma}}_h\|_{0,T}+h_T\,\|\phi-\phi_h\|_{0,T}\Big\}.$$

The foregoing inequality together with (4.51) imply (4.54), and thus the proof is completed. □

**Lemma 4.18.** *There exists $\widetilde{c}_5>0$, which depends only on $\vartheta_2$, $L_f$, $\beta$ (cf. (4.3), (4.4), (1.5)), $|\boldsymbol{k}|$, and other constants, all them independent of $h$, such that for each $e\in\mathcal{E}_h(\Omega)$ there holds*

$$h_e\,\|[\![\widetilde{\boldsymbol{\sigma}}_h\cdot\boldsymbol{\nu}_e]\!]\|^2_{0,e}\leq\widetilde{c}_5\sum_{T\subseteq\omega_e}\Big\{\|\phi-\phi_h\|^2_{1,T}+\|\boldsymbol{u}(\phi-\phi_h)\|^2_{0,T}+\|\phi_h(\boldsymbol{u}-\boldsymbol{u}_h)\|^2_{0,T} +\quad\|(\vartheta(\phi)-\vartheta(\phi_h))\nabla\phi\|^2_{0,T}+h_T^2\,\|\phi-\phi_h\|^2_{0,T}\Big\},$$

*where $\omega_e$ is the union of the two triangles in $\mathcal{T}_h$ having $e$ as an edge.*

*Proof.* See [7, Lemma 3.21]. □

**Lemma 4.19.** *There exists $\widetilde{c}_6>0$, which depends only on $\vartheta_2$, $L_f$, $\beta$ (cf. (4.3), (4.4), (1.5)), $|\boldsymbol{k}|$, and other constants, all them independent of $h$, such that for each $e\in\mathcal{E}_h(\Gamma_{\mathrm{N}})$ there holds*

$$h_e\,\|\widetilde{\boldsymbol{\sigma}}_h\cdot\boldsymbol{\nu}\|^2_{0,e}\leq\widetilde{c}_6\Big\{\|\phi-\phi_h\|^2_{1,T}+\|\boldsymbol{u}(\phi-\phi_h)\|^2_{0,T}+\|\phi_h(\boldsymbol{u}-\boldsymbol{u}_h)\|^2_{0,T} +\quad\|(\vartheta(\phi)-\vartheta(\phi_h))\nabla\phi\|^2_{0,T}+h_T^2\,\|\phi-\phi_h\|^2_{0,T}\Big\},$$

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

*Proof.* See [7, Lemma 3.22]. □

In order to complete the proof of global efficiency given by (4.44), it only remains to estimate properly the three terms: $\|\boldsymbol{u}(\phi-\phi_h)\|^2_{0,T}$, $\|\phi_h(\boldsymbol{u}-\boldsymbol{u}_h)\|^2_{0,T}$ and $\|(\vartheta(\phi)-\vartheta(\phi_h))\nabla\phi\|^2_{0,T}$, appearing in the upper bounds provided by the last four lemmas, which is indeed the purpose of the following lemma.

**Lemma 4.20.** *There exist positive constants $\widetilde{c}_7, \widetilde{c}_8$, independent of $h$, such that*

$$\sum_{T\in\mathcal{T}_h}\|\boldsymbol{u}\,(\phi-\phi_h)\|^2_{0,T}\leq\widetilde{c}_7\,\|\phi-\phi_h\|^2_{1,\Omega}\quad\text{and}\quad\sum_{T\in\mathcal{T}_h}\|\phi_h(\boldsymbol{u}-\boldsymbol{u}_h)\|^2_{0,T}\leq\widetilde{c}_8\,\|\boldsymbol{u}-\boldsymbol{u}_h\|^2_{1,\Omega},\tag{4.55}$$

*where $\widetilde{c}_7$ depends on $\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}}$, $\|\boldsymbol{f}\|_{\infty,\Omega}$ and $r$ (cf. (4.9)), and $\widetilde{c}_8$ depends on $r$. In addition, assuming $\nabla\phi\in\mathrm{L}^4(\Omega)$, there exists a positive constant $\widetilde{c}_9$, independent of $h$, such that*

$$\sum_{T\in\mathcal{T}_h}\|(\vartheta(\phi)-\vartheta(\phi_h))\nabla\phi\|^2_{0,T}\leq\widetilde{c}_9\,\|\phi-\phi_h\|^2_{1,\Omega}\,,\tag{4.56}$$

*where $\widetilde{c}_9$ depends on $L_\vartheta$ (cf. (4.4)) and $\|\nabla\phi\|_{\mathrm{L}^4(\Omega)}$.*

*Proof.* The estimates given in (4.55) were established in [7, eq. (3.71)-(3.72)]. On the other hand, using the Lipschitz continuity assumption on $\vartheta$ (*cf.* (4.4)), but restricted to each triangle $T\in\mathcal{T}_h$ instead of $\Omega$, employing Cauchy-Schwarz's inequality and the compactness (and hence continuity) of the injection $i:\mathrm{H}^1(\Omega)\to\mathrm{L}^4(\Omega)$ (*cf.* [1, Theorem 6.3], [92, Theorem 1.3.5]), we deduce that

$$\begin{aligned}\sum_{T\in\mathcal{T}_h}\|(\vartheta(\phi)-\vartheta(\phi_h))\nabla\phi\|^2_{0,T}&\leq\sum_{T\in\mathcal{T}_h}L^2_\vartheta\,\|\phi-\phi_h\|^2_{\mathrm{L}^4(T)}\,\|\nabla\phi\|^2_{\mathrm{L}^4(T)}\\&\leq\quad L^2_\vartheta\,\|\phi-\phi_h\|^2_{\mathrm{L}^4(\Omega)}\,\|\nabla\phi\|^2_{\mathrm{L}^4(\Omega)}\leq\widetilde{c}_9\,\|\phi-\phi_h\|^2_{1,\Omega}\,,\end{aligned}$$

where $\widetilde{c}_9$ depends only on $\|i\|$, $L_\vartheta$, and $\|\nabla\phi\|_{\mathrm{L}^4(\Omega)}$, which gives (4.56) and finishes the proof. □

By virtue of the estimates (4.46), (4.47), (4.48), Lemmas 2.14, 4.15, 4.17, 4.18, and 4.19, and the final estimates given by (4.55) and (4.56), we deduce (4.44). Finally, assuming now that $\boldsymbol{\sigma}\in\mathbb{L}^4(\Omega)$ and proceeding as at the end of the proof of [7, Theorem 3.13], we find that

$$\left\|\frac{1}{\mu(\phi)}\boldsymbol{\sigma}^{\mathtt{d}}-\frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\|_{0,\Omega}\leq C\left\{\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{0,\Omega}+\|\phi-\phi_h\|_{1,\Omega}\right\},\tag{4.57}$$

where $C$ is a positive constant, independent of $h$, that depends only on $\mu_1$ (*cf.* (4.3)), $L_\mu$ (*cf.* (4.4)) and $\|\boldsymbol{\sigma}\|_{\mathbb{L}^4(\Omega)}$. In this way, combining (4.57) and (4.44), we arrive at (4.45), which completes the proof of Theorem 4.3.

## 4.4 A second a posteriori error estimator

In this section we derive another *a posteriori* error estimator for our augmented mixed-primal finite element scheme (4.8), with the same discrete spaces introduced in the Section 4.2.3. In turn, the reliability of our new estimator can be proved without resorting to Helmholtz decompositions. More precisely, this second estimator arises simply employing the alternative definition of the functional $E_h$

(*cf.* (4.19)) and bounding $\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}$ in the preliminary estimate for the total error given by (4.32) (*cf.* Theorem 4.2). Then, with the same notations and discrete spaces introduced in Sections 4.2 and 4.3, we now set for each $T \in \mathcal{T}_h$ the local error indicator

$$\begin{aligned}\widetilde{\theta}_T^2 \,:=\, \|\boldsymbol{f}\phi_h \,-\, (\boldsymbol{K}^{-1}\boldsymbol{u}_h \,-\, \mathbf{div}\boldsymbol{\sigma}_h)\|_{0,T}^2 \,+\, \left\|\nabla \boldsymbol{u}_h \,-\, \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\|_{0,T}^2 \,+\, h_T^2\,\|g \,-\, (\beta\phi_h \,-\, \mathrm{div}\widetilde{\boldsymbol{\sigma}}_h)\|_{0,T}^2 \\ +\, \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Omega)} h_e\,\|[\![\widetilde{\boldsymbol{\sigma}}_h\cdot\boldsymbol{\nu}_e]\!]\|_{0,e}^2 \,+\, \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{N}})} h_e\,\|\widetilde{\boldsymbol{\sigma}}_h\cdot\boldsymbol{\nu}\|_{0,e}^2 \,+\, \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{D}})} \|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|_{0,e}^2\,,\end{aligned} \tag{4.58}$$

and define the following global residual error estimator

$$\widetilde{\boldsymbol{\theta}}^{\,2} \,:=\, \sum_{T\in\mathcal{T}_h} \widetilde{\theta}_T^2 \,+\, \|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|_{1/2,\Gamma_{\mathrm{D}}}^2\,. \tag{4.59}$$

Throughout the rest of this section, we establish *quasi-local* reliability and efficiency for the estimator $\widetilde{\boldsymbol{\theta}}$. The name *quasi-local* refers here to the fact that the last term defining $\widetilde{\boldsymbol{\theta}}$ can not be decomposed into local quantities associated to each triangle $T \in \mathcal{T}_h$ (unless it is either conveniently bounded or previously modified, as we will see below).

**Theorem 4.4.** *Assume that the data $\boldsymbol{k}$, $g$, $\vartheta$, $\boldsymbol{u}_{\mathrm{D}}$, and $\boldsymbol{f}$ are sufficiently small so that there holds*

$$C_4\,|\boldsymbol{k}| \,+\, C_5\,\|g\|_{\delta,\Omega} \,+\, C_6\,\vartheta_2 \,+\, C_7\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} \,+\, C_8\,\|\boldsymbol{f}\|_{\infty,\Omega} \,<\, \frac{1}{2},$$

*where $C_4$, $C_5$, $C_6$, $C_7$ and $C_8$ are the constants given in (4.20). Then, there exists a constant $\widetilde{C}_{\mathtt{rel}} > 0$, which depends only on $\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}}$, $\|\boldsymbol{f}\|_{\infty,\Omega}$ and other constants, all them independent of $h$, such that*

$$\|\phi - \phi_h\|_{1,\Omega}^2 \,+\, \|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H^2 \,\leq\, \widetilde{C}_{\mathtt{rel}}\,\widetilde{\boldsymbol{\theta}}^{\,2}\,. \tag{4.60}$$

*Proof.* Using the alternative definition of the functional $E_h$ (*cf.* (4.19)), and then applying the Cauchy-Schwarz inequality, we deduce that

$$\begin{aligned}\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'} \,\leq\, C\,\Big\{\|\boldsymbol{u}_{\mathrm{D}} \,-\, \boldsymbol{u}_h\|_{1/2,\Gamma_{\mathrm{D}}} \,+\, \left\|\nabla\boldsymbol{u}_h \,-\, \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\|_{0,\Omega} \\ +\, \|\boldsymbol{f}\phi_h \,-\, (\boldsymbol{K}^{-1}\boldsymbol{u}_h \,-\, \mathbf{div}\boldsymbol{\sigma}_h)\|_{0,\Omega}\Big\}\,,\end{aligned} \tag{4.61}$$

where $C$ is a positive constant independent of $h$. Then, replacing (4.61) back into (4.32) (*cf.* Theorem 4.2), and employing again the upper bound for $\|\widetilde{E}_h\|_{\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'}$ (*cf.* Lemma 4.4), and the definition of $\boldsymbol{\theta}_0$ (*cf.* Lemma 4.1), we obtain (4.60) and finish the proof. □

**Theorem 4.5.** *Assume that $\nabla\phi \in \mathrm{L}^4(\Omega)$. Then, there exists a constant $C^*_{\mathtt{eff}} > 0$, which depends only on parameters, $\|\boldsymbol{K}^{-1}\|_\infty$, $|\boldsymbol{k}|$, $\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}}$, $\|\boldsymbol{f}\|_{\infty,\Omega}$, $\|\nabla\phi\|_{\mathrm{L}^4(\Omega)}$, and other constants, all them independent of $h$, such that*

$$C^*_{\mathtt{eff}}\,\widetilde{\boldsymbol{\theta}}^{\,2} \,\leq\, \|\phi-\phi_h\|_{1,\Omega}^2 + \|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega}^2 + \|\mathbf{div}(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_{0,\Omega}^2 + \left\|\frac{1}{\mu(\phi)}\boldsymbol{\sigma}^{\mathtt{d}} - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\|_{0,\Omega}^2 + \text{h.o.t.} \tag{4.62}$$

*where* h.o.t. *stands for one or several terms of higher order. Moreover, assuming $\boldsymbol{\sigma} \in \mathbb{L}^4(\Omega)$, there exists a constant $\widetilde{C}_{\mathtt{eff}} > 0$, which depends only on parameters, $\|\boldsymbol{K}^{-1}\|_\infty$, $|\boldsymbol{k}|$, $\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2,\Gamma_{\mathrm{D}}}$, $\|\boldsymbol{f}\|_{\infty,\Omega}$, $\|\nabla\phi\|_{\mathrm{L}^4(\Omega)}$, $\|\boldsymbol{\sigma}\|_{\mathbb{L}^4(\Omega)}$, and other constants, all them independent of $h$, such that*

$$\widetilde{C}_{\mathtt{eff}}\,\widetilde{\boldsymbol{\theta}}^{\,2} \,\leq\, \|\phi - \phi_h\|_{1,\Omega}^2 \,+\, \|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|_H^2 \,+\, \text{h.o.t.} \tag{4.63}$$

*Proof.* We proceed as in the proof of [7, Theorem 4.2]. In fact, thanks to the trace theorem in $\mathbf{H}^1(\Omega)$, there exists $c > 0$, depending on $\Gamma_{\mathrm{D}}$ and $\Omega$, such that

$$\|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|^2_{1/2,\Gamma_{\mathrm{D}}} \leq c\,\|\boldsymbol{u} - \boldsymbol{u}_h\|^2_{1,\Omega}\,.$$

Next, the estimates (4.62) and (4.63) are obtained by applying the same arguments employed in the proof of Theorem 4.3 (*cf.* Section 4.3.3), and hence we omit further details. □

We point out here that, in order to use the indicator $\widetilde{\boldsymbol{\theta}}$ (*cf.* (4.59)) in an adaptive algorithm that solves (4.8), we need to estimate the expression $\|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|^2_{1/2,\Gamma_{\mathrm{D}}}$ through local terms. To this end, as well as in [7], we now employ an interpolation argument and replace the aforementioned expression by a suitable upper bound, which yields a reliable and fully local *a posteriori* error estimate.

**Theorem 4.6.** *Assume that the data $\boldsymbol{k}$, $g$, $\vartheta$, $\boldsymbol{u}_{\mathrm{D}}$, and $\boldsymbol{f}$ are sufficiently small so that there holds*

$$C_4\,|\boldsymbol{k}| + C_5\,\|g\|_{\delta,\Omega} + C_6\,\vartheta_2 + C_7\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} + C_8\,\|\boldsymbol{f}\|_{\infty,\Omega} < \frac{1}{2},$$

*where $C_4$, $C_5$, $C_6$, $C_7$ and $C_8$ are the constants given in (4.20) In turn, let $\widehat{\boldsymbol{\theta}}^2 := \sum_{T\in\mathcal{T}_h} \widehat{\theta}^2_T$, where for each $T \in \mathcal{T}_h$ we set*

$$\widehat{\theta}^2_T := \|\boldsymbol{f}\phi_h - (\boldsymbol{K}^{-1}\boldsymbol{u}_h - \mathbf{div}\boldsymbol{\sigma}_h)\|^2_{0,T} + \left\|\nabla\boldsymbol{u}_h - \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}^{\mathtt{d}}_h\right\|^2_{0,T} + h^2_T\,\|g - (\beta\phi_h - \mathrm{div}\widetilde{\boldsymbol{\sigma}}_h)\|^2_{0,T}$$
$$+ \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Omega)} h_e\,\|[\![\widetilde{\boldsymbol{\sigma}}_h\cdot\boldsymbol{\nu}_e]\!]\|^2_{0,e} + \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{N}})} h_e\,\|\widetilde{\boldsymbol{\sigma}}_h\cdot\boldsymbol{\nu}\|^2_{0,e} + \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{D}})} \|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|^2_{1,e}\,.$$

*Then, there exists a constant $\widehat{C}_{\mathtt{rel}} > 0$, which depends only on parameters, $\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}}$, $\|\boldsymbol{f}\|_{\infty,\Omega}$ and other constants, all them independent of $h$, such that*

$$\|\phi - \phi_h\|^2_{1,\Omega} + \|(\boldsymbol{\sigma},\boldsymbol{u}) - (\boldsymbol{\sigma}_h,\boldsymbol{u}_h)\|^2_H \leq \widehat{C}_{\mathtt{rel}}\,\widehat{\boldsymbol{\theta}}^2\,.$$

*Proof.* The proof reduces to bound the term $\|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|_{1/2,\Gamma_{\mathrm{D}}}$. To this end, it suffices to apply the fact that $H^{1/2}(\Gamma_{\mathrm{D}})$ is the interpolation space with index $1/2$ between $H^1(\Gamma_{\mathrm{D}})$ and $L^2(\Gamma_{\mathrm{D}})$, and proceed as in [7, Theorem 4.3]. □

## 4.5 Residual-based a posteriori error estimators: The 3D case

In this section we extend the results from Sections 4.3 and 4.4 to the three-dimensional version of (4.8). Analogously, as in Section 4.3, given a tetrahedron $T \in \mathcal{T}_h$, we let $\mathcal{E}_h(T)$ be the set of its faces, and let $\mathcal{E}_h$ be the set of all faces of the triangulation $\mathcal{T}_h$. Then, we write $\mathcal{E}_h = \mathcal{E}_h(\Omega) \cup \mathcal{E}_h(\Gamma)$, where $\mathcal{E}_h(\Omega) := \{e \in \mathcal{E}_h : e \subseteq \Omega\}$ and $\mathcal{E}_h(\Gamma) := \{e \in \mathcal{E}_h : e \subseteq \Gamma\}$. Also, for each face $e \in \mathcal{E}_h$ we fix a unit normal $\boldsymbol{\nu}_e$ to $e$, so that given $\boldsymbol{\tau} \in \mathbb{L}^2(\Omega)$ such that $\boldsymbol{\tau}|_T \in \mathbb{C}(T)$ on each $T \in \mathcal{T}_h$, and given $e \in \mathcal{E}_h(\Omega)$, we let $[\![\boldsymbol{\tau}\times\boldsymbol{\nu}_e]\!]$ be the corresponding jump of the tangential traces across $e$, that is $[\![\boldsymbol{\tau}\times\boldsymbol{\nu}_e]\!] := (\boldsymbol{\tau}|_T - \boldsymbol{\tau}|_{T'})|_e \times \boldsymbol{\nu}_e$, where $T$ and $T'$ are the elements of $\mathcal{T}_h$ having $e$ as a common face. In what follows, when no confusion arises, we simple write $\boldsymbol{\nu}$ instead of $\boldsymbol{\nu}_e$.

Now, we recall that the curl of a 3D vector $\boldsymbol{v} := (v_1, v_2, v_3)$ is the 3D vector

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

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

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

We now set for each $T \in \mathcal{T}_h$ the local *a posteriori* error indicator $\theta_T^2$ as follows

$$\begin{aligned}
\theta_T^2 \,:=\, & \|\boldsymbol{f}\phi_h \,-\, (\boldsymbol{K}^{-1}\boldsymbol{u}_h \,-\, \mathbf{div}\boldsymbol{\sigma}_h)\|_{0,T}^2 \,+\, \left\|\nabla \boldsymbol{u}_h \,-\, \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\|_{0,T}^2 \,+\, h_T^2\, \|g \,-\, (\beta\phi_h \,-\, \mathrm{div}\widetilde{\boldsymbol{\sigma}}_h)\|_{0,T}^2 \\
& +\, h_T^2 \left\|\mathbf{curl}\left\{\frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}}\right\}\right\|_{0,T}^2 \,+\, \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Omega)} h_e \left\|\left[\!\left[\frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}} \times \boldsymbol{\nu}\right]\!\right]\right\|_{0,e}^2 \\
& +\, \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Omega)} h_e\, \|[\![\widetilde{\boldsymbol{\sigma}}_h \cdot \boldsymbol{\nu}_e]\!]\|_{0,e}^2 + \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{N}})} h_e\, \|\widetilde{\boldsymbol{\sigma}}_h \cdot \boldsymbol{\nu}\|_{0,e}^2 \\
& +\, \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{D}})} \|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|_{0,e}^2 + \sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\mathrm{D}})} h_e \left\|\nabla \boldsymbol{u}_{\mathrm{D}} \times \boldsymbol{\nu} \,-\, \frac{1}{\mu(\phi_h)}\boldsymbol{\sigma}_h^{\mathtt{d}} \times \boldsymbol{\nu}\right\|_{0,e}^2,
\end{aligned} \tag{4.64}$$

whereas $\widetilde{\theta}_T^2$ stays exactly as in (4.58). In this way, the corresponding global *a posteriori* error estimators are defined as (4.11) and (4.59), that is

$$\boldsymbol{\theta}^2 \,:=\, \sum_{T\in\mathcal{T}_h} \theta_T^2 \quad \text{and} \quad \widetilde{\boldsymbol{\theta}}^2 \,:=\, \sum_{T\in\mathcal{T}_h} \widetilde{\theta}_T^2 \,+\, \|\boldsymbol{u}_{\mathrm{D}} - \boldsymbol{u}_h\|_{1/2,\Gamma_{\mathrm{D}}}^2.$$

We now establish the analogue of Theorems 4.1 and 4.4, respectively.

**Theorem 4.7** (Reliability of $\boldsymbol{\theta}$). *Assume that $\Omega$ is a connected domain and that $\Gamma_{\mathrm{N}}$ is the boundary of a convex extension of $\Omega$. In addition, assume that the data $\boldsymbol{k}$, $g$, $\vartheta$, $\boldsymbol{u}_{\mathrm{D}}$, and $\boldsymbol{f}$ are sufficiently small so that there holds*

$$C_4\,|\boldsymbol{k}| \,+\, C_5\,\|g\|_{\delta,\Omega} \,+\, C_6\,\vartheta_2 \,+\, C_7\,\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} \,+\, C_8\,\|\boldsymbol{f}\|_{\infty,\Omega} \,<\, \frac{1}{2},$$

*where $C_4$, $C_5$, $C_6$, $C_7$ and $C_8$ are the constants given below in (4.20). Then, there exists a constant $C_{\mathtt{rel}} > 0$, which depends only on parameters, $\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}}$, $\|\boldsymbol{f}\|_{\infty,\Omega}$, and other constants, all them independent of $h$, such that*

$$\|\phi - \phi_h\|_{1,\Omega} \,+\, \|(\boldsymbol{\sigma}, \boldsymbol{u}) - (\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\|_H \,\leq\, C_{\mathtt{rel}}\,\boldsymbol{\theta}.$$

The proof of Theorem 4.7 follows from a very similar analysis to the Section 4.3.2, except in a few points to be described throughout the following discussion. Indeed, we first need to use a 3D version of the stable Helmholtz decomposition, provided by Lemma 4.5, which was established recently in

[64, Theorem 3.2]. We remark that the proof of [64, Theorem 3.2] makes use of several estimates available in [10] and combines similar arguments to those from the proofs of [64, Theorem 3.1] and [7, Lemma 3.9]. Then, the associated discrete Helmholtz decomposition and the functional $E_h$ are set and rewritten exactly as in (4.38) and (4.39), respectively. Secondly, in order to derive the upper bound to $\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}$, we need to employ the 3D analogue of the integration by parts formula on the boundary given by (4.40) (*cf.* Lemma 4.7). In fact, by employing the identities from [77, Chapter I, eq. (2.17) and Theorem 2.11], we find that in the 3D case, there holds

$$\langle \underline{\mathbf{curl}}\, \boldsymbol{\chi}\, \boldsymbol{\nu}, \boldsymbol{\phi} \rangle \;=\; -\langle \nabla \boldsymbol{\phi} \times \boldsymbol{\nu}, \boldsymbol{\chi} \rangle \qquad \forall\, \boldsymbol{\chi} \in \mathbb{H}^1(\Omega), \quad \forall\, \boldsymbol{\phi} \in \mathbf{H}^1(\Omega). \tag{4.65}$$

On the other hand, the integration by parts formula on each tetrahedron $T \in \mathcal{T}_h$, which is employed in the proof of the 3D analogue of Lemma 4.8 (see also [7, Lemma 3.11]), becomes (*cf.* [77, Chapter I, Theorem 2.11])

$$\int_T \underline{\mathbf{curl}}\, \boldsymbol{\tau} : \boldsymbol{\chi} \;-\; \int_T \boldsymbol{\tau} : \underline{\mathbf{curl}}\, \boldsymbol{\chi} \;=\; \langle \boldsymbol{\tau} \times \boldsymbol{\nu}, \boldsymbol{\chi} \rangle_{\partial T} \qquad \forall\, \boldsymbol{\tau} \in \mathbb{H}(\underline{\mathbf{curl}}; \Omega), \qquad \forall\, \boldsymbol{\chi} \in \mathbb{H}^1(\Omega), \tag{4.66}$$

where $\langle \cdot, \cdot \rangle_{\partial T}$ is the duality pairing between $\mathbb{H}^{-1/2}(T)$ and $\mathbb{H}^{1/2}(T)$, and, as usual, $\mathbb{H}(\underline{\mathbf{curl}}; \Omega)$ corresponds to the space of tensors in $\mathbb{L}^2(\Omega)$ whose $\underline{\mathbf{curl}}$ belongs to $\mathbb{L}^2(\Omega)$. Notice that the identities (4.65) and (4.66) explain the appearing of the expressions $\dfrac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \times \boldsymbol{\nu}$ and $\nabla \boldsymbol{u}_{\mathrm{D}} \times \boldsymbol{\nu} - \dfrac{1}{\mu(\phi_h)} \boldsymbol{\sigma}_h^{\mathtt{d}} \times \boldsymbol{\nu}$ in the 3D definition of $\boldsymbol{\theta}$ (*cf.* (4.64)).

**Theorem 4.8** (Reliability of $\widetilde{\boldsymbol{\theta}}$)**.** *Assume that the data $\boldsymbol{k}$, $g$, $\vartheta$, $\boldsymbol{u}_{\mathrm{D}}$, and $\boldsymbol{f}$ are sufficiently small so that there holds*

$$C_4\, |\boldsymbol{k}| \;+\; C_5\, \|g\|_{\delta,\Omega} \;+\; C_6\, \vartheta_2 \;+\; C_7\, \|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}} \;+\; C_8\, \|\boldsymbol{f}\|_{\infty,\Omega} \;<\; \frac{1}{2},$$

*where $C_4$, $C_5$, $C_6$, $C_7$ and $C_8$ are the constants given in (4.20). Then, there exists a constant $\widetilde{C}_{\mathtt{rel}} > 0$, which depends only on $\|\boldsymbol{u}_{\mathrm{D}}\|_{1/2+\delta,\Gamma_{\mathrm{D}}}$, $\|\boldsymbol{f}\|_{\infty,\Omega}$ and other constants, all them independent of $h$, such that*

$$\|\phi - \phi_h\|_{1,\Omega}^2 \;+\; \|(\boldsymbol{\sigma}, \boldsymbol{u}) - (\boldsymbol{\sigma}_h, \boldsymbol{u}_h)\|_H^2 \;\leq\; \widetilde{C}_{\mathtt{rel}}\, \widetilde{\boldsymbol{\theta}}^2 \,.$$

The proof of Theorem 4.8 proceeds similarly as in the proof of Theorem 4.4. In fact, notice that the upper bounds for $\|\widetilde{E}_h\|_{\mathrm{H}^1_{\Gamma_{\mathrm{D}}}(\Omega)'}$ and $\|E_h\|_{\mathbb{H}_{\mathrm{N}}(\mathbf{div},\Omega)'}$ given by (4.34) (*cf.* Lemma 4.4) and (4.61) (*cf.* Theorem 4.4), respectively, are also valid in 3D and hence the corresponding reliability of $\widetilde{\boldsymbol{\theta}}$ is obtained.

We end this section by remarking that the efficiency of the estimators $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}}$ follows as in Section 4.3.3 and Theorem 4.5, respectively. In particular, we remark that the 3D version of estimates provided in Lemmas 4.13, 4.14, and 4.15 can be derived from [67, Lemmas 4.9, 4.10, 4.11 and 4.13].

## 4.6 Numerical tests

This section serves to illustrate the properties of the estimators introduced in Sections 4.3-4.5. Fixed point iterations were used for the linearization of the coupled mixed-primal scheme, and a residual tolerance of 1e-7 was prescribed for the termination of the Picard algorithm. All linear solves are performed with the unsymmetric multifrontal direct solver UMFPACK. In addition, all tests in this Section use a classical adaptive mesh refinement procedure based on the equi-distribution of the

error indicators, where the diameter of each element in the new adapted mesh (contained in a generic element $K$ on the initial coarse mesh) is proportional to the diameter of the initial element times the ratio $\frac{\bar{\eta}_h}{\eta_K}$, where $\bar{\eta}_h$ is the mean value of a given indicator $\eta$ over the initial mesh (*cf.* [102]).

**Example 1: accuracy assessment.** Our first example focuses on the case where, under uniform mesh refinement, the convergence rates are affected by the singularities of the exact solutions. A non-convex domain $\Omega := (0,1)^2\backslash[0,\frac{1}{2}]^2$ is considered, and its boundary $\partial\Omega$ is split into $\Gamma_{\rm N} := (1,0)\times\{0\}$ and $\Gamma_{\rm D} := \partial\Omega\backslash\Gamma_{\rm N}$. We construct a sequence of nested unstructured triangulations, where measured errors and experimental convergence rates will be computed as usual

$$e(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_{\mathbf{div},\Omega},\ e(\phi) = \|\phi-\phi_h\|_{1,\Omega},\ e(\boldsymbol{u}) = \|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\Omega},\ r(\cdot) = -2\log(e(\cdot)/\hat{e}(\cdot))[\log(N/\hat{N})]^{-1},$$

with $e$ and $\hat{e}$ denoting errors produced on two consecutive meshes representing $N$ and $\hat{N}$ degrees of freedom, respectively. In addition, the total error, the modified error suggested by (4.44) and (4.62), and the effectivity and quasi-effectivity indexes associated to a given global estimator $\boldsymbol{\eta}$ are defined, respectively, as

$$\mathtt{e} = \left\{[e(\boldsymbol{\sigma})]^2+[e(\boldsymbol{u})]^2+[e(\phi)]^2\right\}^{1/2},\qquad \mathtt{eff}(\boldsymbol{\eta}) = \frac{\mathtt{e}}{\boldsymbol{\eta}},$$

$$\mathtt{m} = \left\{[e(\boldsymbol{u})]^2+[e(\phi)]^2+\|\mathbf{div}\boldsymbol{\sigma}-\mathbf{div}\boldsymbol{\sigma}_h\|_{0,\Omega}^2+\left\|\frac{\boldsymbol{\sigma}^{\mathtt{d}}}{\mu(\phi)}-\frac{\boldsymbol{\sigma}_h^{\mathtt{d}}}{\mu(\phi_h)}\right\|_{0,\Omega}^2\right\}^{1/2},\qquad \mathtt{qeff}(\boldsymbol{\eta}) = \frac{\mathtt{m}}{\boldsymbol{\eta}}.$$

An exact solution to (1.5) is given as follows

$$\begin{aligned}&\phi(x_1,x_2) = mx_1x_2(1-x_2)(x_1-1/2)^2(x_2-1/2)^2+b,\\ &\boldsymbol{u}(x_1,x_2) = \begin{pmatrix}\sin(\pi x_1)\cos(\pi x_2)\\ -\cos(\pi x_1)\sin(\pi x_2)\end{pmatrix},\quad \boldsymbol{\sigma}(x_1,x_2) = \mu(\phi)\boldsymbol{u}-\left[\mu(\phi)\frac{\partial u_1}{\partial x_1}+\frac{(x_1-1)^2}{(x_2-a_1)(x_2-a_2)}\right]\mathbb{I},\end{aligned}\tag{4.67}$$

where $\boldsymbol{K}^{-1} = K^{-1}\mathbb{I}$, $\boldsymbol{k} = (0,-1)^{\mathtt{t}}$, $\mu(\phi) = (1-a\phi)^{-2}$, $f_{\rm bk}(\phi) = a\phi(1-a\phi)^2$, $\vartheta(\phi) = \phi+(1-a\phi)^2$, and the source terms are

$$\boldsymbol{f}(x_1,x_2) = \phi^{-1}(K^{-1}\boldsymbol{u}-\mathbf{div}\boldsymbol{\sigma}),\quad g(x_1,x_2) = \beta\phi-\mathrm{div}(\vartheta(\phi)\nabla\phi)+\boldsymbol{u}\cdot\nabla\phi+f_{\rm bk}'(\phi)\boldsymbol{k}\cdot\nabla\phi.$$

Notice that the only difference with respect to (4.1) is a non-homogeneous concentration flux $\widetilde{\boldsymbol{\sigma}}\cdot\boldsymbol{\nu} = s$ imposed on $\Gamma_{\rm N}$, where $s$ is manufactured according to (4.67). Therefore, the relevant term in the *a posteriori* error estimators will be replaced by

$$\sum_{e\in\mathcal{E}_h(T)\cap\mathcal{E}_h(\Gamma_{\rm N})} h_e\|\widetilde{\boldsymbol{\sigma}}_h\cdot\boldsymbol{\nu}-s\|_{0,e}^2,$$

whose estimation from below and above follows in a straightforward way. The model parameters specifying (4.67) correspond to $m = 200$, $b = 0.008$, $a = 0.35$, $K = 0.01$, $\beta = 0.35$, and $a_1 = -0.05$, $a_2 = 1.1$. Notice that the pressure defining the isotropic part of the stress in (4.67) exhibits a singularity near the upper right corner of the domain, at $(a_1,a_2)$ (see Fig. 4.3). As a consequence, optimal convergence for the stress is no longer evidenced under uniform mesh refinement (see first rows of Table 4.1). In turn, if an adaptive mesh refinement step (employing the residual error indicators $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}}$) is applied, optimal convergence can be restored, as shown in the last two blocks of Table 4.1. Approximated solutions obtained after six adaptation steps are collected in Figure 4.1, and a few adapted meshes produced using the two indicators are depicted in Figure 4.2. It is observed that the

| D.o.f. | h | $e(\boldsymbol{\sigma})$ | $r(\boldsymbol{\sigma})$ | $e(\boldsymbol{u})$ | $r(\boldsymbol{u})$ | $e(\phi)$ | $r(\phi)$ | $i_P$ | e | m | eff($\boldsymbol{\theta}$) | qeff($\boldsymbol{\theta}$) | eff($\widetilde{\boldsymbol{\theta}}$) | qeff($\widetilde{\boldsymbol{\theta}}$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Augmented $\mathbf{RT}_0 - \mathbf{P}_1 - \mathrm{P}_1$ scheme with quasi-uniform refinement | | | | | | | | | | | | | | |
| 105 | 0.53 | 140.05 | – | 3.40 | – | 1.43 | – | 12 | 140.10 | 139.42 | 1.11 | 1.10 | 1.10 | 1.10 |
| 192 | 0.49 | 225.45 | -6.44 | 8.81 | -12.90 | 1.35 | 0.80 | 16 | 225.63 | 222.53 | 0.55 | 0.55 | 0.55 | 0.55 |
| 492 | 0.30 | 139.34 | 0.97 | 3.36 | 1.95 | 0.87 | 0.88 | 14 | 139.38 | 138.77 | 1.07 | 1.07 | 1.07 | 1.07 |
| 1488 | 0.16 | 74.51 | 1.01 | 0.82 | 2.27 | 0.47 | 0.98 | 12 | 74.52 | 74.36 | 0.92 | 0.92 | 0.92 | 0.92 |
| 4902 | 0.09 | 44.05 | 0.88 | 0.28 | 1.77 | 0.26 | 0.98 | 13 | 44.05 | 43.98 | 0.96 | 0.96 | 0.96 | 0.95 |
| 17800 | 0.04 | 21.46 | 1.05 | 0.12 | 1.25 | 0.14 | 0.88 | 13 | 21.46 | 21.43 | 0.98 | 0.98 | 0.98 | 0.98 |
| 67800 | 0.02 | 11.47 | 1.14 | 0.05 | 1.36 | 0.07 | 1.33 | 14 | 11.48 | 11.46 | 1.00 | 1.00 | 0.99 | 0.99 |
| Augmented $\mathbf{RT}_0 - \mathbf{P}_1 - \mathrm{P}_1$ scheme with adaptive refinement according to $\boldsymbol{\theta}$ | | | | | | | | | | | | | | |
| 64 | 0.61 | 203.22 | – | 8.93 | – | 1.40 | – | 16 | 203.42 | 197.77 | 0.49 | 0.48 | – | – |
| 171 | 0.35 | 138.78 | 0.77 | 3.07 | 2.17 | 1.18 | 0.34 | 16 | 138.82 | 138.09 | 1.24 | 1.23 | – | – |
| 303 | 0.33 | 99.71 | 1.15 | 1.98 | 1.52 | 1.10 | 0.26 | 14 | 99.74 | 99.37 | 0.95 | 0.94 | – | – |
| 499 | 0.30 | 68.36 | 1.51 | 1.27 | 1.78 | 1.09 | 0.03 | 13 | 68.38 | 68.07 | 1.05 | 1.05 | – | – |
| 1014 | 0.21 | 39.04 | 1.58 | 0.78 | 1.35 | 0.85 | 0.69 | 13 | 39.06 | 38.83 | 1.17 | 1.17 | – | – |
| 3763 | 0.10 | 16.51 | 1.31 | 0.26 | 1.66 | 0.37 | 1.26 | 13 | 16.52 | 16.45 | 1.07 | 1.07 | – | – |
| 14690 | 0.05 | 7.81 | 1.09 | 0.11 | 1.27 | 0.17 | 1.07 | 13 | 7.81 | 7.79 | 1.03 | 1.03 | – | – |
| Augmented $\mathbf{RT}_0 - \mathbf{P}_1 - \mathrm{P}_1$ scheme with adaptive refinement according to $\widetilde{\boldsymbol{\theta}}$ | | | | | | | | | | | | | | |
| 64 | 0.61 | 203.22 | – | 8.93 | – | 1.40 | – | 16 | 203.42 | 197.77 | – | – | 0.49 | 0.47 |
| 171 | 0.35 | 138.78 | 0.77 | 3.07 | 2.17 | 1.18 | 0.34 | 16 | 138.82 | 138.09 | – | – | 1.24 | 1.23 |
| 303 | 0.33 | 99.63 | 1.15 | 1.97 | 1.54 | 1.10 | 0.26 | 14 | 99.66 | 99.29 | – | – | 0.95 | 0.94 |
| 535 | 0.28 | 71.12 | 1.18 | 1.41 | 1.18 | 1.08 | 0.04 | 13 | 71.14 | 70.61 | – | – | 1.06 | 1.05 |
| 1145 | 0.21 | 37.86 | 1.65 | 0.60 | 2.24 | 0.80 | 0.79 | 13 | 37.88 | 37.64 | – | – | 1.08 | 1.08 |
| 4270 | 0.10 | 16.93 | 1.22 | 0.24 | 1.35 | 0.38 | 1.11 | 13 | 16.94 | 16.88 | – | – | 1.05 | 1.05 |
| 16790 | 0.05 | 8.28 | 1.04 | 0.10 | 1.25 | 0.18 | 1.10 | 13 | 8.28 | 8.26 | – | – | 1.02 | 1.01 |

Table 4.1: Test 1: convergence history, Picard iteration count, error e and quasi-error m, effectivity and quasi-effectivity indexes for the approximation of the coupled Brinkman-transport problem, under quasi-uniform, and adaptive refinement according to the indicators introduced in Sections 4.3 and 4.4.

agglomeration of points follows the regions of high concentration gradients occurring near $\Gamma_{\mathrm{N}}$, as well as the sharp pressure profile localized at $(a_1, a_2)$.

**Example 2: sedimentation below downward-facing inclined walls.** This test illustrates the properties of the second estimator (4.59) in a 2D setting, where we simulate the sedimentation of a mixture within an heterogeneous porous medium. The domain consists of an isosceles trapeze of height 3, maximal width 2.82, and walls having an angle of inclination of $4/9\pi$ with respect to the horizontal axis. The permeability of the medium is constant $K_0$, except for 20 randomly placed spots (consisting of disks with radii 2.5e-3) of much lower permeability $K_1$. Viscosity, hindering sedimentation, and

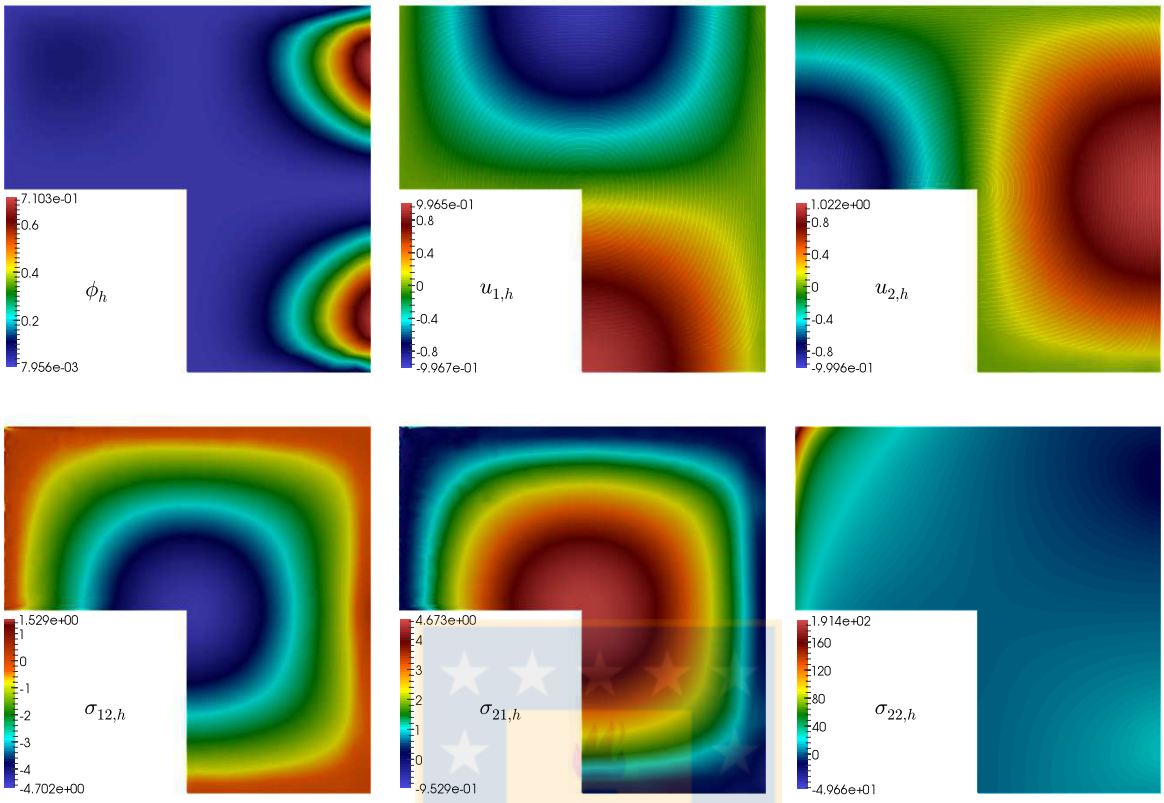


Figure 4.1: Test 1: approximate solutions obtained with the lowest order method, after six steps of adaptive mesh refinement following the second indicator $\widetilde{\boldsymbol{\theta}}$. Concentration, velocity components, and stress components are depicted.

compaction coefficients (all concentration-dependent) are respectively specified as

$$\mu(\phi) = \mu_0(1-\phi/\phi_{\max})^{-\eta_1}, \quad f_{\rm bk}(\phi) = u_\infty \phi(1-\phi/\phi_{\max})^{\eta_2}, \quad \vartheta(\phi) = \frac{\sigma_0 \alpha}{\phi_c^\alpha \Delta\rho G}\phi^{\alpha-2} f_{\rm bk} + u_\infty, \quad (4.68)$$

where the adimensional model parameters and remaining constants assume the values $\mu_0$ =2.5e-4, $\sigma_0$ = 5.5e-4, $G$ = 9.81, $\alpha$ = 5, $\beta$ = 0.25, $\eta$ = 2, $\phi_c$ = 0.07, $\phi_{\max}$ = 0.95, $K_0$ = 10, $K_1$ = 0.01, $\boldsymbol{k} = (0,-1)^T$, $\boldsymbol{f} = (0,-1/2)^T$, $u_\infty$ = 2.5e-3, $\Delta\rho$ = 1562. From the physical bounds of the concentration we find $\mu_1 = \mu_0$ and $\mu_2 = 5\mu_1$, yielding the following stabilization coefficients $\kappa_1 = 1/5\mu_0^2$ = 5e-5, $\alpha_{\boldsymbol{K}} = 0.1$, $\tilde{\delta}$ = 4.88e-3, $\kappa_2$ = 2.38e-6.

A pseudo time-advancing algorithm is employed to capture the transient nature of the phenomenon (this can be achieved by setting $g = \beta\phi^k$, where $\phi^k$ is the concentration distribution at the previous pseudo timestep). The initial guess for the concentration is a high concentration $\phi = 0.75$ on the top of the domain and a random perturbation of amplitude 0.05 around $\phi = 0.15$. We assume that the vessel is open on the top and closed elsewhere on $\partial\Omega$, so that a clear fluid $\phi = \phi_c$ and zero normal stresses $\boldsymbol{\sigma\nu} = \boldsymbol{0}$ are prescribed on top, whereas on the remainder of the boundary we set zero fluxes $\widetilde{\boldsymbol{\sigma}} \cdot \boldsymbol{\nu} = 0$ and no-slip conditions $\boldsymbol{u} = \boldsymbol{0}$.

The adaptive algorithm applies mesh refinement according to the second *a posteriori* error indicator (4.59), and it is invoked at the end of each pseudo-time step. We point out that due to the roughness of the permeability for coarser meshes, a continuation technique is applied on the viscosity scaling $\mu_0$

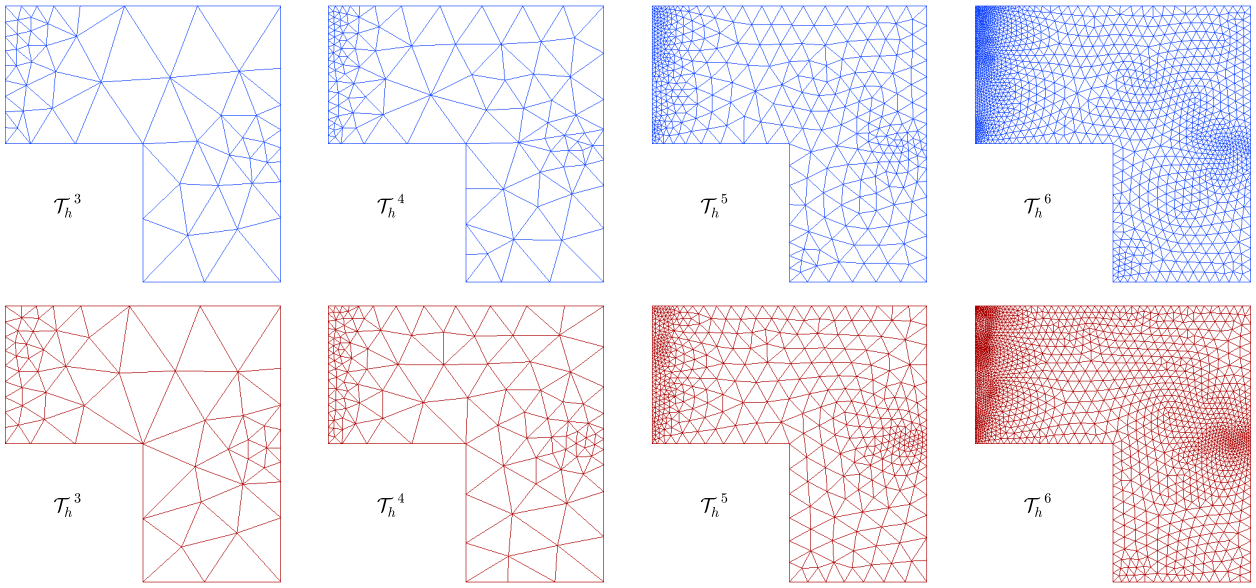


Figure 4.2: Test 1: from left to right, four snapshots of successively refined meshes according to the indicators $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}}$ (top and bottom panels, respectively).

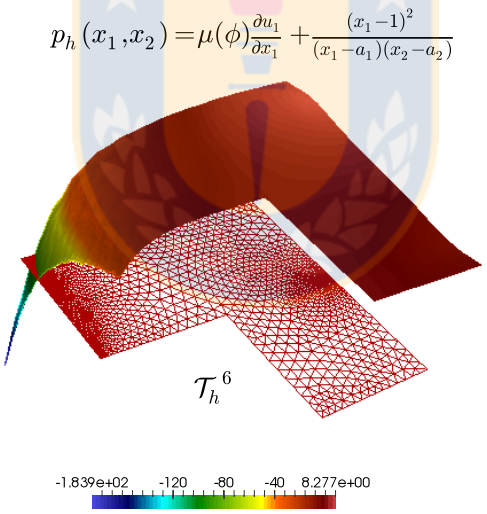


Figure 4.3: Test 1: approximate postprocessed pressure and adapted mesh near $(a_1, a_2)$, after six refinement steps following the second estimator $\widetilde{\boldsymbol{\theta}}$.

(using $\tilde{\mu}_0 = 8\mu_0$ as initial guess, and halving it until reaching $\mu_0$). A set of snapshots of the numerical solution obtained after ten pseudo-time steps are displayed in Figure 4.4. Apart from the main flow features expected in the pure-fluid case (acceleration of the deposition near the inclined walls, as discussed in [98] and simulated in [96]), we also observe tortuous concentration and velocity patterns produced by a combination of tight flow-transport coupling, the highly heterogeneous coefficients, and the random initial distribution. The velocity plots (second row) indicate that the flow tends to avoid the regions of low permeability, and recirculation zones are formed near the transition from clear to

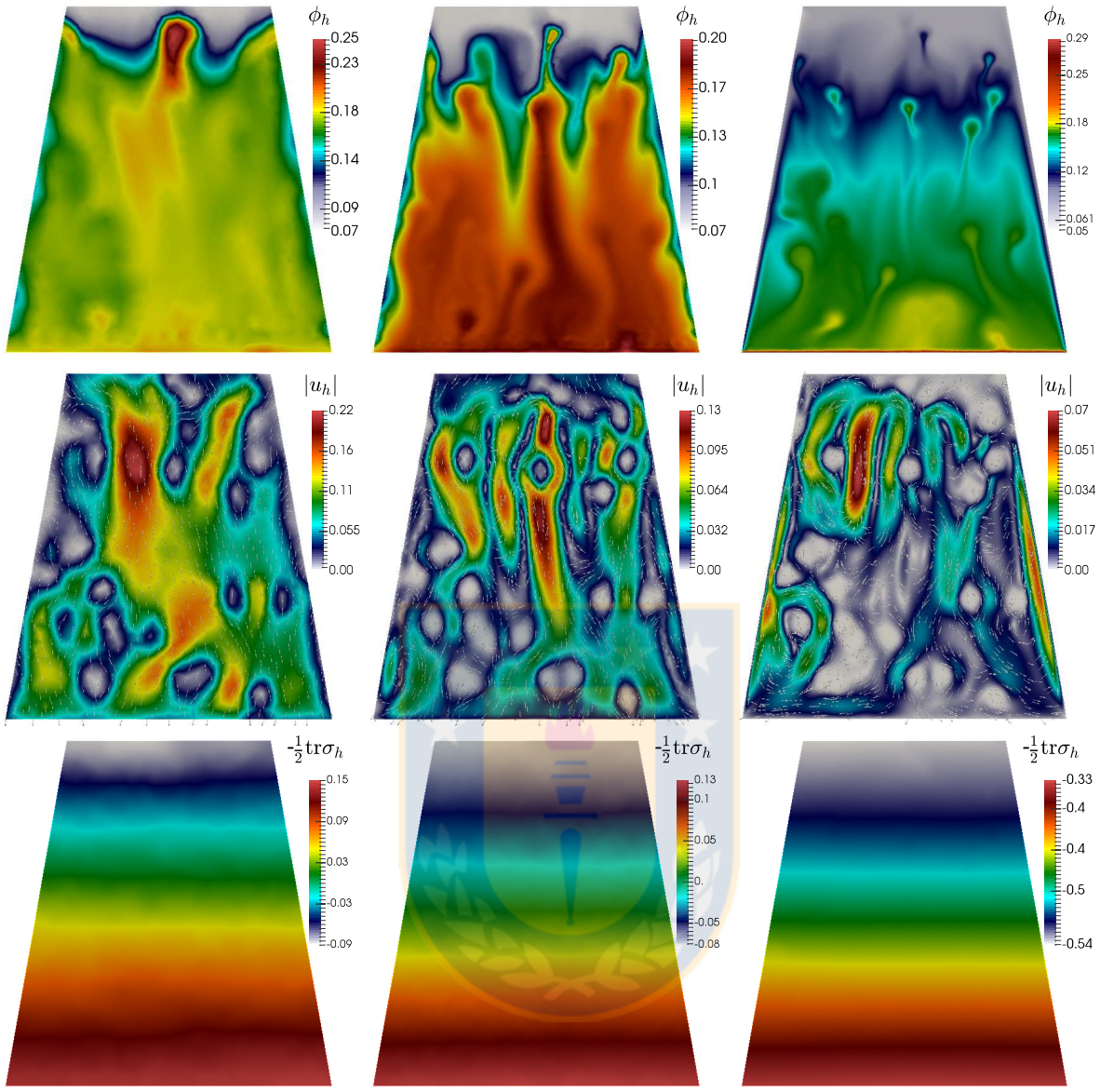


Figure 4.4: Test 2: approximate solutions at 3 (left), 6 (middle) and 12 (right) pseudo-time steps. A lowest order method and mesh adaptive refinement guided by (4.59) were used.

high-concentration mixture. In addition, the concentration plots (panels in the first row) suggest that solid particles remain attached to the low-permeability spots and reverse plumes are formed. We also show a sequence of refined meshes after two, four, and six steps (see Figure 4.5), where it is seen that the *a posteriori* error indicator yields more refinement near the high gradients of concentration and the aforementioned recirculation zones. As these irregularities spread throughout large portions of the domain, a substantial gain in computational cost with respect to a uniformly refined mesh is not expected.

**Example 3: sedimentation in a clarifier-thickener unit.** We close this section with a numerical test that illustrates the performance of the proposed numerical scheme and the first a posteriori error indicator (4.64) on a 3D computation. The example reproduces the steady-state of a sedimentation-

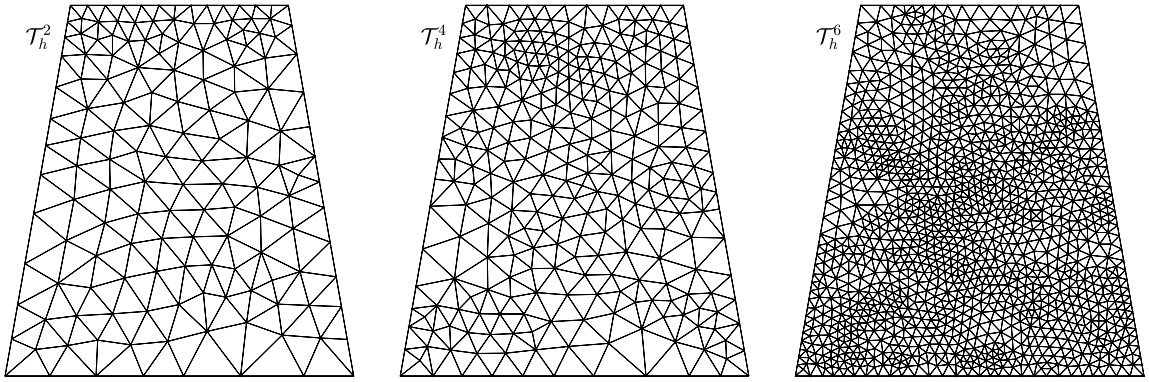


Figure 4.5: Test 2: adapted meshes at 2, 4, and 6 steps, generated following the second estimator (4.59).

consolidation process in a clarifier-thickener unit. Model parameters and domain configuration are adapted from those in [4, Example 3], but here the device has a radial length of 14.6 m and a total height of 7.6 m. It features a feed inlet $\Gamma_{\rm in}$ consisting of an horizontal disk of radius 1.5 m, an underflow outlet for the discharge of sediment $\Gamma_{\rm out}$ (an horizontal disk of smaller radius 0.5 m), and a peripheral overflow annular region $\Gamma_{\rm ofl}$ (see a sketch in the top left panel of Figure 4.6). A suspension is fed through $\Gamma_{\rm in}$ with velocity $\boldsymbol{u} = \boldsymbol{u}_{\rm in} = (0, 0, -u_{3,\rm in})^T$ and having a concentration of $\phi = \phi_{\rm in}$. At the outlet $\Gamma_{\rm out}$ we set $\boldsymbol{u} = \boldsymbol{u}_{\rm out} = (0, 0, -u_{3,\rm out})^T$; at the overflow annulus we impose zero normal pseudo-stresses; and on the remainder of $\partial\Omega$ we put no slip boundary data for the velocity and zero-flux conditions for the concentration.

The concentration-dependent coefficients are defined as in (4.68), and the remaining parameters are chosen as in Example 2, except for $u_{3,\rm in}$ =1.25e-2, $u_{3,\rm out}$ =1.25e-3, $\phi_c = 0.1$, $u_\infty$ =2.2e-3, $\phi_{\rm in} = 0.08$, $\sigma_0$ =5.5e-2, and $\beta$ =1e-3. Again, the bounds for the concentration imply that the stabilization parameters assume the following values $\kappa_1 = 0.256$, $\kappa_2 = 0.25$.

The proposed primal-mixed method is used to generate the approximate solutions depicted in Figure 4.6 (where we show only half of the domain, for visualization purposes). As in [4, 31], we can observe that the mixture concentrates near the outlet boundary $\Gamma_{\rm out}$. The velocity arrows show recirculating patterns, and a very small underflow. In contrast with Example 2, here the Picard iterations until convergence are embedded inside the adaptive refinement loop, consisting of solving, estimating, marking and refining using the error equi-distribution strategy mentioned above. The plots in Figure 4.7 show a sequence of three adapted meshes, forming a clustering of elements near the zones of high concentration gradients (connecting inflow and underflow boundaries), where also the velocity and postprocessed pressure profiles are more pronounced.

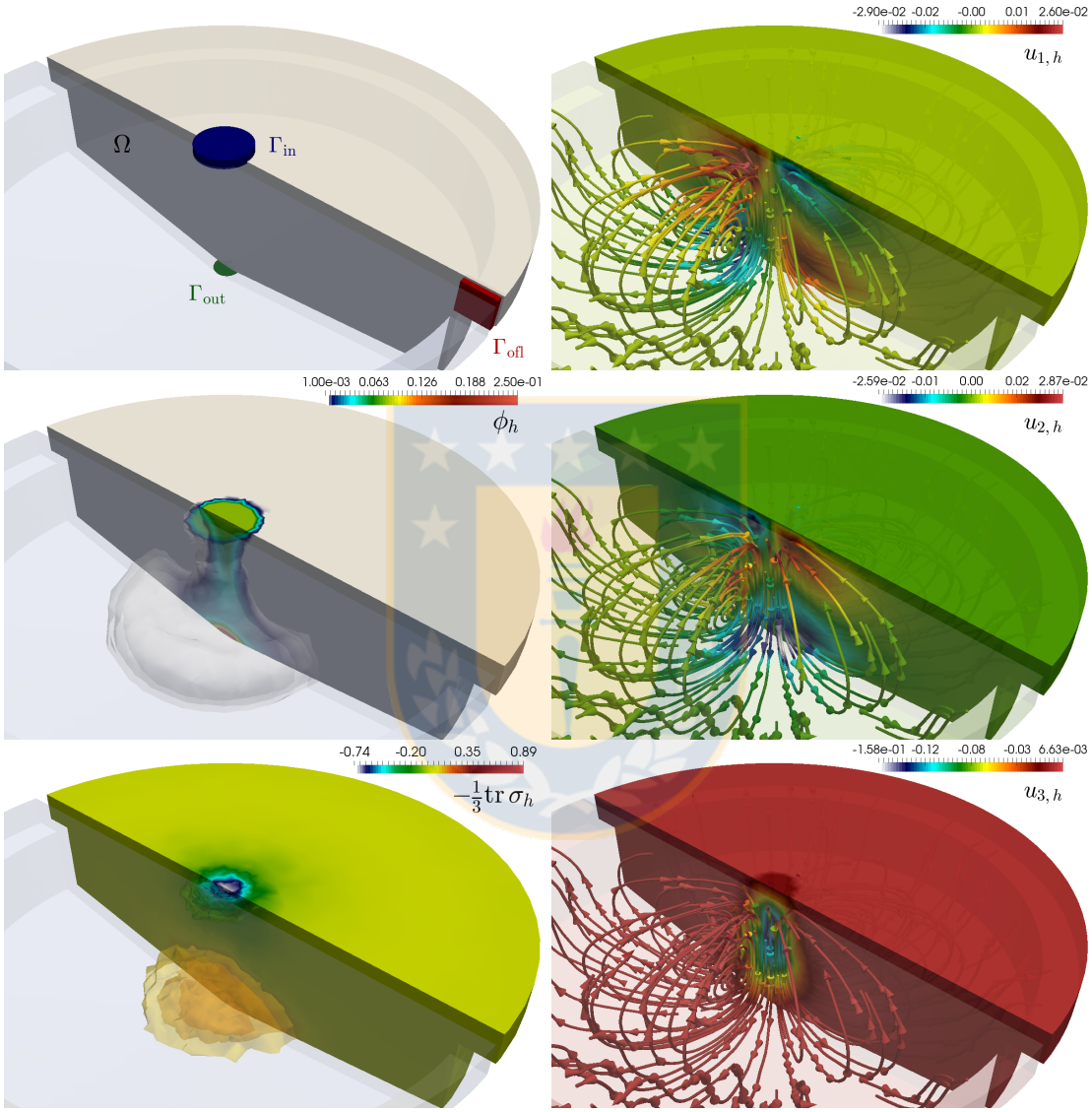


Figure 4.6: Test 3: sketch of the clipped domain and different boundaries in a clarifier-thickener device (top left panel), and snapshots of the approximate concentration, postprocessed pressure, and velocity components and streamlines computed with the proposed lowest order mixed-primal method.

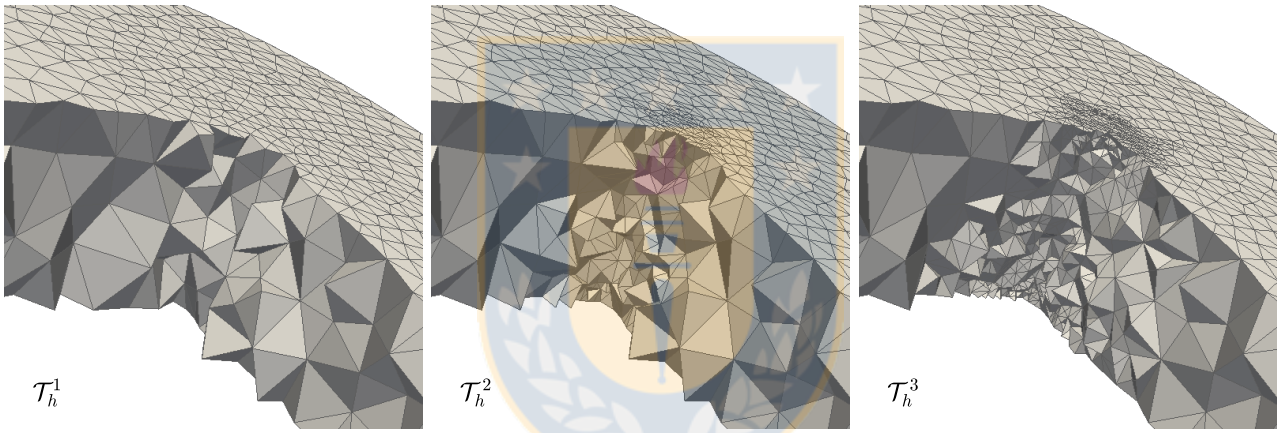


Figure 4.7: Test 3: zoom on the produced meshes after the first three steps of adaptive refinement using the first estimator as defined in (4.64).

# CHAPTER 5

# A vorticity-based fully-mixed formulation for the 3D Brinkman-Darcy problem

## 5.1 Introduction

This chapter is motivated by the numerical approximation of flow patterns in an heterogeneous media composed by a porous medium, where Darcy equations govern the flow behavior of a non-viscous incompressible fluid, and a much more permeable domain, where the laminar flow exhibits viscous effects and can be described by the linear Brinkman model. The two domains are separated by an essentially fixed interface, across which the flow passes from viscous to a non-viscous regime. According to the discussion in [14], we can assume that the Brinkman domain consists of an array of low concentration fixed particles, whereas the Darcy domain is a *classical* porous medium constituted by connected porous matrices. Such a scenario is often encountered in e.g. the modelling of surface and subsurface flow in porous media, petroleum reservoirs, or perfusion of physiological fluids into soft tissues, focusing typically on filtration or other similar processes of interest. We are also interested in accurately recovering the additional vorticity field (vectorial for three-dimensional flows and perpendicular to the plane of the flow and therefore considered scalar in 2D), which yields better information on circulation effects of the free fluid, sometimes observed near interfaces.

At the interface of the two domains, and depending on the specific form of the problem at hand, one typically requires preservation of physical quantities such as normal velocities, normal stresses, and so on. An abundant body of literature is devoted to different ways of treating the interface conditions, from both mathematical and numerical perspectives. These basically include sequential sub-structuring methods, where decoupled subproblems are solved on each subdomain, followed by an updating of the interface field values, then using these values as boundary data to solve a local problem on the other subdomain, and iterating in some adequate manner (see e.g. [16, 49, 85, 94]); and monolithic, fully coupled approaches where all sought fields are computed at once, for instance by a single operator acting on the two media or with the aid of Lagrange multipliers specifically designed to impose continuity of fields to be conserved across the interface (see for instance [13, 35, 52, 59]). Our method follows the latter strategy. Up to our knowledge, the coupling of Brinkman and Darcy flows has been only addressed in terms of the primal unknowns of velocity and pressure [25, 52, 85]. Vorticity-based formulations for the Stokes-Darcy coupling were introduced in [19] and later studied in [18, 51]. Differences with respect to these contributions include a slightly different formulation (we do not assume that the fluid boundary coincides with the interface between both domains, which is used

in [18, 19]); the analysis itself differs in that here we set pressure continuity across the interface using a Lagrange multiplier, and the normal stress conditions are weakly imposed. In addition, the proposed treatment does not require higher regularity of the fluid pressure as in e.g. [18]. At the discrete level, that work involves a family of nonconforming discretizations consisting in Nédélec elements for vorticity, piecewise constant elements for velocity, and Crouzeix-Raviart elements for the pressure. In contrast, here we use a finite element family where the **curl** of the subspace approximating the vorticity must be contained in the space where the discrete velocity of the fluid lives, and hence Raviart-Thomas and Nédélec finite elements for velocities and vorticity, respectively, become feasible choices. In turn, the pressures and the Lagrange multiplier are approximated, respectively, by discontinuous and continuous piecewise polynomials. Finally, our numerical tests also include the 3D case.

A general advantage of formulations involving vorticity is that this additional field can be accessed directly, without postprocessing; and it is straightforward to include non-inertial effects by modifying initial and boundary data [11, 12]. For instance, for external flows it is known that boundary conditions are better suited for vorticity than for pressure. Moreover, in many flow regimes the vorticity is concentrated in a specific region of the domain, which suggests the use of vorticity as guide to mesh refinement.

The contents of the chapter are organized as follows. In the remainder of the present section we recall basic terminology and some properties of functional spaces, and introduce further standard notations. In Section 5.2 we describe the coupled problem of interest and derive a first version of its mixed variational formulation. The solvability analysis of the later is carried out in Section 5.3. We first identify the non-trivial solutions of the associated homogeneous problem, and then reformulate the original continuous formulation in order to be able to guarantee unique solvability of it. The classical Babuška-Brezzi is then applied in such a way that the continuous inf-sup conditions of the main bilinear form are established by employing a known approach that has been recently referred as $T$-coercivity. Then, in Section 5.4 we introduce the associated Galerkin scheme and adapt the arguments from the continuous case to prove that, under suitable assumptions on the finite element subspaces involved, it is well-posed. Next, in Section 5.5 we modify the mixed formulation by incorporating a residual arising from the Brinkman momentum equation, and show that the resulting augmented scheme, yielding a strongly elliptic main bilinear form, does not require the aforementioned constraint. Finally, several numerical examples illustrating the good performance of the mixed finite element methods and confirming the theoretical rates of convergence are provided in Section 5.6.

We end this section by specifying some notations to be employed throughout the chapter. In particular, we utilize standard simplified terminology for Sobolev spaces and norms. For instance, if $\mathcal{O} \subseteq \mathrm{R}^3$ is a domain, $\mathcal{S} \subseteq \mathrm{R}^3$ is a Lipschitz surface, and $r \in \mathrm{R}$, we define

$$\mathbf{H}^r(\mathcal{O}) := [\mathrm{H}^r(\mathcal{O})]^3 \quad \text{and} \quad \mathbf{H}^r(\mathcal{S}) := [\mathrm{H}^r(\mathcal{S})]^3 .$$

However, when $r = 0$ we usually write $\mathbf{L}^2(\mathcal{O})$ and $\mathbf{L}^2(\mathcal{S})$ instead of $\mathbf{H}^0(\mathcal{O})$ and $\mathbf{H}^0(\mathcal{S})$, respectively. The corresponding norms are denoted by $\|\cdot\|_{r,\mathcal{O}}$ (for $\mathrm{H}^r(\mathcal{O})$ and $\mathbf{H}^r(\mathcal{O})$) and $\|\cdot\|_{r,\mathcal{S}}$ (for $\mathrm{H}^r(\mathcal{S})$ and $\mathbf{H}^r(\mathcal{S})$). In general, given any Hilbert space H, we use $\mathbf{H}$ to denote $\mathrm{H}^3$. In turn, in the realm of mixed methods (see [26]) one usually needs the Hilbert spaces

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

normed, respectively, with

$$\|\boldsymbol{v}\|_{\mathrm{div};\mathcal{O}} := \left\{ \|\boldsymbol{v}\|_{0,\mathcal{O}}^2 + \|\mathrm{div}\, \boldsymbol{v}\|_{0,\mathcal{O}}^2 \right\}^{1/2}, \quad \|\boldsymbol{v}\|_{\mathbf{curl};\mathcal{O}} := \left\{ \|\boldsymbol{v}\|_{0,\mathcal{O}}^2 + \|\mathbf{curl}\boldsymbol{v}\|_{0,\mathcal{O}}^2 \right\}^{1/2},$$

where, for any vector field $\boldsymbol{v} := (v_1, v_2, v_3)^{\mathtt{t}} \in \mathbf{L}^2(\mathcal{O})$,

$$\operatorname{div} \boldsymbol{v} := \sum_{i=1}^{3} \partial_i v_i \quad \text{and} \quad \mathbf{curl}\boldsymbol{v} := \nabla \times \boldsymbol{v} = \begin{pmatrix} \partial_2 v_3 - \partial_3 v_2 \\ \partial_3 v_1 - \partial_1 v_3 \\ \partial_1 v_2 - \partial_2 v_1 \end{pmatrix} .$$

In addition, we also recall the orthogonal decomposition

$$\mathrm{L}^2(\mathcal{O}) \;=\; \mathrm{L}^2_0(\mathcal{O}) \;\oplus\; P_0(\mathcal{O})\,, \tag{5.1}$$

where $P_0(\mathcal{O})$ is the space of constant functions on $\mathcal{O}$, and

$$\mathrm{L}^2_0(\mathcal{O}) \;=\; P_0(\mathcal{O})^{\perp} := \Big\{ q \in \mathrm{L}^2(\mathcal{O}) : \quad \int_{\mathcal{O}} q = 0 \Big\}\,. \tag{5.2}$$

Equivalently, each $q \in \mathrm{L}^2(\mathcal{O})$ can be uniquely decomposed as $q = q_0 + c$, with

$$q_0 := q - \frac{1}{|\mathcal{O}|}\int_{\mathcal{O}} q \;\in\; \mathrm{L}^2_0(\mathcal{O}) \quad \text{and} \quad c := \frac{1}{|\mathcal{O}|}\int_{\mathcal{O}} q \;\in\; \mathrm{R}\,. \tag{5.3}$$

Certainly, $\mathrm{L}^2_0(\mathcal{O})$ is endowed with the usual norm of $\mathrm{L}^2(\mathcal{O})$, and it is easy to see that there holds

$$\|q\|^2_{0,\mathcal{O}} \;=\; \|q_0\|^2_{0,\mathcal{O}} \;+\; |\mathcal{O}|\, c^2\,. \tag{5.4}$$

Finally, in what follows $\mathbf{0}$ stands for a generic null vector (including the null functional and operator), and we use $C$ and $c$, with or without subscripts, bars, tildes or hats, to denote generic constants independent of the discretization parameters, which may take different values in different occurrences.

## 5.2 The coupled problem and its mixed formulation

We first let $\Omega_{\mathrm{B}}$ and $\Omega_{\mathrm{D}}$ be bounded and simply connected polyhedral Lipschitz domains in $\mathrm{R}^3$ such that $\partial\Omega_{\mathrm{B}} \cap \partial\Omega_{\mathrm{D}} =: \Sigma \neq \emptyset$ and $\Omega_{\mathrm{B}} \cap \Omega_{\mathrm{D}} = \emptyset$, and set $\Omega := \overline{\Omega}_B \cup \overline{\Omega}_D$ with boundary $\Gamma = \partial\Omega$ split into $\Gamma_{\mathrm{B}}$ and $\Gamma_{\mathrm{D}}$ (see the sketch in Figure 5.1). Note that the interface $\Sigma$ between $\Omega_{\mathrm{B}}$ and $\Omega_{\mathrm{D}}$ does not necessarily coincide with $\partial\Omega_{\mathrm{B}}$ (as it was assumed in e.g. [18, 51]). Then, given source terms $\boldsymbol{f}_{\mathrm{D}} \in \mathbf{L}^2(\Omega_{\mathrm{D}})$ and $\boldsymbol{f}_{\mathrm{B}} \in \mathbf{L}^2(\Omega_{\mathrm{B}})$, we are interested in the Brinkman-Darcy coupled problem, which is formulated in what follows in terms of the fluid velocity $\boldsymbol{u}_{\mathrm{B}}$, the fluid pressure $p_{\mathrm{B}}$, the fluid vorticity $\boldsymbol{\omega}_{\mathrm{B}}$, the Darcy velocity $\boldsymbol{u}_{\mathrm{D}}$, and the Darcy pressure $p_{\mathrm{D}}$. More precisely, the sets of equations in the Brinkman and Darcy domains $\Omega_{\mathrm{B}}$ and $\Omega_{\mathrm{D}}$, are given, respectively, by

$$\left.\begin{array}{rcl} \alpha\,\boldsymbol{u}_{\mathrm{B}} + \nu\,\mathbf{curl}\,\boldsymbol{\omega}_{\mathrm{B}} + \nabla p_{\mathrm{B}} &=& \boldsymbol{f}_{\mathrm{B}} \\ \boldsymbol{\omega}_{\mathrm{B}} - \mathbf{curl}\boldsymbol{u}_{\mathrm{B}} &=& \mathbf{0} \\ \operatorname{div}\boldsymbol{u}_{\mathrm{B}} &=& 0 \end{array}\right\} \quad \text{in} \quad \Omega_{\mathrm{B}}\,, \tag{5.5}$$

and

$$\left.\begin{array}{rcl} \mu\boldsymbol{u}_{\mathrm{D}} + \nabla p_{\mathrm{D}} &=& \boldsymbol{f}_{\mathrm{D}} \\ \operatorname{div}\boldsymbol{u}_{\mathrm{D}} &=& 0 \end{array}\right\} \quad \text{in} \quad \Omega_{\mathrm{D}}\,, \tag{5.6}$$

where $\nu > 0$ is the kinematic viscosity of the fluid, $\mu > 0$ depends on this viscosity and on the permeability of the porous medium, which is assumed to be homogeneous, and $\alpha > 0$ is a parameter

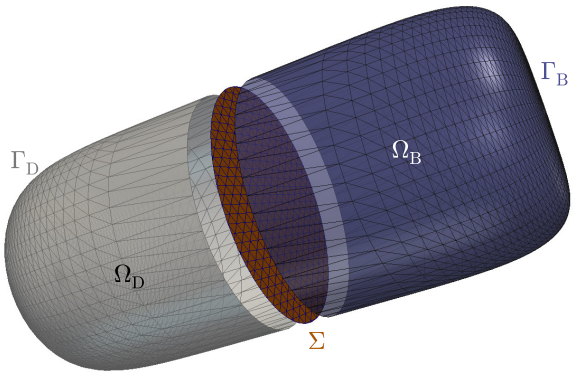


Figure 5.1: Sketch of the domains occupied by the incompressible fluid and by the porous medium ($\Omega_B$ and $\Omega_D$, respectively), interface $\Sigma$, and corresponding boundaries.

related to the relaxation time (typically proportional to the inverse of the timestep after a Rothe type time discretization). In turn, the corresponding transmission conditions become

$$\boldsymbol{u}_D \cdot \boldsymbol{n} = \boldsymbol{u}_B \cdot \boldsymbol{n} \quad \text{and} \quad p_D = p_B \quad \text{on} \quad \Sigma\,, \tag{5.7}$$

where $\boldsymbol{n}$ stands for the outward normal at $\Omega_B$ and $\Omega_D$, whereas the boundary conditions reduce to

$$\boldsymbol{\omega}_B \times \boldsymbol{n} = 0 \quad \text{on} \quad \partial\Omega_B = \Sigma \cup \Gamma_B\,, \quad \boldsymbol{u}_B \cdot \boldsymbol{n} = 0 \quad \text{on} \quad \Gamma_B\,, \quad \text{and} \quad \boldsymbol{u}_D \cdot \boldsymbol{n} = 0 \quad \text{on} \quad \Gamma_D\,. \tag{5.8}$$

Evidently, this choice of boundary data (especially those applied on the vorticity) is driven mainly by easiness of the subsequent analysis, but we stress that other conditions could be incorporated without compromising the main ideas in this work.

We now aim to derive the mixed variational formulation of (5.5) - (5.8). We begin by testing the first equation in (5.5) with functions in the space

$$\mathbf{H}_B(\mathrm{div};\Omega_B) := \left\{ \boldsymbol{v}_B \in \mathbf{H}(\mathrm{div};\Omega_B) : \quad \boldsymbol{v}_B \cdot \boldsymbol{n} = 0 \quad \text{on} \quad \Gamma_B \right\}.$$

To this end, we need to recall that the fact that $\boldsymbol{v}_B \cdot \boldsymbol{n} = 0$ on $\Gamma_B$ guarantees that $\boldsymbol{v}_B \cdot \boldsymbol{n}|_\Sigma$ belongs to $\mathrm{H}^{-1/2}(\Sigma)$ for each $\boldsymbol{v}_B \in \mathbf{H}_B(\mathrm{div};\Omega_B)$ (see the beginning of Section 5.3 below for further details on this issue). In this way, integrating by parts and using the respective boundary conditions, we find that

$$\alpha \int_{\Omega_B} \boldsymbol{u}_B \cdot \boldsymbol{v}_B + \nu \int_{\Omega_B} \boldsymbol{v}_B \cdot \mathbf{curl}\boldsymbol{\omega}_B - \int_{\Omega_B} p_B \,\mathrm{div}\, \boldsymbol{v}_B + \langle \boldsymbol{v}_B \cdot \boldsymbol{n}, \lambda \rangle_\Sigma = \int_{\Omega_B} \boldsymbol{f}_B \cdot \boldsymbol{v}_B \quad \forall\, \boldsymbol{v}_B \in \mathbf{H}_B(\mathrm{div};\Omega_B), \tag{5.9}$$

where, thanks to the second transmission condition in (5.7), we have introduced the auxiliary unknown

$$\lambda := p_D|_\Sigma = p_B|_\Sigma \in \mathrm{H}^{1/2}(\Sigma)\,,$$

and $\langle \cdot, \cdot \rangle_\Sigma$ denotes the duality pairing of $\mathrm{H}^{-1/2}(\Sigma)$ and $\mathrm{H}^{1/2}(\Sigma)$ with respect to the $\mathrm{L}^2(\Sigma)$-inner product. Furthermore, it will become clear below that $\lambda$ can also be seen as the Lagrange multiplier enforcing the continuity of pressure across the interface $\Sigma$. Next, we define

$$\mathbf{H}_0(\mathbf{curl};\Omega_B) := \left\{ \boldsymbol{z}_B \in \mathbf{H}(\mathbf{curl};\Omega_B) : \quad \boldsymbol{z}_B \times \boldsymbol{n} = \mathbf{0} \quad \text{on} \quad \partial\Omega_B \right\},$$

so that testing the second equation in (5.5) with functions in this space, and integrating by parts, we obtain

$$\int_{\Omega_{\rm B}} \boldsymbol{\omega}_{\rm B}\cdot\boldsymbol{z}_{\rm B} - \int_{\Omega_{\rm B}} \boldsymbol{u}_{\rm B}\cdot\mathbf{curl}\boldsymbol{z}_{\rm B} = 0 \qquad \forall\, \boldsymbol{z}_{\rm B}\in \mathbf{H}_0(\mathbf{curl};\Omega_{\rm B})\,. \tag{5.10}$$

In turn, the third equation in (5.5) is initially tested as

$$\int_{\Omega_{\rm B}} q_{\rm B}\,\mathrm{div}\,\boldsymbol{u}_{\rm B} = 0 \qquad \forall\, q_{\rm B}\in \mathrm{L}^2(\Omega_{\rm B})\,. \tag{5.11}$$

On the other hand, in order to deal with the equations in the Darcy domain, we now set

$$\mathbf{H}_{\rm D}(\mathrm{div};\Omega_{\rm D}) := \Big\{\boldsymbol{v}_{\rm D}\in\mathbf{H}(\mathrm{div};\Omega_{\rm D}):\quad \boldsymbol{v}_{\rm D}\cdot\boldsymbol{n}=0\quad\text{on}\quad \Gamma_{\rm D}\Big\}\,,$$

and test the first equation of (5.6) with functions in this space. Thus, integrating by parts, using the corresponding boundary conditions, noting that the normal $\boldsymbol{n}$ on $\Sigma$ points inward $\Omega_{\rm D}$, and recalling that $\lambda := p_{\rm D}|_\Sigma$, we get

$$\mu\int_{\Omega_{\rm D}} \boldsymbol{u}_{\rm D}\cdot\boldsymbol{v}_{\rm D} - \int_{\Omega_{\rm D}} p_{\rm D}\,\mathrm{div}\,\boldsymbol{v}_{\rm D} - \langle \boldsymbol{v}_{\rm D}\cdot\boldsymbol{n},\lambda\rangle_\Sigma = \int_{\Omega_{\rm D}} \boldsymbol{f}_{\rm D}\cdot\boldsymbol{v}_{\rm D} \qquad \forall\,\boldsymbol{v}_{\rm D}\in\mathbf{H}_{\rm D}(\mathrm{div};\Omega_{\rm D})\,. \tag{5.12}$$

In addition, similarly as for the incompressibility condition in $\Omega_{\rm B}$, the second equation in (5.6) is initially tested as

$$\int_{\Omega_{\rm D}} q_{\rm D}\,\mathrm{div}\,\boldsymbol{v}_{\rm D} = 0 \qquad \forall\, q_{\rm D}\in \mathrm{L}^2(\Omega_{\rm D})\,. \tag{5.13}$$

We end the present derivation with the weak imposition of the essential transmission condition given by the first equation in (5.7), that is

$$\langle \boldsymbol{u}_{\rm B}\cdot\boldsymbol{n} - \boldsymbol{u}_{\rm D}\cdot\boldsymbol{n},\xi\rangle_\Sigma = 0 \qquad \forall\,\xi\in \mathrm{H}^{1/2}(\Sigma)\,. \tag{5.14}$$

Consequently, reordering (5.9) - (5.14) in a suitable way, namely placing each set of equations {(5.9), (5.10), (5.12)} and {(5.11), (5.13), (5.14)} into a single equation each, we arrive at the mixed formulation of (5.5) - (5.8): Find $\vec{\boldsymbol{u}} := (\boldsymbol{u}_{\rm B},\boldsymbol{\omega}_{\rm B},\boldsymbol{u}_{\rm D})\in\mathbf{H}$ and $\vec{p} := (p_{\rm B},p_{\rm D},\lambda)\in\mathbf{Q}$ such that

$$\begin{aligned} \mathbf{a}(\vec{\boldsymbol{u}},\vec{\boldsymbol{v}}) + \mathbf{b}(\vec{\boldsymbol{v}},\vec{p}) &= \mathbf{F}(\vec{\boldsymbol{v}}) \qquad \forall\,\vec{\boldsymbol{v}} := (\boldsymbol{v}_{\rm B},\boldsymbol{z}_{\rm B},\boldsymbol{v}_{\rm D})\in\mathbf{H}\,,\\ \mathbf{b}(\vec{\boldsymbol{u}},\vec{q}) &= \mathbf{G}(\vec{q}) \qquad \forall\,\vec{q} := (q_{\rm B},q_{\rm D},\xi)\in\mathbf{Q}\,, \end{aligned} \tag{5.15}$$

where

$$\mathbf{H} := \mathbf{H}_{\rm B}(\mathrm{div};\Omega_{\rm B})\times\mathbf{H}_0(\mathbf{curl};\Omega_{\rm B})\times\mathbf{H}_{\rm D}(\mathrm{div};\Omega_{\rm D})\,,\quad \mathbf{Q} := \mathrm{L}^2(\Omega_{\rm B})\times\mathrm{L}^2(\Omega_{\rm D})\times\mathrm{H}^{1/2}(\Sigma)\,,$$

$\mathbf{a}:\mathbf{H}\times\mathbf{H}\to\mathrm{R}$ and $\mathbf{b}:\mathbf{H}\times\mathbf{Q}\to\mathrm{R}$ are the bilinear forms defined by

$$\begin{aligned} \mathbf{a}(\vec{\boldsymbol{u}},\vec{\boldsymbol{v}}) := \alpha\int_{\Omega_{\rm B}}\boldsymbol{u}_{\rm B}\cdot\boldsymbol{v}_{\rm B} + \nu\int_{\Omega_{\rm B}}\boldsymbol{\omega}_{\rm B}\cdot\boldsymbol{z}_{\rm B} + \nu\int_{\Omega_{\rm B}}\boldsymbol{v}_{\rm B}\cdot\mathbf{curl}\boldsymbol{\omega}_{\rm B}\\ - \nu\int_{\Omega_{\rm B}}\boldsymbol{u}_{\rm B}\cdot\mathbf{curl}\boldsymbol{z}_{\rm B} + \mu\int_{\Omega_{\rm D}}\boldsymbol{u}_{\rm D}\cdot\boldsymbol{v}_{\rm D} \qquad \forall\,(\vec{\boldsymbol{u}},\vec{\boldsymbol{v}})\in\mathbf{H}\times\mathbf{H}\,, \end{aligned} \tag{5.16}$$

$$\mathbf{b}(\vec{\boldsymbol{v}},\vec{q}) := -\int_{\Omega_{\rm B}} q_{\rm B}\,\mathrm{div}\,\boldsymbol{v}_{\rm B} - \int_{\Omega_{\rm D}} q_{\rm D}\,\mathrm{div}\,\boldsymbol{v}_{\rm D} + \langle \boldsymbol{v}_{\rm B}\cdot\boldsymbol{n} - \boldsymbol{v}_{\rm D}\cdot\boldsymbol{n},\xi\rangle_\Sigma \qquad \forall\,(\vec{\boldsymbol{v}},\vec{q})\in\mathbf{H}\times\mathbf{Q}\,, \tag{5.17}$$

and $\mathbf{F}\in\mathbf{H}'$ and $\mathbf{G}\in\mathbf{Q}'$ are the functionals defined by

$$\mathbf{F}(\vec{\boldsymbol{v}}) := \int_{\Omega_{\rm B}}\boldsymbol{f}_{\rm B}\cdot\boldsymbol{v}_{\rm B} + \int_{\Omega_{\rm D}}\boldsymbol{f}_{\rm D}\cdot\boldsymbol{v}_{\rm D} \quad \forall\,\vec{\boldsymbol{v}}\in\mathbf{H}\,,\quad\text{and}\quad \mathbf{G} = \mathbf{0}\,. \tag{5.18}$$

## 5.3 Solvability analysis of the mixed formulation

In this section we analyze the solvability of (5.15). For this purpose, we first recall some definitions and technical results concerning Sobolev spaces on $\Gamma_{\mathrm{D}}$, $\Gamma_{\mathrm{B}}$, and $\Sigma$. We begin by mentioning that, given $\eta \in \mathrm{H}^{-1/2}(\partial\Omega_{\mathrm{D}})$, its restriction to $\Gamma_{\mathrm{D}}$, say $\eta|_{\Gamma_{\mathrm{D}}}$, is defined as

$$\langle \eta|_{\Gamma_{\mathrm{D}}}, \rho\rangle_{\Gamma_{\mathrm{D}}} := \langle \eta, E_{\mathrm{D},0}(\rho)\rangle_{\partial\Omega_{\mathrm{D}}} \qquad \forall\, \rho \in \mathrm{H}^{1/2}_{00}(\Gamma_{\mathrm{D}})\,,$$

where $E_{\mathrm{D},0} : \mathrm{H}^{1/2}(\Gamma_{\mathrm{D}}) \to \mathrm{L}^2(\partial\Omega_{\mathrm{D}})$ is the extension by zero in $\Sigma := \partial\Omega_{\mathrm{D}}\backslash\Gamma_D$, and

$$\mathrm{H}^{1/2}_{00}(\Gamma_{\mathrm{D}}) := \Big\{ \rho \in \mathrm{H}^{1/2}(\Gamma_{\mathrm{D}}) : \quad E_{\mathrm{D},0}(\rho) \in \mathrm{H}^{1/2}(\partial\Omega_{\mathrm{D}}) \Big\}\,,$$

which is endowed with the natural norm $\|\rho\|_{1/2,00,\Gamma_{\mathrm{D}}} := \|E_{\mathrm{D},0}(\rho)\|_{1/2,\partial\Omega_{\mathrm{D}}}$. It is quite clear, then, that $\eta|_{\Gamma_{\mathrm{D}}}$ belongs to $\mathrm{H}^{-1/2}_{00}(\Gamma_{\mathrm{D}})$, the dual of $\mathrm{H}^{1/2}_{00}(\Gamma_{\mathrm{D}})$, and that $\eta = 0$ on $\Gamma_{\mathrm{D}}$ (equivalently $\eta|_{\Gamma_{\mathrm{D}}} = 0$) if and only if

$$\langle \eta, E_{\mathrm{D},0}(\rho)\rangle_{\partial\Omega_{\mathrm{D}}} = 0 \qquad \forall\, \rho \in \mathrm{H}^{1/2}_{00}(\Gamma_{\mathrm{D}})\,.$$

Hereafter, $\langle\cdot,\cdot\rangle_{\Gamma_{\mathrm{D}}}$ (resp. $\langle\cdot,\cdot\rangle_{\partial\Omega_{\mathrm{D}}}$) stands for the duality pairing of the spaces $\mathrm{H}^{-1/2}_{00}(\Gamma_{\mathrm{D}})$ and $\mathrm{H}^{1/2}_{00}(\Gamma_{\mathrm{D}})$ (resp. $\mathrm{H}^{-1/2}(\partial\Omega_{\mathrm{D}})$ and $\mathrm{H}^{1/2}(\partial\Omega_{\mathrm{D}})$) with respect to the $\mathrm{L}^2(\Gamma_{\mathrm{D}})$ (resp. $\mathrm{L}^2(\partial\Omega_{\mathrm{D}})$) inner product. Furthermore, it is not difficult to show (see, e.g. [59, Section 2]) that there holds the decomposition

$$H^{1/2}(\partial\Omega_{\mathrm{D}}) = E_{\mathrm{D}}\big(\mathrm{H}^{1/2}(\Sigma)\big) \oplus E_{\mathrm{D},0}\big(\mathrm{H}^{1/2}_{00}(\Gamma_{\mathrm{D}})\big)\,,$$

where $E_{\mathrm{D}} : \mathrm{H}^{1/2}(\Sigma) \to \mathrm{H}^{1/2}(\partial\Omega_{\mathrm{D}})$ is the bounded linear extension defined by $E_{\mathrm{D}}(\xi) := z_\xi|_{\partial\Omega_{\mathrm{D}}} \quad \forall\, \xi \in \mathrm{H}^{1/2}(\Sigma)$, with $z_\xi \in \mathrm{H}^1(\Omega_{\mathrm{D}})$ being the unique weak solution of the boundary value problem with mixed boundary conditions:

$$\Delta z_\xi = 0 \quad \text{in} \quad \Omega_{\mathrm{D}}\,, \quad z_\xi = \xi \quad \text{on} \quad \Sigma\,, \quad \nabla z_\xi \cdot \boldsymbol{n} = 0 \quad \text{on} \quad \Gamma_{\mathrm{D}}\,.$$

In this way, given $\varphi \in \mathrm{H}^{1/2}(\partial\Omega_{\mathrm{D}})$, there exist unique $\xi_\varphi \in \mathrm{H}^{1/2}(\Sigma)$ and $\rho_\varphi \in \mathrm{H}^{1/2}_{00}(\Gamma_{\mathrm{D}})$ such that

$$\varphi = E_{\mathrm{D}}(\xi_\varphi) + E_{\mathrm{D},0}(\rho_\varphi)\,, \tag{5.19}$$

and hence

$$\langle \eta, \varphi\rangle_{\partial\Omega_{\mathrm{D}}} = \langle \eta, E_{\mathrm{D}}(\xi_\varphi)\rangle_{\partial\Omega_{\mathrm{D}}} + \langle \eta, E_{\mathrm{D},0}(\rho_\varphi)\rangle_{\partial\Omega_{\mathrm{D}}}\,, \tag{5.20}$$

which can be rewritten as

$$\langle \eta, \varphi\rangle_{\partial\Omega_{\mathrm{D}}} = \langle \eta_\Sigma, \xi_\varphi\rangle_\Sigma + \langle \eta_{\mathrm{D}}, \rho_\varphi\rangle_{\Gamma_{\mathrm{D}}}\,,$$

where $\eta_\Sigma \in \mathrm{H}^{-1/2}(\Sigma)$ and $\eta_{\mathrm{D}} \in \mathrm{H}^{-1/2}_{00}(\Gamma_{\mathrm{D}})$ are defined accordingly. In addition, it is clear from (5.19) and the definitions of $E_{\mathrm{D}}$ and $E_{\mathrm{D},0}$ that actually $\xi_\varphi = \varphi|_\Sigma$ for each $\varphi \in \mathrm{H}^{1/2}(\partial\Omega_{\mathrm{D}})$. In particular, when $\eta = 0$ on $\Gamma_{\mathrm{D}}$, the foregoing equations yield $\langle \eta, \varphi\rangle_{\partial\Omega_{\mathrm{D}}} = \langle \eta, E_{\mathrm{D}}(\xi_\varphi)\rangle_{\partial\Omega_{\mathrm{D}}} =: \langle \eta_\Sigma, \xi_\varphi\rangle_\Sigma = \langle \eta_\Sigma, \varphi|_\Sigma\rangle_\Sigma$, and hence $\eta$ can be identified with a functional $\eta_\Sigma \in \mathrm{H}^{-1/2}(\Sigma)$. In other words, one simply says that $\eta|_\Sigma = \eta_\Sigma \in \mathrm{H}^{-1/2}(\Sigma)$. Note that an interesting application of this result arises when we consider $\boldsymbol{v}_{\mathrm{D}} \in \mathbf{H}_{\mathrm{D}}(\mathrm{div};\Omega_{\mathrm{D}})$ and define $\eta := \boldsymbol{v}_{\mathrm{D}} \cdot \boldsymbol{n} \in \mathrm{H}^{-1/2}(\partial\Omega_{\mathrm{D}})$. In fact, since $\boldsymbol{v}_{\mathrm{D}} \cdot \boldsymbol{n} = 0$ on $\Gamma_{\mathrm{D}}$, we readily deduce that $\boldsymbol{v}_{\mathrm{D}} \cdot \boldsymbol{n}|_\Sigma \in \mathrm{H}^{-1/2}(\Sigma)$. Moreover, the analogue conclusion obtained by exchanging $\Omega_{\mathrm{D}}$, $\Gamma_{\mathrm{D}}$, and $\mathbf{H}_{\mathrm{D}}(\mathrm{div};\Omega_{\mathrm{D}})$ by $\Omega_{\mathrm{B}}$, $\Gamma_{\mathrm{B}}$, and $\mathbf{H}_{\mathrm{B}}(\mathrm{div};\Omega_{\mathrm{B}})$, respectively, is precisely what we used in Section 5.2 for the derivation of (5.9).

We are now in position to provide the following preliminary result, which establishes a continuous inf-sup condition on $\mathbf{H}_{\mathrm{D}}(\mathrm{div};\Omega_{\mathrm{D}}) \times \big(\mathrm{L}^2_0(\Omega_{\mathrm{D}}) \times \mathrm{H}^{1/2}(\Sigma)\big)$.

**Lemma 5.1.** *There exists $\beta_{\mathrm{D}} > 0$ such that*

$$S_{\mathrm{D}}(q_{\mathrm{D}},\xi) := \sup_{\substack{\boldsymbol{v}_{\mathrm{D}}\in\mathbf{H}_{\mathrm{D}}(\mathrm{div};\Omega_{\mathrm{D}})\\ \boldsymbol{v}_{\mathrm{D}}\neq\mathbf{0}}} \frac{\displaystyle\int_{\Omega_{\mathrm{D}}} q_{\mathrm{D}}\,\mathrm{div}\,\boldsymbol{v}_{\mathrm{D}} + \langle \boldsymbol{v}_{\mathrm{D}}\cdot\boldsymbol{n},\xi\rangle_{\Sigma}}{\|\boldsymbol{v}_{\mathrm{D}}\|_{\mathrm{div};\Omega_{\mathrm{D}}}} \geq \beta_{\mathrm{D}}\left\{\|q_{\mathrm{D}}\|_{0,\Omega_{\mathrm{D}}} + \|\xi\|_{1/2,\Sigma}\right\} \tag{5.21}$$

*for all $(q_{\mathrm{D}},\xi) \in \mathrm{L}^2_0(\Omega_{\mathrm{D}}) \times \mathrm{H}^{1/2}(\Sigma)$.*

*Proof.* It proceeds almost verbatim as the 2D version provided in [75, Lemma 3.3]. However, for sake of completeness, most details are given in what follows. The first part of the proof reduces to show that the operator $\mathrm{div} : \mathbf{H}_{\mathrm{D}}(\mathrm{div};\Omega_{\mathrm{D}}) \to \mathrm{L}^2_0(\Omega_{\mathrm{D}})$ is surjective, for which, given $q_{\mathrm{D}} \in \mathrm{L}^2_0(\Omega_{\mathrm{D}})$, it suffices to define the pre-image $\widetilde{\boldsymbol{v}}_{\mathrm{D}} := \nabla z \in \mathbf{H}_{\mathrm{D}}(\mathrm{div};\Omega_{\mathrm{D}})$, where $z \in \mathrm{H}^1(\Omega_{\mathrm{D}})$ is the unique weak solution of the Neumann boundary value problem

$$\Delta z = q_{\mathrm{D}} \quad \text{in} \quad \Omega_{\mathrm{D}}\,, \quad \nabla z\cdot\boldsymbol{n} = 0 \quad \text{on} \quad \partial\Omega_{\mathrm{D}}\,, \quad \int_{\Omega_{\mathrm{D}}} z = 0\,. \tag{5.22}$$

Indeed, it is clear that $\mathrm{div}\,\widetilde{\boldsymbol{v}}_{\mathrm{D}} = \Delta z = q_{\mathrm{D}}$ in $\Omega_{\mathrm{D}}$, and the continuous dependence result for (5.22) establishes that $\|\widetilde{\boldsymbol{v}}_{\mathrm{D}}\|_{0,\Omega_{\mathrm{D}}} = |z|_{1,\Omega_{\mathrm{D}}} \leq \|z\|_{1,\Omega_{\mathrm{D}}} \leq c\|q_{\mathrm{D}}\|_{0,\Omega_{\mathrm{D}}}$, whence we readily deduce that $\|\widetilde{\boldsymbol{v}}_{\mathrm{D}}\|_{\mathrm{div};\Omega_{\mathrm{D}}} \leq \widetilde{C}\,\|q_{\mathrm{D}}\|_{0,\Omega_{\mathrm{D}}}$. In this way, since

$$S_{\mathrm{D}}(q_{\mathrm{D}},\xi) \geq \sup_{\substack{\boldsymbol{v}_{\mathrm{D}}\in\mathbf{H}_{\mathrm{D}}(\mathrm{div};\Omega_{\mathrm{D}})\\ \boldsymbol{v}_{\mathrm{D}}\neq\mathbf{0}}} \frac{\displaystyle\int_{\Omega_{\mathrm{D}}} q_{\mathrm{D}}\,\mathrm{div}\,\boldsymbol{v}_{\mathrm{D}}}{\|\boldsymbol{v}_{\mathrm{D}}\|_{\mathrm{div};\Omega_{\mathrm{D}}}} - \|\xi\|_{1/2,\Sigma} \geq \frac{\left|\displaystyle\int_{\Omega_{\mathrm{D}}} q_{\mathrm{D}}\,\mathrm{div}\,\widetilde{\boldsymbol{v}}_{\mathrm{D}}\right|}{\|\widetilde{\boldsymbol{v}}_{\mathrm{D}}\|_{\mathrm{div};\Omega_{\mathrm{D}}}} - \|\xi\|_{1/2,\Sigma}\,,$$

the foregoing identity and estimates imply the existence of $C > 0$ such that

$$S_{\mathrm{D}}(q_{\mathrm{D}},\xi) \geq C\,\|q_{\mathrm{D}}\|_{0,\Omega_{\mathrm{D}}} - \|\xi\|_{1/2,\Sigma}\,. \tag{5.23}$$

In turn, the main ingredient of the second part has to do with the construction of a proper extension of an arbitrary $\phi \in \mathrm{H}^{-1/2}(\Sigma)$ to a functional $\eta \in \mathrm{H}^{-1/2}(\partial\Omega_{\mathrm{D}})$. More precisely, given $\xi \in \mathrm{H}^{1/2}(\Sigma)$, we consider $\phi \in \mathrm{H}^{-1/2}(\Sigma)$ and, following the previous analysis and notations, we simply define $\eta \in \mathrm{H}^{-1/2}(\partial\Omega_{\mathrm{D}})$ as

$$\langle \eta,\varphi\rangle_{\partial\Omega_{\mathrm{D}}} := \langle \phi,\xi_{\varphi}\rangle_{\Sigma} = \langle \phi,\varphi|_{\Sigma}\rangle_{\Sigma} \qquad \forall\,\varphi \in \mathrm{H}^{1/2}(\partial\Omega_{\mathrm{D}})\,,$$

which yields $\|\eta\|_{-1/2,\partial\Omega_{\mathrm{D}}} \leq \|\phi\|_{-1/2,\Sigma}$. It follows straightforwardly from (5.19) and (5.20) that

$$\langle \eta, E_{\mathrm{D},0}(\rho)\rangle_{\partial\Omega_{\mathrm{D}}} = 0 \quad \forall\,\rho \in \mathrm{H}^{1/2}_{00}(\Gamma_{\mathrm{D}}) \quad \text{and} \quad \langle \eta, E_{\mathrm{D}}(\xi)\rangle_{\partial\Omega_{\mathrm{D}}} = \langle \phi,\xi\rangle_{\Sigma} \quad \forall\,\xi \in \mathrm{H}^{1/2}(\Sigma)\,,$$

which says, equivalently, that $\eta = 0$ on $\Gamma_{\mathrm{D}}$ and $\eta = \phi$ on $\Sigma$. Next, we let $z_{\eta} \in \mathrm{H}^1(\Omega_{\mathrm{D}})$ be the unique weak solution of the boundary value problem

$$\Delta z_{\eta} = \frac{1}{|\Omega_{\mathrm{D}}|}\,\langle \eta,1\rangle_{\partial\Omega_{\mathrm{D}}} \quad \text{in} \quad \Omega_{\mathrm{D}}\,, \quad \nabla z_{\eta}\cdot\boldsymbol{n} = \eta \quad \text{on} \quad \partial\Omega_{\mathrm{D}}\,, \quad \int_{\Omega_{\mathrm{D}}} z_{\eta} = 0\,,$$

define $\boldsymbol{w}_{\mathrm{D}} := \nabla z_{\eta}$ in $\Omega_{\mathrm{D}}$, and observe that $\mathrm{div}\,\boldsymbol{w}_{\mathrm{D}} = \frac{1}{|\Omega_{\mathrm{D}}|}\,\langle \eta,1\rangle_{\partial\Omega_{\mathrm{D}}} \in P_0(\Omega_{\mathrm{D}})$ (which yields $\boldsymbol{w}_{\mathrm{D}} \in \mathbf{H}(\mathrm{div};\Omega_{\mathrm{D}})$), $\boldsymbol{w}_{\mathrm{D}}\cdot n = \eta$ on $\partial\Omega_{\mathrm{D}}$, and $\|\boldsymbol{w}_{\mathrm{D}}\|_{\mathrm{div};\Omega_{\mathrm{D}}} \leq C\,\|\eta\|_{-1/2,\partial\Omega_{\mathrm{D}}} \leq C\,\|\phi\|_{-1/2,\Sigma}$. It follows that $\boldsymbol{w}_{\mathrm{D}} \in \mathbf{H}_{\mathrm{D}}(\mathrm{div};\Omega_{\mathrm{D}})$ and hence

$$S_{\mathrm{D}}(q_{\mathrm{D}},\xi) \geq \frac{\left|\displaystyle\int_{\Omega_{\mathrm{D}}} q_{\mathrm{D}}\,\mathrm{div}\,\boldsymbol{w}_{\mathrm{D}} + \langle \boldsymbol{w}_{\mathrm{D}}\cdot\boldsymbol{n},\xi\rangle_{\Sigma}\right|}{\|\boldsymbol{w}_{\mathrm{D}}\|_{\mathrm{div};\Omega_{\mathrm{D}}}} = \frac{|\langle \phi,\xi\rangle_{\Sigma}|}{\|\boldsymbol{w}_{\mathrm{D}}\|_{\mathrm{div};\Omega_{\mathrm{D}}}} \geq c\,\frac{|\langle \phi,\xi\rangle_{\Sigma}|}{\|\phi\|_{-1/2,\Sigma}}\,,$$

which, being valid for any $\phi \in \mathrm{H}^{-1/2}(\Sigma)$, implies that $S_{\mathrm{D}}(q_{\mathrm{D}},\xi) \geq c\,\|\xi\|_{1/2,\Sigma}$. This inequality and (5.23) yield (5.21), thus completing the proof.

□

The following result is basically a "*mirror reflection*" through $\Sigma$ of the previous lemma.

**Lemma 5.2.** *There exists $\beta_{\mathrm{B}} > 0$ such that*

$$S_{\mathrm{B}}(q_{\mathrm{B}},\xi) := \sup_{\substack{\boldsymbol{v}_{\mathrm{B}} \in \mathbf{H}_{\mathrm{B}}(\mathrm{div};\Omega_{\mathrm{B}}) \\ \boldsymbol{v}_{\mathrm{B}} \neq \mathbf{0}}} \frac{\displaystyle\int_{\Omega_{\mathrm{B}}} q_{\mathrm{B}}\,\mathrm{div}\,\boldsymbol{v}_{\mathrm{B}} - \langle \boldsymbol{v}_{\mathrm{B}}\cdot\boldsymbol{n},\xi\rangle_{\Sigma}}{\|\boldsymbol{v}_{\mathrm{B}}\|_{\mathrm{div};\Omega_{\mathrm{B}}}} \geq \beta_{\mathrm{B}}\left\{\|q_{\mathrm{B}}\|_{0,\Omega_{\mathrm{B}}} + \|\xi\|_{1/2,\Sigma}\right\} \tag{5.24}$$

*for all $(q_{\mathrm{B}},\xi) \in \mathrm{L}^2_0(\Omega_{\mathrm{B}}) \times \mathrm{H}^{1/2}(\Sigma)$.*

*Proof.* It proceeds exactly as the proof of Lemma 5.1 by replacing $\Omega_{\mathrm{D}}$, $\Gamma_{\mathrm{D}}$, and $\mathbf{H}_{\mathrm{D}}(\mathrm{div};\Omega_{\mathrm{D}})$ by $\Omega_{\mathrm{B}}$, $\Gamma_{\mathrm{B}}$, and $\mathbf{H}_{\mathrm{B}}(\mathrm{div};\Omega_{\mathrm{B}})$, respectively.

□

Lemma 5.1 and 5.2 imply the following continuous inf-sup condition for $\mathbf{b}$.

**Lemma 5.3.** *There exists $\beta > 0$ such that*

$$S(\vec{q}) := \sup_{\substack{\vec{\boldsymbol{v}} \in \mathbf{H} \\ \vec{\boldsymbol{v}} \neq \mathbf{0}}} \frac{\mathbf{b}(\vec{\boldsymbol{v}},\vec{q})}{\|\vec{\boldsymbol{v}}\|_{\mathbf{H}}} \geq \beta\left\{\|q_{\mathrm{B},0}\|_{0,\Omega_{\mathrm{B}}} + \|q_{\mathrm{D},0}\|_{0,\Omega_{\mathrm{D}}} + \|\xi - c_{\mathrm{B}}\|_{1/2,\Sigma} + \|\xi - c_{\mathrm{D}}\|_{1/2,\Sigma}\right\} \tag{5.25}$$

*for all $\vec{q} := (q_{\mathrm{B}},q_{\mathrm{D}},\xi) \in \mathbf{Q}$, where, according to (1.1), $q_{\mathrm{B}} = q_{\mathrm{B},0} + c_{\mathrm{B}}$ and $q_{\mathrm{D}} = q_{\mathrm{D},0} + c_{\mathrm{D}}$, with $q_{\mathrm{B},0} \in \mathrm{L}^2_0(\Omega_{\mathrm{B}})$, $q_{\mathrm{D},0} \in \mathrm{L}^2_0(\Omega_{\mathrm{D}})$, and $c_{\mathrm{B}} := \frac{1}{|\Omega_{\mathrm{B}}|}\int_{\Omega_{\mathrm{B}}} q_{\mathrm{B}}$, $c_{\mathrm{D}} := \frac{1}{|\Omega_{\mathrm{D}}|}\int_{\Omega_{\mathrm{D}}} q_{\mathrm{D}} \in R$.*

*Proof.* Given $\vec{q} := (q_{\mathrm{B}},q_{\mathrm{D}},\xi) \in \mathbf{Q}$, with $q_{\mathrm{B}}$ and $q_{\mathrm{D}}$ decomposed as indicated above, we integrate by parts in $\Omega_{\mathrm{B}}$ and $\Omega_{\mathrm{D}}$, respectively, to deduce that

$$\int_{\Omega_{\mathrm{B}}} q_{\mathrm{B}}\,\mathrm{div}\,\boldsymbol{v}_{\mathrm{B}} - \langle \boldsymbol{v}_{\mathrm{B}}\cdot\boldsymbol{n},\xi\rangle_{\Sigma} = \int_{\Omega_{\mathrm{B}}} q_{\mathrm{B},0}\,\mathrm{div}\,\boldsymbol{v}_{\mathrm{B}} - \langle \boldsymbol{v}_{\mathrm{B}}\cdot\boldsymbol{n},\xi - c_{\mathrm{B}}\rangle_{\Sigma} \qquad \forall\,\boldsymbol{v}_{\mathrm{B}} \in \mathbf{H}_{\mathrm{B}}(\mathrm{div};\Omega_{\mathrm{B}})\,,$$

and

$$\int_{\Omega_{\mathrm{D}}} q_{\mathrm{D}}\,\mathrm{div}\,\boldsymbol{v}_{\mathrm{D}} + \langle \boldsymbol{v}_{\mathrm{D}}\cdot\boldsymbol{n},\xi\rangle_{\Sigma} = \int_{\Omega_{\mathrm{D}}} q_{\mathrm{D},0}\,\mathrm{div}\,\boldsymbol{v}_{\mathrm{D}} + \langle \boldsymbol{v}_{\mathrm{D}}\cdot\boldsymbol{n},\xi - c_{\mathrm{D}}\rangle_{\Sigma} \qquad \forall\,\boldsymbol{v}_{\mathrm{D}} \in \mathbf{H}_{\mathrm{D}}(\mathrm{div};\Omega_{\mathrm{D}})\,.$$

Hence, bearing in mind the definitions of the bilinear form $\mathbf{b}$ (*cf.* (5.17)) and the operators $S_{\mathrm{D}}$ and $S_{\mathrm{B}}$ (*cf.* (5.21), (5.24)), and employing the foregoing equations, we easily find that

$$S(\vec{q}) \geq S_{\mathrm{D}}(q_{\mathrm{D},0},\xi - c_{\mathrm{D}}) \qquad \text{and} \qquad S(\vec{q}) \geq S_{\mathrm{B}}(q_{\mathrm{B},0},\xi - c_{\mathrm{B}})$$

Consequently, these inequalities, along with straightforward applications of Lemmata 5.1 and 5.2, imply (5.25) and complete the proof.

□

Having proved a first property for $\mathbf{b}$, we now observe that the bilinear form $\mathbf{a}$ satisfies a positiveness condition. More precisely, it follows directly from its definition (*cf.* (5.16)) that

$$\mathbf{a}(\vec{\boldsymbol{v}},\vec{\boldsymbol{v}}) \,=\, \alpha\,\|\boldsymbol{v}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} \,+\, \nu\,\|\boldsymbol{z}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} \,+\, \mu\,\|\boldsymbol{v}_{\mathrm{D}}\|^2_{0,\Omega_{\mathrm{D}}} \qquad \forall\,\vec{\boldsymbol{v}} \,:=\, (\boldsymbol{v}_{\mathrm{B}},\boldsymbol{z}_{\mathrm{B}},\boldsymbol{v}_{\mathrm{D}}) \in \mathbf{H}\,. \tag{5.26}$$

A first result concerning the solvability of our mixed formulation (5.15) is established next.

**Theorem 5.1.** *Let* $(\vec{\boldsymbol{u}},\vec{p}) \,:=\, \big((\boldsymbol{u}_{\mathrm{B}},\boldsymbol{\omega}_{\mathrm{B}},\boldsymbol{u}_{\mathrm{D}}),(p_{\mathrm{B}},p_{\mathrm{D}},\lambda)\big) \in \mathbf{H}\times\mathbf{Q}$ *be a solution of the homogeneous problem associated to* (5.15)*, that is with* $\mathbf{F}=\mathbf{G}=\mathbf{0}$*. Then* $\vec{\boldsymbol{u}}=\mathbf{0}$ *and there exists* $c\in R$ *such that* $\vec{p}\,=\,(c,c,c)$*.*

*Proof.* We first notice from the second equation of (5.15) with $\vec{q}=\vec{p}$ that $\mathbf{b}(\vec{\boldsymbol{u}},\vec{p})\,=\,0$, and hence, taking $\vec{\boldsymbol{v}}=\vec{\boldsymbol{u}}$ in the first equation of (5.15) and using the identity (5.26), we arrive at

$$0 \,=\, \mathbf{a}(\vec{\boldsymbol{u}},\vec{\boldsymbol{u}}) \,=\, \alpha\,\|\boldsymbol{u}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} \,+\, \nu\,\|\boldsymbol{\omega}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} \,+\, \mu\,\|\boldsymbol{u}_{\mathrm{D}}\|^2_{0,\Omega_{\mathrm{D}}}\,,$$

from which it follows that $\vec{\boldsymbol{u}}\,=\,\mathbf{0}$. In this way, the first equation of (5.15) becomes now $\mathbf{b}(\vec{\boldsymbol{v}},\vec{p})=0$ for all $\vec{\boldsymbol{v}}\in\mathbf{Q}$, which, according to the continuous inf-sup condition for $\mathbf{b}$ given by Lemma 5.3, yields $p_{\mathrm{B},0}=0$, $p_{\mathrm{D},0}=0$, and $\lambda \,=\, \frac{1}{|\Omega_{\mathrm{B}}|}\int_{\Omega_{\mathrm{B}}} p_{\mathrm{B}} \,=\, \frac{1}{|\Omega_{\mathrm{D}}|}\int_{\Omega_{\mathrm{D}}} p_{\mathrm{D}} \,=:\, c\in\mathrm{R}$, so that $p_{\mathrm{B}}=p_{\mathrm{B},0}+c=c$ and $p_{\mathrm{D}}=p_{\mathrm{D},0}+c=c$.

□

As a straightforward consequence of Theorem 5.1 we conclude that whenever (5.15) has solution, it is not unique. Therefore, in order to overcome this drawback, we need to remove the constant $c\in\mathrm{R}$ from the solutions of the associated homogeneous system, for which from now on we propose to look for the unknown $\vec{p}$ in the space

$$\mathbf{Q}_0 \;:=\; \mathrm{L}^2_0(\Omega_{\mathrm{B}})\times\mathrm{L}^2(\Omega_{\mathrm{D}})\times\mathrm{H}^{1/2}(\Sigma)\,. \tag{5.27}$$

Alternatively, one could also consider $\mathbf{Q}_0 \,:=\, \mathrm{L}^2(\Omega_{\mathrm{B}})\times\mathrm{L}^2_0(\Omega_{\mathrm{D}})\times\mathrm{H}^{1/2}(\Sigma)$ or $\mathbf{Q}_0 \,:=\, \mathrm{L}^2(\Omega_{\mathrm{B}})\times\mathrm{L}^2(\Omega_{\mathrm{D}})\times\mathrm{H}^{1/2}_0(\Sigma)$, where $\mathrm{H}^{1/2}_0(\Sigma) := \big\{\xi\in\mathrm{H}^{1/2}(\Sigma): \quad \langle 1,\xi\rangle_\Sigma = 0\big\}$.

Throughout the rest of the chapter we stay with (5.27) and consider, instead of (5.15), the following mixed formulation: Find $\vec{\boldsymbol{u}} \,:=\, (\boldsymbol{u}_{\mathrm{B}},\boldsymbol{\omega}_{\mathrm{B}},\boldsymbol{u}_{\mathrm{D}})\in\mathbf{H}$ and $\vec{p} \,:=\, (p_{\mathrm{B}},p_{\mathrm{D}},\lambda)\in\mathbf{Q}_0$ such that

$$\begin{aligned} \mathbf{a}(\vec{\boldsymbol{u}},\vec{\boldsymbol{v}}) \,+\, \mathbf{b}(\vec{\boldsymbol{v}},\vec{p}) \;&=\; \mathbf{F}(\vec{\boldsymbol{v}}) \qquad \forall\,\vec{\boldsymbol{v}} \,:=\, (\boldsymbol{v}_{\mathrm{B}},\boldsymbol{z}_{\mathrm{B}},\boldsymbol{v}_{\mathrm{D}})\in\mathbf{H}\,,\\ \mathbf{b}(\vec{\boldsymbol{u}},\vec{q}) \;&=\; \mathbf{G}(\vec{q}) \qquad \forall\,\vec{q} \,:=\, (q_{\mathrm{B}},q_{\mathrm{D}},\xi)\in\mathbf{Q}_0\,. \end{aligned} \tag{5.28}$$

Note that the second equation of (5.15), which is tested against $\vec{q}\in\mathbf{Q}$, is equivalent to the present second equation of (5.28), which is tested against $\vec{q}\in\mathbf{Q}_0$. In fact, one implication is obvious because of the inclusion $\mathbf{Q}_0\subseteq\mathbf{Q}$. Conversely, assume that the second equation of (5.28) holds. Then, given $c\in\mathrm{R}$, we integrate by parts and, noting that $(\mathbf{0},q_{\mathrm{D}}-c,\xi-c)\in\mathbf{Q}_0 \,:=\, \mathrm{L}^2_0(\Omega_{\mathrm{B}})\times\mathrm{L}^2(\Omega_{\mathrm{D}})\times\mathrm{H}^{1/2}(\Sigma)$, we find that

$$\mathbf{b}(\vec{\boldsymbol{u}},(c,q_{\mathrm{D}},\xi)) \,=\, \mathbf{b}(\vec{\boldsymbol{u}},(\mathbf{0},q_{\mathrm{D}}-c,\xi-c)) \,=\, 0\,,$$

which yields $\mathbf{b}(\vec{\boldsymbol{u}},(q_{\mathrm{B}},q_{\mathrm{D}},\xi)) \,=\, 0 \,=\, \mathbf{G}(\vec{q})$ for all $\vec{q} \,:=\, (q_{\mathrm{B}},q_{\mathrm{D}},\xi)\in\mathbf{Q}$, thus confirming that the second equation of (5.15) holds.

We now aim to establish the well-posedness of (5.28) by applying the classical Babuška-Brezzi theory. We begin with the inf-sup condition for $\mathbf{b}$ on $\mathbf{H}\times\mathbf{Q}_0$.

**Lemma 5.4.** *There exists* $\widetilde{\beta} > 0$ *such that*

$$S(\vec{q}) := \sup_{\substack{\vec{v}\in\mathbf{H}\\ \vec{v}\neq\mathbf{0}}} \frac{\mathbf{b}(\vec{v},\vec{q})}{\|\vec{v}\|_{\mathbf{H}}} \geq \widetilde{\beta}\,\|\vec{q}\|_{\mathbf{Q}} \qquad \forall\,\vec{q}\in\mathbf{Q}_0\,. \tag{5.29}$$

*Proof.* Given $\vec{q} = (q_{\mathrm{B}}, q_{\mathrm{D}}, \xi) \in \mathbf{Q}_0 := \mathrm{L}^2_0(\Omega_{\mathrm{B}}) \times \mathrm{L}^2(\Omega_{\mathrm{D}}) \times \mathrm{H}^{1/2}(\Sigma)$, we obtain from Lemma 5.3 that

$$S(\vec{q}) := \sup_{\substack{\vec{v}\in\mathbf{H}\\ \vec{v}\neq\mathbf{0}}} \frac{\mathbf{b}(\vec{v},\vec{q})}{\|\vec{v}\|_{\mathbf{H}}} \geq \beta\left\{\|q_{\mathrm{B}}\|_{0,\Omega_{\mathrm{B}}} + \|q_{\mathrm{D},0}\|_{0,\Omega_{\mathrm{D}}} + \|\xi\|_{1/2,\Sigma} + \|\xi - c_{\mathrm{D}}\|_{1/2,\Sigma}\right\}, \tag{5.30}$$

where, according to (1.1), $q_{\mathrm{D}} = q_{\mathrm{D},0} + c_{\mathrm{D}}$, with $q_{\mathrm{D},0} \in \mathrm{L}^2_0(\Omega_{\mathrm{D}})$ and $c_{\mathrm{D}} := \frac{1}{|\Omega_{\mathrm{D}}|}\int_{\Omega_{\mathrm{D}}} q_{\mathrm{D}} \in \mathrm{R}$. In turn, a simple application of the triangle inequality shows that

$$|\Sigma|\,|c_{\mathrm{D}}| = \|c_{\mathrm{D}}\|_{1/2,\Sigma} \leq \|\xi\|_{1/2,\Sigma} + \|\xi - c_{\mathrm{D}}\|_{1/2,\Sigma}\,,$$

which, combined with (5.30) and the fact that $\|q_{\mathrm{D}}\|^2_{0,\Omega_{\mathrm{D}}} = \|q_{\mathrm{D},0}\|^2_{0,\Omega_{\mathrm{D}}} + |\Omega_{\mathrm{D}}|\,c^2_{\mathrm{D}}$ (*cf.* (5.4)), imply (5.29) and finish the proof.

□

Next, we address the coerciveness of $\mathbf{a}$ on the kernel $\mathbf{V}$ of $\mathbf{b}$. Indeed, we first deduce from the definitions of $\mathbf{b}$ (*cf.* (5.17)) and $\mathbf{Q}_0$ (*cf.* (5.27)) that

$$\mathbf{V} = \mathbf{V}_{\mathrm{B,D}} \cap \mathbf{V}_{\Sigma}\,, \tag{5.31}$$

with

$$\mathbf{V}_{\mathrm{B,D}} := \left\{\vec{v} := (\boldsymbol{v}_{\mathrm{B}}, \boldsymbol{z}_{\mathrm{B}}, \boldsymbol{v}_{\mathrm{D}}) \in \mathbf{H} : \quad \mathrm{div}\,\boldsymbol{v}_{\mathrm{B}} \in P_0(\Omega_{\mathrm{B}}) \quad \text{and} \quad \mathrm{div}\,\boldsymbol{v}_{\mathrm{D}} = 0 \text{ in } \Omega_{\mathrm{D}}\right\}, \tag{5.32}$$

and

$$\mathbf{V}_{\Sigma} := \left\{\vec{v} := (\boldsymbol{v}_{\mathrm{B}}, \boldsymbol{z}_{\mathrm{B}}, \boldsymbol{v}_{\mathrm{D}}) \in \mathbf{H} : \quad \boldsymbol{v}_{\mathrm{B}}\cdot\boldsymbol{n} = \boldsymbol{v}_{\mathrm{D}}\cdot\boldsymbol{n} \quad \text{on} \quad \Sigma\right\}. \tag{5.33}$$

**Lemma 5.5.** *There exists* $\widetilde{\varrho} > 0$ *such that*

$$\sup_{\substack{\vec{w}\in\mathbf{V}\\ \vec{w}\neq\mathbf{0}}} \frac{\mathbf{a}(\vec{v},\vec{w})}{\|\vec{w}\|_{\mathbf{H}}} \geq \widetilde{\varrho}\,\|\vec{v}\|_{\mathbf{H}} \qquad \forall\,\vec{v}\in\mathbf{V}\,, \tag{5.34}$$

*and*

$$\sup_{\substack{\vec{v}\in\mathbf{V}\\ \vec{v}\neq\mathbf{0}}} \frac{\mathbf{a}(\vec{v},\vec{w})}{\|\vec{v}\|_{\mathbf{H}}} \geq \widetilde{\varrho}\,\|\vec{w}\|_{\mathbf{H}} \qquad \forall\,\vec{w}\in\mathbf{V}\,, \tag{5.35}$$

*Proof.* We begin by recalling from [74, Lemma 3.2] that there exists $\varrho_0 > 0$ such that

$$\|\boldsymbol{v}_{\mathrm{B}}\|_{0,\Omega_{\mathrm{B}}} \geq \varrho_0\,\|\boldsymbol{v}_{\mathrm{B}}\|_{\mathrm{div},\Omega_{\mathrm{B}}} \quad \forall\,\boldsymbol{v}_{\mathrm{B}}\in\mathbf{H}(\mathrm{div};\Omega_{\mathrm{B}}) \quad \text{such that} \quad \mathrm{div}\,\boldsymbol{v}_{\mathrm{B}}\in P_0(\Omega_{\mathrm{B}})\,. \tag{5.36}$$

Hence, thanks to the foregoing inequality and (5.26), we find that

$$\mathbf{a}(\vec{v},\vec{v}) \geq \widetilde{\varrho}_1\left\{\|\boldsymbol{v}_{\mathrm{B}}\|^2_{\mathrm{div};\Omega_{\mathrm{B}}} + \|\boldsymbol{z}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} + \|\boldsymbol{v}_{\mathrm{D}}\|^2_{\mathrm{div};\Omega_{\mathrm{D}}}\right\} \quad \forall\,\vec{v} := (\boldsymbol{v}_{\mathrm{B}}, \boldsymbol{z}_{\mathrm{B}}, \boldsymbol{v}_{\mathrm{D}}) \in \mathbf{V}_{\mathrm{B,D}}\,, \tag{5.37}$$

with $\widetilde{\varrho}_1 := \min\{\alpha\varrho_0^2, \nu, \mu\} > 0$. Next, given a particular $\vec{\boldsymbol{v}} := (\boldsymbol{v}_{\mathrm{B}}, \boldsymbol{z}_{\mathrm{B}}, \boldsymbol{v}_{\mathrm{D}}) \in \mathbf{V}$, we certainly have $\boldsymbol{z}_{\mathrm{B}} \in \mathbf{H}_0(\mathbf{curl}; \Omega_{\mathrm{B}})$, and thus, due to a well-known result (see, e.g. [77, Chapter I, Section 2.3, Remark 2.5]), there holds $\mathbf{curl}\, \boldsymbol{z}_{\mathrm{B}} \in \mathbf{H}_0(\mathrm{div}; \Omega_{\mathrm{B}})$, where

$$\mathbf{H}_0(\mathrm{div}; \Omega_{\mathrm{B}}) := \left\{ \boldsymbol{v}_{\mathrm{B}} \in \mathbf{H}(\mathrm{div}; \Omega_{\mathrm{B}}) : \quad \boldsymbol{v}_{\mathrm{B}} \cdot \boldsymbol{n} = 0 \quad \text{on} \quad \partial\Omega_{\mathrm{B}} \right\}.$$

In this way, denoting

$$T_0(\vec{\boldsymbol{v}}) := (\mathbf{curl}\, \boldsymbol{z}_{\mathrm{B}}, \boldsymbol{z}_{\mathrm{B}}, \mathbf{0}),$$

which clearly belongs to $\mathbf{V}$, we find, according to the definition of $\mathbf{a}$ (*cf.* (5.16)), that

$$\mathbf{a}(\vec{\boldsymbol{v}}, T_0(\vec{\boldsymbol{v}})) = (\alpha - \nu) \int_{\Omega_{\mathrm{B}}} \boldsymbol{v}_{\mathrm{B}} \cdot \mathbf{curl} \boldsymbol{z}_{\mathrm{B}} + \nu \, \|\mathbf{curl} \boldsymbol{z}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} + \nu \, \|\boldsymbol{z}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}},$$

which, applying Cauchy-Schwarz's inequality and simple algebraic manipulations, yields

$$\mathbf{a}(\vec{\boldsymbol{v}}, T_0(\vec{\boldsymbol{v}})) \geq -\frac{|\alpha - \nu|^2}{2\nu} \|\boldsymbol{v}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} + \frac{\nu}{2} \|\mathbf{curl} \boldsymbol{z}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} + \nu \, \|\boldsymbol{z}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}}. \tag{5.38}$$

Therefore, introducing now $T(\vec{\boldsymbol{v}}) := c\, \vec{\boldsymbol{v}} + c_0\, T_0(\vec{\boldsymbol{v}})$, with suitable chosen positive constants $c$ and $c_0$ (depending on $\widetilde{\varrho}_1$, $\alpha$, and $\nu$), and utilizing (5.37) and (5.38), we obtain that

$$T(\vec{\boldsymbol{v}}) \in \mathbf{V}, \quad \|T(\vec{\boldsymbol{v}})\|_{\mathbf{H}} \leq C \, \|\vec{\boldsymbol{v}}\|_{\mathbf{H}}, \quad \text{and} \quad \mathbf{a}(\vec{\boldsymbol{v}}, T(\vec{\boldsymbol{v}})) \geq \widetilde{\varrho}_2 \, \|\vec{\boldsymbol{v}}\|^2_{\mathbf{H}},$$

with $C$ and $\widetilde{\varrho}_2$ positive constants depending on $\widetilde{\varrho}_1$, $\alpha$, and $\nu$ as well. Then, we can write

$$\sup_{\substack{\vec{\boldsymbol{w}} \in \mathbf{V} \\ \vec{\boldsymbol{w}} \neq \mathbf{0}}} \frac{\mathbf{a}(\vec{\boldsymbol{v}}, \vec{\boldsymbol{w}})}{\|\vec{\boldsymbol{w}}\|_{\mathbf{H}}} \geq \frac{\mathbf{a}(\vec{\boldsymbol{v}}, T(\vec{\boldsymbol{v}}))}{\|T(\vec{\boldsymbol{v}})\|_{\mathbf{H}}},$$

which, due to the foregoing estimates, gives (5.34). On the other hand, introducing the operator $\widetilde{T} : \mathbf{H} \to \mathbf{H}$ as $\widetilde{T}(\vec{\boldsymbol{v}}) := (-\boldsymbol{v}_{\mathrm{B}}, \boldsymbol{z}_{\mathrm{B}}, -\boldsymbol{v}_{\mathrm{D}}) \quad \forall\, \vec{\boldsymbol{v}} := (\boldsymbol{v}_{\mathrm{B}}, \boldsymbol{z}_{\mathrm{B}}, \boldsymbol{v}_{\mathrm{D}}) \in \mathbf{H}$, we realize that $\|\widetilde{T}(\vec{\boldsymbol{v}})\|_{\mathbf{H}} = \|\vec{\boldsymbol{v}}\|_{\mathbf{H}}$, $\mathbf{a}(\vec{\boldsymbol{v}}, \vec{\boldsymbol{w}}) = \mathbf{a}(\widetilde{T}(\vec{\boldsymbol{w}}), \widetilde{T}(\vec{\boldsymbol{v}})) \quad \forall\, \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \in \mathbf{H}$, $\widetilde{T}(\vec{\boldsymbol{v}}) \in \mathbf{V} \quad \forall\, \vec{\boldsymbol{v}} \in \mathbf{V}$, and $\widetilde{T} : \mathbf{V} \to \mathbf{V}$ is an isomorphism. It follows easily that

$$\sup_{\substack{\vec{\boldsymbol{v}} \in \mathbf{V} \\ \vec{\boldsymbol{v}} \neq \mathbf{0}}} \frac{\mathbf{a}(\vec{\boldsymbol{v}}, \vec{\boldsymbol{w}})}{\|\vec{\boldsymbol{v}}\|_{\mathbf{H}}} = \sup_{\substack{\vec{\boldsymbol{v}} \in \mathbf{V} \\ \vec{\boldsymbol{v}} \neq \mathbf{0}}} \frac{\mathbf{a}(\widetilde{T}(\vec{\boldsymbol{w}}), \widetilde{T}(\vec{\boldsymbol{v}}))}{\|\widetilde{T}(\vec{\boldsymbol{v}})\|_{\mathbf{H}}} = \sup_{\substack{\vec{\boldsymbol{v}} \in \mathbf{V} \\ \vec{\boldsymbol{v}} \neq \mathbf{0}}} \frac{\mathbf{a}(\widetilde{T}(\vec{\boldsymbol{w}}), \vec{\boldsymbol{v}})}{\|\vec{\boldsymbol{v}}\|_{\mathbf{H}}},$$

which, thanks to (5.34), yields (5.35) and completes the proof.

□

As a consequence of the previous analysis we can state the following main result.

**Theorem 5.2.** *Assume that* $\mathbf{f}_{\mathrm{D}} \in \mathbf{L}^2(\Omega_{\mathrm{D}})$ *and* $\mathbf{f}_{\mathrm{B}} \in \mathbf{L}^2(\Omega_{\mathrm{B}})$. *Then there exists a unique* $(\vec{\boldsymbol{u}}, \vec{p}) := \big((\boldsymbol{u}_{\mathrm{B}}, \boldsymbol{\omega}_{\mathrm{B}}, \boldsymbol{u}_{\mathrm{D}}), (p_{\mathrm{B}}, p_{\mathrm{D}}, \lambda)\big) \in \mathbf{H} \times \mathbf{Q}_0$ *solution of the modified mixed formulation* (5.28). *Moreover, there exists* $C > 0$ *such that*

$$\|\vec{\boldsymbol{u}}\|_{\mathbf{H}} + \|\vec{p}\|_{\mathbf{Q}} \leq C \left\{ \|\boldsymbol{f}_{\mathrm{D}}\|_{0,\Omega_{\mathrm{D}}} + \|\boldsymbol{f}_{\mathrm{B}}\|_{0,\Omega_{\mathrm{B}}} \right\}. \tag{5.39}$$

*Proof.* Thanks to Lemma 5.4 and 5.5, the proof is a straightforward application of the continuous Babuška-Brezzi theory. In particular, it is clear from the definition of $\mathbf{F}$ (*cf.* (5.18)) that $\|\mathbf{F}\|_{\mathbf{H}'}$ is bounded by the right hand side of (5.39).

□

We end this section by remarking that the way of proving the inf-sup conditions for the bilinear form $\mathbf{a}$ (*cf.* Lemma 5.5), namely using suitable operators $T$ and $\widetilde{T}$ to get a lower bound of the suprema involved, corresponds basically to what has been recently denominated in the literature as $T$-coercivity (see, e.g. [22] and [39]). Nevertheless, the same idea, without any particular name of it, had been employed previously at least in the context of fluid-solid interaction problems (see, e.g. [69, 70], and [71]).

## 5.4 The mixed finite element method

In this section we introduce and analyze a mixed finite element scheme for (5.28). More precisely, we first define the associated Galerkin scheme and establish suitable assumptions on the finite element subspaces ensuring that it becomes well posed. Then, we provide specific examples satisfying the required hypotheses. In what follows, given an integer $k \geq 0$ and a subset $S$ of $\mathrm{R}^3$, we denote by $P_k(S)$ the space of polynomials in $S$ of total degree $\leq k$. In addition, according to the notation introduced in Section 5.1, we let $\mathbf{P}_k(S) = [P_k(S)]^3$.

### 5.4.1 Preliminaries and main results

We begin by selecting a set of arbitrary discrete spaces, namely

$$\begin{array}{lll} \mathbf{H}^{\mathrm{B}_h} \subseteq \mathbf{H}_{\mathrm{B}}(\mathrm{div};\Omega_{\mathrm{B}})\,, & \mathbf{H}^{\mathrm{B}}_{0,h} \subseteq \mathbf{H}_0(\mathbf{curl};\Omega_{\mathrm{B}})\,, & \mathbf{H}^{\mathrm{D}_h} \subseteq \mathbf{H}_{\mathrm{D}}(\mathrm{div};\Omega_{\mathrm{D}})\,, \\ \mathrm{Q}^{\mathrm{B}_h} \subseteq \mathrm{L}^2(\Omega_{\mathrm{B}})\,, & \mathrm{Q}^{\mathrm{D}}_h \subseteq \mathrm{L}^2(\Omega_{\mathrm{D}})\,, \quad \text{and} & \mathrm{Q}^{\Sigma}_h \subseteq \mathrm{H}^{1/2}(\Sigma)\,. \end{array} \tag{5.40}$$

In addition, in order to deal with the mean value condition for the Brinkman pressure $p_{\mathrm{B}}$, and also to handle the assumptions guaranteeing the discrete inf-sup condition for $\mathbf{b}$, we need to define

$$\mathrm{Q}^{\mathrm{B}}_{h,0} := \mathrm{Q}^{\mathrm{B}}_h \cap \mathrm{L}^2_0(\Omega_{\mathrm{B}}) \quad \text{and} \quad \mathrm{Q}^{\mathrm{D}}_{h,0} := \mathrm{Q}^{\mathrm{D}}_h \cap \mathrm{L}^2_0(\Omega_{\mathrm{D}})\,. \tag{5.41}$$

Hence, setting the global spaces

$$\mathbf{H}_h := \mathbf{H}^{\mathrm{B}}_h \times \mathbf{H}^{\mathrm{B}}_{0,h} \times \mathbf{H}^{\mathrm{D}}_h \quad \text{and} \quad \mathbf{Q}_{0,h} := \mathrm{Q}^{\mathrm{B}}_{h,0} \times \mathrm{Q}^{\mathrm{D}}_h \times \mathrm{Q}^{\Sigma}_h\,, \tag{5.42}$$

the Galerkin scheme for (5.28) becomes: Find $\vec{\boldsymbol{u}}_h := (\boldsymbol{u}^{\mathrm{B}}_h, \boldsymbol{\omega}^{\mathrm{B}}_h, \boldsymbol{u}^{\mathrm{D}}_h) \in \mathbf{H}_h$ and $\vec{p}_h := (p^{\mathrm{B}}_h, p^{\mathrm{D}}_h, \lambda_h) \in \mathbf{Q}_{0,h}$ such that

$$\begin{array}{lcll} \mathbf{a}(\vec{\boldsymbol{u}}_h, \vec{\boldsymbol{v}}_h) + \mathbf{b}(\vec{\boldsymbol{v}}_h, \vec{p}_h) & = & \mathbf{F}(\vec{\boldsymbol{v}}_h) & \forall\, \vec{\boldsymbol{v}}_h := (\boldsymbol{v}^{\mathrm{B}}_h, \boldsymbol{z}^{\mathrm{B}}_h, \boldsymbol{v}^{\mathrm{D}}_h) \in \mathbf{H}_h\,, \\ \mathbf{b}(\vec{\boldsymbol{u}}_h, \vec{q}_h) & = & \mathbf{G}(\vec{q}_h) & \forall\, \vec{q}_h := (q^{\mathrm{B}}_h, q^{\mathrm{D}}_h, \xi_h) \in \mathbf{Q}_{0,h}\,. \end{array} \tag{5.43}$$

We now aim to derive general hypotheses on the finite element subspaces introduced in (5.40) ensuring, by means of the discrete Babuška-Brezzi theory, that the Galerkin scheme (5.43) becomes well-posed. Our approach consists of adapting to the present discrete case the arguments employed in Section 5.3 for the analysis of the continuous problem, mainly those from the proofs of Lemmas 5.4 and 5.5. We begin by observing that in order to have meaningful spaces $\mathrm{Q}^{\mathrm{B}}_{h,0}$ and $\mathrm{Q}^{\mathrm{D}}_{h,0}$ (*cf.* (5.41)), we need to be able to eliminate constants polynomials from $\mathrm{Q}^{\mathrm{B}}_h$ and $\mathrm{Q}^{\mathrm{D}}_h$. This request is certainly satisfied if we assume that:

**(H.0)** $P_0(\Omega_{\mathrm{B}}) \subseteq \mathrm{Q}^{\mathrm{B}}_h$ *and* $P_0(\Omega_{\mathrm{D}}) \subseteq \mathrm{Q}^{\mathrm{D}}_h$,

which, in turn, yields the analogue orthogonal decompositions suggested by (1.1), that is

$$\mathrm{Q}_h^{\mathrm{B}} \,=\, \mathrm{Q}_{h,0}^{\mathrm{B}} \,\oplus\, P_0(\Omega_{\mathrm{B}}) \quad \text{and} \quad \mathrm{Q}_h^{\mathrm{D}} \,=\, \mathrm{Q}_{h,0}^{\mathrm{D}} \,\oplus\, P_0(\Omega_{\mathrm{D}})\,. \tag{5.44}$$

Next, according to the same arguments utilized in the proof of Lemma 5.4, which actually are determined by those employed in the proofs of Lemmata 5.1, 5.2, and 5.3, we realize that in order to show the discrete inf-sup condition for $\mathbf{b}$ on $\mathbf{H}_h \times \mathbf{Q}_{0,h}$, we need to assume the following hypothesis:

**(H.1)** *there holds* $P_0(\Sigma) \subseteq \mathrm{Q}_h^{\Sigma}$ *and there exist* $\widetilde{\beta}_{\mathrm{B}}$, $\widetilde{\beta}_{\mathrm{D}} > 0$, *independent of* $h$, *such that*

$$S_h^{\mathrm{B}}(q_h^{\mathrm{B}}, \xi_h) \,:=\, \sup_{\substack{\boldsymbol{v}_h^{\mathrm{B}} \in \mathbf{H}_h^{\mathrm{B}} \\ \boldsymbol{v}_h^{\mathrm{B}} \neq \mathbf{0}}} \frac{\displaystyle\int_{\Omega_{\mathrm{B}}} q_h^{\mathrm{B}} \,\mathrm{div}\, \boldsymbol{v}_h^{\mathrm{B}} \,-\, \langle \boldsymbol{v}_h^{\mathrm{B}} \cdot \boldsymbol{n}, \xi_h \rangle_{\Sigma}}{\|\boldsymbol{v}_h^{\mathrm{B}}\|_{\mathrm{div};\Omega_{\mathrm{B}}}} \,\geq\, \widetilde{\beta}_{\mathrm{B}} \left\{ \|q_h^{\mathrm{B}}\|_{0,\Omega_{\mathrm{B}}} \,+\, \|\xi_h\|_{1/2,\Sigma} \right\} \tag{5.45}$$

*for all* $(q_h^{\mathrm{B}}, \xi_h) \in \mathrm{Q}_{h,0}^{\mathrm{B}} \times \mathrm{Q}_h^{\Sigma}$, *and*

$$S_h^{\mathrm{D}}(q_h^{\mathrm{D}}, \xi_h) \,:=\, \sup_{\substack{\boldsymbol{v}_h^{\mathrm{D}} \in \mathbf{H}_h^{\mathrm{D}} \\ \boldsymbol{v}_h^{\mathrm{D}} \neq \mathbf{0}}} \frac{\displaystyle\int_{\Omega_{\mathrm{D}}} q_h^{\mathrm{D}} \,\mathrm{div}\, \boldsymbol{v}_h^{\mathrm{D}} \,+\, \langle \boldsymbol{v}_h^{\mathrm{D}} \cdot \boldsymbol{n}, \xi_h \rangle_{\Sigma}}{\|\boldsymbol{v}_h^{\mathrm{D}}\|_{\mathrm{div};\Omega_{\mathrm{D}}}} \,\geq\, \widetilde{\beta}_{\mathrm{D}} \left\{ \|q_h^{\mathrm{D}}\|_{0,\Omega_{\mathrm{D}}} \,+\, \|\xi_h\|_{1/2,\Sigma} \right\} \tag{5.46}$$

*for all* $(q_h^{\mathrm{D}}, \xi_h) \in \mathrm{Q}_{h,0}^{\mathrm{D}} \times \mathrm{Q}_h^{\Sigma}$.

On the other hand, we now look at the discrete kernel of $\mathbf{b}$, which is defined by

$$\mathbf{V}_h \,:=\, \left\{ \vec{\boldsymbol{v}}_h \,:=\, (\boldsymbol{v}_h^{\mathrm{B}}, \boldsymbol{z}_h^{\mathrm{B}}, \boldsymbol{v}_h^{\mathrm{D}}) \in \mathbf{H}_h : \quad \mathbf{b}(\vec{\boldsymbol{v}}_h, \vec{q}_h) = 0 \quad \forall\, \vec{q}_h \,:=\, (q_h^{\mathrm{B}}, q_h^{\mathrm{D}}, \xi_h) \in \mathbf{Q}_{0,h} \right\}. \tag{5.47}$$

Actually, in order to have a more explicit definition of $\mathbf{V}_h$, similarly as obtained for the continuous kernel $\mathbf{V}$ (*cf.* (5.31)), we now introduce the following assumption

**(H.2)** $\mathrm{div}\, \mathbf{H}_h^{\mathrm{B}} \subseteq \mathrm{Q}_h^{\mathrm{B}}$ *and* $\mathrm{div}\, \mathbf{H}_h^{\mathrm{D}} \subseteq \mathrm{Q}_h^{\mathrm{D}}$,

which, together with (5.42) and (5.44), implies

$$\mathbf{V}_h \,=\, \mathbf{V}_{\mathrm{B,D}}^h \,\cap\, \mathbf{V}_{\Sigma}^h\,, \tag{5.48}$$

with

$$\mathbf{V}_{\mathrm{B,D}}^h \,:=\, \left\{ \vec{\boldsymbol{v}}_h \,:=\, (\boldsymbol{v}_h^{\mathrm{B}}, \boldsymbol{z}_h^{\mathrm{B}}, \boldsymbol{v}_h^{\mathrm{D}}) \in \mathbf{H}_h : \quad \mathrm{div}\, \boldsymbol{v}_h^{\mathrm{B}} \in P_0(\Omega_{\mathrm{B}}) \quad \text{and} \quad \mathrm{div}\, \boldsymbol{v}_h^{\mathrm{D}} = 0 \text{ in } \Omega_{\mathrm{D}} \right\}, \tag{5.49}$$

and

$$\mathbf{V}_{\Sigma}^h \,:=\, \left\{ \vec{\boldsymbol{v}}_h \,:=\, (\boldsymbol{v}_h^{\mathrm{B}}, \boldsymbol{z}_h^{\mathrm{B}}, \boldsymbol{v}_h^{\mathrm{D}}) \in \mathbf{H}_h : \quad \langle \boldsymbol{v}_h^{\mathrm{B}} \cdot \boldsymbol{n} - \boldsymbol{v}_h^{\mathrm{D}} \cdot \boldsymbol{n}, \xi_h \rangle_{\Sigma} = 0 \quad \forall\, \xi_h \in \mathrm{Q}_h^{\Sigma} \right\}. \tag{5.50}$$

Since $\mathbf{V}_{\mathrm{B,D}}^h \subseteq \mathbf{V}_{\mathrm{B,D}}$ (*cf.* (5.32)), it is clear that inequality (5.37) is also valid in $\mathbf{V}_{\mathrm{B,D}}^h$ and hence in the discrete kernel $\mathbf{V}_h$. Consequently, in order to show the discrete coerciveness of $\mathbf{a}$ on $\mathbf{V}_h$ by adapting the procedure utilized in the proof of Lemma 5.5, it just remains to assume the following hypothesis

**(H.3)** $\mathbf{curl}\,\mathbf{H}^{\mathrm{B}}_{0,h} \subseteq \mathbf{H}^{\mathrm{B}}_{h}$.

Having established hypotheses **(H.0)**, **(H.1)**, **(H.2)**, and **(H.3)**, we now reconfirm that they suffice to show that our Galerkin scheme (5.43) is well-posed and convergent. We begin with the discrete inf-sup condition for $\mathbf{b}$.

**Lemma 5.6.** *There exists $\widehat{\beta} > 0$, independent of $h$, such that*

$$S_h(\vec{q}_h) := \sup_{\substack{\vec{\boldsymbol{v}}_h \in \mathbf{H}_h \\ \vec{\boldsymbol{v}}_h \neq \mathbf{0}}} \frac{\mathbf{b}(\vec{\boldsymbol{v}}_h, \vec{q}_h)}{\|\vec{\boldsymbol{v}}_h\|_{\mathbf{H}}} \geq \widehat{\beta}\,\|\vec{q}_h\|_{\mathbf{Q}} \qquad \forall\, \vec{q}_h \in \mathbf{Q}_{0,h}\,. \tag{5.51}$$

*Proof.* Given $\vec{q}_h := (q^{\mathrm{B}}_h, q^{\mathrm{D}}_h, \xi_h) \in \mathbf{Q}_{0,h}$, we let $q^{\mathrm{D}}_{h,0} \in \mathrm{Q}^{\mathrm{D}}_{h,0}$ and $c_{\mathrm{D}} \in \mathrm{R}$ such that $q^{\mathrm{D}}_h = q^{\mathrm{D}}_{h,0} + c_{\mathrm{D}}$. Then, reasoning as in the proof of Lemma 5.3, which in this case reduces to integrate by parts in $\Omega_{\mathrm{D}}$ only (since $q^{\mathrm{B}}_h$ is already in $\mathrm{Q}^{\mathrm{B}}_{h,0}$), we find, using the notations from **(H.1)**, that

$$S_h(\vec{q}_h) \geq S^{\mathrm{B}}_h(q^{\mathrm{B}}_h, \xi_h) \quad \text{and} \quad S_h(\vec{q}_h) \geq S^{\mathrm{D}}_h(q^{\mathrm{D}}_{h,0}, \xi_h - c_{\mathrm{D}})\,.$$

In this way, since thanks to the first assumption in **(H.1)** we have that $\xi_h - c_{\mathrm{D}}$ belongs to $\mathrm{Q}^{\Sigma}_h$, the foregoing inequalities and a straightforward application of the discrete inf-sup conditions (5.45) and (5.46), imply

$$S_h(\vec{q}_h) \geq \frac{1}{2}\,(\widetilde{\beta}_{\mathrm{B}} + \widetilde{\beta}_{\mathrm{D}})\,\Big\{\|q^{\mathrm{B}}_h\|_{0,\Omega_{\mathrm{B}}} + \|\xi_h\|_{1/2,\Sigma} + \|q^{\mathrm{D}}_{h,0}\|_{0,\Omega_{\mathrm{D}}} + \|\xi_h - c_{\mathrm{D}}\|_{1/2,\Sigma}\Big\}\,.$$

The proof is concluded by employing the triangle inequality, exactly as we did for Lemma 5.4. □

The discrete inf-sup condition for $\mathbf{a}$ on $\mathbf{V}_h$ is proved next. Since $\mathbf{V}_h$ is finite dimensional, it suffices to show one of the discrete analogues of the inequalities provided in Lemma 5.5.

**Lemma 5.7.** *There exists $\widehat{\varrho} > 0$, independent of $h$, such that*

$$\sup_{\substack{\vec{\boldsymbol{w}}_h \in \mathbf{V}_h \\ \vec{\boldsymbol{w}}_h \neq \mathbf{0}}} \frac{\mathbf{a}(\vec{\boldsymbol{v}}_h, \vec{\boldsymbol{w}}_h)}{\|\vec{\boldsymbol{w}}_h\|_{\mathbf{H}}} \geq \widehat{\varrho}\,\|\vec{\boldsymbol{v}}_h\|_{\mathbf{H}} \qquad \forall\, \vec{\boldsymbol{v}}_h \in \mathbf{V}_h\,. \tag{5.52}$$

*Proof.* Given $\vec{\boldsymbol{v}}_h := (\boldsymbol{v}^{\mathrm{B}}_h, \boldsymbol{z}^{\mathrm{B}}_h, \boldsymbol{v}^{\mathrm{D}}_h) \in \mathbf{V}_h$, we know from (5.37) that

$$\mathbf{a}(\vec{\boldsymbol{v}}_h, \vec{\boldsymbol{v}}_h) \geq \widetilde{\varrho}_1\,\Big\{\|\boldsymbol{v}^{\mathrm{B}}_h\|^2_{\mathrm{div};\Omega_{\mathrm{B}}} + \|\boldsymbol{z}^{\mathrm{B}}_h\|^2_{0,\Omega_{\mathrm{B}}} + \|\boldsymbol{v}^{\mathrm{D}}_h\|^2_{\mathrm{div};\Omega_{\mathrm{D}}}\Big\}\,.$$

In addition, thanks to the result in [77, Chapter I, Section 2.3, Remark 2.5] and our assumption **(H.3)**, we find that $\mathbf{curl}\,\boldsymbol{z}^{\mathrm{B}}_h \in \mathbf{H}_0(\mathrm{div};\Omega_{\mathrm{B}}) \cap \mathbf{H}^{\mathrm{B}}_h$, and hence $T_0(\vec{\boldsymbol{v}}_h) := (\mathbf{curl}\,\boldsymbol{z}^{\mathrm{B}}_h, \boldsymbol{z}^{\mathrm{B}}_h, \mathbf{0})$ clearly belongs to $\mathbf{V}_h$ (*cf.* (5.48)). The rest of the proof proceeds as in Lemma 5.5. Moreover, it is easy to realize that the constants $c$ and $c_0$ defining now $T(\vec{\boldsymbol{v}}_h) := c\,\vec{\boldsymbol{v}}_h + c_0\,T_0(\vec{\boldsymbol{v}}_h)$ can be taken exactly as those chosen in the proof of that lemma, so that the resulting constant $\widehat{\varrho}$ of the present result coincides with $\widetilde{\varrho}$ in (5.34) and (5.35). □

The following main result is a direct consequence of the previous analysis.

**Theorem 5.3.** *Assume that* $\mathbf{f}_{\mathrm{D}} \in \mathbf{L}^2(\Omega_{\mathrm{D}})$ *and* $\mathbf{f}_{\mathrm{B}} \in \mathbf{L}^2(\Omega_{\mathrm{B}})$. *In addition, suppose that* **(H.0)**, **(H.1)**, **(H.2)**, *and* **(H.3)** *hold. Then there exists a unique* $(\vec{\boldsymbol{u}}_h, \vec{p}_h) := \big((\boldsymbol{u}_h^{\mathrm{B}}, \boldsymbol{\omega}_h^{\mathrm{B}}, \boldsymbol{u}_h^{\mathrm{D}}), (p_h^{\mathrm{B}}, p_h^{\mathrm{D}}, \lambda_h)\big) \in \mathbf{H}_h \times \mathbf{Q}_{0,h}$ *solution of the Galerkin scheme* (5.43). *Moreover, there exist* $C_1$, $C_2 > 0$, *independent of* $h$, *such that*

$$\|\vec{\boldsymbol{u}}_h\|_{\mathbf{H}} + \|\vec{p}_h\|_{\mathbf{Q}} \leq C_1 \Big\{ \|\boldsymbol{f}_{\mathrm{D}}\|_{0,\Omega_{\mathrm{D}}} + \|\boldsymbol{f}_{\mathrm{B}}\|_{0,\Omega_{\mathrm{B}}} \Big\}, \tag{5.53}$$

*and*

$$\|(\vec{\boldsymbol{u}}, \vec{p}) - (\vec{\boldsymbol{u}}_h, \vec{p}_h)\|_{\mathbf{H}\times\mathbf{Q}} \leq C_2 \Big\{ \mathrm{dist}(\vec{\boldsymbol{u}}, \mathbf{H}_h) + \mathrm{dist}(\vec{p}, \mathbf{Q}_{0,h}) \Big\}. \tag{5.54}$$

*Proof.* Thanks to Lemma 5.6 and 5.7, the proof results as a straightforward application of the discrete Babuška-Brezzi theory. □

### 5.4.2 Specific finite element subspaces

We now specify concrete examples of finite element subspaces satisfying the hypotheses introduced in the previous section. For this purpose, we now let $\mathcal{T}_h$ be a regular family of triangulations of $\bar{\Omega}_{\mathrm{B}} \cup \bar{\Omega}_{\mathrm{D}}$ by tetrahedra $K$ of diameter $h_K$ with mesh size $h := \max\{h_K : \quad K \in \mathcal{T}_h\}$, such that $\mathcal{T}_h(\Omega_\star) := \Big\{K \in \mathcal{T}_h : \quad K \subseteq \bar{\Omega}_\star\Big\}$ is a triangulation of $\Omega_\star$ for each $\star \in \big\{\mathrm{B}, \mathrm{D}\big\}$. Then, we denote by $\mathcal{T}_h(\Sigma)$ the triangulation on $\Sigma$ induced by $\mathcal{T}_h$ (either from $\Omega_{\mathrm{B}}$ or $\Omega_{\mathrm{D}}$). Also, for reasons that will become clear below, we introduce an independent triangulation $\mathcal{T}_{\tilde{h}}(\Sigma)$ of $\Sigma$ by triangles $\widetilde{T}$ of diameter $h_{\widetilde{T}}$, and define $\tilde{h} := \max\big\{h_{\widetilde{T}} : \quad \widetilde{T} \in \mathcal{T}_{\tilde{h}}(\Sigma)\big\}$.

#### 5.4.2.1 Definition of subspaces

We first introduce the finite element subspaces

$$\begin{aligned}
\mathbf{H}_h^\star &:= \Big\{ \boldsymbol{v}_h^\star \in \mathbf{H}_\star(\mathrm{div}; \Omega_\star) : \quad \boldsymbol{v}_h^\star|_K \in \mathbb{RT}_0(K) \quad \forall\, K \in \mathcal{T}_h(\Omega_\star) \Big\}, \\
\mathrm{Q}_h^\star &:= \Big\{ q_h \in \mathrm{L}^2(\Omega_\star) : \quad q_h|_K \in P_0(K) \quad \forall\, K \in \mathcal{T}_h(\Omega_\star) \Big\}, \\
\mathrm{Q}_{h,0}^\star &:= \mathrm{Q}_h^\star \cap \mathrm{L}_0^2(\Omega_\star),
\end{aligned}$$

where $\star \in \{\mathrm{B}, \mathrm{D}\}$, and for any $K \in \mathcal{T}_h(\Omega_\star)$

$$\mathbb{RT}_0(K) := \mathbf{P}_0(K) \oplus P_0(K)\, \boldsymbol{x}$$

is the local Raviart-Thomas space of lowest order. In addition, we set

$$\mathbf{H}_{0,h}^{\mathrm{B}} := \Big\{ \boldsymbol{z}_h^{\mathrm{B}} \in \mathbf{H}_0(\mathbf{curl}; \Omega_{\mathrm{B}}) : \quad \boldsymbol{z}_h^{\mathrm{B}}|_K \in \mathbb{ND}_1(K) \quad \forall\, K \in \mathcal{T}_h(\Omega_{\mathrm{B}}) \Big\},$$

where for any $K \in \mathcal{T}_h(\Omega_{\mathrm{B}})$

$$\mathbb{ND}_1(K) := \mathbf{P}_0(K) \oplus \mathbf{P}_0(K) \times \boldsymbol{x}$$

is the local edge space of Nédélec, that is

$$\mathbb{ND}_1(K) := \Big\{ w : K \to \mathbb{C}^3 : \quad w(\boldsymbol{x}) = \boldsymbol{a} + \boldsymbol{b} \times \boldsymbol{x} \quad \forall\, \boldsymbol{x} \in K,\, \boldsymbol{a},\, \boldsymbol{b} \in \mathbb{C}^3 \Big\}.$$

Finally for the interface $\Sigma$ we consider the finite element subspace

$$\mathrm{Q}_{\tilde{h}}^{\Sigma} := \Big\{ \lambda_{\tilde{h}} \in \mathcal{C}^0(\Sigma) : \quad \lambda_{\tilde{h}}|_{\tilde{T}} \in P_1(\tilde{T}) \quad \forall\, \tilde{T} \in \mathcal{T}_{\tilde{h}}(\Sigma) \Big\}.$$

It is easy to check that these subspaces satisfy the hypotheses **(H.0)**, **(H.2)** and **(H.3)**.

On the other hand, for purposes of the analysis, we also need to define

$$\Phi_h(\Sigma) := \Big\{ \psi_h \in \mathrm{L}^2(\Sigma) : \quad \psi_h|_T \in P_0(T) \quad \forall\, T \in \mathcal{T}_h(\Sigma) \Big\}.$$

#### 5.4.2.2 Approximation properties

In what follows $\star$ is a mute symbol taken in $\{\mathrm{B}, \mathrm{D}\}$. We let $\Pi_h^\star : \mathbf{H}^1(\Omega_\star) \to \mathbf{H}_h^\star$ be the usual Raviart-Thomas interpolation operator, that is, given a sufficiently smooth vector field $\boldsymbol{v} : \Omega_\star \to \mathrm{R}^3$, we define $\Pi_h^\star(\boldsymbol{v})$ as the only element of $\mathbf{H}_h^\star$ such that

$$\int_F \Pi_h^\star(\boldsymbol{v}) \cdot \boldsymbol{n} = \int_F \boldsymbol{v} \cdot \boldsymbol{n} \quad \forall\, F \in \mathcal{E}_h^\star, \tag{5.55}$$

where $\mathcal{E}_h^\star$ is the set of faces of the triangulation $\mathcal{T}_h(\Omega_\star)$. We now recall some properties of $\Pi_h^\star$ and its local counterparts $\Pi_K^\star$ for each $K \in \mathcal{T}_h(\Omega_\star)$ (see, e.g [63]):

**(a)** $\Pi_h^\star$ is well defined in $\mathbf{H}^\delta(\Omega_\star) \cap \mathbf{H}(\mathrm{div}; \Omega_\star)$ for any $\delta \in (0, 1)$.

**(b)** There holds $\mathrm{div}\, \Pi_h^\star(\boldsymbol{v}) = \mathcal{P}_h^\star(\mathrm{div}\, \boldsymbol{v})$, where $\mathcal{P}_h^\star : \mathrm{L}^2(\Omega_\star) \to \mathrm{Q}_h^\star$ is the orthogonal projector. Equivalently

$$\int_{\Omega_\star} q_h\, \mathrm{div}\, \Pi_h^\star(\boldsymbol{v}) = \int_{\Omega_\star} q_h\, \mathrm{div}\, (\boldsymbol{v}) \qquad \forall\, q_h \in \mathrm{Q}_h^\star.$$

**(c)** For each face $F$ of $K$ there holds $\Pi_K^\star(\boldsymbol{v}) \cdot \boldsymbol{n}_F = \mathcal{P}_F(\boldsymbol{v} \cdot \boldsymbol{n}_F)$, where $\boldsymbol{n}_F$ is the unit outward normal on $F$ and $\mathcal{P}_F : \mathrm{L}^2(F) \to P_0(F)$ is the orthogonal projector.

**(d)** Given $\delta \in (0, 1)$ and $\boldsymbol{v} \in \mathbf{H}^\delta(\Omega_\star) \cap \mathbf{H}(\mathrm{div}; \Omega_\star)$, there holds

$$\|\boldsymbol{v} - \Pi_K^\star(\boldsymbol{v})\|_{0,K} \leq C h_K^\delta \Big\{ |\boldsymbol{v}|_{\delta,K} + \|\mathrm{div}\, (\boldsymbol{v})\|_{0,K} \Big\} \qquad \forall\, K \in \mathcal{T}_h(\Omega_\star). \tag{5.56}$$

Next, for any $\epsilon > 0$ we introduce the Sobolev space

$$\mathbf{H}^\epsilon(\mathbf{curl}; \Omega_\mathrm{B}) := \left\{ \boldsymbol{v} \in \mathbf{H}^\epsilon(\Omega_\mathrm{B}) : \quad \mathbf{curl}\, \boldsymbol{v} \in \mathbf{H}^\epsilon(\Omega_\mathrm{B}) \right\},$$

and endow it with its Hilbertian norm

$$\|\boldsymbol{v}\|_{\mathbf{H}^\epsilon(\mathbf{curl};\Omega_\mathrm{B})} := \Big\{ \|\boldsymbol{v}\|_{\epsilon,\Omega_\mathrm{B}}^2 + \|\mathbf{curl}(\boldsymbol{v})\|_{\epsilon,\Omega_\mathrm{B}}^2 \Big\}^{1/2}.$$

Then for each face $F$ of $\mathcal{T}_h(\Omega_\mathrm{B})$ we let $\boldsymbol{t}_F$ be a unit tangential vector on $F$. It follows from [10, Lemma 4.7] that if $\epsilon > 1/2$ the interpolation operator $\Pi_h : \mathbf{H}^\epsilon(\mathbf{curl}; \Omega_\mathrm{B}) \to \mathbf{H}_{0,h}^\mathrm{B}$ associated with the face finite element, which is characterized by

$$\int_F \Pi_h(\boldsymbol{v}) \cdot \boldsymbol{t}_F = \int_F \boldsymbol{v} \cdot \boldsymbol{t}_F \quad \forall \text{ faces } F \text{ of } \mathcal{T}_h(\Omega_\mathrm{B}),$$

is well defined and uniformly bounded. In addition, the following property of $\Pi_h$ holds.

**Lemma 5.8.** *There exists $C > 0$, independent of $h$, such that*

$$\|\boldsymbol{v} - \Pi_h(\boldsymbol{v})\|_{\mathbf{curl};\Omega_\mathrm{B}} \leq C h^\epsilon \|\boldsymbol{v}\|_{\mathbf{H}^\epsilon(\mathbf{curl};\Omega_\mathrm{B})} \tag{5.57}$$

*for all $\boldsymbol{v} \in \mathbf{H}^\epsilon(\mathbf{curl}; \Omega_\mathrm{B})$ and for all $\epsilon \in (1/2, 1]$.*

*Proof.* See, [3, Proposition 5.6]. □

The approximation properties of the finite element subspaces involved are then established as follows (see, e.g [26, 87]):

($\mathrm{AP}_h^{\boldsymbol{u}_\star}$) there exists $C > 0$, independent of $h$, such that for each $\delta \in (0,1]$ and for each $\boldsymbol{v} \in \mathbf{H}^\delta(\Omega_\star)$, with $\mathrm{div}\,(\boldsymbol{v}) \in \mathbf{H}^\delta(\Omega_\star)$, there holds

$$\|\boldsymbol{v} - \Pi_h^\star(\boldsymbol{v})\|_{\mathrm{div};\Omega_\star} \leq C h^\delta \Big\{ \|\boldsymbol{v}\|_{\delta,\Omega_\star} + \|\mathrm{div}\,(\boldsymbol{v})\|_{\delta,\Omega_\star} \Big\} \qquad (\star \in \{\mathrm{B},\mathrm{D}\}).$$

($\mathrm{AP}_h^{p_\star}$) there exists $C > 0$, independent of $h$, such that for each $\delta \in (0,1]$ and for each $q \in \mathrm{H}^\delta(\Omega_\star)$, there holds

$$\|q - \mathcal{P}_h^\star(q)\|_{0,\Omega_\star} \leq C h^\delta \|q\|_{\delta,\Omega_\star} \qquad (\star \in \{\mathrm{B},\mathrm{D}\}).$$

($\mathrm{AP}_h^{\boldsymbol{\omega}_\mathrm{B}}$) there exists $C > 0$, independent of $h$, such that for each $\delta \in (1/2,1]$ and for each $\boldsymbol{z}_\mathrm{B} \in \mathbf{H}^\delta(\mathbf{curl};\Omega_\mathrm{B})$, there holds

$$\|\boldsymbol{z}_\mathrm{B} - \Pi_h(\boldsymbol{z}_\mathrm{B})\|_{\mathbf{curl};\Omega_\mathrm{B}} \leq C\, h^\delta\, \|\boldsymbol{z}_\mathrm{B}\|_{\mathbf{H}^\delta(\mathbf{curl};\Omega_\mathrm{B})}.$$

($\mathrm{AP}_{\tilde{h}}^{\lambda}$) there exists $C > 0$, independent of $\tilde{h}$, such that for each $\delta \in (0,1]$ and for each $\xi \in \mathrm{H}^{1/2+\delta}(\Sigma)$, there holds

$$\|\xi - \mathcal{P}_{\tilde{h}}(\xi)\|_{1/2,\Sigma} \leq C\, \tilde{h}^\delta\, \|\xi\|_{1/2+\delta,\Sigma},$$

where $\mathcal{P}_{\tilde{h}} : \mathrm{H}^{1/2}(\Sigma) \to \mathrm{Q}_{\tilde{h}}^\Sigma$ is the orthogonal projector.

($\mathrm{AP}_h{}^{\psi}$) there exists $C > 0$, independent of $h$, such that for each $\delta \in (0,1]$ and for each $\varphi \in \mathrm{H}^{-1/2+\delta}(\Sigma)$, there holds

$$\|\varphi - \mathcal{P}_h^{-1/2}(\varphi)\|_{-1/2,\Sigma} \leq C\, h_\Sigma^\delta\, \|\varphi\|_{-1/2+\delta,\Sigma},$$

where $\mathcal{P}_h^{-1/2} : \mathrm{H}^{-1/2}(\Sigma) \to \Phi_h(\Sigma)$ is the orthogonal projector.

#### 5.4.2.3 Stable discrete liftings

In this section, as usual we let $\star$ be a mute symbol taken in $\{\mathrm{B},\mathrm{D}\}$, and provide sufficient conditions for the existence of a stable discrete lifting $\mathcal{L}_h : \Phi_h(\Sigma) \to \mathbf{H}_h^\star$. To this end, we proceed as in [63, Theorem 4.1], and assume first that $\mathcal{T}_h(\Omega_\star)$ is quasi-uniform in a neighborhood of $\Sigma$. This means that there exists a neighborhood of $\Sigma$, say $\Omega_\Sigma$, and a constant $c > 0$, independent of $h$, such that, denoting

$$\mathcal{T}_{h,\Sigma}^\star := \{K \in \mathcal{T}_h(\Omega_\star) : \quad K \cap \Omega_\Sigma \neq \emptyset\}, \tag{5.58}$$

there holds

$$\max_{K \in \mathcal{T}_{h,\Sigma}^\star} h_K \leq c \min_{K \in \mathcal{T}_{h,\Sigma}^\star} h_K.$$

Now, because of the regularity of $\mathcal{T}_h(\Omega_\star)$, the quasi-uniformity assumption around $\Sigma$ implies that the partition $\mathcal{T}_h(\Sigma)$ inherited from $\mathcal{T}_h(\Omega_\star)$ is quasi-uniform as well, which implies that $\Phi_h(\Sigma)$ satisfies the inverse inequality (see, [63, Lemma 4.6])

$$\|\psi_h\|_{-1/2+\delta,\Sigma} \leq C\, h_\Sigma^{-\delta} \|\psi_h\|_{-1/2,\Sigma} \qquad \forall\, \psi_h \in \Phi_h(\Sigma), \qquad \forall\, \delta \in [0,1/2], \tag{5.59}$$

where $h_\Sigma := \max\big\{h_T : \quad T \in \mathcal{T}_h(\Sigma)\big\}$.

**Lemma 5.9.** *There exist a linear operator $\mathcal{L}_h : \Phi_h(\Sigma) \to \mathbf{H}^\star_h$ and a constant $C_\mathcal{L} > 0$, independent of $h$, such that for each $\psi_h \in \Phi_h(\Sigma)$ there hold*

$$\mathcal{L}_h(\psi_h)\cdot\boldsymbol{n} = \psi_h \quad \text{on} \quad \Sigma, \quad \|\mathcal{L}_h(\psi_h)\|_{\mathrm{div},\Omega_\star} \leq C_\mathcal{L}\,\|\psi_h\|_{-1/2,\Sigma}, \quad \text{and} \quad \mathrm{div}\,\mathcal{L}_h(\psi_h) \in P_0(\Omega_\star).$$

*Proof.* Let $\psi_h \in \Phi_h(\Sigma)$, and let $z \in \mathrm{H}^1(\Omega_\star)$ be the unique solution of the boundary value problem

$$\Delta z = \frac{1}{|\Omega_\star|}\langle \psi_h, 1\rangle_\Sigma \quad \text{in} \quad \Omega_\star, \qquad \nabla z\cdot\boldsymbol{n} = \left\{ \begin{array}{ccc} \psi_h & \text{on} & \Sigma \\ 0 & \text{on} & \Gamma_\star \end{array} \right. , \qquad \int_{\Omega_\star} z = 0.$$

The corresponding continuous dependence result says that $\|z\|_{1,\Omega_\star} \leq c_1\,\|\psi_h\|_{-1/2,\Sigma}$. In turn, the elliptic regularity result (*cf.* [79]) establishes that there exists $\delta > 0$ such that

$$\|z\|_{1+\delta,\Omega_\star} \leq c_2\,\|\psi_h\|_{-1/2+\delta,\Sigma}.$$

In addition, notice that

$$\mathrm{div}\,(\nabla z) = \Delta z = \frac{1}{|\Omega_\star|}\langle \psi_h, 1\rangle_\Sigma \in \mathrm{R} \quad \text{in} \quad \Omega_\star.$$

It follows that $\nabla z \in \mathbf{H}^\delta(\Omega_\star)\cap\mathbf{H}(\mathrm{div};\Omega_\star)$, and hence we can define

$$\mathcal{L}_h(\psi_h) := \Pi^\star_h(\nabla z).$$

Next, from properties **(b)** and **(c)** of the Raviart-Thomas interpolation operator, we find that

$$\mathrm{div}\,\mathcal{L}_h(\psi_h) = \mathrm{div}\,\Pi^\star_h(\nabla z) = \mathcal{P}^\star_h(\mathrm{div}\,\nabla z) = \mathcal{P}^\star_h(\Delta z) = \frac{1}{|\Omega_\star|}\langle \psi_h, 1\rangle_\Sigma \quad \text{in} \quad \Omega_\star, \tag{5.60}$$

and

$$\mathcal{L}_h(\psi_h)\cdot\boldsymbol{n} = \Pi^\star_h(\nabla z)\cdot\boldsymbol{n} = \mathcal{P}_{h,\Sigma}(\nabla z\cdot\boldsymbol{n}) = \mathcal{P}_{h,\Sigma}(\psi_h) = \psi_h \quad \text{on} \quad \Sigma\,, \tag{5.61}$$

where $\mathcal{P}_{h,\Sigma} : \mathrm{L}^2(\Sigma) \to \Phi_h(\Sigma)$ is the orthogonal projector. It remains to show that $\mathcal{L}_h$ is uniformly bounded. To this end, we first observe that

$$\|\mathcal{L}_h(\psi_h)\|^2_{\mathrm{div},\Omega_\star} = \|\mathcal{L}_h(\psi_h)\|^2_{0,\Omega_\star} + \Big\|\frac{1}{|\Omega_\star|}\langle \psi_h, 1\rangle_\Sigma\Big\|^2_{0,\Omega_\star} \leq \|\mathcal{L}_h(\psi_h)\|^2_{0,\Omega_\star} + c_3\,\|\psi_h\|^2_{-1/2,\Sigma}. \tag{5.62}$$

Now, we divide $\Omega_\star$ into two regions

$$\Omega^1_{\star,h} := \cup\Big\{K \in \mathcal{T}_h(\Omega_\star): \quad K \notin \mathcal{T}^\star_{h,\Sigma}\Big\} \subseteq \Omega_\star\backslash\Omega_\Sigma, \qquad \Omega^2_{\star,h} := \Omega_\star\backslash\Omega^1_{\star,h} = \cup\Big\{K \in \mathcal{T}^\star_{h,\Sigma}\Big\},$$

where we recall that $\mathcal{T}^\star_{h,\Sigma} := \Big\{K \in \mathcal{T}_h(\Omega_\star): \quad K\cap\Omega_\Sigma \neq \emptyset\Big\}$. Then, since $\Omega_\star\backslash\Omega_\Sigma$ is strictly contained in $\Omega_\star$, the interior elliptic regularity result [86, Theorem 4.16] implies that $z|_{\Omega_\star\backslash\Omega_\Sigma} \in \mathrm{H}^2(\Omega_\star\backslash\Omega_\Sigma)$ and

$$\|z\|_{2,\Omega_\star\backslash\Omega_\Sigma} \leq c_4\,\|\psi_h\|_{-1/2,\Sigma}.$$

It follows that

$$\begin{aligned}
\|\mathcal{L}_h(\psi_h)\|_{0,\Omega_\star} &\leq \|\mathcal{L}_h(\psi_h)\|_{0,\Omega^1_{\star,h}} + \|\mathcal{L}_h(\psi_h)\|_{0,\Omega^2_{\star,h}} \\
&= \|\Pi^\star_h(\nabla z)\|_{0,\Omega^1_{\star,h}} + \|\Pi^\star_h(\nabla z)\|_{0,\Omega^2_{\star,h}} \\
&\leq c_5\,\|\nabla z\|_{1,\Omega^1_{\star,h}} + \|\nabla z\|_{0,\Omega^2_{\star,h}} + \|\nabla z - \Pi^\star_h(\nabla z)\|_{0,\Omega^2_{\star,h}} \\
&\leq c_5\,\|z\|_{2,\Omega^1_{\star,h}} + \|z\|_{1,\Omega^2_{\star,h}} + \|\nabla z - \Pi^\star_h(\nabla z)\|_{0,\Omega^2_{\star,h}} \\
&\leq c_5\,c_4\,\|\psi_h\|_{-1/2,\Sigma} + c_1\,\|\psi_h\|_{-1/2,\Sigma} + \|\nabla z - \Pi^\star_h(\nabla z)\|_{0,\Omega^2_{\star,h}}.
\end{aligned} \tag{5.63}$$

On the other hand, applying estimate (5.56) and inverse inequality (5.59), we obtain that

$$\begin{aligned}
\|\nabla z - \Pi_h^\star(\nabla z)\|^2_{0,\Omega^2_{\star,h}} &= \sum_{K\in\mathcal{T}^\star_{h,\Sigma}} \|\nabla z - \Pi_K^\star(\nabla z)\|^2_{0,K} \\
&\leq c_6 \sum_{K\in\mathcal{T}^\star_{h,\Sigma}} h_K^{2\delta}\Big\{|\nabla z|^2_{\delta,K} + \Big\|\frac{1}{|\Omega_\star|}\langle\psi_h,1\rangle_\Sigma\Big\|^2_{0,K}\Big\} \\
&\leq c_7 \max_{K\in\mathcal{T}^\star_{h,\Sigma}} h_K^{2\delta}\Big\{\|z\|^2_{1+\delta,\Omega^2_{\star,h}} + \|\psi_h\|^2_{-1/2,\Sigma}\Big\} \\
&\leq c_7 \max_{K\in\mathcal{T}^\star_{h,\Sigma}} h_K^{2\delta}\Big\{\|z\|^2_{1+\delta,\Omega_\star} + \|\psi_h\|^2_{-1/2,\Sigma}\Big\} \\
&\leq c_8 \max_{K\in\mathcal{T}^\star_{h,\Sigma}} h_K^{2\delta}\Big\{\|\psi_h\|^2_{-1/2+\delta,\Sigma} + \|\psi_h\|^2_{-1/2,\Sigma}\Big\} \\
&\leq c_8 \max_{K\in\mathcal{T}^\star_{h,\Sigma}} h_K^{2\delta}\Big\{h_\Sigma^{-2\delta}\,\|\psi_h\|^2_{-1/2,\Sigma} + \|\psi_h\|^2_{-1/2,\Sigma}\Big\} \\
&\leq c_9\,\|\psi_h\|^2_{-1/2,\Sigma}\,,
\end{aligned} \tag{5.64}$$

where the fact that $h_K \leq c\,h_\Sigma \quad \forall\, K \in \mathcal{T}^\star_{h,\Sigma}$ has been used in the last inequality. In this way, from (5.62), (5.63) and (5.64) we conclude that

$$\|\mathcal{L}_h(\psi_h)\|_{\mathrm{div},\Omega_\star} \leq C_{\mathcal{L}}\,\|\psi_h\|_{-1/2,\Sigma} \qquad \forall\,\psi_h \in \Phi_h(\Sigma),$$

which, together with the identities (5.60) and (5.61), complete the proof. □

We remark at this point that the quasi-uniformity assumption of $\mathcal{T}_h(\Omega_*)$ around $\Sigma$, which is needed here for the stable discrete lifting provided by Lemma 5.9, has been removed recently in [2, Theorem 2.1] for the case of locally refined meshes, when the lifting is from the whole boundary of the given domain. However, it is not clear from the analysis in [2] whether that result is also valid for a discrete lifting from part of the boundary (as it is required in the present case).

We now assume that the family of independent triangulations $\mathcal{T}_{\tilde{h}}(\Sigma)$ is also quasi-uniform, which implies that $\mathrm{Q}^\Sigma_{\tilde{h}}$ satisfies the inverse inequality, that is there exists a constant $C > 0$, independent of $\tilde{h}$, such that for each $\delta \in [0,1)$ there holds (*cf.* [69, Lemma 7.4])

$$\|\xi\|_{1/2+\delta,\Sigma} \leq C\,\tilde{h}^{-\delta}\|\xi\|_{1/2,\Sigma} \qquad \forall\,\xi \in \mathrm{Q}^\Sigma_{\tilde{h}}. \tag{5.65}$$

Then, we have the following result.

**Lemma 5.10.** *There exist $C_0$, $\beta > 0$, independent of $h_\Sigma$ and $\tilde{h}$, such that for all $h_\Sigma \leq C_0\,\tilde{h}$ there holds*

$$\sup_{\substack{\psi_h\in\Phi_h(\Sigma)\\ \psi_h\neq 0}} \frac{\langle\psi_h,\xi_{\tilde{h}}\rangle_\Sigma}{\|\psi_h\|_{-1/2,\Sigma}} \geq \beta\,\|\xi_{\tilde{h}}\|_{1/2,\Sigma} \qquad \forall\,\xi_{\tilde{h}} \in \mathrm{Q}^\Sigma_{\tilde{h}}. \tag{5.66}$$

*Proof.* We proceed similarly as in [63, Lemma 4.11]. In fact, given $\xi_{\tilde{h}} \in \mathrm{Q}^\Sigma_{\tilde{h}}\backslash\{0\}$, we let $z \in \mathrm{H}^1(\Omega_\star)$ be the unique solution of the boundary value problem with mixed boundary conditions:

$$-\Delta z + z = 0 \quad \text{in} \quad \Omega_\star, \quad z = \xi_{\tilde{h}} \quad \text{on} \quad \Sigma, \quad \nabla z\cdot\boldsymbol{n} = 0 \quad \text{on} \quad \Gamma_\star.$$

Notice that the corresponding continuous dependence result gives

$$\|z\|_{1,\Omega_\star} \leq C_1\,\|\xi_{\tilde{h}}\|_{1/2,\Sigma}, \tag{5.67}$$

and thanks to the trace theorem and a simple integration by parts procedure, we also have that

$$C_2\,\|\xi_{\tilde{h}}\|^2_{1/2,\Sigma} \leq \|z\|^2_{1,\Omega_\star} \;=\; \langle \nabla z\cdot\boldsymbol{n},\xi_{\tilde{h}}\rangle_\Sigma. \tag{5.68}$$

On the other hand, since $\mathrm{Q}^\Sigma_{\tilde{h}} \subset \mathrm{H}^1(\Sigma)$, we obtain that $z \in \mathrm{H}^{1+\delta}(\Omega_\star)$ for some $\delta > 0$ (see [79]), and there holds

$$\|\nabla z\cdot\boldsymbol{n}\|_{-1/2+\delta,\Sigma} \leq C_3\|z\|_{1+\delta,\Omega_\star} \leq C_4\,\|\xi_{\tilde{h}}\|_{1/2+\delta,\Sigma}. \tag{5.69}$$

We now let $\psi^*_h := \mathcal{P}_h^{-1/2}(\nabla z\cdot\boldsymbol{n}) \in \Phi_h(\Sigma)$. Then, applying the approximation property $(\mathrm{AP}^\psi_h)$, the regularity estimate (5.69), and the inverse inequality (5.65), we deduce that

$$\|\nabla z\cdot\boldsymbol{n}-\psi^*_h\|_{-1/2,\Sigma} \leq\, C_5\,h^\delta_\Sigma\|\nabla z\cdot\boldsymbol{n}\|_{-1/2+\delta,\Sigma} \leq\, C_6\,h^\delta_\Sigma\|\xi_{\tilde{h}}\|_{1/2+\delta,\Sigma} \leq\, C_7\left(\frac{h_\Sigma}{\tilde{h}}\right)^\delta\|\xi_{\tilde{h}}\|_{1/2,\Sigma}.$$

Next, using that $\|\nabla z\|_{\mathrm{div},\Omega_\star} = \|z\|_{1,\Omega_\star}$, it follows that

$$\|\psi^*_h\|_{-1/2,\Sigma} \;=\; \|\mathcal{P}_h^{-1/2}(\nabla z\cdot\boldsymbol{n})\|_{-1/2,\Sigma} \leq\; \|\nabla z\cdot\boldsymbol{n}\|_{-1/2,\Sigma} \;\leq\; \|\nabla z\|_{\mathrm{div},\Omega_\star} \;=\; \|z\|_{1,\Omega_\star}\,,$$

which together with the estimate (5.67), imply

$$\|\psi^*_h\|_{-1/2,\Sigma} \leq C_8\,\|\xi_{\tilde{h}}\|_{1/2,\Sigma}.$$

Now, using (5.68) and the foregoing estimates, we find that

$$\begin{aligned}
\langle\psi^*_h,\xi_{\tilde{h}}\rangle_\Sigma &= \langle\nabla z\cdot\boldsymbol{n},\xi_{\tilde{h}}\rangle_\Sigma - \langle\nabla z\cdot\boldsymbol{n}-\psi^*_h,\xi_{\tilde{h}}\rangle_\Sigma\\
&\geq \left\{C_2 - C_7\left(\frac{h_\Sigma}{\tilde{h}}\right)^\delta\right\}\|\xi_{\tilde{h}}\|^2_{1/2,\Sigma}\\
&\geq \left\{\frac{C_2}{C_8} - \frac{C_7}{C_8}\left(\frac{h_\Sigma}{\tilde{h}}\right)^\delta\right\}\|\xi_{\tilde{h}}\|_{1/2,\Sigma}\,\|\psi^*_h\|_{-1/2,\Sigma}.
\end{aligned}$$

Consequently, we can write

$$\sup_{\substack{\psi_h\in\Phi_h(\Sigma)\\ \psi_h\neq 0}}\frac{\langle\psi_h,\xi_{\tilde{h}}\rangle_\Sigma}{\|\psi_h\|_{-1/2,\Sigma}} \;\geq\; \frac{\langle\psi^*_h,\xi_{\tilde{h}}\rangle_\Sigma}{\|\psi^*_h\|_{-1/2,\Sigma}} \;\geq\; \left\{\frac{C_2}{C_8} - \frac{C_7}{C_8}\left(\frac{h_\Sigma}{\tilde{h}}\right)^\delta\right\}\|\xi_{\tilde{h}}\|_{1/2,\Sigma},$$

from which, taking $h_\Sigma \leq C_0\,\tilde{h}$ with $C_0 := \left(\frac{C_2}{2C_7}\right)^{1/\delta}$, we conclude the proof. □

#### 5.4.2.4 Verification of the discrete inf-sup conditions

We are now in a position to prove the discrete inf-sup conditions required by hypotheses **(H.1)**. To this end, we assume from now on that $\mathcal{T}_h(\Omega_\mathrm{D})$ and $\mathcal{T}_h(\Omega_\mathrm{B})$ are quasi-uniform in a neighborhood $\Omega_\Sigma$ of $\Sigma$, and that $\mathcal{T}_{\tilde{h}}(\Sigma)$ is quasi-uniform.

**Lemma 5.11.** *There exist $C_0$, $\tilde{\beta}_\mathrm{D} > 0$, independent of $h$, $h_\Sigma$ and $\tilde{h}$, such that for all $h_\Sigma \leq C_0\,\tilde{h}$, there holds*

$$S^\mathrm{D}_h(q^\mathrm{D}_h,\xi_{\tilde{h}}) := \sup_{\substack{\boldsymbol{v}^\mathrm{D}_h\in\mathbf{H}^\mathrm{D}_h\\ \boldsymbol{v}^\mathrm{D}_h\neq\mathbf{0}}}\frac{\displaystyle\int_{\Omega_\mathrm{D}} q^\mathrm{D}_h\,\mathrm{div}\,\boldsymbol{v}^\mathrm{D}_h \;+\; \langle\boldsymbol{v}^\mathrm{D}_h\cdot\boldsymbol{n},\xi_{\tilde{h}}\rangle_\Sigma}{\|\boldsymbol{v}^\mathrm{D}_h\|_{\mathrm{div};\Omega_\mathrm{D}}} \geq \widetilde{\beta}_\mathrm{D}\left\{\|q^\mathrm{D}_h\|_{0,\Omega_\mathrm{D}} \;+\; \|\xi_{\tilde{h}}\|_{1/2,\Sigma}\right\} \tag{5.70}$$

*for all $(q^\mathrm{D}_h,\xi_{\tilde{h}}) \in \mathrm{Q}^\mathrm{D}_{h,0}\times\mathrm{Q}^\Sigma_{\tilde{h}}$.*

*Proof.* We begin by observing that

$$S_h^{\mathrm{D}}(q_h^{\mathrm{D}},\xi_{\tilde{h}}) \geq \sup_{\substack{\boldsymbol{v}_h^{\mathrm{D}}\in\mathbf{H}_h^{\mathrm{D}}\\ \boldsymbol{v}_h^{\mathrm{D}}\neq\mathbf{0}}} \frac{\displaystyle\int_{\Omega_{\mathrm{D}}} q_h^{\mathrm{D}}\,\mathrm{div}\,\boldsymbol{v}_h^{\mathrm{D}}}{\|\boldsymbol{v}_h^{\mathrm{D}}\|_{\mathrm{div};\Omega_{\mathrm{D}}}} - \|\xi_{\tilde{h}}\|_{1/2,\Sigma}.$$

Then according to the results in [26, Chapter IV] (see also [63, Section 4.2]), we know that there exists $C_{\mathrm{D}}>0$, independent of $h$, $h_\Sigma$ and $\tilde{h}$, such that

$$\sup_{\substack{\boldsymbol{v}_h^{\mathrm{D}}\in\mathbf{H}_h^{\mathrm{D}}\\ \boldsymbol{v}_h^{\mathrm{D}}\neq\mathbf{0}}} \frac{\displaystyle\int_{\Omega_{\mathrm{D}}} q_h^{\mathrm{D}}\,\mathrm{div}\,\boldsymbol{v}_h^{\mathrm{D}}}{\|\boldsymbol{v}_h^{\mathrm{D}}\|_{\mathrm{div};\Omega_{\mathrm{D}}}} \geq C_{\mathrm{D}}\,\|q_h^{\mathrm{D}}\|_{0,\Omega_{\mathrm{D}}} \qquad \forall\, q_h^{\mathrm{D}}\in \mathrm{Q}_{0,h}^{\mathrm{D}},$$

and hence

$$S_h^{\mathrm{D}}(q_h^{\mathrm{D}},\xi_{\tilde{h}}) \geq C_{\mathrm{D}}\,\|q_h^{\mathrm{D}}\|_{0,\Omega_{\mathrm{D}}} - \|\xi_{\tilde{h}}\|_{1/2,\Sigma} \qquad \forall\,(q_h^{\mathrm{D}},\xi_{\tilde{h}})\in \mathrm{Q}_{h,0}^{\mathrm{D}}\times\mathrm{Q}_{\tilde{h}}^{\Sigma}. \tag{5.71}$$

On the other hand, we know from Lemma 5.9, that there exist a linear operator $\mathcal{L}_h:\Phi_h(\Sigma)\to\mathbf{H}_h^{\mathrm{D}}$ and a constant $C_{\mathcal{L}}>0$, independent of $h$, such that for each $\psi_h\in\Phi_h(\Sigma)$ there hold

$$\mathcal{L}_h(\psi_h)\cdot\boldsymbol{n}=\psi_h \quad \text{on} \quad \Sigma, \quad \|\mathcal{L}_h(\psi_h)\|_{\mathrm{div},\Omega_{\mathrm{D}}}\leq C_{\mathcal{L}}\,\|\psi_h\|_{-1/2,\Sigma}, \quad \text{and} \quad \mathrm{div}\,\mathcal{L}_h(\psi_h)\in P_0(\Omega_{\mathrm{D}}).$$

In this way, we deduce that

$$S_h^{\mathrm{D}}(q_h^{\mathrm{D}},\xi_{\tilde{h}}) \geq \frac{\displaystyle\int_{\Omega_{\mathrm{D}}} q_h^{\mathrm{D}}\,\mathrm{div}\,\mathcal{L}_h(\psi_h) + \langle\mathcal{L}_h(\psi_h)\cdot\boldsymbol{n},\xi_{\tilde{h}}\rangle_\Sigma}{\|\mathcal{L}_h(\psi_h)\|_{\mathrm{div};\Omega_{\mathrm{D}}}} \qquad \forall\,\psi_h\in\Phi_h(\Sigma),$$

from which, using that $\mathrm{div}\,\mathcal{L}_h(\psi_h)\in P_0(\Omega_{\mathrm{D}})$ and that $q_h^{\mathrm{D}}\in\mathrm{Q}_{h,0}^{\mathrm{D}}$, it follows that

$$S_h^{\mathrm{D}}(q_h^{\mathrm{D}},\xi_{\tilde{h}}) \geq \frac{|\langle\mathcal{L}_h(\psi_h)\cdot\boldsymbol{n},\xi_{\tilde{h}}\rangle_\Sigma|}{\|\mathcal{L}_h(\psi_h)\|_{\mathrm{div},\Omega_{\mathrm{D}}}} \geq \frac{1}{C_{\mathcal{L}}}\frac{|\langle\psi_h,\xi_{\tilde{h}}\rangle_\Sigma|}{\|\psi_h\|_{-1/2,\Sigma}} \qquad \forall\,\psi_h\in\Phi_h(\Sigma),$$

and hence

$$S_h^{\mathrm{D}}(q_h^{\mathrm{D}},\xi_{\tilde{h}}) \geq \frac{1}{C_{\mathcal{L}}}\sup_{\substack{\psi_h\in\Phi_h(\Sigma)\\ \psi_h\neq 0}}\frac{\langle\psi_h,\xi_{\tilde{h}}\rangle_\Sigma}{\|\psi_h\|_{-1/2,\Sigma}} \qquad \forall\,\xi_{\tilde{h}}\in\mathrm{Q}_{\tilde{h}}^{\Sigma}. \tag{5.72}$$

Therefore, (5.72) and a straightforward application of Lemma 5.10 imply the existence of $\widetilde{C}_{\mathrm{D}}>0$, independent of $h$, $h_\Sigma$ and $\tilde{h}$, such that for all $h_\Sigma\leq C_0\tilde{h}$ there holds

$$S_h^{\mathrm{D}}(q_h^{\mathrm{D}},\xi_{\tilde{h}}) \geq \widetilde{C}_{\mathrm{D}}\,\|\xi_{\tilde{h}}\|_{1/2,\Sigma} \qquad \forall\,\xi_{\tilde{h}}\in\mathrm{Q}_{\tilde{h}}^{\Sigma}. \tag{5.73}$$

Finally, it easy to see that estimates (5.71) and (5.73) imply the discrete inf-sup condition (5.70), thus finishing the proof. □

**Lemma 5.12.** *There exist $C_0$, $\tilde{\beta}_{\mathrm{B}}>0$, independent of $h$, $h_\Sigma$ and $\tilde{h}$, such that for all $h_\Sigma\leq C_0\,\tilde{h}$ there holds*

$$S_h^{\mathrm{B}}(q_h^{\mathrm{B}},\xi_{\tilde{h}}) := \sup_{\substack{\boldsymbol{v}_h^{\mathrm{B}}\in\mathbf{H}_h^{\mathrm{B}}\\ \boldsymbol{v}_h^{\mathrm{B}}\neq\mathbf{0}}} \frac{\displaystyle\int_{\Omega_{\mathrm{B}}} q_h^{\mathrm{B}}\,\mathrm{div}\,\boldsymbol{v}_h^{\mathrm{B}} - \langle\boldsymbol{v}_h^{\mathrm{B}}\cdot\boldsymbol{n},\xi_{\tilde{h}}\rangle_\Sigma}{\|\boldsymbol{v}_h^{\mathrm{B}}\|_{\mathrm{div};\Omega_{\mathrm{B}}}} \geq \widetilde{\beta}_{\mathrm{B}}\left\{\|q_h^{\mathrm{B}}\|_{0,\Omega_{\mathrm{B}}} + \|\xi_{\tilde{h}}\|_{1/2,\Sigma}\right\} \tag{5.74}$$

*for all $(q_h^{\mathrm{B}},\xi_{\tilde{h}})\in\mathrm{Q}_{h,0}^{\mathrm{B}}\times\mathrm{Q}_{\tilde{h}}^{\Sigma}$.*

*Proof.* It proceeds exactly as the proof of Lemma 5.11 by replacing $\Omega_{\mathrm{D}}$, $\Gamma_{\mathrm{D}}$, $\mathrm{Q}_{h,0}^{\mathrm{D}}$ and $\mathbf{H}_h^{\mathrm{D}}$ by $\Omega_{\mathrm{B}}$, $\Gamma_{\mathrm{B}}$, $\mathrm{Q}_{h,0}^{\mathrm{B}}$ and $\mathbf{H}_h^{\mathrm{B}}$, respectively. □

The following theorem provides the rate of convergence of our Galerkin scheme (5.43).

**Theorem 5.4.** *Let* $\mathbf{H}_h := \mathbf{H}_h^{\mathrm{B}} \times \mathbf{H}_{0,h}^{\mathrm{B}} \times \mathbf{H}_h^{\mathrm{D}}$ *and* $\mathrm{Q}_{h,0} := \mathrm{Q}_{h,0}^{\mathrm{B}} \times \mathrm{Q}_h^{\mathrm{D}} \times \mathrm{Q}_{\tilde{h}}^{\Sigma}$ *be the subspaces specified above, and let* $(\vec{\boldsymbol{u}}, \vec{p}) := ((\boldsymbol{u}_{\mathrm{B}}, \boldsymbol{\omega}_{\mathrm{B}}, \boldsymbol{u}_{\mathrm{D}}), (p_{\mathrm{B}}, p_{\mathrm{D}}, \lambda)) \in \mathbf{H} \times \mathbf{Q}_0$ *and* $(\vec{\boldsymbol{u}}_h, \vec{p}_h) := ((\boldsymbol{u}_h^{\mathrm{B}}, \boldsymbol{\omega}_h^{\mathrm{B}}, \boldsymbol{u}_h^{\mathrm{D}}), (p_h^{\mathrm{B}}, p_h^{\mathrm{D}}, \lambda_{\tilde{h}})) \in \mathbf{H}_h \times \mathbf{Q}_{0,h}$ *be the unique solutions of the continuous and discrete problems* (5.28) *and* (5.43)*, respectively. Assume that* $\boldsymbol{u}_\star \in \mathbf{H}^\delta(\Omega_\star), \mathrm{div}\, \boldsymbol{u}_\star \in \mathrm{H}^\delta(\Omega_\star), p_\star \in \mathrm{H}^\delta(\Omega_\star)$ *where* $\star \in \{\mathrm{B}, \mathrm{D}\}$, $\boldsymbol{\omega}_{\mathrm{B}} \in \mathbf{H}^\delta(\mathbf{curl}; \Omega_{\mathrm{B}})$ *and* $\lambda \in \mathrm{H}^{1/2+\delta}(\Sigma)$*, for some* $\delta \in (1/2, 1]$*. Then, there exists* $C > 0$ *and* $\widetilde{C} > 0$ *independent of* $h$ *and* $\tilde{h}$ *such that*

$$\begin{aligned}\|(\vec{\boldsymbol{u}}, \vec{p}) - (\vec{\boldsymbol{u}}_h, \vec{p}_h)\|_{\mathbf{H}} \leq\ & Ch^\delta \Big\{ \|\boldsymbol{u}_{\mathrm{B}}\|_{\delta,\Omega_{\mathrm{B}}} + \|\mathrm{div}\,(\boldsymbol{u}_{\mathrm{B}})\|_{\delta,\Omega_{\mathrm{B}}} + \|\boldsymbol{\omega}_{\mathrm{B}}\|_{\mathbf{H}^\delta(\mathbf{curl};\Omega_{\mathrm{B}})} + \|\boldsymbol{u}_{\mathrm{D}}\|_{\delta,\Omega_{\mathrm{D}}} \\ & + \|\mathrm{div}\,(\boldsymbol{u}_{\mathrm{D}})\|_{\delta,\Omega_{\mathrm{D}}} + \|p_{\mathrm{B}}\|_{\delta,\Omega_{\mathrm{B}}} + \|p_{\mathrm{D}}\|_{\delta,\Omega_{\mathrm{D}}} \Big\} + \widetilde{C}\tilde{h}^\delta \|\lambda\|_{\delta+1/2,\Sigma}.\end{aligned}$$

*Proof.* It follows from the Céa estimate (5.54) and the approximation properties $(\mathrm{AP}_h^{\boldsymbol{u}_\star})$, $(\mathrm{AP}_h^{p_\star})$, $(\mathrm{AP}_h^{\boldsymbol{\omega}_{\mathrm{B}}})$ and $(\mathrm{AP}_h^{\lambda})$. □

We end this section by remarking that the analysis from Section 5.4.2 can be extended without difficulties, to Raviart-Thomas and Nédélec spaces of higher order.

## 5.5 An augmented mixed formulation

In this section we propose an augmented variational formulation of problem (5.28). Indeed, though many finite element subspaces $\mathbf{H}_{0,h}^{\mathrm{B}} \subseteq \mathbf{H}_0(\mathbf{curl}; \Omega_{\mathrm{B}})$ and $\mathbf{H}_h^{\mathrm{B}} \subseteq \mathbf{H}_{\mathrm{B}}(\mathrm{div}; \Omega_{\mathrm{B}})$ do satisfy **(H.3)**, we would like to explore the possibility of getting rid of that assumption. To this end, we suggest to enrich the mixed variational formulation (5.28) with a residual arising from the Brinkman momentum equation in (5.5). More precisely, we include into the variational problem (5.28) the following Galerkin least-squares equation in $\Omega_{\mathrm{B}}$:

$$\kappa \int_{\Omega_{\mathrm{B}}} (\alpha\, \boldsymbol{u}_{\mathrm{B}} + \nu\, \mathbf{curl}\, \boldsymbol{\omega}_{\mathrm{B}} + \nabla p_{\mathrm{B}} - \boldsymbol{f}_{\mathrm{B}}) \cdot \mathbf{curl}\, \boldsymbol{z}_{\mathrm{B}} = 0 \quad \forall\, \boldsymbol{z}_{\mathrm{B}} \in \mathbf{H}_0(\mathbf{curl}; \Omega_{\mathrm{B}}), \tag{5.75}$$

where $\kappa$ is a positive parameter to be specified later. Actually, integrating by parts, and using again that $\mathbf{curl}\, \boldsymbol{z}_{\mathrm{B}} \in \mathbf{H}_0(\mathrm{div}; \Omega_{\mathrm{B}})$ for each $\boldsymbol{z}_{\mathrm{B}} \in \mathbf{H}_0(\mathbf{curl}; \Omega_{\mathrm{B}})$ (*cf.* [77, Chapter I, Section 2.3, Remark 2.5]), we easily find that

$$\int_{\Omega_{\mathrm{B}}} \nabla p_{\mathrm{B}} \cdot \mathbf{curl}\, \boldsymbol{z}_{\mathrm{B}} = 0 \qquad \forall\, \boldsymbol{z}_{\mathrm{B}} \in \mathbf{H}_0(\mathbf{curl}; \Omega_{\mathrm{B}}),$$

whence (5.75) can be recast in the form

$$\kappa\, \alpha \int_{\Omega_{\mathrm{B}}} \boldsymbol{u}_{\mathrm{B}} \cdot \mathbf{curl}\, \boldsymbol{z}_{\mathrm{B}} + \kappa\, \nu \int_{\Omega_{\mathrm{B}}} \mathbf{curl}\, \boldsymbol{\omega}_{\mathrm{B}} \cdot \mathbf{curl}\, \boldsymbol{z}_{\mathrm{B}} = \kappa \int_{\Omega_{\mathrm{B}}} \boldsymbol{f}_{\mathrm{B}} \cdot \mathbf{curl}\, \boldsymbol{z}_{\mathrm{B}} \quad \forall\, \boldsymbol{z}_{\mathrm{B}} \in \mathbf{H}_0(\mathbf{curl}; \Omega_{\mathrm{B}}). \tag{5.76}$$

In this way, adding (5.76) to the first equation of (5.28), we obtain the following augmented variational formulation: Find $\vec{\boldsymbol{u}} := (\boldsymbol{u}_{\mathrm{B}}, \boldsymbol{\omega}_{\mathrm{B}}, \boldsymbol{u}_{\mathrm{D}}) \in \mathbf{H}$ and $\vec{p} := (p_{\mathrm{B}}, p_{\mathrm{D}}, \lambda) \in \mathbf{Q}_0$ such that

$$\begin{aligned} \mathcal{A}(\vec{\boldsymbol{u}}, \vec{\boldsymbol{v}}) + \mathcal{B}(\vec{\boldsymbol{v}}, \vec{p}) &= \mathcal{F}(\vec{\boldsymbol{v}}) \qquad \forall\, \vec{\boldsymbol{v}} := (\boldsymbol{v}_{\mathrm{B}}, \boldsymbol{z}_{\mathrm{B}}, \boldsymbol{v}_{\mathrm{D}}) \in \mathbf{H}, \\ \mathcal{B}(\vec{\boldsymbol{u}}, \vec{q}) &= \mathcal{G}(\vec{q}) \qquad \forall\, \vec{q} := (q_{\mathrm{B}}, q_{\mathrm{D}}, \xi) \in \mathbf{Q}_0, \end{aligned} \tag{5.77}$$

where

$$\mathcal{A}(\vec{\boldsymbol{u}},\vec{\boldsymbol{v}}) := \alpha\int_{\Omega_{\mathrm{B}}}\boldsymbol{u}_{\mathrm{B}}\cdot\boldsymbol{v}_{\mathrm{B}} + \nu\int_{\Omega_{\mathrm{B}}}\boldsymbol{\omega}_{\mathrm{B}}\cdot\boldsymbol{z}_{\mathrm{B}} + \kappa\,\nu\int_{\Omega_{\mathrm{B}}}\mathbf{curl}\,\boldsymbol{\omega}_{\mathrm{B}}\cdot\mathbf{curl}\,\boldsymbol{z}_{\mathrm{B}}$$
$$+\nu\int_{\Omega_{\mathrm{B}}}\boldsymbol{v}_{\mathrm{B}}\cdot\mathbf{curl}\boldsymbol{\omega}_{\mathrm{B}} + (\kappa\,\alpha-\nu)\int_{\Omega_{\mathrm{B}}}\boldsymbol{u}_{\mathrm{B}}\cdot\mathbf{curl}\boldsymbol{z}_{\mathrm{B}} + \mu\int_{\Omega_{\mathrm{D}}}\boldsymbol{u}_{\mathrm{D}}\cdot\boldsymbol{v}_{\mathrm{D}} \qquad \forall\,(\vec{\boldsymbol{u}},\vec{\boldsymbol{v}})\in\mathbf{H}\times\mathbf{H}\,, \tag{5.78}$$

$$\mathcal{F}(\vec{\boldsymbol{v}}) := \int_{\Omega_{\mathrm{B}}}\boldsymbol{f}_{\mathrm{B}}\cdot\boldsymbol{v}_{\mathrm{B}} + \int_{\Omega_{\mathrm{D}}}\boldsymbol{f}_{\mathrm{D}}\cdot\boldsymbol{v}_{\mathrm{D}} + \kappa\int_{\Omega_{\mathrm{B}}}\boldsymbol{f}_{\mathrm{B}}\cdot\mathbf{curl}\,\boldsymbol{z}_{\mathrm{B}} \qquad \forall\,\vec{\boldsymbol{v}}\in\mathbf{H}\,, \tag{5.79}$$

$\mathcal{B}=\mathbf{b}$, and $\mathcal{G}=\mathbf{G}=\mathbf{0}$.

In what follows we address the solvability of (5.77). We first observe that the continuous inf-sup condition for $\mathcal{B}$ on $\mathbf{H}\times\mathbf{Q}_0$ is already proved by Lemma 5.4. In turn, the continuous kernel of $\mathcal{B}$ is certainly given by $\mathbf{V}$ (*cf.* (5.31) - (5.33)). Then, we have the following result establishing the ellipticity of $\mathcal{A}$ on $\mathbf{V}_{\mathrm{B,D}}$ and hence on $\mathbf{V}$.

**Lemma 5.13.** *Assume that the stabilization parameter* $\kappa\in\big(0,2\delta\big)$ *with* $\delta\in\left(0,\dfrac{2\nu}{\alpha}\right)$. *Then, there exists* $\varrho>0$, *depending on* $\kappa$ *and* $\delta$, *such that*

$$\mathcal{A}(\vec{\boldsymbol{v}},\vec{\boldsymbol{v}}) \;\geq\; \varrho\,\|\vec{\boldsymbol{v}}\|^2_{\mathbf{H}} \qquad \forall\,\vec{\boldsymbol{v}}\in\mathbf{V}_{\mathrm{B,D}}\,. \tag{5.80}$$

*Proof.* Given $\vec{\boldsymbol{v}}:=(\boldsymbol{v}_{\mathrm{B}},\boldsymbol{z}_{\mathrm{B}},\boldsymbol{v}_{\mathrm{D}})\in\mathbf{V}_{\mathrm{B,D}}$, we obtain from the definition of $\mathcal{A}$ (*cf.* (5.78)) and the Cauchy-Schwarz inequality, that

$$\mathcal{A}(\vec{\boldsymbol{v}},\vec{\boldsymbol{v}}) = \alpha\,\|\boldsymbol{v}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} + \nu\,\|\boldsymbol{z}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} + \kappa\,\nu\,\|\mathbf{curl}\,\boldsymbol{z}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} + \kappa\,\alpha\int_{\Omega_{\mathrm{B}}}\boldsymbol{v}_{\mathrm{B}}\cdot\mathbf{curl}\,\boldsymbol{z}_{\mathrm{B}} + \mu\,\|\boldsymbol{v}_{\mathrm{D}}\|^2_{0,\Omega_{\mathrm{D}}}$$
$$\geq \alpha\,\|\boldsymbol{v}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} + \nu\,\|\boldsymbol{z}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} + \kappa\,\nu\,\|\mathbf{curl}\,\boldsymbol{z}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} - \kappa\alpha\|\boldsymbol{v}_{\mathrm{B}}\|_{0,\Omega_{\mathrm{B}}}\|\mathbf{curl}\,\boldsymbol{z}_{\mathrm{B}}\|_{0,\Omega_{\mathrm{B}}} + \mu\,\|\boldsymbol{v}_{\mathrm{D}}\|^2_{0,\Omega_{\mathrm{D}}}.$$

Next, for each $\delta>0$ we find that

$$-\kappa\,\alpha\,\|\boldsymbol{v}_{\mathrm{B}}\|_{0,\Omega_{\mathrm{B}}}\,\|\mathbf{curl}\,\boldsymbol{z}_{\mathrm{B}}\|_{0,\Omega_{\mathrm{B}}} \;\geq\; -\frac{\kappa\,\alpha}{2\delta}\,\|\boldsymbol{v}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} - \frac{\delta\,\kappa\,\alpha}{2}\,\|\mathbf{curl}\,\boldsymbol{z}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}},$$

which, replaced back into the foregoing estimate, yields

$$\mathcal{A}(\vec{\boldsymbol{v}},\vec{\boldsymbol{v}}) \;\geq\; \alpha\left(1-\frac{\kappa}{2\delta}\right)\|\boldsymbol{v}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} + \nu\,\|\boldsymbol{z}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} + \kappa\left(\nu-\frac{\delta\,\alpha}{2}\right)\|\mathbf{curl}\,\boldsymbol{z}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} + \mu\,\|\boldsymbol{v}_{\mathrm{D}}\|^2_{0,\Omega_{\mathrm{D}}}.$$

Next, using (5.36) and noting that $\|\boldsymbol{v}_{\mathrm{D}}\|^2_{0,\Omega_{\mathrm{B}}} = \|\boldsymbol{v}_{\mathrm{D}}\|^2_{\mathrm{div};\Omega_{\mathrm{D}}}$, we obtain

$$\mathcal{A}(\vec{\boldsymbol{v}},\vec{\boldsymbol{v}}) \;\geq\; \alpha\left(1-\frac{\kappa}{2\delta}\right)\varrho_0^2\,\|\boldsymbol{v}_{\mathrm{B}}\|^2_{\mathrm{div},\Omega_{\mathrm{B}}} + \nu\,\|\boldsymbol{z}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} + \kappa\left(\nu-\frac{\delta\,\alpha}{2}\right)\|\mathbf{curl}\,\boldsymbol{z}_{\mathrm{B}}\|^2_{0,\Omega_{\mathrm{B}}} + \mu\,\|\boldsymbol{v}_{\mathrm{D}}\|^2_{\mathrm{div},\Omega_{\mathrm{D}}}. \tag{5.81}$$

Hence, since $1-\frac{\kappa}{2\,\delta}>0$ and $\nu-\frac{\delta\,\alpha}{2}>0$, we conclude that

$$\mathcal{A}(\vec{\boldsymbol{v}},\vec{\boldsymbol{v}}) \;\geq\; \varrho\|\vec{\boldsymbol{v}}\|^2_{\mathbf{H}} \quad \vec{\boldsymbol{v}}\in\mathbf{V}_{\mathrm{B,D}},$$

where $\varrho := \min\left\{\alpha\left(1-\frac{\kappa}{2\delta}\right)\varrho_0^2,\nu,\kappa\left(\nu-\frac{\delta\,\alpha}{2}\right),\mu\right\}$. □

Note that, taking in particular $\kappa=\delta=\dfrac{\nu}{\alpha}$, we obtain the optimal ellipticity constant

$$\varrho := \frac{1}{2}\,\min\left\{\alpha\varrho_0^2,2\nu,\kappa\nu,2\mu\right\}.$$

The foregoing analysis yields the following main result.

**Theorem 5.5.** *Assume that* $\mathbf{f}_{\mathrm{D}} \in \mathbf{L}^2(\Omega_{\mathrm{D}})$, $\mathbf{f}_{\mathrm{B}} \in \mathbf{L}^2(\Omega_{\mathrm{B}})$, *and that* $\kappa$ *satisfies the assumption from Lemma 5.13. Then there exists a unique* $(\vec{\boldsymbol{u}}, \vec{p}) := \big((\boldsymbol{u}_{\mathrm{B}}, \boldsymbol{\omega}_{\mathrm{B}}, \boldsymbol{u}_{\mathrm{D}}), (p_{\mathrm{B}}, p_{\mathrm{D}}, \lambda)\big) \in \mathbf{H} \times \mathbf{Q}_0$ *solution of the augmented mixed formulation (5.77). Moreover, there exists* $C > 0$ *such that*

$$\|\vec{\boldsymbol{u}}\|_{\mathbf{H}} + \|\vec{p}\|_{\mathbf{Q}} \leq C \Big\{ \|\boldsymbol{f}_{\mathrm{D}}\|_{0,\Omega_{\mathrm{D}}} + \|\boldsymbol{f}_{\mathrm{B}}\|_{0,\Omega_{\mathrm{B}}} \Big\} . \tag{5.82}$$

*Proof.* Thanks to Lemmata 5.4 and 5.13, the proof is a straightforward application of the continuous Babuška-Brezzi theory. □

We now look at the Galerkin scheme of (5.77). More precisely, employing the same generic finite elements subspaces and related notations introduced in Section 5.4.1, we now consider the augmented mixed finite element scheme: Find $\vec{\boldsymbol{u}}_h := (\boldsymbol{u}_h^{\mathrm{B}}, \boldsymbol{\omega}_h^{\mathrm{B}}, \boldsymbol{u}_h^{\mathrm{D}}) \in \mathbf{H}_h$ and $\vec{p}_h := (p_h^{\mathrm{B}}, p_h^{\mathrm{D}}, \lambda_h) \in \mathbf{Q}_{0,h}$ such that

$$\begin{aligned} \mathcal{A}(\vec{\boldsymbol{u}}_h, \vec{\boldsymbol{v}}_h) + \mathcal{B}(\vec{\boldsymbol{v}}_h, \vec{p}_h) &= \mathcal{F}(\vec{\boldsymbol{v}}_h) \qquad \forall\, \vec{\boldsymbol{v}}_h := (\boldsymbol{v}_h^{\mathrm{B}}, \boldsymbol{z}_h^{\mathrm{B}}, \boldsymbol{v}_h^{\mathrm{D}}) \in \mathbf{H}_h \,, \\ \mathcal{B}(\vec{\boldsymbol{u}}_h, \vec{q}_h) &= \mathcal{G}(\vec{q}_h) \qquad \forall\, \vec{q}_h := (q_h^{\mathrm{B}}, q_h^{\mathrm{D}}, \xi_h) \in \mathbf{Q}_{0,h} \,. \end{aligned} \tag{5.83}$$

Then, assuming that hypotheses **(H.0)**, **(H.1)**, and **(H.2)** from Section 5.4 are satisfied, we certainly deduce that $\mathcal{B}$ verifies the discrete inf-sup condition on $\mathbf{H}_h \times \mathbf{Q}_{0,h}$ (*cf.* Lemma 5.6), the discrete kernel of $\mathcal{B}$ is given again by $\mathbf{V}_h = \mathbf{V}_{\mathrm{B},\mathrm{D}}^h \cap \mathbf{V}_\Sigma^h$ (*cf.* (5.48) - (5.50)), and hence, since $\mathbf{V}_{\mathrm{B},\mathrm{D}}^h$ is contained in $\mathbf{V}_{\mathrm{B},\mathrm{D}}$, the bilinear form $\mathcal{A}$ is elliptic in $\mathbf{V}_{\mathrm{B},\mathrm{D}}^h$ (*cf.* Lemma 5.13) and therefore in $\mathbf{V}_h$. Consequently, a straightforward application of the discrete Babuška-Brezzi theory allows to conclude the following result.

**Theorem 5.6.** *Assume that* $\mathbf{f}_{\mathrm{D}} \in \mathbf{L}^2(\Omega_{\mathrm{D}})$ *and* $\mathbf{f}_{\mathrm{B}} \in \mathbf{L}^2(\Omega_{\mathrm{B}})$. *In addition, suppose that* **(H.0)**, **(H.1)**, *and* **(H.2)** *hold. Then there exists a unique* $(\vec{\boldsymbol{u}}_h, \vec{p}_h) := \big((\boldsymbol{u}_h^{\mathrm{B}}, \boldsymbol{\omega}_h^{\mathrm{B}}, \boldsymbol{u}_h^{\mathrm{D}}), (p_h^{\mathrm{B}}, p_h^{\mathrm{D}}, \lambda_h)\big) \in \mathbf{H}_h \times \mathbf{Q}_{0,h}$ *solution of the augmented Galerkin scheme (5.83). Moreover, there exist* $C_1$, $C_2 > 0$, *independent of* $h$, *such that*

$$\|\vec{\boldsymbol{u}}_h\|_{\mathbf{H}} + \|\vec{p}_h\|_{\mathbf{Q}} \leq C_1 \Big\{ \|\boldsymbol{f}_{\mathrm{D}}\|_{0,\Omega_{\mathrm{D}}} + \|\boldsymbol{f}_{\mathrm{B}}\|_{0,\Omega_{\mathrm{B}}} \Big\} , \tag{5.84}$$

*and*

$$\|(\vec{\boldsymbol{u}}, \vec{p}) - (\vec{\boldsymbol{u}}_h, \vec{p}_h)\|_{\mathbf{H}\times\mathbf{Q}} \leq C_2 \Big\{ \mathrm{dist}(\vec{\boldsymbol{u}}, \mathbf{H}_h) + \mathrm{dist}(\vec{p}, \mathbf{Q}_{0,h}) \Big\} . \tag{5.85}$$

We end this section by providing a specific example of finite element subspaces satisfying (**H.0**), (**H.1**), and (**H.2**), but not (**H.3**), whence only the augmented formulation described in this section can be employed with them. In fact, for each $K \in \mathcal{T}_h(\Omega_{\mathrm{B}})$ we now let $\mathbb{ND}_2(K)$ be the local Nédélec space of order 2, that is

$$\mathbb{ND}_2(K) := \mathbf{P}_1(K) \oplus \widetilde{\mathbf{P}}_1(K) \times \boldsymbol{x} \,,$$

where $\widetilde{\mathbf{P}}_1(K)$ is the space of polynomials of degree $= 1$, and introduce

$$\mathbf{H}_{0,h}^{\mathrm{B}} := \Big\{ \boldsymbol{z}_h^{\mathrm{B}} \in \mathbf{H}_0(\mathbf{curl}; \Omega_{\mathrm{B}}) : \quad \boldsymbol{z}_h^{\mathrm{B}}|_K \in \mathbb{ND}_2(K) \quad \forall\, K \in \mathcal{T}_h(\Omega_{\mathrm{B}}) \Big\} . \tag{5.86}$$

In turn, the remaining subspaces $\mathbf{H}_h^{\mathrm{B}}$, $\mathbf{H}_h^{\mathrm{D}}$, $\mathrm{Q}_{h,0}^{\mathrm{B}}$, $\mathrm{Q}_h^{\mathrm{D}}$, and $\mathrm{Q}_h^{\Sigma}$ are exactly as those defined in Section 5.4.2.1, that is with local Raviart-Thomas spaces of order 0 and discontinuous piecewise constant polynomials on the domains, and with continuous piecewise polynomials of degree $\leq 1$ on the interface $\Sigma$. Then, it is not difficult to see that there exist $\boldsymbol{z}_h^{\mathrm{B}} \in \mathbf{H}_{0,h}^{\mathrm{B}}$ for which $\mathbf{curl}\, \boldsymbol{z}_h^{\mathrm{B}}$ does not belong to $\mathbf{H}_h^{\mathrm{B}}$, thus confirming that (**H.3**) does not hold in this case. Numerical results illustrating optimal convergence rates of the augmented formulation with the aforementioned finite element subspaces are reported below in Section 5.6.

## 5.6 Numerical results

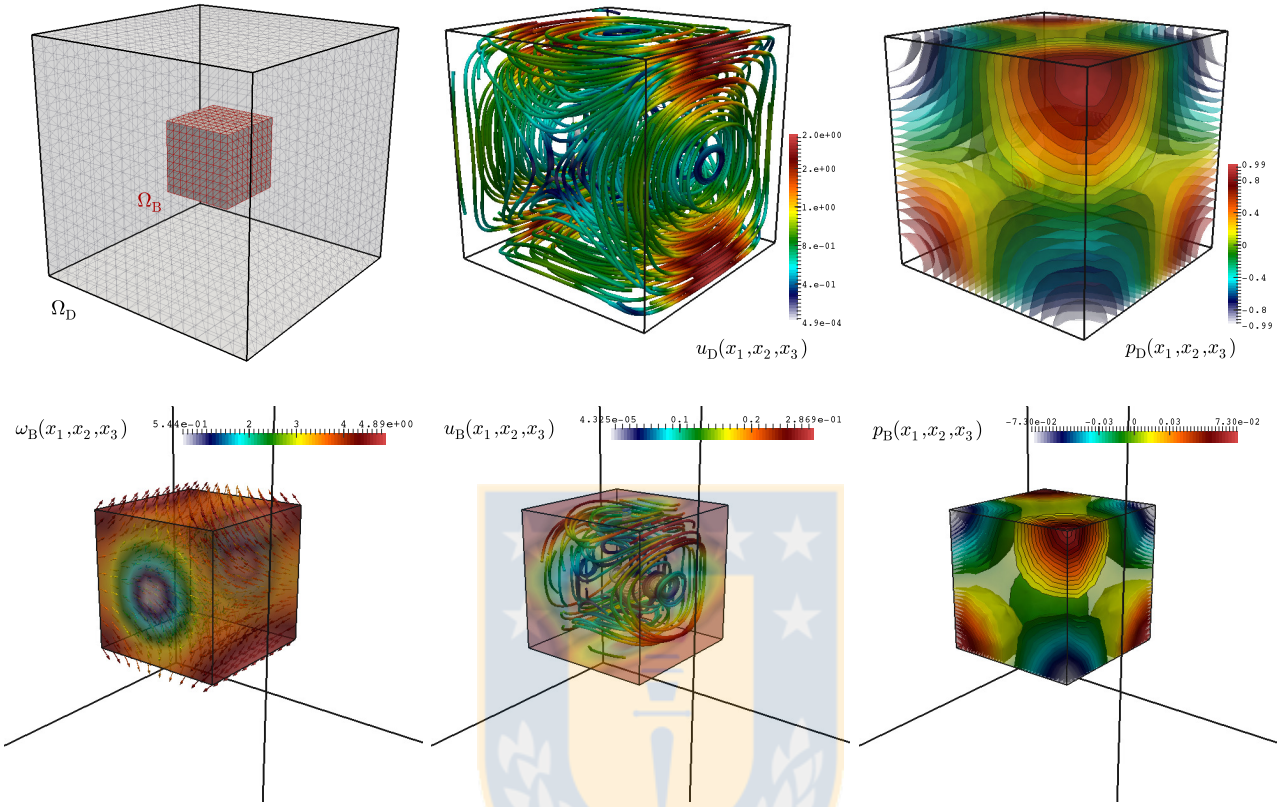


Figure 5.2: Example 1: Two-domain geometry and mesh (top left), approximated Darcy velocity streamlines (top middle), approximated Darcy pressure isosurfaces (top right), zoom of approximated Brinkman vorticity vectors (bottom left), zoom of approximated Brinkman velocity streamlines (bottom middle), and isosurfaces of the computed Brinkman pressure (bottom right).

In this section we provide three computer experiments confirming the convergence rates anticipated by Theorem 5.4 and illustrating the applicability of the method in surface-subsurface flow problems.

### 5.6.1 Accuracy of the mixed and augmented formulations on two embedded cubes

We start by evaluating the convergence of the fully-mixed and the augmented finite element methods applied to (5.5)-(5.6) and defined on the two cubes $\Omega_B = [-r_B, r_B]^3$ and $\Omega_D = [-r_D, r_D]^3$, with $r_D = \frac{1}{2}, r_B = \frac{3}{20}$. Notice that this particular domain configuration does not fall exactly in the theoretical framework analyzed in this chapter. However, both the continuous and discrete study could be carried out using the analogous tools as those used here. We employ the model parameters $\alpha = \mu = 1, \nu = 0.01$, yielding the stabilization constant $\kappa = 2\nu/\alpha = 0.02$ suggested by Lemma 5.13. The convergence of the

method is assessed by computing errors between the following manufactured smooth exact solutions

$$\boldsymbol{\omega}_{\mathrm{B}}(x_1,x_2,x_3)=\begin{pmatrix}-3\pi\sin(\pi x_1)\cos(\pi x_2)\cos(\pi x_3)\\ 3\pi\cos(\pi x_1)\sin(\pi x_2)\cos(\pi x_3)\\ 0\end{pmatrix},\boldsymbol{u}(x_1,x_2,x_3)=\begin{pmatrix}\cos(\pi x_1)\sin(\pi x_2)\sin(\pi x_3)\\ \sin(\pi x_1)\cos(\pi x_2)\sin(\pi x_3)\\ -2\sin(\pi x_1)\sin(\pi x_2)\cos(\pi x_3)\end{pmatrix},$$

$$p(x_1,x_2,x_3)=\sin(\pi x_1)\sin(\pi x_2)\sin(\pi x_3),\ \boldsymbol{u}_{\mathrm{B}}=\boldsymbol{u}|_{\Omega_{\mathrm{B}}},\ \boldsymbol{u}_{\mathrm{D}}=\boldsymbol{u}|_{\Omega_{\mathrm{D}}},\ p_{\mathrm{B}}=p|_{\Omega_{\mathrm{B}}},\ p_{\mathrm{D}}=p|_{\Omega_{\mathrm{D}}},\ \lambda=p|_{\Sigma},$$

and their finite element approximations using a $\mathbb{RT}_0-\mathbb{ND}_1-\mathbb{RT}_0-\mathbf{P}_0-\mathbf{P}_0-\mathbf{P}_1$ family (and using the fully-mixed and augmented formulations), and also an augmented method based on the $\mathbb{RT}_0-\mathbb{ND}_2-\mathbb{RT}_0-\mathbf{P}_0-\mathbf{P}_0-\mathbf{P}_1$ family, which, in particular, does not satisfy assumption (**H.3**). The computations are carried out on a sequence of successively refined tetrahedral meshes $\mathcal{T}_{h_{B_i}}$ and $\mathcal{T}_{h_{D_i}}$ of sizes $h_{B_i}=r_B2^{1-i}$ and $h_{D_i}=r_D2^{-i}$, respectively, $i=0,1,\ldots$. We adequately choose forcing terms $\boldsymbol{f}_{\mathrm{B}}=\alpha\boldsymbol{u}_{\mathrm{B}}+\mathbf{curl}\boldsymbol{\omega}_{\mathrm{B}}+\nabla p_{\mathrm{B}}$, $\boldsymbol{f}_{\mathrm{D}}=\mu\boldsymbol{u}_{\mathrm{D}}+\nabla p_{\mathrm{D}}$, and suitable *nonhomogeneous* slip velocity on $\partial\Omega$ and nonhomogeneous Dirichlet data for the tangential vorticity on $\partial\Omega_{\mathrm{B}}$, such that (5.5)-(5.6) holds. For sake of convenience we define a conforming partition for $\Sigma$, that is $\mathcal{T}_{\hat{h}}=\mathcal{T}_h$. The approximate solutions are depicted in Figure 5.2 and the error history, written in terms of the quantities

$$e(\boldsymbol{u}_{\mathrm{B}}):=\|\boldsymbol{u}_{\mathrm{B}}-\boldsymbol{u}_{\mathrm{B}h}\|_{\mathrm{div},\Omega_{\mathrm{B}}},\ e(\boldsymbol{\omega}_{\mathrm{B}}):=\|\boldsymbol{\omega}_{\mathrm{B}}-\boldsymbol{\omega}_{\mathrm{B}h}\|_{\mathbf{curl},\Omega_{\mathrm{B}}},\ e(\boldsymbol{u}_{\mathrm{D}}):=\|\boldsymbol{u}_{\mathrm{D}}-\boldsymbol{u}_{\mathrm{D}h}\|_{\mathrm{div},\Omega_{\mathrm{D}}},$$

$$e(p_{\mathrm{B}}):=\|p_{\mathrm{B}}-p_{\mathrm{B}h}\|_{0,\Omega_{\mathrm{B}}},\ e(p_{\mathrm{D}}):=\|p_{\mathrm{D}}-p_{\mathrm{D}h}\|_{0,\Omega_{\mathrm{D}}},\ e(\lambda):=\|\lambda-\lambda_h\|_{1/2,\Sigma},\ r(\cdot):=\frac{\log(e(\cdot)/\hat{e}(\cdot))}{\log(h/\hat{h})},$$

is reported in Table 5.1, where $e,\hat{e}$ denote errors on two consecutive meshes of sizes $h=\max\{h_B,h_D\}$ and $\hat{h}$. We observe that all studied methods deliver optimal convergence rates for vorticity, velocity and pressure in the corresponding norms.

### 5.6.2 Flow into a cracked porous medium

Our second example focuses on the simulation of flow in a porous medium with a smoothed V-shaped crack, similar to the 2D simulations presented in the Stokes-Darcy examples of [18, Section 7.1] and [35, Section 6.3]. The full domain is the box $\Omega=[0,2]\times[0,0.2]\times[0,1]$, the Brinkman domain on the top is $0.75\leq x_1\leq 1.25$ and goes down to $x_3=0.5$. Viscosity and porosity correspond to the case of water flowing in a mixture of calcarenite and sand: $\nu=0.01,\mu=10000$, and we set $\alpha=0.001$. The external forces on both domains correspond to gravity $\boldsymbol{f}_{\mathrm{D}}=\boldsymbol{f}_{\mathrm{B}}=(0,0,-0.98)^{\mathtt{t}}$, and a constant flowrate $\boldsymbol{u}_{\mathrm{D}}\cdot\boldsymbol{n}=(10,0,0)^{\mathtt{t}}\cdot\boldsymbol{n}$, is imposed on the right wall $\Gamma_{\mathrm{D}}^{\mathrm{in}}$, at $x_1=0$ (see sketch in figure 5.3), representing a subsurface flow in the $x_1$-direction. Normal Darcy velocities are set to zero everywhere else on $\Gamma_{\mathrm{D}}$. As in [18] we impose a smooth vorticity profile on the top of $\Gamma_{\mathrm{B}}$ $\boldsymbol{\omega}_{\mathrm{B}}\times\boldsymbol{n}=(0,1/16-(x_1-1)^2,0)^{\mathtt{t}}\times\boldsymbol{n}$, which takes into account the wind on the surface, and we also assume a compatible normal velocity on that same surface $\boldsymbol{u}_{\mathrm{B}}\cdot\boldsymbol{n}=(0,0,-x_1/16+[(x_1-1)^3]/3)^{\mathtt{t}}\cdot\boldsymbol{n}$. Everywhere else we set zero normal fluid velocity and zero tangential vorticity. A tetrahedral mesh with conforming interface is generated having 57426 vertices and 307544 elements, which in total correspond to 962639 degrees of freedom for $\mathbb{RT}_0-\mathbb{ND}_1-\mathbb{RT}_0-\mathbf{P}_0-\mathbf{P}_0-\mathbf{P}_1$ finite elements. Figure 5.3 depicts the domain configuration along with the approximate solutions, matching qualitatively the results from [18, 35].

### 5.6.3 Perpendicular infiltration through a porous capsule

In this test we present a model of coupled surface and subsurface flow where the top domain is the flow region and the bottom half of the domain represents e.g. a pellet, or a capsule. On the top left

| $h$ | $e(\boldsymbol{u}_{\mathrm{B}})$ | $r(\boldsymbol{u}_{\mathrm{B}})$ | $e(\boldsymbol{\omega}_{\mathrm{B}})$ | $r(\boldsymbol{\omega}_{\mathrm{B}})$ | $e(\boldsymbol{u}_{\mathrm{D}})$ | $r(\boldsymbol{u}_{\mathrm{D}})$ | $e(p_{\mathrm{B}})$ | $r(p_{\mathrm{B}})$ | $e(p_{\mathrm{D}})$ | $r(p_{\mathrm{D}})$ | $e(\lambda)$ | $r(\lambda)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fully mixed scheme (5.43) | | | | | | | | | | | | |
| 0.70711 | 1.02802 | – | 0.08636 | – | 0.65565 | – | 0.00404 | – | 0.64650 | – | 0.51608 | – |
| 0.38079 | 0.66329 | 0.63216 | 0.04547 | 0.86511 | 0.30143 | 0.96758 | 0.00167 | 0.86588 | 0.24919 | 1.54026 | 0.37415 | 0.94712 |
| 0.30610 | 0.45239 | 1.30206 | 0.03253 | 1.13929 | 0.21869 | 1.46952 | 0.00130 | 1.44871 | 0.12438 | 1.18266 | 0.22579 | 0.93668 |
| 0.18503 | 0.29153 | 0.93254 | 0.02240 | 0.79149 | 0.16048 | 0.61483 | 0.00073 | 1.21498 | 0.05130 | 1.15932 | 0.17396 | 0.95387 |
| 0.14412 | 0.18023 | 1.02275 | 0.01527 | 0.85417 | 0.10498 | 0.89264 | 0.00042 | 1.15692 | 0.02337 | 1.14603 | 0.10985 | 0.96014 |
| 0.05487 | 0.11716 | 0.95707 | 0.00833 | 0.95944 | 0.05076 | 0.99002 | 0.00027 | 0.96289 | 0.01207 | 0.98594 | 0.05139 | 0.96765 |
| 0.03564 | 0.07258 | 0.97681 | 0.00561 | 0.98670 | 0.03123 | 0.99197 | 0.00019 | 0.98441 | 0.00896 | 0.99455 | 0.02987 | 0.97732 |
| Augmented mixed scheme (5.83) | | | | | | | | | | | | |
| 0.70711 | 1.02681 | – | 0.08574 | – | 0.65549 | – | 0.00416 | – | 0.64537 | – | 0.51899 | – |
| 0.38079 | 0.62020 | 0.98418 | 0.04306 | 0.87429 | 0.26781 | 0.94902 | 0.00158 | 0.95434 | 0.24503 | 0.94234 | 0.38461 | 0.97729 |
| 0.30610 | 0.42963 | 0.99011 | 0.02763 | 0.91368 | 0.17061 | 0.96547 | 0.00109 | 0.93939 | 0.10942 | 0.96471 | 0.21733 | 1.07908 |
| 0.18503 | 0.27689 | 0.94556 | 0.01916 | 0.94842 | 0.12903 | 0.95084 | 0.00066 | 0.98741 | 0.04873 | 0.96933 | 0.16430 | 0.98544 |
| 0.14412 | 0.16540 | 0.96134 | 0.01344 | 0.96083 | 0.08211 | 0.95171 | 0.00039 | 0.98177 | 0.01998 | 0.96297 | 0.08127 | 0.97476 |
| 0.05487 | 0.10214 | 0.98608 | 0.00703 | 0.95798 | 0.04714 | 0.90989 | 0.00024 | 0.97506 | 0.00987 | 0.97250 | 0.04550 | 0.97732 |
| 0.03564 | 0.06071 | 0.96110 | 0.00416 | 0.98465 | 0.02595 | 1.01103 | 0.00016 | 0.98411 | 0.00593 | 1.00141 | 0.02831 | 0.97460 |
| Augmented mixed scheme (5.83) with vorticity space as in (5.86) | | | | | | | | | | | | |
| 0.70711 | 1.02643 | – | 0.06194 | – | 0.65421 | – | 0.00381 | – | 0.64501 | – | 0.49673 | – |
| 0.38079 | 0.60941 | 0.96599 | 0.02469 | 1.12048 | 0.26554 | 0.95410 | 0.00155 | 0.95308 | 0.24394 | 0.95725 | 0.34426 | 1.02005 |
| 0.30610 | 0.42550 | 0.98731 | 0.01822 | 1.51635 | 0.17022 | 0.92217 | 0.00102 | 0.94694 | 0.10922 | 0.96820 | 0.20871 | 0.97632 |
| 0.18503 | 0.23411 | 1.01620 | 0.01014 | 1.85309 | 0.12850 | 0.98436 | 0.00061 | 0.93551 | 0.04855 | 0.95735 | 0.15809 | 0.99401 |
| 0.14412 | 0.13869 | 0.98807 | 0.00607 | 1.92486 | 0.08173 | 0.96728 | 0.00039 | 0.93754 | 0.01980 | 0.96443 | 0.08067 | 0.95866 |
| 0.05487 | 0.08619 | 0.97615 | 0.00413 | 1.93952 | 0.04684 | 0.93863 | 0.00022 | 0.98965 | 0.00982 | 0.94538 | 0.04491 | 0.97210 |
| 0.03564 | 0.05175 | 0.99054 | 0.00279 | 1.97943 | 0.02433 | 0.98461 | 0.00013 | 1.90241 | 0.00563 | 1.10394 | 0.02635 | 0.96955 |

Table 5.1: Example 1: Error history associated to fully mixed and augmented $\mathbb{RT}_0 - \mathbb{ND}_1 - \mathbb{RT}_0 - \mathbf{P}_0 - \mathbf{P}_0 - \mathbf{P}_1$ discretizations (top and middle rows), and augmented $\mathbb{RT}_0 - \mathbb{ND}_2 - \mathbb{RT}_0 - \mathbf{P}_0 - \mathbf{P}_0 - \mathbf{P}_1$ FE family (bottom row) for problem (5.5)-(5.6) on a 3D domain.

octant of $\Omega_{\mathrm{B}}$, denoted by $\Gamma_{\mathrm{B}}^{\mathrm{in}}$, we consider an inflow rate of $\boldsymbol{u}_{\mathrm{B}} \cdot \boldsymbol{n} = -0.01$ and on $\Gamma_{\mathrm{D}}^{\mathrm{out}}$ (see the domain sketch in Figure 5.4) we set an outflow of fluid at rate $\boldsymbol{u}_{\mathrm{B}} \cdot \boldsymbol{n} = 0.01$. Also on $\Gamma_{\mathrm{B}}^{\mathrm{in}}$, we impose a smooth vorticity $\boldsymbol{\omega}_{\mathrm{B}} \times \boldsymbol{n} = (0, -0.01 x_1 x_2 x_3, 0)^{\mathtt{t}} \times \boldsymbol{n}$. On the remainder of $\partial\Omega$ we set zero normal velocities and tangential vorticity. As in the previous example, we take into account the gravity force acting on both domains $\boldsymbol{f}_{\mathrm{D}} = \boldsymbol{f}_{\mathrm{B}} = (0, 0, -0.98)^{\mathtt{t}}$, and employ the model parameters $\alpha = 10, \nu = 0.001, \mu = 10000$. The mesh for $\Omega$ consists of 32768 vertices and 191452 tetrahedral elements representing 700835 degrees of freedom. As expected, from Figure 5.4 we observe flow patterns entering the domain through $\Gamma_{\mathrm{B}}^{\mathrm{in}}$, percolating through $\Sigma$, and leaving the domain through $\Gamma_{\mathrm{D}}^{\mathrm{out}}$. These results have been obtained with the augmented mixed scheme (5.83).

### 5.6.4 Flow simulations imposing Dirichlet conditions for the velocity

Finally, we perform a test quite similar to Examples 2 and 3, but this time we impose Dirichlet conditions for the Brinkman and Darcy velocities on the external boundaries (which implies, in particular, that no boundary datum is required for the vorticity field), and employ an augmented formulation using

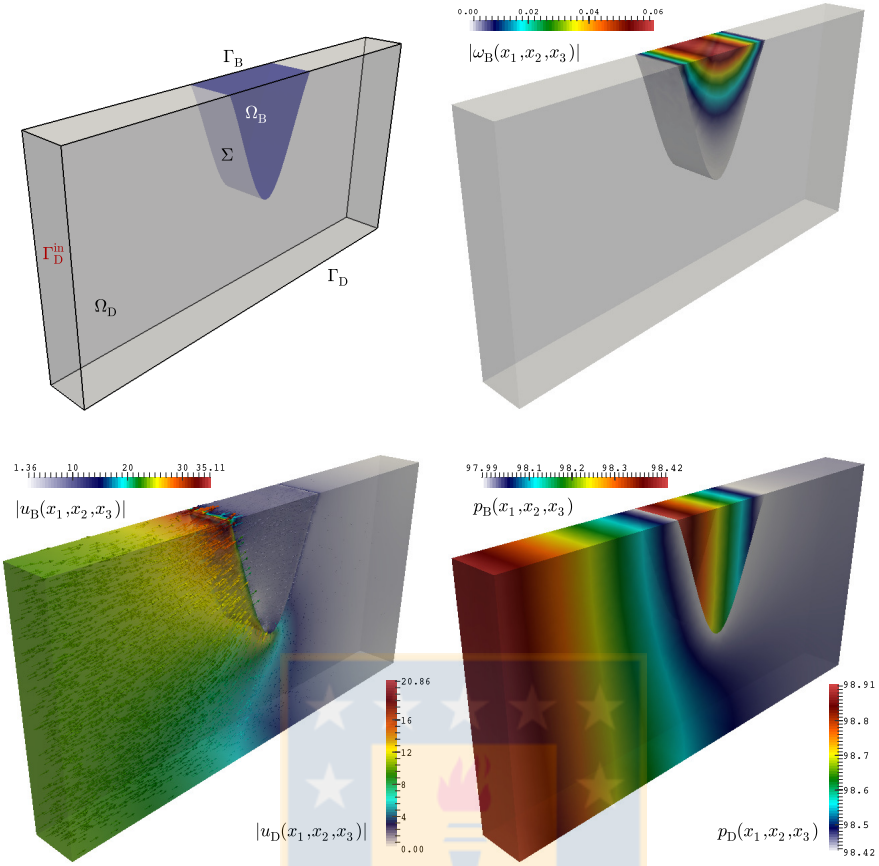


Figure 5.3: Example 2: Two-domain geometry and boundaries (top left), approximated Brinkman vorticity magnitude (top right), approximated velocity magnitude and vectors (bottom left), and computed pressure profiles (bottom right) for the Brinkman-Darcy coupling, using a fully-mixed scheme.

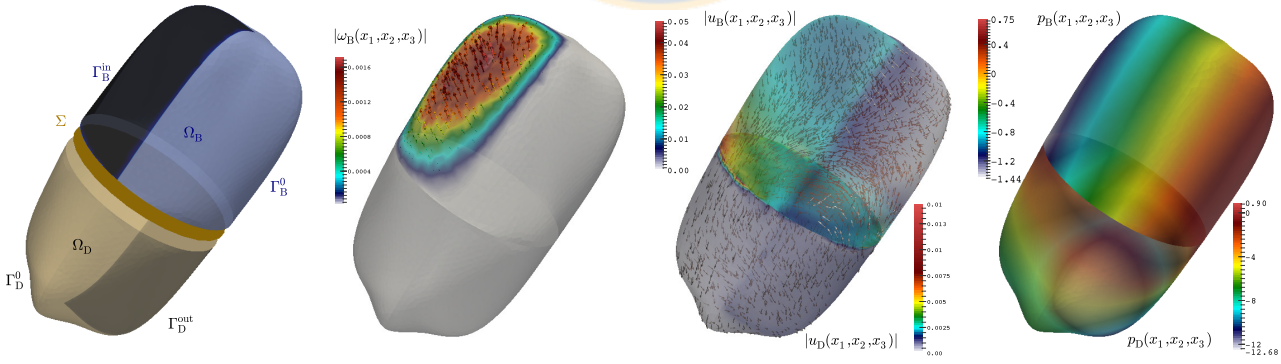


Figure 5.4: Example 3: From left to right: Two-domain geometry and boundaries, approximated Brinkman vorticity magnitude and vectors, approximated velocity magnitude and vectors, and computed pressure profile for the Brinkman-Darcy coupling, using an augmented finite element formulation.

a $\mathbb{RT}_0 - \mathbb{ND}_2 - \mathbb{RT}_0 - \mathbf{P}_0 - \mathbf{P}_0 - \mathbf{P}_1$ FE family. Now the Brinkman domain $\Omega_B = [0,3]\times[0,0.2]\times[1,3/2]$ is on top of the Darcy domain $\Omega_D = [0,3]\times[0,0.2]\times[0,1]$ (as in e.g. a two-layer subsurface flow). These domains are discretized into structured tetrahedral meshes of 63195 and 94847 elements, respec-

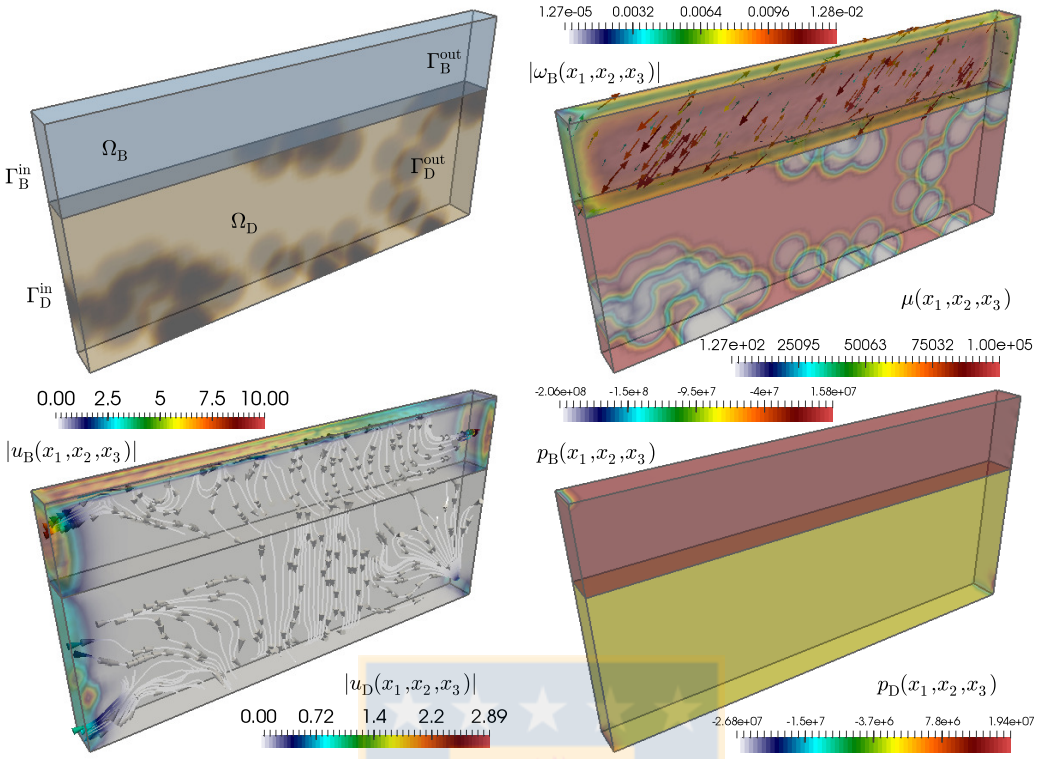


Figure 5.5: Example 4: Two-domain geometry and boundaries (top left), approximated Brinkman vorticity and Darcy permeability field (top right), approximated velocity (bottom left), and computed pressure profile (bottom right) for the Brinkman-Darcy coupling imposing Dirichlet boundary conditions for the velocity.

tively. The parameter $\mu$ (viscosity over permeability of the Darcy domain) is now highly heterogeneous (see top-left panel of Figure 5.5), and the remaining parameters take the values $\alpha = 1/100$, $\nu = 1$, $\kappa = 2\nu/\alpha$. We also assume the presence of a slight current along with gravity in the Brinkman domain, i.e. $\boldsymbol{f}_B = (1, 0, -0.98)^t$. The boundary conditions for velocity are: $\boldsymbol{u}_B = (100(x_3 - 1)(3/2 - x_3), 0, 0)^t$ on $\Gamma_B^{in} \cup \Gamma_B^{out}$, $\boldsymbol{u}_B = (100(1 - x_1^2/9), 0, 0)^t$ on the top surface, $\boldsymbol{u}_D = (1, 0, 0)^t$ on $\Gamma_D^{in}$, and on the sink at the lower part of $\Gamma_D^{out}$, and no-slip Darcy velocity ($\boldsymbol{u}_D = \boldsymbol{0}$) on the bottom surface. No other conditions are imposed (e.g. on the front or back sides of the domain). Notice that Dirichlet conditions are implemented via a penalization strategy. We stress that even if this setting is not covered by our analysis, the obtained results (see the remaining panels in Figure 5.5) suggest that the augmented formulation proposed herein is capable to successfully handle problems involving Dirichlet velocity data. Nevertheless, the performance of the method in conditions where vorticity develops into more involved patterns, remains to be addressed.

## Conclusions

In this thesis we developed mixed finite element methods for a set of partial differential equations of physical interest in transport problems through viscous flows in porous media. Theoretical results that guarantee the well-posedness of the proposed methods, as well as the corresponding numerical tests and simulations, have been provided.

The main conclusions of this work are:

1. We have introduced an augmented mixed–primal finite element method for a coupled flow–transport problem, where the flow is described by the Stokes equations (with variable viscosity) and the transport of species is governed by a convection–diffusion equation. The original problem was reformulated through an augmented variational approach for the fluid flow coupled with a primal formulation for the transport model. Next, by a fixed point strategy together with suitable regularity assumptions, we were able to develop the corresponding analysis of solvability. Consequently, we derived an augmented mixed–primal finite element method, and then we showed its well-posedness. In particular, a feasible choice of finite element subspaces for the associated Galerkin scheme, was given by Raviart-Thomas elements of order $k$ for the pseudostress, and piecewise polinomials of degree $\leq k+1$ for both velocity of fluid and scalar field. Afterwards, suitable Strang-type inequalities were utilized to rigorously derive *a priori* error estimates in their natural norms. Finally, several numerical experiments validating the good performance of the method and confirming the corresponding rates of convergence were reported. The numerical examples reported includes a benchmark test of thermal convection employing a Boussinesq approximation, as well as the simulation of the steady state of a clarifying-thickening process.

2. We provided the *a posteriori* error analysis for the augmented mixed–primal finite element method associated to the flow–transport coupling. Here two efficient and reliable residual-based *a posteriori* error estimators for that scheme, were derived. In particular, to show the reliability of the first estimator, a Helmholtz decomposition for $\mathbb{H}(\mathbf{div};\Omega)$ with homogeneous Neumann boundary condition was introduced for the 2D case. Finally, several numerical results illustrating the reliability and efficiency of the estimators, and showing the expected behavior of the associated adaptive algorithm were provided.

3. We extended the approach employed for the analysis of coupled flow–transport problem to the more realistic case of a sedimentation-consolidation system. More precisely, we introduced an augmented mixed–primal formulation for the Brinkman problem with variable viscosity; coupled with a scalar nonlinear convection–diffusion equation. Then, we showed the well-posedness of

the continuous and discrete formulations. In particular, for the Galerkin scheme, we employed Raviart-Thomas spaces of order $k$ for the Cauchy stress, and continuous piecewise polynomials of degree $\leq k+1$ for the velocity and also for the scalar field associated to the concentration. Next, we derived the corresponding *a priori* error analysis. Finally, we presented some numerical results confirmig the predicted rates of convergence, and highlighting the good performance of the method. In particular, we applied the proposed method to simulate a batch sedimentation within a cylinder with a contraction, and also to the simulation of the steady state of flow patterns, under the so-called Coanda effect, in sedimentation processes.

4. By similar arguments to those employed in the *a posteriori* error analysis for the flow–transport coupling, we derived two efficient and reliable residual-based *a posteriori* error estimators for the augmented mixed–primal finite element method associated to the sedimentation-consolidation system. In particular, in this work we presented the 2D and 3D versions of these estimators for that scheme. Finally, we provided several numerical test confirmig the reliability and efficiency of the estimators, and illustrating the good performance of the corresponding adaptive algorithm.

5. We derived a vorticity-based fully-mixed finite element method to numerically approximate the flow patterns in an heterogeneous media composed by a porous medium, where Darcy equations govern the flow behavior of a non-viscous incompressible fluid, and a much more permeable medium, where the lamimnar flow exhibits viscous effects described by the linear Brinkman model. The system was formulated in terms of velocity and pressure in the porous medium, together with vorticity, velocity and pressure of the viscous fluid. In contrast with others works available in the literature such as [18] and [19], we do not assume here that the fluid boundary coincides with the interface between both domains. In addition, here we established the pressure continuity across the interface using a Lagrange multiplier, whereas the normal stress conditions were weakly imposed. On the other hand, for the resulting mixed variational formulation, higher regularity of the fluid pressure as in e.g. [18] was not required. Next, we applied the classical Babuška-Brezzi theory together with the so-called $T$-coercivity approach to show the well-posedness of the continuous formulation and the corresponding discrete scheme. For the associated mixed finite element method it was required that the **curl** of the finite element subspace approximating the velocity be contained in the space where the discrete velocity of the fluid lives. In this way, a feasible choise of finite element subspaces for the Galerkin scheme was given by Raviart-Thomas for the velocities, Nédélec elements for vorticity, continuous piecewise polynomials for the pressures, and discontinuous piecewise polynomials for the Lagrange multiplier. Alternatively, we showed that the aforementioned constraint can be avoided by augmenting the mixed formulation with a residual-type term arising from the Brinkman momentum equation. Finally, we reported several numerical examples illustrating the satisfactory performance of the methods and confirming the theoretical rates of convergence.

## Future works

1. We plan to derive the *a posteriori* error analysis for the proposed methods in Chapter 5. We expect to provided reliable and efficient residual-based *a posteriori* error estimators for the fully-mixed finite element method and also for the augmented mixed finite element method, respectively.

2. We are interested in extending the results and techniques of Chapters 3 and 5 to analyze the solvability of a Brinkman-Darcy-Transport coupling. At first glance, the extension of such techniques looks immediate, however even though the analysis for the coupled Brinkman-Darcy problem can be easily derived from Chapter 5, a suitable handle to estimate the trilinear term $\int_\Omega \phi \boldsymbol{u} \cdot \nabla \psi$ appearing in the primal formulation for the transport problem, is required for both the continuous and discrete levels. In particular, the global velocity $\boldsymbol{u}$ initially belongs to $\mathbf{H}(\mathrm{div};\Omega)$, and we can not exploit augmentation techniques to recover a $\mathbf{H}^1(\Omega)$ velocity as was the case in Chapter 3. It is because of this that we expect to introduce suitable assumptions on $\boldsymbol{u}$ and to derive appropiate fixed-point strategies for the continuous and discrete case, respectively, in order to establish the solvability analysis.

3. We plan to study the transient state of the Brinkman-Darcy-Transport coupling in order to simulate more realistic physical phenomena concerning biomedical and environment applications such as, the flow of air in the lungs, perfusion of soft living tissues, $CO_2$ sequestration, contamination of the groundwater, and the flow of air in the atmospheric boundary layer over vegetation, among others.

4. Finally, a longer term objective of much interest, is the extension of the mathematical and numerical techniques introduced in **Chapter 3**, to the study of *degenerate* diffusion equations modelling the compaction mechanisms leading to physically relevant sedimentation-consolidation models. We anticipate the use of entropy solutions, discretizations involving low regularity coefficients, and conservative methods.

## Conclusiones

En esta tesis se desarrollaron métodos de elementos finitos mixtos para un conjunto de ecuaciones diferenciales parciales de interés físico en problemas de transporte a través de flujos viscosos en medios porosos. A su vez, todos los resultados teóricos que garantizan el buen planteamiento de los métodos propuestos así como los tests y las simulaciones numéricas correspondientes, han sido proporcionados.

Las principales conclusiones de este trabajo son:

1. Se introdujó un método de elementos finitos mixto-primal aumentado para un problema acoplado de flujo con transporte, donde la dinámica del fluido está descrita por las ecuaciones de Stokes (con viscosidad variable) y el transporte de especies se rige por una ecuación de convección–difusión. El problema original fue reformulado mediante un enfoque variacional aumentado para el flujo del fluido acoplado con una formulación primal para el modelo de transporte. Seguidamente, a través de una estrategía de punto fijo junto con supuestos de regulidad adecuados, fue posible desarrollar el análisis de solubilidad correspondiente. Consecuentemente, se derivó un método de elementos finitos mixto-primal aumentado, y luego se probó que el mismo estaba bien puesto. En particular, se estableció que es posible emplear los espacios de Raviart-Thomas de grado $k$ para el pseudo-esfuerzo, así como polinomios a trozos de grado $\leq k+1$ para la velocidad del fluido y el campo escalar. Después, desigualdades adecuadas de tipo Strang fueron utilizadas para derivar rigurosamente las estimaciones de error *a priori* en las normas naturales. Finalmente, se reportaron varios experimentos numéricos que validaron el buen desempeño del método y que confirmaron los ordenes de convergencia correspondientes. Los ejemplos numéricos reportados incluyen un test de convección térmica empleando una aproximación de Boussinesq, así como la simulación del estado estacionario de un proceso de clarificación-espesamiento.

2. Se proporcionó el análisis de error *a posteriori* para el método de elementos finitos mixto-primal aumentado asociado al acoplamiento de flujo con transporte. Aquí se derivaron dos estimadores de error *a posteriori*, de tipo residual, confiables y eficientes para ese esquema. En particular, para probar la confabilidad del primer estimador, una descomposición de Helmholtz para $\mathbb{H}(\mathbf{div};\Omega)$ con condición de frontera de Neumann homogénea, fue introducida para el caso 2D. Finalmente, se proporcionaron varios resultados numéricos que ilustraron la confiabiidad y la eficiencia de los estimadores, y que también mostraron el comportamiento esperado del algoritmo adaptativo asociado.

3. Se logró extender el enfoque empleado para el análisis del problema acoplado de flujo-transporte al caso más realista de un sistema de sedimentación-consolidación. Más precisamente, se intro-

dujó una formulación mixta-primal aumentada para el problema de Brinkman con viscosidad variable; acoplado con una ecuación escalar no lineal de convección-difusión. Luego, se demostró que las formulaciones continuas y discretas estaban bien puestas. En particular, para el esquema de Galerkin, se emplearon espacios de Raviart-Thomas de orden $k$ para el tensor de Cauchy, y polinomios continuos a trozos de grado $\leq k+1$ para la velocidad y también para el campo escalar asociado a la concentración. Seguidamente, se derivó el análisis de error *a priori* correspondiente. Finalmente, se presentaron algunos resultados numéricos que confirmaron los ordenes de convergencia predichos, y que destacaron el buen desempeño del método. En particular, el método propuesto se aplicó para simular una sedimentación de lotes dentro de un cilindro con contracción, y también para la simulación del estado estacionario de patrones de flujo, bajo el llamado efecto Coanda, en procesos de sedimentación.

4. Mediante argumentos similares a los que se emplearon en el análisis de error *a posteriori* para el acoplamiento de flujo con transporte, se logró derivar dos estimadores de error *a posteriori*, de tipo residual, confiables y eficientes para el método de elementos finitos mixto-primal aumentado asociado al sistema de sedimentación-consolidación. En particular, en este trabajo se presentaron las versiones 2D y 3D de estos estimadores para ese esquema. Finalmente, se proporcionaron varios tests numéricos que confirmaron la confiabilidad y la eficiencia de los estimadores, y que ilustraron el buen desempeño del algoritmo adaptativo correspondiente.

5. Se derivó un método de elementos finitos completamente mixto basado en vorticidad para aproximar los patrones de flujo en un medio heterogéneo compuesto por un medio poroso, donde las ecuaciones de Darcy gobiernan el comportamiento del flujo de un fluido incompresible no viscoso, y un medio más permeable, donde los flujos laminares exhiben efectos viscosos descritos mediante el modelo lineal de Brinkman. El sistema fue formulado en términos de la velocidad y la presión en el medio poroso, junto con la vorticidad, la velocidad y la presión del fluido viscoso. En contraste con otros trabajos disponibles en la literatura como [18] y [19], en este trabajo no se asumió que la frontera del fluido coincide con la interfaz entre ambos dominios. Además, aquí la continuidad de la presión a través de la interfaz se estableció utilizando un multiplicador de Lagrange, mientras que las condiciones de las tensiones normales fueron impuestas débilmente. Por otra parte, para la formulación variacional mixta resultante no se necesitó alta regularidad de la presión del fluido como por ejemplo en [18]. Después, se aplicó la teoría clásica de Babuška-Brezzi junto con el enfoque conocido como $T$-coercividad para mostrar que la formulación continua y el esquema discreto correspondiente estaban bien puestos. Para el método de elementos finitos mixto asociado se requerió que el **curl** del subespacio de elementos finitos que aproxima la velocidad esté contenido en el espacio donde vive la velocidad discreta del fluido. De esta manera, para el esquema de Galerkin se recomienda elegir los subespacios de elementos finitos dados por; Raviart-Thomas para las velocidades, elementos de Nédélec para la vorticidad, polinomios continuos a trozos para las presiones, y discontinuos a trozos para el multiplicador de Lagrange. Alternativamente, se mostró que la restricción mencionada anteriormente se puede evitar aumentando la formulación mixta con un término de tipo residual que surge a partir de la ecuación de momentum de Brinkman. Finalmente, se reportaron varios ejemplos numéricos que ilustraron el desempeño satisfactorio de los métodos y que confirmaron los ordenes teóricos de convergencia.

## Trabajos futuros

1. Se derivará el análisis de error *a posteriori* para los métodos introducidos en el Capítulo 5. Se espera proporcionar estimadores de error *a posteriori*, de tipo residual, confiables y eficientes para el método de elementos finitos completamente mixto y también para el aumentado, respectivamente.

2. Interesa extender los resultados y las técnicas introducidas en los Capítulos 3 y 5 para analizar la solubilidad de un acoplamiento Brinkman-Darcy con transporte. A primera vista, la extensión de tales técnicas parece inmediata, sin embargo aún cuando el análisis para el problema acoplado Brinkman-Darcy se puede derivar facilmente a partir del Capítulo 5, se requiere de un manejo adecuado, tanto a nivel continuo como discreto, para estimar el término trilineal $\int_\Omega \phi \boldsymbol{u} \cdot \nabla \psi$ que aparece en la formulación para el problema de transporte. En particular, la velocidad global $\boldsymbol{u}$ inicialmente vive en $\mathbf{H}(\mathrm{div};\Omega)$, y no se pueden explotar técnicas de aumento para recuperar una velocidad en $\mathbf{H}^1(\Omega)$ como lo fue en el Capítulo 3. Es debido a esto que se espera introducir supuestos adecuados sobre $\boldsymbol{u}$ y derivar estrategias de punto fijo apropiadas para el caso continuo y discreto, respectivamente, con la finalidad de establecer el análisis de solubilidad.

3. Se estudiará el estado transitorio del acoplamiento Brinkman-Darcy-Transporte, con la finalidad de simular fenómenos físicos más realistas en relación con la salud humana y el medio ambiente, tales como; el flujo de aire en los pulmones, perfusión de tejidos blandos, captación de $CO_2$, contaminación de aguas subterraneas, y el flujo de aire en la capa límite atmosférica sobre la vegetación, entre otros.

4. Finalmente, un objetivo a largo plazo de mucho interés es la extensión de las técnicas matemáticas y numéricas, introducidas en el **Capítulo 3**, al estudio de las ecuaciones de difusión degene rada que modelan los mecanismos de compactación que conducen a modelos, físicamente relevantes, de sedimentación-consolidación. Anticipamos que se utilizaran soluciones de entropía, discretizaciones que involucran bajos coeficientes de regularidad, y métodos conservativos.

# References


[1] R. A. Adams and J. J. F. Fournier, *Sobolev Spaces*, Academic Press, Elsevier Ltd, 2003.

[2] M. Ainsworth, J. Guzmán, and F. J. Sayas, *Discrete extension operators for mixed finite element spaces on locally refined meshes*, Mathematics of Computation, 85 (2016), pp. 2639–2650.

[3] A. Alonso and A. Valli, *An optimal domain decomposition preconditioner for low-frequency time-harmonic Maxwell equations*, Mathematics of Computation, 68 (1999), pp. 607–631.

[4] M. Álvarez, G. N. Gatica, and R. Ruiz-Baier, *An augmented mixed-primal finite element method for a coupled flow-transport problem*, ESAIM: Mathematical Modelling and Numerical Analysis, 49 (2015), pp. 1399–1427.

[5] ——, *A mixed-primal finite element approximation of a sedimentation-consolidation system*, Mathematical Models and Methods in Applied Sciences, 26 (2016), pp. 867–900.

[6] ——, *A posteriori error analysis for a sedimentation-consolidation system*, Preprint 2016-26, Centro de Investigación en Ingeniería Matemática ($CI^2MA$), Universidad de Concepción, Chile, (2016).

[7] ——, *A posteriori error analysis for a viscous flow-transport problem*, ESAIM: Mathematical Modelling and Numerical Analysis, 50 (2016), pp. 1789–1816.

[8] ——, *Analysis of a vorticity-based fully-mixed formulation for the 3D brinkman-darcy problem*, Computer Methods in Applied Mechanics and Engineering, 307 (2016), pp. 68–95.

[9] P. R. Amestoy, 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.

[10] C. Amrouche, C. Bernardi, M. Dauge, and V. Girault, *Vector potentials in three-dimensional nonsmooth domains*, Mathematical Methods in the Applied Sciences, 21 (1998), pp. 823–864.

[11] V. Anaya, G. N. Gatica, D. Mora, and R. Ruiz-Baier, *An augmented velocity - vorticity - pressure formulation for the Brinkman equations*, International Journal for Numerical Methods in Fluids, 79 (2015), pp. 109–137.

[12] V. Anaya, D. Mora, R. Oyarzúa, and R. Ruiz-Baier, *A priori and a posteriori error analysis of a mixed scheme for the Brinkman problem*, Numerische Mathematik, 133 (2016), pp. 781–817.

[13] T. Arbogast and D. S. Brunson, *A computational method for approximating a Darcy-Stokes system governing a vuggy porous medium*, Computational Geosciences, 11 (2007), pp. 207–218.

[14] J. L. Auriault, *On the domain of validity of Brinkman's equation*, Transport in Porous Media, 79 (2009), pp. 215–223.

[15] I. Babuška and G. N. Gatica, *A residual-based a posteriori error estimator for the Stoke-Darcy coupled problem*, SIAM Journal on Numerical Analysis, 48 (2010), pp. 498–523.

[16] L. Badea, M. Discacciati, and A. Quarteroni, *Numerical analysis of the Navier-Stokes/Darcy coupling*, Numerische Mathematik, 115 (2010), pp. 195–227.

[17] C. Bernardi, L. E. Alaoui, and Z. Mghazli, *A posteriori analysis of a space and time discretization of a nonlinear model for the flow in partially saturated porous media*, IMA Journal of Numerical Analysis, 34 (2014), pp. 1002–1036.

[18] C. Bernardi, F. Hecht, and F. Z. Nouri, *A new finite-element discretization of the Stokes problem coupled with the Darcy equations*, IMA Journal of Numerical Analysis, 30 (2010), pp. 61–93.

[19] C. Bernardi, F. Hecht, and O. Pironneau, *Coupling Darcy and Stokes equations for porous media with cracks*, ESAIM: Mathematical Modelling and Numerical Analysis, 39 (2005), pp. 7–35.

[20] F. Betancourt, R. Bürger, R. Ruiz-Baier, H. Torres, and C. A. Vega, *On numerical methods for hyperbolic conservation laws and related equations modelling sedimentation of solid-liquid suspensions*, in : G.-Q. Chen, H. Holden and K.H. Karlsen (eds.), Hyperbolic Conservation Laws and Related Analysis with Applications, Berlin, 2014, Springer-Verlag, pp. 23–68.

[21] D. Boffi, F. Brezzi, and M. Fortin, *Mixed Finite Element Methods and Applications*, Springer-Verlag, 2013.

[22] A. S. Bonnet-Ben Dhia, L. Chesnel, and P. Ciarlet, *T-coercivity for scalar interface problems between dielectrics and metamaterials*, ESAIM: Mathematical Modelling and Numerical Analysis, 46 (2012), pp. 1363–1387.

[23] A. E. Boycott, *Sedimentation of blood corpuscules*, Nature, 104 (1920), p. 532.

[24] M. Braack and T. Richter, *Solving multidimensional reactive flow problems with adaptive finite elements*, in : Reactive flows, diffusion and transport, Berlin, 2007, Springer-Verlag, pp. 93–112.

[25] M. Braack and F. Schieweck, *Equal-order finite elements with local projection stabilization for the Darcy-Brinkman equations*, Computer Methods in Applied Mechanics and Engineering, 200 (2011), pp. 1126–1136.

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

[27] M. Bulíček and P. Pustějovská, *Existence analysis for a model describing flow of an incompressible chemically reacting non-Newtonian fluid*, SIAM Journal on Numerical Analysis, 46 (2014), pp. 3223–3240.

[28] R. Bürger, S. Kumar, and R. Ruiz-Baier, *Discontinuous finite volume element discretization for coupled flow-transport problems arising in models of sedimentation*, Journal of Computational Physics, 299 (2015), pp. 446–471.

[29] R. Bürger, C. Liu, and W. L. Wendland, *Existence and stability for mathematical models of sedimentation–consolidation processes in several space dimensions*, Journal of Computational Physics, 264 (2001), pp. 288–310.

[30] R. Bürger, R. Ruiz-Baier, K. Schneider, and H. Torres, *A multiresolution method for the simulation of sedimentation in inclined channels*, International Journal of Numerical Analysis and Modelling, 9 (2012), pp. 479–504.

[31] R. Bürger, R. Ruiz-Baier, and H. Torres, *A stabilized finite volume element formulation for sedimentation-consolidation processes*, SIAM Journal on Scientific Computing, 34 (2012), pp. B265–B289.

[32] R. Bürger, W. L. Wendland, and F. Concha, *Model equations for gravitational sedimentation-consolidation processes*, 80 (2000), pp. 79–92.

[33] M. C. Bustos, F. Concha, R. Bürger, and E. M. Tory, *Sedimentation and Thickening*, Kluwer Academic Publishers, Dordrecht, 1999.

[34] Z. Cai, B. Lee, and P. Wang, *Least-squares methods for incompressible Newtonian fluid flow: linear stationary problems*, SIAM Journal on Numerical Analysis, 42 (2004), pp. 843–859.

[35] J. Camaño, G. N. Gatica, R. Oyarzúa, R. Ruiz-Baier, and P. Venegas, *New fully-mixed finite element methods for the Stokes-Darcy coupling*, Computer Methods in Applied Mechanics and Engineering, 295 (2015), pp. 362–395.

[36] J. Camaño, R. Oyarzúa, and G. Tierra, *Analysis of an augmented mixed-FEM for the Navier Stokes problem*, Mathematics of Computation, DOI: https://doi.org/10.1090/mcom/3124, (2016).

[37] C. Carstensen, *A posteriori error estimate for the mixed finite element method*, Mathematics of Computation, 66 (1997), pp. 465–476.

[38] C. Carstensen and G. Dolzmann, *A posteriori error estimates for mixed FEM in elasticity*, Numerische Mathematik, 81 (1998), pp. 187–209.

[39] L. Chesnel and P. Ciarlet, *T-coercivity and continuous Galerkin methods: application to transmission problems with sign changing coefficients*, Numerische Mathematik, 124 (2013), pp. 1–129.

[40] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, Nort-Holland, 1978.

[41] ——, *Linear and Nonlinear Functional Analysis with Applications*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2013.

[42] P. Clément, *Approximation by finite element functions using local regularization*, Modélisation Mathématique et Analyse Numérique, 9 (1975), pp. 77–84.

[43] E. Colmenares, G. N. Gatica, and R. Oyarzúa, *An augmented fully-mixed finite element method for the stationary Boussinesq problem*, Calcolo, DOI: http://link.springer.com/article/10.1007/s10092-016-0182-3.

[44] ——, *Analysis of an augmented mixed-primal formulation for the stationary Boussinesq problem*, Numerical Methods for Partial Differential Equations, 32 (2016), pp. 445–478.

[45] C. Cox, H. Lee, and D. Szurley, *Finite element approximation of the non-isothermal Stokes-Oldroyd equations*, International Journal of Numerical Analysis and Modelling, 4 (2007), pp. 425–440.

[46] M. Dauge, *Elliptic Boundary Value Problems on Corner Domains. Smoothness and Asymptotics of Solutions. Lecture Notes in Mathematics*, 1341 (1998).

[47] G. de Vahl Davis, *Natural convection of air in a square cavity: A benchmark numerical solution*, International Journal for Numerical Methods in Fluids, 3 (1983), pp. 249–264.

[48] D. A. Di Pietro, E. Flauraud, M. Vohralík, and S. Yousef, *A posteriori error estimates, stopping criteria, and adaptivity for multiphase compositional Darcy flows in porous media*, Journal of Computational Physics, 276 (2014), pp. 163–187.

[49] M. Discacciati and A. Quarteroni, *Navier-Stokes/Darcy coupling: modeling, analysis, and numerical approximation*, Revista Matemaática Complutense, 22 (2009), pp. 315–426.

[50] C. Domínguez, G. N. Gatica, and S. Medahi, *A posteriori error analysis of a fully-mixed finite element method for a two-dimensional fluid-solid interaction problem*, Journal of Computational Mathematics, 33 (2015), pp. 606–641.

[51] F. El Chami, G. Mansour, and T. Sayah, *Error studies of the coupling Darcy-Stokes system with velocity-pressure formulation*, Calcolo, 49 (2012), pp. 73–93.

[52] V. J. Ervin, E. W. Jenkins, and S. Sun, *Coupled generalized nonlinear Stokes flow with flow through a porous medium*, SIAM Journal on Numerical Analysis, 47 (2009), pp. 929–952.

[53] M. Farhloul, S. Nicaise, and L. Paquet, *A mixed formulation of Boussinesq equations: Analysis of nonsingular solutions*, Mathematics of Computation, 69 (2000), pp. 965–986.

[54] ——, *A refined mixed finite element method for the Boussinesq equations in polygonal domains*, IMA Journal of Numerical Analysis, 21 (2001), pp. 525–551.

[55] M. Farhloul and A. Zine, *A dual mixed formulation for non-isothermal Oldroyd-Stokes problem*, Mathematical Modelling of Natural Phenomena, 6 (2011), pp. 130–156.

[56] L. E. Figueroa, G. N. Gatica, and N. Heuer, *A priori and a posteriori error analysis of an augmented mixed finite element method for incompressible fluid flows*, Computer Methods in Applied Mechanics and Engineering, 198 (2008), pp. 280–291.

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

[58] T. Fusegi and J. M. Hyun, *A numerical study of 3D natural convection in a cube: effects of the horizontal thermal boundary conditions*, 8 (1991), pp. 221–230.

[59] J. Galvis and M. Sarkis, *Non-matching mortar discretization analysis for the coupling Stokes-Darcy equations*, Electronic Transactions on Numerical Analysis, 26 (2007), pp. 350–384.

[60] A. I. Garralda-Guillen, G. N. Gatica, A. Márquez, and M. Ruiz, *A posteriori error analysis of twofold saddle point variational formulations for nonlinear boundary value problems*, IMA Journal of Numerical Analysis, 34 (2014), pp. 326–361.

[61] G. N. Gatica, *A note on the efficiency of residual-based a-posteriori error estimators for some mixed finite element methods*, Electronic Transactions on Numerical Analysis, 17 (2004), pp. 218–233.

[62] ——, *Analysis of a new augmented mixed finite element method for linear elasticity allowing* $\mathbb{RT}_0 - \mathbb{P}_1 - \mathbb{P}_0$ *approximations*, ESAIM: Mathematical Modelling and Numerical Analysis, 40 (2006), pp. 1–28.

[63] ——, *A Simple Introduction to the Mixed Finite Element Method: Theory and Applications*, Springerbriefs in Mathematics, Springer Cham Heidelberg New York Dordrecht London, 2014.

[64] ——, *A note on stable Helmholtz decomposition in 3D*, Preprint 2016-03, Centro de Investigación en Ingeniería Matemática (CI$^2$MA), Universidad de Concepción, Chile, (2016).

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

[66] ——, *Analysis of a pseudostress-based mixed finite element method for the Brinkman model of porous media flow*, Numerische Mathematik, 126 (2014), pp. 635–677.

[67] G. N. Gatica, L. F. Gatica, and F. A. Sequeira, *A priori and a posteriori error analyses of a pseudostress-based mixed formulation for linear elasticity*, Computers and Mathematics with Applications, 71 (2016), pp. 635–677.

[68] G. N. Gatica and G. C. Hsiao, *On the coupled BEM and FEM for a nonlinear exterior Dirichlet problem in* $\mathrm{R}^2$, Numerische Mathematik, 61 (1992), pp. 171–214.

[69] G. N. Gatica, G. C. Hsiao, and S. Meddahi, *A coupled mixed finite element method for the interaction problem between and electromagnetic field and elastic body*, SIAM Journal on Numerical Analysis, 48 (2010), pp. 1338–1368.

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

[71] ——, *Analysis of the coupling of Lagrange and Arnold-Falk-Winther finite elements for a fluid-solid interaction problem in 3D*, SIAM Journal on Numerical Analysis, 50 (2012), pp. 1648–1674.

[72] 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.

[73] ——, *A priori and a posteriori error analyses of a velocity-pseudostress formulation for a class of quasi-Newtonian Stokes flows*, Computer Methods in Applied Mechanics and Engineering, 200 (2011), pp. 1619–1636.

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

[75] ——, *A twofold saddle point approach for the coupling of fluid flow with nonlinear porous media flow*, IMA Journal of Numerical Analysis, 32 (2012), pp. 845–887.

[76] G. N. Gatica and W. Wendland, *Coupling of mixed finite elements and boundary elements for linear and nonlinear elliptic problems*, Applicable Analysis, 63 (1996), pp. 39–75.

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

[78] P. Grisvard, *Elliptic Problems in Nonsmooth Domains. Monographs and Studies in Mathematics*, vol. 24, Pitman (Advanced Publishing Program), Boston, MA, 1985.

[79] ——, *Singularities in Boundary Value Problems. Recherches en Mathématiques Appliquées*, vol. 22, Springer, Berlin, 1992.

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

[81] A. Khalili, A. J. Basu, U. Pietrzyk, and B. B. Jørgensen, *Advective transport through permeable sediments: a new numerical and experimental approach*, 132 (1999), pp. 221–227.

[82] M. G. Larson and A. Målqvist, *Goal oriented adaptivity for coupled flow and transport problems with applications in oil reservoir simulations*, Computer Methods in Applied Mechanics and Engineering, 196 (2007), pp. 3546–3561.

[83] M. G. Larson, R. Söderlund, and F. Bengzon, *Adaptive finite element approximation of coupled flow and transport problems with applications in heat transfer*, International Journal for Numerical Methods in Fluids, 57 (2008), pp. 1397–1420.

[84] Y. Le Pentrec and G. Lauriat, *Effects of the heat transfer at the side walls on natural convection in cavities*, Journal of Heat Transfer, 112 (1990), pp. 370–378.

[85] M. Lesinigo, C. D'Angelo, and A. Quarteroni, *A multiscale Darcy-Brinkman model for fluid flow in fractured porous media*, Numerische Mathematik, 117 (2011), pp. 717–752.

[86] W. McLean, *Strongly Elliptic Systems and Boundary Integral Equations*, Cambridge University Press, Cambridge.

[87] P. Monk, *Finite Element Methods for Maxwell's Equations*, Oxford University Press, New York.

[88] Nečas, *Introduction to the Theory of Nonlinear Elliptic Equations, Reprint of the 1983 edition. A Wiley-Interscience Publication*, John Wiley & Sons, Ltd., Chichester.

[89] M. Ouriemi, P. Aussillous, and E. Guazzelli, *Sediment dynamics. Part 1. Bed-load transport by laminar shearing flows*, Journal on Fluid Mechanics, 636 (2009), pp. 295–319.

[90] R. Oyarzúa, T. Qin, and D. Schötzau, *An exactly divergence-free finite element method for a generalized Boussinesq problem*, IMA Journal of Numerical Analysis, 34 (2014), pp. 1104–1135.

[91] K. V. Parchevsky, *Numerical simulation of sedimentation in the presence of 2D comprenssible convection and reconstruction of the particle-readius distribution function*, Journal of Engineering Mathematics, 41 (2001), pp. 203–219.

[92] A. Quarteroni and A. Valli, *Numerical Approximation of Partial Differential Equations*, vol. 23.

[93] R. Rao, L. Mondy, and S. Albotelli, *Instabilities during batch sedimentation in geometries containing obstacles: A numerical and experimental study*, International Journal for Numerical Methods in Fluids, 55 (2007), pp. 723–735.

[94] B. Rivière, *Analysis of a discontinuous finite element method for the coupled Stokes and Darcy problems*, Journal of Scientific Computing, 22 (2005), pp. 479–500.

[95] J. E. Roberts and J. M. Thomas, *Mixed and Hybrid Methods*, In Handbook of Numerical Analysis, edited by P.G. Ciarlet and J.L Lions, vol. II, Finite Elements Methods (Part 1), Nort-Holland, Amsterdam, 1991.

[96] R. Ruiz-Baier and H. Torres, *Numerical solution of a multidimensional sedimentation problem using finite volume-element methods*, Applied Numerical Mathematics, 95 (2015), pp. 280–291.

[97] S. B. Savage, *Gravity flow of cohesionless granular materials in chutes and channels*, Journal on Fluid Mechanics, 92 (1979), pp. 53–96.

[98] U. Schaflinger, *Experiments on sedimentation beneath downward-facing inclined walls*, International Journal of Multiphase Flows, 11 (1985), pp. 189–199.

[99] S. Sun and M. F. Wheeler, *Local problem-based a posteriori error estimators for discontinuous Galerkin approximations of reactive transport*, Computational Geosciences, 11 (2007), pp. 87–101.

[100] R. Tarpagkou and A. Pantokratos, *CFD methodology for sedimentation tanks: The effcects of secondary phase on fluid phase using DPM coupled calculations*, Applied Mathematical Modelling, 37 (2013), pp. 3478–3494.

[101] D. J. Tritton, *Physical Fluid Dynamics*, Van Nostrand Reinhold, 1980.

[102] R. Verfürth, *A posteriori error estimation and adaptive-mesh-refinement techniques*, Journal of Computational and Applied Mathematics, 50 (1994), pp. 67–83.

[103] ——, *A Review of A Posteriori Error Estimation and Adaptive-Mesh-Refinement Techniques*, Wiley-Teubner (Chichester), 1996.

[104] M. Vohralík and M. F. Wheeler, *A posteriori error estimates, stopping criteria, and adaptivity for two-phase flows*, Computational Geosciences, 17 (2013), pp. 789–812.