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

NUMERICAL METHODS FOR THE SIMULATION OF VISCOUS FLOW AND TRANSPORT IN POROUS MEDIA

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

Paul Esteban Méndez Silva
concepción-chile
2020

Profesor Guía: Raimund Bürger
CI^{2}MA y Departamento de Ingeniería Matemática
Universidad de Concepción, Chile

Cotutor: Ricardo Ruiz-Baier
School of Mathematics
Monash University, Australia

Numerical methods for the simulation of viscous flow and transport in porous media

Paul Esteban Méndez Silva

Directores de Tesis: Raimund Bürger, Universidad de Concepción, Chile.
Ricardo Ruiz-Baier, Monash University, Australia.

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

Comisión evaluadora

Prof. Silvia Jerez Galiano, Centro de Investigación en Matemáticas, CIMAT, México.

Prof. Rolf Stenberg, Aalto University, Finland.

Prof. Sander Rhebergen, University of Waterloo, Canada

Comisión examinadora

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

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

Firma: _____
Prof. Fernando Betancourt, Universidad de Concepción, Chile.

Firma: _____
Prof. Carlos Zambra, Universidad de Talca, Chile.

Calificación: _____

Concepción, 12 de Marzo de 2020

Abstract

This thesis is concerned with the mathematical and numerical analysis of partial differential equations (PDE)-based models for the coupling of flow equations and transport arising from problems related with the simulation of transport phenomena and chemical interactions within saturated porous media. This framework is encountered in a vast variety of engineering applications such as polymer flooding in petroleum extraction, wastewater treatment, food and chemicals processing, chromatography and others. Among the applications mentioned, those that motivated the development of this work are mainly related to the design of equipment used in water treatment. This includes settlers, clarifiers/thickeners, and filtration equipment. However, we point out, that other applications where envision and developed during the work on the general models. In fact, this dissertation also includes some results related to traffic flow, bioconvection and thermohaline circulation. Other extensions which require more extensive modifications or additions, such as fluid-structure interactions are briefly discussed in the "current and future works" section, at the end of this thesis.

As a quick overview, in **Chapter** 2 we begin by studying the phenomenon of sedimentation, in the first instance, through polydisperse sedimentation models, considering from the numerical point of view a finite volume method with entropy conservation properties. In **Chapter** 3, we introduce models for the coupling of flow and transport equations motivated by the study of double-diffusive flows. Here we change the approach of the numerical approximation, to focus on the finite element method, with divergence-free approximations for velocity. The analysis and numerical scheme designed for the non-stationary setting is then extended, to develop a second approach for the sedimentation phenomenon, which in turn, motivated **Chapter** 4. It is a complete three-dimensional model for clarifiers, where we incorporate the one-dimensional Kynch density function describing hindered settling, used in the first approach, in a transport equation coupled with a Navier-Stokes-Brinkman model for the flux. Finally **Chapter** 5, discuss the application to the modelling of soil-based water filters of a similar scheme adapted to the context of an axisymmetric domain and a non-stationary system.

The main contents of this thesis are structure as follows:

In **Chapter** 1, we briefly introduce the topics that will be addressed in this thesis. We also discuss the relevant literature and related papers and present a summary of the main contributions of this thesis work. The chapter closes by introducing the notation that will be used regularly in the following chapters.

Chapter 2 discusses entropy conservative schemes for diffusively corrected multiclass kinematic flow models. As a new contribution, we demonstrate, firstly, that these schemes can naturally be extended to initial-boundary value problems with zero-flux boundary conditions in one space dimension,

including an explicit bound on the growth of the total entropy. Secondly, it is shown that the model assumptions are satisfied by certain diffusively corrected multiclass kinematic flow models of arbitrary size that describe traffic flow or the settling of dispersions and emulsions, where the latter application gives rise to zero-flux boundary conditions. Numerical examples illustrate the behaviour and accuracy of entropy stable schemes for these applications.

In **Chapter** 3 we present our model for double-diffusive flows, which adopts the form of the incompressible Brinkman-Navier-Stokes equations for the viscous flow of an incompressible Newtonian fluid in a porous medium, coupled to a pair of advection-diffusion equations with cross-diffusion that describe the diffusion of heat and solute. The solvability analysis of these governing equations results as a combination of compactness arguments and fixed-point theory. Also, an $\boldsymbol{H}(\mathrm{div})$-conforming discretisation is formulated by a modification of existing methods for Brinkman flows. The well-posedness of the discrete Galerkin formulation is also discussed, and convergence properties are derived rigorously. Computational tests confirm the predicted rates of error decay and illustrate the applicability of the methods.

Then in **Chapter** 4 we take this model as a basis for the development of the non-stationary setting model, applied to the simulation of sedimentation-consolidation of solid particles in an incompressible fluid under the effect of gravity and in the presence of a slowly rotating arm assisting the removal of sediment on the bottom of clarifier-thickener units. The governing equations now include an initial-boundary value problem for the Navier-Stokes equations describing the flow of the mixture coupled with a nonlinear parabolic equation describing the volume fraction of solids. The novelty of the treatment consists in the inclusion of terms that account for the influence of the rake motion on the momentum balance and the removal of solids. We also adopt techniques of the immersed boundary finite element method (see e.g. [34]) for the analysis and numerical approximation of those terms. An $\boldsymbol{H}(\mathrm{div})$-conforming method for the coupled problem is proposed, a rigorous proof of convergence is provided, and the validity of the new model and the performance of the scheme are demonstrated numerically by several computational tests. Our aim after completing the presentation of these two approaches is to have complementary models that can help in gaining a better understanding of the sedimentation process inside clarifiers/thickeners.

In **Chapter** 5 a related model, but in an axisymmetric domain is developed with an explicit application to the modelling of soil-based water filtering devices. The governing equations are the Brinkman-Navier-Stokes equations for the flow of the fluid through a porous medium coupled with a convection-diffusion equation for the transport of the contaminants plus a system of ordinary differential equations accounting for the degradation of the adsorption properties of each contaminant. These equations are written in meridional axisymmetric form and the corresponding weak formulation adopts a mixed-primal structure. As in the previous cases, a second-order, (but axisymmetric) divergence-conforming discretisation of this problem is introduced and the solvability, stability, and spatio-temporal convergence of the numerical method are analysed.

Chapter 6 is devoted to discuss the conclusions of this work. We also describe the new topics that we are addressing as an extension of the models presented in this thesis. In this regard, we place special emphasis on a fluid-structure model for the study of blood coagulation in veins. The purpose of the investigation initiated in this subject is to study the effects of platelet count, shear rate and injury size on the initiation of blood coagulation. The model consists of a system of advection-

diffusion-reaction equations describing the spatio-temporal distributions of blood coagulation factors and platelet subtypes during thrombus development, coupled with Navier-Stokes equations to describe the dynamics of blood flow in the vessel. We address the problem of the fluid-structure interaction within the blood vessel using the immersed boundary method that was briefly introduced in **Chapter** 4. We describe the spatio-temporal discretisation, including a semi-implicit scheme for time integration and show some qualitative preliminary results. We conclude this chapter by discussing some of the problems we would like to address in the future.

Resumen

El objetivo de esta tesis es el análisis matemático y numérico de modelos basados en ecuaciones diferenciales parciales (PDE) para el acoplamiento de ecuaciones de flujo y transporte, que surgen de problemas relacionados con la simulación de fenómenos de transporte e interacciones químicas dentro de medios porosos saturados. Este marco teórico se encuentra en una gran variedad de aplicaciones de ingeniería tales como la inyección de polímeros en extracción de petróleo, tratamiento de aguas residuales, procesamiento de químicos y alimentos, cromatografía, entre otros. De entre estas aplicaciones, las que motivaron el desarrollo del presente trabajo, son aquellas relacionadas con el diseño de equipos para el tratamiento de agua, tales como sedimentadores, clarificadores/espesadores y filtros. Sin embargo, hacemos notar, que a lo largo de este trabajo se identificaron muchas otras aplicaciones para los modelos matemáticos generales. Es por esto, que se incluye además resultados concernientes a fenómenos de tráfico vehicular, bioconvección y circulación termohalina. Otras extensiones más complejas de los modelos estudiados, tales como interacciones fluido-estructura se discuten brevemente en la sección de trabajos en marcha y a futuro, al final de la tesis.

Haciendo una vista rápida de la tesis, comenzamos estudiando el fenómeno de sedimentación, en primera instancia, a través de modelos de sedimentación polidispersa, considerando desde el punto de vista numérico, un método de volúmenes finitos con propiedades de conservación de entropía. Este desarrollo y aplicaciones adicionales de los modelos más generales de flujo cinemático multiclase corregido por difusión se presentan en el **Capítulo** 2. En el **Capítulo** 3, introducimos modelos para el acoplamiento de ecuaciones de flujo y transporte motivados por el estudio de flujos doble difusivos. Aquí cambiamos el enfoque de la aproximación numérica, para centramos en el método de elementos finitos, con aproximaciones de divergencia libre para la velocidad. Luego, el análisis y el esquema numérico diseñados para el caso no estacionario se extienden para tratar un segundo enfoque para el problema de sedimentación, el cual motiva el **Capítulo** 4. Este corresponde a un modelo tridimensional para clarificadores/espesadores, donde incorporamos la función de densidad unidimensional de Kynch que describe sedimentación obstaculizada, en una ecuación de transporte acoplada con el modelo de flujo incompresible de Navier-Stokes-Brinkman. El **Capítulo** 5, trata sobre la aplicación al modelado de filtros de agua basados en suelos de un esquema similar adaptado al contexto de un dominio axisimétrico y un sistema no estacionario.

Los contenidos principales de la tesis se estructuran como sigue:

En el **Capítulo** 1, introducimos brevemente los tópicos que serán abordados en esta tesis. También discutimos la literatura relevante y los trabajos relacionados, y presentamos un resumen de las principales contribuciones de este trabajo de tesis. El capítulo se cierra introduciendo la notación que se usará de forma recurrente en los siguientes capítulos.

En el **Capítulo** 2 abordamos los esquemas que conservan entropía para flujo cinemático multiclase corregido por difusión. Como nueva contribución, demostramos, primero, que estos esquemas se pueden extender de forma natural a problemas de valores iniciales y de frontera con condiciones de borde de flujo cero en una dimensión espacial, incluyendo una cota explícita sobre el crecimiento de la entropía total. Segundo, se muestra que las suposiciones del modelo, son satisfechas por modelos de flujo cinemático multiclase corregidos por difusión de tamaño arbitrario, tales como los que describen el flujo de tráfico o la sedimentación de dispersiones y emulsiones. Esta última aplicación es la que requiere el uso de condiciones de borde tipo flujo cero. Los ejemplos numéricos que acompañan ese capítulo, ilustran el comportamiento y la precisión de los esquemas de entropía estable para las aplicaciones mencionadas.

En el **Capítulo** 3 presentamos nuestro modelo para flujos doble-difusivos, el cual adopta la forma de las ecuaciones para el flujo en un medio poroso de un fluido viscoso no Newtoniano incompresible de Navier-Stokes-Brinkman, acoplado a un par de ecuaciones de advección-difusión con difusión cruzada que describe la transferencia de calor y de masa de un soluto. El análisis de existencia y unicidad de las ecuaciones gobernantes resulta de una combinación de argumentos de compacidad y teoría de punto fijo. Adicionalmente, se formula una discretización $\boldsymbol{H}(\mathrm{div})$-conforme, partiendo de una modificación de métodos existentes para flujos de Brinkman. También discutimos la existencia y unicidad de soluciones para la formulación discreta de Galerkin, y las propiedades de convergencia de la misma se derivan de forma rigurosa. Los ensayos computacionales confirman las tasas de convergencia del error e ilustran la aplicabilidad de los métodos para la simulación de problemas de bio-convección bacteriana y circulación termohalina.

Después, en el **Capítulo** 4 tomamos este modelo y desarrollamos su extensión natural al contexto no-estacionario, aplicado a un modelo macroscópico para la simulación de sedimentación-consolidación de partículas sólidas en un fluido incompresible bajo el efecto de la gravedad y en presencia de un brazo que gira lentamente, asistiendo con la remoción de sedimentos del fondo de una unidad de clarificación-espesamiento. El modelo gobernante es un problema de valores iniciales y de contorno para las ecuaciones de Navier-Stokes describiendo el flujo de una mezcla acoplado con ecuaciones no-lineales parabólicas que describen la fracción en volumen de sólidos. La novedad de nuestro tratamiento consiste en la inclusión de términos que modelan la influencia del movimiento rotatorio de la rastra, tanto en el balance de momento como en el balance de masa de la fase sólida. Además, adaptamos técnicas del método de elementos finitos con frontera inmersa (ver por ejemplo [34]) para el análisis y aproximación numérica de los términos mencionados. Se propone un método $\boldsymbol{H}(\mathrm{div})$ conforme para el problema acoplado, junto con una prueba rigurosa de convergencia, y se demuestra mediante varios ensayos numéricos la validez del nuevo modelo y su rendimiento. Nuestro objetivo al completar la presentación de estos dos enfoques, es que sean modelos complementarios, que puedan ayudar a adquirir un mejor entendimiento del proceso de sedimentación en clarificadores/espesadores.

En el **Capítulo** 5, presentamos un modelo relacionado, pero en el dominio axisimétrico con aplicación explícita al modelado de equipos de filtrado de agua basados en suelos. Las ecuaciones gobernantes son las de Navier-Stokes-Brinkman para el flujo de fluido, acoplados con una ecuación de convección-difusión para el transporte de contaminantes, más un sistema de ecuaciones diferenciales ordinarias para la degradación de las propiedades de adsorción de cada contaminante. Estas ecuaciones están escritas en forma axisimétrica meridional y la correspondiente formulación débil adopta una estructura

mixta-primal. Como en los casos anteriores, desarrollamos un esquema discreto de segundo orden (pero axisimétrico) y se analiza la existencia de soluciones, estabilidad y convergencia espacio-temporal del método numérico.

En el **Capítulo** 6 discutimos las principales conclusiones de este trabajo. También describimos los nuevos temas que estamos abordando como una extensión de los modelos presentados en esta tesis. Ponemos especial énfasis en un modelo de fluido-estructura para el estudio de coagulación en venas. El propósito de esta investigación es el estudio de los efectos del conteo de plaquetas, velocidad de corte y tamaño de la herida en la iniciación de la coagulación. El modelo consiste de un sistema de ecuaciones de advección-difusión-reacción describiendo la distribución espacio-temporal de los factores de coagulación y subtipos de plaquetas durante el desarrollo tromboso, acoplados con las ecuaciones de Navier-Stokes para la descripción de la dinámica del flujo sanguíneo en el vaso. Abordamos el problema de la discretización del modelo fluido-estructura haciendo uso del método de frontera inmersa, que se presentó brevemente en el **Capítulo** 4. Describimos además un método discreto semi-implício para la integración temporal y mostramos algunos resultados cualitativos preliminares. Cerramos este capítulo discutiendo algunos de los tópicos que nos interesa investigar en el futuro.

Agradecimientos

Quiero comenzar expresando mi sincera gratitud a mi director de tesis, profesor Raimund Bürger, primero por su gestión como director del programa, por toda su predisposición a asistirme con la información y la documentación que requería desde el primer contacto que tuve con el programa. Gracias también por su guía y apoyo como profesor y luego como director de tesis.

Agradezco de igual manera a mi co-director Ricardo Ruiz-Baier por toda su dedicación a este trabajo, por su constante apoyo y paciencia; agradezco su invitación para visitar y trabajar con él en Instituto de Matemáticas de la Universidad de Oxford, la amabilidad con la que me recibió y el entusiasmo con el que guió mi trabajo durante mi estadía.

A Rodolfo Rodriguez gracias por las enseñanzas de sus clases y su gestión como director de programa. De igual manera gracias al prof. Gabriel Gatica, por su gestión como director del CI^{2}MA, por asegurarse que siempre tuviéramos un espacio de trabajo y equipamiento adecuados, así como su asistencia para mi participación/organización de diferentes eventos.

Quiero agradecer a todos mis profesores en el programa: Raimund Bürger, Rodolfo Rodriguez, Leonardo Figueroa, Mauricio Sepúlveda, Manuel Solano, Rodolfo Araya, Pep Mulet, por sus enseñanzas y la cordialidad que siempre me mostraron. De igual forma, gracias por toda su ayuda, al personal administrativo del CI^{2}MA y del Departamento de Ingeniería Matemática de la Universidad de Concepción, Sra. Cecilia Leiva, Sra. Lorena Carrasco, Sra. Paola Castro, Sr. Jorge Muñoz y Sr. Iván Tobar.

También mi agradecimiento para todas las personas que me brindaron su amistad en estos años chilenos y no chilenos, por contribuir a hacer de mi estancia en Chile una gran experiencia. Gracias a todos mis amigos del doctorado, todos aquellos con los que compartí clase y/o esos agradables momentos de convivencia en el CI^{2}MA: Néstor, Rodrigo, Daniel, Víctor, Juan Paulo, Romel, Yolanda, Yissedt, Adrián, Joaquín, Eduardo, Mario, Ramiro, Sergio, Paulo, Iván y Mauricio. Un agradecimiento especial para mis compañeros de generación: Bryan, Cristian, Willian y Rafael, gracias por su amistad, su asistencia siempre que lo necesité, por ayudarme a mantenerme con ánimo positivo y por todos los buenos momentos compartidos.

Agradezco a CONICYT (PFCHA/DOCTORADO NACIONAL/2016-21160005) y al programa de Becas Convocatoria Abierta 2016, Senescyt, Ecuador por financiar mi investigación, pasantía y estudios de doctorado.

Finalmente, no podría terminar sin extender mi más profundo agradecimiento a mi familia, a mis padres, mi hermano Miguel Angel y mi hermana Gabriela. Gracias por su amor y constante aliento.

Paul Esteban Méndez Silva

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

The simulation of transport phenomena and chemical interactions within saturated porous media is a framework encountered in a vast variety of scientific and engineering applications, including the flow of chemical pollutants in saturated soil, subsurface drilling and petroleum extraction, crystal growth, chemical and food processing, polymer flooding as part of the process of enhanced oil recovery in petroleum engineering [46], chromatography [141], or water decontamination and removal of pollutants such as heavy metals or radioactive ions [160], and numerous other applications [24, 86, 88, 119, 136, 139, 150, 166].

The aim of this thesis work is the mathematical and numerical analysis of partial differential equations (PDE)-based models for the coupling of flow equations and transport. Challenges in these physico-mathematical problems concern the strong nonlinearities involved, as well as a dynamical behaviour characterised by very different time scales. Among the applications mentioned, those that motivated the development of this thesis are mainly related to the design of equipment used in water treatment. This includes settlers, clarifiers/thickeners, and filtration equipment. We begin by studying the phenomenon of sedimentation, firstly, through polydisperse sedimentation models, considering from the numerical point of view, a finite volume method with entropy conservation properties. The numerical scheme showed errors and errors rates comparable with the other tested methods (Kurganov-Tadmor and component-wise Global Lax-Friedrichs) and it seemed to have some computational performance advantages for some applications, however, we also identified some shortcomings of the method, including explicit requirements over the form of the diffusion term, and the need of fine-tuning stabilization parameters. We remark also that the method is a one-dimensional simplification of the sedimentation process. All of this motivated us to explore complementary approaches.

The second approach to sedimentation modelling, which motivates **Chapter** 4, is a complete three-dimensional model for clarifiers/thickeners, where we incorporate the one-dimensional Kynch density function in a transport equation coupled with the Brinkman-Navier-Stokes equations for incompressible flow. We also consider the presence of a rotating rake assisting the sedimentation. Previously, in **Chapter** 3, we discuss a double-diffusive flow model that will be used as the basis for the development of the scheme. We believe that the two approaches, studied here, can help to gain a better understanding of the operation of clarification units.

The other research line of this work involves the study of finite element formulations for partial

differential equations modelling the coupling of flow and transport. As we mentioned earlier, the first model studied in this line, takes as a motivational problem the more general double-diffusive phenomena in a stationary setting. Double-diffusive flows originate in combining heat and mass transfer interacting with flow within porous structures. The model adopts the form of the incompressible Brinkman-Navier-Stokes equations for the viscous flow of an incompressible Newtonian fluid in a porous medium, coupled to a pair of advection-diffusion equations with cross-diffusion that describe the diffusion of heat and solute.

We then extend the previous model and its analysis to an axisymmetric domain with an explicit application to the modelling of soil-based water filtering devices. The governing equations are the Brinkman-Navier-Stokes equations for the flow of the fluid through a porous medium coupled with a convection-diffusion equation for the transport of the contaminants plus a system of ordinary differential equations accounting for the degradation of the adsorption properties of each contaminant. These equations are written in the meridional axisymmetric form and the corresponding weak formulation adopts a mixed-primal structure. Under this model, it is assumed that each site has a maximum capacity for each contaminant, which we take to be uniform across the two layers of filter media. In this way, the adsorption is noncompetitive and the saturation of a site by one contaminant does not prevent adsorption of the other contaminants at the same site. It is also assumed that the adsorption process is irreversible for all contaminants and all filter layers, so that once adsorbed the contaminants remain attached to the filter media with no desorption back into the fluid.

Note that an important component of this thesis work is devoted to the solvability of the associated PDEs using fixed-point theories. We also work in the construction of accurate, robust and reliable methods for the discretisation of these equations, and special emphasis is placed in $\boldsymbol{H}(\mathrm{div})$-conforming formulations for the flow equations, whereas for the formulation of the transport problem (resulting in a scalar or vectorial advection-diffusion equation) we study entropy stable schemes for stand-alone systems as well as finite element primal formulations when coupled with the flow equations. The main advantage of an $\boldsymbol{H}(\mathrm{div})$-conforming formulation is that it produces exactly divergence-free velocity approximations, which are of particular importance in ensuring that solutions to the flow equations remain locally conservative as well as energy stable (see e.g. [62]). Moreover, the error estimates of velocity can be derived in a pressure-robust manner (see [100]).

The $\boldsymbol{H}(\mathrm{div})$-conforming discretisation is introduced in **Chapter** 3 by a modification of existing methods for Brinkman flows. The well-posedness of the discrete Galerkin formulation is also discussed, and convergence properties in space are derived rigorously. In **Chapter** 4, we extend this analysis to the non-stationary case, and in consequence, we derive spatio-temporal convergence properties. In **Chapter** 5, this analysis is modified to account for the axisymmetric domain, which involves the work with weighted spaces. In all cases, the validity of the models and the performance of the schemes are demonstrated numerically by several computational tests.

Finally, in **Chapter** 6, we discuss the main conclusions derived from each chapter. Moreover, we present our current and ongoing investigation, which incorporates fluid-structure interactions to the framework developed through this thesis, aimed at addressing problems with biological applications. Our motivation is to study the effects of platelet count, shear rate and injury size on the initiation of blood coagulation. The work is an extension of the mathematical model for clot growth dynamics proposed in [39], to which we add fluid-structure interactions through the use of the immersed

boundary method with Lagrange multiplier introduced in [37]. The base model consists of advection-diffusion-reaction equations describing the spatio-temporal distributions of blood coagulation factors and platelet subtypes during thrombus development, coupled with the Navier-Stokes equations to describe the dynamics of blood flow in the vessel. We describe the spatio-temporal discretisation, including a semi-implicit scheme for time integration and show some qualitative preliminary results. The chapter concludes by discussing some perspectives regarding future work, motivated by our results, and suggestions/proposals from collaborators we met during the course of this thesis work.

1.1 Related Work

To put this work further into the proper perspective, we mention that a large number of references to the well-posedness and numerical analysis of degenerate convection-diffusion equations are provided in [99]. However, the existence and uniqueness of entropy solutions, and the convergence of numerical methods have so far only been established in the scalar case ($N = 1$); important contributions in this direction include [57, 73, 103–106, 110] (this list is far from being complete). This state of matters is in agreement with the well-known lack of corresponding results for general first-order systems of conservation laws (2.4) considering that (2.1) reduces to (2.4) wherever $\boldsymbol{K} = \mathbf{0}$. That said, we mention that degenerate convection-diffusion systems (2.1) arise in a number of applications such as multiclass vehicular traffic [29, 30, 38, 53, 54, 158], settling of polydisperse solid-liquid suspensions [38, 43, 54, 151], settling of dispersions of droplets and emulsions [1, 51, 146], and chromatography [52, 70]. In particular, in these applications systems of convection-diffusion equations (rather than scalar equations) arise because one wishes to describe the segregation of different classes of units of the disperse phase (cars, particles, droplets, etc.), with the consequence that the number of species N in these applications can be arbitrarily large. These applications motivate the interest in developing efficient solvers for the numerical solution of (2.1), (2.2) or (2.1)–(2.3) even if there is no closed well-posedness theory for these systems. Common numerical schemes are based on a space discretisation which can be finite volumes or discontinuous Galerkin methods [127], while the time discretisation could be fully explicit or implicit-explicit (IMEX; see for example [38, 51, 54]). On the explicit side, a well-known scheme is the Kurganov-Tadmor high-resolution central difference scheme [116]. The original KT scheme was proposed alongside high-order convex combinations of Runge Kutta time stepping. The latter concept was developed further on, resulting in the so called Strong Stability Preserving Runge-Kutta (SSPRK) methods. These schemes allow for a high-order time discretisation while preserving the strong stability properties of first-order Euler time stepping, which makes them attractive for solving hyperbolic partial differential equations by the method of lines [85].

In other hand, concerning the well-posedness of double diffusive systems (under suitable assumptions), we first restrict the discussion to classical Boussinesq-type equations. The solvability of the associated PDEs goes back to Lorca and Boldrini [125, 126]. These works include existence, regularity, and conditions for uniqueness addressing both stationary and non-stationary cases. These results hold for temperature-dependent viscosity and thermal conductivity. Related to the context of our specific problem, the analysis of solutions to double-diffusive problems has been addressed e.g. in [88, 124].

A diversity of numerical methods is available for classical Boussinesq equations as well as for their generalisations to temperature-dependent coefficients. We mention for instance the stabilised finite

elements (using projection-based techniques) proposed and rigorously analysed in [6, 59], the mixed formulations analysed in [5, 9, 10, 63], but also the stability of splitting schemes (for discontinuous Galerkin, spectral, and vorticity-based finite element formulations) and some more applicative examples have been explored in [2, 8, 46, 119, 120, 136, 149, 150, 163]. Mixed-primal and fully-mixed schemes using $\boldsymbol{H}(\mathrm{div})$-conforming velocity approximations have been studied in [137, 138].

One advantage of including a diffusion matrix in the model is that it allows us to study cross-diffusion effects, such as the Soret and Dufour effects. Even when in some applications these can be neglected as their contributions can be orders of magnitude smaller than those described by terms arising from Fourier's or Fick's law, these effects can be significant when species are introduced at a surface in a fluid domain and have different densities in comparison to the surrounding fluid. These mechanisms are important as well in applications related to the transport of moisture in fibrous insulations or grain storage insulations and the dispersion of contaminants through water-saturated soil, bio-chemical contaminants transport in environmental problems, and underground disposal of nuclear waste and crystal growth processes [24].

Other contributions to this area include the finite volume discretisations for thermal and solutal buoyancy within Darcy-Brinkman flows introduced in [86], the error analysis for spectral methods applied to bioconvection in [66], or the vorticity-based Brinkman and nonlinear advection-reaction-diffusion system analysed via fixed-point and compactness arguments in [14], that also includes a mixed-primal scheme featuring divergence-free discrete velocities. Penalty Petrov-Galerkin methods were used for the solution of double-diffusion convective problems in [93]. In [155] the authors introduce least-squares schemes specifically tailored for Rayleigh-Bénard convective flows, and the averaging finite element method has been employed in [165] for solidification problems having the same structure as the models we examine here.

On the side of numerical schemes that provide divergence-free velocity fields approximations for incompressible flows, the work of V. Jhon et al [100] provides a good review of the different approaches that have been proposed and the importance of the divergence-free property for the computation of pressure robust velocity approximations. Some of the methods currently available include conforming finite element pairs obtained using exterior calculus techniques [74] or by enriching $H(\mathrm{div})$-conforming elements locally [91], discontinuous Galerkin (DG) methods with postprocessing [25] and hybridizable DG finite-element methods (HDG) with and without postprocessing [60, 143]. In [96], the authors present a space-time HDG method for the Navier-Stokes problem on time-dependent domains that results in pointwise divergence-free and H(div)-conforming velocity fields. It is shown that the scheme is momentum conserving, energy-stable, and pressure-robust. The study of more efficient schemes for the numerical approximation of the problems developed in this thesis, keeping the good properties of the numerical method currently used, is a topic that we want to address in the future.

Early models for the clarifying process with and without swirl effects are reviewed in [152], where mainly axisymmetric configurations were employed. More recently, a fairly complete model can be found in [65], where the authors couple the momentum equations for fluid flow with a transport equation for solids. The realisable $k-\epsilon$ model, in conjunction with scalable wall functions, is used to model turbulence. The removal of sludge from the clarifier floor by means of a spinning rake is modelled through a rotating sink term added to the right-hand side of the transport equation. References that are related to the rake mechanism in applications of mineral processing include [61,

84, 89, 90, 121, 147, 156, 167].

In the model studied in **Chapter** 4, we also include appropriate drag terms, much as in [161], that account for the indirect effects of the rake on the flow patterns. This consists basically in penalising the moving structure and computing (or as we do here, simply imposing) its velocity and its reconfiguration in an adequate manner. Volume penalisation techniques can be frequently found in the relevant literature. See for instance [108], where the authors propose high-order methods for the modelling of solid obstacles as porous structures whose permeability tends to zero and the flow is regarded in a unified domain, and the momentum on the obstacle is simply obtained from integration of the penalised velocity over the obstacle domain. Other modelling and numerical approaches one could use to incorporate the interaction between the rake and the flow include immerse boundary and fictitious domain finite elements [35], level set methods and their variants [69, 135], other unfitted finite element schemes [22]; or formulations based on remodelling, such as the arbitrary Lagrangian-Eulerian (ALE) setting [164].

A time dependent Boussinesq model with nonlinear viscosity depending on the temperature is proposed in [4]. The authors analyze first and second order numerical schemes based on finite element methods and derive an optimal a priori error estimate for each numerical scheme. A related non-stationary phase-change Boussinesq model is presented in [163], where a second order finite element method for the primal formulation of the problem in terms of velocity, temperature, and pressure is constructed, and conditions for its stability are provided.

The coupling of advection-diffusion-reaction systems with Brinkman equations in their velocity-vorticity-pressure formulation, is studied in [120]. The equations are discretised in space using mixed FE methods on unstructured meshes, whereas the time integration hinges on an operator splitting strategy that uses the differences in scales between the reaction, advection, and diffusion processes. The authors compare several coupling strategies in terms of memory usage, iteration count, speed of calculation, and dynamics of the energy norm.

Regarding our axisymmteric model applied to the study of filtration equipment, we mention that several studies treat the axisymmetric formulation of the Stokes and Navier-Stokes flows, including the discretisation employing spectral, mortar, and stabilized finite elements (see e.g. [18, 26, 27, 31, 72], and references cited in these works). More recently, mixed formulations of Brinkman flow including the numerical analysis of finite element (FE) approximations were studied. Anaya et al. [15] presented an augmented finite element approximation for the Brinkman equations based on an extension of the vorticity-based Stokes problem. A related recent model in [16] incorporates a stream function and vorticity formulation of axisymmetric Brinkman flow, for which a conforming mixed FE approximation is employed.

The numerical analysis of the axisymmetric Darcy and Stokes-Darcy flow using Raviart-Thomas (RT) and Brezzi-Douglas-Marini (BDM) finite elements was presented in [71, 72]. In [71], the authors established the stability of the RT and BDM approximations for an axisymmetric Darcy flow problem by extending the Stenberg criteria, and they also derive a priori error estimates.

Other contributions to the design of numerical methods for axisymmetric formulations of coupled flow and transport problems include [12, 55]. Furthermore, in [47] a semi-discrete discontinuous finite volume element (FVE) scheme is proposed and the unique solvability of both the nonlinear continuous

problem and the semi-discrete counterpart is discussed. An FVE method is also proposed in [55] to discretise a Stokes equation for flow coupled with a parabolic equation modelling sedimentation. The method is based on a stabilized discontinuous Galerkin formulation for the concentration field, and a multiscale stabilized pair of $\mathcal{P}_1$-$\mathcal{P}_1$ elements for velocity and pressure, respectively. A mixed variational formulation of a Darcy-Forchheimer flow coupled with a energy equation is semi-discretised in [12] using Raviart-Thomas elements for fluxes and piecewise constant elements for the pressure, a posteriori error estimates are also established.

The technological application behind the water filter model goes back to the observation that it is possible to remove arsenic from water by passing it through iron-rich laterite soil [131, 132]. The arsenic is removed through an adsorption process, which may be enhanced by chemically treating the laterite to increase its porosity and surface area, improving the adsorption efficiency [130]. Clearly, the formulation of accurate mathematical models of these filters, in addition to their efficient computational solution, would greatly aid in the development of improved filters and guidelines for their safe operation. The development and analysis of such a model forms the basis of the work [134], where the authors examined the removal of a single contaminant (arsenic; case $m = 1$ in our notation) in a cylindrical filter of uniform media. The authors utilised a Darcy-Brinkman equation, coupled with an advection-diffusion-adsorption equation to model the flow of the contaminated water through the filter and the removal of the arsenic through adsorption. In practice, however, there are likely $m > 1$ contaminants present, which calls for a filter consisting of multiple (up to m) layers in order to allow for their removal. In this work we attempt to study the filtration process in a soil-based water filter consisting of two distinct layers of differing media, in the presence of multiple contaminant species.

Problems of a similar nature abound in the literature. For example, [86] considers the numerical solution, via a finite volume method, of a double diffusive problem within a porous medium. The paper [150] considers a similar double diffusive problem, however, much like our proposed layered filter, the authors allow for the possibility of heterogeneous stratified porous media. While many of the studies concerning double diffusive problems consider entirely closed domains filled with porous media, a large number of application cases, such as our filter, feature partial enclosures with openings or infiltrations. The article [154] introduces such a feature, with the addition of 'free ports' to their model domain. Considering other potential variants, the authors of [166] extend the usual double diffusive problem by a first-order reaction process between the diffusing species and the fluid. This reaction process necessitates the addition of a sink term to the equation governing the species concentration that plays a role similar to that on the right-hand side of (5.1c).

1.2 Contributions of this thesis

In **Chapter** 2, we extend the analysis of a class of entropy stable schemes for the numerical solution of initial value problems of nonlinear, possibly strongly degenerate systems of convection-diffusion equations, recently proposed in [S. Jerez, C. Parés, *Entropy stable schemes for degenerate convection-diffusion equations*, SIAM J. Numer. Anal. vol. 55 (2017) pp. 240–264]. As a new contribution, we demonstrate, firstly, that these schemes can naturally be extended to initial-boundary value problems with zero-flux boundary conditions in one space dimension, including an explicit bound on the growth of the total entropy. Secondly, it is shown that the model assumptions are satisfied by certain diffusively

corrected multiclass kinematic flow models of arbitrary size that describe traffic flow or the settling of dispersions and emulsions. The contents of this chapter gave rise to the following paper:

[48] Bürger, R., Méndez, P. E., Parés, C., *On entropy stable schemes for degenerate parabolic multispecies kinematic flow models.* **Numer Methods Partial Differential Eq**. 1– 26; (2019)

In **Chapter** 3 we propose a model for double-diffusive flows, that includes possible cross-diffusion terms. The main differences between the available well-posedness results and analysis of $\boldsymbol{H}(\mathrm{div})$-conforming methods for classical Boussinesq equations and the double-diffusive equations (3.1) are, of course, caused by the vector-valued nature of the quantities (the components of $\vec{m}$) that diffuse in (3.1) while in the classical Boussinesq formulation there is only one scalar diffusive quantity (for instance, solely temperature). Some of the arguments related to the well-posedness analysis of the continuous problem, in particular those related to handling non-homogenous Dirichlet data by a lifting argument [125, 137], carry over almost verbatim from the scalar to the vectorial case. However, the bilinear form associated with the term $-\mathbf{div}(\mathbb{D}\nabla\vec{m})$ must be coercive so that stability is ensured. This requirement, in turn, imposes restrictions on the choice of the diffusion matrix $\mathbb{D}$; this matrix must be positive definite (though not necessarily symmetric). These properties are essential for the proof of existence of a discrete solution, however, it is still possible to study cross-diffusion effects, such as those of Soret and Dufour. In addition, an $\boldsymbol{H}(\mathrm{div})$-conforming discretisation is formulated by a modification of existing methods for Brinkman flows. The well-posedness of the discrete Galerkin formulation is also discussed, and convergence properties in space are derived rigorously. The contents of this chapter gave rise to the following paper

[49] Bürger, R., Méndez, P.E., Ruiz-Baier, R., *On $\boldsymbol{H}(\mathrm{div})$-conforming methods for double-diffusion equations in porous media.* **SIAM Journal on Numerical Analysis**, 57, 1318–1343 (2019)

In **Chapter** 4 we introduce a new model for the simulation of sedimentation-consolidation of solid particles in an incompressible fluid under the effect of gravity and in the presence of a slowly rotating arm assisting the removal of sediment on the bottom of clarifier-thickener units. The governing model is now an initial-boundary value problem for the Navier-Stokes equations describing the flow of the mixture coupled with a nonlinear parabolic equation describing the volume fraction of solids. The novelty of the treatment consists in the inclusion of terms that account for the influence of the rake motion on the momentum balance and the removal of solids. We also adapt techniques of the immersed boundary finite element method (see e.g. [34]) for the analysis and numerical approximation of those terms. We derive rigorously the spatio-temporal convergence properties of the divergence conforming numerical scheme, and demonstrate its properties through several numerical tests. The contents of this chapter gave rise to the following preprint

[50] Bürger, R., Méndez, P.E., Ruiz-Baier, R., *A second-order H(div)-conforming scheme for the simulation of sedimentation and flow in circular clarifiers with a rotating rake.* **Centro de Investigación en Ingeniería Matemática** (CI^{2}MA), Preprint 2019-39, Universidad de Concepción, Chile 2019.

In **Chapter** 5 we extend the previous model to an axisymmetric domain with an explicit application to the modelling of soil-based water filtering devices. Furthermore, we derived an optimal a priori error estimate for the $\boldsymbol{H}(\mathrm{div})$-conforming second-order numerical scheme in time and space, where the main difficulty is the fully discrete analysis verifying that each of the terms is bounded optimally in the corresponding weighted spaces. Some numerical examples illustrate the main features of the problem and the properties of the numerical scheme. The contents of this chapter gave rise to the following preprint

1.3 Notations

From **Chapter** 3 and onward we will consider the following notations:

Let Ω be an open and bounded domain in $\mathbb{R}^d$, $d = 2, 3$ with Lipschitz boundary $\Gamma = \partial\Omega$. We denote by $L^p(\Omega)$ and $W^{r,p}(\Omega)$ the usual Lebesgue and Sobolev spaces with respective norms $\|\cdot\|_{L^p(\Omega)}$ and $\|\cdot\|_{W^{r,p}(\Omega)}$. If $p = 2$ we write $H^r(\Omega)$ in place of $W^{r,p}(\Omega)$, and denote the corresponding norm by $\|\cdot\|_{r,\Omega}$, ($\|\cdot\|_\Omega$ for $H^0(\Omega) = L^2(\Omega)$). The space $L^2_0(\Omega)$ denotes the restriction of $L_2(\Omega)$ to functions with zero mean value over Ω. For $r \geq 0$, we write the H^r-seminorm as $|\cdot|_{r,\Omega}$ and we denote by $(\cdot,\cdot)_\Omega$ the usual inner product in $L^2(\Omega)$. In addition, $H^{1/2}(\Gamma)$ is the space of traces of functions $H^1(\Omega)$ and $H^{-1/2}(\Gamma)$ is its dual. Spaces of vector-valued functions (in dimension d) are denoted in bold face, and we denote by $\mathbb{L}$ the corresponding tensorial counterpart of the scalar variable or functional space L. Next, we denote by $L^s(0,\mathcal{T}; W^{m,p}(\Omega))$ the Banach space of all L^s-integrable functions from $[0,\mathcal{T}]$ into $W^{m,p}(\Omega)$, with norm

$$\|v\|_{L^s(0,\mathcal{T};W^{m,p}(\Omega))} = \begin{cases} \left(\int_0^{\mathcal{T}} \|v(t)\|^s_{W^{m,p}(\Omega)}\,\mathrm{d}t\right)^{1/s} & \text{if } 1 \leq s < \infty, \\ \operatorname{esssup}_{t\in[0,\mathcal{T}]} \|v(t)\|_{W^{m,p}(\Omega)} & \text{if } s = \infty. \end{cases}$$

Let us denote by $\mathcal{T}_h$ a regular partition of Ω into simplices K (triangles in 2D or tetrahedra in 3D) of diameter h_K. The mesh size will be denoted by $h = \max\{h_K,\ K \in \mathcal{T}_h\}$, and for any interior edge e in $\mathcal{E}_h$ (the set of faces in $\mathcal{T}_h$), we will label K^- and K^+ the elements adjacent to it, while h_e will stand for the maximum diameter of the edge. We suppose that $\boldsymbol{v}, w$ are smooth vector and scalar fields defined over $\mathcal{T}_h$. Then, by $(\boldsymbol{v}^\pm, w^\pm)$ we will denote the traces of $(\boldsymbol{v}, w)$ on e being the extensions from the interiors of the elements K^+ and K^-, respectively. Let $\boldsymbol{n}_e$ denote the outward unit normal vector to e on K, we define the tangential component of $\boldsymbol{u}$ on each face e as $\boldsymbol{u}_{\boldsymbol{\tau}} = \boldsymbol{u} - (\boldsymbol{u}\cdot\boldsymbol{n}_e)\boldsymbol{n}_e$. We introduce the average $\{\!\{\cdot\}\!\}$ and jump $[\![\cdot]\!]$ operators as follows:

$$\{\!\{\boldsymbol{v}\}\!\} = (\boldsymbol{v}^- + \boldsymbol{v}^+)/2, \quad \{\!\{w\}\!\} = (w^- + w^+)/2,$$
$$[\![\boldsymbol{v}]\!] = (\boldsymbol{v}^- - \boldsymbol{v}^+), \quad [\![w]\!] = (w^- - w^+),$$

whereas for boundary jumps and averages we adopt the convention that $\{\!\{\boldsymbol{v}\}\!\} = [\![\boldsymbol{v}]\!] = \boldsymbol{v}$, and $\{\!\{w\}\!\} = [\![w]\!] = w$. In addition, we will use the symbol ∇_h to denote the broken gradient operator and $\boldsymbol{\varepsilon}_h$ to denote its symmetrised counterpart. Finally, given a positive integer k and a set $\mathcal{O} \subset \mathbb{R}^n$, $\mathcal{P}_k(\mathcal{O})$ stands for the space of polynomials of degree $\leq k$ defined on $\mathcal{O}$.

Introducción

La simulación de fenómenos de transporte e interacciones químicas dentro de medios porosos saturados es un marco que se encuentra en una amplia variedad de aplicaciones científicas y de ingeniería, incluyendo el flujo de contaminantes químicos en suelo saturado, perforación sub-superficial y extracción de petróleo, crecimiento de cristales, procesamiento de químicos y alimentos, inundación con polímeros como parte de procesos de recuperación mejorada de petróleo en ingeniería de petróleos [46], cromatografía [141], o saneamiento de agua y remoción de contaminantes tales como metales pesados o iones radioactivos [160], entre otras aplicaciones [24, 86, 88, 119, 136, 139, 150, 166].

Este trabajo de tesis tiene como objetivo analizar modelos de ecuaciones diferenciales parciales (EDP) para el acoplamiento de ecuaciones de flujo y transporte. Algunos de los retos que implica el trabajo en estos problemas físico-matemáticos, incluye las fuertes no-linealidades y el comportamiento dinámico caracterizado por escalas de tiempo diferentes. De entre las aplicaciones anteriormente mencionadas, las que motivaron el desarrollo de esta tesis son aquellas relacionadas con el diseño de equipamiento usado en el tratamiento de agua. Esto incluye sedimentadores, clarificadores/espesadores y equipo de filtrado. Comenzamos con el estudio del fenómeno de sedimentación, analizando en primera instancia modelos de sedimentación polidispersa. Consideramos desde el punto de vista del método numérico, un esquema de volúmenes finitos con propiedades de conservación de entropía. Los errores y la tasa de convergencia, para este esquema, resultan comparables a los otros métodos evaluados (Kurganov-Tadmor y Lax-Friedrichs global por componentes), y el esquema muestra ventajas en cuanto a rendimiento computacional para algunas aplicaciones, sin embargo, también se pudieron identificar algunas desventajas, tales como las restricciones sobre las propiedades del término de difusión y la necesidad de calibrar los parámetros de estabilización. Notemos además que el método es una simplificación en una dimensión del proceso de sedimentación. Todo esto nos motivó a explorar enfoques complementarios para la modelación de este fenómeno.

El segundo enfoque para el problema de sedimentación, que motiva el **Capítulo** 4, es un modelo tridimensional para clarificadores/espesadores, donde incorporamos la ecuación unidimensional para la densidad de Kynch, usada anteriormente, en una ecuación de transporte acoplada con las ecuaciones para flujo incompresible de Navier-Stokes-Brinkman, considerando además la presencia de una rastra giratoria. Previamente, en el **Capítulo** 3, discutimos un modelo el flujo doble-difusivo que será usado como base para el desarrollo del esquema. Creemos que los dos enfoques, aquí estudiados, pueden ayudar a adquirir un mejor entendimiento de la operación de unidades de clarificación.

La otra línea de investigación de este trabajo involucra el estudio de formulaciones de elementos finitos para las ecuaciones diferenciales parciales que describen el modelamiento de flujo acoplado con transporte. El primer modelo estudiado, está motivado por el fenómeno de flujo doble-difusivo estacionario. El flujo doble-difusivo se origina cuando combinamos transferencia de masa y calor interactuando con flujo dentro de una matriz porosa. El modelo adopta la forma de ecuaciones de flujo viscoso incompresible de Navier-Stokes-Brinkman en un medio poroso, acoplado con un par de ecua-

ciones de advección-difusión con términos de difusión cruzada, para describir la difusión de masa de un soluto y energía.

En el **Capítulo** 5, extendemos el modelo anterior y su análisis a un dominio axisimétrico con aplicación explícita al modelado de equipos de filtración basados en suelos. Las ecuaciones gobernantes son la ecuación de Navier-Stokes-Brinkman para el flujo a través del medio poroso, acopladas con las ecuaciones de convección-difusión para el transporte de contaminantes, más un sistema de ecuaciones diferenciales ordinarias para las propiedades de adsorción de cada contaminante. Estas ecuaciones se escriben en forma axisimétrica meridional y la formulación débil correspondiente adopta una estructura mixta-primal. Bajo este modelo, se asume que cada sitio tiene una capacidad máxima para cada contaminante, la cual se toma uniforme a lo largo de las dos capas de medio filtrante. De esta forma, el proceso de adsorción es no competitivo y la saturación de un sitio para un contaminante no previene la adsorción de otros contaminantes en el mismo sitio. También se asume que el proceso de adsorción es irreversible para todos los contaminantes y capas de filtrado, y una vez que un contaminante es adsorbido permanece adherido al medio filtrante.

Señalamos que un importante componente de esta tesis se dedica a estudiar la existencia de soluciones de las EDP asociadas, usando teorías de punto fijo. También trabajamos en la construcción de métodos numéricos precisos, robustos y confiables para la discretización de estas ecuaciones, con especial énfasis en formulaciones $H(\mathrm{div})$-conformes para las ecuaciones de flujo, mientras que para la formulación del problema de transporte (resultando en ecuaciones de advección-difusión escalares o vectoriales) estudiamos esquemas de entropía estable para sistemas sin acoplamiento con una ecuación de flujo y formulaciones primales de elementos finales para el caso de sistemas acoplados. La principal ventaja de una formulación $H(\mathrm{div})$-conforme radica en que esta produce aproximación de la velocidad, libres de divergencia, las cuales son de gran importancia, ya que permiten asegurar que las ecuaciones de flujo permanezcan localmente conservativas y estables en energía (ver v.g. [62]). Además, permite derivar estimaciones de error de la velocidad que son robustas a los errores en la presión (ver [100]).

La discretización $H(\mathrm{div})$ conforme se introduce en el **Capítulo** 3 como una modificación de métodos existentes para flujos de Brinkman. También se discute, el análisis de existencia y unicidad de soluciones para la formulación discreta de Galerkin, y se deriva de forma rigurosa sus propiedades de convergencia. En el **Capítulo** 4, extendemos este análisis al caso no estacionario, y en consecuencia, derivamos propiedades de convergencia espacio-temporal. En el **Capítulo** 5, modificamos el análisis considerando una formulación axisimétrica, que involucra el trabajo con espacios con peso. En todos los casos, la validez de los modelos y su rendimiento computacional se muestra numéricamente a través de varios ensayos computacionales.

Finalmente, en el **Capítulo** 6, discutimos las conclusiones principales de cada sección. Adicionalmente, presentamos el trabajo de investigación actualmente en curso, que añade a la base de los modelos que se presentan a lo largo de esta tesis, interacciones fluido-estructura, para abordar problemas con aplicaciones biológicas. El trabajo es una extensión del modelo matemático para la dinámica de la coagulación, propuesto en [39]. Proponemos la adición de interacciones fluido-estructura al modelo original, utilizando el método de frontera inmersa con multiplicador de Lagrange que se introdujo en [37]. El modelo base consiste de un sistema de ecuaciones de advección-difusión-reacción que describen la distribución espacio-temporal de los factores de coagulación y los subtipos de plaquetas durante el desarrollo tromboso, acoplado con las ecuaciones de Navier-Stokes para la descripción de

la dinámica de flujo sanguíneo en los vasos. Describimos la discretización espacio-temporal del modelo, incluyendo un esquema semi-implicito usado para la integración termporal y mostramos algunos resultados cualitativos preliminares. Cerramos el capítulo discutiendo las perspectivas con respecto al trabajo futuro, motivadas a partir de nuestro resultados, pero también de sugerencias/propuestas de colaboradores que conocimos a lo largo del desarrollo de esta tesis.

1.4 Trabajos relacionados

Con el fin de poner este trabajo en perspectiva, mencionamos que una gran cantidad de referencias con respecto a la existencia de soluciones bien puestas y el análisis numérico de ecuaciones de convección-difusión degeneradas se pueden encontrar en [99]. Sin embargo la existencia y unicidad de soluciones de entropía, y la convergencia de los métodos numéricos solo ha sido establecida hasta el momento para el caso escalar ($N = 1$); contribuciones importantes en esta dirección incluyen [57,73,103–106,110] (esta lista está lejos de ser completa). El estado del tema está en acuerdo con la bien conocida falta de resultados correspondientes a los sistemas más generales de primer orden de leyes de conservación (2.4), considerando que (2.1) se reduce a (2.4) cuando $\boldsymbol{K} = \mathbf{0}$. Dicho esto, mencionamos que los sistemas de convección-difusión degenerados (2.1) surgen en un gran número de aplicaciones tales como tráfico vehicular multiclase [29,30,38,53,54,158], sedimentación de suspensiones sólido-líquidas polidispersas [38,43,54,151], sedimentación de dispersiones de gotas y emulsiones [1,51,146], y cromatografía [52,70]. En particular, estas aplicaciones de las ecuaciones de advección-difusión (en lugar de las escalares) surgen dado que se desea describir la segregación de diferentes clases de unidades de una fase dispersa (autos, partículas, gotas, etc.), con la consecuencia de que un número de especies N en estas aplicaciones puede se arbitrariamente grande. Estas aplicaciones motivan el intéers en desarrollar métodos eficientes para la solución numérica de (2.1), (2.2) o (2.1)–(2.3) incluso si no existe teoría cerrada repecto a la existencia y unicidad de soluciones para estos sistemas.

Algunos esquemas numéricos comunes son basados en una discretización espacial que puede ser por medio del método de volúmenes finitos o Galerkin discontínuo [127], mientras que la discretización en espacio puede ser totalmente explícita o implícita-explícita (IMEX; ver por ejemplo [38,51,54]). Por el lado explícito, un esquema bien conocido es el de diferencias centradas de Kurganov-Tadmor (KT) [116]. El esquema KT original fue propuesto junto con combinaciones convexas de el método de paso temporal de Runge Kutta. Este concepto fue luego desarrollado, resultando en los esquemas llamados *Strong Stability Preserving Runge-Kutta* (SSPRK). Estos esquemas permite una discretización de alto orden en tiempo mientras preservan las propiedades fuertes de estabilidad del método de Euler de primer orden. Esto los hace atractivos para la resolución por el método de líneas de ecuaciones diferenciales parciales hiperbólicas [85].

Por otro lado, con respecto a la existencia de soluciones bien puestas para sistemas doble difusivos (bajo suposiciones razonables), primero nos restringimos a la discusión de ecuaciones clásicas de tipo Boussinesq. La existencia de soluciones asociadas a estas EDPs se remonta a Lorca and Boldrini [125,126]. Estos trabajos incluyen existencia, regularidad, y condiciones para la unicidad abordando tanto el casos estacionario como el no estacionario. Estos resultados se cumplen para viscosidad y conductividad térmica dependientes de la temperatura. Relacionados con el contexto de nuestro problema en específico, el análisis de soluciones de problemas doble-difusivos ha sido abordado por

ejemplo en [88, 124].

Existe una diversidad de métodos numéricos disponibles para las ecuaciones de Boussinesq y sus generalizaciones con coeficientes dependientes de la temperatura. Mencionamos por ejemplo los elementos finitos estabilizados (usando técnicas basadas en proyección) propuestas y rigurosamente analizadas en [6, 59], formulaciones mixtas analizadas en [5, 9, 10, 63], y la estabilidad de esquemas de particion (para Galerkin discontinuo, espectral y formulaciones de elementos finitos basadas en la vorticidad) y más ejemplos de aplicación han sido explorados en [2, 8, 46, 119, 120, 136, 149, 150, 163]. Esquemas Mixto-primales y totalmente-mixtos usando aproximaciones de la velocidad $\boldsymbol{H}(\mathrm{div})$-conformes se han estudiado en [137, 138].

Una de las ventajas de incluir una difusión matricial en el modelo es que esto permite el estudio de efectos tales como el de Soret y Dufour. Incluso cuando en algunas aplicaciones estos se pueden despreciar dado que sus contribuciones son algunos órdenes de magnitud inferiores que aquellos descritos por términos que surgen de las leyes de Fick y Fourier, estos efectos pueden ser significativos cuando se introducen especies en la superficie de un dominio fluido y tienen diferentes densidades en comparación al fluido circundante. Estos mecanismos son importantes también en aplicaciones relacionadas con el transporte de humedad en aislamiento fibroso o almacenamiento de granos, transporte de contaminantes químicos, deposito subterráneo de desperdicio nuclear y procesos de crecimiento cristalino [24].

Otras contribuciones en esta área incluyen discretizaciones de volúmenes finitos para la flotación térmica y por concentración de un soluto dentro de los flujos de Darcy-Brinkman introducidos en [86], el análisis de error para métodos espectrales aplicados a bioconvección en [66], o los sistemas basados en vorticidad de Brinkman y sistemas no lineales de advección-reacción-difusión analizados vía punto fijo y argumentos de compacidad en [14], esto incluye un esquema mixto-primal con velocidad libres de divergencia. En [93] se usaron métodos de Petrov-Galerkin Penalizados para la solución de problemas convectivos de doble difusión. En [155] los autores introducen esquemas de mínimos cuadrados específicamente diseñados para flujos convectivos de Rayleigh-Bénard, y en [165] se emplea el método de elementos finitos promediados para abordar problemas de solidificación que tienen una estructura similar a los modelos examinados aquí.

Por el lado de los esquemas numéricos que resultan en aproximaciones libres de divergencia del campo de velocidad, el trabajo de V. Jhon et al [100] provee un buen resumen de los diferentes enfoques que han sido propuestos y la importancia de la propiedad de divergencia libre para el cálculo de aproximaciones de la velocidad robustas a errores en la presión. Algunos de los métodos actualmente disponibles incluyen pares de elementos finitos conformes obtenidos usando técnicas de cálculo exterior [74] o enriqueciendo elementos localmente $H(\mathrm{div})$-conformes [91], métodos de Galerkin discontinuo (DG) con post-procesamiento [25] y de elementos finitos DG hibridizables (HDG) con y sin post-procesamiento [60, 143]. En [96], los autores presentan un método HDG en espacio-tiempo para el problema de Navier-Stokes en dominios dependientes del tiempo que resulta en campos de velocidad $H(\mathrm{div})$-conformes y libres de divergencia punto a punto. Se demuestra además que este esquema conserva momento, es estable en energía y robusto respecto a la presión. El estudio de esquemas numéricos eficientes para la aproximación de las soluciones de los problemas presentados en esta tesis, de forma que se mantengan las buenas propiedades del método actualmente usado, es un tema que queremos abordar en investigaciones futuras.

Modelos tempranos para el proceso de clarificación con y sin efectos de agitación se revisan en [152], donde se emplean configuraciones principalmente axisimétricas. Más recientemente, un modelo bastante completo puede encontrarse en [65], deonde los autores acoplan las ecuaciones de momento para el flujo de fluido con las ecuaciones de transporte para sólidos. Se utiliza el modelo realizable $k-\epsilon$ junto con funciones escalables de pared, para modelar la turbulencia. La remoción de lodo del piso del clarificador por medio de una rastra rotativa se modela a través de un término de sifón rotativo añadido al lado derecho de las ecuaciones de transporte. Referencias adicionales, relacionadas con el mecanismo de rastra y sus aplicaciones el procesamiento de minerales incluyen [61, 84, 89, 90, 121, 147, 156, 167].

En el modelo estudiado en el **Capítulo** 4, también incluimos términos de arrastre apropiados, similares a los utilizados en [161], que capturan los efectos indirectos de la rastra en los patrones de flotación. Esto consiste básicamente en penalizar el movimiento de la estructura y calcular (o como realizamos aquí, simplemente imponer) su velocidad y su configuración de una manera adecuada. Las técnicas de penalización de un volumen pueden encontrarse frecuentemente en la literatura relevante. Ver por ejemplo [108], donde los autores proponen métodos de alto orden para la modelación de obstáculos sólidos como estructuras porosas con una permeabilidad que tiende a cero, el flujo se maneja en un dominio unificado, y el momento del obstáculo se obtiene simplemente d la integración de la velocidad penalizada sobe el dominio del obstáculo. Otros enfoques numéricos que pueden ser utilizados para incorporar la interacción entre la rastra y el flujo incluyen los métodos de elementos finitos de frontera inmersa y dominio ficticio [35], métodos de conjunto de nivel y sus variantes [69, 135], otros esquemas de elementos finitos no ajustados [22]; o formulaciones basadas en remodelar, tales como el esquema arbitrario Lagrangiano-Euleriano (ALE) [164].

Un modelo de Boussinesq con viscosidad no lineal dependiente de la temperatura se propone en [4]. Los autores analizan esquemas numéricos basados en métodos de elementos finitos de primer y segundo orden y derivan estimaciones de error a priori óptimas para cada esquema. Un modelo relacionado de Boussinesq para cambio de fase se presenta en [163], donde un método de elementos finitos de segundo orden para la formulación primal del problema en términos de la velocidad, temperatura y presión se construye y se discuten las condiciones para su estabilidad.

El acoplamiento de sistemas de advección-difusión-reaccion con las ecuaciones de Brinkman en su formulación de velocidad-vorticidad-presión, se estudia en [120]. Las ecuaciones se discretizan en espacion usando métodos de elementos finitos en mallas no estructuradas, mientras que la integración en tiempo se basa en una estrategia de partición de operado que usa diferencias en las escalas entre los procesos de advección, difusión y reacción. Los autores comparan varias estrategias de acoplamiento en términos de usos de memoria, número de iteraciones, velocidad de cómputo y dinámica de la norma de energía.

Con respecto a nuestro modelo axisimétrico aplicado al estudio de equipo de filtración de agua, mencionamos que varios estudios han tratado la formulación axisimétrica de flujos de Stokes y Navier-Stokes, incluyendo la discretización empleando métodos espectrales, mortero, y elementos finitos estabilizados (ver por ejemplo [18, 26, 27, 31, 72], y las referencias citadas en estos trabajos). Más recientemente, se estudiaron formulaciones mixtas de flujos de Brinkman incluyen el análisis numérico de aproximaciones de elementos finitos (FE). Anaya et al. [15] presentó una aproximación aumentada de elementos finitos para las ecuaciones de Brinkman desarrollada como una extensión de la formulación basada en vorticidad para el problema de Stokes. Un modelo relacionado fue recientemente presentado

en [16] e incorpora una función de corriente y una formulación de vorticidad para el flujo axisimétrico de Brinkman, además se emplea una aproximación mixta conforme de elementos finitos, para la solución numérica del problema.

El análisis numérico de problemas de flujo axisimétrico de Darcy y Stokes-Darcy usando elementos finitos de Raviart-Thomas (RT) y Brezzi-Douglas-Marini (BDM) fue presentado en [71,72]. En [71], el autor establece la estabilidad de las aproximaciones RT y BDM para el problema de flujo axisimétrico de Darcy, extendiendo el criterio de Stenberg, y además se derivan estimaciones de error a priori.

Otras contribuciones al diseño de métodos para formulaciones axisimétricas de problemas de flujo y transporte acoplados incluyen [12, 55]. Además, en [47] se propone un esquema semi-discreto de volúmenes-elementos finitos discontinuos (FVE) y se discute la existencia y unicidad de soluciones para los casos no lineal continuo y su contraparte semi-discreta. Un método FVE también se propone en [55] para discretizar la ecuación de Stokes para el flujo acoplado con una ecuación parabólica modelando sedimentación. El método se basa en una formulación de Galerkin discontinuo estabilizado para el campo de concentración, y un par estabilizado multi-escala de elementos $\mathcal{P}_1$-$\mathcal{P}_1$ para la velocidad y presión, respectivamente. Una formulación variacional mixta del flujo de Darcy-Forchheimer acoplado con una ecuación de energía es semi-discretizado en [12] usando elementos finitos de Raviart-Thomas y elementos constantes por pedazos para la presión, también se establecen estimaciones de error a posteriori.

La aplicación tecnológica detrás del modelos de los filtros de agua, se remonta a la observación de que es posible remover arsénico de agua, haciéndola circular a través de suelo de laterita enriquecido con hierro [131, 132]. El arsénico se remueve a través de un proceso de adsorción, el cuál se puede mejorar por medio del tratamiento químico de la laterita para incrementar su porosidad y área superficial, aumentando la eficiencia de adsorción [130]. Claramente, la formulación de modelos matemáticos precisos para estos filtros, además de su eficiencia computacional, ayudaría de forma importante al desarrollo y mejora de estos filtros y a desarrollar guías para su operación segura. El desarrollo y análisis de estos modelos forma la base del trabajo [134], donde los autores examinan la remoción de un contaminante (arsénico, caso $m = 1$ en nuestra notación) en un filtro cilíndrico con medio uniforme. Los autores utilizan un ecuación de Darcy-Brinkman, acoplado con la ecuación de advección-difusión-adsorción para el modelamiento del flujo de agua contaminada a través del filtro y la remoción de arsénico por adsorción. En la práctica, sin embargo, existen $m > 1$ contaminantes, los cual induce a usar filtros consistentes de múltiples (hasta m) capas con el fin de permitir su remoción. En este trabajo nosotros intentamos estudiar el proceso de filtración en filtros de agua basados en suelos consistentes de dos capas de medios distintos, en la presencia de múltiples especies contaminantes.

Problemas con una naturaleza similar abundan en la literatura. Por ejemplo, [86] considera la solución numérica, vía un método de volúmenes finitos, de un problema doble-difusivo dentro de un medio poroso. La publicación [150] considera un problema doble-difusivo similar, sin embargo, a semejanza del filtro en capas que proponemos, los autores permiten la posibilidad de medio poroso estratificado heterogéneo. Mientras muchos estudios concernientes a problemas doble-difusivos consideran dominios cerrados llenos con medio poroso, un gran número de aplicaciones, tales como nuestro filtro, presentan confinamientos parciales con aberturas o infiltraciones. El artículo [154] introduce esta propiedad, con la adición de 'puertos libres' al dominio de su modelo. Considerando otras variaciones potenciales, los autores de [166] extienden el problema doble-difusivo usual con un proceso de reacción de primer

orden entre la especie que se difunde y el fluido. Este proceso de reacción necesita adicionalmente la adición de un término de sifón a la ecuación que gobierna la concentración de especies y que juega un rol similar al lado derecho de (5.1c).

1.5 Contribuciones de esta tesis

En el **Capítulo** 2, extendemos el análisis de una clase de esquemas de entropía estable para la solución numérica de problemas no lineales de valor inicial, recientemente propuestos en [S. Jerez, C. Parés, *Entropy stable schemes for degenerate convection-diffusion equations*, SIAM J. Numer. Anal. vol. 55 (2017) pp. 240–264]. Como nueva contribución, demostramos, primero, que estos esquemas pueden extenderse de forma natular a problemas de valores iniciales y de frontera con condiciones de borde de flujo cero en una dimensión espacial, incluyendo una cota explícita para el crecimiento de la entropía total. Segundo, mostramos que las suposiciones del modelo se satisface para ciertos modelos de flujo cinemático multiclase corregidos por difusividad de tamaño arbitrario, que describen el flujo vehicular o la sedimentación de dispersiones y emulsiones. Los contenidos de este capítulo dieron origen a la siguiente publicación:

[48] Bürger, R., Méndez, P. E., Parés, C., *On entropy stable schemes for degenerate parabolic multispecies kinematic flow models.* **Numer Methods Partial Differential Eq**. 1–26; (2019)

En el **Capítulo** 3 proponemos un modelo para flujos doble-difusivos, que incluye la posibilidad de difusión cruzada. Las principales diferencias entre el análisis de métodos para las ecuaciones clásicas de Boussinesq y las ecuaciones doble-difusivas están, por supuesto, en la naturaleza vectorial de las cantidades a difundirse mientras que en la formulación clásica de Boussinesq solo se considera una cantidad escalar (por ejemplo temperatura). Algunos de los argumentos relacionados con el análisis de existencia y unicidad de soluciones, en particular aquellos relacionados con el manejo de condiciones de frontera tipo Dirichlet no homogéneas por argumentos de lifting [125, 137], se trasladan casi sin cambios del caso escalar al vectorial. Sin embargo, la forma bilineal asociada al término de difusión debe ser coerciva de forma que se garantice la estabilidad. Este requerimiento, a su vez, impone restricciones sobre la elección de la matriz de difusión $\mathbb{D}$; esta matriz debe ser positiva definida (aunque no necesariamente simétrica). Estas propiedades son esenciales para la prueba de existencia de una solución discreta. A pesar de esto, aún es posible estudiar efectos de difusión cruzada, tales como los de Soret y Dufour. Adicionalmente, formulamos una discretización $\boldsymbol{H}(\mathrm{div})$-conforme modificando métodos conocidos para flujos de Brinkman. Discutimos, la existencia de soluciones para la formulación de Galerkin y derivamos de forma rigurosa propiedades de convergencia en espacio. Los contenidos de este capítulo dieron lugar a la siguiente publicación:

[49] Bürger, R., Méndez, P.E., Ruiz-Baier, R., *On $\boldsymbol{H}(\mathrm{div})$-conforming methods for double-diffusion equations in porous media.* **SIAM Journal on Numerical Analysis**, 57, 1318–1343 (2019)

En el **Capítulo** 4 introducimos un nuevo modelo para la simulación de sedimentación-consolidación de partículas sólidas en un fluido incompresible bajo los efectos de la gravedad y en presencia de un

brazo rotatorio que gira lentamente asistiendo con la remoción de sedimento del fondo de unidades de clarificación/espesamiento. El modelo gobernante es ahora un problema de valores iniciales y de frontera para las ecuaciones de Navier-Stokes describiendo el flujo de una mezcla acoplados con ecuaciones no lineales parabólicas describiendo la fracción en volumen de sólidos. La novedad de nuestro tratamiento consiste en la inclusión de términos que modelan la influencia del movimiento de la rastra en el balance de momento y en la remoción de sólidos. Además, adaptamos técnicas del método de elementos finitos de frontera inmersa (ver v.g. [34]) para el análisis y aproximación numérica de esos términos. Derivamos rigurosamente las propiedades de convergencia espacio-temporal del método y demostramos estas propiedades a través de varios ensayos computacionales. Los contenidos de este capítulo dieron origen a la siguiente pre-publicación:

[50] Bürger, R., Méndez, P.E., Ruiz-Baier, R., *A second-order H(div)-conforming scheme for the simulation of sedimentation and flow in circular clarifiers with a rotating rake.* **Centro de Investigación en Ingeniería Matemática** (CI^{2}MA), Preprint 2019-39, Universidad de Concepción, Chile 2019.

En el **Capítulo** 5 extendemos el modelo previo a un dominio axisimétrico con aplicación explícita al modelado de equipos de filtración basados en suelos. Además, derivamos estimaciones de errores a priori óptimas para el esquema numérico de segundo orden $\boldsymbol{H}(\mathrm{div})$-conforme en tiempo y espacio. La principal dificultad en este caso, está dada por el análisis discreto donde es necesario verificar que cada término este acotado de forma óptima en los espacios con peso correspondientes. Algunos ejemplos numéricos ilustran las propiedades principales del problema y del esquema numérico. Los contenidos de este capítulo dieron origen a la siguiente pre-publicación:

[23] Baird, G., Bürger, R., Méndez, P.E., Ruiz-Baier, R., *Second-order schemes for axisymmetric Navier-Stokes-Brinkman and transport equations modelling water filters.* **Centro de Investigación en Ingeniería Matemática** (CI^{2}MA), Preprint 2019-23, Universidad de Concepción, Chile 2019.

1.6 Notación

Desde el **Capítulo** 3 y en adelante consideraremos la siguiente notación: Sea Ω un dominio acotado y abierto en $\mathbb{R}^d$, $d = 2, 3$ con frontera Lipschitz $\Gamma = \partial\Omega$. Denotamos por $L^p(\Omega)$ y $W^{r,p}(\Omega)$ los espacios usuales de Lebesgue y Sobolev con sus normas respectivas $\|\cdot\|_{L^p(\Omega)}$ y $\|\cdot\|_{W^{r,p}(\Omega)}$. Si $p = 2$ escribimos $H^r(\Omega)$ en lugar de $W^{r,p}(\Omega)$, y denotamos la norma correspondiente por $\|\cdot\|_{r,\Omega}$, ($\|\cdot\|_\Omega$ para $H^0(\Omega) = L^2(\Omega)$). El espacio $L^2_0(\Omega)$ denota la restricción de $L_2(\Omega)$ a funciones con valor medio cero sobre Ω. Para $r \geq 0$, escribimos la seminorma H^r como $|\cdot|_{r,\Omega}$ y denotamos por $(\cdot,\cdot)_\Omega$ el producto interior usual en $L^2(\Omega)$. Adicionalmente, $H^{1/2}(\Gamma)$ es el espacio de trazas de funciones $H^1(\Omega)$ y $H^{-1/2}(\Gamma)$ es su dual. Los espacios de funciones con valores vectoriales (en dimensión d) se denotan en negrita, y denotamos por $\mathbb{L}$ sus correspondientes contrapartes tensoriales. A continuación, denotamos por $L^s(0, \mathcal{T}; W^{m,p}(\Omega))$

el espacio de Banach de todas las funciones L^s-integrables de $[0,\mathcal{T}]$ en $W^{m,p}(\Omega)$, con norma

$$\|v\|_{L^s(0,\mathcal{T};W^{m,p}(\Omega))} = \begin{cases} \left(\int_0^{\mathcal{T}} \|v(t)\|^s_{W^{m,p}(\Omega)}\,\mathrm{d}t\right)^{1/s} & \text{if } 1 \le s < \infty, \\ \operatorname{esssup}_{t\in[0,\mathcal{T}]}\|v(t)\|_{W^{m,p}(\Omega)} & \text{if } s = \infty. \end{cases}$$

Denotamos por $\mathcal{T}_h$ una partición regular de Ω compuesta por elementos triangulares K en dos dimensiones, tetrahedros en tres dimensiones de diámetro h_K. El tamaño de malla será denotado por $h = \text{máx}\{h_K,\ K \in \mathcal{T}_h\}$, y para cualquier cara interior e en $\mathcal{E}_h$ (el conjunto de caras en $\mathcal{T}_h$), etiquetaremos como K^- y K^+ sus elementos adyacentes, mientras h_e corresponderá al diámetro máximo de la cara.

Asumimos que $\boldsymbol{v}, w$ con campos escalares y vectoriales suaves definidos sobre $\mathcal{T}_h$. Entonces, por $(\boldsymbol{v}^\pm, w^\pm)$ denotaremos las trazas de $(\boldsymbol{v}, w)$ en e correspondiendo a las extenciones desde los interiores de los elementos K^+ y K^-, respectivamente. Sea $\boldsymbol{n}_e$ el vector normal unitario exterior a e en K, definimos la componente tangencial de $\boldsymbol{u}$ en cada cara e como $\boldsymbol{u}_{\boldsymbol{\tau}} = \boldsymbol{u} - (\boldsymbol{u}\cdot\boldsymbol{n}_e)\boldsymbol{n}_e$. Introducimos los operadores promedio $\{\!\{\cdot\}\!\}$ y salto $[\![\cdot]\!]$ como sigue:

$$\begin{aligned} \{\!\{\boldsymbol{v}\}\!\} &= (\boldsymbol{v}^- + \boldsymbol{v}^+)/2, \quad \{\!\{w\}\!\} = (w^- + w^+)/2, \\ [\![\boldsymbol{v}]\!] &= (\boldsymbol{v}^- - \boldsymbol{v}^+), \quad [\![w]\!] = (w^- - w^+), \end{aligned}$$

mientras que para los saltos y promedios en las fronteras adoptamos la definición $\{\!\{\boldsymbol{v}\}\!\} = [\![\boldsymbol{v}]\!] = \boldsymbol{v}$, y $\{\!\{w\}\!\} = [\![w]\!] = w$. Además, usaremos el símbolo ∇_h para denotar el operador de gradiente a trozos y $\boldsymbol{\varepsilon}_h$ para denotar su contraparte simétrica. Finalmente, dado un entero positivo k y un conjunto $\mathcal{O} \subset \mathbb{R}^n$, $\mathcal{P}_k(\mathcal{O})$ corresponde al espacio de polinomios de grado $\le k$ definidos en $\mathcal{O}$.

CHAPTER 2

On entropy stable schemes for degenerate parabolic multispecies kinematic flow models

In this chapter we analyse entropy stable schemes for the numerical solution of initial value problems of nonlinear, possibly strongly degenerate systems of convection-diffusion equations, recently proposed in [S. Jerez, C. Parés, *Entropy stable schemes for degenerate convection-diffusion equations*, SIAM J. Numer. Anal. vol. 55 (2017) pp. 240–264]. These schemes extend the theoretical framework by [E. Tadmor, *The numerical viscosity of entropy stable schemes for systems of conservation laws. I*, Math. Comp. vol. 49 (1987) pp. 91–103] to convection-diffusion systems. As a new contribution, we demonstrate, firstly, that these schemes can naturally be extended to initial-boundary value problems with zero-flux boundary conditions in one space dimension, including an explicit bound on the growth of the total entropy. Secondly, it is shown that these assumptions are satisfied by certain diffusively corrected multiclass kinematic flow models of arbitrary size that describe traffic flow or the settling of dispersions and emulsions, where the latter application gives rise to zero-flux boundary conditions. Numerical examples illustrate the behavior and accuracy of entropy stable schemes for these applications.

2.1 Introduction

2.1.1 Scope

This chapter concerns numerical schemes for systems of degenerate convection-diffusion equations in one space dimension of the form

$$\boldsymbol{u}_t + \boldsymbol{f}(\boldsymbol{u})_x = \big(\boldsymbol{K}(\boldsymbol{u})\boldsymbol{u}_x\big)_x, \quad x \in I \subset \mathbb{R}, \quad t \in \mathbb{R}_+, \tag{2.1}$$

where $I = \mathbb{R}$ or I is a bounded interval, $\boldsymbol{u} = (u_1, \ldots, u_N)^{\mathrm{T}} \colon I \times \mathbb{R}_+ \to \Omega \subset \mathbb{R}^N$ is the vector of unknown functions of position x and time t, $\boldsymbol{f} = (f_1, \ldots, f_N)^{\mathrm{T}}$ is a given flux vector, and $\boldsymbol{K}(\boldsymbol{u}) \in \mathbb{R}^{N \times N}$ is a positive semidefinite diffusion matrix defined in Ω. We allow that $\boldsymbol{K}(\boldsymbol{u}) = \boldsymbol{0}$ on a set of $\boldsymbol{u}$-values of positive N-dimensional measure, so (2.1) is, in general, strongly degenerate. Equation (2.1) is equipped with the initial condition

$$\boldsymbol{u}(x, 0) = \boldsymbol{u}_0(x), \quad x \in I; \tag{2.2}$$

if I is bounded, that is $I = [0, L]$ with $L > 0$, then we impose, in addition, the zero-flux boundary condition

$$\big(\boldsymbol{f}(\boldsymbol{u}) - \boldsymbol{K}(\boldsymbol{u})\boldsymbol{u}_x\big)\big|_{x=0} = \boldsymbol{0}, \quad \big(\boldsymbol{f}(\boldsymbol{u}) - \boldsymbol{K}(\boldsymbol{u})\boldsymbol{u}_x\big)\big|_{x=L} = \boldsymbol{0}. \tag{2.3}$$

For the problem (2.1), (2.2), whose solutions are in general discontinuous, Jerez and Parés [99] devised so-called entropy stable finite difference schemes. These schemes extend the concept of entropy stable methods for systems of conservation laws due to Tadmor [153]. Such schemes are obtained by adding an adequate amount of artificial dissipation to an entropy conservative method so that the entropy satisfies a system of differential equalities arising from a spatially discrete but continuous in time entropy inequality. Entropy conservative methods capture correctly the appearance and propagation of shocks but they may may produce oscillations around shocks, which are handled by the artificial dissipation. It was shown in [99] that a necessary condition for such a method to be feasible for (2.1) is that the first-order system of conservation laws

$$\boldsymbol{u}_t + \boldsymbol{f}(\boldsymbol{u})_x = \boldsymbol{0} \tag{2.4}$$

has a convex entropy function $\eta = \eta(\boldsymbol{u})$ and entropy flux $g = g(\boldsymbol{u})$, for which the entropy inequality

$$\eta(\boldsymbol{u})_t + g(\boldsymbol{u})_x \leq 0 \tag{2.5}$$

is valid (in the sense of distributions) for solutions of (2.4) [118]. It is well known that for $N \geq 3$, the existence of an entropy pair (η, g) for the first-order system (2.4) is an exceptional property since the gradient of g, denoted by $g_{\boldsymbol{u}}$ and which we assume to be a column vector, the gradient of the entropy function, $\eta_{\boldsymbol{u}}$, and the Jacobian of $\boldsymbol{f}$, denoted by $\boldsymbol{f}_{\boldsymbol{u}}$, must satisfy the compatibility condition

$$g_{\boldsymbol{u}}^{\mathrm{T}} = \eta_{\boldsymbol{u}}^{\mathrm{T}} \boldsymbol{f}_{\boldsymbol{u}}. \tag{2.6}$$

Such an entropy pair exists, however, in the exceptional case $\boldsymbol{f}_{\boldsymbol{u}}$ is symmetrizable. In fact, the existence of an entropy pair and the computation of an entropy-conservative flux is a general limitation for the application of entropy-stable methods in the context of systems of conservation laws. Nevertheless, there are many real-world models for which entropy pairs and entropy conservative numerical fluxes are available, including Euler and related systems, shallow water and related systems, and some multiphase fluid models (see, e.g., [58, 75–77, 94]). In fact, an application to the shallow water model was also considered in [99]. To highlight the principal advantage of entropy stable schemes for (2.1), we may follow the reasoning of [77] (which is expressed in similar form in many other works) advanced for the first-order system (2.4). Namely, convergence results for numerical schemes (even first-order schemes) approximating solutions of (2.4) are difficult to obtain since a global well-posedness theory for (2.4) is currently not available. Thus it is reasonable to require that numerical schemes be entropy stable, i.e., satisfy a discrete version of the entropy inequality (2.5). In particular, such a scheme satisfies a discrete form of a bound of the total entropy (as will be specified in Section 2.2.1 below in the context of the more general equation (2.1)), and will be stable in a suitable L^p space. As Fjordholm et al. [77] further point out, no entropy stability results for high-order numerical schemes for approximating (2.4), based on the total variation diminishing (TVD), essentially non-oscillatory (ENO), weighted essentially oscillatory (WENO), and discontinuous Galerkin (DG) procedures are available (however, entropy stable streamline diffusion finite element methods were proposed in [97]).

This underlines the interest in the construction of schemes that have entropy stability "built in", which will be achieved in this work for the initial(-boundary) value problems (2.1), (2.2) and (2.1)–(2.3).

As was derived in [99], the specific limitation in the case of problems with a diffusion term is the additional requirement of positive definiteness of the matrix $\boldsymbol{K}(\boldsymbol{u})\eta_{\boldsymbol{u},\boldsymbol{u}}^{-1}$, where $\eta_{\boldsymbol{u},\boldsymbol{u}}^{-1}$ is the inverse of the Hessian $\eta_{\boldsymbol{u},\boldsymbol{u}}$ of $\eta(\boldsymbol{u})$. Thus, the class of convection-diffusion problems to which the scheme developed in [99] can actually be applied seems fairly narrow, but it does include a class of diffusively corrected applicative kinematic flow models [38, 53, 54], for instance of vehicular traffic or of polydisperse sedimentation. These models can be expressed by (2.1) on a bounded interval I with an arbitrarily large number N of species. It is therefore the purpose of this chapter to demonstrate that the entropy stable schemes of [99] can successfully be applied to these models, under modifications due to the presence of boundary conditions but maintaining the principal property of entropy stability.

2.1.2 Outline of the chapter

The remainder of this chapter is organized as follows. In Section 2.2 we summarize from [99] the construction of entropy stable schemes for (2.1) and extend the discussion to the zero-flux initial-boundary value problem (IBVP) (2.1)–(2.3). Specifically, we discuss in Section 2.2.1 properties of the continuous problem, and motivate a global entropy inequality for solutions of (2.1)–(2.3). With the goal to design numerical methods for (2.1), we treat in Section 2.2.2 the spatial discretisation of that equation in the interior of the domain and derive an entropy-conservative numerical flux. The resulting semi-discrete scheme is equipped with a small amount of extra viscosity to prevent oscillations, as is detailed in Section 2.2.3. Then, in Section 2.2.4, we outline the numerical scheme that arises from the previous discussion if we wish to solve the zero-flux IBVP (2.1)–(2.3). Results include a time-continuous, spatially discrete entropy inequality. The treatment of Sections 2.2.2 to 2.2.4 presupposes that an entropy conservative numerical flux is given, for which we provide in Section 2.2.5 a sample definition that follows Tadmor [153], and which is utilized in the numerical examples. In Section 2.3 we outline two applicative models to which the entropy stable schemes are applied, namely in Section 2.3.1 a diffusively corrected multi-class version of the well-known Lighthill-Whitham-Richards model (DCMCLWR model) that gives rise to the initial value problem (2.1), (2.2), and in Section 2.3.2 a model of settling of dispersions of droplets and colloidal particles that motivates the IBVP (2.1)–(2.3). Both problems are introduced along with the corresponding entropy conservative numerical flux. Numerical examples for both applicative models are introduced in Section 2.4, starting with a description of the time discretisation and the computation of approximate numerical errors for all cases (in Section 2.4.1). Examples 2.1 to 2.4 (Sections 2.4.2 to 2.4.5) deal with the DCMCLWR traffic model, and Examples 2.5 and 2.6 (Sections 2.4.6 and 2.4.7) are related to the settling model.

2.2 Entropy stable schemes

2.2.1 Preliminaries

If there exists a vector-valued function $\mathcal{K}\colon \Omega \to \mathbb{R}^N$ such that $\mathcal{K}_{\boldsymbol{u}} = \boldsymbol{K}$, where $\mathcal{K}_{\boldsymbol{u}}$ denotes the Jacobian of the function $\mathcal{K}$, then the system (2.1) can be written in the form

$$\boldsymbol{u}_t + \boldsymbol{f}(\boldsymbol{u})_x = \mathcal{K}(\boldsymbol{u})_{xx}.$$

This is always the case for scalar equations with $\boldsymbol{K}(\boldsymbol{u}) = K(u)$ if we define

$$\mathcal{K}(u) \coloneqq \int_0^u K(\xi)\,\mathrm{d}\xi.$$

Let us suppose that the system of conservation laws obtained by dropping the viscous term, i.e., (2.4), is equipped with an entropy pair (η, g) consisting of an entropy function η and an entropy flux g such that $\eta, g\colon \Omega \to \mathbb{R}$, η is strictly convex, and (2.6) holds. We then define the so-called entropy variables $\boldsymbol{v}$ as in [116], namely $\boldsymbol{v}(\boldsymbol{u}) \coloneqq \eta_{\boldsymbol{u}}(\boldsymbol{u})$. Then, in order to study the evolution of the entropy for a solution of (2.1), let us first express the diffusion term in terms of the entropy variables. Clearly,

$$\big(\boldsymbol{K}(\boldsymbol{u})\boldsymbol{u}_x\big)_x = \big(\hat{\boldsymbol{K}}(\boldsymbol{v})\boldsymbol{v}_x\big)_x, \tag{2.7}$$

where we define

$$\hat{\boldsymbol{K}}(\boldsymbol{v}) \coloneqq \boldsymbol{K}\eta_{\boldsymbol{u},\boldsymbol{u}}^{-1}, \tag{2.8}$$

where $\eta_{\boldsymbol{u},\boldsymbol{u}}$ is the Hessian matrix of η. The matrix on the left-hand side of (2.7) is evaluated at $\boldsymbol{u} = \eta_{\boldsymbol{u}}^{-1}(\boldsymbol{v})$. Once the diffusion term is rewritten, we multiply (2.1) by the vector of entropy variables $\boldsymbol{v}$ to obtain

$$0 = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}_t + \boldsymbol{v}^{\mathrm{T}}\boldsymbol{f}_{\boldsymbol{u}}(\boldsymbol{u})\boldsymbol{u}_x - \boldsymbol{v}^{\mathrm{T}}\big(\hat{\boldsymbol{K}}(\boldsymbol{v})\boldsymbol{v}_x\big)_x = \eta(\boldsymbol{u})_t + g(\boldsymbol{u})_x - \big(\boldsymbol{v}^{\mathrm{T}}\hat{\boldsymbol{K}}(\boldsymbol{v})\boldsymbol{v}_x\big)_x + \boldsymbol{v}_x^{\mathrm{T}}\hat{\boldsymbol{K}}(\boldsymbol{v})\boldsymbol{v}_x.$$

Therefore, if the matrix $\hat{\boldsymbol{K}}$ is positive semidefinite, i.e.

$$\boldsymbol{w}^{\mathrm{T}}\hat{\boldsymbol{K}}(\boldsymbol{v})\boldsymbol{w} \geq 0 \quad \text{for all } \boldsymbol{w} \in \mathbb{R}^N, \tag{2.9}$$

the following entropy inequality is satisfied:

$$\eta(\boldsymbol{u})_t + g(\boldsymbol{u})_x - \big(\boldsymbol{v}^{\mathrm{T}}\hat{\boldsymbol{K}}(\boldsymbol{v})\boldsymbol{v}_x\big)_x \leq 0. \tag{2.10}$$

In the case that $I = \mathbb{R}$ and we consider the initial value problem (2.1), (2.2) under the additional assumption that $\boldsymbol{u} \to 0$ when $x \to \pm\infty$, then the total entropy decreases, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}} \eta(\boldsymbol{u})\,\mathrm{d}x \leq 0.$$

(This also includes the case of a finite interval I with solution $\boldsymbol{u}$ that is compactly supported in I at all times.) On the other hand, considering the IBVP (2.1)–(2.3) and assuming that $\boldsymbol{u}$ and $\boldsymbol{v}$ have well-defined traces at the boundaries $x = 0$ and $x = L$, which we denote by $\boldsymbol{u}(0,t)$ and $\boldsymbol{u}(L,t)$, as well

as that the boundary condition (2.3) is well defined in the sense of traces, we can argue as follows. Integrating (2.10) over I, utilizing that $\hat{\boldsymbol{K}}(\boldsymbol{v})\boldsymbol{v}_x = \boldsymbol{K}(\boldsymbol{u})\boldsymbol{u}_x$ and the boundary condition (2.3), we get

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}} \eta(\boldsymbol{u})\,\mathrm{d}x + g\big(\boldsymbol{u}(L,t)\big) - g\big(\boldsymbol{u}(0,t)\big) &\leq \big(\boldsymbol{v}^{\mathrm{T}}\hat{\boldsymbol{K}}(\boldsymbol{v})\boldsymbol{v}_x\big)\big|_{x=L} - \big(\boldsymbol{v}^{\mathrm{T}}\hat{\boldsymbol{K}}(\boldsymbol{v})\boldsymbol{v}_x\big)\big|_{x=0} \\ &= \big(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{K}(\boldsymbol{u})\boldsymbol{u}_x\big)\big|_{x=L} - \big(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{K}(\boldsymbol{u})\boldsymbol{u}_x\big)\big|_{x=0} \\ &= \boldsymbol{v}(L,t)^{\mathrm{T}}\boldsymbol{f}\big(\boldsymbol{u}(L,t)\big) - \boldsymbol{v}(0,t)^{\mathrm{T}}\boldsymbol{f}\big(\boldsymbol{u}(0,t)\big).\end{aligned}$$

In terms of the so-called entropy potential function $\varphi \coloneqq \boldsymbol{v}^{\mathrm{T}}\boldsymbol{f} - g$, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}} \eta(\boldsymbol{u})\,\mathrm{d}x \leq \varphi\big(\boldsymbol{u}(L,t)\big) - \varphi\big(\boldsymbol{u}(0,t)\big). \tag{2.11}$$

Note that the function φ, and therefore the right-hand side of (2.11), do not depend on the particular choice of the diffusion matrix $\boldsymbol{K}(\boldsymbol{u})$.

Remark 2.1. *We emphasize that the requirement that the matrix $\hat{\boldsymbol{K}}$ defined by (2.8) should be positive semidefinite is the most severe restriction of the applicability of the approach. In fact, for a* general *positive semidefinite matrix $\boldsymbol{K} = \boldsymbol{K}(\boldsymbol{u})$, the product $\boldsymbol{K}\eta_{\boldsymbol{u},\boldsymbol{u}}^{-1}$ is, in general,* not *positive semidefinite unless $\boldsymbol{K}$ and $\eta_{\boldsymbol{u},\boldsymbol{u}}^{-1}$ or equivalently, $\boldsymbol{K}$ and $\eta_{\boldsymbol{u},\boldsymbol{u}}$ possess the same set of eigenvectors. That latter property is, however, valid if the diffusion term can be expressed as $\boldsymbol{K}(\boldsymbol{u}) = k(\boldsymbol{u})\mathbb{I}$, where $k(\boldsymbol{u}) \geq 0$ is a scalar function and $\mathbb{I}$ is the $N \times N$ identity matrix. Then*

$$\hat{\boldsymbol{K}}(\boldsymbol{v}) = k(\boldsymbol{u})\eta_{\boldsymbol{u},\boldsymbol{u}}^{-1} \tag{2.12}$$

is indeed positive semidefinite, since $\eta_{\boldsymbol{u},\boldsymbol{u}}$ is positive definite. Therefore, in this case, (2.10) holds.

2.2.2 Entropy conservative numerical method

We first consider the case of the initial-value problem (2.1), (2.2) on a standard spatial mesh defined by cells $\mathcal{I}_j \coloneqq [x_{j-1}, x_j)$, where $x_j = j\Delta x$, $\Delta x = L/M$ for some integer M, and $\boldsymbol{u}_j(t)$ denotes the cell average of $\boldsymbol{u}(\cdot,t)$ on $\mathcal{I}_j$. We will first discretize (2.1) in the interior of the computational domain, and handle the boundary conditions in Section 2.2.4. To this end, we first consider an entropy-conservative (EC) numerical flux $\boldsymbol{F}_{j+1/2}$, i.e. a numerical flux satisfying

$$[\![\boldsymbol{v}]\!]_{j+1/2}^{\mathrm{T}}\boldsymbol{F}_{j+1/2} = [\![\varphi]\!]_{j+1/2}, \tag{2.13}$$

where we employ the following notation to denote the average and jump of any variable ω:

$$[\![\omega]\!]_{j+1/2} \coloneqq \omega_{j+1} - \omega_j, \quad \{\!\{\omega\}\!\}_{j+1/2} \coloneqq (\omega_j + \omega_{j+1})/2.$$

Tadmor [153] showed that if the numerical flux $\boldsymbol{F}_{j+1/2}$ satisfies (2.13), then the solution of the semidiscrete method for (2.4),

$$\boldsymbol{u}_j'(t) = -\frac{1}{\Delta x}(\boldsymbol{F}_{j+1/2} - \boldsymbol{F}_{j-1/2}),$$

where $\cdot' \equiv \mathrm{d}\cdot/\mathrm{d}t$, satisfies the equality

$$\eta(\boldsymbol{u})_j'(t) = -\frac{1}{\Delta x}(G_{j+1/2} - G_{j-1/2})$$

for some numerical entropy flux $G_{j+1/2}$ consistent with g. Once an EC numerical flux (for (2.4)) has been chosen, we propose the following semidiscrete method for (2.1):

$$\boldsymbol{u}_j'(t) = -\frac{1}{\Delta x}\big(\boldsymbol{F}_{j+1/2} - \boldsymbol{F}_{j-1/2}\big) + \frac{1}{\Delta x^2}\big(\hat{\boldsymbol{K}}_{j+1/2}[\![\boldsymbol{v}]\!]_{j+1/2} - \hat{\boldsymbol{K}}_{j-1/2}[\![\boldsymbol{v}]\!]_{j-1/2}\big), \tag{2.14}$$

where

$$\hat{\boldsymbol{K}}_{j+1/2} = \hat{\boldsymbol{K}}(\{\!\{\boldsymbol{v}\}\!\}_{j+1/2}). \tag{2.15}$$

Let us show that a semidiscrete counterpart of (2.10) is satisfied. Multiplying (2.14) from the left by $\boldsymbol{v}_j^{\mathrm{T}}$ yields

$$\eta(\boldsymbol{u})_j'(t) = -\frac{1}{\Delta x}\boldsymbol{v}_j^{\mathrm{T}}(\boldsymbol{F}_{j+1/2} - \boldsymbol{F}_{j-1/2}) + \frac{1}{\Delta x^2}\boldsymbol{v}_j^{\mathrm{T}}\big(\hat{\boldsymbol{K}}_{j+1/2}[\![\boldsymbol{v}]\!]_{j+1/2} - \hat{\boldsymbol{K}}_{j-1/2}[\![\boldsymbol{v}]\!]_{j-1/2}\big).$$

The following identities are obtained by straightforward algebraic manipulations:

$$\begin{aligned}
\boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{F}_{j+1/2} &= \{\!\{\boldsymbol{v}\}\!\}_{j+1/2}^{\mathrm{T}}\boldsymbol{F}_{j+1/2} - \frac{1}{2}[\![\boldsymbol{v}]\!]_{j+1/2}^{\mathrm{T}}\boldsymbol{F}_{j+1/2},\\
\boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{F}_{j-1/2} &= \{\!\{\boldsymbol{v}\}\!\}_{j-1/2}^{\mathrm{T}}\boldsymbol{F}_{j-1/2} + \frac{1}{2}[\![\boldsymbol{v}]\!]_{j-1/2}^{\mathrm{T}}\boldsymbol{F}_{j-1/2},\\
\boldsymbol{v}_j^{\mathrm{T}}\hat{\boldsymbol{K}}_{j+1/2}[\![\boldsymbol{v}]\!]_{j+1/2} &= \{\!\{\boldsymbol{v}\}\!\}_{j+1/2}^{\mathrm{T}}\hat{\boldsymbol{K}}_{j+1/2}[\![\boldsymbol{v}]\!]_{j+1/2} - \frac{1}{2}[\![\boldsymbol{v}]\!]_{j+1/2}^{\mathrm{T}}\hat{\boldsymbol{K}}_{j+1/2}[\![\boldsymbol{v}]\!]_{j+1/2},\\
\boldsymbol{v}_j^{\mathrm{T}}\hat{\boldsymbol{K}}_{j-1/2}[\![\boldsymbol{v}]\!]_{j-1/2} &= \{\!\{\boldsymbol{v}\}\!\}_{j-1/2}^{\mathrm{T}}\hat{\boldsymbol{K}}_{j-1/2}[\![\boldsymbol{v}]\!]_{j-1/2} + \frac{1}{2}[\![\boldsymbol{v}]\!]_{j-1/2}^{\mathrm{T}}\hat{\boldsymbol{K}}_{j-1/2}[\![\boldsymbol{v}]\!]_{j-1/2}.
\end{aligned} \tag{2.16}$$

From (2.13) we now conclude that

$$\begin{aligned}
\boldsymbol{v}_j^{\mathrm{T}}\big(\boldsymbol{F}_{j+1/2} - \boldsymbol{F}_{j-1/2}\big) &= \{\!\{\boldsymbol{v}\}\!\}_{j+1/2}^{\mathrm{T}}\boldsymbol{F}_{j+1/2} - \{\!\{\boldsymbol{v}\}\!\}_{j-1/2}^{\mathrm{T}}\boldsymbol{F}_{j-1/2} - \frac{1}{2}\big([\![\varphi]\!]_{j+1/2} + [\![\varphi]\!]_{j-1/2}\big)\\
&= \{\!\{\boldsymbol{v}\}\!\}_{j+1/2}^{\mathrm{T}}\boldsymbol{F}_{j+1/2} - \{\!\{\boldsymbol{v}\}\!\}_{j-1/2}^{\mathrm{T}}\boldsymbol{F}_{j-1/2} + \{\!\{g\}\!\}_{j+1/2} - \{\!\{g\}\!\}_{j-1/2}\\
&\quad - \big\{\!\{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{f}\}\!\big\}_{j+1/2} + \big\{\!\{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{f}\}\!\big\}_{j-1/2},
\end{aligned}$$

while in light of (2.9) we get

$$\boldsymbol{v}_j^{\mathrm{T}}\big(\hat{\boldsymbol{K}}_{j+1/2}[\![\boldsymbol{v}]\!]_{j+1/2} - \hat{\boldsymbol{K}}_{j-1/2}[\![\boldsymbol{v}]\!]_{j-1/2}\big) \le \{\!\{\boldsymbol{v}\}\!\}_{j+1/2}^{\mathrm{T}}\hat{\boldsymbol{K}}_{j+1/2}[\![\boldsymbol{v}]\!]_{j+1/2} - \{\!\{\boldsymbol{v}\}\!\}_{j-1/2}^{\mathrm{T}}\hat{\boldsymbol{K}}_{j-1/2}[\![\boldsymbol{v}]\!]_{j-1/2}. \tag{2.17}$$

We arrive at the semi-discrete entropy inequality

$$\begin{aligned}
&\eta(\boldsymbol{u})_j'(t) + \frac{1}{\Delta x}(G_{j+1/2} - G_{j-1/2})\\
&- \frac{1}{\Delta x^2}\big(\{\!\{\boldsymbol{v}\}\!\}_{j+1/2}^{\mathrm{T}}\hat{\boldsymbol{K}}_{j+1/2}[\![\boldsymbol{v}]\!]_{j+1/2} - \{\!\{\boldsymbol{v}\}\!\}_{j-1/2}^{\mathrm{T}}\hat{\boldsymbol{K}}_{j-1/2}[\![\boldsymbol{v}]\!]_{j-1/2}\big) \le 0,
\end{aligned} \tag{2.18}$$

where the following numerical entropy flux is obviously consistent with (2.10):

$$G_{j+1/2} = \{\!\{g\}\!\}_{j+1/2} + \{\!\{\boldsymbol{v}\}\!\}_{j+1/2}^{\mathrm{T}}\boldsymbol{F}_{j+1/2} - \big\{\!\{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{f}\}\!\big\}_{j+1/2}. \tag{2.19}$$

2.2.3 Additional numerical diffusion

In regions where the diffusion matrix $\boldsymbol{K}$ vanishes, the numerical methods (2.14) or (2.29) reduce to entropy conservative methods whose solutions may exhibit strong oscillations near discontinuities. So to prevent these oscillations some extra numerical diffusion has to be added, either in conservative variables or in entropy variables. Hence the complete scheme is given by

$$\begin{aligned}\boldsymbol{u}_j'(t) = &-\frac{1}{\Delta x}\big(\boldsymbol{F}_{j+1/2} - \boldsymbol{F}_{j-1/2}\big) + \frac{1}{\Delta x^2}\big(\hat{\boldsymbol{K}}_{j+1/2}[\![\boldsymbol{v}]\!]_{j+1/2} - \hat{\boldsymbol{K}}_{j-1/2}[\![\boldsymbol{v}]\!]_{j-1/2}\big)\\ &+ \frac{\varepsilon}{\Delta x^2}\big([\![\boldsymbol{v}]\!]_{j+1/2} - [\![\boldsymbol{v}]\!]_{j-1/2}\big),\end{aligned} \tag{2.20}$$

where we choose the extra viscosity

$$\varepsilon = \alpha \Delta x \tag{2.21}$$

with a suitable constant $\alpha > 0$.

It can be checked easily that the numerical method (2.20) (with the extra viscosity given by (2.21)) satisfies an entropy inequality similar to (2.18) if the numerical entropy flux is replaced by

$$\tilde{G}_{j+1/2} = G_{j+1/2} - \alpha \{\!\{\boldsymbol{v}\}\!\}_{j+1/2}^{\mathrm{T}} [\![\boldsymbol{v}]\!]_{j+1/2}, \tag{2.22}$$

with $G_{j+1/2}$ given by (2.19).

2.2.4 Discretisation of the initial-boundary value problem with zero-flux boundary conditions

The zero-flux IBVP (2.1)–(2.3) is discretized in space by the following variant of (2.20):

$$\boldsymbol{u}_j'(t) = -\frac{1}{\Delta x}(\boldsymbol{J}_{j+1/2} - \boldsymbol{J}_{j-1/2}), \quad j = 1, \dots, M, \tag{2.23}$$

where we implement (2.3) by setting the total numerical flux to zero at the boundaries,i.e., we utilize

$$\boldsymbol{J}_{j+1/2} = \begin{cases} \boldsymbol{F}_{j+1/2} - \dfrac{1}{\Delta x}\big((\hat{\boldsymbol{K}} + \alpha \Delta x \mathbb{I})[\![\boldsymbol{v}]\!]_{j+1/2}\big) & \text{for } j = 1, \dots, M-1,\\ \boldsymbol{0} & \text{for } j = 0 \text{ and } j = M. \end{cases} \tag{2.24}$$

Then the scheme (2.23), (2.24) satisfies the semi-discrete entropy inequality (2.18) for $j = 2, \dots, M-1$. On the other hand, for $j = 1$ we obtain by calculations similar to (2.16)–(2.17), and utilizing (2.13) for $j = 1$, from

$$\eta(\boldsymbol{u})_1'(t) + \frac{1}{\Delta x}\boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{F}_{3/2} - \frac{1}{\Delta x^2}\boldsymbol{v}_1^{\mathrm{T}}\hat{\boldsymbol{K}}_{3/2}[\![\boldsymbol{v}]\!]_{3/2} = 0$$

the inequality

$$\eta(\boldsymbol{u})_1'(t) + \frac{1}{\Delta x}\left(\{\!\{\boldsymbol{v}\}\!\}_{3/2}^{\mathrm{T}}\boldsymbol{F}_{3/2} - \frac{1}{2}[\![\varphi]\!]_{3/2}\right) - \frac{1}{\Delta x^2}\{\!\{\boldsymbol{v}\}\!\}_{3/2}^{\mathrm{T}}\hat{\boldsymbol{K}}_{3/2}[\![\boldsymbol{v}]\!]_{3/2} \le 0.$$

A straightforward calculation and taking into account (2.22) for $j = 1$ reveal that

$$\{\!\{\boldsymbol{v}\}\!\}_{3/2}^{\mathrm{T}}\boldsymbol{F}_{3/2} - \frac{1}{2}[\![\varphi]\!]_{3/2} = \tilde{G}_{3/2} - g_1 + \boldsymbol{v}_1^{\mathrm{T}}\boldsymbol{f}_1 = \tilde{G}_{3/2} + \varphi\big(\boldsymbol{u}_1(t)\big),$$

hence we obtain

$$\eta(\boldsymbol{u})_1'(t) + \frac{1}{\Delta x}\big(\tilde{G}_{3/2} + \varphi\big(\boldsymbol{u}_1(t)\big)\big) - \frac{1}{\Delta x^2}\{\!\{\boldsymbol{v}\}\!\}_{3/2}^{\mathrm{T}}\hat{\boldsymbol{K}}_{3/2}[\![\boldsymbol{v}]\!]_{3/2} \leq 0. \tag{2.25}$$

For $j = M$ we deduce by analogous arguments from

$$\eta(\boldsymbol{u})_M'(t) - \frac{1}{\Delta x}\boldsymbol{v}_M^{\mathrm{T}}\boldsymbol{F}_{M-1/2} + \frac{1}{\Delta x^2}\boldsymbol{v}_M^{\mathrm{T}}\hat{\boldsymbol{K}}_{M-1/2}[\![\boldsymbol{v}]\!]_{M-1/2} = 0$$

the inequality

$$\eta(\boldsymbol{u})_M'(t) - \frac{1}{\Delta x}\big(\tilde{G}_{M-1/2} + \varphi\big(\boldsymbol{u}_M(t)\big)\big) + \frac{1}{\Delta x^2}\{\!\{\boldsymbol{v}\}\!\}_{M-1/2}^{\mathrm{T}}\hat{\boldsymbol{K}}_{M-1/2}[\![\boldsymbol{v}]\!]_{M-1/2} \leq 0. \tag{2.26}$$

Let us now define

$$\eta(\boldsymbol{u})_{\mathrm{tot}}(t) \coloneqq \Delta x \sum_{j=1}^{M} \eta(\boldsymbol{u})_j(t).$$

Then, summing (2.25), (2.18) for $j = 2, \ldots, M-1$, and (2.26), and multiplying the result by Δx, we obtain the inequality

$$\eta(\boldsymbol{u})_{\mathrm{tot}}'(t) \leq \varphi\big(\boldsymbol{u}_M(t)\big) - \varphi\big(\boldsymbol{u}_1(t)\big), \tag{2.27}$$

which is a discrete analogue of (2.11).

2.2.5 Construction of an entropy conservative (EC) numerical flux

Following Tadmor [153], we may obtain an entropy conservative (EC) numerical flux by solving the following integral:

$$\boldsymbol{F}_{j+1/2} = \int_0^1 \boldsymbol{f}\big(\boldsymbol{u}\left(\boldsymbol{v}_j + s\left(\boldsymbol{v}_{j+1} - \boldsymbol{v}_j\right)\right)\big)\,\mathrm{d}s. \tag{2.28}$$

Remark 2.2. *An alternative way of constructing entropy stable schemes could be the following. Suppose that, given $\boldsymbol{u}_{\mathrm{L}}$ and $\boldsymbol{u}_{\mathrm{R}}$, there exists an approximation of $\boldsymbol{\eta}_{\boldsymbol{u},\boldsymbol{u}}$, denoted by $\boldsymbol{H}(\boldsymbol{u}_{\mathrm{L}}, \boldsymbol{u}_{\mathrm{R}})$, that satisfies the Roe-like property $\boldsymbol{v}_{\mathrm{R}} - \boldsymbol{v}_{\mathrm{L}} = \boldsymbol{H}(\boldsymbol{u}_{\mathrm{L}}, \boldsymbol{u}_{\mathrm{R}})(\boldsymbol{u}_{\mathrm{R}} - \boldsymbol{u}_{\mathrm{L}})$. We may then consider the numerical method*

$$\boldsymbol{u}_j'(t) = -\frac{1}{\Delta x}\big(\boldsymbol{F}_{j+1/2} - \boldsymbol{F}_{j-1/2}\big) + \frac{1}{\Delta x^2}\big(\boldsymbol{K}_{j+1/2}[\![\boldsymbol{u}]\!]_{j+1/2} - \boldsymbol{K}_{j-1/2}[\![\boldsymbol{u}]\!]_{j-1/2}\big), \tag{2.29}$$

where $\boldsymbol{K}_{j+1/2} = \hat{\boldsymbol{K}}_{j+1/2}\boldsymbol{H}_{j+1/2}$. Here, $\hat{\boldsymbol{K}}_{j+1/2}$ is given by (2.15) and $\boldsymbol{H}_{j+1/2} = \boldsymbol{H}(\boldsymbol{u}_j, \boldsymbol{u}_{j+1})$. The equality

$$\boldsymbol{K}_{j+1/2}[\![\boldsymbol{u}]\!]_{j+1/2} = \hat{\boldsymbol{K}}_{j+1/2}[\![\boldsymbol{v}]\!]_{j+1/2}$$

allows one to prove the entropy inequality (2.18) reasoning as in the previous case.

2.3 Applicative models

2.3.1 A diffusively corrected multi-class traffic model (DCMCLWR model)

We consider the system (2.1) with a flux function defined by

$$\boldsymbol{f}(\boldsymbol{u}) = \mathcal{V}(u)\big(V_1^{\max}u_1, \dots, V_N^{\max}u_N\big)^{\mathrm{T}}, \tag{2.30}$$

where $V_i^{\max}$ is the preferential (maximum) velocity of species i (driver class i); $u = u_1 + \cdots + u_N$ is the total density; and $\mathcal{V}$ is a hindrance function that is usually assumed to satisfy

$$\mathcal{V}(0) = 1, \quad \mathcal{V}(u_{\max}) = 0, \quad \mathcal{V}'(u) < 0 \quad \text{for } 0 < u < u_{\max},$$

where $u_{\max}$ is a maximum density. We assume, furthermore, that $V_1^{\max} > V_2^{\max} > \ldots > V_N^{\max}$. Under these assumptions on $\boldsymbol{f}$, the first-order system (2.4) corresponds to the multiclass extension, introduced in [29, 162], of the well-known Lighthill-Whitham-Richards (LWR) single-class kinematic traffic model [123, 144]. An entropy pair (η, g) for this multiclass model is given by [30]

$$\eta(\boldsymbol{u}) = \sum_{i=1}^{N} \frac{u_i(\log(u_i) - 1)}{V_i^{\max}}, \quad g(\boldsymbol{u}) = \mathcal{V}(u)\sum_{i=1}^{N} u_i \log(u_i) - \tilde{\mathcal{V}}(u), \tag{2.31}$$

where $\tilde{\mathcal{V}}(u)$ is any primitive of $\mathcal{V}(u)$. Using $\boldsymbol{v}(\boldsymbol{u}) \coloneqq \eta_{\boldsymbol{u}}(\boldsymbol{u})$ (see Section 2.2.1) we then obtain the entropy variables $\boldsymbol{v} = (v_1, \dots, v_N)^{\mathrm{T}}$ given by

$$v_i = \frac{\log(u_i)}{V_i^{\max}} \Leftrightarrow u_i = \exp(V_i^{\max}v_i), \quad i = 1, \dots, N.$$

In addition, the following notation will be used:

$$v \coloneqq \sum_{i=1}^{N} \exp(V_i^{\max}v_i).$$

Notice that the transformation $\boldsymbol{u} \to \boldsymbol{v}$ is one-to-one from $(0, \infty)^N$ to $\mathbb{R}^N$, but is not defined when $u_i = 0$. Now we associate the behavior of drivers with an anticipation distance $L_{\min}$. Then the reaction of a driver at (x, t) depends on $u(x + L_{\min}, t)$. Using a Taylor expansion of $\mathcal{V}(u(x + L_{\min}, t))$ around $u(x, t)$, we obtain

$$\mathcal{V}\big(u(x + L_{\min}, t)\big) = \mathcal{V}(u) + \mathcal{V}'(u)(L_{\min}\partial_x u) + \mathcal{O}(L_{\min}^2),$$

where all quantities on the right-hand side are evaluated at (x, t). Neglecting the $\mathcal{O}(L_{\min}^2)$ term and inserting the remaining expression into (2.1), we have

$$\partial_t u_i(x, t) + \partial_x\big(u_i(x, t)V_i^{\max}\mathcal{V}(u)\big) = \partial_x\big(-L_{\min}\mathcal{V}'(u)u_i(x, t)V_i^{\max}\partial_x u(x, t)\big).$$

To further simplify the model we remove the dependencies on individual driver classes. Hence, we propose to use the positive semidefinite diffusion matrix

$$\boldsymbol{K}(\boldsymbol{u}) = \beta(u)\boldsymbol{I}, \tag{2.32}$$

where $\boldsymbol{I}$ denotes the $N \times N$ identity matrix and $\beta(u) \geq 0$ is a scalar function. Since we define the diffusion term based on the entropy variables form, following (2.12), the diffusion matrix is given by

$$\hat{\boldsymbol{K}}(\boldsymbol{v}) = \boldsymbol{K}(\boldsymbol{u})\eta_{\boldsymbol{u},\boldsymbol{u}}(\boldsymbol{u})^{-1} = \beta(v)\,\mathrm{diag}\big(V_1^{\max}\exp(V_1^{\max}v_1), \ldots, V_N^{\max}\exp(V_N^{\max}v_N)\big).$$

Consequently, in terms of the driver class densities we have

$$\begin{aligned}
&\hat{\boldsymbol{K}}_{j+1/2}[\![\boldsymbol{v}(\boldsymbol{u})]\!]_{j+1/2} \\
&= \hat{\boldsymbol{K}}\{\!\{\boldsymbol{v}(\boldsymbol{u})\}\!\}_{j+1/2}[\![\boldsymbol{v}(\boldsymbol{u})]\!]_{j+1/2} \\
&= \beta\left(\sum_{i=1}^{N}\sqrt{u_{i,j}u_{i,j+1}}\right)\mathrm{diag}\left(\sqrt{u_{1,j}u_{1,j+1}}, \ldots, \sqrt{u_{N,j}u_{N,j+1}}\right)\begin{pmatrix}\log u_{1,j+1} - \log u_{1,j} \\ \vdots \\ \log u_{N,j+1} - \log u_{N,j}\end{pmatrix}.
\end{aligned} \tag{2.33}$$

In agreement with the definition above, the extra viscosity term is given by

$$\frac{\varepsilon}{\Delta x^2}\big([\![\boldsymbol{v}(\boldsymbol{u})]\!]_{j+1/2} - [\![\boldsymbol{v}(\boldsymbol{u})]\!]_{j-1/2}\big), \tag{2.34}$$

where the ith component of $[\![\boldsymbol{v}(\boldsymbol{u})]\!]_{j+1/2}$ is given by $[\![v_i(\boldsymbol{u})]\!]_{j+1/2} = 1/V_i^{\max}(\log(u_{i,j+1}) - \log(u_{i,j}))$.

On the other hand, we will use the hindrance function $\mathcal{V}(u) = 1 - u$ due to Greenshields [87]. Replacing this function in (2.30),from (2.28) we obtain that

$$\begin{aligned}
F_{j+1/2,i} &= \int_0^1 \mathcal{V}\left(\sum_{k=1}^{N} u_k\left(v_{k,j} + s\left(v_{k,j+1} - v_{k,j}\right)\right)\right) V_i^{\max} u_i\left(v_{i,j} + s\left(v_{i,j+1} - v_{i,j}\right)\right)\mathrm{d}s \\
&= \int_0^1 \left(1 - \sum_{k=1}^{N} u_k\left(v_{k,j} + s\left(v_{k,j+1} - \boldsymbol{v}_j\right)\right)\right) V_i^{\max} u_i\left(v_{i,j} + s\left(v_{i,j+1} - v_{i,j}\right)\right)\,\mathrm{d}s.
\end{aligned} \tag{2.35}$$

Since

$$u_i\big(v_{i,j} + s(v_{i,j+1} - v_{i,j})\big) = \exp\Big(\log(u_{j,i}) + s\big(\log(u_{j+1,i}) - \log(u_{j,i})\big)\Big) = u_{j,i}\left(\frac{u_{j,i+1}}{u_{j,i}}\right)^s,$$

we can rewrite (2.35) as

$$F_{j+1/2,i} = \int_0^1 \left(1 - \sum_{k=1}^{N} u_{j,k}\left(\frac{u_{j,k+1}}{u_{j,k}}\right)^s\right) v_i^{\max} u_{j,i}\left(\frac{u_{j,i+1}}{u_{j,i}}\right)^s\,\mathrm{d}s.$$

Evaluating the integral in closed form, we get the entropy stable numerical flux

$$\boldsymbol{F}_{j+1/2} = (F_{j+1/2,1}, \ldots, F_{j+1/2,N})^{\mathrm{T}} \tag{2.36}$$

where

$$F_{j+1/2,i} = v_i^{\max}\left(\frac{u_{j+1,i} - u_{j,i}}{\log(u_{j+1,i}) - \log(u_{j,i})} - \sum_{k=1}^{N}\frac{u_{j+1,k}u_{j+1,i} - u_{j,k}u_{j,i}}{\log(u_{j+1,k}u_{j+1,i}) - \log(u_{j,k}u_{j,i})}\right), \quad i = 1, \ldots, N. \tag{2.37}$$

Equations (2.33), (2.34) and (2.37) complete the definition of the semi-discrete numerical scheme (2.20).

Remark 2.3. *In order to get rid of the singularity of the entropy variables when one of the terms in the* log *differences in* (2.37) *is zero or the difference is zero, we use the following third-order approach:*

$$\log(u) - \log(v) \approx \frac{u - v}{0.5(u + v)}.$$

which means the logarithmic mean on (2.37)*, is replaced by*

$$\frac{u - v}{\log(u) - log(v)} \approx \frac{1}{2}(u + v)$$

when appropriate. An alternative stable numerical algorithm used to compute the logarithmic mean is given in [98, App. B].

Remark 2.4. *Although we considered other forms for the hindrance function, the integral* (2.28) *is difficult to compute in general or can result in a numerically unstable flux [78]. Indeed equation* (2.28) *results in a closed form only for a limited selection of functions, such as functions of the form* $\mathcal{V}(u) = (1 - u)^n$ *with* $n \in \mathbb{N}$*. We are aware that the development of entropy stable flows for more general forms of flow functions in multispecies kinematic flow models needs more extensive study.*

2.3.2 Settling of dispersions of droplets and colloidal particles

The settling of a dispersion of droplets or that of a suspension of colloidal solid particles dispersed in a fluid can be modeled by system of convection-diffusion equations of the form (2.1) for $I = [0, L]$, where t is time, x is depth, and $\boldsymbol{u}(x, t)$ is the vector of volume fractions of particles u_i of class i, $i = 1, \ldots, N$ [51]. The problem (2.1), (2.2) is completed with the zero-flux boundary condition (2.3).

Particles are characterized by their diameter d_i and settling velocities $V_1 > V_2 > \ldots > V_N$. Moreover, we assume that the flux vector $\boldsymbol{f}(\boldsymbol{u})$ has the form

$$\boldsymbol{f}(\boldsymbol{u}) = \mathcal{V}(u)(V_1 u_1, \ldots, V_N u_N)^{\mathrm{T}},$$

where again $u := u_1 + \cdots + u_N$. According to [1], the Stokes terminal velocities V_i are given by

$$V_i = \frac{(\rho_{\mathrm{d}} - \rho_{\mathrm{c}}) g d_i^2}{18 \mu_{\mathrm{c}}}, \quad i = 1, \ldots, N,$$

where ρ and μ, respectively, denote density and viscosity, and the indices d and c respectively, refer to the disperse or continuous phase, and in this formula $g = 9.81\,\mathrm{m/s^2}$ is the acceleration of gravity. A common choice for the so-called hindered settling function $\mathcal{V}(u)$ is given by the Richardson-Zaki [144] expression:

$$\mathcal{V}(u) = \begin{cases} (1 - u)^{n_{\mathrm{RZ}}} & \text{if } u \leq 1, \\ 0 & \text{if } u > 1. \end{cases}$$

The diffusion matrix is again defined by (2.32), where $\beta(u) = D_0 \mathcal{V}(u)$ for some constant $D_0 > 0$. For the numerical examples we choose $n_{\mathrm{RZ}} = 2$, and follow the same procedure as in the previous

application. The numerical diffusion is given by (2.33) and the numerical flux function, obtained from (2.28), is now given by (2.36) with

$$\begin{aligned} F_{i,j+1/2} = V_i \Bigg(& \frac{u_{j+1,i} - u_{j,i}}{\log(u_{j+1,i}) - \log(u_{j,i})} - 2\sum_{k=1}^{N} \frac{u_{j+1,k}u_{j+1,i} - u_{j,k}u_{j,i}}{\log(u_{j+1,k}u_{j+1,i}) - \log(u_{j,k}u_{j,i})} \\ & + \sum_{k=1}^{N} \frac{u_{j+1,k}^2 u_{j+1,i} - u_{j,k}^2 u_{j,i}}{\log(u_{j+1,k}^2 u_{j+1,i}) - \log(u_{j,k}^2 u_{j,i})} \\ & + \sum_{\substack{k,l=1 \\ k\neq l}}^{N} \frac{u_{j+1,k}u_{j+1,l}u_{j+1,i} - u_{j,k}u_{j,l}u_{j,i}}{\log(u_{j+1,k}u_{j+1,l}u_{j+1,i}) - \log(u_{j,k}u_{j,l}u_{j,i})} \Bigg), \quad i = 1, \dots, N. \end{aligned} \tag{2.38}$$

2.4 Numerical examples

2.4.1 Preliminaries

For the time integration in all examples, we use a second-order strong stability preserving Runge-Kutta scheme (SSPRK2, also known as Heun's method), i.e., for a given spatial discretisation $\boldsymbol{h}(\boldsymbol{U})$ such as the semidiscrete form (2.20) along with definitions (2.33), (2.34) and (2.37) or (2.38) and where $\boldsymbol{U}(t)$ represents the vector of numerical solutions at all spatial positions at time t, i.e., $\boldsymbol{U}(t) = (\boldsymbol{u}_1(t), \dots, \boldsymbol{u}_M(t))^{\mathrm{T}}$, the integration scheme for the system $\boldsymbol{U}'(t) = \boldsymbol{h}(\boldsymbol{U})$ is given as follows, where we assume that we wish to advance the solution from $\boldsymbol{U}^n \approx \boldsymbol{U}(t_n)$ to $\boldsymbol{U}^{n+1} \approx \boldsymbol{U}(t_{n+1})$, where $t_{n+1} = t_n + \Delta t$:

$$\begin{aligned} \boldsymbol{U}^{(1)} &= \boldsymbol{U}^n + \Delta t \boldsymbol{h}(\boldsymbol{U}^n), \\ \boldsymbol{U}^{(2)} &= \boldsymbol{U}^{(1)} + \Delta t \boldsymbol{h}(\boldsymbol{U}^{(1)}), \\ \boldsymbol{U}^{n+1} &= \frac{1}{2}(\boldsymbol{U}^n + \boldsymbol{U}^{(2)}), \quad n = 0, 1, 2, \dots. \end{aligned}$$

We choose the time step Δt at each iteration t_n according to the following CFL condition:

$$\frac{\Delta t}{\Delta x} \max_{1\le j\le M} \rho\big(\boldsymbol{f}_u(\boldsymbol{u}_j^n)\big) + \frac{\Delta t}{2\Delta x^2} \max_{1\le j\le M} \rho\big(\boldsymbol{K}(\boldsymbol{u}_j^n)\big) = C_{\mathrm{CFL}} \tag{2.39}$$

where $\rho(\cdot)$ is the spectral radius. In all cases, we calculate the approximate total L^1 error at a given time t as follows. We assume that the spatial computational domain is subdivided into M equal-sized cells of width Δx, and that we calculate approximate errors by utilizing a reference solution defined on a mesh with $M_{\mathrm{ref}} > M$ cells, where we assume that $R := M_{\mathrm{ref}}/M$ is an integer. Then we calculate the projection of the reference solution onto the coarser grid,

$$\tilde{u}_{j,i}^{\mathrm{ref}}(t) = \frac{1}{R}\sum_{k=1}^{R} u_{R(j-1)+k,i}^{\mathrm{ref}}(t), \quad j = 1, \dots, M, \quad i = 1, \dots, N,$$

and then calculate the total approximate total L^1 error by summing the corresponding errors of each species, that is,

$$e_M^{\mathrm{tot}} = \frac{1}{M}\sum_{i=1}^{N}\sum_{j=1}^{M} \big|\tilde{u}_{j,i}^{\mathrm{ref}}(t) - u_{j,i}^{M}(t)\big|.$$

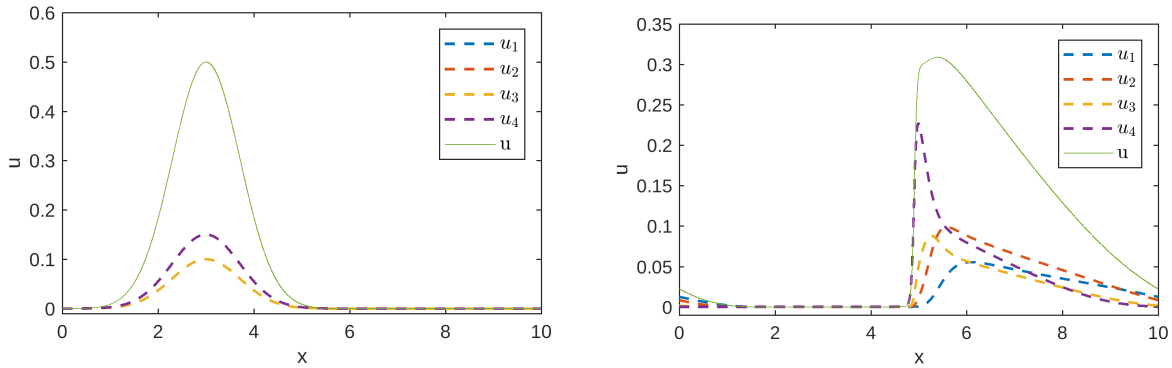


Figure 2.1: Example 2.1 (traffic model, non-degenerate diffusion, $N = 4$): (left) initial condition (2.41), (right) reference numerical solution at simulated time $t = 0.1\,\mathrm{h}$ obtained by the ES scheme with $\alpha = 1.5$ and $M_{\mathrm{ref}} = 12800$ (figure produced by the author).

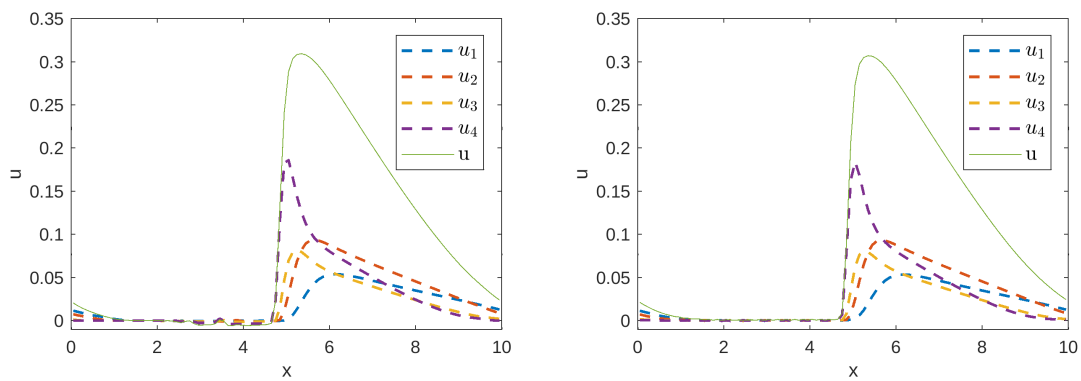


Figure 2.2: Example 2.1 (traffic model, non-degenerate diffusion, $N = 4$): numerical solution at simulated time $t = 0.1\,\mathrm{h}$ obtained by the entropy stable scheme with $M = 100$ and (left) with zero extra viscosity, (right) with extra viscosity (2.21) with $\alpha = 1.5$ (figure produced by the author).

The corresponding (approximate) convergence rate between successive grids with discretisations $M/2$ and M is given by

$$\theta_M := \log_2(e^{\mathrm{tot}}_{M/2}/e^{\mathrm{tot}}_M).$$

2.4.2 Example 2.1 (traffic model, non-degenerate diffusion, $N = 4$)

First, we test the entropy conserving scheme on a regular grid. We consider a circular road of length $L = 10\,\mathrm{mi}$ and $N = 4$ driver classes with the velocities $V_1^{\max} = 60\,\mathrm{mi/h}$, $V_2^{\max} = 55\,\mathrm{mi/h}$, $V_3^{\max} = 50\,\mathrm{mi/h}$, and $V_4^{\max} = 45\,\mathrm{mi/h}$, along with a uniform anticipation length of $L_{\min} = 0.03\,\mathrm{mi}$ and the non-degenerate diffusion term defined by (2.32) and

$$\beta(u) = \frac{L_{\min}}{N}(\overline{V^{\max}}), \quad \overline{V^{\max}} := \frac{1}{N}\sum_{i=1}^{N} V_i^{\max}. \tag{2.40}$$

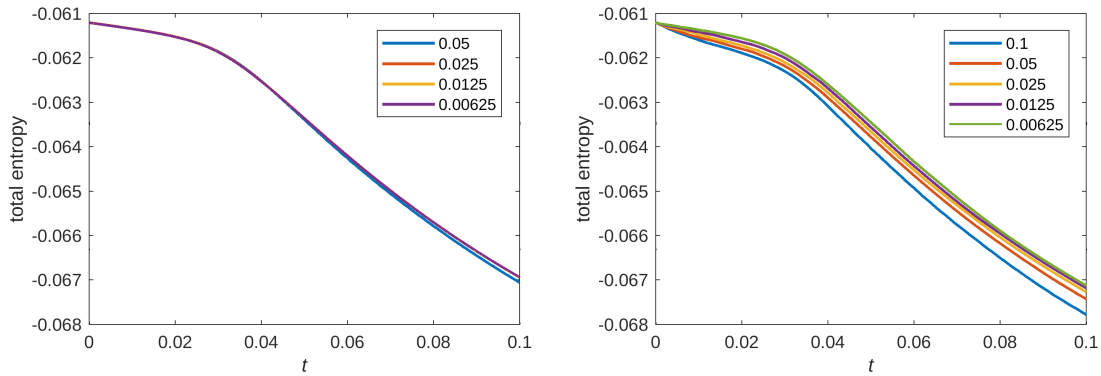


Figure 2.3: Example 2.1 (traffic model, non-degenerate diffusion, $N = 4$): total entropy $\mathcal{E}_n^{\text{tot}}$ of the numerical solution at different mesh sizes (left) without extra viscosity and (right) with extra viscosity (2.21) with $\alpha = 1.5$ (figure produced by the author).

	KT		LLF		ES ($\alpha = 1.5$)	
M	e_M^{tot}	θ_M	e_M^{tot}	θ_M	e_M^{tot}	θ_M
100	4.024e-2	—	1.445e-1	—	3.140e-2	—
200	1.524e-2	1.401	8.830e-2	0.711	1.379e-2	1.188
400	5.881e-3	1.374	5.119e-2	0.786	6.857e-3	1.008
800	4.232e-3	0.475	2.868e-2	0.836	3.212e-3	1.094
1600	3.637e-3	0.219	1.574e-2	0.866	1.369e-3	1.230
3200	3.350e-3	0.055	8.718e-3	0.853	5.476e-4	1.322

Table 2.1: Example 2.1 (traffic model, non-degenerate diffusion, $N = 4$): approximate total L^1 errors (e_M^{tot}) and convergence rates (θ_M) at simulated time $t = 0.1$ (table produced by the author).

The initial traffic platoon (see Figure 2.1 (left)) is given by

$$\boldsymbol{u}_0(x,0) = p(x)(0.2, 0.3, 0.2, 0.3)^{\mathrm{T}}, \quad p(x) = 0.5\exp(-(x-3)^2). \tag{2.41}$$

Numerical approximations are computed with $C_{\text{CFL}} = 0.25$ at simulated time $t = 0.1\,\text{h}$ using the method of lines of the semidiscretisation given by the numerical flux (2.37) combined with the numerical diffusion (2.33). The performances of the entropy stable (ES) scheme without and with extra viscosity are compared in Figure 2.2. Here and in Examples 2.2 to 2.4 we also verify that the method is indeed entropy stable by plotting the following total entropy for $t = t_n = n\Delta t$:

$$\mathcal{E}_n^{\text{tot}} := \sum_{j=1}^{M} \overline{\eta(\boldsymbol{u}_j(t_n))}\Delta x,$$

see Figure 2.3 for this example.We observe that $\mathcal{E}_n^{\text{tot}}$ decreases for the base scheme without extra viscosity (corresponding to $\alpha = 0$), as expected from (2.18). It is also clear from Figures 2.2 and 2.3, that the extra viscosity helps to prevent oscillations while preserving the general entropy decay

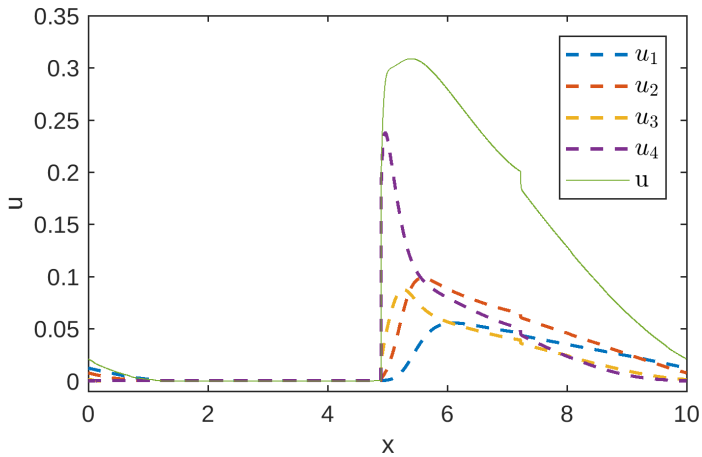


Figure 2.4: Example 2.2 (traffic model, continuous degenerate diffusion, $N = 4$): reference numerical solution at simulated time $t = 0.1$ h obtained by the ES scheme with $\alpha = 1.5$ and $M_{\text{ref}} = 12800$ (figure produced by the author).

of the solution. It is important to note that nonphysical negative values due to oscillations preclude the computation of meaningful total entropy values for coarser grids. Moreover, in this example the approximate total L^1 errors were computed by using a numerical reference solution (ES scheme with $M_{\text{ref}} = 12800$, $\alpha = 1.5$), and are shown in Table 2.1. For comparison solutions obtained with Kurganov-Tadmor (KT) scheme [116] and local Lax-Friedrichs (LLF) scheme [122], augmented by the expression (2.33) to handle the degenerate diffusion, are also presented. The KT and LLF schemes are known for the simplicity of their Riemann-solver-free approach, which makes them a computationally efficient universal tool for a wide variety of applications [115]. The main disadvantage of the LLF scheme lies in its large numerical dissipation, an issue that KT schemes try to solve by using more accurate information of the local propagation speeds [116]. From a point of view of finding a balance between computational cost while controlling the amount of numerical dissipation, we find these methods constitute a good reference point to compare the ES method against. As will be presented in this and the following examples, in general the ES scheme matches the reference methods with respect to the absolute error obtained, and performs better regarding computational cost, while preserving a numerical equivalent of entropy inequality (2.10). The disadvantage are the strong requirements necessary for its application, and the experimentation-based adjustment of the viscosity parameter.

With respect to the error table, we observe that the ES scheme exhibits convergence rates that are consistently slightly larger than one.

2.4.3 Example 2.2 (traffic model, continuous degenerate diffusion, $N = 4$)

In Example 2.2, under the same initial conditions as in Example 2.1, we test the model with the diffusion matrix (2.32), where we define

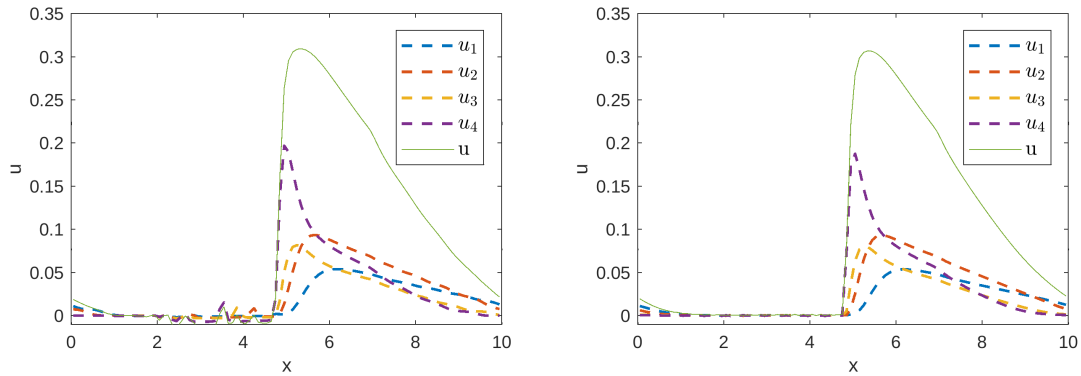


Figure 2.5: Example 2.2 (traffic model, continuous degenerate diffusion, $N = 4$): numerical solution at simulated time $t = 0.1$ h obtained by the entropy stable scheme with $M = 100$ and (left) with zero extra viscosity, (right) with extra viscosity (2.21) with $\alpha = 1.5$ (figure produced by the author).

	KT			LLF			ES ($\alpha = 1.5$)		
M	e_M^{tot}	θ_M	CPU[s]	e_M^{tot}	θ_M	CPU[s]	e_M^{tot}	θ_M	CPU[s]
100	7.846e-2	—	0.76	1.722e-1	—	0.52	6.108e-2	—	0.36
200	4.327e-2	0.859	3.00	1.166e-1	0.562	2.09	3.254e-2	0.908	1.41
400	1.749e-2	1.306	11.50	8.042e-2	0.537	8.19	1.5425e-2	1.078	5.56
800	7.710e-3	1.182	46.51	5.284e-2	0.606	32.82	9.020e-3	0.773	22.23
1600	4.011e-3	0.943	197.69	3.225e-2	0.712	134.94	5.433e-3	0.731	91.72
3200	2.707e-3	0.567	863.22	1.884-2	0.775	626.73	2.466e-3	1.139	413.56

Table 2.2: Example 2.2 (traffic model, continuous degenerate diffusion, $N = 4$): approximate L^1 errors (e_M^{tot}), convergence rates (θ_M), and CPU times (CPU) at simulated time $t = 0.1$ (table produced by the author).

$$\beta(u) = \begin{cases} 0 & \text{if } u \le u_{\text{c}}, \\ \dfrac{L_{\min}\overline{V^{\max}}}{N}(u - u_{\text{c}}) & \text{if } u > u_{\text{c}}, \end{cases}$$

where $\overline{V^{\max}}$ is defined as in (2.40), and we choose $u_{\text{c}} = 0.2$. The new diffusion matrix now depends on the total density $u = u_1 + \cdots + u_N$ and vanishes when $u \le u_{\text{c}}$, but is still a continuous function of u. Note that since $\beta(u) = 0$ for $u \le u_{\text{c}}$, for these u-values the method (2.14) is reduced to an entropy conservative method for first-order systems of conservation laws that exhibits oscillations. The resulting model is strongly degenerate. Figure 2.6 confirms that also this example exhibits a decrease in approximate total entropy. Approximate L^1-errors for $\boldsymbol{u}$ computed by a numerical reference solution (ES scheme with $M_{\text{ref}} = 12800$, α=1.5) are shown in Table 2.2. That table also shows CPU times. It is worth noting that the ES scheme is the one that executes most rapidly and produces errors that are only slightly larger in some instances that those of the KT scheme at the same discretisation. Thus, we can say that the ES scheme is the most efficient (in terms of error reduction versus CPU time) in this case.

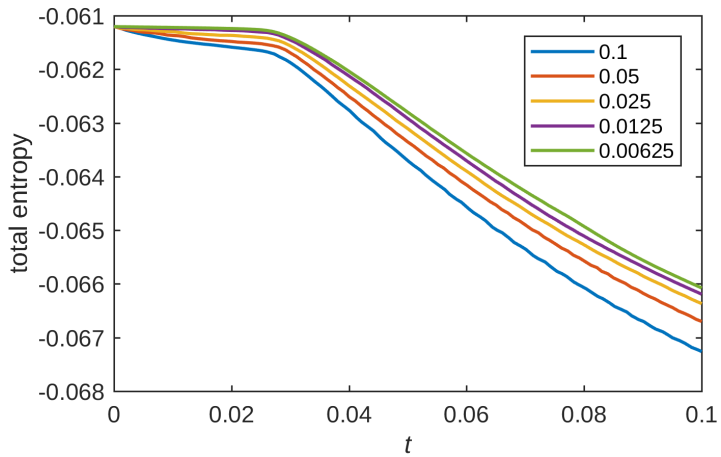


Figure 2.6: Example 2.2 (traffic model, continuous degenerate diffusion, $N = 4$): total entropy $\mathcal{E}_n^{\text{tot}}$ of the numerical solution at different mesh sizes, based on the extra viscosity (2.21) with $\alpha = 1.5$ (figure produced by the author).

2.4.4 Example 2.3 (traffic model, discontinuous degenerate diffusion, $N = 4$)

Under the same initial conditions of Examples 2.1 and 2, now we test the model with the diffusion matrix (2.32) with

$$\beta(u) = \begin{cases} 0 & \text{if } u \le u_{\text{c}}, \\ L_{\min}\overline{V^{\max}}/N & \text{if } u > u_{\text{c}}, \end{cases}$$

where $\overline{V^{\max}}$ is still defined as in (2.40) and we choose $u_{\text{c}} = 0.2$. The resulting model is strongly degenerate, and an additional complication comes from the fact that β, and therefore $\boldsymbol{K}$, are now discontinuous functions of u. Figure 2.7 shows the reference solution obtained for this case, and Figure 2.8 displays numerical solutions with $M = 100$. Entropy stability still holds, as depicted in Figure 2.9. The approximate L^1-errors for $\boldsymbol{u}$ computed by using a numerical reference solution (ES scheme with $M_{\text{ref}} = 12800$, $\alpha = 1.5$) are shown in Table 2.3.

2.4.5 Example 2.4 (traffic model, continuous degenerate diffusion, non-smooth initial datum, $N = 4$)

Under the assumptions of Example 2.2, we replace the smooth initial condition (2.41) by the following function, corresponding to a "platoon" of traffic:

$$\boldsymbol{u}_0(x,0) = p(x)(0.2, 0.3, 0.2, 0.3)^{\text{T}}, \quad p(x) = \begin{cases} 10x & \text{for } 0 < x \le 0.1, \\ 1 & \text{for } 0.1 < x \le 0.9, \\ -10(x-1) & \text{for } 0.9 < x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.42}$$

KT				LLF			ES ($\alpha = 1.5$)		
M	e_M^{tot}	θ_M	CPU[s]	e_M^{tot}	θ_M	CPU[s]	e_M^{tot}	θ_M	CPU[s]
100	4.742e-2	—	0.78	1.486e-1	—	0.53	3.796e-2	—	0.37
200	2.053e-2	1.207	2.97	9.143e-2	0.701	2.26	1.814e-2	1.065	1.41
400	9.797e-3	1.069	12.83	5.348e-2	0.774	8.92	9.037e-3	1.005	6.06
800	6.752e-3	0.535	55.48	3.130e-2	0.773	38.36	4.301e-3	1.071	26.06
1600	4.941e-3	0.450	350.13	1.793e-2	0.804	233.55	2.102e-3	1.033	168.91
3200	4.536e-3	0.123	1452.28	1.043e-2	0.781	984.89	1.333e-3	0.656	656.21

Table 2.3: Example 2.3 (traffic model, continuous degenerate diffusion, $N = 4$): approximate L^1 errors (e_M^{tot}), convergence rates (θ_M), and CPU times (CPU) at simulated time $t = 0.1$ (table produced by the author).

KT				ES ($\alpha = 1.5$)			CU		
M	e_M^{tot}	θ_M	CPU[s]	e_M^{tot}	θ_M	CPU[s]	e_M^{tot}	θ_M	CPU[s]
100	1.365e-1	—	1.92	32.053*	—	97.43	1.216e-1	—	1.78
200	7.765e-2	0.814	3.34	7.408-2	—	2.22	6.948e-2	0.807	4.49
400	3.751e-2	1.050	14.67	3.931e-2	0.914	10.09	3.417e-2	1.024	20.34
800	1.843e-2	1.025	59.91	2.157e-2	0.866	42.54	1.707e-2	1.001	81.32
1600	1.030e-3	0.840	219.62	1.142e-2	0.918	151.49	9.736e-3	0.811	298.57
3200	8.006e-3	0.363	963.64	6.921-3	0.722	785.65	8.562e-3	0.185	1362.84

Table 2.4: Example 2.4 (traffic model, continuous degenerate diffusion, non-smooth initial datum, $N = 4$): approximate L^1 errors (e_M^{tot}), convergence rates (θ_M), and CPU times (CPU) at simulated time $t = 0.2$ (table produced by the author).

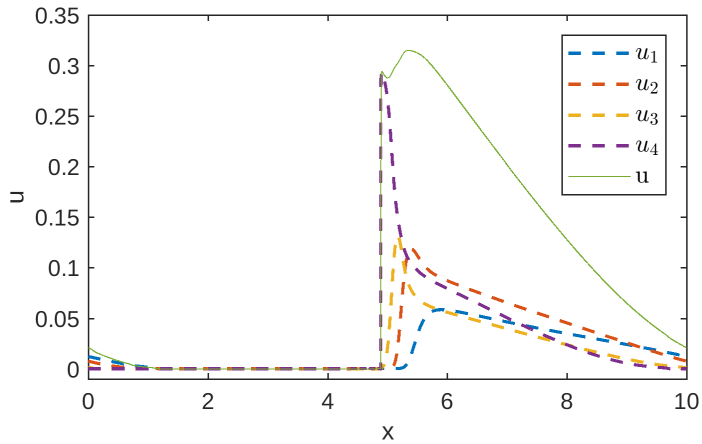


Figure 2.7: Example 2.3 (traffic model, discontinuously degenerate diffusion, $N = 4$): reference numerical solution at simulated time $t = 0.1\,\mathrm{h}$ obtained by the ES scheme with $\alpha = 1.5$ and $M_{\mathrm{ref}} = 12800$ (figure produced by the author).

	i	1	2	3	4	5	6	7	8
20% glycerol	$d_i[\mu m]$	201.430	140.2	99.751	68.986	48.391	34.185	23.810	6.101
	$\phi_i^0[\%]$	0.0859	0.6410	4.4309	7.928	4.7065	1.5710	0.5720	0.1758
50% glycerol	$d_i[\mu m]$	417.819	291.590	202.854	143.384	100.118	68.629	48.259	33.886
	$\phi_i^0[\%]$	0.329	11.380	25.010	9.921	2.305	0.821	0.502	0.183

Table 2.5: Example 2.5 and 2.6 (settling model, discontinuous degenerate diffusion, $N = 8$): droplet particle diameters d_i and initial concentrations ϕ_i^0 (table produced by the author).

Figure 2.10 shows the initial condition and the reference solution for this case, and Figure 2.11 shows the numerical results for the ES and KT schemes with $M = 100$ and $M = 800$. As is shown in Figure 2.11 (top), this set of initial conditions causes strong oscillations near the transition between hyperbolic and parabolic regimes. On the $M = 100$ mesh, these oscillations produce artefacts that remain through time iterations even with high extra viscosity. In order to avoid these artefacts, a finer mesh was required; Figure 2.11 (bottom) compares the entropy conservative scheme solution against a solution by the KT scheme with $M = 800$. In Table 2.4 we show L^1-errors for $\boldsymbol{u}$ computed by a numerical reference solution (ES scheme with $\alpha = 1.5$, $M_{\mathrm{ref}} = 12800$). The large value of the $M = 100$ entry for the ES scheme in that table indicates that additional numerical viscosity was not sufficient to prevent strong oscillations (see Figure 2.11). A numerical solution obtained with the less diffusive central-upwind (CU) scheme by Kurganov et al. [115] is also presented for comparison.Central-upwind schemes improve further on the projection step if one is looking for a less dissipative scheme that could behave closer to the upwind alternatives. This usually results in improvements of the resolution of nonsmooth parts of the solution [113]. However, for this particular case, we could not observe significant differences in the total error with respect to the more simple KT scheme. Entropy stability also holds for this case, see Figure 2.12.

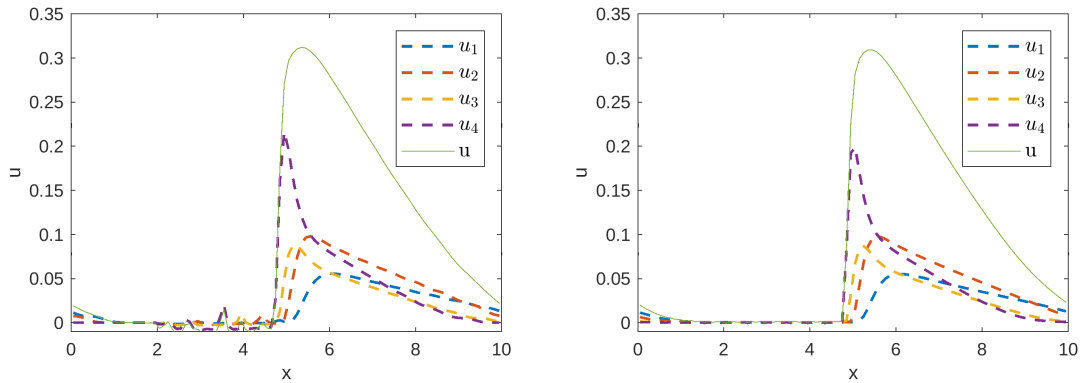


Figure 2.8: Example 2.3 (traffic model, discontinuous degenerate diffusion, $N = 4$): numerical solution at simulated time $t = 0.1$ h obtained by the entropy stable scheme with $M = 100$ and (left) with zero extra viscosity, (right) with extra viscosity (2.21) with $\alpha = 1.5$ (figure produced by the author).

	KT		COMP-GLF		ES ($\alpha = 1 \times 10^{-12}$)	
M	e_M^{tot}	θ_M	e_M^{tot}	θ_M	e_M^{tot}	θ_M
100	2.877e-4	—	9.554e-5	—	2.187e-4	—
200	1.345e-4	1.097	4.694e-5	1.025	1.027e-4	1.090
400	6.869e-5	0.969	2.872e-5	0.709	5.313e-5	0.951
800	3.882e-5	0.823	2.215e-5	0.375	2.202e-5	1.270
1600	2.332e-5	0.735	1.880e-5	0.237	8.378e-6	1.394

Table 2.6: Example 2.5 (settling model, discontinuous degenerate diffusion, $N = 8$): approximate L^1 errors (e_M^{tot}) and convergence rates (θ_M) at simulated time $t = 50$ (table produced by the author).

2.4.6 Example 2.5 (settling model, discontinuous degenerate diffusion, $N = 8$)

In this example we consider the settling of dispersions of glycerol droplets of total initial concentration 50% in a column of biodiesel of depth $L = 20$ mm according to the experimental setup of [1]. The density of biodiesel is $\rho_c = 880\,\text{kg/m}^3$ and that of glycerol is $\rho_d = 1090\,\text{kg/m}^3$. Other parameters are the viscosity $\mu_c = 6.5$ mPa and the diffusivity $D_0 = 10 \times 10^{-7}\,\text{m}^2/\text{s}$. We consider $N = 8$ droplet size classes. The corresponding droplet diameters d_i and initial concentrations u_i^0 have been reconstructed from droplet size histograms [51], see Table 2.5. We also introduce a discontinuous diffusion function $\beta(u)$, namely

$$\beta(u) = \begin{cases} 0 & \text{if } u \le u_c, \\ D_0 \mathcal{V}(u) & \text{if } u > u_c, \end{cases}$$

where u_c is a critical density, or gel point, accounting for the onset of compression effects when entities of the disperse phase start forming permanent contact, for which we choose $u_c = 0.1$ in this example. Numerical results are obtained by the entropy stable (ES), component-wise global Lax-Friedrichs (COMP-GLF, a component-wise WENO scheme with a Lax-Friedrichs-type flux splitting, see [44]

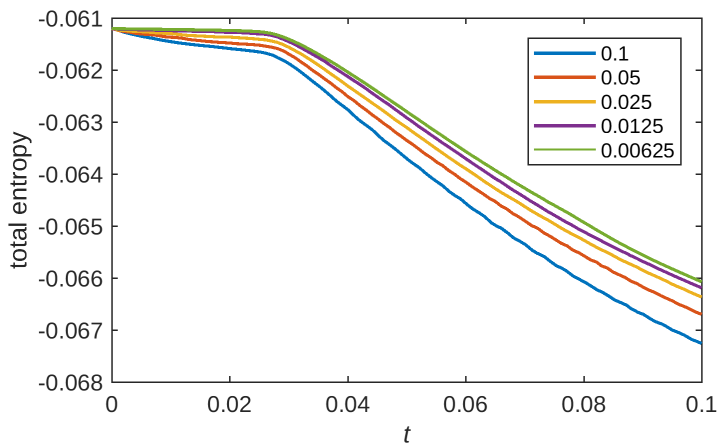


Figure 2.9: Example 2.3 (traffic model, discontinuous degenerate diffusion, $N = 4$): total entropy $\mathcal{E}_n^{\mathrm{tot}}$ of the numerical solution at different mesh sizes, based on the extra viscosity (2.21) with $\alpha = 1.5$ (figure produced by the author).

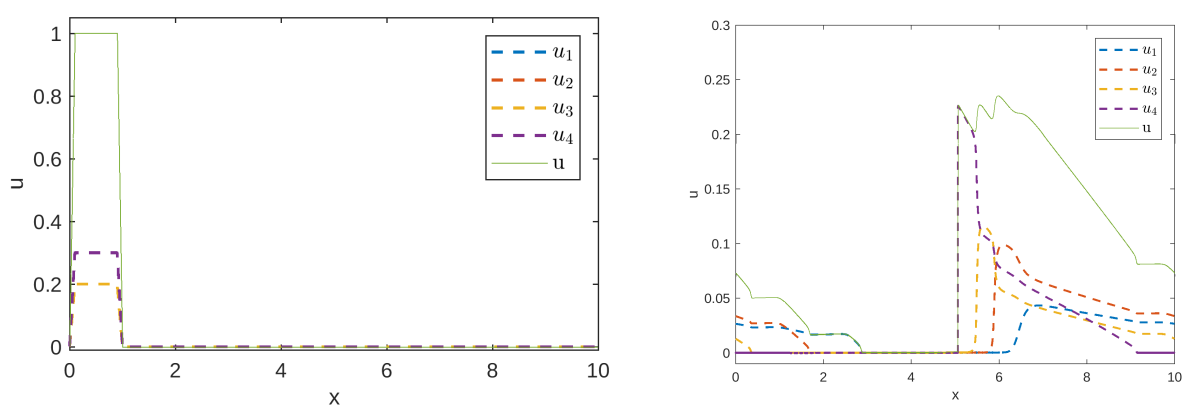


Figure 2.10: Example 2.4 (traffic model, continuous degenerate diffusion, non-smooth initial datum, $N = 4$): (left) initial condition (2.42), (right) reference numerical solution at simulated time $t = 0.1\,\mathrm{h}$ obtained by the ES scheme with $\alpha = 1.5$ and $M_{\mathrm{ref}} = 12800$ (figure produced by the author).

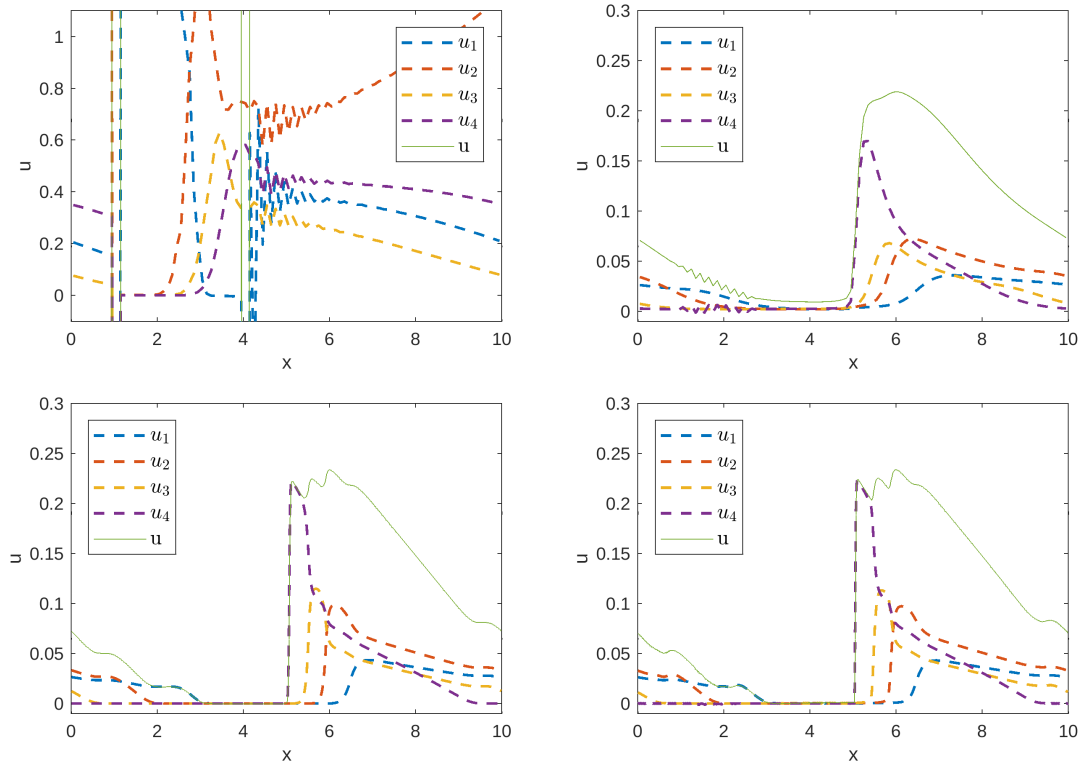


Figure 2.11: Example 2.4 (traffic model, continuous degenerate diffusion, non-smooth initial datum, $N = 4$): numerical solution at simulated time $t = 0.2\,\mathrm{h}$ (top) obtained by the entropy stable scheme with $M = 100$ and (left) with zero extra viscosity, (right) with extra viscosity (2.21) with $\alpha = 8$, (bottom) with $M = 800$ and (left) with the KT scheme, (right) with the entropy stable scheme with extra viscosity (2.21) with $\alpha = 1.5$ (figure produced by the author).

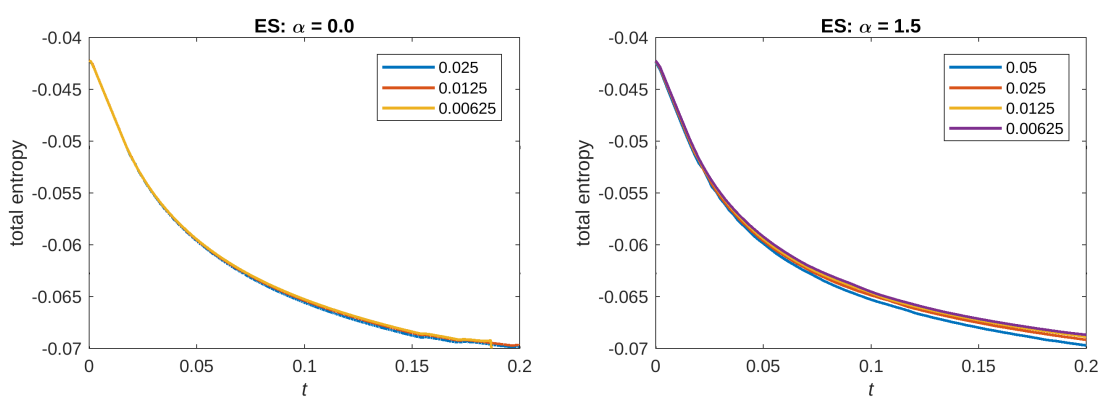


Figure 2.12: Example 2.4 (traffic model, continuous degenerate diffusion, non-smooth initial datum, $N = 4$): total entropy $\mathcal{E}_n^{\mathrm{tot}}$ of the numerical solution of the ES scheme (left)without extra viscosity and (right) with extra viscosity (2.21) with $\alpha = 1.5$ at different mesh sizes (figure produced by the author).

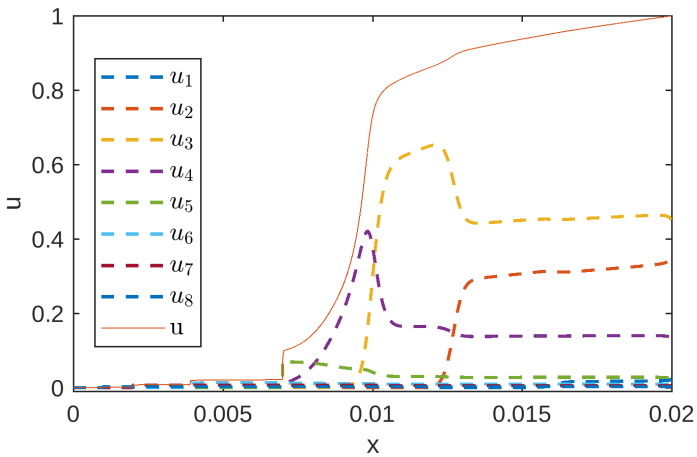


Figure 2.13: Example 2.5 (settling model, discontinuous degenerate diffusion, $N = 8$): reference solution at simulated time $t = 50$ calculated by the COMP-GLF scheme with $M_{\text{ref}} = 6400$ (figure produced by the author).

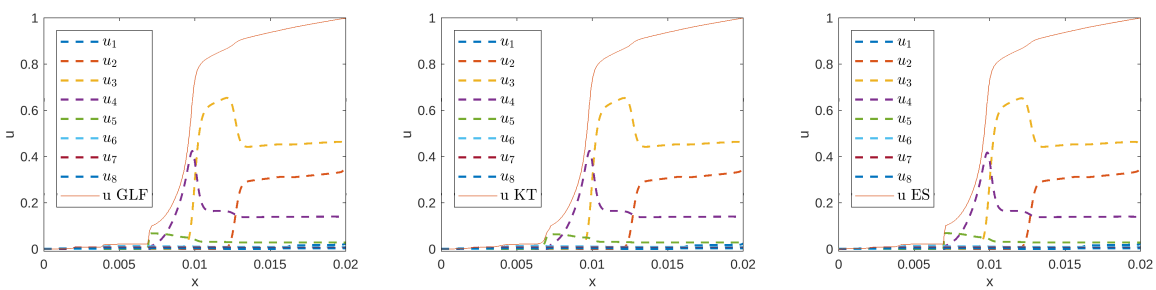


Figure 2.14: Example 2.5 (settling model, discontinuous degenerate diffusion, $N = 8$): comparison of numerical solutions computed using COMP-GLF, KT and ES ($\alpha = 1 \times 10^{-12}$) schemes, $M = 800$ (figure produced by the author).

for more information) and Kurganov-Tadmor (KT) schemes. Comparisons are made with results produced by the COMP-GLF Scheme, the reference solution is computed on a fine grid $M_{\text{ref}} = 6400$ (see Figure 2.13) and all methods are integrated in time by a SSPRK22 method with $C_{\text{CFL}} = 0.3$. Observe that the numerical errors presented in Table 2.6 seem to indicate that the methods are not converging to the same solution. Qualitative results comparing the state of the system at end time, computed with each of the three methods are displayed in Figure 2.14.

For the present problem with its zero-flux boundary condition the growth of the total entropy is bounded by inequality (2.11), whose analogy for the semi-discrete entropy stable scheme is (2.27). To study whether the latter inequality is also valid in the fully discrete case, we plot for this example (Figures 2.15 and 2.16) and the next one (Figure 2.20) the quantity

$$\tilde{\mathcal{E}}_n^{\text{tot},\prime} := \frac{\Delta x}{\Delta t} \sum_{j=1}^{M} \Big(\eta\big(\boldsymbol{u}_j^n\big) - \eta\big(\boldsymbol{u}_j^{n-1}\big) \Big) + \varphi(\boldsymbol{u}_1^n) - \varphi(\boldsymbol{u}_M^n). \tag{2.43}$$

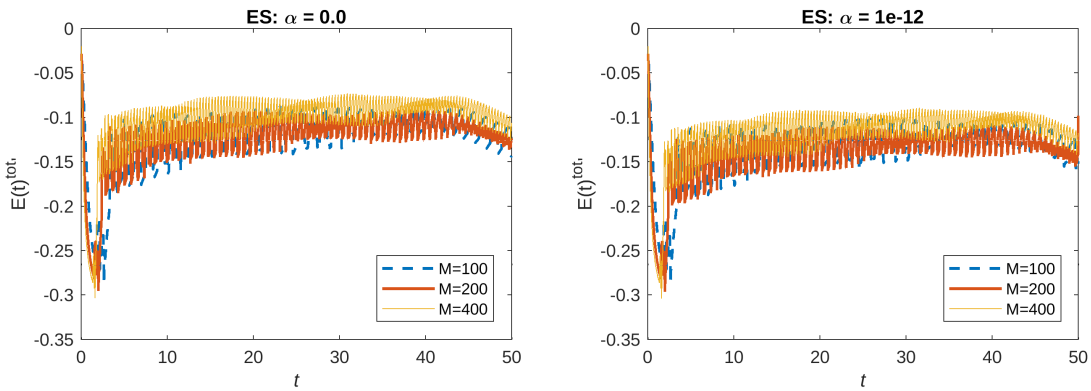


Figure 2.15: Example 2.5 (settling model, discontinuous degenerate diffusion, $N = 8$): $\tilde{\mathcal{E}}^{\text{tot},'}$ for ES scheme at different values of M (left) without extra viscosity, corresponding to $\alpha = 0$, and (right) with extra viscosity (2.21) with $\alpha = 1 \times 10^{-12}$ (figure produced by the author).

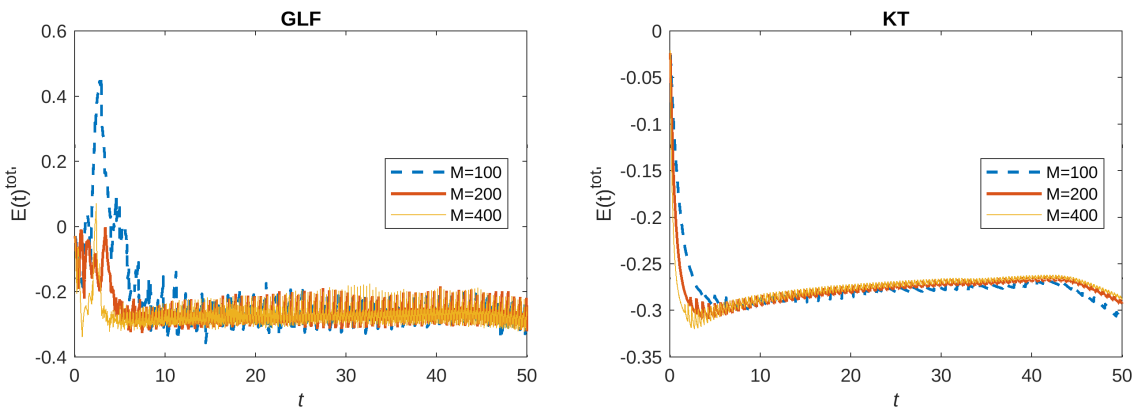


Figure 2.16: Example 2.5 (settling model, discontinuous degenerate diffusion, $N = 8$): $\tilde{\mathcal{E}}^{\text{tot},'}$ for (left) COMP-GLF and (right) KT Schemes at different values of M (figure produced by the author).

Note that, since $\varphi = \boldsymbol{v}^T \boldsymbol{f} - g$ after replacing (2.30) and (2.31), we have

$$\varphi(u) = \boldsymbol{v}(u)^{\mathrm{T}} \boldsymbol{f}(u) - g(u) = \sum_{i=1}^{N} \mathcal{V}(u) \frac{\log(u_i)}{v_i} v_i u_i - \left(\mathcal{V}(u) \sum_{i=1}^{N} u_i \log(u_i) - \tilde{\mathcal{V}}(u) \right) = \tilde{\mathcal{V}}(u).$$

It is interesting to observe(in Figure 2.16) that contrary to the other two schemes, the component-wise global Lax-Friedrichs (COMP-GLF) scheme presents problems to preserve non-positivity of the quantity (2.43) at early stages of the time evolution process.This means that the GLF scheme does not satisfy general entropy stability, a property only the ES schemes have "built in". On the other hand, as in the previous examples, the additional dissipation has little effect on the general behavior of the total entropy of the ES schemes, moreover considering that in this case (with $\alpha = 1 \times 10^{-12}$) the amount of extra viscosity added is minimal.

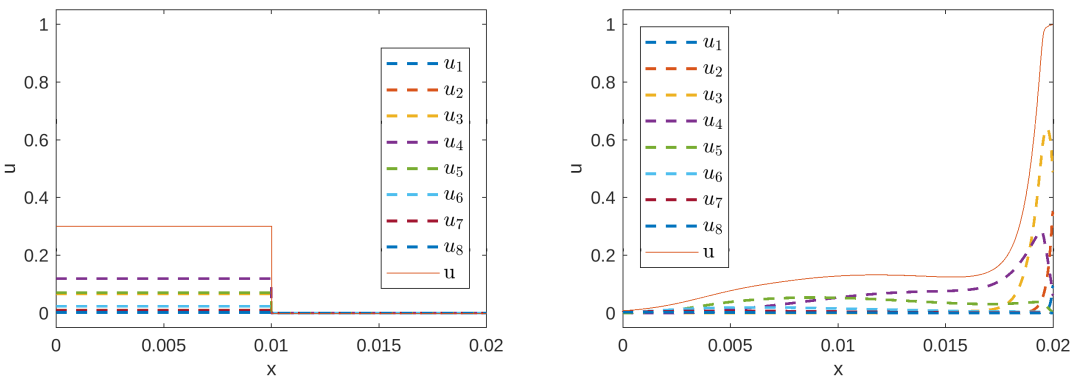


Figure 2.17: Example 2.6 (settling model, continuous diffusion, $N = 8$): (left) initial conditions and (right) reference solution computed with COMP-GLF Scheme and $M = 3200$ at $T = 200$ (figure produced by the author).

	KT		COMP-GLF		ES($\alpha = 1 \times 10^{-12}$)	
M	e_M^{tot}	θ_M	e_M^{tot}	θ_M	e_M^{tot}	θ_M
50	3.5472e-4	—	3.7022e-4	—	2.7353e-4	—
100	2.4303e-4	0.546	2.8147e-4	0.395	2.3220e-4	0.236
200	1.9425e-4	0.322	2.1254e-4	0.405	2.1183e-4	0.132
400	1.7688e-4	0.136	1.7031e-4	0.320	2.0561e-4	0.043
800	1.7687e-4	0.000	1.5841e-4	0.105	2.0409e-4	0.011

Table 2.7: Example 2.6 (settling model, continuous diffusion, $N = 8$): approximate L^1 errors (e_M^{tot}) and convergence rates (θ_M) for Example 2.6 at simulated time $t = 200$ (table produced by the author).

2.4.7 Example 2.6 (settling model, continuous diffusion, $N = 8$)

Now we consider the settling of a dispersion of 20% glycerol with a continuous diffusion function β. We suppose the initial concentration (scaled by a factor 1.5) is present only in the top half of the column as is shown in the left plot of Figure 2.17. Numerical approximations where computed using COMP-GLF, KT and ES schemes. In all cases $C_{\text{CFL}} = 0.1$ is used, and for the ES scheme a value $\alpha = 1 \times 10^{-12}$ is chosen. Qualitative results comparing results for different times are shown in Figures 2.18 and 2.19.

For this example the appearance of thin layers of particles at the bottom of the vessel poses severe difficulties for the numerical schemes to capture them. In fact for large times, a few oscillations at high concentrations start to appear which is the reason why a lower CFL constant value than the ones used on previous examples was required. However, non-smooth artefacts still could be observed in the high-concentration region at simulated time $T = 300$ in Figure 2.19, especially for the COMP-GLF scheme.

Numerical errors and convergence ratescomputed against a numerical reference solution obtained by the COMP-GLF scheme, CFL $= 0.01$ and a mesh of $M_{\text{ref}} = 3200$ can be found in Table 2.7. Here

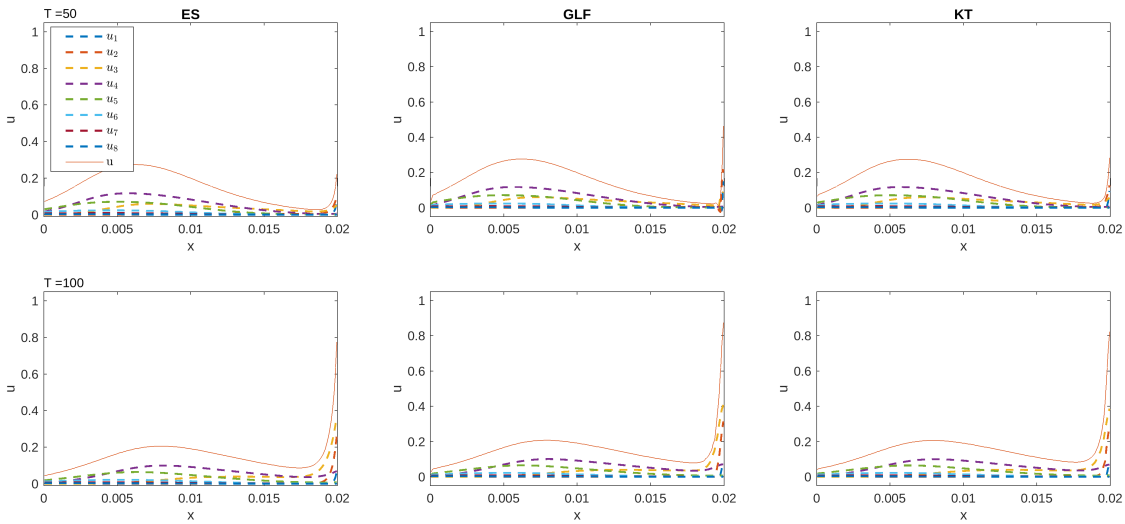


Figure 2.18: Example 2.6 (settling model, continuous diffusion, $N = 8$): numerical solutions at different times, $M = 200$ (figure produced by the author).

the stabilization of the total error for KT and ES schemes seems to suggest that the methods are not converging to the same solution. From Figure 2.20, despite an initial peak, we can see that ES scheme preserves non-positivity of the quantity (2.43), as expected.

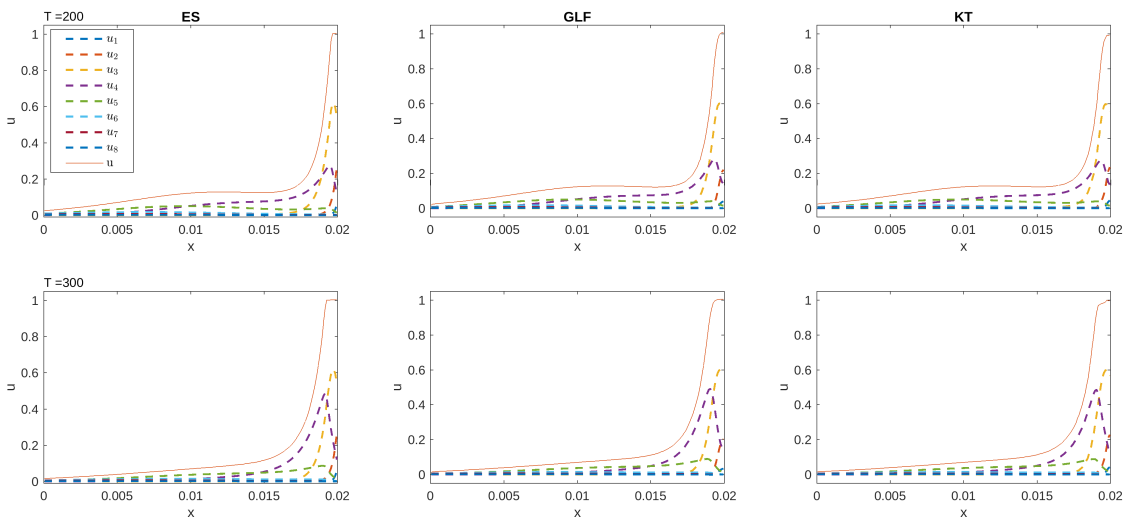


Figure 2.19: Example 2.6 (settling model, continuous diffusion, $N = 8$): numerical solutions at different times, $M = 200$ (figure produced by the author).

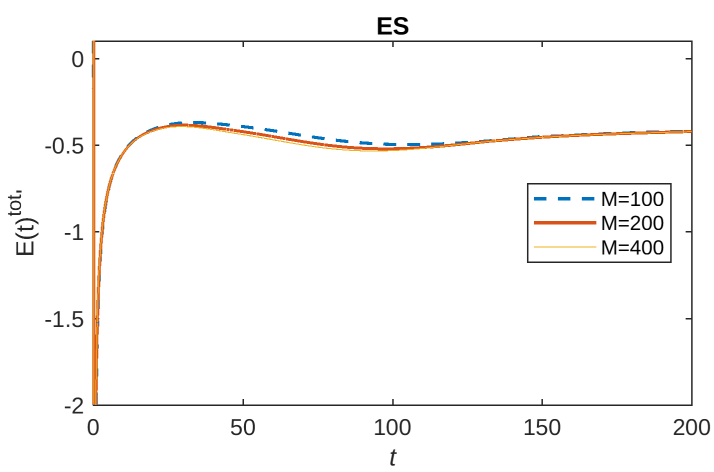


Figure 2.20: Example 2.6 (settling model, continuous diffusion, $N = 8$): $\tilde{\mathcal{E}}^{\mathrm{tot},\prime}(t)$ for ES ($\alpha = 1 \times 10^{-12}$) at different values of M (figure produced by the author).

CHAPTER 3

On $\boldsymbol{H}(\mathrm{div})$-conforming methods for double-diffusion equations in porous media

In this chapter we study a stationary Navier-Stokes-Brinkman problem coupled to a system of advection-diffusion equations, which serves as a model for so-called double-diffusive viscous flow in porous media in which both heat and a solute within the fluid phase are subject to transport and diffusion. The solvability analysis of these governing equations results as a combination of compactness arguments and fixed-point theory. In addition an $\boldsymbol{H}(\mathrm{div})$-conforming discretisation is formulated by a modification of existing methods for Brinkman flows. The well-posedness of the discrete Galerkin formulation is also discussed, and convergence properties are derived rigorously. Computational tests confirm the predicted rates of error decay and illustrate the applicability of the methods for the simulation of bacterial bioconvection and thermohaline circulation problems.

3.1 Introduction

3.1.1 Scope

Double-diffusive flows arise in the flow chemical pollutants in saturated soil, subsurface drilling and petroleum extraction, crystal growth, chemical and food processing, and numerous other applications [24, 86, 88, 119, 136, 139, 150, 166]. This class of models originates in combining heat and mass transfer interacting with flow within porous structures. One of its particularities is the formation of boundary layers due to coupled thermal and compositional mechanisms [59]. This occurs (at least in the case known as augmenting flows) since mass transfer increases the effect of buoyancy due to heat transfer. The difference in the diffusivities of the two fluid components then contributes to redirecting the flow away from the vertical density gradient [155]. Another characteristic phenomenon of double-diffusive flows is cross-diffusion [129, 139], where the flux of the solute is influenced by temperature gradients. This so-called Soret effect usually co-exists with the reciprocal phenomenon, known as the Dufour effect.

The governing equations are posed on a open and bounded spatial domain $\Omega \subseteq \mathbb{R}^d$, $d = 2$ or $d = 3$, with boundary conditions imposed on the boundary $\Gamma = \partial\Omega$ that is assumed to be Lipschitz. The

model adopts the form of the incompressible Brinkman-Navier-Stokes equations for the viscous flow of an incompressible Newtonian fluid in a porous medium, where the velocity $\boldsymbol{u}$ and the pressure p are the unknowns, coupled to a pair of advection-diffusion equations with cross-diffusion that describe the diffusion of heat and solute. Specifically, we assume that a given species (e.g. salt) has a slight solubility within this fluid, and that S denotes its concentration (i.e., weight of solute per unit weight of solution), while T is temperature, and $\vec{m} := (T, S)^{\mathrm{T}}$. The stationary behaviour of this system can be expressed as follows:

$$\begin{aligned} \mathbb{K}^{-1}\boldsymbol{u} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} - \mathbf{div}(\nu(T)\nabla\boldsymbol{u}) + \nabla p - \boldsymbol{F}(\vec{m}) &= \mathbf{0} \quad \text{in } \Omega, \\ \operatorname{div}\boldsymbol{u} &= 0 \quad \text{in } \Omega, \\ -\mathbf{div}(\mathbb{D}\nabla\vec{m}) + \operatorname{div}(\boldsymbol{u}\otimes\vec{m}) &= 0 \quad \text{in } \Omega; \\ \vec{m} = \vec{m}^{\mathrm{D}}, \quad \boldsymbol{u} &= \mathbf{0} \quad \text{on } \Gamma, \end{aligned} \tag{3.1}$$

where $\mathbb{K}(\boldsymbol{x}) > 0$ is the permeability matrix rescaled with viscosity, $\boldsymbol{F}(\vec{m})$ is a given function modeling buoyancy, $\mathbb{D}$ is the 2×2 constant matrix of the thermal conductivity and solutal diffusivity coefficients (possibly with cross-diffusion terms), and ν is a temperature-dependent viscosity function. (Precise assumptions on the model functions and problem data are stated in Section 3.2.)

It is the purpose of this chapter to propose a divergence-conforming finite element method for the double diffusive problem, considering temperature-dependent viscosity and possible cross-diffusion terms subject to the restriction of maintaining the coercivity of the diffusion operator. The formulation includes the Navier-Stokes/Brinkman flow description, which makes this model suitable for the study of flow in saturated porous media and interfaces between porous media and free flow. The numerical scheme is based on $\boldsymbol{H}(\mathrm{div})$-conforming Brezzi-Douglas-Marini (BDM) elements of order k for the velocity, discontinuous elements of order $k-1$ for the pressure, and Lagrangian finite elements of order k for temperature and the concentration of a solute. In particular this formulation produces exactly divergence-free velocity approximations, which are of particular importance in ensuring that solutions to the flow equations remain locally conservative as well as energy stable (see e.g. [62]), and moreover, the error estimates of velocity could be derived in a pressure-robust manner (see [100]). Another consequence of local conservation is that the coupled systems (in the present case, of temperature and reactive concentrations) can be written, at the discrete level, in exact divergence form.

3.1.2 Outline of the chapter

The remainder of this chapter is organized as follows. In Section 3.2 we introduce some recurrent definitions of functional spaces (Section 3.2.1), specify the assumptions on the model coefficients and problem data and state the problem in variational form (Section 3.2.2), and establish auxiliary properties of the bilinear and trilinear forms involved (Section 3.2.3). Section 3.3, which follows closely the analysis of [137], is devoted to the well-posedness analysis of the continuous problem (3.1). The basic idea consists in utilizing the correspondence of solutions $(\boldsymbol{u}, p, \boldsymbol{y})$ of the variational formulation of (3.1) with solutions $(\boldsymbol{u}, \boldsymbol{y})$ of a problem in which the pressure does not appear. The main results of Section 3.3 are Theorems 3.1 and 3.2, stating the existence and uniqueness, respectively, of a variational solution of (3.1) under appropriate assumptions. The $\boldsymbol{H}(\mathrm{div})$-conforming method for (3.1) is then introduced and analyzed in Section 3.4, which is at the core of this chapter. Specifically, in

Section 3.4.1 the method is formulated (based on an appropriate choice of the underlying discrete spaces), and in Section 3.4.2 discrete stability properties of the bilinear and trilinear forms at discrete level are provided. These properties allow us then, in Section 3.4.3, to establish existence of a discrete solution. This follows from the main result of that section, Theorem 3.3, which is based on a fixed-point argument. Finally, in Section 3.4.4 we conduct an a priori error analysis, and in particular establish orders of convergence (in terms of the meshsize) of the discrete solution to the continuous one. In Section 3.5 we present results of three different numerical experiments, namely an accuracy test for a two-dimensional manufactured solution that confirms that the experimentally observed orders of convergence are consistent with those predicted in Section 3.4.4 (Example 3.1, Section 3.5.1), an illustration of the Soret and Dufour effects in a two-dimensional porous cavity setup that validates the method against benchmark data from literature (Example 3.2, Section 3.5.2), and simulations of a non-stationary problem on a three-dimensional domain describing bioconvection of oxytactic bacteria that evaluates the extension of the proposed methods to nonlinear cross-diffusion and reaction terms in the diffusion-advection equations.

3.2 The model problem

3.2.1 Preliminaries

Let Ω be an open and bounded domain in $\mathbb{R}^d$, $d = 2, 3$ with Lipschitz boundary $\Gamma = \partial\Omega$. We will use the vector-valued Hilbert spaces

$$\begin{aligned}
\boldsymbol{H}(\mathrm{div};\Omega) &\coloneqq \big\{\boldsymbol{w} \in \boldsymbol{L}^2(\Omega) : \mathrm{div}\,\boldsymbol{w} \in L^2(\Omega)\big\},\\
\boldsymbol{H}_0(\mathrm{div};\Omega) &\coloneqq \big\{\boldsymbol{w} \in \boldsymbol{H}(\mathrm{div};\Omega) : \boldsymbol{w}\cdot\boldsymbol{n}_{\partial\Omega} = 0 \text{ on } \partial\Omega\big\},\\
\boldsymbol{H}_0(\mathrm{div}^0;\Omega) &\coloneqq \big\{\boldsymbol{w} \in \boldsymbol{H}_0(\mathrm{div};\Omega) : \mathrm{div}\,\boldsymbol{w} = 0 \text{ in } \Omega\big\},
\end{aligned}$$

were $\boldsymbol{n}_{\partial\Omega}$ denotes the outward normal on $\partial\Omega$; and we endow these spaces with the norm $\|\boldsymbol{w}\|^2_{\mathrm{div},\Omega} \coloneqq \|\boldsymbol{w}\|^2_{0,\Omega} + \|\mathrm{div}\,\boldsymbol{w}\|^2_{0,\Omega}$.

3.2.2 Assumptions and weak form of the governing equations

We assume boundary data regularity $\vec{m}^{\mathrm{D}} = (T^{\mathrm{D}}, S^{\mathrm{D}}) \in [H^{1/2}(\Gamma)]^2$, as well as Lipschitz continuity and uniform boundedness of the kinematic (temperature dependent) viscosity, i.e.,

$$\big|\nu(T_1) - \nu(T_2)\big| \le \gamma_\nu |T_1 - T_2| \quad \text{and} \quad \nu_1 \le \nu(T) \le \nu_2, \tag{3.2}$$

where γ_ν, ν_1, ν_2 are positive constants. Moreover, we assume Lipschitz continuity of the function $\boldsymbol{F}(\boldsymbol{y})$ defining the buoyancy term, i.e. there exist $\gamma_{\boldsymbol{F}}, C_{\boldsymbol{F}} > 0$ such that

$$\big|\boldsymbol{F}(\vec{m}_1) - \boldsymbol{F}(\vec{m}_2)\big| \le \gamma_{\boldsymbol{F}} |\vec{m}_1 - \vec{m}_2| \quad \text{and} \quad \big|\boldsymbol{F}(\vec{m})\big| \le C_{\boldsymbol{F}} |\vec{m}|. \tag{3.3}$$

The $d \times d$ permeability matrix $\mathbb{K}$ is assumed symmetric and uniformly positive definite, hence its inverse satisfies $\boldsymbol{v}^{\mathrm{T}}\mathbb{K}^{-1}(\boldsymbol{x})\boldsymbol{v} \ge \alpha_1 |\boldsymbol{v}|^2$ for all $\boldsymbol{v} \in \mathbb{R}^d$ and $\boldsymbol{x} \in \Omega$, for a constant $\alpha_1 > 0$. We also require $\mathbb{D}$ to be positive definite, i.e., $\vec{s}^{\mathrm{T}}\mathbb{D}\vec{s} \ge \alpha_2 |\vec{s}|^2$ for all $\vec{s} \in \mathbb{R}^2$, for a constant $\alpha_2 > 0$.

The variational formulation of problem (3.1) is obtained by testing against suitable functions and integrating by parts, and can be formulated as follows:

$$\begin{aligned}
&\text{Find } (\boldsymbol{u}, p, \vec{m}) \in \boldsymbol{H}_0^1(\Omega) \times L_0^2(\Omega) \times [H^1(\Omega)]^2 \text{ satisfying } \vec{m} = \vec{m}^{\mathrm{D}} \text{ on } \Gamma \text{ and} \\
&a(\vec{m}; \boldsymbol{u}, \boldsymbol{v}) + c(\boldsymbol{u}; \boldsymbol{u}, \boldsymbol{v}) + b(\boldsymbol{v}, p) = d(\vec{m}, \boldsymbol{v}) \quad \text{for all } \boldsymbol{v} \in \boldsymbol{H}_0^1(\Omega), \\
&b(\boldsymbol{u}, q) = 0 \quad \text{for all } q \in L_0^2(\Omega), \\
&a_{\vec{m}}(\vec{m}, \vec{s}) + c_{\vec{m}}(\boldsymbol{u}; \vec{m}, \vec{s}) = 0 \quad \text{for all } \vec{s} \in [H_0^1(\Omega)]^2,
\end{aligned} \tag{3.4}$$

where the involved forms are defined as

$$\begin{aligned}
&a(\vec{s}; \boldsymbol{u}, \boldsymbol{v}) := (\mathbb{K}^{-1}\boldsymbol{u}, \boldsymbol{v})_\Omega + \big(\nu(\vec{s})\nabla\boldsymbol{u}, \nabla\boldsymbol{v}\big)_\Omega, \quad c(\boldsymbol{w}; \boldsymbol{u}, \boldsymbol{v}) := \big((\boldsymbol{w}\cdot\nabla)\boldsymbol{u}, \boldsymbol{v}\big)_\Omega, \\
&b(\boldsymbol{v}, q) := (q, \operatorname{div}\boldsymbol{v})_\Omega, \quad d(\vec{s}, \boldsymbol{v}) := \big(\boldsymbol{F}(\vec{s}), \boldsymbol{v}\big)_\Omega, \\
&a_{\vec{m}}(\vec{m}, \vec{s}) := (\mathbb{D}\nabla\vec{m}, \nabla\vec{s})_\Omega, \quad c_{\vec{m}}(\boldsymbol{v}; \vec{m}, \vec{s}) := \big((\boldsymbol{v}\cdot\nabla)\vec{m}, \vec{s}\big)_\Omega
\end{aligned}$$

for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \boldsymbol{H}^1(\Omega)$, $q \in L^2(\Omega)$, and $\vec{m}, \vec{s} \in [H^1(\Omega)]^2$, where $\nu(\vec{s})$ is understood as the kinematic viscosity depending only on the first component of the vector $\vec{s}$.

3.2.3 Stability properties

First, note that due to (3.2)-(3.3), the following continuity properties hold for all $\boldsymbol{u}, \boldsymbol{v}, \in \boldsymbol{H}^1(\Omega)$, $q \in L^2(\Omega)$, and $\vec{m}, \vec{s} \in [H^1(\Omega)]^2$:

$$\begin{aligned}
\big|a(\cdot, \boldsymbol{u}, \boldsymbol{v})\big| &\le \max\big\{\nu_2, \|\mathbb{K}^{-1}\|_\infty\big\}\big(\|\nabla\boldsymbol{u}\|_{0,\Omega}\|\nabla\boldsymbol{v}\|_{0,\Omega} + \|\boldsymbol{u}\|_{0,\Omega}\|\boldsymbol{v}\|_{0,\Omega}\big) && \text{(3.5a)} \\
&\le C_a\|\boldsymbol{u}\|_{1,\Omega}\|\boldsymbol{v}\|_{1,\Omega}, \\
\big|a_{\vec{m}}(\vec{m}, \vec{s})\big| &\le \hat{C}_a\|\vec{m}\|_{1,\Omega}\|\vec{s}\|_{1,\Omega}, && \text{(3.5b)} \\
\big|b(\boldsymbol{v}, q)\big| &\le C_b\|\boldsymbol{v}\|_{1,\Omega}\|q\|_{0,\Omega}, && \text{(3.5c)} \\
\big|d(\vec{m}, \boldsymbol{v})\big| &\le C_{\boldsymbol{F}}\|\vec{m}\|_{1,\Omega}\|\boldsymbol{v}\|_{1,\Omega}. && \text{(3.5d)}
\end{aligned}$$

In addition, and due to the Lipschitz continuity of ν (stated in (3.2)) and Hölder's inequality, the following property holds for all $\vec{m}_1, \vec{m}_2 \in [H^1(\Omega)]^2$ and $\boldsymbol{u} \in \boldsymbol{W}^{1,\infty}(\Omega)$:

$$\big|a(\vec{m}_1; \boldsymbol{u}, \boldsymbol{v}) - a(\vec{m}_2; \boldsymbol{u}, \boldsymbol{v})\big| \le \gamma_\nu\|\boldsymbol{u}\|_{\boldsymbol{W}^{1,\infty}(\Omega)}\|\vec{m}_1 - \vec{m}_2\|_{1,\Omega}\|\boldsymbol{v}\|_{1,\Omega}. \tag{3.6}$$

On the other hand, standard Sobolev embeddings indicate that for $r \ge 1$ if $d = 2$ or $r \in [1, 6]$ if $d = 3$, there exists $C_r^* > 0$ depending only upon $|\Omega|$ and r such that $\|\boldsymbol{w}\|_{\boldsymbol{L}^r(\Omega)} \le C_r^*\|\boldsymbol{w}\|_{1,\Omega}$ for all $\boldsymbol{w} \in \boldsymbol{H}^1(\Omega)$. Then taking $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \boldsymbol{H}^1(\Omega)$ and $\vec{m}, \vec{s} \in [H^1(\Omega)]^2$, and applying this inequality along with Hölder's inequality with $\frac{1}{r} + \frac{1}{r^*} = \frac{1}{2}$, gives the following bounds

$$\begin{aligned}
\big|c(\boldsymbol{w}; \boldsymbol{u}, \boldsymbol{v})\big| &\le C_r^* C_{r^*}^*\|\boldsymbol{w}\|_{1,\Omega}\|\boldsymbol{u}\|_{1,\Omega}\|\boldsymbol{v}\|_{1,\Omega} = C_v\|\boldsymbol{w}\|_{1,\Omega}\|\boldsymbol{u}\|_{1,\Omega}\|\boldsymbol{v}\|_{1,\Omega}, \\
\big|c_{\vec{m}}(\boldsymbol{w}; \vec{m}, \vec{s})\big| &\le C_6^*\|\boldsymbol{w}\|_{1,\Omega}\|\vec{m}\|_{1,\Omega}\|\vec{s}\|_{[L^3(\Omega)]^2} = \bar{C}_v\|\boldsymbol{w}\|_{1,\Omega}\|\vec{m}\|_{1,\Omega}\|\vec{s}\|_{[L^3(\Omega)]^2}, \\
\big|c_{\vec{m}}(\boldsymbol{w}; \vec{m}, \vec{s})\big| &\le C_6^* C_3^*\|\boldsymbol{w}\|_{1,\Omega}\|\vec{m}\|_{1,\Omega}\|\vec{s}\|_{1,\Omega} = \hat{C}_v\|\boldsymbol{w}\|_{1,\Omega}\|\vec{m}\|_{1,\Omega}\|\vec{s}\|_{1,\Omega}.
\end{aligned} \tag{3.7}$$

Next, Poincaré's inequality together with the properties stated in Section 3.2.2 implies that the bilinear forms $a(\cdot; \cdot, \cdot)$ (for a fixed temperature), and $a_{\vec{m}}(\cdot, \cdot)$ are coercive, that is

$$a(\cdot; \boldsymbol{v}, \boldsymbol{v}) \ge \min\{\nu_1, \alpha_1\}\big(\|\nabla\boldsymbol{v}\|_{0,\Omega}^2 + \|\boldsymbol{v}\|_{0,\Omega}^2\big) \ge \alpha_a\|\boldsymbol{v}\|_{1,\Omega}^2 \quad \text{for all } \boldsymbol{v} \in \boldsymbol{H}_0^1(\Omega), \tag{3.8a}$$

$$a_{\vec{m}}(\vec{s}, \vec{s}) \ge \alpha_2|\vec{s}|_{1,\Omega}^2 \ge \hat{\alpha}_a\|\vec{s}\|_{1,\Omega}^2 \quad \text{for all } \vec{s} \in [H_0^1(\Omega)]^2. \tag{3.8b}$$

Using the definition and characterisation of the kernel $\boldsymbol{Z}$ of $b(\cdot,\cdot)$, namely

$$\boldsymbol{Z} := \left\{ \boldsymbol{v} \in \boldsymbol{H}_0^1(\Omega) : b(\boldsymbol{v}, q) = 0 \ \forall q \in L_0^2(\Omega) \right\} = \left\{ \boldsymbol{v} \in \boldsymbol{H}_0^1(\Omega) : \operatorname{div} \boldsymbol{v} = 0 \text{ in } \Omega \right\},$$

and using integration by parts we can readily observe that

$$c(\boldsymbol{w}; \boldsymbol{v}, \boldsymbol{v}) = 0 \text{ and } c_{\vec{m}}(\boldsymbol{w}; \vec{s}, \vec{s}) = 0 \text{ for all } \boldsymbol{w} \in \boldsymbol{Z}, \boldsymbol{v} \in \boldsymbol{H}^1(\Omega), \vec{s} \in [H^1(\Omega)]^2. \tag{3.9}$$

Remark 3.1. *Note that* (3.8a) *together with* (3.9) *implies the* $\boldsymbol{H}_0^1(\Omega)$*-ellipticity of the bilinear form* $a(\vec{m}, \cdot, \cdot) + c(\boldsymbol{w}, \cdot, \cdot) : \boldsymbol{H}_0^1(\Omega) \times \boldsymbol{H}_0^1(\Omega) \to \mathbb{R}$ *for any given* $\vec{m} \in [H^1(\Omega)]^2$ *and* $\boldsymbol{w} \in \boldsymbol{Z}$*. A similar result holds for the bilinear form* $\vec{a}_m(\cdot,\cdot) + c_{\vec{m}}(\boldsymbol{w}, \cdot, \cdot) : [H_0^1(\Omega)]^2 \times [H_0^1(\Omega)]^2 \to \mathbb{R}$

Moreover, the bilinear form $b(\cdot,\cdot)$ satisfies an inf-sup condition:

$$\sup_{\boldsymbol{v} \in \boldsymbol{H}_0^1(\Omega) \setminus \{\boldsymbol{0}\}} \frac{b(\boldsymbol{v}, q)}{\|\boldsymbol{v}\|_{1,\Omega}} \geq \beta \|q\|_{0,\Omega} \quad \text{for all } q \in L_0^2(\Omega)$$

(see [157] for this well-known property). Finally, for $\boldsymbol{u} \in \boldsymbol{W}^{1,\infty}(\Omega)$ and $\vec{s} \in [H^1(\Omega) \cap L^\infty(\Omega)]^2$ there exists an embedding constant $C_\infty > 0$ such that

$$\|\boldsymbol{u}\|_{1,\Omega} \leq C_\infty \|\boldsymbol{u}\|_{\boldsymbol{W}^{1,\infty}(\Omega)} \quad \text{and} \quad \|\vec{s}\|_{[L^3(\Omega)]^2} \leq C_\infty \|\vec{s}\|_{[L^\infty(\Omega)]^2}. \tag{3.10}$$

3.3 Well-posedness analysis of the continuous problem

We start by stating a well-known equivalence result (see [41, Chapter II, Theorem 1.1], [82, Chapter I, section 4]), adapted to the context of our problem.

Lemma 3.1. *If* $(\boldsymbol{u}, p, \vec{m}) \in \boldsymbol{H}_0^1(\Omega) \times L_0^2(\Omega) \times [H^1(\Omega)]^2$ *solves* (3.4)*, then* $(\boldsymbol{u}, \vec{m}) \in \boldsymbol{Z} \times [H^1(\Omega)]^2$ *satisfies* $\vec{m}|_\Gamma = \vec{m}^{\mathrm{D}}$ *and*

$$\begin{aligned} &a(\vec{m}; \boldsymbol{u}, \boldsymbol{v}) + c(\boldsymbol{u}; \boldsymbol{u}, \boldsymbol{v}) - d(\vec{m}, \boldsymbol{v}) = 0 \quad \text{for all } \boldsymbol{v} \in \boldsymbol{Z}, \\ &a_{\vec{m}}(\vec{m}, \vec{s}) + c_{\vec{m}}(\boldsymbol{u}; \vec{m}, \vec{s}) = 0 \quad \text{for all } \vec{s} \in [H_0^1(\Omega)]^2. \end{aligned} \tag{3.11}$$

Conversely, if $(\boldsymbol{u}, \vec{m}) \in \boldsymbol{Z} \times [H^1(\Omega)]^2$ *is a solution of the reduced problem* (3.11)*, then there exists* $p \in L_0^2(\Omega)$ *such that* $(\boldsymbol{u}, p, \vec{m})$ *is a solution of* (3.4)*.*

In order to deal with the non-homogeneous Dirichlet data appearing in the thermal energy and concentration equation, we utilise a lifting argument adapted from [137]. We write $\vec{m}$ as $\vec{m} = \vec{m}_0 + \vec{m}_1$, where $\vec{m}_0 \in [H_0^1(\Omega)]^2$ and $\vec{m}_1$ is such that

$$\vec{m}_1 \in [H^1(\Omega)]^2 \text{ with } \vec{m}_1|_\Gamma = \vec{m}^{\mathrm{D}}. \tag{3.12}$$

Lemma 3.2. *If* $\vec{s}^{\mathrm{D}} \in [H^{1/2}(\Gamma)]^2$*, then for any* $\varepsilon > 0$ *and* $1 \leq r \leq 6$ *if* $d = 3$ *or any* $r \geq 1$ *if* $d = 2$*, there exists an extension* $\vec{s}_1 \in [H^1(\Omega)]^2$ *of* $\vec{s}^{\mathrm{D}}$ *with* $\|\vec{s}_1\|_{[L^r(\Omega)]^2} \leq \varepsilon$*.*

Proof. It follows similarly as for its scalar counterpart, proven in [125, Lemma 4.1]. □

Lemma 3.3. *Let $(\boldsymbol{u}, \vec{m})$ be a solution to (3.11). Then there exist positive constants $\tilde{C}_{\boldsymbol{u}}, \tilde{C}_{\vec{m}}$ such that $\|\boldsymbol{u}\|_{1,\Omega} \leq \tilde{C}_{\boldsymbol{u}} \|\vec{m}_1\|_{1,\Omega}$ and $\|\vec{m}_0\|_{1,\Omega} \leq \tilde{C}_{\vec{m}} \|\vec{m}_1\|_{1,\Omega}$.*

Proof. If one takes $\boldsymbol{v} = \boldsymbol{u}$ and $\vec{s} = \vec{m}_0$ in (3.11), then we can assert that

$$a(\vec{m}_0 + \vec{m}_1; \boldsymbol{u}, \boldsymbol{u}) + c(\boldsymbol{u}; \boldsymbol{u}, \boldsymbol{u}) - d(\vec{m}, \boldsymbol{u}) = 0,$$
$$a_{\vec{m}}(\vec{m}_0 + \vec{m}_1, \vec{m}_0) + c_{\vec{m}}(\boldsymbol{u}; \vec{m}_0 + \vec{m}_1, \vec{m}_0) = 0.$$

Using Remark 3.1, conditions (3.3), (3.5d), and Hölder's inequality, yields the estimate

$$\alpha_a \|\boldsymbol{u}\|_{1,\Omega}^2 \leq C_{\boldsymbol{F}} \big(\|\vec{m}_0\|_{1,\Omega} + \|\vec{m}_1\|_{1,\Omega}\big) \|\boldsymbol{u}\|_{1,\Omega}. \tag{3.13}$$

Similarly as above, from (3.8b), (3.9), (3.5b) and (3.7) we can derive the relation

$$\hat{\alpha}_a \|\vec{m}_0\|_{1,\Omega}^2 \leq \hat{C}_a \|\vec{m}_1\|_{1,\Omega} \|\vec{m}_0\|_{1,\Omega} + \bar{C}_v \|\boldsymbol{u}\|_{1,\Omega} \|\vec{m}_1\|_{[L^3(\Omega)]^2} \|\vec{m}_0\|_{1,\Omega}. \tag{3.14}$$

Then, substituting (3.14) back into (3.13), we obtain

$$\|\boldsymbol{u}\|_{1,\Omega} \leq \frac{C_{\boldsymbol{F}}}{\alpha_a} \left(\frac{\hat{C}_a + \hat{\alpha}_a}{\hat{\alpha}_a} \|\vec{m}_1\|_{1,\Omega} + \frac{\bar{C}_v}{\hat{\alpha}_a} \|\boldsymbol{u}\|_{1,\Omega} \|\vec{m}_1\|_{[L^3(\Omega)]^2} \right),$$

which in turn implies that

$$\|\boldsymbol{u}\|_{1,\Omega} \left(1 - \frac{\bar{C}_v}{\hat{\alpha}_a} \|\vec{m}_1\|_{[L^3(\Omega)]^2} \right) \leq \frac{C_{\boldsymbol{F}} (\hat{C}_a + \hat{\alpha}_a)}{\alpha_a \hat{\alpha}_a} (\|\vec{m}_1\|_{1,\Omega}).$$

In view of Lemma 3.2, we may assume that $\frac{\bar{C}_v}{\hat{\alpha}_a} \|\vec{m}_1\|_{[L^3(\Omega)]^2} \leq 1/2$. Then we have

$$\|\boldsymbol{u}\|_{1,\Omega} \leq \frac{2 C_{\boldsymbol{F}} (\hat{C}_a + \hat{\alpha}_a)}{\alpha_a \hat{\alpha}_a} \|\vec{m}_1\|_{1,\Omega}. \tag{3.15}$$

Inserting (3.15) into (3.14), we are then left with

$$\begin{aligned} |\vec{m}_0|_{1,\Omega} &\leq \frac{C_a}{\hat{\alpha}_a} \|\vec{m}_1\|_{1,\Omega} + \frac{2 \bar{C}_v C_{\boldsymbol{F}} (\hat{C}_a + \hat{\alpha}_a)}{\alpha_a \hat{\alpha}_a} \|\vec{m}_1\|_{[L^3(\Omega)]^2} \|\vec{m}_1\|_{1,\Omega} \\ &\leq \left(\frac{C_a}{\hat{\alpha}_a} + \frac{C_{\boldsymbol{F}} (\hat{C}_a + \hat{\alpha}_a)}{\alpha_a} \right) \|\vec{m}_1\|_{1,\Omega}. \end{aligned}$$

□

Theorem 3.1. *Assume that the conditions of Section 3.2.2 hold. Then there is a lifting $\vec{m}_1 \in [H^1(\Omega)]^2$ of $\vec{m}^{\mathrm{D}} \in [H^{1/2}(\Gamma)]^2$ satisfying (3.12) and such that problem (3.11) has a solution $(\boldsymbol{u}, \vec{m} = \vec{m}_0 + \vec{m}_1) \in \boldsymbol{H}_0^1(\Omega) \times [H^1(\Omega)]^2$. Furthermore, there exist constants $C_{\boldsymbol{u}}, C_{\vec{m}} > 0$ only depending on the stability constants of Section 3.2.3 such that $\|\boldsymbol{u}\|_{1,\Omega} \leq C_{\boldsymbol{u}} \|\vec{m}_1\|_{1,\Omega}$ and $\|\vec{m}_0\|_{1,\Omega} \leq C_{\vec{m}} \|\vec{m}_1\|_{1,\Omega}$.*

Proof. The result follows as an adequate modification of the proof in [125, Section 4], after applying Lemma 3.3 and Brouwer's fixed-point theorem. □

The assumption of additional regularity (justified for velocity in e.g. [157, Sect. 1.3], and for temperature and concentration in [64, 112, 126], for example), along with a smallness condition allows us to establish uniqueness of solution, stated as follows.

Theorem 3.2. *Let* $(\boldsymbol{u},\vec{m}) \in \big[\boldsymbol{Z}\cap \boldsymbol{W}^{1,\infty}(\Omega)\big]\times[H^1(\Omega)\cap L^\infty(\Omega)]^2$ *be a solution of the reduced problem* (3.11)*, and assume that*

$$\max\big\{\|\boldsymbol{u}\|_{\boldsymbol{W}^{1,\infty}(\Omega)}, \|\vec{m}\|_{[L^\infty(\Omega)]^2}, \gamma_{\boldsymbol{F}}\big\} \le M, \tag{3.16}$$

for a sufficiently small constant $M>0$*. Then such solution is unique.*

Proof. Let $(\boldsymbol{u},\vec{m})$, $(\tilde{\boldsymbol{u}},\tilde{\vec{m}})$ be two solutions of problem (3.11), both satisfying assumption (3.16). Subtracting the corresponding variational formulations, we have

$$a(\vec{m},\boldsymbol{u},\boldsymbol{v}) - a(\tilde{\vec{m}},\tilde{\boldsymbol{u}},\boldsymbol{v}) + c(\boldsymbol{u},\boldsymbol{u},\boldsymbol{v}) - c(\tilde{\boldsymbol{u}},\tilde{\boldsymbol{u}},\boldsymbol{v}) - \big(d(\vec{m},\boldsymbol{v}) - d(\tilde{\vec{m}},\boldsymbol{v})\big) = 0, \tag{3.17a}$$

$$a_{\vec{m}}(\vec{m},\vec{s}) - a_{\vec{m}}(\tilde{\vec{m}},\vec{s}) + c_{\vec{m}}(\boldsymbol{u};\vec{m},\vec{s}) - c_{\vec{m}}(\tilde{\boldsymbol{u}};\tilde{\vec{m}},\vec{s}) = 0 \tag{3.17b}$$

for all $\boldsymbol{v}\in\boldsymbol{Z}$, $\vec{s}\in[H_0^1(\Omega)]^2$. One next notices that in (3.17) one can write

$$\begin{aligned}
a(\vec{m},\boldsymbol{u},\boldsymbol{v}) - a(\tilde{\vec{m}},\tilde{\boldsymbol{u}},\boldsymbol{v}) &= a(\vec{m},\boldsymbol{u}-\tilde{\boldsymbol{u}},\boldsymbol{v}) + a(\vec{m},\tilde{\boldsymbol{u}},\boldsymbol{v}) - a(\tilde{\vec{m}},\tilde{\boldsymbol{u}},\boldsymbol{v}),\\
c(\boldsymbol{u},\boldsymbol{u},\boldsymbol{v}) - c(\tilde{\boldsymbol{u}},\tilde{\boldsymbol{u}},\boldsymbol{v}) &= c(\boldsymbol{u},\boldsymbol{u}-\tilde{\boldsymbol{u}},\boldsymbol{v}) + c(\boldsymbol{u},\tilde{\boldsymbol{u}},\boldsymbol{v}) - c(\tilde{\boldsymbol{u}},\tilde{\boldsymbol{u}},\boldsymbol{v}),\\
c_{\vec{m}}(\boldsymbol{u};\vec{m},\vec{s}) - c_{\vec{m}}(\tilde{\boldsymbol{u}};\tilde{\vec{m}},\vec{s}) &= c_{\vec{m}}(\boldsymbol{u};\vec{m}-\tilde{\vec{m}},\vec{s}) + c_{\vec{m}}(\boldsymbol{u};\tilde{\vec{m}},\vec{s}) - c_{\vec{m}}(\tilde{\boldsymbol{u}};\tilde{\vec{m}},\vec{s}),
\end{aligned}$$

and then we can choose as test function $\boldsymbol{v}=\boldsymbol{u}-\tilde{\boldsymbol{u}}\in\boldsymbol{Z}$, and exploit (3.9) to obtain

$$\begin{aligned}
a(\vec{m},\boldsymbol{u}-\tilde{\boldsymbol{u}},\boldsymbol{u}-\tilde{\boldsymbol{u}}) &+ \big(a(\vec{m},\tilde{\boldsymbol{u}},\boldsymbol{u}-\tilde{\boldsymbol{u}}) - a(\tilde{\vec{m}},\tilde{\boldsymbol{u}},\boldsymbol{u}-\tilde{\boldsymbol{u}})\big)\\
&+ \big(c(\boldsymbol{u};\tilde{\boldsymbol{u}},\boldsymbol{u}-\tilde{\boldsymbol{u}}) - c(\tilde{\boldsymbol{u}};\tilde{\boldsymbol{u}},\boldsymbol{u}-\tilde{\boldsymbol{u}})\big) - \big(d(\vec{m},\boldsymbol{u}-\tilde{\boldsymbol{u}}) - d(\tilde{\vec{m}},\boldsymbol{u}-\tilde{\boldsymbol{u}})\big) = 0.
\end{aligned}$$

Applying the coercivity of the bilinear form $a(\cdot,\cdot)$ in (3.8), we readily get

$$\begin{aligned}
\alpha_a\|\boldsymbol{u}-\tilde{\boldsymbol{u}}\|_{1,\Omega}^2 \le &\big|a(\vec{m},\tilde{\boldsymbol{u}},\boldsymbol{u}-\tilde{\boldsymbol{u}}) - a(\tilde{\vec{m}},\tilde{\boldsymbol{u}},\boldsymbol{u}-\tilde{\boldsymbol{u}})\big|\\
&+ \big|c(\boldsymbol{u};\tilde{\boldsymbol{u}},\boldsymbol{u}-\tilde{\boldsymbol{u}}) - c(\tilde{\boldsymbol{u}};\tilde{\boldsymbol{u}},\boldsymbol{u}-\tilde{\boldsymbol{u}})\big|\\
&+ \big|d(\vec{m},\boldsymbol{u}-\tilde{\boldsymbol{u}}) - d(\tilde{\vec{m}},\boldsymbol{u}-\tilde{\boldsymbol{u}})\big|.
\end{aligned} \tag{3.18}$$

Analogously, we can take $\vec{s}=\vec{m}-\tilde{\vec{m}}\in[H_0^1(\Omega)]^2$ in (3.17b), and employ the coercivity of the form $a_{\vec{m}}(\cdot,\cdot,\cdot)$ in (3.8), to eventually obtain

$$\hat{\alpha}_a\|\vec{m}-\tilde{\vec{m}}\|_{1,\Omega}^2 \le \big|c_{\vec{m}}(\boldsymbol{u}-\tilde{\boldsymbol{u}};\tilde{\vec{m}},\vec{m}-\tilde{\vec{m}})\big|.$$

On the other hand, from relation (3.6) and assumption (3.16) it follows that

$$\Big|a(\vec{m},\tilde{\boldsymbol{u}},\boldsymbol{u}-\tilde{\boldsymbol{u}}) - a(\tilde{\vec{m}},\tilde{\boldsymbol{u}},\boldsymbol{u}-\tilde{\boldsymbol{u}})\Big| \le \gamma_\nu M\|\vec{m}-\tilde{\vec{m}}\|_{1,\Omega}\|\boldsymbol{u}-\tilde{\boldsymbol{u}}\|_{1,\Omega}, \tag{3.19}$$

and hence replacing (3.19) in (3.18) and taking into account the continuity of the forms $c(\cdot;\cdot,\cdot)$ (stated in (3.7)) and the Lipschitz condition (3.3), we arrive at the bound

$$\begin{aligned}
\alpha_a\|\boldsymbol{u}-\tilde{\boldsymbol{u}}\|_{1,\Omega}^2 \le \gamma_\nu M\|\vec{m}-\tilde{\vec{m}}\|_{1,\Omega}\|\boldsymbol{u}-\tilde{\boldsymbol{u}}\|_{1,\Omega} &+ C_v\|\tilde{\boldsymbol{u}}\|_{1,\Omega}\|\boldsymbol{u}-\tilde{\boldsymbol{u}}\|_{1,\Omega}^2\\
&+ \gamma_{\boldsymbol{F}}\|\boldsymbol{u}-\tilde{\boldsymbol{u}}\|_{1,\Omega}\|\vec{m}-\tilde{\vec{m}}\|_{1,\Omega}.
\end{aligned}$$

Proceeding in a similar manner, we can also derive the estimate

$$\hat{\alpha}_a\|\vec{m}-\tilde{\vec{m}}\|_{1,\Omega}^2 \le \hat{C}_v\|\boldsymbol{u}-\tilde{\boldsymbol{u}}\|_{1,\Omega}\|\tilde{\vec{m}}\|_{[L^3(\Omega)]^2}\|\vec{m}-\tilde{\vec{m}}\|_{1,\Omega}.$$

Now employing (3.10) in combination with Young's inequality, we have

$$\alpha_a\|\boldsymbol{u}-\tilde{\boldsymbol{u}}\|_{1,\Omega}^2 \le M\left(\frac{\gamma_\nu}{2}+C_vC_\infty+\frac{1}{2}\right)\|\boldsymbol{u}-\tilde{\boldsymbol{u}}\|_{1,\Omega}^2+\frac{M}{2}(\gamma_\nu+1)\|\vec{m}-\tilde{\vec{m}}\|_{1,\Omega}^2,$$
$$\hat{\alpha}_a\|\vec{m}-\tilde{\vec{m}}\|_{1,\Omega}^2 \le \frac{1}{2}\bar{C}_vC_\infty M\left(\|\boldsymbol{u}-\tilde{\boldsymbol{u}}\|_{1,\Omega}^2+\|\vec{m}-\tilde{\vec{m}}\|_{1,\Omega}^2\right).$$

Adding these inequalities and defining $\tilde{C} := (1+\gamma_\nu+\bar{C}_vC_\infty)/2$, we get

$$\big(\alpha_a - M(C_vC_\infty+\tilde{C})\big)\|\boldsymbol{u}-\tilde{\boldsymbol{u}}\|_{1,\Omega}^2+(\hat{\alpha}_a-M\tilde{C})\|\vec{m}-\tilde{\vec{m}}\|_{1,\Omega}^2 \le 0,$$

and thus uniqueness holds as long as $M < \min\{\alpha_a/(C_vC_\infty+\tilde{C}), \hat{\alpha}_a/\tilde{C}\}$. □

3.4 Finite element discretisation

3.4.1 Formulation of the $\boldsymbol{H}(\mathrm{div})$-conforming method

Let us consider a family of regular partitions, denoted $\mathcal{T}_h$, of Ω into simplices K (triangles in 2D or tetrahedra in 3D) of diameter h_K. For $k \in \mathbb{N}_0$ and a mesh $\mathcal{T}_h$ on Ω, we consider the discrete spaces (see e.g. [40])

$$\begin{aligned}
\boldsymbol{V}_h &:= \left\{\, \boldsymbol{v}_h \in \boldsymbol{H}_0(\mathrm{div};\Omega) : \boldsymbol{v}_h|_K \in [\mathcal{P}_k(K)]^d \quad \forall K \in \mathcal{T}_h \,\right\},\\
\mathcal{Q}_h &:= \left\{\, q_h \in L_0^2(\Omega) : q_h|_K \in \mathcal{P}_{k-1}(K) \quad \forall K \in \mathcal{T}_h \,\right\},\\
\mathcal{M}_h &:= \left\{\, \vec{s}_h \in [C(\bar{\Omega})]^2 : \vec{s}_h|_K \in [\mathcal{P}_k(K)]^2 \quad \forall K \in \mathcal{T}_h \,\right\}, \qquad \mathcal{M}_{h,0} := \mathcal{M}_h \cap [H_0^1(\Omega)]^2,
\end{aligned}$$

which in particular satisfy $\mathrm{div}\,\boldsymbol{V}_h \subset \mathcal{Q}_h$ (cf. [109]). Here $\boldsymbol{V}_h$ is the space of divergence-conforming BDM elements. Associated with these finite-dimensional spaces, we state the following Galerkin formulation for problem (3.1):

$$\begin{aligned}
&\text{Find } (\boldsymbol{u}_h, p_h, \vec{m}_h) \in \boldsymbol{V}_h \times \mathcal{Q}_h \times \mathcal{M}_h \text{ such that } \vec{m}_h|_\Gamma = \vec{m}_h^{\mathrm{D}}\\
&\text{and for all } (\boldsymbol{v}_h, q_h, \vec{s}_h) \in \boldsymbol{V}_h \times \mathcal{Q}_h \times \mathcal{M}_{h,0},\\
&a^h(\vec{m}_h;\boldsymbol{u}_h,\boldsymbol{v}_h)+c^h(\boldsymbol{u}_h;\boldsymbol{u}_h,\boldsymbol{v}_h)+b(\boldsymbol{v}_h,p_h) = d(\vec{m}_h,\boldsymbol{v}_h),\\
&b(\boldsymbol{u}_h,q_h) = 0,\\
&a_{\vec{m}}(\vec{m}_h,\vec{s}_h)+c_{\vec{m}}(\boldsymbol{u}_h;\vec{m}_h,\vec{s}_h) = 0.
\end{aligned} \tag{3.20}$$

Here $\vec{m}_h^{\mathrm{D}} := \mathcal{I}_\Gamma \vec{m}^{\mathrm{D}}$ and $\mathcal{I}_\Gamma$ is the nodal interpolation operator defined in Section 3.4.4, the discrete versions of the trilinear forms $a(\cdot;\cdot,\cdot)$ and $c(\cdot;\cdot,\cdot)$ are defined using a symmetric interior penalty and an upwind approach, respectively (see e.g. [19, 46, 109]):

$$\begin{aligned}
&a^h(\vec{s}_h;\boldsymbol{u}_h,\boldsymbol{v}_h)\\
&:= \int_\Omega \big(\mathbb{K}^{-1}\boldsymbol{u}_h\cdot\boldsymbol{v}_h+\nu(\vec{s}_h)\nabla_h\boldsymbol{u}_h:\nabla_h\boldsymbol{v}_h\big)\\
&\quad -\sum_{e\in\mathcal{E}_h}\int_e\left(\{\!\{\nu(\vec{s}_h)\nabla_h\boldsymbol{u}_h\boldsymbol{n}_e\}\!\}\cdot[\![\boldsymbol{v}_h]\!]-\{\!\{\nu(\vec{s}_h)\nabla_h\boldsymbol{v}_h\boldsymbol{n}_e\}\!\}\cdot[\![\boldsymbol{u}_h]\!]+\frac{a_0}{h_e}\nu(\vec{s}_h)[\![\boldsymbol{u}_h]\!]\cdot[\![\boldsymbol{v}_h]\!]\right),\\
&c^h(\boldsymbol{w}_h;\boldsymbol{u}_h,\boldsymbol{v}_h) := \int_\Omega(\boldsymbol{w}_h\cdot\nabla\boldsymbol{u}_h)\cdot\boldsymbol{v}_h+\sum_{K\in\mathcal{T}_h}\int_{\partial K\setminus\Gamma}\hat{\boldsymbol{w}}_h^{\mathrm{up}}(\boldsymbol{u}_h)\cdot\boldsymbol{v}_h,
\end{aligned}$$

where the fluxes are defined as $\hat{\boldsymbol{w}}_h^{\mathrm{up}}(\boldsymbol{u}_h) := \frac{1}{2}(\boldsymbol{w}_h \cdot \boldsymbol{n}_K - |\boldsymbol{w}_h \cdot \boldsymbol{n}_K|)(\boldsymbol{u}_h^e - \boldsymbol{u}_h)$, and $\boldsymbol{u}_h^e$ is the trace of $\boldsymbol{u}$ taken from within the exterior of K. As in the continuous case, we define the discrete kernel of the bilinear form $b(\cdot,\cdot)$ as

$$\boldsymbol{Z}_h := \{\, \boldsymbol{v}_h \in \boldsymbol{V}_h : b(\boldsymbol{v}_h, q_h) = 0 \;\forall q_h \in \mathcal{Q}_h \,\} = \{\, \boldsymbol{v}_h \in \boldsymbol{V}_h : \operatorname{div} \boldsymbol{v}_h = 0 \text{ in } \Omega \,\}.$$

3.4.2 Discrete stability properties

For sake of the subsequent analysis, we introduce the following, parameter and mesh dependent broken norms

$$\begin{aligned}
\|\boldsymbol{v}\|_{*,\mathcal{T}_h}^2 &:= \sum_{K\in\mathcal{T}_h} \|\nabla \boldsymbol{v}\|_{0,K}^2 + \sum_{e\in\mathcal{E}_h} \frac{1}{h_e} \|[\![\boldsymbol{v}]\!]\|_{0,e}^2,\\
\|\boldsymbol{v}\|_{1,\mathcal{T}_h}^2 &:= \sigma \|\boldsymbol{v}\|_{0,\Omega}^2 + \nu_2 \|\boldsymbol{v}\|_{*,\mathcal{T}_h}^2 \quad \text{for all } \boldsymbol{v} \in \boldsymbol{H}^1(\mathcal{T}_h),\\
\|\boldsymbol{v}\|_{2,\mathcal{T}_h}^2 &:= \|\boldsymbol{v}\|_{1,\mathcal{T}_h}^2 + \sum_{K\in\mathcal{T}_h} h_K^2 |\boldsymbol{v}|_{2,K}^2 \quad \text{for all } \boldsymbol{v} \in \boldsymbol{H}^2(\mathcal{T}_h),
\end{aligned}$$

where $\sigma = \|\mathbb{K}^{-1}\|_{\infty,\Omega}$ and ν_2 is defined in (3.2). We also recall the broken version of the well-known Sobolev embedding result (see e.g. [83, Lemma 6.2], [102, Prop. 4.5] or [67, Th. 5.3]): for any $r > 1$ if $d = 2$ or $1 \le r \le 6$, if $d = 3$ there exists a constant $C_{\mathrm{emb}} > 0$ such that

$$\|\boldsymbol{v}\|_{\boldsymbol{L}^r(\Omega)} \le C_{\mathrm{emb}} \|\boldsymbol{v}\|_{1,\mathcal{T}_h} \qquad \text{for all } \boldsymbol{v} \in \boldsymbol{H}^1(\mathcal{T}_h). \tag{3.21}$$

Moreover, we will use the broken space

$$\boldsymbol{C}^1(\mathcal{T}_h) := \big\{ \boldsymbol{u} \in \boldsymbol{H}^1(\mathcal{T}_h) : \boldsymbol{u}|_K \in \boldsymbol{C}^1(\bar{K}), K \in \mathcal{T}_h \big\},$$

equipped with an appropriate norm $\|\boldsymbol{u}\|_{\boldsymbol{W}^{1,\infty}(\mathcal{T}_h)} := \max_{K\in\mathcal{T}_h} \|\boldsymbol{u}\|_{\boldsymbol{W}^{1,\infty}(K)}$. Finally, we will also use an augmented H^1-norm defined as

$$\|\vec{s}\|_{1,\mathcal{E}_h} := \|\vec{s}\|_{1,\Omega}^2 + \sum_{e\in\mathcal{E}_h} \frac{1}{h_e} \|\vec{s}\|_{0,e}^2 \qquad \text{for all } \vec{s} \in [H^1(\Omega)]^2.$$

Using these norms, and the local trace inequalities

$$\begin{aligned}
\|\boldsymbol{v}\|_{0,\partial K} &\le C(h_K^{-1/2} \|\boldsymbol{v}\|_{0,K} + h_K^{1/2} |\boldsymbol{v}|_{1,K}) \quad \text{for all } \boldsymbol{v} \in \boldsymbol{H}^1(K),\\
\|p\|_{0,\partial K} &\le C h_K^{-1/2} \|p\|_{0,K} \quad \text{for all } p \in \mathbb{P}_k(K),
\end{aligned}$$

we can establish continuity of the trilinear and bilinear forms involved, stated in the following lemma that can be proved following [137, Section 3.3.2] and [19, Section 4]:

Lemma 3.4. *The following properties hold:*

$$\big|a^h(\cdot, \boldsymbol{u}, \boldsymbol{v})\big| \le C \|\boldsymbol{u}\|_{2,\mathcal{T}_h} \|\boldsymbol{v}\|_{1,\mathcal{T}_h} \qquad \text{for all } \boldsymbol{u} \in \boldsymbol{H}^2(\mathcal{T}_h),\ \boldsymbol{v} \in \boldsymbol{V}_h, \tag{3.22a}$$

$$\big|a^h(\cdot, \boldsymbol{u}, \boldsymbol{v})\big| \le \tilde{C}_a \|\boldsymbol{u}\|_{1,\mathcal{T}_h} \|\boldsymbol{v}\|_{1,\mathcal{T}_h} \qquad \text{for all } \boldsymbol{u}, \boldsymbol{v} \in \boldsymbol{V}_h, \tag{3.22b}$$

$$\big|b(\boldsymbol{v}, q)\big| \le \tilde{C}_b \|\boldsymbol{v}\|_{1,\mathcal{T}_h} \|q\|_{0,\Omega} \qquad \text{for all } \boldsymbol{v} \in \boldsymbol{H}^1(\mathcal{T}_h),\ q \in L_0^2(\Omega), \tag{3.22c}$$

and for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \boldsymbol{H}^1(\mathcal{T}_h)$ *and* $\vec{s}, \vec{m} \in [H^1(\Omega)]^2$,

$$\left|d(\vec{m}, \boldsymbol{v})\right| \leq C_{\boldsymbol{F}} \|\vec{m}\|_{1,\Omega} \|\boldsymbol{v}\|_{1,\mathcal{T}_h}, \tag{3.23a}$$

$$\left|c_{\vec{m}}(\boldsymbol{w}; \vec{m}, \vec{s})\right| \leq \tilde{C}_1 \|\boldsymbol{w}\|_{1,\mathcal{T}_h} \|\vec{s}\|_{1,\Omega} \|\vec{m}\|_{1,\Omega}, \tag{3.23b}$$

$$\left|c_{\vec{m}}(\boldsymbol{w}; \vec{m}, \vec{s})\right| \leq \tilde{C}_2 \|\boldsymbol{w}\|_{1,\mathcal{T}_h} \|\vec{m}\|_{[L^3(\Omega)]^2} \|\nabla \vec{s}\|_{0,\Omega}. \tag{3.23c}$$

Moreover, for $\vec{s}_1, \vec{s}_2 \in [H^1(\Omega)]^2$, $\boldsymbol{u} \in \boldsymbol{C}^1(\mathcal{T}_h)$ and $\boldsymbol{v} \in \boldsymbol{V}_h$, there holds

$$\left|a^h(\vec{s}_1; \boldsymbol{u}, \boldsymbol{v}) - a^h(\vec{s}_2; \boldsymbol{u}, \boldsymbol{v})\right| \leq \tilde{C}_{\mathrm{Lip}} \gamma_\nu \|\vec{s}_1 - \vec{s}_2\|_{1,\mathcal{E}_h} \|\boldsymbol{u}\|_{\boldsymbol{W}^{1,\infty}(\mathcal{T}_h)} \|\boldsymbol{v}\|_{1,\mathcal{T}_h}, \tag{3.24}$$

where the constant $\tilde{C}_{\mathrm{Lip}} > 0$ is independent of h (cf. [137, Lemma 3.3]). A related result follows for $c^h(\cdot;\cdot,\cdot)$ as in [137, Lemma 3.4]. Let $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{u} \in \boldsymbol{H}^2(\mathcal{T}_h)$ and $\boldsymbol{v} \in V_h$. Then there exists $\tilde{C}_v > 0$ independently of h such that

$$\left|c^h(\boldsymbol{w}_1; \boldsymbol{u}, \boldsymbol{v}) - c^h(\boldsymbol{w}_2; \boldsymbol{u}, \boldsymbol{v})\right| \leq \tilde{C}_v \|\boldsymbol{w}_1 - \boldsymbol{w}_2\|_{1,\mathcal{T}_h} \|\boldsymbol{u}\|_{1,\mathcal{T}_h} \|\boldsymbol{v}\|_{1,\mathcal{T}_h}. \tag{3.25}$$

While the coercivity of the form $a_{\vec{m}}(\cdot,\cdot)$ in the discrete setting is readily implied by (3.8), there also holds (cf. [109, Lemma 3.2])

$$a^h(\cdot, \boldsymbol{v}, \boldsymbol{v}) \geq \tilde{\alpha}_a \|\boldsymbol{v}\|^2_{1,\mathcal{T}_h} \quad \text{for all } \boldsymbol{v} \in \boldsymbol{V}_h, \tag{3.26}$$

provided that $a_0 > 0$ is sufficiently large and independent of the meshsize.

Let $\boldsymbol{w} \in \boldsymbol{H}_0(\mathrm{div}^0; \Omega)$, then, according to [137] we can write

$$c^h(\boldsymbol{w}; \boldsymbol{u}, \boldsymbol{u}) = \frac{1}{2} \sum_{e \in \mathcal{E}_h^i} \int_e |\boldsymbol{w} \cdot \boldsymbol{n}_e| |[\![\boldsymbol{v}]\!]|^2 \geq 0 \quad \text{for all } \boldsymbol{u} \in \boldsymbol{V}_h, \tag{3.27}$$

as well as the following relation

$$c_{\vec{m}}(\boldsymbol{w}; \vec{s}_h, \vec{s}_h) = 0 \quad \text{for all } \vec{s}_h \in \mathcal{M}_h, \tag{3.28}$$

which arises from integration by parts and holds at the discrete level since the produced discrete velocities are exactly divergence free. Finally, we recall from [109] the following discrete inf-sup condition for $b(\cdot,\cdot)$, where $\tilde{\beta}$ is independent of h:

$$\sup_{\boldsymbol{v}_h \in \boldsymbol{V}_h \setminus \{\boldsymbol{0}\}} \frac{b(\boldsymbol{v}_h, q_h)}{\|\boldsymbol{v}_h\|_{1,\mathcal{T}_h}} \geq \tilde{\beta} \|q_h\|_{0,\Omega} \quad \text{for all } q_h \in \mathcal{Q}_h. \tag{3.29}$$

3.4.3 Existence of discrete solutions

Due to the discrete stability properties stated in the previous section, a discrete analogue of Lemma 3.1 holds.

Lemma 3.5. *If* $(\boldsymbol{u}_h, p_h, \vec{m}_h) \in \boldsymbol{V}_h \times \mathcal{Q}_h \times \mathcal{M}_h$ *is a solution of* (3.20)*, then* $\boldsymbol{u}_h \in \boldsymbol{Z}_h$*, and* $(\boldsymbol{u}_h, \vec{m}_h)$ *is a solution of the discrete reduced problem*

$$\begin{aligned} &a^h(\vec{m}_h; \boldsymbol{u}_h, \boldsymbol{v}) + c^h(\boldsymbol{u}_h; \boldsymbol{u}_h, \boldsymbol{v}) - d(\vec{m}_h, \boldsymbol{v}) = 0, \\ &a_{\vec{m}}(\vec{m}_h, \vec{s}) + c_{\vec{m}}(\boldsymbol{u}_h; \vec{m}_h, \vec{s}) = 0 \quad \text{for all } (\boldsymbol{v}, \vec{s}) \in \boldsymbol{Z}_h \times \mathcal{M}_{h,0}. \end{aligned} \tag{3.30}$$

Conversely, if $(\boldsymbol{u}_h, \vec{m}_h) \in \boldsymbol{Z}_h \times \mathcal{M}_{h,0}$ *is a solution of* (3.30)*, then there exists a unique pressure* $p_h \in \mathcal{Q}_h$ *such that* $(\boldsymbol{u}_h, p_h, \vec{m}_h)$ *is a solution to* (3.20)*.*

As in the continuous case, we also perform a boundary lifting of $\vec{m}_h$ by setting $\vec{m}_h = \vec{m}_{h,0} + \vec{m}_{h,1}$ with $\vec{m}_{h,0} \in \mathcal{M}_{h,0}$, and

$$\vec{m}_{h,1} \in \mathcal{M}_h, \quad \vec{m}_{h,1}|_\Gamma = \vec{m}_h^{\mathrm{D}}. \tag{3.31}$$

Lemma 3.6. *Let $(\boldsymbol{u}_h, \vec{m}_h)$ be a solution of (3.30) with $\vec{m}_h = \vec{m}_{h,0} + \vec{m}_{h,1}$ as in (3.31). Assume that*

$$C_{\mathrm{dep}} \|\vec{m}_{h,1}\|_{[L^3(\Omega)]^2} \leq \frac{1}{2}, \quad \text{where} \quad C_{\mathrm{dep}} = \frac{\tilde{C}_{\boldsymbol{F}} \tilde{C}_2}{\tilde{\alpha}_a \hat{\alpha}_a}. \tag{3.32}$$

Then there exist constants $\tilde{C}_{\boldsymbol{u}}, \tilde{C}_{\vec{m}} > 0$ only depending on the stability constants from Section 3.4.2, such that

$$\|\boldsymbol{u}_h\|_{1,\mathcal{T}_h} \leq \tilde{C}_u \|\vec{m}_{h,1}\|_{1,\Omega} \quad \text{and} \quad \|\vec{m}_h\|_{1,\Omega} \leq \tilde{C}_{\vec{m}} \|\vec{m}_{h,1}\|_{1,\Omega}. \tag{3.33}$$

Proof. We choose $(\boldsymbol{v}, \vec{s}) = (\boldsymbol{u}_h, \vec{m}_{h,0})$ in (3.30) and use (3.27)–(3.28) to obtain

$$a^h(\vec{m}_h; \boldsymbol{u}_h, \boldsymbol{u}_h) = d(\vec{m}_h, \boldsymbol{u}_h), \quad a_{\vec{m}}(\vec{m}_{h,0}, \vec{m}_{h,0}) + a_{\vec{m}}(\vec{m}_{h,1}, \vec{m}_{h,0}) = -c_{\vec{m}}(\boldsymbol{u}_h; \vec{m}_{h,1}, \vec{m}_{h,0}).$$

Invoking the coercivity of the forms $a_h(\cdot;\cdot,\cdot)$ and $a_{\vec{m}}(\cdot,\cdot)$ in (3.26), (3.8b) and the boundedness of $c_{\vec{m}}(\cdot;\cdot,\cdot)$, $d(\cdot,\cdot)$ stated in (3.23c), (3.23a), we have

$$\tilde{\alpha}_a \|\boldsymbol{u}_h\|_{1,\mathcal{T}_h} \leq \tilde{C}_{\boldsymbol{F}} \big(\|\vec{m}_{h,0}\|_{1,\Omega} + \|\vec{m}_{h,1}\|_{1,\Omega} \big), \tag{3.34a}$$

$$\hat{\alpha}_a \|\vec{m}_{h,0}\|_{1,\Omega} \leq \hat{C}_a \|\vec{m}_{h,1}\|_{1,\Omega} + \tilde{C}_2 \|\vec{m}_{h,1}\|_{[L^3(\Omega)]^2} \|\boldsymbol{u}\|_{1,\mathcal{T}_h}. \tag{3.34b}$$

Substituting equation (3.34b) into (3.34a) then leads to

$$\tilde{\alpha}_a \|\boldsymbol{u}_h\|_{1,\mathcal{T}_h} \leq \tilde{C}_{\boldsymbol{F}} \left(\|\vec{m}_{h,1}\|_{1,\Omega} + \frac{\hat{C}_a}{\hat{\alpha}_a} \|\vec{m}_{h,1}\|_{1,\Omega} + \frac{\tilde{C}_2}{\hat{\alpha}_a} \|\vec{m}_{h,1}\|_{[L^3(\Omega)]^2} \|\boldsymbol{u}\|_{1,\mathcal{T}_h} \right),$$

$$\|\boldsymbol{u}_h\|_{1,\mathcal{T}_h} \leq C_{\mathrm{dep}} \|\vec{m}_{h,1}\|_{[L^3(\Omega)]^2} + \frac{C_{\boldsymbol{F}}}{\tilde{\alpha}_a} \left(1 + \frac{\hat{C}_a}{\hat{\alpha}_a} \right) \|\vec{m}_{h,1}\|_{1,\Omega} \leq \tilde{C}_{\boldsymbol{u}} \|\vec{m}_{h,1}\|_{1,\Omega},$$

where $\tilde{C}_{\boldsymbol{u}} = 2 \frac{C_{\boldsymbol{F}}}{\tilde{\alpha}_a} (1 + \frac{\hat{C}_a}{\hat{\alpha}_a})$. Finally, the definition of the discrete liftings and an application of triangle inequality imply that

$$\begin{aligned}
\|\vec{m}_h\|_{1,\Omega} &\leq \frac{\hat{C}_a}{\hat{\alpha}_a} \|\vec{m}_{h,1}\|_{1,\Omega} + \frac{\tilde{C}_2}{\hat{\alpha}_a} \|\vec{m}_{h,1}\|_{[L^3(\Omega)]^2} \|\boldsymbol{u}_h\|_{1,\mathcal{T}_h} + \|\vec{m}_{h,1}\|_{1,\Omega} \\
&\leq \frac{\hat{C}_a + \hat{\alpha}_a}{\hat{\alpha}_a} \|\vec{m}_{h,1}\|_{1,\Omega} + \frac{\tilde{C}_2}{\hat{\alpha}_a} \|\vec{m}_{h,1}\|_{[L^3(\Omega)]^2} 2 C_{\boldsymbol{F}} \frac{\hat{C}_a + \hat{\alpha}_a}{\tilde{\alpha}_a \hat{\alpha}_a} \|\vec{m}_{h,1}\|_{1,\Omega} \\
&\leq 2 \frac{\hat{C}_a + \hat{\alpha}_a}{\hat{\alpha}_a} \|\vec{m}_{h,1}\|_{1,\Omega} \leq \tilde{C}_{\vec{m}} \|\vec{m}_{h,1}\|_{1,\Omega}.
\end{aligned}$$

□

Theorem 3.3. *Let $\vec{m}_{h,1}$ be a discrete lifting satisfying (3.32). Then there exists a discrete solution $(\boldsymbol{u}_h, \vec{m}_h) \in \boldsymbol{Z}_h \times \mathcal{M}_h$ to (3.30) satisfying the stability bound (3.33).*

Proof. We shall make use of Brouwer's fixed-point theorem in the following form: *Let $\mathcal{K} \neq \varnothing$ be a nonempty compact convex subset of a finite dimensional normed space, and let $\mathcal{L}\colon \mathcal{K} \to \mathcal{K}$ be a continuous mapping. Then $\mathcal{L}$ has at least one fixed point in $\mathcal{K}$.* Let us then start by defining the following finite-dimensional set, where $\tilde{C}_{\boldsymbol{u}}$ is the constant from (3.33):

$$\mathcal{K}_1 = \big\{\boldsymbol{w}_h \in \boldsymbol{Z}_h \colon \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} \leq \tilde{C}_{\boldsymbol{u}} \|\vec{m}_{h,1}\|_{1,\Omega}\big\},$$

Note that $\mathcal{K}_1$ is convex and compact. Next we define the mapping $\boldsymbol{T}\colon \mathcal{K}_1 \to \mathcal{K}_1$, $\boldsymbol{w}_h \mapsto \boldsymbol{T}(\boldsymbol{w}_h) = \boldsymbol{u}_h$, where $\boldsymbol{u}_h$ is the first component of the solution of the following linearised version of problem (3.30):

$$\begin{aligned}
&\text{Find } (\boldsymbol{u}_h, \vec{m}_h) \in \boldsymbol{Z}_h \times \mathcal{M}_h \text{ such that for all } (\boldsymbol{v}, \vec{s}) \in \boldsymbol{Z}_h \times \mathcal{M}_{h,0}\colon \\
&a^h(\vec{m}_h; \boldsymbol{u}_h, \boldsymbol{v}) + c^h(\boldsymbol{w}_h; \boldsymbol{u}_h, \boldsymbol{v}) - d(\vec{m}_h, \boldsymbol{v}) = 0, \\
&a_{\vec{m}}(\vec{m}_{h,0}, \vec{s}) + c_{\vec{m}}(\boldsymbol{w}_h; \vec{m}_{h,0}, \vec{s}) = -a_{\vec{m}}(\vec{m}_{h,1}, \vec{s}) - c_{\vec{m}}(\boldsymbol{w}_h; \vec{m}_{h,1}, \vec{s}).
\end{aligned} \tag{3.35}$$

Clearly, we have the equivalence

$$\boldsymbol{T}(\boldsymbol{u}_h) = \boldsymbol{u}_h \iff (\boldsymbol{u}_h, \vec{m}_h) \in \boldsymbol{Z}_h \times \mathcal{M}_h \text{ satisfies (3.30)},$$

and owing to Lemma 3.5, we also get

$$\boldsymbol{T}(\boldsymbol{u}_h) = \boldsymbol{u}_h \iff (\boldsymbol{u}_h, \vec{m}_h, p_h) \in \boldsymbol{V}_h \times \mathcal{M}_h \times \mathcal{Q}_h \text{ satisfies (3.20)}.$$

In order to prove that the discrete fixed-point operator $\boldsymbol{T}$ is well-defined, we define the following sets, where $\tilde{C}_{\boldsymbol{u}}$ and $\tilde{C}_{\vec{m}}$ are the constants from (3.33):

$$\begin{aligned}
\mathcal{K} &\coloneqq \big\{(\boldsymbol{w}_h, \boldsymbol{\varphi}_h) \in \boldsymbol{Z}_h \times \mathcal{M}_h : \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} \leq \tilde{C}_{\boldsymbol{u}} \|\vec{m}_{h,1}\|_{1,\Omega}, \|\boldsymbol{\varphi}_h\|_{1,\Omega} \leq \tilde{C}_{\vec{m}} \|\vec{m}_{h,1}\|_{1,\Omega}\big\}, \\
\mathcal{K}_2 &\coloneqq \big\{\boldsymbol{\varphi}_h \in \mathcal{M}_h : \|\boldsymbol{\varphi}_h\|_{1,\Omega} \leq \tilde{C}_{\vec{m}} \|\vec{m}_{h,1}\|_{1,\Omega}\big\},
\end{aligned}$$

and introduce the discrete operator $\boldsymbol{R}\colon \mathcal{K} \to \mathcal{K}_1$, $(\boldsymbol{w}_h, \boldsymbol{\varphi}_h) \mapsto \boldsymbol{R}((\boldsymbol{w}_h, \boldsymbol{\varphi}_h)) = \boldsymbol{u}_h$, where $\boldsymbol{u}_h$ is the unique solution to the problem

$$\begin{aligned}
&\text{Find } \boldsymbol{u}_h \in \boldsymbol{Z}_h \text{ such that for all } \boldsymbol{v} \in \boldsymbol{Z}_h, \\
&a^h(\boldsymbol{\varphi}_h; \boldsymbol{u}_h, \boldsymbol{v}) + c^h(\boldsymbol{w}_h; \boldsymbol{u}_h, \boldsymbol{v}) - d(\boldsymbol{\varphi}_h, \boldsymbol{v}) = 0.
\end{aligned} \tag{3.36}$$

and similarly define the discrete map $\boldsymbol{S}\colon \mathcal{K}_1 \to \mathcal{K}_2$, $\boldsymbol{w}_h \mapsto \boldsymbol{S}(\boldsymbol{w}_h) = \vec{m}_h$, where $\vec{m}_h \in \mathcal{M}_h$ is the unique solution of the problem

$$\begin{aligned}
&\text{Find } \vec{m}_h \in \mathcal{M}_h \text{ such that for all } \vec{s} \in \mathcal{M}_{h,0}, \\
&a_{\vec{m}}(\vec{m}_{h,0}, \vec{s}) + c_{\vec{m}}(\boldsymbol{w}_h; \vec{m}_{h,0}, \vec{s}) = -a_{\vec{m}}(\vec{m}_{h,1}, \vec{s}) - c_{\vec{m}}(\boldsymbol{w}_h; \vec{m}_{h,1}).
\end{aligned} \tag{3.37}$$

Clearly, $\boldsymbol{T}$ can be rewritten as $\boldsymbol{T}(\boldsymbol{w}_h) = \boldsymbol{R}(\boldsymbol{w}_h, \boldsymbol{S}(\boldsymbol{w}_h))$, so to prove its well-definiteness, it suffices to show that $\boldsymbol{R}$ and $\boldsymbol{S}$ are well-defined. We begin with operator $\boldsymbol{R}$. Since for any $\boldsymbol{w}_h \in \boldsymbol{Z}_h$ and $\boldsymbol{\varphi}_h \in [H^1(\Omega)]^2$ the bilinear form $a^h(\boldsymbol{\varphi}_h; \cdot, \cdot) + c^h(\boldsymbol{w}_h, \cdot, \cdot)$ is $\boldsymbol{V}_h$-elliptic (thanks to (3.26) and (3.27)), existence and uniqueness follow from the Lax-Milgram lemma. Moreover, selecting $\boldsymbol{v} = \boldsymbol{u}_h$ in (3.36), we can appeal to the coercivity of $a^h(\cdot; \cdot, \cdot)$, the positivity of $c^h(\cdot; \cdot, \cdot)$ (3.27), condition (3.32), the bound for $d(\cdot, \cdot)$ stated in (3.23a), and the bounds within the definition of $\mathcal{K}$ to deduce that

$$\begin{aligned}
&\|\boldsymbol{u}_h\|_{1,\mathcal{T}_h}^2 \leq \frac{C_{\boldsymbol{F}}}{\tilde{\alpha}_a} \|\boldsymbol{\varphi}_h\|_{1,\Omega} \|\boldsymbol{u}_h\|_{1,\mathcal{T}_h}, \\
&\|\boldsymbol{u}_h\|_{1,\mathcal{T}_h} \leq \frac{C_{\boldsymbol{F}} \tilde{C}_{\vec{m}}}{\tilde{\alpha}_a} \|\vec{m}_{h,1}\|_{1,\Omega} \leq 2C_{\boldsymbol{F}} \frac{\tilde{\alpha}_a + \hat{C}_a}{\tilde{\alpha}_a \hat{\alpha}_a} \|\vec{m}_{h,1}\|_{1,\Omega} \leq \tilde{C}_{\boldsymbol{u}} \|\vec{m}_{h,1}\|_{1,\Omega},
\end{aligned}$$

which implies that $\boldsymbol{u}_h \in \mathcal{K}_1$.

Analogously, for $\boldsymbol{S}$ we note that thanks to (3.8b) and (3.28), the bilinear form $a_{\vec{m}}(\cdot,\cdot) + c^h(\boldsymbol{w}_h,\cdot,\cdot)$ is $\mathcal{M}_{h,0}$-elliptic, hence for a fixed discrete lifting $\vec{m}_{h,1}$, the homogeneous counterpart to the linear problem (3.37) has a unique solution. Proceeding as done above for (3.35), we use once more the coercivity of $a_{\vec{m}}(\cdot,\cdot)$ (3.5b), (3.28), condition (3.32), the bound (3.23c) for $c_{\vec{m}}(\cdot;\cdot,\cdot)$, and the definition of $\mathcal{K}_1$ to find that

$$\begin{aligned}
\|\vec{m}_{h,0}\|_{1,\Omega}^2 &\leq \frac{\tilde{C}_a}{\hat{\alpha}_a}\|\vec{m}_{h,1}\|_{1,\Omega}\|\vec{m}_{h,0}\|_{1,\Omega} + \frac{\tilde{C}_2}{\hat{\alpha}_a}\|\vec{m}_{h,1}\|_{[L^3(\Omega)]^2}\|\boldsymbol{w}_h\|_{1,\mathcal{T}_h}\|\vec{m}_{h,0}\|_{1,0},\\
\|\vec{m}_{h,0}\|_{1,\Omega} &\leq \frac{\hat{C}_a}{\hat{\alpha}_a}\|\vec{m}_{h,1}\|_{1,\Omega} + \frac{\tilde{C}_2}{\hat{\alpha}_a}\|\vec{m}_{h,1}\|_{[L^3(\Omega)]^2}2C_{\boldsymbol{F}}\frac{\tilde{\alpha}_a+\hat{C}_a}{\tilde{\alpha}_a\hat{\alpha}_a}\|\vec{m}_{h,1}\|_{1,\Omega}\\
&\leq 2\frac{\hat{\alpha}_a+\hat{C}_a}{\hat{\alpha}_a}\|\vec{m}_{h,1}\|_{1,\Omega}.
\end{aligned}$$

We then employ triangle inequality to obtain

$$\|\vec{m}_h\|_{1,\Omega} \leq 2\frac{\hat{\alpha}_a+\hat{C}_a}{\hat{\alpha}_a}\|\vec{m}_{h,1}\|_{1,\Omega} + \|\vec{m}_{h,1}\|_{1,\Omega} \leq \tilde{C}_{\vec{m}}\|\vec{m}_{h,1}\|_{1,\Omega},$$

hence establishing that $\vec{m}_h \in \mathcal{K}_2$.

In order to apply Brouwer's theorem, it remains to show that $\boldsymbol{R}$ and $\boldsymbol{S}$ are continuous operators. Let us assume we are given $(\boldsymbol{w},\boldsymbol{\varphi}) \in \mathcal{K}$ and a sequence $\{(\boldsymbol{w}_l,\boldsymbol{\varphi}_l)\}_{l\in\mathbb{N}} \subset \mathcal{K}$ such that $\|\boldsymbol{w}_l - \boldsymbol{w}\|_{1,\mathcal{T}_h} \to 0$ and $\|\boldsymbol{\varphi}_l - \boldsymbol{\varphi}\|_{1,\Omega} \to 0$ as $l \to \infty$.

From the definition of $\boldsymbol{R}$ (cf. (3.36)) the following relations can be derived:

$$\begin{aligned}
a^h(\boldsymbol{\varphi}_l;\boldsymbol{u}_l,\boldsymbol{v}) + c^h(\boldsymbol{w}_l;\boldsymbol{u}_l,\boldsymbol{v}) - d(\boldsymbol{\varphi}_l,\boldsymbol{v}) &= 0,\\
a^h(\boldsymbol{\varphi};\boldsymbol{u},\boldsymbol{v}) + c^h(\boldsymbol{w};\boldsymbol{u},\boldsymbol{v}) - d(\boldsymbol{\varphi},\boldsymbol{v}) &= 0 \quad \text{for all } \boldsymbol{v} \in \boldsymbol{Z}_h.
\end{aligned}$$

Subtracting these two systems from each other and rearranging terms yields

$$\begin{aligned}
a^h(\boldsymbol{\varphi}_l;\boldsymbol{u}-\boldsymbol{u}_l,\boldsymbol{v}) + c^h(\boldsymbol{w}_l;\boldsymbol{u}-\boldsymbol{u}_l,\boldsymbol{v}) = &-a^h(\boldsymbol{\varphi};\boldsymbol{u},\boldsymbol{v}) + a^h(\boldsymbol{\varphi}_l;\boldsymbol{u},\boldsymbol{v}) - c^h(\boldsymbol{w};\boldsymbol{u},\boldsymbol{v})\\
&+ c^h(\boldsymbol{w}_l;\boldsymbol{u},\boldsymbol{v}) + d(\boldsymbol{\varphi}_l,\boldsymbol{v}) - d(\boldsymbol{\varphi},\boldsymbol{v})
\end{aligned}$$

for all $\boldsymbol{v} \in \boldsymbol{Z}_h$. We can take in particular $\boldsymbol{v} = \boldsymbol{u} - \boldsymbol{u}_l$, and exploit the coercivity of $a^h(\cdot;\cdot,\cdot)$, the fact that $c^h(\cdot,\boldsymbol{u}-\boldsymbol{u}_l,\boldsymbol{u}-\boldsymbol{u}_l) > 0$, the boundedness of $c^h(\cdot;\cdot,\cdot)$ (3.25) in combination with the bounds for $d(\cdot,\cdot)$, as well as property (3.24), to eventually get

$$\begin{aligned}
\|\boldsymbol{u}-\boldsymbol{u}_l\|_{1,\mathcal{T}_h} \leq \frac{1}{\tilde{\alpha}_a}\big(&\tilde{C}_{\mathrm{Lip}}\gamma_\nu\|\boldsymbol{\varphi}-\boldsymbol{\varphi}_l\|_{1,\mathcal{E}_h}\|\boldsymbol{u}\|_{\boldsymbol{W}^{1,\infty}(\mathcal{T}_h)}\\
&+ \tilde{C}_v\|\boldsymbol{w}-\boldsymbol{w}_l\|_{1,\mathcal{T}_h}\|\boldsymbol{u}\|_{1,\mathcal{T}_h} + \gamma_{\boldsymbol{F}}\|\boldsymbol{\varphi}-\boldsymbol{\varphi}_l\|_{1,\Omega}\big)\\
\leq C\big(&\|\boldsymbol{\varphi}-\boldsymbol{\varphi}_l\|_{1,\mathcal{E}_h}\|\boldsymbol{u}\|_{\boldsymbol{W}^{1,\infty}(\mathcal{T}_h)} + \|\boldsymbol{w}-\boldsymbol{w}_l\|_{1,\mathcal{T}_h}\|\boldsymbol{u}\|_{1,\mathcal{T}_h} + \|\boldsymbol{\varphi}-\boldsymbol{\varphi}_l\|_{1,\Omega}\big),
\end{aligned}$$

and hence $\|\boldsymbol{u}-\boldsymbol{u}_l\|_{1,\mathcal{T}_h} \to 0$ as $l \to \infty$.

Next we consider the definition of $\boldsymbol{S}$ (3.37) and again we consider the relations

$$a_{\vec{m}}(\vec{m}_l,\vec{s}) + c_{\vec{m}}(\boldsymbol{w}_l;\vec{m}_l,\vec{s}) = 0, \quad a_{\vec{m}}(\vec{m},\vec{s}) + c_{\vec{m}}(\boldsymbol{w};\vec{m},\vec{s}) = 0 \quad \text{for all } \vec{s} \in \mathcal{M}_{h,0}.$$

Subtracting the second system from the first leads to

$$a_{\vec{m}}(\vec{m}_l - \vec{m}, \vec{s}) + c_{\vec{m}}(\boldsymbol{w}_l; \vec{m} - \vec{m}_l, \vec{s}) = -c_{\vec{m}}(\boldsymbol{w}; \vec{m}, \vec{s}) - c_{\vec{m}}(\boldsymbol{w}_l; \vec{m}, \vec{s}).$$

Now we take $\vec{s} = \vec{m} - \vec{m}_l \in \mathcal{M}_{h,0}$ and immediately note that $c_{\vec{m}}(\boldsymbol{w}_l; \vec{m} - \vec{m}_l, \vec{m} - \vec{m}_l) = 0$, thanks to (3.28). Using the coercivity of $a_{\vec{m}}(\cdot;\cdot,\cdot)$ in (3.5b) together with the boundedness of $c_{\vec{m}}(\cdot;\cdot,\cdot)$, we have

$$\|\vec{m} - \vec{m}_l\|_{1,\Omega}^2 \leq \frac{\tilde{C}_2}{\hat{\alpha}_a} \|\boldsymbol{w} - \boldsymbol{w}_l\|_{1,\mathcal{T}_h} \|\vec{m}\|_{[L^3(\Omega)]^2} \|\vec{m} - \vec{m}_l\|_{1,\Omega},$$

hence $\|\vec{m} - \vec{m}_l\|_{1,\Omega} \leq C\|\boldsymbol{w} - \boldsymbol{w}_l\|_{1,\mathcal{T}_h} \|\vec{m}\|_{[L^3(\Omega)]^2}$ and thus $\|\vec{m} - \vec{m}_l\|_{1,\Omega} \to 0$ as $l \to \infty$. □

Remark 3.2. *The application of Theorem 3.3 relies on the particular choice of the discrete boundary datum and the associated discrete liftings. Furthermore, the construction of the liftings may be computationally expensive. As in [137], we focus mainly on nodal interpolation of the boundary data, however, the discussion of other alternatives and its shortcomings in [137, Section 4.2] is still of great relevance for this work.*

Remark 3.3. *Unlike conforming discretisations, one cannot directly establish a discrete version of Theorem 3.2. In fact we were not able to control the augmented norm $\|\cdot\|_{1,\mathcal{E}_h}$ in a way reciprocal to that used to prove that theorem. However, even when uniqueness of the discrete counterpart remains an open problem, our non-exhaustive selection of numerical examples did not present any difficulties in this regard.*

3.4.4 A priori error analysis

Let us denote by $\mathcal{I}_h : [C(\bar{\Omega})]^2 \to [\mathcal{M}_h]^2$ the classical nodal interpolation operator with respect to a unisolvent set of Lagrangian interpolation nodes associated to the conforming space $\mathcal{M}_h$ and by $\mathcal{I}_\Gamma$ the restriction of $\mathcal{I}_h$ to the boundary nodes. By $\Pi_h \boldsymbol{u}$ we denote the BDM projection of $\boldsymbol{u}$, and $\mathcal{L}_h p$ is the L^2-projection of p onto $\mathcal{Q}_h$. Under adequate regularity assumptions, the following approximation properties hold (see [41, 109]):

$$\begin{aligned}
&\|\boldsymbol{u} - \Pi_h \boldsymbol{u}\|_{2,\mathcal{T}_h} \leq C(\sqrt{\sigma} h^{k+1} + \sqrt{\nu_2} h^k) \|\boldsymbol{u}\|_{k+1,\Omega}, \\
&\|\vec{m} - \mathcal{I}_h \vec{m}\|_{1,\Omega} \leq C h^k \|\vec{m}\|_{k+1,\Omega}, \\
&\|p - \mathcal{L}_h p\|_{0,\Omega} \leq C h^k \|p\|_{k,\Omega}.
\end{aligned} \tag{3.38}$$

The following preliminary trace result can be proven as in [137, Lemma 4.3].

Lemma 3.7. *Assume that $\vec{m}^{\mathrm{D}} \in [C(\bar{\Gamma})]^2$ and $\vec{m}_h^{\mathrm{D}} = \mathcal{I}_\Gamma \vec{m}^{\mathrm{D}}$. Then there is a lifting $\vec{m}_{h,1} \in \mathcal{M}_h$ such that $\vec{m}_{h,1}|_\Gamma = \vec{m}_h^{\mathrm{D}}$ and*

$$\|\vec{m}_{h,1}\|_{1,\Omega} \leq C_{\mathrm{lift}} \|\vec{m}_h^{\mathrm{D}}\|_{1/2,\Gamma}, \tag{3.39}$$

where the constant $C_{\mathrm{lift}} > 0$ is independent of the meshsize.

Remark 3.4. *If one assumes that $C_{\mathrm{dep}} C_{\mathrm{emb}} C_{\mathrm{lift}} \|\vec{m}_h^{\mathrm{D}}\|_{1/2,\Gamma} \leq 1/2$ with C_{Lip}, C_{emb}, and C_{dep} defined by (3.39), (3.21), and (3.32), respectively; then, by Theorem 3.1, there exists a solution $(\boldsymbol{u}_h, \vec{m}_h)$ to (3.11) with $\vec{m}_h = \vec{m}_{h,0} + \vec{m}_{h,1}$ satisfying the stability bounds*

$$\|\boldsymbol{u}_h\|_{1,\mathcal{T}_h} \leq \tilde{C}_{\boldsymbol{u}} C_{\mathrm{lift}} \|\vec{m}_h^{\mathrm{D}}\|_{1/2,\Gamma} \quad \text{and} \quad \|\vec{m}_h\|_{1,\mathcal{T}_h} \leq \tilde{C}_{\vec{m}} C_{\mathrm{lift}} \|\vec{m}_h^{\mathrm{D}}\|_{1/2,\Gamma}. \tag{3.40}$$

If we assume additional regularity of the exact solution $\vec{m} \in [H^2(\Omega)]^2$, then $\|\vec{m}_h^{\mathrm{D}}\|_{1/2,\Gamma}$ is bounded independently of h (cf. [137, Lemma 4.7 and Remarks 4.8 and 4.9]).

Theorem 3.4. *Let us consider liftings satisfying (3.31), and let us assume the data are sufficiently small (3.32). Let also $(\boldsymbol{u}, p, \vec{m})$, $(\boldsymbol{u}_h, p_h, \vec{m}_h)$ be the solutions of (3.4) and (3.20), respectively. Assume the condition*

$$\max\{\|\boldsymbol{u}\|_{\boldsymbol{W}^{1,\infty}(\Omega)}, \|\vec{m}\|_{[L^\infty(\Omega)]^2}, \gamma_{\boldsymbol{F}}\} \leq \min(M, \tilde{M}), \tag{3.41}$$

with M sufficiently small as specified in (3.16), and $\tilde{M}$ is bounded by the data of the problem in a way that will be made explicit in the proof. Furthermore, suppose that for $k = 1$, $\boldsymbol{u} \in \boldsymbol{C}^1(\bar{\Omega}) \cap \boldsymbol{H}^2(\Omega) \cap \boldsymbol{Z}$, $p \in H^1(\Omega)$, and $\vec{m} \in [L^\infty(\Omega)]^2 \cap [H^2(\Omega)]^2$, and that for $k \geq 2$ there holds $\boldsymbol{u} \in \boldsymbol{H}^{k+1}(\Omega) \cap \boldsymbol{Z}$, $p \in H^k(\Omega)$, and $\vec{m} \in [H^{k+1}(\Omega)]^2$. Then there exist constants $C > 0$ independent of the meshsize such that

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{2,\mathcal{T}_h} + \|\vec{m} - \vec{m}_h\|_{1,\Omega} \leq Ch^k\big(\|\boldsymbol{u}\|_{k+1,\Omega} + \|\vec{m}\|_{k+1,\Omega}\big), \tag{3.42}$$

$$\|p - p_h\|_{0,\Omega} \leq Ch^k\big(\|p\|_{k,\Omega} + \|\boldsymbol{u}\|_{k+1,\Omega} + \|\vec{m}\|_{k+1,\Omega}\big). \tag{3.43}$$

Proof. An application of integration by parts together with the assumed velocity regularity readily implies that the exact solution $(\boldsymbol{u}, p, \vec{m})$ satisfies:

$$a^h(\vec{m}; \boldsymbol{u}, \boldsymbol{v}_h) + c^h(\boldsymbol{u}; \boldsymbol{u}, \boldsymbol{v}_h) - b(\boldsymbol{v}_h, p) - d(\vec{m}, \boldsymbol{v}_h) = 0 \quad \text{for all } \boldsymbol{v}_h \in \boldsymbol{V}_h \tag{3.44}$$

(see for example [109, Lemma 3.1]). We then write a discrete analogue of (3.44) and subtract the result, leading to the following Galerkin orthogonality

$$\begin{aligned} a^h(\vec{m}; \boldsymbol{u}, \boldsymbol{v}_h) - a^h(\vec{m}_h; \boldsymbol{u}_h, \boldsymbol{v}_h) + c^h(\boldsymbol{u}; \boldsymbol{u}, \boldsymbol{v}_h) - c^h(\boldsymbol{u}_h; \boldsymbol{u}_h, \boldsymbol{v}_h) \\ -b(\boldsymbol{v}_h, p - p_h) - d(\vec{m} - \vec{m}_h, \boldsymbol{v}_h) = 0. \end{aligned} \tag{3.45}$$

In addition, it is not difficult to verify that

$$b(\boldsymbol{u} - \boldsymbol{u}_h, q_h) = 0, \quad a_{\vec{m}}(\vec{m} - \vec{m}_h, \boldsymbol{\varphi}_h) + c_{\vec{m}}(\boldsymbol{u}, \vec{m}, \boldsymbol{\varphi}_h) - c_{\vec{m}}(\boldsymbol{u}_h, \vec{m}_h, \boldsymbol{\varphi}_h) = 0 \tag{3.46}$$

for all $(q_h, \boldsymbol{\varphi}_h) \in \mathcal{Q}_h \times \mathcal{M}_{h,0}$. Let us define the errors

$$\begin{aligned} e_{\boldsymbol{u}} &:= (\boldsymbol{u} - \Pi_h \boldsymbol{u}) + (\Pi_h \boldsymbol{u} - \boldsymbol{u}_h) = E_{\boldsymbol{u}} + \xi_{\boldsymbol{u}}, \\ e_p &:= (p - \mathcal{L}_h p) + (\mathcal{L}_h p - p_h) = E_p + \xi_p, \\ e_{\vec{m}} &:= (\vec{m} - \mathcal{I}_h \vec{m}) + (\mathcal{I}_h \vec{m} - \vec{m}_h) = E_{\vec{m}} + \xi_{\vec{m}}, \end{aligned}$$

so after testing (3.45) against $\boldsymbol{v}_h = \xi_{\boldsymbol{u}}$ and rearranging terms we end up with

$$a^h(\vec{m}_h, \xi_{\boldsymbol{u}}, \xi_{\boldsymbol{u}}) + c^h(\boldsymbol{u}_h, \xi_{\boldsymbol{u}}, \xi_{\boldsymbol{u}}) = I_0 + I_1 + I_2,$$

where,

$$\begin{aligned} I_0 &:= d(\vec{m}, \xi_{\boldsymbol{u}}) - d(\vec{m}_h, \xi_{\boldsymbol{u}}), \\ I_1 &:= [a^h(\vec{m}_h, \boldsymbol{u}, \xi_{\boldsymbol{u}}) - a^h(\mathcal{I}_h \vec{m}; \boldsymbol{u}, \xi_{\boldsymbol{u}})] \\ &\quad + [a^h(\mathcal{I}_h \vec{m}; \boldsymbol{u}, \xi_{\boldsymbol{u}}) - a^h(\vec{m}; \boldsymbol{u}, \xi_{\boldsymbol{u}})] - a^h(\vec{m}_h; E_{\boldsymbol{u}}, \xi_{\boldsymbol{u}}), \\ I_2 &:= [c^h(\boldsymbol{u}_h; \boldsymbol{u}, \xi_{\boldsymbol{u}}) - c^h(\Pi_h \boldsymbol{u}; \boldsymbol{u}, \xi_{\boldsymbol{u}})] \\ &\quad + [c^h(\Pi_h \boldsymbol{u}; \boldsymbol{u}, \xi_{\boldsymbol{u}}) - c^h(\boldsymbol{u}; \boldsymbol{u}, \xi_{\boldsymbol{u}})] - c^h(\boldsymbol{u}_h; E_{\boldsymbol{u}}, \xi_{\boldsymbol{u}}). \end{aligned} \tag{3.47}$$

The rest of the proof will be devoted to finding appropriate bounds for these terms. Starting with I_0, we combine (3.3) and the triangular inequality to get

$$I_0 \leq \gamma_{\boldsymbol{F}} \|\vec{m} - \vec{m}_h\|_{1,\Omega} \|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h} \leq \gamma_{\boldsymbol{F}} \big(\|\xi_{\vec{m}}\|_{1,\Omega} + \|E_{\vec{m}}\|_{1,\Omega}\big) \|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h}.$$

Next, from (3.24), the continuity of a^h, and the small data assumption in (3.41) we get

$$I_1 \leq \tilde{C}_{\mathrm{Lip}} \gamma_\nu \tilde{M} \big(\|\xi_{\vec{m}}\|_{1,\Omega} \|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h} + \|E_{\vec{m}}\|_{1,\Omega} \|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h}\big) + \tilde{C}_a \|E_{\boldsymbol{u}}\|_{1,\mathcal{T}_h} \|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h}.$$

Moreover, from (3.25), (3.25), (3.10), (3.40), and again assumption (3.41), we obtain

$$\begin{aligned} I_2 &\leq \tilde{C}_v \|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h}^2 \|\boldsymbol{u}\|_{1,\Omega} + \tilde{C}_v \|E_{\boldsymbol{u}}\|_{1,\mathcal{T}_h} \|\boldsymbol{u}\|_{1,\mathcal{T}_h} \|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h} + \tilde{C}_v \|\boldsymbol{u}_h\|_{1,\mathcal{T}_h} \|E_{\boldsymbol{u}}\|_{1,\mathcal{T}_h} \|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h} \\ &\leq \tilde{C}_v C_\infty \tilde{M} \big(\|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h}^2 + \|E_{\boldsymbol{u}}\|_{1,\mathcal{T}_h} \|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h}\big) + \tilde{C}_v \tilde{C}_{\boldsymbol{u}} C_{\mathrm{lift}} \|\vec{m}_h^{\mathrm{D}}\|_{1/2,\Gamma} \|E_{\boldsymbol{u}}\|_{1,\mathcal{T}_h} \|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h}. \end{aligned}$$

Inserting the bounds on I_0, I_1 and I_2 into (3.47), also using the coercivity of the left-hand, thanks to (3.26)-(3.27); and applying Young's inequality we arrive at

$$\begin{aligned} \tilde{\alpha}_a \|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h}^2 \leq{}& \big((1 + C_{\mathrm{Lip}}) \tilde{M} \|E_{\vec{m}}\|_{1,\Omega} + (\tilde{C}_a + \tilde{C}_v \tilde{C}_{\boldsymbol{u}} \|\vec{m}_h^{\mathrm{D}}\|_{H^{1/2}(\Gamma)}) \|E_{\boldsymbol{u}}\|_{1,\mathcal{T}_h}\big) \|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h} \\ &+ \left(\tilde{M}\left(\frac{1 + \tilde{C}_{\mathrm{Lip}} \gamma_\nu}{2} + \tilde{C}_v C_\infty\right)\right) \|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h}^2 + \frac{1 + \tilde{C}_{\mathrm{Lip}} \gamma_\nu}{2} \tilde{M} \|\xi_{\vec{m}}\|_{1,\mathcal{T}_h}^2. \end{aligned} \tag{3.48}$$

We handle (3.46) in a similar way and take $\boldsymbol{\varphi}_h = \xi_{\vec{m}}$ as test function. This leads to

$$\begin{aligned} a_{\vec{m}}(\xi_{\vec{m}}, \xi_{\vec{m}}) + c_{\vec{m}}(\vec{m}_h; \vec{m}_h, \xi_{\vec{m}}) = {}& -a_{\vec{m}}(E_{\vec{m}}, \xi_{\vec{m}}) - c_{\vec{m}}(\xi_{\boldsymbol{u}}; \vec{m}, \xi_{\vec{m}}) \\ & - c_{\vec{m}}(E_{\boldsymbol{u}}; \vec{m}, \xi_{\vec{m}}) - c_{\vec{m}}(\boldsymbol{u}_h; E_{\vec{m}}, \xi_{\vec{m}}). \end{aligned}$$

In addition, on the left-hand side we use the coercivity of $a_{\vec{m}}$, properties (3.9), (3.23b), (3.40), the embedding (3.10), as well as assumption (3.41) to get

$$\begin{aligned} \hat{\alpha}_a \|\xi_{\vec{m}}\|_{1,\Omega}^2 \leq{}& \hat{C}_a \|E_{\vec{m}}\|_{1,\Omega} \|\xi_{\vec{m}}\|_{1,\Omega} + \tilde{C}_1 C_\infty \tilde{M} \big(\|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h} \|\xi_{\vec{m}}\|_{1,\Omega} + \|E_{\boldsymbol{u}}\|_{1,\mathcal{T}_h} \|\xi_{\vec{m}}\|_{1,\Omega}\big) \\ &+ \tilde{C}_1 \tilde{C}_v C_{\mathrm{lift}} \|\vec{m}_h^{\mathrm{D}}\|_{1/2,\Gamma} \|E_{\vec{m}}\|_{1,\Omega} \|\xi_{\vec{m}}\|_{1,\Omega}, \end{aligned}$$

and after applying Young's inequality and regrouping terms, we have

$$\begin{aligned} \hat{\alpha}_a \|\xi_{\vec{m}}\|_{1,\Omega}^2 \leq{}& \big((\hat{C}_a + \tilde{C}_1 \tilde{C}_{\boldsymbol{u}} C_{\mathrm{lift}} \|\vec{m}_h^{\mathrm{D}}\|_{1/2,\Gamma}) \|E_{\vec{m}}\|_{1,\Omega} + \tilde{C}_1 C_\infty \tilde{M} \|E_{\boldsymbol{u}}\|_{1,\mathcal{T}_h}\big) \|\xi_{\vec{m}}\|_{1,\Omega} \\ &+ \frac{1}{2} \tilde{C}_1 C_\infty \tilde{M} \big(\|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h}^2 + \|\xi_{\vec{m}}\|_{1,\Omega}^2\big). \end{aligned} \tag{3.49}$$

Adding (3.48) and (3.49) and defining $\check{C} := (1 + \tilde{C}_{\mathrm{Lip}} \gamma_\nu + \tilde{C}_1 C_\infty)/2$ we obtain

$$\begin{aligned} &\big(\tilde{\alpha}_a - \tilde{M}(\check{C} + \tilde{C}_v C_\infty)\big) \|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h}^2 + (\hat{\alpha}_a - \tilde{M}\check{C}) \|\xi_{\vec{m}}\|_{1,\Omega}^2 \\ &\leq C\big(\|E_{\vec{m}}\|_{1,\Omega} + \|E_{\boldsymbol{u}}\|_{1,\mathcal{T}_h}\big)\big(\|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h} + \|\xi_{\vec{m}}\|_{1,\mathcal{T}_h}\big). \end{aligned}$$

Hence, if we choose $\tilde{M}$ such that $\tilde{M} < \min\{\tilde{\alpha}_a/(\check{C} + \tilde{C}_v C_\infty), \hat{\alpha}_a/\check{C}\}$ (note that this constant depends only on the data of the problem), then we readily obtain $\|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h} + \|\xi_{\vec{m}}\|_{1,\Omega} \leq C(\|E_{\vec{m}}\|_{1,\Omega} + \|E_{\boldsymbol{u}}\|_{1,\mathcal{T}_h})$. Using now the approximation properties in (3.38), we straightforwardly get (3.42).

For the pressure estimate we consider the discrete inf-sup condition (3.29) as well as (3.22c). It follows that

$$\begin{aligned}\|\xi_p\|_{0,\Omega} &\leq \frac{1}{\tilde{\beta}} \sup_{\boldsymbol{v}_h \in \boldsymbol{V}_h \setminus \{\boldsymbol{0}\}} \frac{b(\boldsymbol{v}_h, \xi_p)}{\|\boldsymbol{v}_h\|_{1,\mathcal{T}_h}} \leq \frac{1}{\tilde{\beta}} \sup_{\boldsymbol{v}_h \in \boldsymbol{V}_h \setminus \{\boldsymbol{0}\}} \frac{b(\boldsymbol{v}_h, e_p)}{\|\boldsymbol{v}_h\|_{1,\mathcal{T}_h}} + \frac{1}{\tilde{\beta}} \sup_{\boldsymbol{v}_h \in \boldsymbol{V}_h \setminus \{\boldsymbol{0}\}} \frac{b(\boldsymbol{v}_h, E_p)}{\|\boldsymbol{v}_h\|_{1,\mathcal{T}_h}} \\ &\leq \frac{1}{\tilde{\beta}} \sup_{\boldsymbol{v}_h \in \boldsymbol{V}_h \setminus \{\boldsymbol{0}\}} \frac{b(\boldsymbol{v}_h, e_p)}{\|\boldsymbol{v}_h\|_{1,\mathcal{T}_h}} + \frac{\tilde{C}_b}{\tilde{\beta}} \|E_p\|_{0,\Omega}. \end{aligned} \tag{3.50}$$

Now for any $\boldsymbol{v}_h \in \boldsymbol{V}_h$, (3.45) implies the bound $b(\boldsymbol{v}_h, e_p) \leq I_3 + I_4 + I_5$, where

$$\begin{aligned} I_3 &= |d(\vec{m}, \boldsymbol{v}_h) - d(\vec{m}_h, \boldsymbol{v}_h)|, \\ I_4 &= |a^h(\vec{m}; \boldsymbol{u}, \boldsymbol{v}_h) - a^h(\vec{m}_h; \boldsymbol{u}, \boldsymbol{v}_h)| + |a^h(\vec{m}_h, e_{\boldsymbol{u}}, \boldsymbol{v}_h)|, \\ I_5 &= |c^h(\boldsymbol{u}; \boldsymbol{u}, \boldsymbol{v}_h) - c^h(\boldsymbol{u}_h; \boldsymbol{u}, \boldsymbol{v}_h)| + |c^h(\boldsymbol{u}_h; e_{\boldsymbol{u}}, \boldsymbol{v}_h)|. \end{aligned}$$

Hence we can use property (3.3) to deduce that $I_3 \leq \gamma_{\boldsymbol{F}} \|e_{\vec{m}}\|_{1,\Omega} \|\boldsymbol{v}_h\|_{1,\mathcal{T}_h}$. From (3.24), (3.22a), and assumption (3.41), it then follows that

$$\begin{aligned} I_4 &\leq \tilde{C}_{\mathrm{Lip}} \gamma_\nu \|e_{\vec{m}}\|_{1,\Omega} \|\boldsymbol{u}\|_{\boldsymbol{W}^{1,\infty}(\Omega)} \|\boldsymbol{v}_h\|_{1,\mathcal{T}_h} + C \|e_{\boldsymbol{u}}\|_{2,\mathcal{T}_h} \|\boldsymbol{v}_h\|_{1,\mathcal{T}_h}, \\ &\leq \tilde{C}_{\mathrm{Lip}} \gamma_\nu \tilde{M} \|e_{\vec{m}}\|_{1,\Omega} \|\boldsymbol{v}_h\|_{1,\mathcal{T}_h} + C \|e_{\boldsymbol{u}}\|_{2,\mathcal{T}_h} \|\boldsymbol{v}_h\|_{1,\mathcal{T}_h}. \end{aligned}$$

Now we use (3.25), (3.10), (3.40) and the bound in (3.25) to get

$$\begin{aligned} I_5 &\leq \tilde{C}_v \|\boldsymbol{u}_h\|_{1,\mathcal{T}_h} \|e_{\boldsymbol{u}}\|_{1,\mathcal{T}_h} \|\boldsymbol{v}_h\|_{1,\mathcal{T}_h} + \tilde{C}_v \|\boldsymbol{u}\|_{1,\mathcal{T}_h} \|e_{\boldsymbol{u}}\|_{1,\mathcal{T}_h} \|\boldsymbol{v}_h\|_{1,\mathcal{T}_h} \\ &\leq \tilde{C}_v C_\infty \tilde{M} \|e_{\boldsymbol{u}}\|_{2,\mathcal{T}_h} \|\boldsymbol{v}_h\|_{1,\mathcal{T}_h} + \tilde{C}_v \tilde{C}_{\boldsymbol{u}} C_{\mathrm{lift}} \|\vec{m}_h^{\mathrm{D}}\|_{H^{1/2}(\Gamma)} \|e_{\boldsymbol{u}}\|_{2,\mathcal{T}_h} \|\boldsymbol{v}_h\|_{1,\mathcal{T}_h}. \end{aligned}$$

The estimates on I_3, I_4 and I_5 therefore yield

$$\big|b(\boldsymbol{v}_h, e_p)\big| \leq C\big(\|e_{\vec{m}}\|_{1,\Omega} + \|e_{\boldsymbol{u}}\|_{2,\mathcal{T}_h}\big) \|\boldsymbol{v}_h\|_{1,\mathcal{T}_h}. \tag{3.51}$$

Hence (3.43) follows by replacing (3.51) in (3.50) and using the approximation properties (3.38). □

Notice that, thanks to the divergence-free property of the discrete velocities, the bound (3.42) confirms that the family of methods proposed here is pressure-robust (see also the discussion in [100]). This can be also observed numerically, for instance in Table 3.3 where the magnitude of the pressure errors does not affect the magnitude of the velocity errors.

3.5 Numerical tests

The following set of examples provides numerical confirmation of the convergence rates anticipated in Theorem 3.4. We further validate the proposed method by comparing our produced results against benchmark solutions found in the literature, and we present one test oriented to applications inherent to doubly-diffusive flows in porous media. The linearisation of the system of equations associated with the assembled form of (3.20) is carried out by Newton's method, setting a relative tolerance of 1E-8 on the residuals. In turn, the solution of the resulting linear systems present at each Newton step is conducted using the bi-conjugate gradient stabilised Krylov solver (BiCGStab). In the implementation

k	DoF	$\mathtt{e}_{\boldsymbol{u}}$	rate	$\mathtt{e}_p$	rate	$\mathtt{e}_T$	rate	$\mathtt{e}_S$	rate	$\lVert \mathrm{div}\, \boldsymbol{u}_h \rVert_{\infty,\Omega}$
1	195	0.6798	–	1.5670	–	0.3498	–	0.2721	–	1.33E-15
	707	0.3779	0.847	1.1370	0.563	0.1975	0.824	0.1385	0.974	4.88E-15
	2691	0.1873	1.012	0.6614	0.787	0.1019	0.954	0.0696	0.992	9.77E-15
	10499	0.0923	1.021	0.3485	0.925	0.0513	0.988	0.0348	0.998	2.13E-14
	41475	0.0459	1.007	0.1771	0.977	0.0257	0.997	0.0174	0.999	4.62E-14
2	523	0.3258	1.657	1.7741	1.243	0.1221	1.101	0.0338	1.767	9.03E-14
	1971	0.0847	1.943	0.6826	1.378	0.0326	1.905	0.0089	1.928	2.23E-13
	7651	0.0179	2.237	0.2159	1.661	0.0083	1.968	0.0023	1.979	4.82E-13
	30147	0.0038	2.238	0.0587	1.877	0.0021	1.991	0.0006	1.994	9.96E-13
	119683	0.0008	2.108	0.0151	1.964	0.0005	1.998	0.0001	1.998	2.01E-12

Table 3.1: Example 3.1 (accuracy test): experimental errors and convergence rates for the approximate solutions $\boldsymbol{u}_h$, p_h, T_h and S_h; and ℓ^∞-norm of the vector formed by the divergence of the discrete velocity computed for each discretisation. Values are displayed for the first and second order schemes for a flow regime with $\nu_2 = \sigma = 1$ (table produced by the author).

k	DoF	$\mathtt{e}_{\boldsymbol{u}}$	rate	$\mathtt{e}_p$	rate	$\mathtt{e}_T$	rate	$\mathtt{e}_S$	rate	$\lVert \mathrm{div}\, \boldsymbol{u}_h \rVert_{\infty,\Omega}$
1	195	2.1490	–	14.352	–	0.3498	–	0.2721	–	1.55E-15
	707	1.2041	0.835	10.710	0.429	0.1975	0.824	0.1385	0.974	4.00E-15
	2691	0.5958	1.015	6.3981	0.749	0.1019	0.954	0.0696	0.992	8.88E-15
	10499	0.2925	1.026	3.4170	0.904	0.0513	0.988	0.0348	0.998	2.31E-14
	41475	0.1453	1.010	1.7461	0.968	0.0257	0.997	0.0174	1.000	4.26E-14
2	523	1.0380	1.652	17.152	1.119	0.1221	1.101	0.0338	1.767	9.24E-14
	1971	0.2688	1.949	6.7861	1.338	0.0326	1.905	0.0089	1.928	2.29E-13
	7651	0.0568	2.241	2.1562	1.654	0.0083	1.968	0.0023	1.979	4.87E-13
	30147	0.0121	2.239	0.5875	1.876	0.0021	1.991	0.0006	1.994	1.01E-12
	119683	0.0028	2.108	0.1507	1.963	0.0005	1.998	0.0001	1.998	2.00E-12

Table 3.2: Example 3.1 (accuracy test): errors and convergence rates under a Stokes regime with $\nu_2 = 10, \sigma = 0$ (table produced by the author).

of the method, the normal component of the velocity is fixed in the form of an essential boundary condition, whereas its tangential component is incorporated as a natural boundary condition and imposed *à la* Nitsche (see e.g. [92]). Moreover, the condition of zero mean value for the pressure approximation is implemented using a real Lagrange multiplier. All tests were implemented using the open-source finite element library FEniCS [7].

3.5.1 Example 3.1: accuracy test

In our first computational test we examine the convergence of the Galerkin method (3.20), taking as computational domain the square $\Omega = (-1,1)^2$, and considering a sequence of uniformly refined meshes $\{\mathcal{T}_{h,l}\}_l$ of mesh size $h_l = 2^{-l}\sqrt{2}$. We take a buoyancy term of the form $\boldsymbol{F}(\vec{m}) = (T + N_r S)\boldsymbol{g}$, where N_r is the solutal to thermal buoyancy ratio; and choose an exponential form for the viscosity $\nu(T) = \nu_2 \exp(-T)$, $\boldsymbol{g} = (0,1)^T$, $\mathbb{K}^{-1} = \sigma\mathbb{I}$, $\mathbb{D} = 1000\mathbb{I}$, $a_0 = \sqrt{\sigma}10^k$. Following the approach of manufactured solutions, we prescribe boundary data and additional external forces and adequate

k	DoF	$\mathtt{e}_{\boldsymbol{u}}$	rate	$\mathtt{e}_p$	rate	$\mathtt{e}_T$	rate	$\mathtt{e}_S$	rate	$\lVert \mathrm{div}\, \boldsymbol{u}_h \rVert_{\infty,\Omega}$
1	195	5.3102	–	287.42	–	0.3498	–	0.2721	–	1.78E-15
	707	1.6182	1.715	148.01	0.958	0.1975	0.825	0.1385	0.974	4.44E-15
	2691	0.4303	1.911	72.992	1.021	0.1019	0.954	0.0696	0.993	1.07E-14
	10499	0.1324	1.701	36.721	0.991	0.0514	0.988	0.0348	0.998	2.13E-14
	41475	0.0516	1.359	18.472	0.992	0.0257	0.997	0.0174	1.000	4.26E-14
2	523	1.9250	2.483	270.41	2.175	0.1221	1.101	0.0338	1.767	9.49E-14
	1971	0.5142	1.905	51.930	2.38	0.0326	1.905	0.0089	1.928	2.27E-13
	7651	0.1364	1.914	11.504	2.175	0.0083	1.968	0.0023	1.979	4.94E-13
	30147	0.0389	1.808	3.1610	1.863	0.0021	1.991	0.0006	1.994	9.99E-13
	119683	0.0104	1.900	1.0190	1.633	0.0005	1.998	0.0001	1.998	2.03E-12

Table 3.3: Example 3.1 (accuracy test): errors and convergence rates for the approximate solutions for a Darcy regime, with $\nu_2 = 1, \sigma = 10000$ (table produced by the author).

source terms so that the closed-form solutions to (3.1) are given by the smooth functions

$$\boldsymbol{u}(x,y) = \big(\sin(\pi x)\cos(\pi y), -\cos(\pi x)\sin(\pi y)\big)^{\mathrm{T}}, \quad p(x,y) = \cos(\pi x)\exp(y),$$
$$T(x,y) = 0.5 + 0.5\cos(xy), \quad S(x,y) = 0.1 + 0.3\exp(xy).$$

Relative errors in their natural norms, along with the corresponding convergence rates computed as

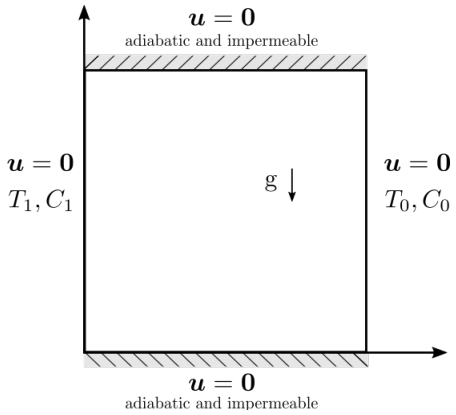
$$\mathtt{e}_{\boldsymbol{u}} = \lVert \boldsymbol{u} - \boldsymbol{u}_h \rVert_{1,\mathcal{T}_h} / \lVert \boldsymbol{u} \rVert_{1,\mathcal{T}_h}, \quad \mathtt{e}_p = \lVert p - p_h \rVert_{0,\Omega} / \lVert p \rVert_{0,\Omega}, \quad \mathtt{e}_T = \lVert T - T_h \rVert_{1,\Omega} / \lVert T \rVert_{1,\Omega},$$
$$\mathtt{e}_S = \lVert S - S_h \rVert_{1,\Omega} / \lVert S \rVert_{1,\Omega}, \quad \mathtt{rate} = \log(e_{(\cdot)}/\tilde{e}_{(\cdot)})[\log(h/\tilde{h})]^{-1},$$

where $e, \tilde{e}$ denote errors generated on two consecutive meshes of sizes h and $\tilde{h}$, respectively; are listed in Table 3.1 for $k = 1, 2$, where the model constants are chosen as stated above. We can observe that the total error is dominated by the pressure approximation, and that the discrete velocities are divergence free. The tabulated values also indicate an optimal $\mathcal{O}(h^k)$ convergence, consistently with the theoretical bounds stated in Theorem 3.4. We also conduct two additional series of accuracy tests focusing on the cases where the viscosity and permeability coefficients scale differently, changing from Stokes to Darcy regimes. These values are collected in Tables 3.2 and 3.3, respectively. Apart from an increase of the pressure error, we can see that the experimental rates of convergence remain close to the optimal behaviour.

3.5.2 Example 3.2: Soret and Dufour effects in a porous cavity

Using the following dimensionless variables: $x = x^*/H, y = y^*/X, \boldsymbol{u} = \boldsymbol{u}H/\nu, p = p^*H/\rho\nu, T = (T^* - T_0)/(T_1 - T_0)$ and $C = (C^* - c_0)/(C_1 - C_0)$ (where H is the cavity height and ν the kinematic viscosity of the fluid), we can write the equations describing transport phenomena in a square porous cavity with thermal and concentration diffusion in the form (3.1). We set $\mathbb{K} = \mathrm{Da}\,\mathbb{I}$, $\nu(T) = 1$ and $\boldsymbol{F}(\boldsymbol{y}) = (\mathrm{Gr}_T\, T + \mathrm{Gr}_C\, C)\boldsymbol{g}$, where $\boldsymbol{g} = (0, -1)^{\mathrm{T}}$ points in the direction of gravity, $\vec{m} = (T, C)^{\mathrm{T}}$, and the diffusion coefficients are given by

$$\mathbb{D} = \begin{bmatrix} R_k/\mathrm{Pr} & \mathrm{Du} \\ \mathrm{Sr} & 1/\mathrm{Sc} \end{bmatrix}.$$



Ra		100	200	400	1000	2000
Nu	Present Study	3.10	4.97	7.84	13.72	20.31
	Ref. [59]	3.15	5.02	7.83	14.01	20.00
	Ref. [86]	3.11	4.96	7.77	13.47	19.90
Sh	Present Study	13.58	20.73	30.91	49.42	66.80
	Ref. [59]	13.54	20.11	27.96	48.01	71.25
	Ref. [86]	13.25	19.86	28.41	48.32	69.29

Table 3.4: Example 3.2 (porous cavity): (left) sketched domain with boundary conditions, (right) comparison of average Nusselt and Sherwood numbers for $N = 0$, $\mathrm{Le} = 10$ with thermal Rayleigh numbers on Darcy's regime (table produced by the author).

Here, R_k is the thermal conductivity ratio, $\mathrm{Gr}_T, \mathrm{Gr}_C$ are the thermal and solutal Grashof numbers respectively, $\mathrm{Da} = \kappa/H^2$ is the Darcy number, $\mathrm{Pr} = \nu/\alpha$ the Prandtl number, $\mathrm{Sc} = \nu/D_C$ the Schmidt number, and the ratio $\mathrm{Le} = \mathrm{Sc}/\mathrm{Pr}$ the Lewis number.

For a preliminary validation we conduct a series of computational tests using a buoyancy ratio $N := \mathrm{Gr}_C/\mathrm{Gr}_T = 0$. The computational domain is the unit square $\Omega := (0,1)^2$, considering no-slip velocity conditions on Γ. Temperature and concentration are kept at T_0, C_0 and T_1, C_1 at the right and left walls respectively, where $T_0 < T_1$ and $C_0 < C_1$. Horizontal walls are adiabatic and impermeable, as depicted on the left of Table 3.4. In this subsection we will use $k = 2$ and a mesh with 20000 elements. We compute Nusselt and Sherwood numbers and compare these outputs against well-known benchmark data from [59] and [86]. The average values of Nu and Sh values on the left vertical wall are, respectively

$$\mathrm{Nu} = \int_0^1 \frac{\partial T}{\partial x}\bigg|_{x=0} \mathrm{d}y, \quad \mathrm{Sh} = \int_0^1 \frac{\partial C}{\partial x}\bigg|_{x=0} \mathrm{d}y.$$

For the values $R_k = 1.0$, $\mathrm{Da} = 10^{-7}$, $\mathrm{Le} = 10$, $\mathrm{Sr} = 0$, $\mathrm{Du} = 0$, and $\mathrm{Pr} = 10$, results for different thermal Rayleigh values are computed and summarised on the right panel of Table 3.4 along with the results from [59, 86]. For $\mathrm{Ra} \leq 1000$, the values of Nu and Sh are within a relative error of 3%, for the last value $\mathrm{Ra} = 2000$, within 6%.

Keeping the remaining parameters fixed, we now set $\mathrm{Ra} = 100$, $\mathrm{Le} = 0.8$ and $N = 1$. The effect of Dufour parameter on the flow, thermal and concentration fields are portrayed in Figure 3.1 for $\mathrm{Du} \in \{0.1, 1\}$. The velocity field and isotherms are in qualitative agreement with those in [24, Fig. 2]. In Figure 3.2 we repeat the plots keeping $\mathrm{Du} = 0$ and with Soret values of $\mathrm{Sr} \in \{0.1, 1\}$. As expected, the result is almost symmetric with an exchange of behaviour between temperature and concentration. Moreover, in both cases an increment of Sr or Du drives an increase of velocity in the recirculation patterns. Finally, in Figure 3.3 we fix $\mathrm{Du} = 0.5$, $\mathrm{Sr} = 0.5$ and test the effect of buoyancy by setting $N = -5$ and alternatively, $N = 5$. We can see the reversion of flow direction caused by the difference in buoyancy of the species. Note that in the last case $\mathbb{D}$ is not positive definite and solvability of the coupled problem cannot be guaranteed. Nevertheless, convergence of the Newton iterations

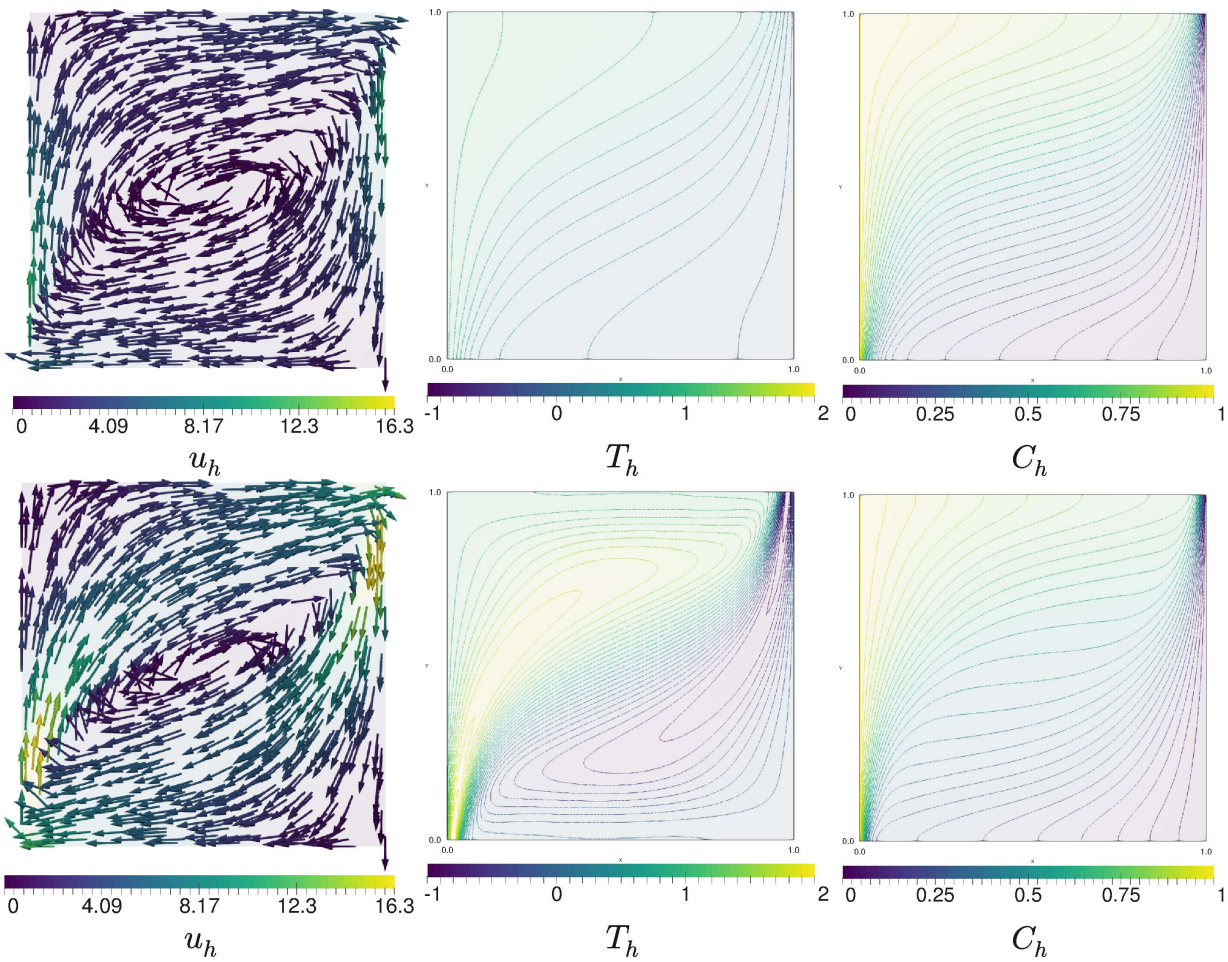


Figure 3.1: Example 3.2 (porous cavity): (left) velocity field, (middle) isotherms and (right) concentration contours for (top) $\mathrm{Du} = 0.1$, (bottom) $\mathrm{Du} = 1$ (figure produced by the author).

was observed for a broad range of parameters ($\mathrm{Sr}, \mathrm{Pr} \in [10^{-3}, 10^{3}]$, $N \in [1, 10]$, $\mathrm{Da} \in [10^{-7}, 1]$, $\mathrm{Ra} \in [100, 2000]$). The convergence of Newton iterates is lost only when the Soret number Sr takes values greater than 5 (and provided that $N \geq 0$ and $\mathrm{Du} = 0$).

3.5.3 Example 3.3: bioconvection of oxytactic bacteria

With the notation $\vec{m} = (c_1, c_2)^{\mathrm{T}}$ the oxytactic bacteria bioconvection phenomenon (see [119, 120]), can be modelled by (3.1), with diffusion, reaction, and remaining concentration-dependent coefficients given by

$$\mathbb{D}(\vec{m}) = \begin{bmatrix} D_1 & -\alpha r(c_2) c_1 \\ 0.0 & D_2 \end{bmatrix}, \quad g(\vec{m}) = \beta r(c_2) \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad \boldsymbol{F}(\vec{m}) = \gamma c_1 \boldsymbol{g}$$

$$\boldsymbol{g} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad r(c_2) = \frac{1}{2}\left(1 + \frac{c_2 - c_2^*}{\sqrt{(c_2 - c_2^*)^2 + \varepsilon^2}}\right).$$

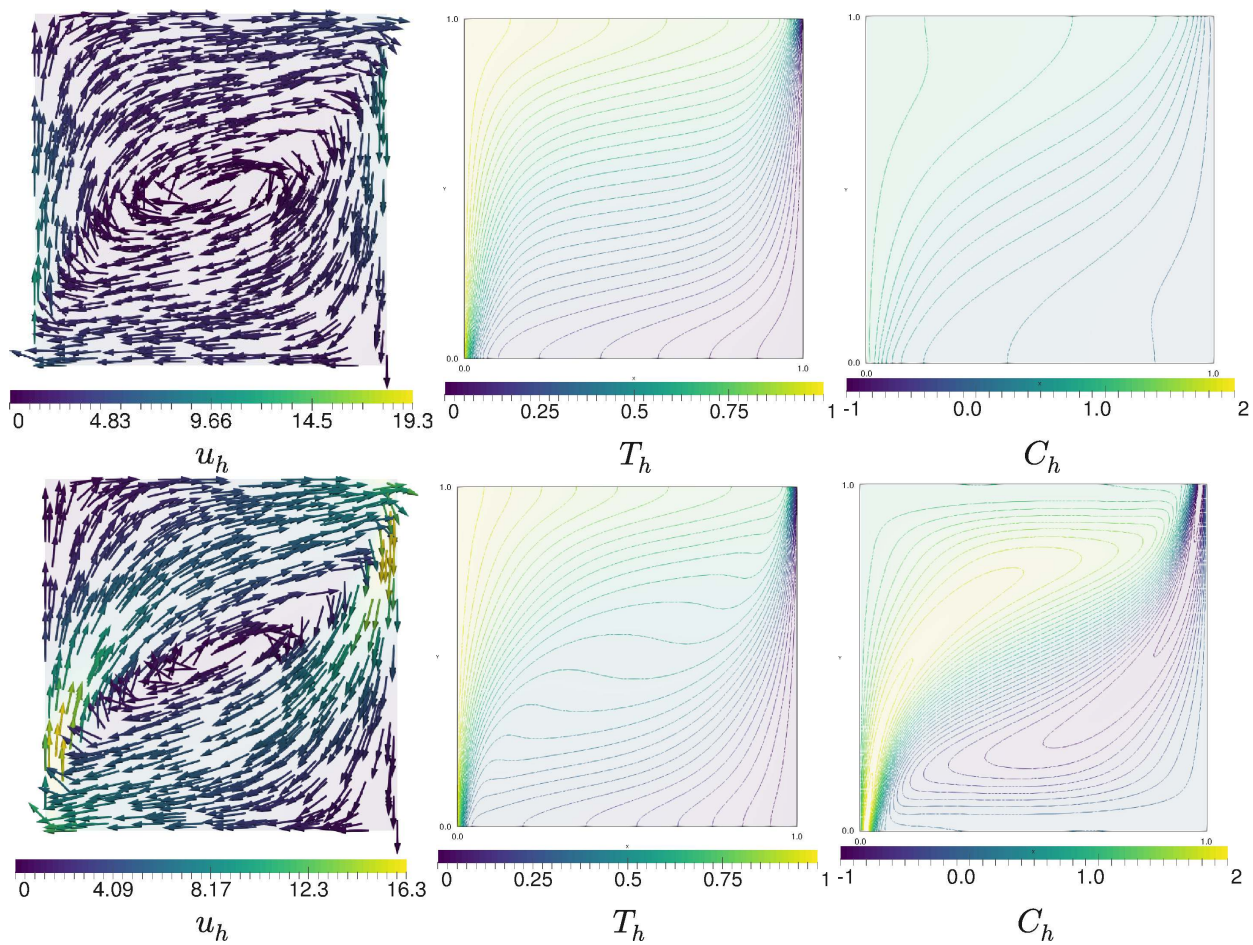


Figure 3.2: Example 3.2 (porous cavity): (left) velocity field, (middle) isotherms and (right) concentration contours for (top) $\mathrm{Sr} = 0.1$, (bottom) $\mathrm{Sr} = 1$ (figure produced by the author).

We consider a rectangular prism with square base $[0,1] \times [0,1]$ and height 0.75, discretised into a tetrahedral mesh of 48000 cells. Fixing the parameters $\beta = 0.1$, $D_1 = 0.01$, $D_2 = 0.2$, $\gamma = 5000$, $\alpha = 0.25$, $\mathrm{Sc} = 10^{-2}$, and $\mu = 2$, we use a pseudo timestep, using $\Delta t = 0.1$ to compute intermediate state solutions, starting from a distribution of bacteria packed in a ball of radius 0.2 and placed near the top of the vessel. Snapshots (at advanced time) of the numerical solution are displayed in Figures 3.4 and 3.5. We observe how the bacteria propagate downwards, producing recirculating zones as indicated by the velocity field. The first snapshot shows that the oxygen concentration has more variation on the top layers due to the competition between consumption of the high bacterial concentration, recirculating flow, and diffusion. Later on, oxygen concentration follows the flow direction, showing higher values downwards in the centre of the recirculating zones. The pressure distributes from low on the top, to high on the bottom, also decreasing its magnitude as the bacteria reaches the vessel's bottom.

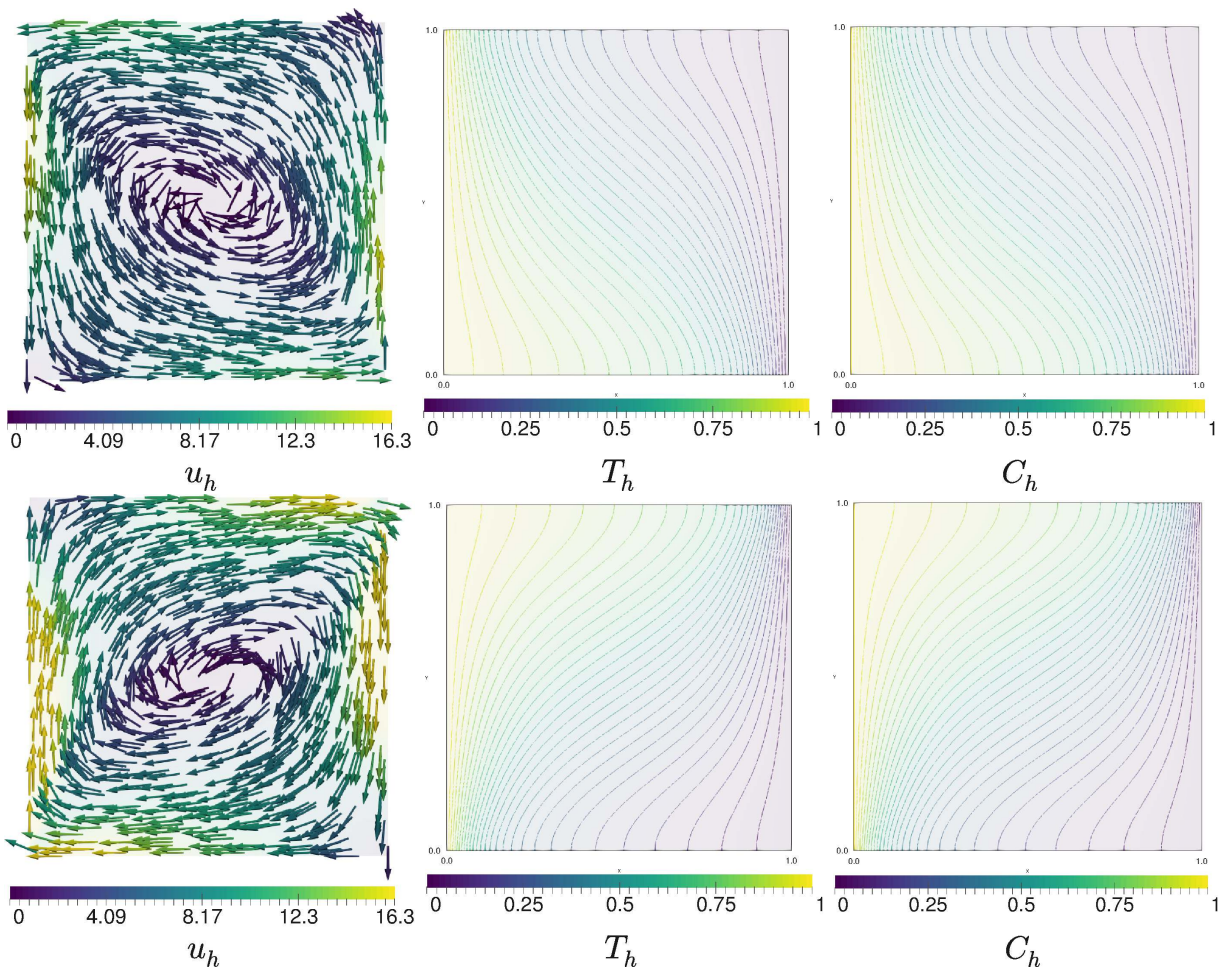


Figure 3.3: Example 3.2 (porous cavity): (left) velocity field, (middle) isotherms and (right) concentration contours for (top) $N = -5$, (bottom) $N = 5$ (figure produced by the author).

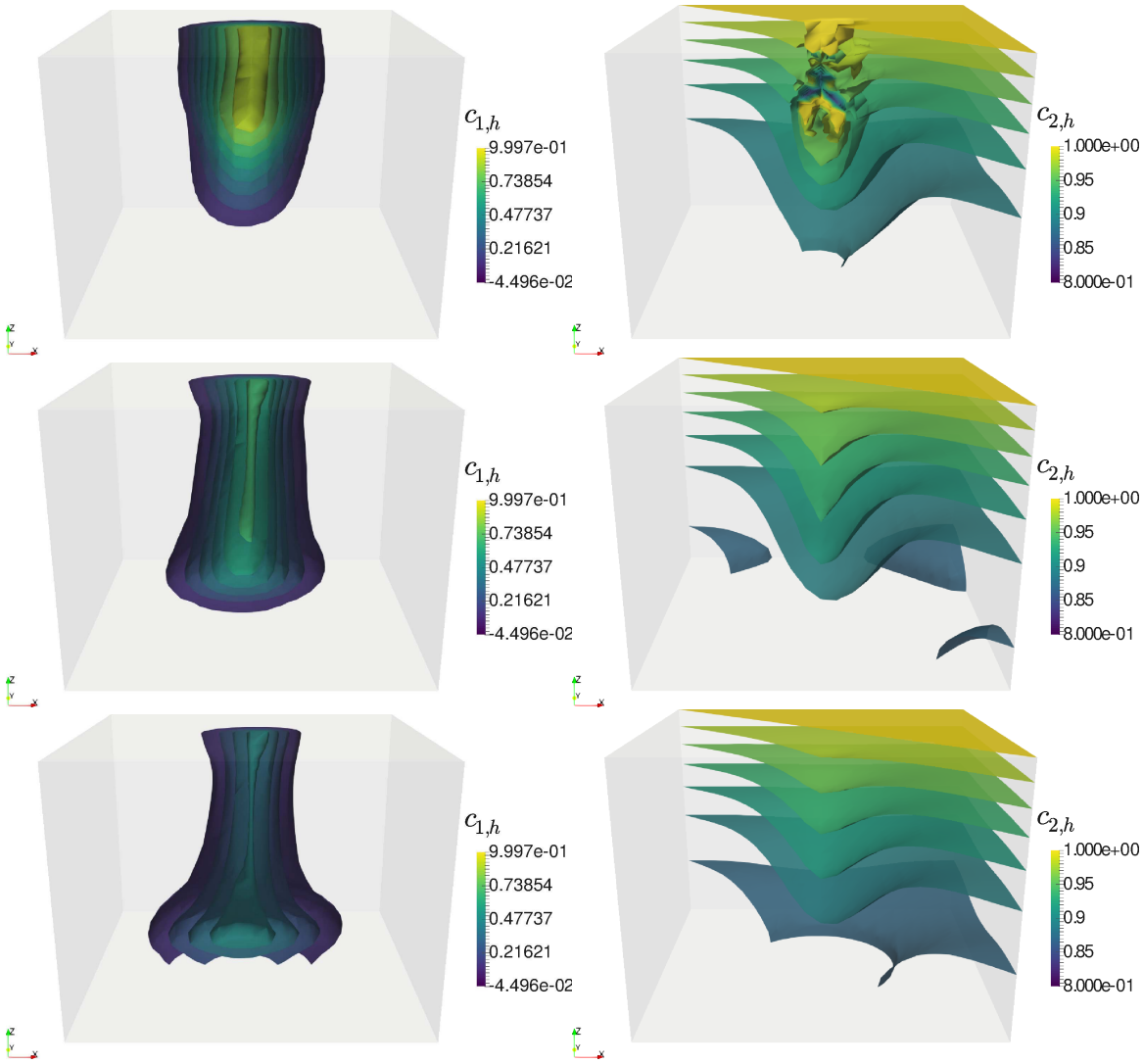


Figure 3.4: Example 3.3 (bioconvection): patterns generated by the bacterial chemotaxis towards oxygen concentration. Snapshots of the obtained solutions at times (top) $t = 0.1$, (middle) $t = 0.3$ and (bottom) $t = 0.5$ (figure produced by the author).

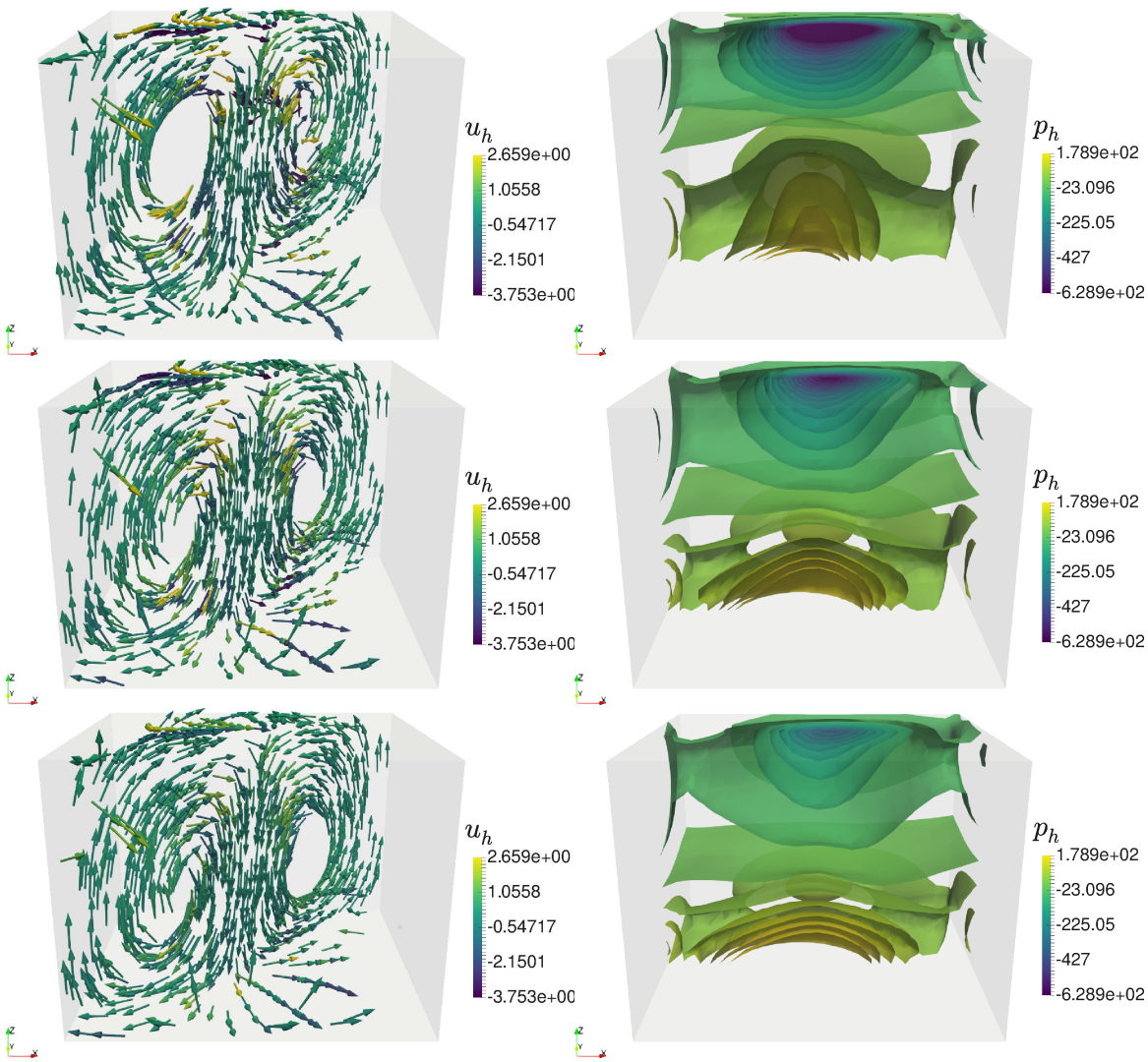


Figure 3.5: Example 3.3 (bioconvection): patterns generated by the bacterial chemotaxis towards oxygen concentration. Snapshots of the obtained solutions at times (top) $t = 0.1$, (middle) $t = 0.3$ and (bottom) $t = 0.5$ (figure produced by the author).

CHAPTER 4

Convergence of $\boldsymbol{H}(\mathrm{div})$-conforming schemes for a new model of sedimentation in circular clarifiers with a rotating rake

In this chapter we introduce a macroscopic model for simulating the sedimentation-consolidation of solid particles in an incompressible fluid under the effect of gravity and in the presence of a slowly rotating arm assisting the removal of sediment on the bottom of clarifier-thickener units. The governing model is an initial-boundary value problem for the Navier-Stokes equations describing the flow of the mixture coupled with a nonlinear parabolic equation describing the volume fraction of solids. The rotating structure is accounted for by suitable drag laws on the momentum balance of the mixture and on the mass balance of the solid phase. An $H(\mathrm{div})$-conforming method for the coupled problem is proposed, a rigorous proof of convergence is provided, and the validity of the new model and the performance of the scheme are demonstrated numerically by several computational tests.

4.1 Introduction

4.1.1 Scope

We advance a phenomenological model of solid-fluid interaction in a continuously operated clarifier-thickener, which is an equipment widely used in the mining industry, wastewater treatment plants, and other applications. The new approach accounts for the effect of the rotating rake structure, the influence of the settling solid particles, and the three-dimensional incompressible flow of the mixture. A large variety of these devices are used in industry, but most clarifier-thickeners are circular tanks of $1,50\,\mathrm{m}$ to $150\,\mathrm{m}$ in diameter equipped with a feed inlet and overflow and discharge outlets for continuous operation. In many devices, a pair of rotating rake arms that move over the gently sloped bottom help to move the concentrated slurry toward the centre of the tank, where it is removed. Clear liquid overflows the top of the tank and is collected through a circumferential launder (see Figure 4.1). Although there are many main types of thickeners or clarifiers such as bridge support, column support, and traction devices, for the purpose of the present modelling framework these are all considered equivalent.

The mathematical modelling and numerical simulation of this kind of processes is challenging due to

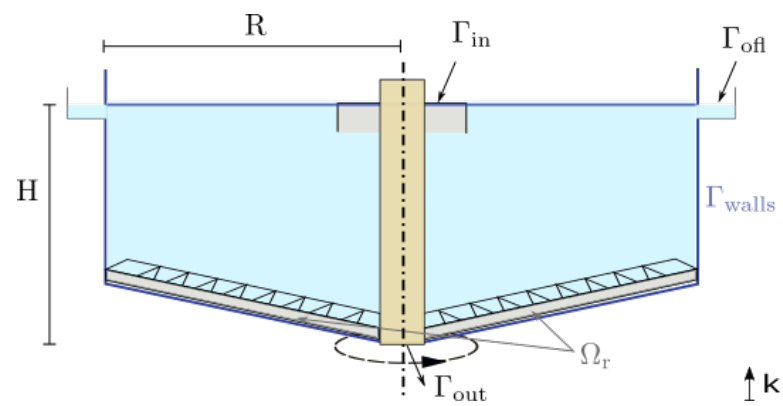


Figure 4.1: Schematic view of the clarifier unit, indicating height H, maximal radius R, and the location of the rotating rake; as well as the separation of the boundary into the walls, the outlet, the feedwell inlet, and the overflow weir (figure produced by the author).

the intrinsic multiscale and highly nonlinear nature of the sedimentation-consolidation mechanisms, complicated geometries and boundary conditions, as well as the feedback interaction between the mixture flow and the motion of the rake (the fluid applies a load on the solid structure, implying a deformation, generating stresses, and eventually modifying the flow). For instance, simplified models that would be based on geometrical symmetry are in this case of very restricted applicability, since the settling of the particles occurs in the vertical direction while the rotation of the rake acts in two horizontal directions, and the velocity distribution under typical operating conditions is quite far from unidirectional.

We consider process of sedimentation and transport of a suspension consisting of a phase of finely divided solid particles dispersed in a viscous fluid. This mixture is contained in a clarifier tank with a moving rake. For the sedimentation-consolidation of the suspension we assume that the particles are relatively small with respect to the tank size and possess the same density. It is assumed that the mixture is composed of incompressible solid and liquid phases, that the mixture velocity is relatively small, and that the suspension is already flocculated before the process starts (see [56, 148]). The motion of the mixture is governed by the incompressible Navier-Stokes equations coupled with the transport equation for the solids as follows,

$$\rho_{\text{f}}\left(\frac{\partial \boldsymbol{u}}{\partial t} + \mathbf{div}(\boldsymbol{u}\otimes\boldsymbol{u})\right) - \mathbf{div}\big(\nu(c)\boldsymbol{\varepsilon}(\boldsymbol{u})\big) + \nabla p = \boldsymbol{f}_{\text{g}}(c) + \boldsymbol{f}_{\text{r}}(\boldsymbol{u},\boldsymbol{x},t), \tag{4.1a}$$

$$\operatorname{div}\boldsymbol{u} = 0, \tag{4.1b}$$

$$\frac{\partial c}{\partial t} - \operatorname{div}\big(D(c)\nabla c - c\boldsymbol{u} - f_{\text{bk}}(c)\boldsymbol{k}\big) = -g_{\text{r}}(c,\boldsymbol{x},t) \quad \text{in } \Omega\times(0,\mathcal{T}). \tag{4.1c}$$

Here the sought quantities are the mixture velocity $\boldsymbol{u}$, the pressure p and the local solids fraction c as functions of time $t\in[0,\mathcal{T}]$ and spatial position $\boldsymbol{x}\in\Omega\subset\mathbb{R}^3$, where the spatial domain Ω represents the interior of the clarifier-thickener. Moreover, ρ_{f} is the fluid density, $\boldsymbol{\varepsilon}(\boldsymbol{u}) = \frac{1}{2}(\nabla\boldsymbol{u}+\nabla\boldsymbol{u}^{\text{T}})$ is the strain rate tensor, and $\boldsymbol{k}$ is the upwards-pointing unit vector. The material behaviour is described by the concentration-dependent viscosity ν, the Kynch batch flux density function f_{bk}, and the diffusion function D. These three quantities are nonlinear given functions of c that are specified in Section 4.2.1.

The term $\boldsymbol{f}_{\mathrm{g}}(c)$ represents the body force and is given by $\boldsymbol{f}_{\mathrm{g}}(c) = \boldsymbol{g}\,(\rho_{\mathrm{s}} - \rho_{\mathrm{f}})\,c$ as in [65], where $\boldsymbol{g} = -g\boldsymbol{k}$ and g is the acceleration of gravity. The terms $\boldsymbol{f}_{\mathrm{r}}(\boldsymbol{u}, \boldsymbol{x}, t)$ and $g_{\mathrm{r}}(c, \boldsymbol{x}, t)$ describe the action of the rotating rake, and are specified in Section 4.2.2. The system (4.1) is supplied with initial and boundary conditions that are made precise in Section 4.2.3.

It is the purpose of this chapter to advance a novel discretisation for the resulting initial-boundary value problem that is of second-order in space and time. The discretisation employs divergence-conforming BDM elements of order k for the approximation of the velocity, discontinuous elements of order $k-1$ for the pressure, and continuous Lagrange elements of order k for the volume fraction. We use an interior penalty discontinuous Galerkin technique in order to enforce $\boldsymbol{H}^1$-continuity of the velocity (similarly as done in [49]); and employ the second-order backward differentiation formula (BDF2) for the discretisation in time. Our analysis includes the stability of solutions of the associated Galerkin scheme and the derivation of optimal error estimates in time and space for problems with small and sufficiently smooth solutions. These properties constitute a proof of convergence of the fully discrete scheme as the meshwidth and the time step tend to zero. The novelty of the treatment consists in the inclusion of terms that account for the influence of the rake motion on the momentum balance and the removal of solids. We also adapt techniques of the immerse boundary finite element method (see e.g. [34]) for the analysis and numerical approximation of those terms.

4.1.2 Outline of the chapter

We have organised the contents of this chapter in the following manner. Section 4.2 describes the general governing equations, the constitutive relations, and the interaction terms. It also specifies the boundary and initial conditions, and it outlines the weak formulation of the problem for a fixed time. In Section 4.3 we introduce the Galerkin discretisation and define the fully discrete method, briefly addressing stability and convergence properties. Section 4.4 is devoted to the computational results, including parameter calibration, accuracy verification, as well as the simulation of clarifier performance under different operation scenarios.

4.2 Preliminaries

4.2.1 Constitutive functions

The viscosity ν is supposed to be given by the following nonlinear function of c:

$$\nu(c) = \nu_0 + \nu_0(1 - c/c_{\max})^b, \tag{4.2}$$

where ν_0 is the viscosity of the pure fluid, $b > 0$ is a parameter, and $c_{\max}$ is a (nominal) maximum solids volume fraction. We do not consider here the high-order terms that account for microstructural arrangement of the granular material as e.g. in [148].

Moreover, the one-dimensional Kynch batch flux density function describing hindered settling [117], f_{bk}; and the sediment compressibility, $D(c)$; are non linear functions of the concentration c, which can

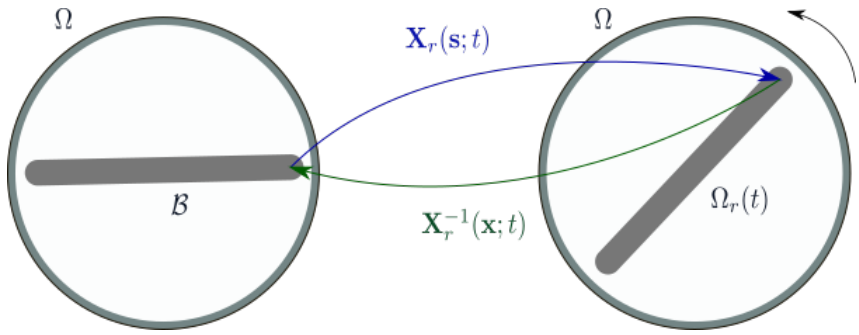


Figure 4.2: Schematic representation of the mapping $\boldsymbol{X}_r$ from the rake reference domain $\mathcal{B}$ to the moving domain $\Omega_r(t)$ in a longitudinal section of the clarifier unit (figure produced by the author).

be taken as follows [47]:

$$D(c) = D_0 + \frac{f_{\mathrm{bk}}(c)\sigma_{\mathrm{e}}'(c)}{(\rho_{\mathrm{s}} - \rho_{\mathrm{f}})gc}, \qquad f_{\mathrm{bk}}(c) = V_\infty \left[c \left(1 - \frac{c}{c_{\max}} \right)^{\eta_F} \right], \tag{4.3}$$

where V_∞ is the Stokes velocity, η_F a material-dependent exponent, $D_0 > 0$ is the constant of hydrodynamic self-diffusion, ρ_{s} and ρ_{f} are the solid and fluid mass densities, respectively, and $\sigma_{\mathrm{e}}(c)$ is the so-called effective solid stress function, which characterises sediment compressibility in the case of flocculated particles. The function σ_{e} is assumed to satisfy $\sigma_{\mathrm{e}}'(c) = \mathrm{d}\sigma_{\mathrm{e}}(c)/\mathrm{d}c \geq 0$ for all c, which ensures that $D(c) \geq D_0 > 0$.

4.2.2 Rotating rake

To include the rotating rake into the computational model, we follow Das et al. [65] using a simplified approach that only takes into account the area of influence of the rake, and characterises the details of its geometry through parameters. The rake area of influence (hereinafter we will refer to it only as rake) $\Omega_{\mathrm{r}}(t)$ can be represented as the image of a mapping $\boldsymbol{X}_{\mathrm{r}}(\cdot;t)$ from a reference domain $\mathcal{B} \subset \mathbb{R}^d$ (see figure 4.2). We denote by $\boldsymbol{s}$ the coordinates in $\mathcal{B}$, then $\boldsymbol{X}_{\mathrm{r}}(\boldsymbol{s};t)$ represents the position of a point in the current domain $\Omega_{\mathrm{r}}(t)$. That is, $\boldsymbol{x} \in \Omega_{\mathrm{r}}(t)$ if and only if there exists $\boldsymbol{s} \in \mathcal{B}$ such that $\boldsymbol{x} = \boldsymbol{X}_{\mathrm{r}}(\boldsymbol{s};t)$. For simplicity we will consider a constant angular velocity ω for the rake, then the rake velocity $\boldsymbol{u}_{\mathrm{r}}(\boldsymbol{s})$, depends only on the distance to the rake centre. Further, we suppose $\boldsymbol{f}_{\mathrm{r}}$ depends on the difference between the fluid velocity and the rake velocity $\boldsymbol{u}_{\mathrm{r}}$; and g_{r} depends on the difference between the concentration in front of the rake and a concentration after removal c_{r}, which is linked to the rake geometry. To express $\boldsymbol{f}_{\mathrm{r}}(\boldsymbol{u},\boldsymbol{x},t)$ in compact form, it is useful to define the function $\zeta : \mathbb{R} \to \mathbb{R}$ given by $\zeta(x) = x^2 \operatorname{sgn} x = x|x|$. Then we define

$$\boldsymbol{f}_{\mathrm{r}}(\boldsymbol{u},\boldsymbol{x},t) := \begin{cases} \beta\rho_{\mathrm{r}}\zeta\big((\boldsymbol{u}_{\mathrm{r}}(\boldsymbol{X}_r^{-1}(\boldsymbol{x};t)) - \boldsymbol{u}(\boldsymbol{x},t)) \cdot \boldsymbol{n}_{\mathrm{r}}\big)\boldsymbol{n}_{\mathrm{r}} & \text{if } \boldsymbol{x} \in \Omega_{\mathrm{r}}(t), \\ \boldsymbol{0} & \text{otherwise,} \end{cases}$$

$$g_{\mathrm{r}}(c,\boldsymbol{x},t) := \begin{cases} \alpha\big(c(\boldsymbol{x},t) - c_{\mathrm{r}}\big) & \text{if } \boldsymbol{x} \in \Omega_{\mathrm{r}}(t), \\ 0 & \text{otherwise,} \end{cases}$$

where α is a removal coefficient, β is the drag coefficient that includes the contact surface to volume ratio, ρ_{r} the rake density and $\boldsymbol{n}_{\mathrm{r}}$ the vector pointing towards the tangential direction with respect to the circular motion of the rake in the (x_1, x_2)-plane. Following the approach of the immersed boundary method [34, 35] as well as a recently proposed model arising in the context of flow-canopy interaction [161], we rewrite these expressions as

$$\begin{aligned} \boldsymbol{f}_{\mathrm{r}}(\boldsymbol{u}, \boldsymbol{x}, t) &= \beta\rho_{\mathrm{r}} \int_{\mathcal{B}} \zeta\big(\boldsymbol{u}_{\mathrm{r}}(\boldsymbol{s}) - \boldsymbol{u}(\boldsymbol{X}_{\mathrm{r}}(\boldsymbol{s}, t); t)\big) \cdot \boldsymbol{n}_{\mathrm{r}}\big)\boldsymbol{n}_{\mathrm{r}}\delta\big(\boldsymbol{x} - \boldsymbol{X}_{\mathrm{r}}(\boldsymbol{s}; t)\big)\,\mathrm{d}\boldsymbol{s}, \\ g_{\mathrm{r}}(c, \boldsymbol{x}, t) &= \alpha \int_{\mathcal{B}} \big(c(\boldsymbol{X}_{\mathrm{r}}(\boldsymbol{s}, t); t) - c_{\mathrm{r}}\big)\delta\big(\boldsymbol{x} - \boldsymbol{X}_{\mathrm{r}}(\boldsymbol{s}; t)\big)\,\mathrm{d}\boldsymbol{s} \quad \text{for all } \boldsymbol{x} \in \Omega \text{ and } t \in (0, \mathcal{T}). \end{aligned} \tag{4.4}$$

Here, δ is the Dirac delta function. Even if the presence of the rotating arm through (4.4) does not resolve stress localisation on the structure, it already represents an extension over the model in [65].

4.2.3 Initial and boundary conditions

The set of governing equations is furnished with the following initial and boundary conditions:

$$\boldsymbol{u}(0) = \boldsymbol{0}, \quad c(0) = c_0 \qquad \text{in } \Omega, \tag{4.5a}$$
$$\boldsymbol{u}(\boldsymbol{x}, t) = \boldsymbol{u}_{\mathrm{in}} \qquad \text{on } \Gamma_{\mathrm{in}},\ t \in [0, \mathcal{T}], \tag{4.5b}$$
$$c(\boldsymbol{x}, t) = c_{\mathrm{in}} \qquad \text{on } \Gamma_{\mathrm{in}},\ t \in [0, \mathcal{T}], \tag{4.5c}$$
$$\boldsymbol{u}(\boldsymbol{x}, t) = \boldsymbol{0} \qquad \text{on } \Gamma_{\mathrm{wall}},\ t \in [0, \mathcal{T}], \tag{4.5d}$$
$$[\nu(c)\boldsymbol{\varepsilon}(\boldsymbol{u}) - p\boldsymbol{I}]\boldsymbol{n} = \boldsymbol{0} \qquad \text{on } \Gamma_{\mathrm{out}} \cup \Gamma_{\mathrm{ofl}},\ t \in [0, \mathcal{T}], \tag{4.5e}$$
$$(D(c)\nabla c - f_{\mathrm{bk}}(c)\boldsymbol{k}) \cdot \boldsymbol{n} = 0 \qquad \text{on } \Gamma_{\mathrm{wall}} \cup \Gamma_{\mathrm{in}},\ t \in [0, \mathcal{T}], \tag{4.5f}$$
$$D(c)\nabla c \cdot \boldsymbol{n} = 0 \qquad \text{on } \Gamma_{\mathrm{out}} \cup \Gamma_{\mathrm{ofl}},\ t \in [0, \mathcal{T}], \tag{4.5g}$$

which represent that at the inlet we impose velocity and volume fraction of solids, on the walls we set no-slip velocity and zero-flux for c, and on the outlet and effluent overflow regions we set zero normal total stress, and zero total flux. The disposition of domain boundaries is exemplified in Figure 4.1.

4.2.4 Weak formulation

The weak formulation of problem (4.1) is obtained by testing against suitable functions and integrating by parts, and can be stated as follows:

Find $(\boldsymbol{u}(t), p(t), c(t)) \in \boldsymbol{H}^1(\Omega) \times L^2(\Omega) \times H^1(\Omega)$ satisfying
the boundary conditions (4.5b) and (4.5c) and for all $\boldsymbol{v} \in \boldsymbol{H}^1_0(\Omega)$, $q \in L^2(\Omega)$ and $l \in H^1(\Omega)$:

$$\begin{aligned} &\big(\partial_t \boldsymbol{u}(t), \boldsymbol{v}\big)_\Omega + a_1\big(c(t); \boldsymbol{u}(t), \boldsymbol{v}\big) + c_1\big(\boldsymbol{u}(t); \boldsymbol{u}(t), \boldsymbol{v}\big) - b\big(\boldsymbol{v}, p(t)\big) = \boldsymbol{F}_{\mathrm{g}}(c(t), \boldsymbol{v}) + \boldsymbol{F}_{\mathrm{r}}(\boldsymbol{u}(t), \boldsymbol{v}), \\ &b\big(\boldsymbol{u}(t), q\big) = 0, \\ &\big(\partial_t c(t), l\big)_\Omega + a_2\big(c(t); c(t), l\big) + c_2\big(\boldsymbol{u}(t); c(t), l\big) - d_2\big(c(t), l\big) = -G_{\mathrm{r}}\big(c(t), l\big). \end{aligned} \tag{4.6}$$

Using [34, Lemma 1], we can consider $\boldsymbol{F}_{\mathrm{r}} \in \boldsymbol{H}^{-1}(\Omega)$, $G_{\mathrm{r}} \in H^{-1}(\Omega)$, and the variational forms that are defined as follows for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \boldsymbol{H}^1(\Omega)$, $q \in L^2(\Omega)$, and $c, l \in H^1(\Omega)$:

$$
\begin{aligned}
&a_1(c;\boldsymbol{u},\boldsymbol{v}) := \big(\nu(c)\boldsymbol{\varepsilon}(\boldsymbol{u}),\boldsymbol{\varepsilon}(\boldsymbol{v})\big)_\Omega, \quad b(\boldsymbol{v},q) := (q,\operatorname{div}\boldsymbol{v})_\Omega, \quad c_1(\boldsymbol{w};\boldsymbol{u},\boldsymbol{v}) := \big((\boldsymbol{w}\cdot\nabla)\boldsymbol{u},\boldsymbol{v}\big)_\Omega,\\
&\boldsymbol{F}_{\mathrm{g}}(c,v) = \big(\boldsymbol{g}\,(\rho_{\mathrm{f}}-\rho_{\mathrm{s}})\,c,\boldsymbol{v}\big)_\Omega,\\
&\boldsymbol{F}_{\mathrm{r}}(\boldsymbol{u},\boldsymbol{v}) = \beta\rho_{\mathrm{r}}\int_{\mathcal{B}}\zeta\big((\boldsymbol{u}_{\mathrm{r}}(\boldsymbol{s})-\boldsymbol{u}(\boldsymbol{X}_{\mathrm{r}}(\boldsymbol{s};t),t))\cdot\boldsymbol{n}_{\mathrm{r}}\big)\boldsymbol{n}_{\mathrm{r}}\boldsymbol{v}\big(\boldsymbol{X}_{\mathrm{r}}(\boldsymbol{s};t)\big)\,\mathrm{d}s,\\
&a_2(c;c,l) := \big(D(c)\nabla c,\nabla l\big)_\Omega, \quad c_2(\boldsymbol{v};c,l) := (\boldsymbol{v}\cdot\nabla c,l)_\Omega\,,\\
&d_2(c,l) = \big(f_{\mathrm{bk}}(c)\boldsymbol{k},\nabla l\big)_\Omega - \langle f_{\mathrm{bk}}(c)\boldsymbol{k}\cdot\boldsymbol{n},l\rangle_{\Gamma_{\mathrm{out}}\cup\Gamma_{\mathrm{ofl}}}, \quad G_{\mathrm{r}}(c,l) := \alpha\int_{\mathcal{B}}\big(c(\boldsymbol{X}_{\mathrm{r}}(\boldsymbol{s};t),t)-c_{\mathrm{r}}\big)l\big(\boldsymbol{X}_{\mathrm{r}}(\boldsymbol{s};t)\big)\,\mathrm{d}s.
\end{aligned}
$$

Although some related results are available from the literature, for instance the existence of strong and weak solutions for the periodic motion of a rigid body in an incompressible fluid [80], the solvability analysis of (4.6) is still an open problem. We will proceed to the semidiscrete analysis under the assumption that the continuous problem is well-posed and that the weak solutions are regular enough.

4.3 Numerical method

4.3.1 Definition of the discrete problem

For the space discretisation, we will consider a family of regular partitions, denoted $\mathcal{T}_h$, of $\Omega \subset \mathbb{R}^d$ into simplices K (triangles in 2D or tetrahedra in 3D) of diameter h_K. For $k \geq 1$ and a mesh $\mathcal{T}_h$ on Ω, let us consider the discrete spaces (see e.g. [40, 49])

$$
\begin{aligned}
\boldsymbol{V}_h &:= \Big\{\, \boldsymbol{v}_h \in \boldsymbol{H}(\operatorname{div};\Omega) : \boldsymbol{v}_h|_K \in [\mathcal{P}_k(K)]^{\mathrm{d}} \quad \forall K \in \mathcal{T}_h \,\Big\},\\
\mathcal{Q}_h &:= \big\{\, q_h \in L^2(\Omega) : q_h|_K \in \mathcal{P}_{k-1}(K) \quad \forall K \in \mathcal{T}_h \,\big\},\\
\mathcal{M}_h &:= \big\{\, s_h \in C(\bar{\Omega}) : l_h|_K \in \mathcal{P}_k(K) \quad \forall K \in \mathcal{T}_h \,\big\},
\end{aligned}
$$

which in particular satisfy $\operatorname{div}\boldsymbol{V}_h \subset \mathcal{Q}_h$ (cf. [109]). Here $\mathcal{P}_k(K)$ denotes the local space spanned by polynomials of degree up to k and $\boldsymbol{V}_h$ is the space of divergence-conforming BDM elements. Associated with these finite-dimensional spaces, we state the following semi-discrete Galerkin formulation for problem (4.1):

$$
\begin{aligned}
&\text{Find } (\boldsymbol{u}_h,p_h,c_h) \in \boldsymbol{V}_h\times\mathcal{Q}_h\times\mathcal{M}_h \text{ such that for all } (\boldsymbol{v}_h,q_h,l_h) \in \boldsymbol{V}_h\times\mathcal{Q}_h\times\mathcal{M}_h:\\
&(\partial_t\boldsymbol{u}_h,\boldsymbol{v}_h)_\Omega + a_1^h(c_h;\boldsymbol{u}_h,\boldsymbol{v}_h) + c_1^h(\boldsymbol{u}_h;\boldsymbol{u}_h,\boldsymbol{v}_h) - b(\boldsymbol{v}_h,p_h) = \boldsymbol{F}_{\mathrm{g}}(c_h,v_h) + \boldsymbol{F}_{\mathrm{r}}(\boldsymbol{u}_h,\boldsymbol{v}_h),\\
&b(\boldsymbol{u}_h,q_h) = 0,\\
&(\partial_t c_h,l_h)_\Omega + a_2(c_h;c_h,l_h) + c_2(\boldsymbol{u}_h;c_h,l_h) - d_2(c_h,l_h) = -G_{\mathrm{r}}(c_h,l_h).
\end{aligned}
\tag{4.7}
$$

Here the discrete versions of the trilinear forms $a_1^h(\cdot;\cdot,\cdot)$ and $c_1^h(\cdot;\cdot,\cdot)$ are defined using a symmetric interior penalty and an upwind approach, respectively (see e.g. [19, 109]):

$$\begin{aligned} a_1^h(c_h;\boldsymbol{u}_h,\boldsymbol{v}_h) &:= \int_\Omega \big(\nu(c_h)\boldsymbol{\varepsilon}_h(\boldsymbol{u}_h):\boldsymbol{\varepsilon}_h(\boldsymbol{v}_h)\big) + \sum_{e\in\mathcal{E}_h}\int_e \Big(-\{\!\{\nu(c_h)\boldsymbol{\varepsilon}_h(\boldsymbol{u}_h)\boldsymbol{n}_e\}\!\}\cdot[\![\boldsymbol{v}_h]\!] \\ &\qquad - \{\!\{\nu(c_h)\boldsymbol{\varepsilon}_h(\boldsymbol{v}_h)\boldsymbol{n}_e\}\!\}\cdot[\![\boldsymbol{u}_h]\!] + \frac{a_0}{h_e}\nu(c_h)[\![\boldsymbol{u}_h]\!]\cdot[\![\boldsymbol{v}_h]\!]\Big), \\ c_1^h(\boldsymbol{w}_h;\boldsymbol{u}_h,\boldsymbol{v}_h) &:= \int_\Omega (\boldsymbol{w}_h\cdot\nabla)\boldsymbol{u}_h\cdot\boldsymbol{v}_h + \sum_{K\in\mathcal{T}_h}\int_{\partial K\setminus\Gamma}\hat{\boldsymbol{w}}_h^{\mathrm{up}}(\boldsymbol{u}_h)\cdot\boldsymbol{v}_h, \end{aligned}$$

where the upwind flux is defined as $\hat{\boldsymbol{w}}_h^{\mathrm{up}}(\boldsymbol{u}_h) := \frac{1}{2}(\boldsymbol{w}_h\cdot\boldsymbol{n}_k - |\boldsymbol{w}_h\cdot\boldsymbol{n}_K|)(\boldsymbol{u}_h^e - \boldsymbol{u}_h)$, and $\boldsymbol{u}_h^e$ is the trace of $\boldsymbol{u}_h$ taken from within the exterior of K.

Let us introduce a partition of the interval $[0,\mathcal{T}]$ into N subintervals $[t_{n-1},t_n]$ of length Δt. We will use an implicit, BDF2 formula. That is, all first-order time derivatives are approximated using the centred operator

$$\partial_t\boldsymbol{u}_h(t^{n+1}) \approx \frac{1}{\Delta t}\left(\frac{3}{2}\boldsymbol{u}_h^{n+1} - 2\boldsymbol{u}_h^n + \frac{1}{2}\boldsymbol{u}_h^{n-1}\right),$$

(similarly for $\partial_t c$) whereas for the first time step a first-order backward Euler method is used from t^0 to t^1, starting from the interpolates $\boldsymbol{u}_h^0, c_h^0$ of the initial data. The resulting set of nonlinear equations is solved with an iterative Newton-Raphson method with exact Jacobian.

4.3.2 Spatio-temporal accuracy of the discretisation

For sake of the subsequent analysis, we assume Lipschitz continuity of the concentration-dependent viscosity

$$\nu \in \mathrm{Lip}(\mathbb{R}_+); \quad \exists \nu_{\min},\nu_{\max}:\ \forall c\in\mathbb{R}_+:\ \nu_{\min}\le\nu(c)\le\nu_{\max}.$$

Moreover, the flux $f_{\mathrm{bk}}(c)$ is assumed to be Lipschitz continuous, and the diffusion coefficient $D = D(c)$ is supposed to be a nonlinear function satisfying

$$D \in \mathrm{Lip}(\mathbb{R}_+); \quad \exists D_1,D_2>0:\ \forall c\in\mathbb{R}_+:\ D_1\le D(c)\le D_2. \tag{4.8}$$

For simplicity, we impose the following modified boundary conditions:

$$\boldsymbol{u}(\boldsymbol{x},t) = \boldsymbol{0}, \quad c(\boldsymbol{x},t) = 0, \quad (D(c)\nabla c - f_{\mathrm{bk}}(c)\boldsymbol{k})\cdot\boldsymbol{n} = 0 \quad \text{on } \Gamma,\ t\in[0,\mathcal{T}],$$

and we emphasise that the analysis can be extended to the non-homogeneous case following, for instance, lifting arguments.

We utilise the following mesh dependent broken norms

$$\begin{aligned} &\|\boldsymbol{v}\|_{*,\mathcal{T}_h}^2 := \sum_{K\in\mathcal{T}_h}\|\nabla\boldsymbol{v}\|_{0,K}^2 + \sum_{e\in\mathcal{E}_h}\frac{1}{h_e}\|[\![\boldsymbol{v}]\!]\|_{0,e}^2, \quad \|\boldsymbol{v}\|_{1,\mathcal{T}_h}^2 := \|\boldsymbol{v}\|_{0,\Omega}^2 + \|\boldsymbol{v}\|_{*,\mathcal{T}_h}^2 \quad \text{for all } \boldsymbol{v}\in\boldsymbol{H}^1(\mathcal{T}_h), \\ &\|\boldsymbol{v}\|_{2,\mathcal{T}_h}^2 := \|\boldsymbol{v}\|_{1,\mathcal{T}_h}^2 + \sum_{K\in\mathcal{T}_h} h_K^2|\boldsymbol{v}|_{2,K}^2 \quad \text{for all } \boldsymbol{v}\in\boldsymbol{H}^2(\mathcal{T}_h). \end{aligned}$$

We also recall the broken version of the well-known Sobolev embedding result (see e.g. [83, Lemma 6.2], [102, Prop. 4.5] or [67, Th. 5.3]): for any $r > 1$ if $d = 2$ or $1 \le r \le 6$, if $d = 3$ there exists a constant $C_{\mathrm{emb}} > 0$ such that

$$\|\boldsymbol{v}\|_{\boldsymbol{L}^r(\Omega)} \le C_{\mathrm{emb}} \|\boldsymbol{v}\|_{1,\mathcal{T}_h} \qquad \text{for all } \boldsymbol{v} \in \boldsymbol{H}^1(\mathcal{T}_h). \tag{4.9}$$

Furthermore, we will use the broken space

$$\boldsymbol{C}^1(\mathcal{T}_h) := \big\{\boldsymbol{u} \in \boldsymbol{H}^1(\mathcal{T}_h) : \boldsymbol{u}|_K \in \boldsymbol{C}^1(\bar{K}), K \in \mathcal{T}_h\big\},$$

equipped with an appropriate norm $\|\boldsymbol{u}\|_{\boldsymbol{W}^{1,\infty}(\mathcal{T}_h)} := \max_{K\in\mathcal{T}_h} \|\boldsymbol{u}\|_{\boldsymbol{W}^{1,\infty}(K)}$. Using the discrete norms, embedding (4.9) and local trace inequalities, we can establish continuity of the trilinear and bilinear forms involved, stated in the following lemma that can be proved following [19, Section 4]:

Lemma 4.1. *The following properties hold:*

$$\big|a_1^h(\cdot, \boldsymbol{u}, \boldsymbol{v})\big| \le \tilde{C}_a \|\boldsymbol{u}\|_{1,\mathcal{T}_h} \|\boldsymbol{v}\|_{1,\mathcal{T}_h} \quad \text{for all } \boldsymbol{u}, \boldsymbol{v} \in \boldsymbol{V}_h, \tag{4.10a}$$

$$\big|b(\boldsymbol{v}, q)\big| \le \tilde{C}_b \|\boldsymbol{v}\|_{1,\mathcal{T}_h} \|q\|_{0,\Omega} \quad \text{for all } \boldsymbol{v} \in \boldsymbol{H}^1(\mathcal{T}_h),\ q \in L^2_0(\Omega), \tag{4.10b}$$

$$\big|c_2(\boldsymbol{w}; c, l)\big| \le \tilde{C}_1 \|\boldsymbol{w}\|_{1,\mathcal{T}_h} \|l\|_{1,\Omega} \|c\|_{1,\Omega} \quad \text{for all } \boldsymbol{w} \in \boldsymbol{H}^1(\mathcal{T}_h) \text{ and } l, c \in H^1(\Omega). \tag{4.10c}$$

Moreover, for $c_1, c_2 \in H^1(\Omega)$, $c \in W^{1,\infty}(\Omega)$, $\boldsymbol{u} \in \boldsymbol{C}^1(\mathcal{T}_h) \cap \boldsymbol{H}^1_0(\Omega)$ and $\boldsymbol{v} \in \boldsymbol{V}_h$, there holds

$$\begin{aligned} &\big|a_1^h(c_1; \boldsymbol{u}, \boldsymbol{v}) - a_1^h(c_2; \boldsymbol{u}, \boldsymbol{v})\big| \le \tilde{C}_{\mathrm{Lip}} \|c_1 - c_2\|_{1,\Omega} \|\boldsymbol{u}\|_{\boldsymbol{W}^{1,\infty}(\mathcal{T}_h)} \|\boldsymbol{v}\|_{1,\mathcal{T}_h}, \\ &\big|a_2(c_1, c, l) - a_2(c_2, c, l)\big| \le \hat{C}_{\mathrm{Lip}} \|c_1 - c_2\|_{1,\Omega} \|c\|_{W^{1,\infty}(\Omega)} \|l\|_{1,\Omega}, \end{aligned} \tag{4.11}$$

where the constant $\tilde{C}_{\mathrm{Lip}} > 0$ is independent of h (cf. [49]). A related result follows for $c_1^h(\cdot;\cdot,\cdot)$ as in [137, Lemma 3.4]. On the other hand, let $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{u} \in \boldsymbol{H}^2(\mathcal{T}_h)$ and $\boldsymbol{v} \in V_h$. Then there exists $\tilde{C}_u > 0$ independently of h such that

$$\big|c_1^h(\boldsymbol{w}_1; \boldsymbol{u}, \boldsymbol{v}) - c_1^h(\boldsymbol{w}_2; \boldsymbol{u}, \boldsymbol{v})\big| \le \tilde{C}_u \|\boldsymbol{w}_1 - \boldsymbol{w}_2\|_{1,\mathcal{T}_h} \|\boldsymbol{u}\|_{1,\mathcal{T}_h} \|\boldsymbol{v}\|_{1,\mathcal{T}_h}. \tag{4.12}$$

Moreover, while the coercivity of the form $a_2(\cdot,\cdot,\cdot)$ is readily implied by (4.8),

$$a_2(\cdot, c, c) \ge \hat{\alpha}_a \|c\|^2_{1,\Omega} \quad \text{for all } c \in H^1(\Omega), \tag{4.13}$$

there also holds (cf. [109, Lemma 3.2])

$$a_1^h(\cdot, \boldsymbol{v}, \boldsymbol{v}) \ge \tilde{\alpha}_a \|\boldsymbol{v}\|^2_{1,\mathcal{T}_h} \quad \text{for all } \boldsymbol{v} \in \boldsymbol{V}_h, \tag{4.14}$$

provided that $a_0 > 0$ is sufficiently large and independent of the meshsize.

Furthermore, based on the assumptions on D, we have

$$\big|a_2(\cdot; c, l)\big| \le \hat{C}_a \|c\|_{1,\Omega} \|l\|_{1,\Omega} \quad \text{for all } c, l \in H^1(\Omega). \tag{4.15}$$

In addition, if we let $\boldsymbol{w} \in \boldsymbol{H}_0(\mathrm{div}^0; \Omega) := \{\boldsymbol{w} \in \boldsymbol{H}(\mathrm{div}, \Omega) : \boldsymbol{w} \cdot \boldsymbol{n} = 0 \text{ on } \partial\Omega, \mathrm{div}\, \boldsymbol{w} = \boldsymbol{0} \text{ in } \Omega\}$, then according to [137] we can write

$$c_1^h(\boldsymbol{w}; \boldsymbol{u}, \boldsymbol{u}) = \frac{1}{2} \sum_{e \in \mathcal{E}_h^i} \int_e |\boldsymbol{w} \cdot \boldsymbol{n}_e| [\![\boldsymbol{u}]\!]^2 \ge 0 \quad \text{for all } \boldsymbol{u} \in \boldsymbol{V}_h, \tag{4.16}$$

as well as the relation

$$c_2(\boldsymbol{w}; l_h, l_h) = 0 \quad \text{for all } l_h \in \mathcal{M}_h, \tag{4.17}$$

which arises from integration by parts and holds at the discrete level since the produced discrete velocities are exactly divergence free. Based on the assumptions on f_{bk}, it is also clear that

$$\left| d_2(c_1, l) - d_2(c_2, l) \right| \leq C_d \|c_1 - c_2\|_{0,\Omega} \|l\|_{1,\Omega}. \tag{4.18}$$

Finally, we recall from [109] the following discrete inf-sup condition for $b(\cdot,\cdot)$, where $\tilde{\beta}$ is independent of h:

$$\sup_{\boldsymbol{v}_h \in \boldsymbol{V}_h \setminus \{\boldsymbol{0}\}} \frac{b(\boldsymbol{v}_h, q_h)}{\|\boldsymbol{v}_h\|_{1,\mathcal{T}_h}} \geq \tilde{\beta} \|q_h\|_{0,\Omega} \quad \text{for all } q_h \in \mathcal{Q}_h. \tag{4.19}$$

Remark 4.1. *Using the definition and characterisation of the kernel $\boldsymbol{Z}$ of $b(\cdot,\cdot)$, namely*

$$\boldsymbol{Z} := \left\{ \boldsymbol{v} \in \boldsymbol{H}_0^1(\Omega) : b(\boldsymbol{v}, q) = 0 \; \forall q \in L_0^2(\Omega) \right\} = \left\{ \boldsymbol{v} \in \boldsymbol{H}_0^1(\Omega) : \operatorname{div} \boldsymbol{v} = 0 \text{ in } \Omega \right\},$$

and using integration by parts, we can readily observe that

$$c_1(\boldsymbol{w}; \boldsymbol{v}, \boldsymbol{v}) = 0 \quad \text{and} \quad c_2(\boldsymbol{w}; s, s) = 0 \quad \text{for all } \boldsymbol{w} \in \boldsymbol{Z},\ \boldsymbol{v} \in \boldsymbol{H}^1(\Omega), \text{ and } s \in H^1(\Omega).$$

It is also well known (see for instance [41]) that if $(\boldsymbol{u}, p, c) \in \boldsymbol{H}_0^1(\Omega) \times L_0^2 \times H^1$ solves (4.6), then $\boldsymbol{u} \in \boldsymbol{Z}$ is a solution of the following reduced problem:

For all $t \in (0, \mathcal{T}]$, find $(\boldsymbol{u}, c) \in \boldsymbol{Z} \times H^1$ such that

$$\begin{aligned} &\big(\partial_t \boldsymbol{u}(t), \boldsymbol{v}\big)_\Omega + a_1\big(c(t); \boldsymbol{u}(t), \boldsymbol{v}\big) + c_1\big(\boldsymbol{u}(t); \boldsymbol{u}(t), \boldsymbol{v}\big) = \boldsymbol{F}_{\mathrm{g}}(c, \boldsymbol{v}) + \boldsymbol{F}_{\mathrm{r}}(\boldsymbol{u}, \boldsymbol{v}) \quad \text{for all } \boldsymbol{v} \in \boldsymbol{H}_0^1(\Omega), \\ &\big(\partial_t c(t), l\big)_\Omega + a_2\big(c(t); c(t), l\big) + c_2\big(\boldsymbol{u}(t); c(t), l\big) - d_2\big(c(t), l\big) = -G_{\mathrm{r}}(c(t), l) \quad \text{for all } l \in H^1(\Omega). \end{aligned} \tag{4.20}$$

Conversely, if $(\boldsymbol{u}, c) \in \boldsymbol{Z} \times H^1$ is a solution of (4.20), then there exists a pressure $p \in L_0^2$ such that $(\boldsymbol{u}, p, c)$ is a solution of (4.6). As in the continuous case, we define the discrete kernel of the bilinear form $b(\cdot,\cdot)$ as

$$\boldsymbol{Z}_h := \left\{ \boldsymbol{v}_h \in \boldsymbol{V}_h : b(\boldsymbol{v}_h, q_h) = 0 \; \forall q_h \in \mathcal{Q}_h \right\} = \left\{ \boldsymbol{v}_h \in \boldsymbol{V}_h : \operatorname{div} \boldsymbol{v}_h = 0 \text{ in } \Omega \right\},$$

and relying on the inf-sup condition (4.19), we can introduce an equivalent discrete reduce problem.

Let us denote by $\mathcal{I}_h : C(\bar{\Omega}) \to \mathcal{M}_h$ the classical nodal interpolation operator with respect to a unisolvent set of Lagrangian interpolation nodes associated with the conforming space $\mathcal{M}_h$. By $\Pi_h \boldsymbol{u}$ we denote the BDM projection of $\boldsymbol{u}$, and $\mathcal{L}_h p$ is the L^2-projection of p onto $\mathcal{Q}_h$. Under adequate regularity assumptions, the following approximation properties hold (see [109]):

$$\begin{aligned} &\|\boldsymbol{u} - \Pi_h \boldsymbol{u}\|_{1,\mathcal{T}_h} \leq C^* h^{k+1} \|\boldsymbol{u}\|_{k+1,\Omega}, \\ &\|c - \mathcal{I}_h c\|_{1,\Omega} \leq C^* h^k \|c\|_{k+1,\Omega}, \quad \|p - \mathcal{L}_h p\|_{0,\Omega} \leq C^* h^k \|p\|_{k,\Omega}. \end{aligned} \tag{4.21}$$

The following development follows the structure adopted in [4].

Lemma 4.2. *Assume that* $\boldsymbol{u} \in \boldsymbol{H}^2(\Omega)$, $p \in L^2(\Omega)$ *and* $c \in H^1(\Omega)$. *Then we have*

For all $\boldsymbol{v} \in \boldsymbol{V}_h$, $q \in \mathcal{Q}_h$ and $l \in \mathcal{M}_h$:
$\big(\partial_t \boldsymbol{u}(t), \boldsymbol{v}\big)_\Omega + a_1^h\big(c(t); \boldsymbol{u}(t), \boldsymbol{v}\big) + c_1^h\big(\boldsymbol{u}(t); \boldsymbol{u}(t), \boldsymbol{v}\big) - b(\boldsymbol{v}, p(t)) - \boldsymbol{F}_{\mathrm{g}}\big(c(t), \boldsymbol{v}\big) - \boldsymbol{F}_{\mathrm{r}}\big(\boldsymbol{u}(t), \boldsymbol{v}\big) = 0,$
$b(\boldsymbol{u}(t), q) = 0,$
$\big(\partial_t c(t), l\big)_\Omega + a_2\big(c(t); c(t), l\big) + c_2\big(\boldsymbol{u}(t); c(t), l\big) - d_2\big(c(t), l\big) = -G_{\mathrm{r}}\big(c(t), l\big).$

Proof. Since we assume $\boldsymbol{u} \in \boldsymbol{H}^2(\Omega)$, integration by parts yields the required result. See also [19]. The third equation is a straightforward result from the continuous form. □

Now we decompose the errors as follows:

$$\begin{aligned} \boldsymbol{u}_h - \boldsymbol{u} &= E_{\boldsymbol{u}} + \xi_{\boldsymbol{u}} = (\Pi_h \boldsymbol{u} - \boldsymbol{u}) + (\boldsymbol{u}_h - \Pi_h \boldsymbol{u}), \\ p_h - p &= E_p + \xi_p = (\mathcal{L}_h p - p) + (p_h - \mathcal{L}_h p), \\ c_h - c &= E_c + \xi_c = (\mathcal{I}_h c - c) + (c_h - \mathcal{I}_h c). \end{aligned}$$

Assuming that $\boldsymbol{u}_h^0 = \Pi_h \boldsymbol{u}(0)$ and $c_h^0 = \mathcal{I}_h c(0)$, we will use also the notation $E_{\boldsymbol{u}}^n = (\boldsymbol{u}(t_n) - \Pi_h \boldsymbol{u}(t_n))$ and $\xi_{\boldsymbol{u}}^n = (\Pi_h \boldsymbol{u}(t_n) - \boldsymbol{u}_h^n)$, and similar notation for other variables. Note that for the first time iteration of the fully discrete form of system (4.7) we adopt a backward Euler scheme, and so we require error estimates for this step.

In what follows we assume a simpler form for the drag term $\boldsymbol{f}_{\mathrm{r}}$ such that for all $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{v} \in \boldsymbol{H}^1(\Omega)$ we have the following Lipschitz continuity:

$$\big|\boldsymbol{F}_{\mathrm{r}}(\boldsymbol{u}_1, \boldsymbol{v}) - \boldsymbol{F}_{\mathrm{r}}(\boldsymbol{u}_2, \boldsymbol{v})\big| \le \tilde{\gamma}_1 \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_{0,\mathcal{B}} \|\boldsymbol{v}\|_{0,\mathcal{B}}. \tag{4.22}$$

Since $\boldsymbol{X}_{\mathrm{r}}(\boldsymbol{s}, t)$ is a rigid motion, (4.22) can be achieved, for instance, if we consider

$$\boldsymbol{f}_{\mathrm{r}}(\boldsymbol{x}, t) = \beta^* \rho_{\mathrm{r}} \int_{\mathcal{B}} \big((\boldsymbol{u}_{\mathrm{r}}(\boldsymbol{s}) - \boldsymbol{u}(\boldsymbol{X}_{\mathrm{r}}(\boldsymbol{s}, t), t)) \cdot \boldsymbol{n}_{\mathrm{r}}\big) \boldsymbol{n}_{\mathrm{r}} \delta(\boldsymbol{x} - \boldsymbol{X}_{\mathrm{r}}(\boldsymbol{s}, t))\, \mathrm{d}\boldsymbol{s}. \tag{4.23}$$

Furthermore, since $\mathcal{B} \subset \Omega$, we have that $\|\cdot\|_{0,\mathcal{B}} \le \|\cdot\|_{0,\Omega}$ and

$$\big|\boldsymbol{F}_{\mathrm{r}}(\boldsymbol{u}_1, \boldsymbol{v}) - \boldsymbol{F}_{\mathrm{r}}(\boldsymbol{u}_2, \boldsymbol{v})\big| \le \gamma_1 \|\boldsymbol{u}_1 - \boldsymbol{u}_2\|_{0,\Omega} \|\boldsymbol{v}\|_{0,\Omega}. \tag{4.24}$$

By Hölder's inequality for all $c, c_1, c_2, l \in H^1(\Omega)$ and $\boldsymbol{v} \in \boldsymbol{H}^1(\Omega)$ there also hold

$$\boldsymbol{F}_{\mathrm{g}}(c, \boldsymbol{v}) \le \gamma_2 \|c\|_{0,\Omega} \|\boldsymbol{v}\|_{0,\Omega}, \tag{4.25}$$

$$G_{\mathrm{r}}(c_1, l) - G_{\mathrm{r}}(c_2, l) \le \gamma_3 \|c_1 - c_2\|_{0,\Omega} \|l\|_{0,\Omega}. \tag{4.26}$$

The following algebraic relation will be useful in the sequel: for any real numbers a^{n+1}, a^n, a^{n-1} and defining $\Lambda a^n := a^{n+1} - 2a^n + a^{n-1}$, we have

$$2(3a^{n+1} - 4a^n + a^{n-1}, a^n) = |a^{n+1}|^2 + |2a^{n+1} - a^n|^2 + |\Lambda a^n|^2 - |a^n|^2 - |2a^n - a^{n-1}|^2. \tag{4.27}$$

Theorem 4.1. *Let $(\boldsymbol{u}_h^n, c_h^n)$ in $\boldsymbol{X}_h \times \mathcal{M}_h$ be a solution of problem (4.7), using the second-order backward differentiation formula with initial data $(\boldsymbol{u}_h^1, c_h^1)$ and $(\boldsymbol{u}_h^0, c_h^0)$. Then there exist constants $\bar{C}_u > 0$ and $\bar{C}_c > 0$ that are independent of h and Δt such that*

$$\begin{aligned}
&\sup_{2\le n\le N} \|\boldsymbol{u}_h^n\|_{0,\Omega}^2 + \sup_{2\le n\le N} \|2\boldsymbol{u}_h^n - \boldsymbol{u}_h^{n-1}\|_{0,\Omega}^2 + \sum_{n=2}^{N} \|\Lambda \boldsymbol{u}_h^{n-1}\|_{0,\Omega} + \sum_{n=2}^{N} \Delta t \tilde{\alpha}_a \|\boldsymbol{u}_h^n\|_{1,\mathcal{T}_h}^2 \\
&\le \bar{C}_{\boldsymbol{u}} (\|c_h^1\|_{0,\Omega}^2 + \|2c_h^1 - c_h^0\|_{0,\Omega}^2 + \|\boldsymbol{u}_h^1\|_{0,\Omega}^2 + \|2\boldsymbol{u}_h^1 - \boldsymbol{u}_h^0\|_{0,\Omega}^2 + \|\boldsymbol{u}_{\mathrm{r}}\|_{0,\mathcal{B}}^2 + |c_{\mathrm{r}}|^2), \\
&\sup_{2\le n\le N} \|c_h^n\|_{0,\Omega}^2 + \sup_{2\le n\le N} \|2c_h^n - c_h^{n-1}\|_{0,\Omega}^2 + \sum_{n=2}^{N} \|\Lambda c_h^{n-1}\|_{0,\Omega} + 4\sum_{n=2}^{N} \Delta t \hat{\alpha}_a \|c_h^n\|_{1,\Omega}^2 \\
&\le \bar{C}_c (\|c_h^1\|_{0,\Omega}^2 + \|2c_h^1 - c_h^0\|_{0,\Omega}^2 + |c_{\mathrm{r}}|^2).
\end{aligned}$$

Proof. It suffices to take $\boldsymbol{v}_h = 4\Delta t \boldsymbol{u}_h^{n+1}$ and $l_h = 4\Delta t c_h^{n+1}$ in system (4.7), using BDF2 differentiation formula, Sobolev inequalities, summing over n from 1 to $n \le N-1$, and applying Gronwall's lemma, with Δt sufficiently small. Note that by Remark 4.1, all terms containing the bilinear form b are simply removed from the system. □

Theorem 4.2. *Assume that $\boldsymbol{u} \in L^\infty(0,\mathcal{T}; H_0^{k+1}(\Omega))$, $\boldsymbol{u}' \in L^\infty(0,\mathcal{T}; \boldsymbol{H}^1(\Omega))$, $\boldsymbol{u}'' \in L^\infty(0,\mathcal{T}; \boldsymbol{L}^2(\Omega))$, $p \in L^\infty(0,\mathcal{T}; H^k(\Omega))$, $c \in L^\infty(0,\mathcal{T}; H_0^{k+1}(\Omega))$, $c' \in L^\infty(0,\mathcal{T}; H^k(\Omega))$, $c'' \in L^\infty(0,\mathcal{T}; L^2(\Omega))$, with $\gamma_2^2 \le \frac{1}{32}\hat{\alpha}_a \tilde{\alpha}_a^2$, $k \ge 1$ and also that*

$$\max\big\{\|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T};\boldsymbol{W}^{1,\infty}(\Omega))}, \|c\|_{L^\infty(0,T;W^{1,\infty}(\Omega))}\big\} < M,$$

for a sufficiently small constant $M > 0$ (a precise condition for M, can be found on Theorem 4.5). Then there exist positive constants C_u^1, C_c^1, independent of h and Δt, such that

$$\frac{1}{4}\|\xi_{\boldsymbol{u}}^1\|_{0,\Omega}^2 + \frac{1}{4}\Delta t \tilde{\alpha}_a \|\xi_{\boldsymbol{u}}^1\|_{1,\mathcal{T}_h}^2 \le C_u^1 (h^{2k} + \Delta t^4), \quad \frac{1}{8}\|\xi_c^1\|_{0,\Omega}^2 + \frac{1}{4}\Delta t \hat{\alpha}_a \|\xi_c^1\|_{H^1(\Omega)}^2 \le C_c^1 (h^{2k} + \Delta t^4).$$

Proof. First, taking into account the regularity assumptions for $\boldsymbol{u}$, we have for all $\boldsymbol{x}$ a $\gamma \in (0,1)$ that depends on $\boldsymbol{x}$ such that

$$\boldsymbol{u}(0) = \boldsymbol{u}(\Delta t) - \Delta t \boldsymbol{u}'(\Delta t) + \frac{1}{2}\Delta t^2 \boldsymbol{u}''(\Delta t \gamma),$$

then using the reduced problem as stated on Remark 4.1, $\boldsymbol{u}$ satisfies the following error equation

$$\begin{aligned}
\|\xi_{\boldsymbol{u}}^1\|_{0,\Omega}^2 + \Delta t \tilde{\alpha}_a \|\xi_{\boldsymbol{u}}^1\|_{1,\mathcal{T}_h}^2 \le &-\big(\Pi_h \boldsymbol{u}(\Delta t) - \boldsymbol{u}(\Delta t) + \boldsymbol{u}_h^0 - \boldsymbol{u}(0), \xi_{\boldsymbol{u}}^1\big)_\Omega \\
&+ \Delta t \big(a_1^h(c_h^1; \Pi_h \boldsymbol{u}(\Delta t), \xi_{\boldsymbol{u}}^1) - a_1^h(c^1; \boldsymbol{u}(\Delta t), \xi_{\boldsymbol{u}}^1)\big) \\
&- \Delta t \big(c_1^h(\boldsymbol{u}_h^1; \boldsymbol{u}_h^1, \xi_{\boldsymbol{u}}^1) - c_1^h(\boldsymbol{u}(\Delta t), \boldsymbol{u}(\Delta t), \xi_{\boldsymbol{u}}^1)\big) \\
&- \Delta t \big(\boldsymbol{F}_{\mathrm{r}}(\boldsymbol{u}_h^1, \xi_{\boldsymbol{u}}^1) - \boldsymbol{F}_{\mathrm{r}}(\boldsymbol{u}(\Delta t), \xi_{\boldsymbol{u}}^1)\big) \\
&- \Delta t \big(\boldsymbol{F}_{\mathrm{g}}(c_h^1, \xi_{\boldsymbol{u}}^1) - \boldsymbol{F}_{\mathrm{g}}(c(\Delta t), \xi_{\boldsymbol{u}}^1)\big) - \frac{\Delta t^2}{2}\big(\boldsymbol{u}''(\Delta t \gamma), \xi_{\boldsymbol{u}}^1\big),
\end{aligned}$$

which results after choosing $\xi_{\boldsymbol{u}}^1$ as test function in the first equation of Lemma 4.2 and system (4.7), performing an Euler scheme step, subtracting both equations, and adding $\pm a_1^h(c_h^1; \Pi_h \boldsymbol{u}(\Delta t), \xi_{\boldsymbol{u}}^1)$.

Now, by applying the error approximation results from (4.21), Young's inequality, and the stability properties, we get

$$\begin{aligned}\frac{1}{4}\|\xi_{\boldsymbol{u}}^1\|_{0,\Omega}^2+\frac{1}{4}\Delta t\tilde{\alpha}_a\|\xi_{\boldsymbol{u}}^1\|_{1,\mathcal{T}_h}^2 \leq Ch^{2k}\Delta t\Big(&\|\boldsymbol{u}(\Delta t)\|_{\boldsymbol{H}^{k+1}(\Omega)}^2+\|\boldsymbol{u}(0)\|_{\boldsymbol{H}^{k+1}(\Omega)}^2\\ &+\|c(\Delta t)\|_{H^{k+1}(\Omega)}^2+\|p(\Delta t)\|_{H^k(\Omega)}^2\Big)\\ &+C\Delta t^4(\|\boldsymbol{u}''\|_{L^\infty(0,\mathcal{T};\boldsymbol{L}^2(\Omega))}^2)+\frac{4\tilde{C}_{\mathrm{Lip}}^2M^2}{\tilde{\alpha}_a}\Delta t\|\xi_c^1\|_{1,\Omega}^2\\ &+\Delta t\frac{\gamma_2^2}{\tilde{\alpha}_a}\|\xi_c^1\|_{1,\Omega}.\end{aligned}\tag{4.28}$$

Next, we choose ξ_c^1 as test function in the third equation of Lemma 4.2 and system (4.7); we follow the same steps as before, adding to the sum of both equations the term $\pm a_2(c_h^1;\mathcal{I}_h\,c^1,\xi_c^1)$, with Δt sufficiently small $\left(\Delta t\leq\frac{1}{2(12C_d^2+2\tilde{\gamma}_3^2)}\right)$ to obtain

$$\begin{aligned}\frac{1}{4}\|\xi_c^1\|_{0,\Omega}^2+\frac{1}{2}\Delta t\hat{\alpha}_a\|\xi_c^1\|_{1,\Omega}^2 \leq C\Delta th^{2k}\Big(&\|\boldsymbol{u}(\Delta t)\|_{\boldsymbol{H}^{k+1}(\Omega)}^2+\|c(\Delta t)\|_{H^{k+1}(\Omega)}^2+\|c(0)\|_{H^{k+1}(\Omega)}^2\\ &+\|c(\Delta t)\|_{H^{k+1}(\Omega)}^2\|\boldsymbol{u}(\Delta t)\|_{H^1(\Omega)}^2+\|\boldsymbol{u}(\Delta t)\|_{H^{k+1}(\Omega)}^2\|c(\Delta t)\|_{H^1(\Omega)}^2\Big)\\ &+C\Delta t^4(\|c''\|_{L^\infty(0,\Delta t;L^2(\Omega))}^2)+\frac{6\tilde{C}_1^2(1+C^*)^2M^2}{\hat{\alpha}_a}\Delta t\|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h}^2.\end{aligned}\tag{4.29}$$

In this way, from (4.28) we deduce that

$$\tau\|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h}^2\leq C(h^{2k}+\Delta t^4)+\frac{16\tilde{C}_{\mathrm{Lip}}^2M^2}{\tilde{\alpha}_a^2}\tau\|\xi_c^1\|_{1,\Omega}^2+4\frac{\gamma_2^2}{\tilde{\alpha}_a^2}\Delta t\|\xi_c^1\|_{1,\Omega}^2.$$

We insert the previous identity into (4.29) and consider M sufficiently small such that the terms multiplying $\|\xi_c\|_{1,\Omega}^2$, can be absorbed into the left-hand side of the inequality, to get

$$\frac{1}{8}\|\xi_c^1\|_{0,\Omega}^2+\frac{1}{4}\Delta t\hat{\alpha}_a\|\xi_c\|_{1,\Omega}^2\leq C_c^1(h^{2k}+\Delta t^4).\tag{4.30}$$

The first estimate follows by directly substituting (4.30) into (4.28). □

Theorem 4.3. *Let $(\boldsymbol{u},p,c)$ be the solution of (4.6) and $(\boldsymbol{u}_h,p_h,c_h)$ be the solution of (4.7) with BDF2 iteration. Suppose that $\boldsymbol{u}\in L^\infty(0,\mathcal{T};\boldsymbol{H}_0^{k+1}(\Omega))$, $c\in L^\infty(0,\mathcal{T};H_0^{k+1}(\Omega))$, $\boldsymbol{u}'\in L^\infty(0,\mathcal{T};\boldsymbol{H}^k(\Omega))$, $\boldsymbol{u}^{(3)}\in L^2(0,\mathcal{T};\boldsymbol{L}^2(\Omega))$ and $\|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T};\boldsymbol{W}^{1,\infty}(\Omega))}<M$ for a sufficiently small constant $M>0$. Then there exist positive constants $C,\eta_1\geq 0$ independent of h and Δt such that for all $m+1\leq N$,*

$$\begin{aligned}\|\xi_{\boldsymbol{u}}^{m+1}\|_{0,\Omega}^2+\|2\xi_{\boldsymbol{u}}^{m+1}-\xi_{\boldsymbol{u}}^m\|_{0,\Omega}^2+\sum_{n=1}^m\|\Lambda\xi_{\boldsymbol{u}}^n\|_{0,\Omega}^2+\sum_{n=1}^m\Delta t\tilde{\alpha}_a\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2&\\ \leq C(\Delta t^4+h^{2k})+\sum_{n=1}^m\eta_1\Delta t\|\xi_c^{n+1}\|_{0,\Omega}^2.&\end{aligned}$$

Proof. We choose as tests functions $\boldsymbol{v}_h=\xi_{\boldsymbol{u}}^{n+1}$ in the first equation of (4.7), using BDF2 differentiation formula and inserting the terms

$$\pm\frac{1}{2\Delta t}\left(3\boldsymbol{u}(t_{n+1})-4\boldsymbol{u}(t_n)+\boldsymbol{u}(t_{n-1}),\xi_{\boldsymbol{u}}^{n+1}\right)_\Omega,\quad \pm\frac{1}{2\Delta t}\left(3\Pi_h\,\boldsymbol{u}(t_{n+1})-4\Pi_h\,\boldsymbol{u}(t_n)+\Pi_h\,\boldsymbol{u}(t_{n-1}),\xi_{\boldsymbol{u}}^{n+1}\right)_\Omega,$$

and $\pm a_1^h(c_h^{n+1}; \Pi_h\, \boldsymbol{u}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1})$, we get

$$\begin{aligned}
&\frac{1}{2\Delta t}\left(3\xi_{\boldsymbol{u}}^{n+1} - 4\xi_{\boldsymbol{u}}^{n} + \xi_{\boldsymbol{u}}^{n-1}, \xi_{\boldsymbol{u}}^{n+1}\right)_{\Omega} + \frac{1}{2\Delta t}\left(3E_{\boldsymbol{u}}^{n+1} - 4E_{\boldsymbol{u}}^{n} + E_{\boldsymbol{u}}^{n-1}, \xi_{\boldsymbol{u}}^{n+1}\right)_{\Omega} \\
&+ \frac{1}{2\Delta t}\left(3\boldsymbol{u}(t_{n+1}) - 4\boldsymbol{u}(t_n) + \boldsymbol{u}(t_{n-1}), \xi_{\boldsymbol{u}}^{n+1}\right)_{\Omega} + a_1^h(c_h^{n+1}; \xi_{\boldsymbol{u}}^{n+1}, \xi_{\boldsymbol{u}}^{n+1}) + a_1^h(c_h^{n+1}; \Pi_h\, \boldsymbol{u}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1}) \\
&+ c_1^h(\boldsymbol{u}_h^{n+1}, \boldsymbol{u}_h^{n+1}, \xi_{\boldsymbol{u}}^{n+1}) = \boldsymbol{F}_{\mathrm{g}}(c_h^{n+1}, \xi_{\boldsymbol{u}}^{n+1}) + \boldsymbol{F}_{\mathrm{r}}(\boldsymbol{u}_h^{n+1}, \xi_{\boldsymbol{u}}^{n+1}). \qquad (4.31)
\end{aligned}$$

Considering Lemma 4.2 at $t = t_{n+1}$ with $\boldsymbol{v} = \xi_{\boldsymbol{u}}^{n+1}$, and after inserting the term

$$\pm\frac{1}{2\Delta t}\left(3\boldsymbol{u}(t_{n+1}) - 4\boldsymbol{u}(t_n) + \boldsymbol{u}(t_{n-1}), \xi_{\boldsymbol{u}}^{n+1}\right)_{\Omega},$$

we readily deduce the expression

$$\begin{aligned}
&\frac{1}{2\Delta t}\left(3\boldsymbol{u}(t_{n+1}) - 4\boldsymbol{u}(t_n) + \boldsymbol{u}(t_{n-1}), \xi_{\boldsymbol{u}}^{n+1}\right)_{\Omega} + a_1^h(c(t_{n+1}); \boldsymbol{u}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1}) + c_1^h(\boldsymbol{u}(t_{n+1}), \boldsymbol{u}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1}) \\
&= \boldsymbol{F}_{\mathrm{g}}(c^{n+1}, \xi_{\boldsymbol{u}}^{n+1}) + \boldsymbol{F}_{\mathrm{r}}(\boldsymbol{u}^{n+1}, \xi_{\boldsymbol{u}}^{n+1}) - \left(\boldsymbol{u}'(t_{n+1}) - \frac{3\boldsymbol{u}(t_{n+1}) - 4\boldsymbol{u}(t_n) + \boldsymbol{u}(t_{n-1})}{2\Delta t}, \xi_{\boldsymbol{u}}^{n+1}\right)_{\Omega}. \qquad (4.32)
\end{aligned}$$

We can then subtract (4.31) from (4.32) and multiply both sides by $4\Delta t$ to obtain an equality

$$I_1 + I_2 = I_3 + I_4 + I_5 + I_6 + I_7 + I_8,$$

where we define

$$\begin{aligned}
I_1 &:= 2\left(3\xi_{\boldsymbol{u}}^{n+1} - 4\xi_{\boldsymbol{u}}^{n} + \xi_{\boldsymbol{u}}^{n-1}, \xi_{\boldsymbol{u}}^{n+1}\right), \\
I_2 &:= 4\Delta t a_1^h(c_h^{n+1}; \xi_{\boldsymbol{u}}^{n+1}, \xi_{\boldsymbol{u}}^{n+1}), \\
I_3 &:= 4\Delta t\left(\boldsymbol{u}'(t_{n+1}) - \frac{3\boldsymbol{u}(t_{n+1}) - 4\boldsymbol{u}(t_n) + \boldsymbol{u}(t_{n-1})}{2\Delta t}, \xi_{\boldsymbol{u}}^{n+1}\right)_{\Omega}, \\
I_4 &:= -2\left(3E_{\boldsymbol{u}}^{n+1} - 4E_{\boldsymbol{u}}^{n} + E_{\boldsymbol{u}}^{n-1}, \xi_{\boldsymbol{u}}^{n+1}\right), \\
I_5 &:= 4\Delta t\big(\boldsymbol{F}_{\mathrm{g}}(c_h^{n+1}, \xi_{\boldsymbol{u}}^{n+1}) - \boldsymbol{F}_{\mathrm{g}}(c(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1}\big), \\
I_6 &:= 4\Delta t\big(\boldsymbol{F}_{\mathrm{r}}(\boldsymbol{u}_h^{n+1}, \xi_{\boldsymbol{u}}^{n+1}) - \boldsymbol{F}_{\mathrm{r}}(\boldsymbol{u}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1}\big), \\
I_7 &:= -4\Delta t\big(a_1^h(c_h^{n+1}; \Pi_h\, \boldsymbol{u}^{n+1}, \xi_{\boldsymbol{u}}^{n+1}) - a_1^h(c(t_{n+1}); \boldsymbol{u}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1})\big), \\
I_8 &:= -4\Delta t\big(c_1^h(\boldsymbol{u}_h^{n+1}, \boldsymbol{u}_h^{n+1}, \xi_{\boldsymbol{u}}^{n+1}) - c_1^h(\boldsymbol{u}(t_{n+1}), \boldsymbol{u}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1})\big).
\end{aligned}$$

Let us estimate each term I_i, $i \in \{1, \ldots, 8\}$. For I_1, using (4.27) we can assert that

$$I_1 = \|\xi_{\boldsymbol{u}}^{n+1}\|_{0,\Omega}^2 + \|2\xi_{\boldsymbol{u}}^{n+1} - \xi_{\boldsymbol{u}}^{n}\|_{0,\Omega}^2 + \|\Lambda\xi_{\boldsymbol{u}}^{n+1}\|_{0,\Omega}^2 - \|\xi_{\boldsymbol{u}}^{n}\|_{0,\Omega}^2 - \|2\xi_{\boldsymbol{u}}^{n} - \xi_{\boldsymbol{u}}^{n-1}\|_{0,\Omega}^2.$$

Using the ellipticity stated in (4.14), we readily get

$$I_2 \geq 4\Delta t\tilde{\alpha}_a\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2.$$

By using Taylor's formula with integral remainder we have

$$\left|\boldsymbol{u}'(t_{n+1}) - \frac{3\boldsymbol{u}(t_{n+1}) - 4\boldsymbol{u}(t_n) + \boldsymbol{u}(t_{n-1})}{2\Delta t}\right| = \frac{\Delta t^{3/2}}{2\sqrt{3}}\|\boldsymbol{u}^{(3)}\|_{L^2(t^{n-1}, t^{n+1}; \boldsymbol{L}^2(\Omega))},$$

then by combining Cauchy-Schwarz and Young's inequality, we obtain the bound

$$|I_3| \leq \frac{\Delta t^4}{24\varepsilon_1}\|\boldsymbol{u}^{(3)}\|^2_{L^2(t_{n-1},t_{n+1};\boldsymbol{L}^2(\Omega))} + \frac{\Delta t\varepsilon_1}{2}\|\xi^{n+1}_{\boldsymbol{u}}\|^2_{1,\mathcal{T}_h}.$$

Now we insert $\pm 4\Delta t E'_{\boldsymbol{u}}(t_{n+1})$ onto the fourth term, which leads to

$$I_4 = -4\Delta t(E'_{\boldsymbol{u}}(t_{n+1}), \xi^{n+1}_{\boldsymbol{u}})_\Omega + \left(E'_{\boldsymbol{u}}(t_{n+1}) - \frac{3E^{n+1}_{\boldsymbol{u}} - 4E^n_{\boldsymbol{u}} + E^{n-1}_{\boldsymbol{u}}}{2\Delta t}, \xi^{n+1}_{\boldsymbol{u}}\right)_\Omega.$$

Proceeding as before and using (4.21) on the first term of I_4, we get

$$|I_4| \leq \frac{C}{2\varepsilon_2}h^{2k}\|\boldsymbol{u}'\|^2_{\boldsymbol{L}^\infty(0,\mathcal{T};\boldsymbol{H}^k(\Omega))} + \frac{\Delta t\varepsilon_2}{2}\|\xi^{n+1}_{\boldsymbol{u}}\|^2_{1,\mathcal{T}_h} + \frac{\Delta t^4 C}{2\varepsilon_3}\|\boldsymbol{u}^{(3)}\|^2_{L^2(0,\mathcal{T};\boldsymbol{L}^2(\Omega))} + \frac{\Delta t\varepsilon_3}{2}\|\xi^{n+1}_{\boldsymbol{u}}\|^2_{1,\mathcal{T}_h}.$$

Now by (4.25), appealing to (4.21), and inserting $\pm 4\Delta t\boldsymbol{F}_{\mathrm{g}}(\mathcal{I}_h\, c^{n+1}, \xi^{n+1}_{\boldsymbol{u}})$, we are left with

$$|I_5| \leq \frac{2\gamma_2^2\Delta t}{\varepsilon_4}\left(C^* h^{2k}\|c\|^2_{L^\infty(0,\mathcal{T};H^k(\Omega))} + \|\xi^{n+1}_c\|^2_{0,\Omega}\right) + 2\Delta t\varepsilon_4\|\xi^{n+1}_{\boldsymbol{u}}\|^2_{1,\mathcal{T}_h}.$$

In the same manner using (4.24), and inserting $\pm 4\Delta t\boldsymbol{F}_{\mathrm{r}}(\Pi_h\, \boldsymbol{u}^{n+1}, \xi^{n+1}_{\boldsymbol{u}})$, we get

$$|I_6| \leq \frac{2\gamma_1^2\Delta t}{\varepsilon_5}\left(C^* h^{2k}\|\boldsymbol{u}\|^2_{L^\infty(0,\mathcal{T};\boldsymbol{H}^k(\Omega))} + \|\xi_{\boldsymbol{u}}\|^2_{0,\Omega}\right) + 2\Delta t\varepsilon_5\|\xi^{n+1}_{\boldsymbol{u}}\|^2_{1,\mathcal{T}_h}$$

Again inserting $\pm a_1^h(c_h^{n+1}; \boldsymbol{u}(t_{n+1}), \xi^{n+1}_{\boldsymbol{u}})$ and $\pm a_1^h(\mathcal{I}_h\, c^{n+1}; \boldsymbol{u}(t_{n+1}), \xi^{n+1}_{\boldsymbol{u}})$ and using (4.21) we get

$$\begin{aligned}|I_7| \leq{}& \frac{\tilde{C}_a^2\Delta t h^{2k}}{2\varepsilon_6}\|\boldsymbol{u}\|^2_{L^\infty(0,\mathcal{T};\boldsymbol{H}^{k+1}(\Omega))} + \frac{\Delta t\varepsilon_6}{2}\|\xi^{n+1}_{\boldsymbol{u}}\|^2_{1,\mathcal{T}_h} + \frac{\tilde{C}^2_{\mathrm{lip}}M^2}{2\varepsilon_7}\|\xi_c\|^2_{1,\Omega}\\ &+ \frac{\Delta t}{2}\varepsilon_7\|\xi^{n+1}_{\boldsymbol{u}}\|^2_{1,\mathcal{T}_h} + \frac{\tilde{C}^2_{\mathrm{lip}}M^2\Delta t h^{2k}}{2\varepsilon_8}\|c\|^2_{L^\infty(0,T;H^k(\Omega))} + \frac{\varepsilon_8\Delta t}{2}\|\xi^{n+1}_{\boldsymbol{u}}\|^2_{1,\mathcal{T}_h}.\end{aligned}$$

Now we insert into I_8 the three terms

$$\begin{aligned}&\pm c_1^h(\boldsymbol{u}(t_{n+1}), \Pi_h\, \boldsymbol{u}(t_{n+1}), \xi^{n+1}_{\boldsymbol{u}}), \quad \pm c_1^h(\Pi_h\, \boldsymbol{u}(t_{n+1}), \Pi_h\, \boldsymbol{u}(t_{n+1}), \xi^{n+1}_{\boldsymbol{u}}),\\ &\pm c_1^h(\Pi_h\, \boldsymbol{u}(t_{n+1}), \boldsymbol{u}(t_{n+1}), \xi^{n+1}_{\boldsymbol{u}}),\end{aligned}$$

which yields

$$\begin{aligned}I_8 = -4\Delta t\big(&c_1^h(\boldsymbol{u}(t_{n+1}), \Pi_h\, \boldsymbol{u}(t_{n+1}), \xi^{n+1}_{\boldsymbol{u}}) - c_1^h(\Pi_h\, \boldsymbol{u}(t_{n+1}), \Pi_h\, \boldsymbol{u}(t_{n+1}), \xi^{n+1}_{\boldsymbol{u}})\\ &+ c_1^h(\Pi_h\, \boldsymbol{u}(t_{n+1}), \Pi_h\, \boldsymbol{u}(t_{n+1}), \xi^{n+1}_{\boldsymbol{u}}) - c_1^h(\Pi_h\, \boldsymbol{u}(t_{n+1}), \boldsymbol{u}(t_{n+1}), \xi^{n+1}_{\boldsymbol{u}})\\ &+ c_1^h(\Pi_h\, \boldsymbol{u}(t_{n+1}), \boldsymbol{u}(t_{n+1}), \xi^{n+1}_{\boldsymbol{u}}) - c_1^h(\boldsymbol{u}(t_{n+1})\boldsymbol{u}(t_{n+1}), \xi^{n+1}_{\boldsymbol{u}}) + c_1^h(\boldsymbol{u}_h^{n+1}, \xi^{n+1}_{\boldsymbol{u}}, \xi^{n+1}_{\boldsymbol{u}})\big).\end{aligned}$$

The last term is moved to the left-hand side, where we use (4.16); whereas for the remaining terms (which we further rename as $\tilde{I}_8$), the bound (4.12) together with (4.21) imply that

$$\begin{aligned}|\tilde{I}_8| \leq 4\Delta t\Bigg(&C^*\tilde{C}_u C_\infty M\|\xi^{n+1}_{\boldsymbol{u}}\|^2_{1,\mathcal{T}_h} + \frac{h^{2k}C}{2\varepsilon_9}\|\boldsymbol{u}\|^2_{L^\infty(0,T;\boldsymbol{H}^1(\Omega))}\|\boldsymbol{u}\|^2_{L^\infty(0,\mathcal{T};\boldsymbol{H}^{k+1}(\Omega))} + \frac{\varepsilon_9}{2}\|\xi^{n+1}_{\boldsymbol{u}}\|^2_{1,\mathcal{T}_h}\\ &+ \frac{Ch^{2k}}{2\varepsilon_{10}}\|\boldsymbol{u}\|^2_{L^\infty(0,\mathcal{T};\boldsymbol{H}^{k+1}(\Omega))}\|\boldsymbol{u}\|^2_{L^\infty(0,\mathcal{T};\boldsymbol{H}^1(\Omega))} + \frac{\varepsilon_{10}}{2}\|\xi^{n+1}_{\boldsymbol{u}}\|^2_{1,\mathcal{T}_h}\Bigg),\end{aligned}$$

where C^* is a positive constant coming from (4.21). Hence, by choosing $\varepsilon_i = 2\tilde{\alpha}_a/11$ for $i = 1, \dots, 11$, collecting the above estimates, and summing over $1 \le n \le m$ for all $m+1 \le N$ we get

$$\begin{aligned}
&\|\xi_{\boldsymbol{u}}^{m+1}\|_{0,\Omega}^2 + \|2\xi_{\boldsymbol{u}}^{m+1} - \xi_{\boldsymbol{u}}^{m}\|_{0,\Omega}^2 + \sum_{n=1}^{m} \|\Lambda \xi_{\boldsymbol{u}}^{n}\|_{0,\Omega}^2 - 3\|\xi_{\boldsymbol{u}}^{1}\|_{0,\Omega}^2 + \sum_{n=1}^{m} \Delta t \tilde{\alpha}_a \|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2 \\
&\le C(\Delta t^4 + h^{2k}) + \eta_1 \sum_{n=1}^{m} \|\xi_c^{n+1}\|_{0,\Omega}^2 + \sum_{n=1}^{m} \frac{11\gamma_1^2 \Delta t}{\tilde{\alpha}_a} \|\xi_{\boldsymbol{u}}^{n+1}\|_{0,\Omega}^2,
\end{aligned}$$

where $\tilde{C}_u C^* C_\infty M \le \tilde{\alpha}_a/4$ and $\eta_1 = C(\tilde{\alpha}_a, \tilde{C}_{\mathrm{Lip}}, \gamma_1, \gamma_2)$. Finally, using Theorem 4.2, considering Δt sufficiently small and applying Gronwall's lemma, we get the desired result. □

Theorem 4.4. *Let $(\boldsymbol{u}, c)$ be the solution of (4.6) and $(\boldsymbol{u}_h, c_h)$ be the solution of (4.7) using the BDF2 differential operator. If $\boldsymbol{u} \in L^\infty(0,\mathcal{T}; \boldsymbol{H}_0^{k+1}(\Omega))$, $c \in L^\infty(0,\mathcal{T}; H_0^{k+1}(\Omega))$, $c' \in L^\infty(0,\mathcal{T}; H^k(\Omega))$, $c^{(3)} \in L^2(0,\mathcal{T}; L^2(\Omega))$, and $\|c\|_{L^\infty(0,\mathcal{T}; W^{1,\infty}(\Omega))} < M$; then there exist positive constants $C, \eta_2 > 0$, independent of h and Δt, such that for all $m+1 \le N$*

$$\begin{aligned}
&\|\xi_c^{m+1}\|_{0,\Omega}^2 + \|2\xi_c^{m+1} - \xi_c^{m}\|_{0,\Omega}^2 + \sum_{n=1}^{m} \|\Lambda \xi_c^{n+1}\|_{0,\Omega}^2 + \sum_{n=1}^{m} \Delta t \hat{\alpha}_a \|\xi_c^{n+1}\|_{1,\Omega}^2 \\
&\le C(\Delta t^4 + h^{2k}) + \sum_{n=1}^{m} \eta_2 \Delta t \|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2.
\end{aligned}$$

Proof. Proceeding similarly as in the proof of Theorem 4.3, from the second equation of (4.6) we get

$$\begin{aligned}
&\frac{1}{2\Delta t}\left(3\xi_c^{n+1} - 4\xi_c^{n} + \xi_c^{n-1}, \xi_c^{n+1}\right)_\Omega + \frac{1}{2\Delta t}\left(3E_c^{n+1} - 4E_c^{n} + E_c^{n-1}, \xi_c^{n+1}\right)_\Omega \\
&+ \frac{1}{2\Delta t}\left(3c(t_{n+1}) - 4c(t_n) + c(t_{n-1}), \xi_c^{n+1}\right)_\Omega + a_2^h(\xi_c^{n+1}, \xi_c^{n+1}) + a_2^h(\mathcal{I}_h\, c(t_{n+1}), \xi_c^{n+1}) \\
&+ c_2^h(\boldsymbol{u}_h^{n+1}, c_h^{n+1}, \xi_c^{n+1}) - d_2(c_h^{n+1}, \xi_c^{n+1}) = -G_{\mathrm{r}}(c_h^{n+1}, \xi_c^{n+1}),
\end{aligned} \tag{4.33}$$

and considering the third equation in Lemma 4.2, focusing on $t = t_{n+1}$, we immediately obtain

$$\begin{aligned}
&\frac{1}{2\Delta t}\left(3c(t_{n+1}) - 4c(t_n) + c(t_{n-1}), \xi_c^{n+1}\right)_\Omega + a_2\big(c(t_{n+1}), \xi_c^{n+1}\big) + c_2\big(\boldsymbol{u}(t_{n+1}), c(t_{n+1}), \xi_c^{n+1}\big) \\
&- d_2(c^{n+1}, \xi_c^{n+1}) = -G_{\mathrm{r}}(c^{n+1}, \xi_c^{n+1}) - \left(c'(t_{n+1}) - \frac{3c(t_{n+1}) - 4c(t_n) + c(t_{n-1})}{2\Delta t}, \xi_c^{n+1}\right)_\Omega.
\end{aligned} \tag{4.34}$$

Subtracting (4.33) from (4.34) and multiplying both sides of the result by $4\Delta t$ leads to

$$\begin{aligned}
&2\left(3\xi_c^{n+1} - 4\xi_c^{n} + \xi_c^{n-1}, \xi_c^{n+1}\right)_\Omega + 4\Delta t a_2(c_h^{n+1}; \xi_c^{n+1}, \xi_c^{n+1}) \\
&= 4\Delta t\left(c'(t_{n+1}) - \frac{3c(t_{n+1}) - 4c(t_n) + c(t_{n-1})}{2\Delta t}, \xi_c^{n+1}\right)_\Omega - 2\left(3E_c^{n+1} - 4E_c^{n} + E_c^{n-1}, \xi_c^{n+1}\right)_\Omega \\
&\quad - 4\Delta t\big(a_2(c_h^{n+1}; \mathcal{I}_h\, c^{n+1}, \xi_c^{n+1}) - a_2(c^{n+1}; c^{n+1}, \xi_c^{n+1})\big) \\
&\quad - 4\Delta t\big(c_2(\boldsymbol{u}_h^{n+1}, c_h^{n+1}, \xi_{\boldsymbol{u}}^{n+1}) - c_1^h(\boldsymbol{u}(t_{n+1}), c(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1})\big) \\
&\quad + 4\Delta t\big(d_2(c_h^{n+1}, \xi_c^{n+1}) - d_2(c^{n_1}, \xi_c^{n+1})\big) - 4\Delta t\big(G_{\mathrm{r}}(c_h^{n+1}, \xi_c^{n+1}) - G_{\mathrm{r}}(c^{n+1}, \xi_c^{n+1})\big).
\end{aligned} \tag{4.35}$$

As done above, we rewrite (4.35) using auxiliary terms now denoted $\hat{I}_1, \dots, \hat{I}_8$, and derive individual bounds for each term. For the first, second, and third terms, we use (4.27), (4.13), and Taylor expansion together with Young's inequality, respectively, to obtain

$$\begin{aligned}
\hat{I}_1 &= \|\xi_c^{n+1}\|_{0,\Omega}^2 + \|2\xi_c^{n+1} - \xi_c^n\|_{0,\Omega}^2 + \|\Lambda\xi_c^{n+1}\|_{0,\Omega}^2 - \|\xi_c^n\|_{0,\Omega}^2 - \|2\xi_c^n - \xi_c^{n-1}\|_{0,\Omega}^2,\\
\hat{I}_2 &\geq 4\Delta t\hat{\alpha}_a\|\xi_c^{n+1}\|_{1,\Omega}^2,\\
|\hat{I}_3| &\leq \frac{\Delta t^4}{24\varepsilon_1}\|c^{(3)}\|_{L^2(t_{n-1},t_{n+1};L^2(\Omega))}^2 + \frac{\Delta t\varepsilon_1}{2}\|\xi_c^{n+1}\|_{1,\Omega}^2.
\end{aligned}$$

Now we insert $\pm 4\Delta t E_c'(t_{n+1})$ into $\hat{I}_4$ and exploit (4.21). This leads to the bound

$$|\hat{I}_4| \leq \frac{C}{2\varepsilon_2}h^{2k}\|c'\|_{L^\infty(0,\mathcal{T};H^k(\Omega))}^2 + \frac{\Delta t\varepsilon_2}{2}\|\xi_c^{n+1}\|_{1,\Omega}^2 + \frac{\Delta t^4 C}{2\varepsilon_3}\|c^{(3)}\|_{L^2(0,\mathcal{T};L^2(\Omega))}^2 + \frac{\Delta t\varepsilon_3}{2}\|\xi_c^{n+1}\|_{1,\Omega}^2.$$

Employing again (4.21) in combination with (4.15) and (4.11), inserting $\pm a_2(c_h^{n+1}; c(t_{n+1}), \xi_c^{n+1})$ and $\pm a_2(\mathcal{I}_h\, c^{n+1}; c(t_{n+1}), \xi_c^{n+1})$; we have

$$\begin{aligned}
|\hat{I}_5| \leq{}& \frac{\hat{C}_a^2\Delta t h^{2k}}{2\varepsilon_4}\|c\|_{L^\infty(0,\mathcal{T};H^{k+1}(\Omega))}^2 + \frac{\Delta t\varepsilon_4}{2}\|\xi_c^{n+1}\|_{1,\Omega}^2 + \hat{C}_{\mathrm{lip}}M\|\xi_c\|_{1,\Omega}^2\\
&+ \frac{\tilde{C}_{\mathrm{lip}}^2 M^2\Delta t h^{2k}}{2\varepsilon_5}\|c\|_{L^\infty(0,T;H^k(\Omega))}^2 + \frac{\varepsilon_5\Delta t}{2}\|\xi_c^{n+1}\|_{1,\Omega}^2.
\end{aligned}$$

In order to derive a bound for $\hat{I}_6$ we add and subtract the terms

$$\pm c_2(\boldsymbol{u}(t_{n+1}), \mathcal{I}_h\, c(t_{n+1}), \xi_c^{n+1}), \quad \pm c_2(\Pi_h\, \boldsymbol{u}(t_{n+1}), \mathcal{I}_h\, c(t_{n+1}), \xi_c^{n+1}), \quad \pm c_2(\Pi_h\, \boldsymbol{u}(t_{n+1}), c(t_{n+1}), \xi_c^{n+1}),$$

which yields

$$\begin{aligned}
\hat{I}_6 = 4\Delta t\big(&c_2^h(\boldsymbol{u}_h^{n+1}, \mathcal{I}_h\, c(t_{n+1}), \xi_c^{n+1}) - c_2(\Pi_h\, \boldsymbol{u}(t_{n+1}), \mathcal{I}_h\, c(t_{n+1}), \xi_c^{n+1})\\
&+ c_2(\Pi_h\, \boldsymbol{u}(t_{n+1}), \mathcal{I}_h\, c(t_{n+1}), \xi_c^{n+1}) - c_2(\Pi_h\, \boldsymbol{u}(t_{n+1}), c(t_{n+1}), \xi_c^{n+1})\\
&+ c_2^h(\Pi_h\, \boldsymbol{u}(t_{n+1}), c(t_{n+1}), \xi_c^{n+1}) - c_2(\boldsymbol{u}(t_{n+1})c(t_{n+1}), \xi_c^{n+1}) + c_2(\boldsymbol{u}_h^{n+1}, \xi_c^{n+1}, \xi_c^{n+1})\big).
\end{aligned}$$

Using (4.17), (4.10c) and (4.21), we get

$$\begin{aligned}
|\tilde{I}_6| \leq 4\Delta t\Bigg(&\frac{\tilde{C}^2 C^*}{2\varepsilon_6}\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2\|c\|_{L^\infty(0,\mathcal{T};H^1(\Omega))}^2 + \frac{\varepsilon_6}{2}\|\xi_c\|_{1,\Omega}^2\\
&+ \frac{h^{2k}C\tilde{C}^2}{2\varepsilon_7}\|\boldsymbol{u}\|_{L^\infty(0,T;\boldsymbol{H}^1(\Omega))}^2\|c\|_{L^\infty(0,\mathcal{T};H^{k+1}(\Omega))}^2 + \frac{\varepsilon_7}{2}\|\xi_c^{n+1}\|_{1,\Omega}^2\\
&+ \frac{Ch^{2k}\tilde{C}^2}{2\varepsilon_8}\|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T};\boldsymbol{H}^{k+1}(\Omega))}^2\|c\|_{L^\infty(0,\mathcal{T};H^1(\Omega))}^2 + \frac{\varepsilon_8}{2}\|\xi_c^{n+1}\|_{1,\Omega}^2\Bigg).
\end{aligned}$$

Now, using (4.18) and (4.26), we have:

$$\begin{aligned}
|\hat{I}_7| &\leq \frac{C_d^2}{\varepsilon_9}2\Delta t\big(C^* h^{2k}\|c\|_{L^\infty(0,\mathcal{T};H^k(\Omega))}^2 + \|\xi_c\|_{0,\Omega}^2\big) + 2\Delta t\varepsilon_9\|\xi_c^{n+1}\|_{1,\Omega}^2,\\
|\hat{I}_8| &\leq \frac{\gamma_3^2}{\varepsilon_{10}}2\Delta t\big(C^* h^{2k}\|c\|_{L^\infty(0,\mathcal{T};H^k(\Omega))}^2 + \|\xi_c\|_{0,\Omega}^2\big) + 2\Delta t\varepsilon_{10}\|\xi_c^{n+1}\|_{1,\Omega}^2.
\end{aligned}$$

In this manner, and after choosing $\varepsilon_i = 4\hat{\alpha}_a/25$ for $i = 1, \dots, 8$ and $M \le \hat{\alpha}_a/\hat{C}_{\mathrm{Lip}}$, we can collect the above estimates and sum over $1 \le n \le m$, for all $m+1 \le N$, to get

$$\begin{aligned}
&\|\xi_c^{m+1}\|_{0,\Omega}^2 + \|2\xi_c^{m+1} - \xi_c^m\|_{0,\Omega}^2 + \sum_{n=1}^{m} \|\Lambda \xi_c^n\|_{0,\Omega}^2 + \sum_{n=1}^{m} \Delta t \hat{\alpha}_a \|\xi_c^{n+1}\|_{1,\Omega}^2 - 3\|\xi_c^1\|_{0,\Omega}^2 \\
&\le C(\Delta t^4 + h^{2k}) + \sum_{n=1}^{m} \eta_2 \|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2 + \sum_{n=1}^{m} \Delta t \frac{25}{\hat{\alpha}_a}(C_d^2 + \gamma_3^2)\|\xi_c\|_{0,\Omega}^2.
\end{aligned}$$

And the proof is completed by considering Δt sufficiently small and applying Gronwall's lemma. □

Theorem 4.5. *Under the assumptions of Theorems 4.3 and 4.4 with*

$$M := \min\left\{ \frac{\tilde{\alpha}_a}{\hat{C}_{\mathrm{Lip}}}, \frac{\tilde{\alpha}_a}{4\tilde{C}_u C^* C_\infty} \right\}$$

there exist positive constants $\hat{\gamma}_u$ and $\hat{\gamma}_c$ independent of Δt and h, such that for a sufficiently small Δt and all $m+1 \le N$, the following inequalities hold

$$\begin{aligned}
&\left(\|\xi_{\boldsymbol{u}}^{m+1}\|_{0,\Omega}^2 + \|2\xi_{\boldsymbol{u}}^{m+1} - \xi_{\boldsymbol{u}}^m\|_{0,\Omega}^2 + \sum_{n=1}^{m} \|\Lambda \xi_{\boldsymbol{u}}^n\|_{0,\Omega}^2 + \sum_{n=1}^{m} \Delta t \tilde{\alpha}_a \|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2 \right)^{1/2} \le \hat{\gamma}_u(\Delta t^2 + h^k), \\
&\left(\|\xi_c^{m+1}\|_{0,\Omega}^2 + \|2\xi_c^{m+1} - \xi_c^m\|_{0,\Omega}^2 + \sum_{n=1}^{m} \|\Lambda \xi_c^n\|_{0,\Omega}^2 + \sum_{n=1}^{m} \Delta t \hat{\alpha}_a \|\xi_c^{n+1}\|_{1,\Omega}^2 \right)^{1/2} \le \hat{\gamma}_c(\Delta t^2 + h^k).
\end{aligned}$$

Proof. From Theorem 4.3 we have the estimate

$$\sum_{n=1}^{m} \Delta t \|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2 \le C(\Delta t^4 + h^{2k}) + \frac{\eta_1}{\tilde{\alpha}_a} \sum_{n=1}^{m} \Delta t \|\xi_c^{n+1}\|_{0,\Omega}^2,$$

which, substituting back into Theorem 4.4, yields

$$\begin{aligned}
&\|\xi_c^{m+1}\|_{0,\Omega}^2 + \|2\xi_c^{m+1} - \xi_c^m\|_{0,\Omega}^2 + \sum_{n=1}^{m} \|\Lambda \xi_c^n\|_{0,\Omega}^2 + \sum_{n=1}^{m} \Delta t \hat{\alpha}_a \|\xi_c^{n+1}\|_{1,\Omega}^2 \\
&\le C(\Delta t^4 + h^{2k}) + \frac{\eta_1 \eta_2}{\tilde{\alpha}_a} \sum_{n=1}^{m} \Delta t \|\xi_c^{n+1}\|_{0,\Omega}^2.
\end{aligned}$$

For the last term on the right-hand side of this last bound we have

$$\|\xi_c^{m+1}\|_{0,\Omega}^2 \le 2\big(\|\Lambda \xi_c^m\|_{0,\Omega}^2 + \|2\xi_c^m - \xi_c^{m-1}\|_{0,\Omega}^2\big),$$

and considering Δt sufficiently small and applying Gronwall's lemma, we readily deduce that

$$\|\xi_c^{m+1}\|_{0,\Omega}^2 + \|2\xi_c^{m+1} - \xi_c^m\|_{0,\Omega}^2 + \sum_{n=1}^{m} \|\Lambda \xi_c^{n+1}\|_{0,\Omega}^2 + \sum_{n=1}^{m} \Delta t \hat{\alpha}_a \|\xi_c^{n+1}\|_{1,\Omega}^2 \le C(\Delta t^4 + h^{2k}). \tag{4.36}$$

The first bound follow by combining (4.36) and Theorem 4.3. □

Lemma 4.3. *Under the same assumptions of Theorem 4.5, we have*

$$\left(\sum_{n=1}^{m} \Delta t \|p(t_{n+1}) - p_h^{n+1}\|_{0,\Omega}^2\right)^{1/2} \leq \hat{\gamma}_p(\Delta t^2 + h^k).$$

Proof. Owing to the inf-sup condition (4.19), there exists a function $\boldsymbol{w}_h \in \boldsymbol{Z}_h^{\perp}$ such that

$$b\big(\boldsymbol{w}_h, p(t_{n+1}) - p_h^{n+1}\big) = \|p(t_{n+1}) - p_h^{n+1}\|_{0,\Omega}^2, \tag{4.37}$$

$$\|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} \leq \frac{1}{\tilde{\beta}} \|p(t_{n+1}) - p_h^{n+1}\|_{0,\Omega}. \tag{4.38}$$

From (4.7) and Lemma 4.2, proceeding as in the proof of Theorem 4.3, we obtain

$$\begin{aligned}
\Delta t b(\boldsymbol{w}_h, p(t_{n+1}) - p_h^{n+1}) = &-\Delta t \left(\boldsymbol{u}'(t_{n+1}) - \frac{3\boldsymbol{u}_h^{n+1} - 4\boldsymbol{u}_h^{n} + \boldsymbol{u}_h^{n-1}}{2\Delta t}, \boldsymbol{w}_h\right)_{\Omega} \\
&+ \Delta t\big(a_1^h(c_h^{n+1}; \boldsymbol{u}_h^{n+1}, \boldsymbol{w}_h) - a_1^h(c(t_{n+1}); \boldsymbol{u}(t_{n+1}), \boldsymbol{w}_h)\big) \\
&+ \Delta t\big(c_1^h(\boldsymbol{u}_h^{n+1}; \boldsymbol{u}_h^{n+1}, \boldsymbol{w}_h) - c_1^h(\boldsymbol{u}(t_{n+1}); \boldsymbol{u}(t_{n+1}), \boldsymbol{w}_h)\big) \\
&+ \Delta t\big(\boldsymbol{F}_{\mathrm{g}}(c(t_{n+1}), \boldsymbol{w}_h) - \boldsymbol{F}_{\mathrm{g}}(c_h^{n+1}, \boldsymbol{w}_h)\big) \\
&+ \Delta t\big(\boldsymbol{F}_{\mathrm{r}}(\boldsymbol{u}(t_{n+1}), \boldsymbol{w}_h) - \boldsymbol{F}_{\mathrm{r}}(\boldsymbol{u}_h^{n+1}, \boldsymbol{w}_h)\big) \\
\leq &\frac{\Delta t^2}{2\sqrt{3}} \|\boldsymbol{u}^{(3)}\|_{L^2(t_{n-1}, t_{n+1}, \boldsymbol{L}^2(\Omega))} \sqrt{\Delta t} \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} \\
&+ \tilde{C}_a C^* h^k \Delta t \|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T}; \boldsymbol{H}^{k+1}(\Omega))} \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} + \tilde{C}_{\mathrm{lip}} M \Delta t \|\xi_c^{n+1}\|_{1,\Omega} \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} \\
&+ \tilde{C}_{\mathrm{lip}} \Delta t M \|\xi_c\|_{1,\Omega} \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} + \Delta t C \tilde{C}_u C^* C_\infty M \|\xi_c\|_{1,\Omega} \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} \\
&+ \Delta t C \tilde{C}_u h^k \|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T}; \boldsymbol{H}^1(\Omega))} \|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T}; \boldsymbol{H}^{k+1}(\Omega))} \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} \\
&+ \Delta t C \tilde{C}_u h^k \|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T}; \boldsymbol{H}^1(\Omega))} \|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T}, \boldsymbol{H}^{k+1}(\Omega))} \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} \\
&+ \gamma_2 \Delta t h^k C^* \|c\|_{L^\infty(0,\mathcal{T}; H^k(\Omega))} \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} + \gamma_2 \Delta t \|\xi_{\boldsymbol{u}}\|_{0,\Omega} \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} \\
&+ \gamma_1 \Delta t h^k C^* \|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T}; \boldsymbol{H}^k(\Omega))} \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} + \gamma_1 \Delta t \|\xi_{\boldsymbol{u}}\|_{0,\Omega} \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h}.
\end{aligned}$$

Summing over $1 \leq n \leq m$ for all $m + 1 \leq N$ and substituting back into equations (4.37) and (4.38), we obtain

$$\begin{aligned}
&\left(\sum_{n=1}^{m} \Delta t \|p(t_{n+1}) - p_h^{n+1}\|_{0,\Omega}^2\right)^{1/2} \\
&\qquad \leq \frac{C}{\tilde{\beta}} \left((\Delta t^2 + h^k) + \left(\sum_{n=1}^{m} \Delta t \|\xi_c^{n+1}\|_{0,\Omega}^2\right)^{1/2} + \left(\sum_{n=1}^{m} \Delta t \|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2\right)^{1/2}\right).
\end{aligned}$$

The result follows by applying Theorem 4.5. □

4.4 Numerical results

In this section we test the performance of the numerical method and produce some typical solutions in operating conditions. Tetrahedral meshes have been constructed using the freely available mesh

k	DoF	$\mathtt{e}_{\boldsymbol{u}}$	rate	$\mathtt{e}_p$	rate	$\mathtt{e}_s$	rate
1	53	0.004507	–	0.291804	–	0.253207	–
	299	0.002783	0.679	0.192100	0.589	0.153518	0.708
	1265	0.001273	1.150	0.096891	1.006	0.073370	1.085
	4634	0.000631	1.017	0.051713	0.911	0.038362	0.941
	17780	0.000308	1.033	0.026853	0.945	0.018841	1.026
2	132	0.001817	–	0.115142	–	0.089672	–
	797	0.000342	2.349	0.032084	1.799	0.018151	2.249
	3427	8.031e-5	2.133	0.007198	2.197	0.003702	2.337
	12702	1.948e-5	2.056	0.002023	1.844	0.000996	1.905
	49157	4.358e-6	2.159	0.000525	1.941	0.000251	1.987

Table 4.1: Spatial accuracy test: experimental errors and convergence rates for the approximate solutions $\boldsymbol{u}_h$, p_h and c_h. Values are displayed for schemes with first and second order in space (table produced by the author).

manipulator GMSH [81], and the implementation of the H(div)-conforming finite element scheme is carried out using the open source finite element library FEniCS [7]. The linear systems encountered at each Newton-Raphson step are solved with the GMRES method preconditioned with AMG. The Newton iterations stop whenever either the absolute or the relative residuals (measured in the ℓ^2-norm) drop below the fixed tolerance set to 1×10^{-6}. Apart from the main python modules, a dedicated C++ expression is needed to efficiently compile the position of the rake at each time. It depends on the structure dimensions and on the angular velocity.

4.4.1 Numerical verification of convergence

We start with a simple experimental convergence analysis to confirm the error bounds anticipated in Section 4.3.2. Doing this in a 2D domain suffices, so we consider Ω as a circle of radius one and construct a sequence of successively refined meshes on which we compute errors between the approximate solutions obtained with the H(div)-conforming scheme and the closed-form solutions

$$\boldsymbol{u} = \sin(t) \begin{pmatrix} \cos(\pi/2x)\sin(\pi/2y) \\ -\sin(\pi/2x)\cos(\pi/2y) \end{pmatrix}, \quad p = (x^4 - y^4)\exp(-t), \quad c = \frac{1}{2}\cos\left(\frac{\pi}{4}|\boldsymbol{x}|^2\right)\exp(-t),$$

that are used to construct suitable Dirichlet boundary data for velocity and an exact flux for concentration, and manufactured forcing and source terms $\boldsymbol{F}_{ex}$ and g_{ex} appearing on the right-hand side of the momentum equation and of the concentration mass balance, respectively. As $\boldsymbol{u}$ is prescribed everywhere on $\partial\Omega$, for sake of uniqueness we impose $p \in L^2_0(\Omega)$ through a Lagrange multiplier approach. We use a constant viscosity $\nu = 0.01$ and diffusivity $D = 1.0$ with $\boldsymbol{f}_{\mathrm{r}}$ as given in (4.23), $f_{bk}(c) = 1 \times 10^{-2}(1-c)$ and $\boldsymbol{k}$ pointing in the radial outwards direction.

Δt	$\hat{\mathtt{e}}_{\boldsymbol{u}}$	rate	$\hat{\mathtt{e}}_p$	rate	$\hat{\mathtt{e}}_s$	rate
2	5.6194	–	0.5069	–	0.4558	–
1	1.5943	1.817	0.1809	1.487	0.0868	2.391
0.5	0.4433	1.847	0.0523	1.789	0.0193	2.167
0.25	0.1153	1.943	0.0135	1.951	0.0046	2.070
0.125	0.0296	1.959	0.0033	2.000	0.0012	1.994

Table 4.2: Time accuracy test: experimental errors and convergence rates for the approximate solutions $\boldsymbol{u}_h$, p_h and c_h, computed for each refinement level (table produced by the author).

We show orders of convergence in the discrete norm $\|\cdot\|_{1,\mathcal{T}_h}$ for the velocity $\boldsymbol{u}$, in the L^2-norm of the error of p, and in the H^1-norm of the error in c in Table 4.1. For polynomial degrees $k = 1$ and $k = 2$ we observe that the order of convergence predicted by our theory (see Theorem 4.5 and Lemma 4.3) is achieved.

Regarding the convergence of the time advancing scheme, now we set $\mathcal{T} = 4$ and consider a sequence of uniform refined time partitions $\tau_l, l \in \{1, 2, 3, 4, 5\}$ where the time step is 2^{2-l}. Absolute errors are computed as

$$\hat{\mathtt{e}}_{\boldsymbol{u}} = \left(\sum_{n=1}^{m} \Delta t \|\boldsymbol{u}(t_{n+1}) - \boldsymbol{u}_h^{n+1}\|_{1,\mathcal{T}_h}^2 \right)^{1/2}, \quad \hat{\mathtt{e}}_p = \left(\sum_{n=1}^{m} \Delta t \|p(t_{n+1}) - p_h^{n+1}\|_{0,\Omega}^2 \right)^{1/2},$$

$$\hat{\mathtt{e}}_c = \left(\sum_{n=1}^{m} \Delta t \|c(t_{n+1}) - c_h^{n+1}\|_{1,\Omega}^2 \right)^{1/2},$$

and we readily observe from Table 4.2 that the method converges to the exact solution with the expected second-order rate.

4.4.2 Preliminary two-dimensional computation

The typical operation conditions on the clarifier unit are characterised by about 1.2 revolutions per hour, a solid concentration behind the rake of 0.01 g/l, a feed flow rate of 10,000 gpm, a return sludge flow rate of 3000 gpm, an effluent flow rate in the overflow weir of approximately 7000 gpm, and a solid concentration at the inlet of 5 g/l (see [65] and the references therein). The specification of the remaining model parameters, at least in this specific scenario, are much less clear and we need to characterise them in terms of the expected flow conditions. Known issues in the operation process include a strong backflow into the feedwell, a large recirculation zone near the feedwell, the high velocity of the flow exiting the feedwell, and the lack of flow symmetry.

In order to gain insight into the impact of the rake parameters on the simulation we regard the operation from an azimuthal view and consider only the coupled Navier-Stokes/concentration problem in an annular domain of external radius 30m and internal radius 3m, where one can still see the rotating arm, but the vertical sedimentation is not represented. Here the body force term exerted on the fluid ($\boldsymbol{F}_{\mathrm{g}}$) is considered with a radial direction towards the centre of the inner disk. Furthermore the

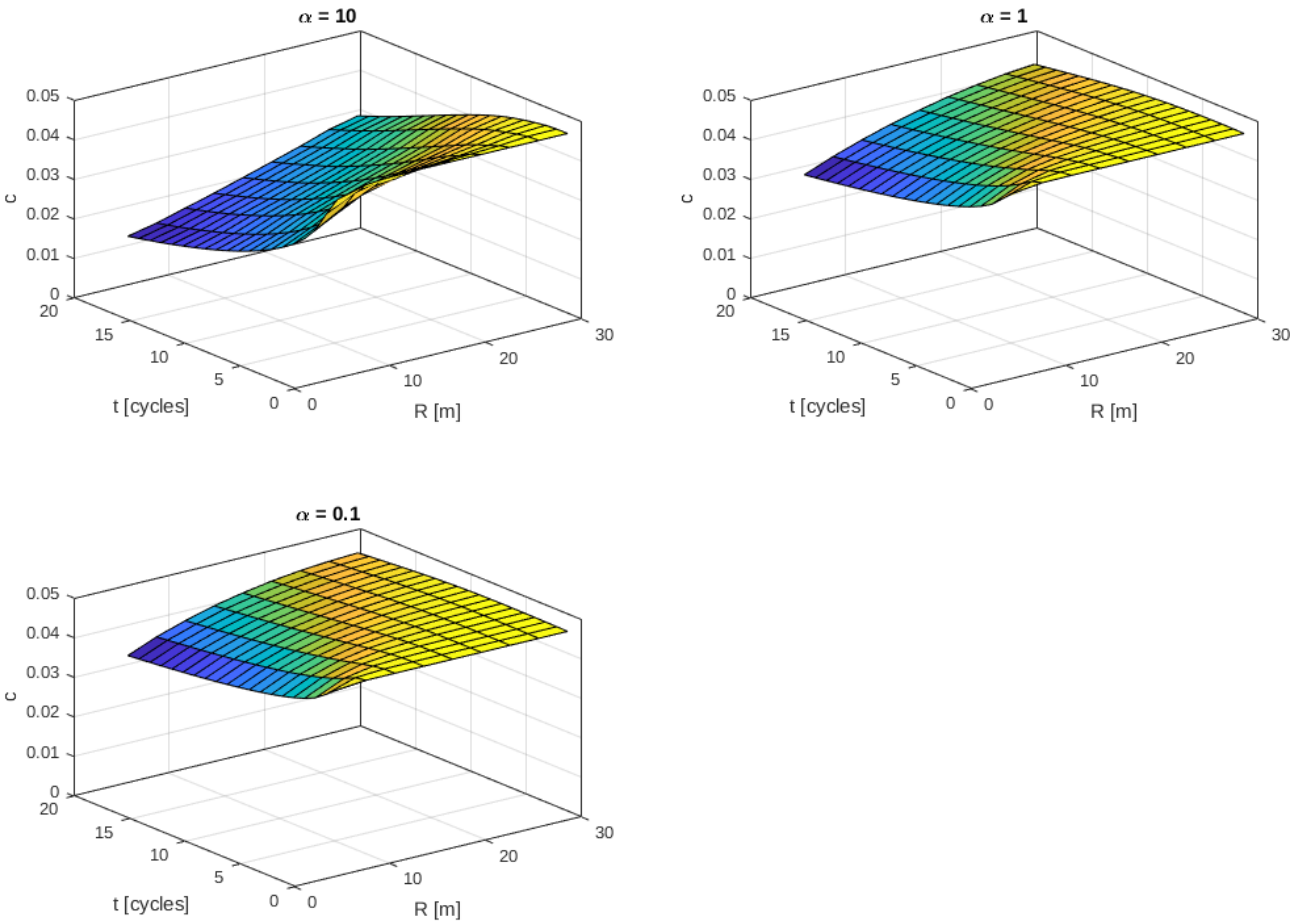


Figure 4.3: Spatio-temporal variation of the average concentration after complete rake cycles at different radius (measured from the centre of the annular domain) and values of α (figure produced by the author).

parameters of the simulation are taken as follows:

$$\rho_{\mathrm{s}} = 2500\,[\mathrm{kg/m^2}], \quad c_0 = 0.05, \quad \rho_{\mathrm{f}} = 1000\,[\mathrm{kg/m^2}], \quad \omega = 1.2\,[\mathrm{rad/min}], \quad c_{\mathrm{r}} = 1 \times 10^{-3},$$
$$g = 1 \times 10^{-3}\,[\mathrm{m/min^2}], \quad D_0 = 1.0\,[\mathrm{m^2/min}], \quad \nu_0 = 0.05\,[\mathrm{kg/(m\,min)}],$$
$$f_{\mathrm{bk}}(c) = 1.0 \times 10^{-3} c(1-c)^2 [\mathrm{m/s}], \quad \sigma_{\mathrm{e}}(c) = \begin{cases} 0 & \text{for } c \le c_{\mathrm{c}} = 0.07 \\ (50.0/c_{\mathrm{c}})[(c/0.07)^5 - 1][\mathrm{Pa}] & \text{for } c > c_{\mathrm{c}}. \end{cases}$$

We start the simulation with a homogeneous initial concentration c_0 and then, we observe how this concentration changes over time for different values of the parameters α and β. From results shown in Figures 4.3 and 4.4, it can be highlighted that the solids removal coefficient α is the most relevant for the concentration profile, while the combined contributions from drag and density do not seem to have a large effect.

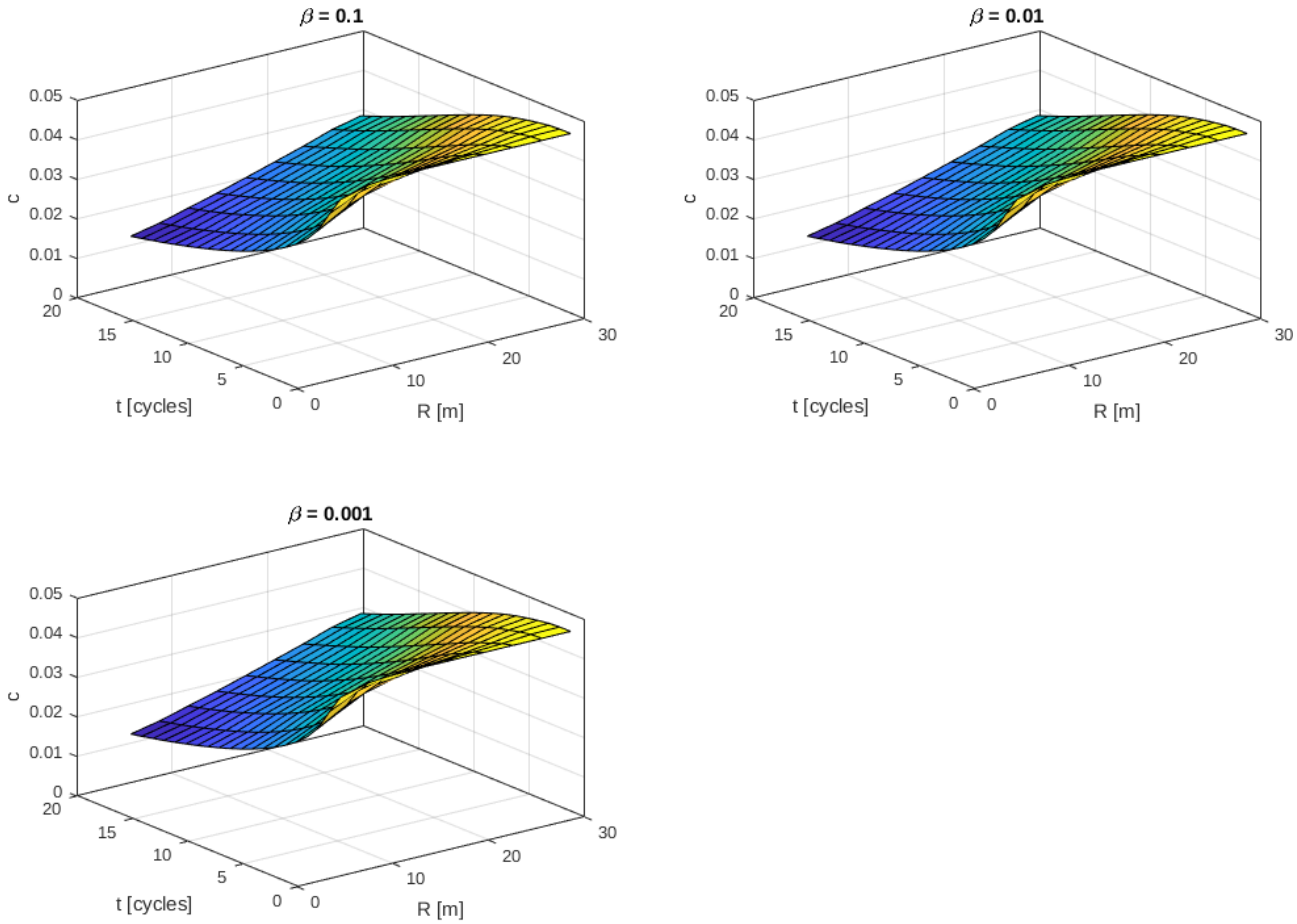


Figure 4.4: Spatio-temporal variation of the average concentration after complete rake cycles at different radius (measured from the centre of the annular domain) and values of β (figure produced by the author).

4.4.3 Performance of clarifier units

Having now a better understanding on the dimension and isolated effects of each mechanism in the coupled problem, we turn to the simulation of the sedimentation of flocculated suspensions in a more realistic geometry. We consider the domain sketched in Figure 4.1, and take $R = 15\,[\mathrm{m}]$ and $H = 7\,[\mathrm{m}]$. We suppose that the tank is initially filled with a homogeneous mixture of concentration $c_0 = 0.02$. Apart from the specifications in (4.2), (4.3), the remaining concentration-dependent and constant parameters needed in the model assume the following form (where the suspension is assumed

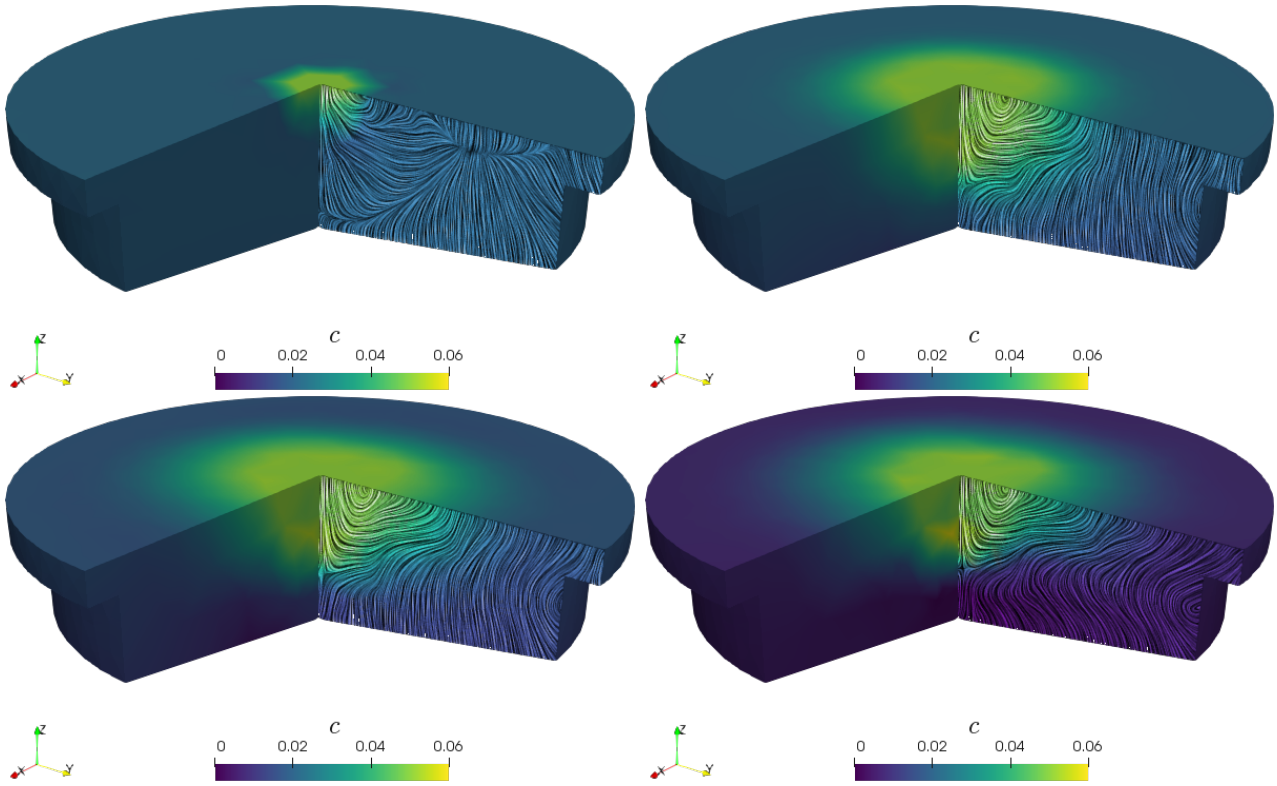


Figure 4.5: Domain cuts showing snapshots of solids concentration and line integral contours of velocity on a slice, focusing on time instants $t = 1, 30, 60$ and 180 [min] (figure produced by the author).

of type *Kaolin flat D*)

$$\rho_{\mathrm{s}} = 2500\,[\mathrm{kg/m^3}], \quad c_{\mathrm{in}} = 0.05, \quad \rho_{\mathrm{f}} = 1000\,[\mathrm{kg/m^3}], \quad \omega = 0.12\,[\mathrm{rad/min}], \quad \alpha = 0.01\,[\mathrm{min^{-1}}],$$
$$\beta\rho_{\mathrm{r}} = 50\,[\mathrm{kg/m^3\,m^{-1}}], \quad c_{\mathrm{r}} = 1\times 10^{-3}, \quad g = 9.8\,[\mathrm{m/s^2}], \quad D_0 = 0.05\,[\mathrm{m^2/min}],$$
$$\boldsymbol{u}_{\mathrm{in}} = -4.2\boldsymbol{k}\,[\mathrm{m/min}], \quad \nu_0 = 0.05\,[\mathrm{kg/(m\,min)}],$$
$$f_{\mathrm{bk}}(c) = 1.0\times 10^{-4} c(1-c)^2 [\mathrm{m/s}], \quad \sigma_{\mathrm{e}}(c) = \begin{cases} 0 & \text{for } c \le c_{\mathrm{c}} = 0.07 \\ (50.0/c_{\mathrm{c}})[(c/0.07)^5 - 1][\mathrm{Pa}] & \text{for } c > c_{\mathrm{c}}. \end{cases}$$

We conduct a series of runs on the 3D geometry where the resulting tetrahedral mesh has 139001 elements and 27510 vertices (representing 1.1M DoFs for the lowest-order H(div)-conforming finite element method). The time stepping scheme uses a fixed timestep of $\Delta t = 0.5\,[\mathrm{min}]$ and we simulate the process until 180 [min]. As mentioned above, one manifestation of performance in the clarifier units is the development of recirculation patterns, and we plot in Figure 4.5 the concentration profiles on a cut of the domain, as well as a slice of a section where we plot line integral contours of velocity, for three different times. The plots indicate a large diffusion of the concentration as it spreads out from the feedwell, and we also see a substantial modification on the flow patterns due to the combined contribution of the rake mechanism and the gravitational settling. The velocity can be seen more clearly from Figure 4.6, showing streamlines at $t = 180\,[\mathrm{min}]$ from different angles, emphasising that the recirculation in the xy plane occurs mainly near the bottom of the vessel, whereas on the top the

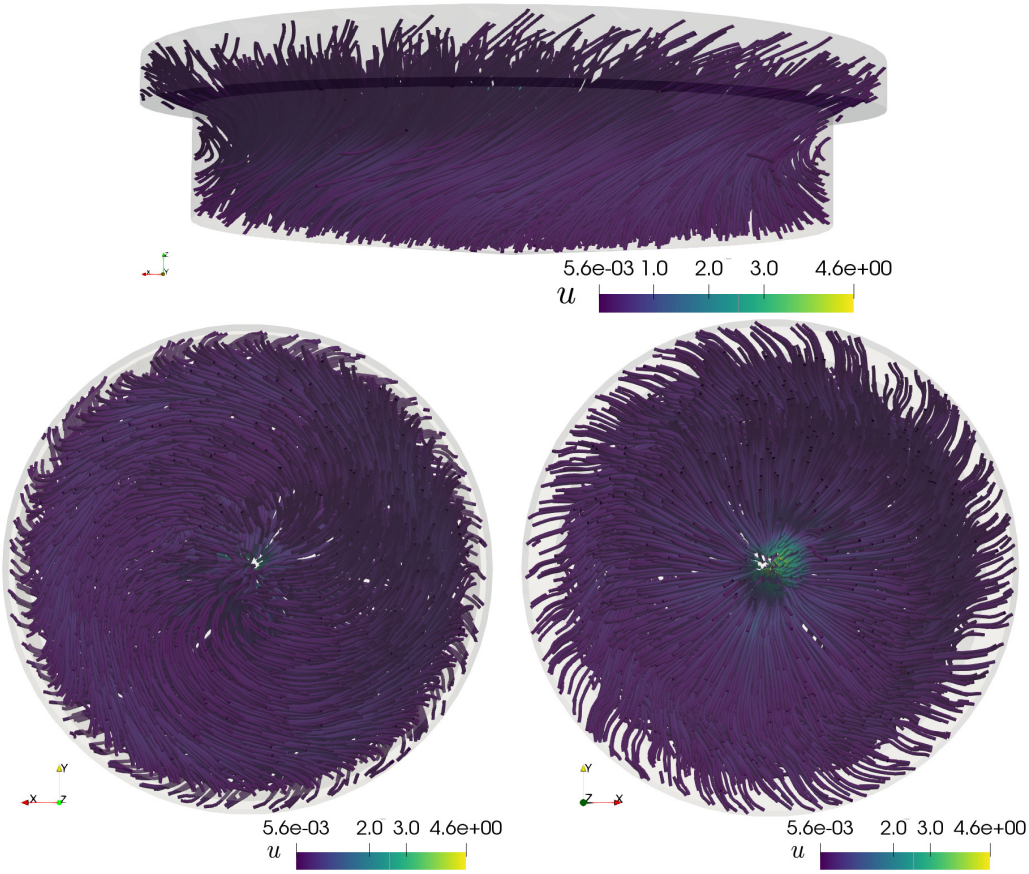


Figure 4.6: Velocity streamlines at 180 [min], shown from the side (top panel), from the bottom (bottom left figure), and from the top (bottom right figure) (figure produced by the author).

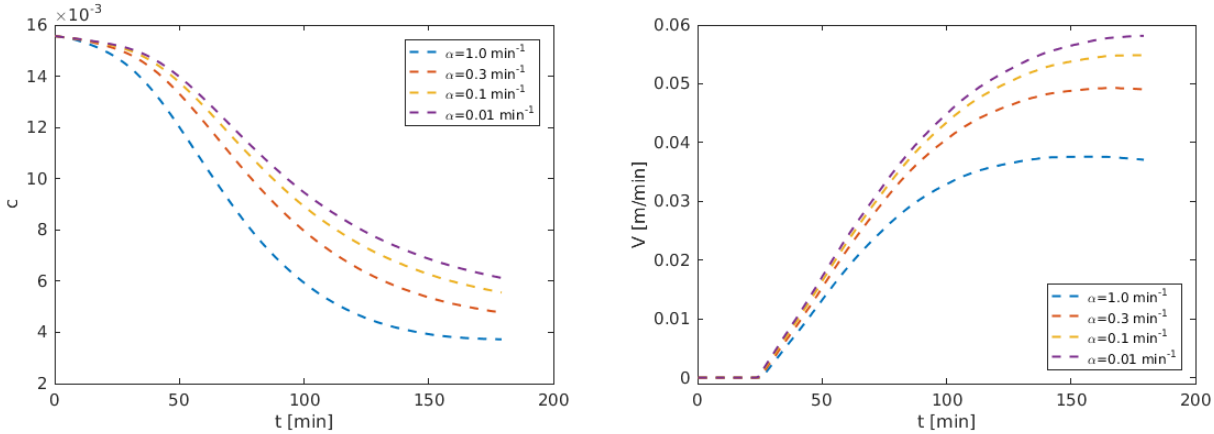


Figure 4.7: Time evolution of the concentration and normal velocity on the overflow for different values of the solids removal coefficient α (figure produced by the author).

velocity is dominated by gravitational forces and a radially spreading concentration-driven flow.

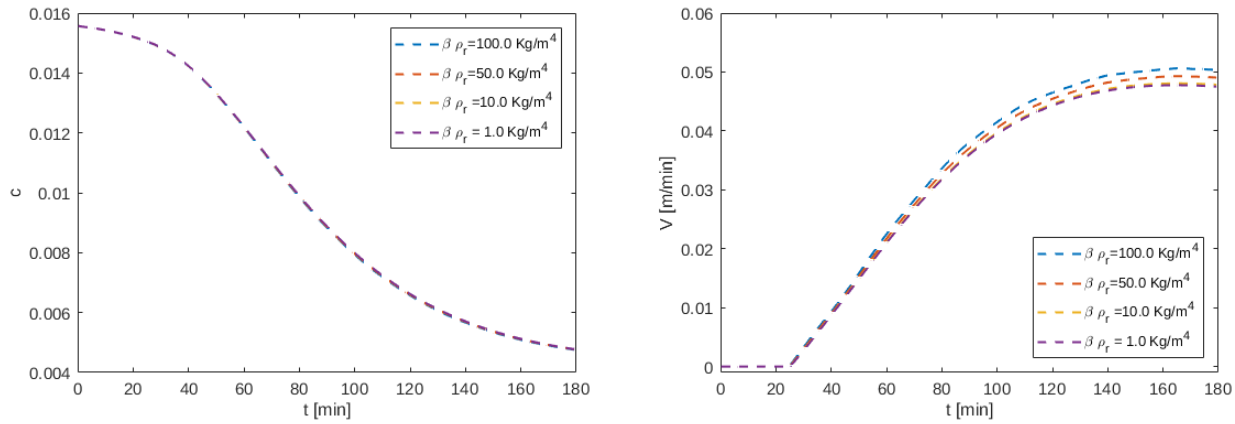


Figure 4.8: Time evolution of the concentration and normal velocity on the overflow for different values of the drag-density coefficient $\beta\rho_{\mathrm{r}}$ (figure produced by the author).

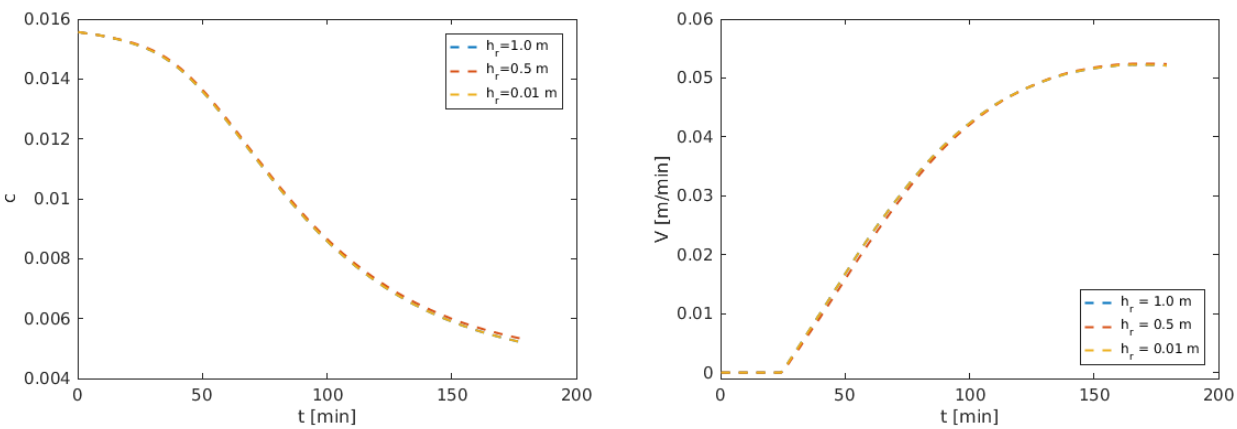


Figure 4.9: Time evolution of the concentration and normal velocity on the overflow for different values of rake height h_{r} (figure produced by the author).

On the other hand, the variation of the flow conditions depending on different factors can be observed from Figures 4.7, 4.8, and 4.9. There we portray the dynamics of the concentration and flow rate on the overflow, that is, respectively

$$\frac{1}{60\pi}\int_{\Gamma_{\mathrm{ofl}}} c\,\mathrm{d}s, \qquad \frac{1}{60\pi}\int_{\Gamma_{\mathrm{ofl}}} \boldsymbol{u}\cdot\boldsymbol{n}\,\mathrm{d}s,$$

according to modifications in the solids removal intensity, on the drag and density of the rotating rake, and on the rake height. Based on the results of this set of simulations, we can identify the solids removal coefficient α as the most sensitive factor on the outputs of overflow concentration and overflow flow rate. On the other hand, the combined contributions from drag and density do not seem to have a large effect on these markers, which is consistent with what we saw in the preliminary 2D test. However, a further inspection reveals that the effects are not necessarily localised but they differ over the height of the device. From Figure 4.10 we can see how the average concentration varies over time

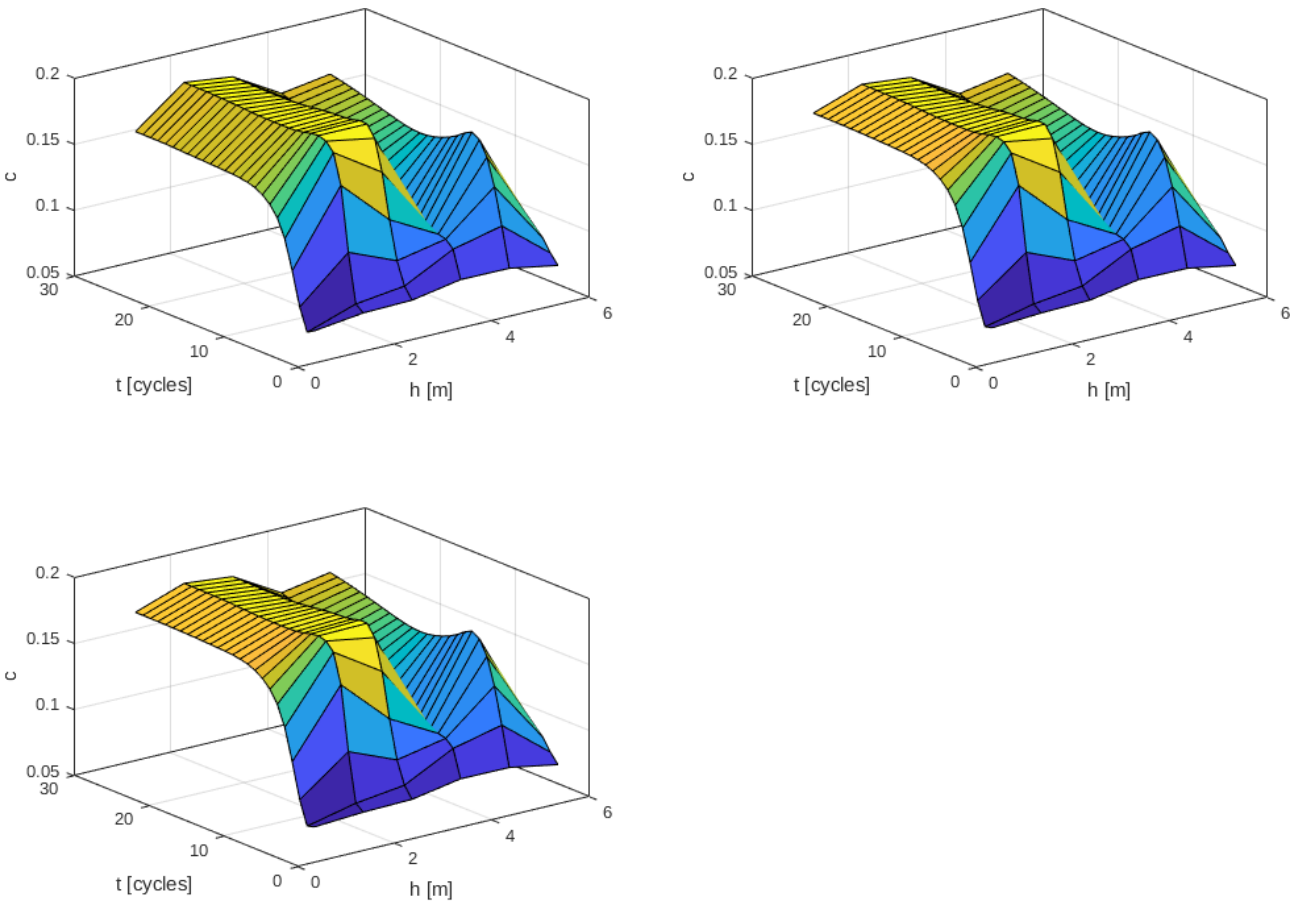


Figure 4.10: Spatio-temporal variation of the average concentration after complete rake cycles at different heights (measured from the bottom) with $\alpha = 0.3, \beta = 50$ (top left), $\alpha = 0.0, \beta = 50$ (top right), $\alpha = 0.0, \beta = 0.0$ (bottom) (figure produced by the author).

(and measured after a given number of cycles of the rotating rake) depending on the solids removal coefficient and on the drag.

CHAPTER 5

Second-order schemes for axisymmetric Navier-Stokes-Brinkman and transport equations modelling water filters

Soil-based water filtering devices can be described by models of viscous flow in porous media coupled with an advection-diffusion-reaction system modelling the transport of distinct contaminant species within water, and being susceptible to adsorption in the medium that represents soil. In this chapter we analyze such models mathematically, and design suitable numerical methods for their approximate solution. The governing equations are the Navier-Stokes-Brinkman equations for the flow of the fluid through a porous medium coupled with a convection-diffusion equation for the transport of the contaminants plus a system of ordinary differential equations accounting for the degradation of the adsorption properties of each contaminant. These equations are written in meridional axisymmetric form and the corresponding weak formulation adopts a mixed-primal structure. A second-order, (axisymmetric) divergence-conforming discretisation of this problem is introduced and the solvability, stability, and spatio-temporal convergence of the numerical method are analysed. Some numerical examples illustrate the main features of the problem and the properties of the numerical scheme.

5.1 Introduction

5.1.1 Scope

We are interested in the analysis and numerical approximation of the flow of a viscous fluid through a porous medium, where it is assumed that the fluid carries a number m of components that are adsorbed by the porous medium. While viscous flow in porous media with adsorption arises in several applications including polymer flooding as part of the process of enhanced oil recovery in petroleum engineering [46], chromatography [141], or water decontamination and removal of pollutants such as heavy metals or radioactive ions [160], the particular formulation in the present work is motivated by a model of a soil-based water filtering device designed to remove contaminants from water by adsorption [134].

The governing equations for this process can be formulated as follows. We assume that the porous medium is represented by a simply connected spatial domain $\Omega \subset \mathbb{R}^3$ whose boundary $\partial\Omega$ is split

into three disjoints parts Γ^{in}, Γ^{wall} and Γ^{out} representing the inlet, walls, and outlet boundaries. For all times $0 < t \leq \mathcal{T}$, we consider the Navier-Stokes-Brinkman equations written in terms of the volume average flow velocity $\boldsymbol{u}(t)\colon \Omega \to \mathbb{R}^3$ and the fluid pressure $p(t)\colon \Omega \to \mathbb{R}$; as well as the balances for contaminant concentration possessing sink terms that depend on the rate of degradation of the adsorption properties of each material, described in terms of the vector of concentrations of $m \geq 2$ distinct types of contaminants $\vec{\theta}(t) = (\theta_1(t), \ldots, \theta_m(t))\colon \Omega \to \mathbb{R}^m$ and of the adsorption capacity relative to each contaminant $\vec{s}(t) = (s_1(t), \ldots, s_m(t))\colon \Omega \to \mathbb{R}^m$. The coupled set of governing equations (three partial differential equations (PDEs) and one ordinary differential equation (ODE)) adopts the form

$$\rho_{\text{f}}(\partial_t \boldsymbol{u} + \boldsymbol{u} \cdot \nabla \boldsymbol{u}) + \mathbb{K}^{-1}\nu \boldsymbol{u} - \mathbf{div}(\nu \boldsymbol{\varepsilon}(\boldsymbol{u}) - p\mathbb{I}) = \boldsymbol{F}(\vec{\theta}), \qquad \text{in } \Omega \times (0, \mathcal{T}], \tag{5.1a}$$

$$\operatorname{div} \boldsymbol{u} = 0, \qquad \text{in } \Omega \times (0, \mathcal{T}], \tag{5.1b}$$

$$\phi \partial_t \vec{\theta} - \operatorname{div}(\mathbb{D}\nabla\vec{\theta}) + (\boldsymbol{u} \cdot \nabla)\vec{\theta} = -\rho_{\text{b}} \partial_t \vec{s}, \qquad \text{in } \Omega \times (0, \mathcal{T}], \tag{5.1c}$$

$$\partial_t \vec{s} = \boldsymbol{G}(\vec{s}, \vec{\theta}) \qquad \text{in } \Omega \times (0, \mathcal{T}], \tag{5.1d}$$

where $\boldsymbol{\varepsilon}(\boldsymbol{u}) = \frac{1}{2}(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^T)$ is the strain rate tensor, $\mathbb{D} = \operatorname{diag}(D_1(\boldsymbol{x}), \ldots, D_m(\boldsymbol{x}))$ denotes a space-dependent and positive definite matrix containing diffusivity coefficients, $\nu > 0$ is the constant fluid viscosity, $\rho_{\text{f}}, \rho_{\text{b}}$ are the constant densities of the fluid phases and of the bulk filter medium, $\phi(\boldsymbol{x})$ is the porosity of the soil constituting the porous medium, and $\mathbb{K}(\boldsymbol{x}) > 0$ is the permeability tensor (assumed symmetric and uniformly positive definite). The source and reaction terms are

$$\boldsymbol{F}(\vec{\theta}) = \boldsymbol{g} \sum_{i=1}^{m} \theta_i; \quad G_i(s_i, \theta_i) = k_i^+(\boldsymbol{x})\big(s_i^{\max} - s_i\big)\theta_i, \quad i = 1, \ldots, m, \tag{5.2}$$

where $\boldsymbol{G} = (G_1, \ldots, G_m)^{\text{T}}$, $\boldsymbol{g}$ is the gravity acceleration, $s_i^{\max}$ is a constant representing the maximum amount of contaminant i that can be absorbed at a given point, and $k_i^+(\boldsymbol{x})$ is a spatially-dependent modulation coefficient accounting for the forward adsorption rate related to the loss of contaminant i due to the filtering process (boundary conditions and further assumptions will be specified in later parts of the chapter).

Thus, the flow of the incompressible fluid through Ω is modelled by the Navier-Stokes-Brinkman equation (5.1a) and the continuity equation (5.1b), which express the conservation of momentum and mass respectively. Equation (5.1c) describes the evolution of $\vec{\theta}$ within Ω, under the effects of advection and diffusion, in addition to adsorption by the filter media. Given the typical operating conditions within the filter, we would expect the effects of advection to dominate those from diffusion, as noted in [134]. The sink term $-\rho_{\text{b}}\partial_t \vec{s}$ in (5.1c) accounts for the net and local removal of each contaminant type due to the filtration process. This adsorption process is described by a multicomponent Langmuir-type model, as given by (5.1d) and (5.2). Under this model, it is assumed that each site has a maximum capacity for each individual contaminant, which we take to be uniform across the two layers of filter media. In this way, the adsorption is noncompetitive and the saturation of a site by one contaminant does not prevent adsorption of the other contaminants at the same site. It is also assumed that the adsorption process is irreversible for all contaminants and all filter layers, so that once adsorbed the contaminants remain attached to the filter media with no desorption back into the fluid. As described previously, for each contaminant we ascribe a spatially dependent adsorption rate $k_i^+(\boldsymbol{x})$, so (5.2) stipulates that the rate of removal of a contaminant at a site is proportional to the concentration of

the contaminant present in the fluid at the site, the remaining capacity of the filter media at the site and the adsorption rate.

While the modelling of a filter calls for a three-dimensional domain, in practice most filter designs display rotational symmetry around their central axis, with the flow also expected to exhibit such symmetry. This property motivates an axisymmetric formulation of the problem, allowing for the reduction from three to two spatial dimensions, which evidently reduces the computational cost associated with its solution. Thus, the model which is eventually analysed herein is a reformulation of (5.1) along with suitable initial and boundary conditions as a meridional axisymmetric PDE-ODE initial-boundary value problem. It is the purpose of this chapter to advance a second-order divergence-conforming discretisation for this problem. Specifically, we introduce an axisymmetric $\boldsymbol{H}(\mathrm{div})$-conforming method based on two-dimensional BDM spaces [40] combined with an implicit, second-order backward differentiation formula for time discretisation. Based on discrete stability properties, we prove that the discrete problem has at least one solution. At the core of this chapter is the derivation of an optimal a priori error estimate for the numerical scheme, where the main difficulty is the fully discrete analysis verifying that each of the terms is bounded optimally in the corresponding weighted spaces. Numerical examples illustrate the model and reconfirm the theoretical order of accuracy.

5.1.2 Outline of the chapter

The remainder of this chapter is organized as follows. In Section 5.2 we introduce the model problem and state some preliminaries for its analysis, starting with a description of the initial and boundary conditions for (5.1) that correspond to the filter model (Section 5.2.1). Next, in Section 5.2.2, we reformulate (5.1) and the corresponding initial and boundary conditions in meridional axisymmetric form, which under suitable assumptions leads to model in two (namely, radial and vertical) space dimensions. We provide in Section 5.2.3 some preliminaries on functional spaces associated with radially symmetric functions. The weak (variational) formulation of the axisymmetric problem is stated in Section 5.2.4. Further assumptions on the model coefficients, as well as a number of inequalities related to the bilinear and trilinear forms involved in the weak formulation, are stated in Section 5.2.5. Section 5.3 outlines the well-posedness analysis (proof of existence and uniqueness of a weak solution) of the axisymmetric problem derived in Section 5.2.4. Section 5.4 is devoted to the description of the spatio-temporal discretisation of the axisymmetric model. We then proceed to specify, in Section 5.4.1, the axisymmetric $\boldsymbol{H}(\mathrm{div})$-conforming method, where we first derive a semi-discrete (continuous in time) Galerkin formulation for the model problem, based on two-dimensional BDM spaces adapted to the axisymmtric setting, and then pass to a fully discrete scheme by applying a second-order time discretisation through an implicit backward differentiation formula. Next, in Section 5.4.2, we establish discrete stability properties of the bilinear and trilinear forms involved in the method. These properties allow us to prove (in Section 5.4.3) the existence of a discrete solution. Then, in Section 5.5, we prove an optimal a priori error estimate for the numerical scheme, where we verify that each of the terms is bounded optimally in the corresponding weighted space. Finally, in Section 5.6 we present numerical examples generated by the method introduced. Example 5.1 (Section 5.6.1) is an accuracy test with a manufactured known exact solution of (5.1) equipped with initial and boundary conditions. Results confirm that the method converges to the exact solution with the expected second-order rate. Next, in Example 5.2 (Section 5.6.2), numerical results are validated against experimental data, and

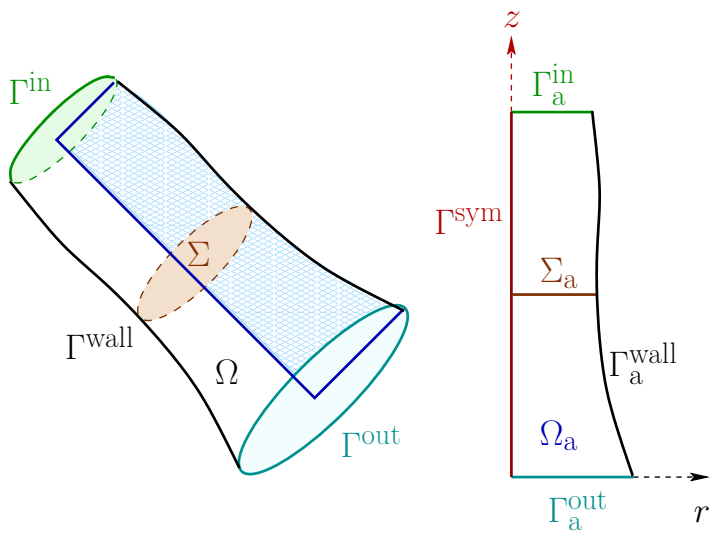


Figure 5.1: Left: schematic representation of the domain Ω, its various boundaries Γ^{in}, Γ^{wall} and Γ^{out}, and the material interface Σ. Right: reduction to the axisymmetric configuration (adapted from [15]).

in Example 5.3 (Section 5.6.3) we solve the full two-layer, two-contaminant filter model.

5.2 Model problem and preliminaries

5.2.1 Initial and boundary conditions

Let us consider a porous skeleton consisting of two different materials separated by an interface, where the matrix is saturated with an incompressible interstitial fluid (see a diagrammatic representation on the left part of Figure 5.1). The coupled set of governing equations (5.1) is posed along with the initial and boundary conditions

$$\boldsymbol{u} = \boldsymbol{u}^{\text{in}}, \quad \vec{\theta} = \vec{\theta}^{\text{in}} \qquad \text{on } \Gamma^{\text{in}} \times (0, \mathcal{T}], \tag{5.3a}$$

$$\boldsymbol{u} = \boldsymbol{0}, \quad \mathbb{D}\nabla\vec{\theta} \cdot \boldsymbol{n} = \vec{0} \qquad \text{on } \Gamma^{\text{wall}} \times (0, \mathcal{T}], \tag{5.3b}$$

$$(\nu\boldsymbol{\varepsilon}(\boldsymbol{u}) - p\mathbb{I})\boldsymbol{n} = \boldsymbol{0}, \quad \mathbb{D}\nabla\vec{\theta} \cdot \boldsymbol{n} = \vec{0} \qquad \text{on } \Gamma^{\text{out}} \times (0, \mathcal{T}], \tag{5.3c}$$

$$\vec{\theta}(0) = \vec{0}, \quad \boldsymbol{u}(0) = \vec{0}, \quad \vec{s}(0) = \vec{0} \qquad \text{in } \Omega. \tag{5.3d}$$

Condition (5.3a) indicates that the contaminated water enters the filter at Γ^{in} with a constant influx velocity, and each contaminant θ_i, $1 \leq i \leq m$ present at a fixed concentration θ_i^{in}; while condition (5.3c) accounts for zero normal stress and zero contaminant flux at the outlet. The system is preliminarily flushed with clean water and so there are no contaminants in the filter. Once the flow is at rest, we consider the initial conditions (5.3d).

The two distinct materials that compose the porous domain will have different permeability, porosity, as well as adsorption rate. Moreover, the diffusivities of the contaminants will vary from one type of porous structure to another. However it is important to remark that these differences in material

properties, at least in the applications we address here, are not large enough to modify the flow regime between the two subdomains and this explains why (5.1a)–(5.1d) are defined on the whole domain Ω. Should this not be the case, one needs to solve explicitly for the coupling of Navier-Stokes/Brinkman or Brinkman/Darcy equations including suitable transmission conditions at the interface (see for instance [17, 72] for formulations tailored to axisymmetric domains).

5.2.2 An axisymmetric formulation

Assuming that the data, the domain and the expected flow properties are all symmetric with respect to a given axis of symmetry denoted Γ^{sym}, we may rewrite the model equations in the *meridional* domain Ω_{a} (see the right part of Figure 5.1). In this case the velocity only possess radial and vertical components and we recall that the divergence operator in axisymmetric coordinates (in radial and height variables r, z) is

$$\mathrm{div}_{\mathrm{a}}\, \boldsymbol{v} := \partial_z v_z + \frac{1}{r}\partial_r(r v_r).$$

Then, making abuse of notation, we may rewrite system of PDEs (5.1) as

$$\rho_{\mathrm{f}}(\partial_t \boldsymbol{u} + \boldsymbol{u}\cdot\nabla\boldsymbol{u}) + \mathbb{K}^{-1}\nu\boldsymbol{u} - \mathbf{div}_{\mathrm{a}}(\nu\boldsymbol{\varepsilon}(\boldsymbol{u})) + \nabla p + \nu(\boldsymbol{u}_r/r^2)\boldsymbol{e}_1 = \boldsymbol{F}(\vec{\theta}), \tag{5.4a}$$

$$\mathrm{div}_{\mathrm{a}}\, \boldsymbol{u} = 0, \tag{5.4b}$$

$$\phi\partial_t\vec{\theta} - \mathrm{div}_{\mathrm{a}}(\mathbb{D}\nabla\vec{\theta}) + (\boldsymbol{u}\cdot\nabla)\vec{\theta} = -\rho_{\mathrm{b}}\partial_t\vec{s}, \tag{5.4c}$$

$$\partial_t\vec{s} = \boldsymbol{G}(\vec{s},\vec{\theta}) \qquad \text{for } (r,z,t)\in\Omega_{\mathrm{a}}\times(0,\mathcal{T}], \tag{5.4d}$$

while the corresponding initial and boundary conditions (5.3) take the form

$$\boldsymbol{u} = \boldsymbol{u}^{\mathrm{in}}, \quad \vec{\theta} = \vec{\theta}^{\mathrm{in}} \qquad \text{on } \Gamma_{\mathrm{a}}^{\mathrm{in}}\times(0,\mathcal{T}], \tag{5.5a}$$

$$\boldsymbol{u} = \boldsymbol{0}, \quad \mathbb{D}\nabla\vec{\theta}\cdot\boldsymbol{n} = \vec{0} \qquad \text{on } \Gamma_{\mathrm{a}}^{\mathrm{wall}}\times(0,\mathcal{T}], \tag{5.5b}$$

$$\boldsymbol{u}\cdot\boldsymbol{n} = 0, \quad \mathbb{D}\nabla\vec{\theta}\cdot\boldsymbol{n} = \vec{0} \qquad \text{on } \Gamma^{\mathrm{sym}}\times(0,\mathcal{T}], \tag{5.5c}$$

$$(\nu\boldsymbol{\varepsilon}(\boldsymbol{u}) - p\mathbb{I})\boldsymbol{n} = \boldsymbol{0}, \quad \mathbb{D}\nabla\vec{\theta}\cdot\boldsymbol{n} = \vec{0}, \qquad \text{on } \Gamma_{\mathrm{a}}^{\mathrm{out}}\times(0,\mathcal{T}], \tag{5.5d}$$

$$\vec{\theta}(0) = \vec{0}, \quad \boldsymbol{u}(0) = \boldsymbol{0}, \quad \vec{s}(0) = \vec{0} \qquad \text{in } \Omega_{\mathrm{a}}, \tag{5.5e}$$

where the condition (5.5c) at the symmetry axis indicates slip velocity and zero normal fluxes.

5.2.3 Preliminaries on spaces of radially symmetric functions

For $\alpha\in\mathbb{R}$ and $1\le p<\infty$, let $L^p_\alpha(\Omega_{\mathrm{a}})$ denote the space of measurable functions v on Ω_{a} such that

$$\|v\|^p_{L^p_\alpha(\Omega_{\mathrm{a}})} := \int_{\Omega_{\mathrm{a}}} |v|^p r^\alpha \,\mathrm{d}r\,\mathrm{d}z \le \infty,$$

and let us denote the scalar product in $L^2_\alpha(\Omega_{\mathrm{a}})$ by $(\cdot,\cdot)_{\alpha,\Omega_{\mathrm{a}}}$. Moreover we introduce $H^q_\alpha(\Omega_{\mathrm{a}})$ as the space of functions in $L^p_\alpha(\Omega_{\mathrm{a}})$ whose derivatives up to order q are also in $L^p_\alpha(\Omega_{\mathrm{a}})$, and we denote by $H^q_{\alpha,j}(\Omega_{\mathrm{a}})$ its restriction to functions with null trace on a given portion Γ^j_{a} of the boundary. By $\vec{L}$ we denote the corresponding vectorial counterpart of the scalar functional space L when the number

of components depends on m. Furthermore, the space $V_1^1(\Omega_\mathrm{a}) := H_1^1(\Omega_\mathrm{a}) \cap L_{-1}^2(\Omega_\mathrm{a})$ is endowed with the following norm and seminorm:

$$\|v\|_{V_1^1(\Omega_\mathrm{a})} := \big(\|v\|^2_{L_1^2(\Omega_\mathrm{a})} + |v|^2_{H_1^1(\Omega_\mathrm{a})} + \|v\|^2_{L^2_{-1}(\Omega_\mathrm{a})}\big)^{1/2},$$
$$|v|_{V_1^1(\Omega_\mathrm{a})} := \big(|v|^2_{H_1^1(\Omega_\mathrm{a})} + \|v\|^2_{L^2_{-1}(\Omega_\mathrm{a})}\big)^{1/2}.$$

Let us define the space

$$\boldsymbol{H}_0(\mathrm{div}_\mathrm{a}; \Omega_\mathrm{a}) := \big\{\boldsymbol{v} \in \boldsymbol{L}_1^2(\Omega_\mathrm{a}) : \mathrm{div}_\mathrm{a}\, \boldsymbol{v} \in L_1^2(\Omega_\mathrm{a}) \quad \text{and} \quad \boldsymbol{v}|_{\partial\Omega_\mathrm{a}} \cdot \boldsymbol{n} = 0\big\},$$

endowed with the following norm

$$\|\boldsymbol{v}\|_{\mathrm{div}_\mathrm{a},\Omega_\mathrm{a}} = \big(\|\boldsymbol{v}\|_{\boldsymbol{L}_1^2(\Omega_\mathrm{a})} + \|\mathrm{div}_\mathrm{a}(\boldsymbol{v})\|_{L_1^2(\Omega_\mathrm{a})}\big)^{1/2}.$$

The essential boundary conditions (5.5a), (5.5b)$_1$, (5.5c)$_1$ suggest to employ the functional spaces

$$\boldsymbol{V}^1_{1,\mathrm{in,wall}}(\Omega_\mathrm{a}) := \big\{\boldsymbol{v} \in V_1^1(\Omega_\mathrm{a}) \times H_1^1(\Omega_\mathrm{a}) : \boldsymbol{v}|_{\Gamma_\mathrm{a}^\mathrm{in} \cup \Gamma_\mathrm{a}^\mathrm{wall}} = \boldsymbol{0} \ \text{ and } \ \boldsymbol{v}|_{\Gamma^\mathrm{sym}} \cdot \boldsymbol{n} = 0\big\},$$
$$\vec{H}^1_{1,\mathrm{in}}(\Omega_\mathrm{a}) = \big\{\vec{\psi} \in \vec{H}_1^1(\Omega_\mathrm{a}) : \vec{\psi}|_{\Gamma_\mathrm{a}^\mathrm{in}} = \vec{0}\big\}.$$

In what follows, to make notation more concise, we write L_1^2 instead of $L_1^2(\Omega_\mathrm{a})$, and proceed similarly for $\boldsymbol{V}_1^1(\Omega_\mathrm{a})$, $\vec{L}_1^2(\Omega_\mathrm{a})$, $\vec{H}_1^1(\Omega_\mathrm{a})$, and other spaces of functions defined on Ω_a as well as their corresponding norms. That is, in the remainder any space of functions and corresponding norm whose domain is not specified is understood to refer to functions defined on Ω_a.

5.2.4 Weak formulation of the axisymmetric problem

For a fixed $t > 0$, the weak (variational) formulation of problem (5.4), (5.5) is obtained after testing against suitable functions and applying integration by parts in axisymmetric coordinates; and it can be formulated as follows:

Find $(\boldsymbol{u}(t), p(t), \vec{\theta}(t), s(t)) \in \boldsymbol{V}_1^1 \times L_1^2 \times \vec{H}_1^1 \times L_1^2$ such that (5.5a) holds, and

$$\big(\rho_\mathrm{f}\partial_t \boldsymbol{u}(t), \boldsymbol{v}\big)_{1,\Omega_\mathrm{a}} + a_1\big(\boldsymbol{u}(t), \boldsymbol{v}\big) + c_1\big(\boldsymbol{u}(t); \boldsymbol{u}(t), \boldsymbol{v}\big) + b\big(\boldsymbol{v}, p(t)\big) = d_1(\vec{\theta}, \boldsymbol{v}) \quad \text{for all } \boldsymbol{v} \in \boldsymbol{V}^1_{1,\mathrm{in,wall}}(\Omega_\mathrm{a}), \tag{5.6a}$$

$$b\big(\boldsymbol{u}(t), q\big) = 0 \quad \text{for all } q \in L_1^2, \tag{5.6b}$$

$$\big(\phi\partial_t \vec{\theta}(t), \vec{\psi}\big)_{1,\Omega_\mathrm{a}} + a_2\big(\vec{\theta}(t), \vec{\psi}\big) + c_2\big(\boldsymbol{u}(t); \vec{\theta}(t), \vec{\psi}\big) + d_2\big(s(t); \vec{\theta}(t), \vec{\psi}\big) = 0 \quad \text{for all } \vec{\psi} \in \vec{H}^1_{1,\mathrm{in}}(\Omega_\mathrm{a}), \tag{5.6c}$$

$$\big(\partial_t \vec{s}(t), \vec{l}\big)_{1,\Omega_\mathrm{a}} + d_3\big(\vec{\theta}(t); \vec{s}(t), \vec{l}\big) - d_4\big(\vec{\theta}(t), \vec{l}\big) = 0 \quad \text{for all } l \in \vec{L}_1^2, \tag{5.6d}$$

where the bilinear, trilinear, and nonlinear forms are defined as follows for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \boldsymbol{V}_1^1$, $q \in L_1^2$, $\vec{s}, \vec{l} \in \vec{L}_1^2$, and $\vec{\theta}, \vec{\psi} \in \vec{H}_1^1$:

$$
\begin{aligned}
a_1(\boldsymbol{u}, \boldsymbol{v}) &:= \int_{\Omega_\mathrm{a}} \mathbb{K}^{-1} \nu \boldsymbol{u} \cdot \boldsymbol{v} r \,\mathrm{d}r \,\mathrm{d}z + \int_{\Omega_\mathrm{a}} \nu \boldsymbol{\varepsilon}(\boldsymbol{u}) : \boldsymbol{\varepsilon}(\boldsymbol{v}) r \,\mathrm{d}r \,\mathrm{d}z + \int_{\Omega_\mathrm{a}} \frac{\nu}{r} u_r v_r \,\mathrm{d}r \,\mathrm{d}z, \\
a_2(\vec{\theta}, \vec{\psi}) &:= \int_{\Omega_\mathrm{a}} \mathbb{D} \nabla \vec{\theta} : \nabla \vec{\psi} r \,\mathrm{d}r \,\mathrm{d}z, \quad b(\boldsymbol{v}, q) := -\int_{\Omega_\mathrm{a}} q \,\mathrm{div}_\mathrm{a}\, \boldsymbol{v} r \,\mathrm{d}r \,\mathrm{d}z, \\
c_1(\boldsymbol{w}; \boldsymbol{u}, \boldsymbol{v}) &:= \int_{\Omega_\mathrm{a}} \rho_\mathrm{f} (\boldsymbol{w} \cdot \nabla) \boldsymbol{u} \cdot \boldsymbol{v} r \,\mathrm{d}r \,\mathrm{d}z, \quad c_2(\boldsymbol{v}; \vec{\theta}, \vec{\psi}) := \int_{\Omega_\mathrm{a}} (\boldsymbol{v} \cdot \nabla) \vec{\theta} \cdot \vec{\psi} r \,\mathrm{d}r \,\mathrm{d}z, \\
d_1(\vec{\psi}, \boldsymbol{v}) &:= \int_{\Omega_\mathrm{a}} \boldsymbol{F}(\vec{\psi}) \cdot \boldsymbol{v} r \,\mathrm{d}r \,\mathrm{d}z, \quad d_2(\vec{s}; \vec{\theta}, \vec{\psi}) := \int_{\Omega_\mathrm{a}} \sum_{i=1}^m (f(\boldsymbol{x}, s_i) \theta_i \psi_i) r \,\mathrm{d}r \,\mathrm{d}z, \\
d_3(\vec{\psi}; \vec{s}, \vec{l}) &:= \int_{\Omega_\mathrm{a}} \sum_{i=1}^m g(\boldsymbol{x}, \psi_i) s_i l_i r \,\mathrm{d}r \,\mathrm{d}z, \\
d_4(\vec{\psi}, \vec{l}) &:= \int_{\Omega_\mathrm{a}} \sum_{i=1}^m g(\boldsymbol{x}, \psi_i) s_i^{\max} l_i r \,\mathrm{d}r \,\mathrm{d}z.
\end{aligned}
$$

5.2.5 Further assumptions and preliminaries

The permeability tensor $\mathbb{K} \in [C(\overline{\Omega_\mathrm{a}})]^{d \times d}$ is assumed symmetric and uniformly positive definite, hence its inverse satisfies

$$\boldsymbol{v}^\mathrm{T} \mathbb{K}^{-1}(\boldsymbol{x}) \boldsymbol{v} \geq \alpha_1 |\boldsymbol{v}|^2 \quad \text{for all } \boldsymbol{v} \in \mathbb{R}^d \text{ and } \boldsymbol{x} \in \Omega_\mathrm{a}, \text{ for a constant } \alpha_1 > 0.$$

We also require $\mathbb{D}$ to be positive definite, i.e.,

$$\vec{\psi}^\mathrm{T} \mathbb{D} \vec{\psi} \geq \alpha_2 |\vec{\psi}|^2 \quad \text{for all } \vec{\psi} \in \mathbb{R}^m, \text{ for a constant } \alpha_2 > 0.$$

We assume there exist constants $f_1, f_2, g_1, g_2 > 0$ such that $f_1 \leq f(\boldsymbol{x}, s) \leq f_2$, $g_1 \leq g(\boldsymbol{x}, \theta) \leq g_2$, and that f and g are Lipschitz continuous and satisfy

$$\big| f(s_1) - f(s_2) \leq |f|_\mathrm{Lip} |s_1 - s_2|, \quad \big| g(\theta_1) - g(\theta_2) \big| \leq |g|_\mathrm{Lip} |\theta_1 - \theta_2|.$$

These assumptions imply that for all $\vec{s}_1, \vec{s}_2, \vec{s}, \vec{l} \in \vec{L}_1^2$ and $\vec{\theta}, \vec{\psi} \in \vec{H}_1^1$ such that $s_i^{\max} \leq s^{\max}$, there hold

$$d_2(\vec{s}; \vec{\theta}, \vec{\theta}) \geq f_1 \|\vec{\theta}\|_{\vec{L}_1^2}^2, \tag{5.7}$$

$$d_2(\vec{s}; \vec{\theta}, \vec{\psi}) \leq f_2 \|\vec{\theta}\|_{\vec{L}_1^2} \|\vec{\psi}\|_{\vec{L}_1^2}, \tag{5.8}$$

$$d_2(\vec{s}_2; \vec{\theta}, \vec{\psi}) - d_2(\vec{s}_1; \vec{\theta}, \vec{\psi}) \leq |f|_\mathrm{Lip} \|\vec{s}_2 - \vec{s}_1\|_{\vec{L}_1^2} \|\vec{\theta}\|_{\vec{H}_1^1} \|\vec{\psi}\|_{\vec{H}_1^1}, \tag{5.9}$$

$$d_3(\vec{\psi}; \vec{s}, \vec{s}) \geq g_1 \|\vec{s}\|_{\vec{L}_1^2}^2, \tag{5.10}$$

$$d_3(\vec{\psi}; \vec{s}, \vec{l}) \leq g_2 \|\vec{s}\|_{\vec{L}_1^2} \|\vec{l}\|_{\vec{L}_1^2}, \tag{5.11}$$

$$d_4(\vec{\psi}, \vec{l}) \leq g_2 s^{\max} \|\vec{l}\|_{\vec{L}_1^2} \leq C_d \|\vec{l}\|_{\vec{L}_1^2}. \tag{5.12}$$

If in addition $\vec{s} \in \vec{H}_1^1$, we also get

$$d_3(\vec{\theta}_2; \vec{s}, \vec{l}) - d_3(\vec{\theta}_1; \vec{s}, \vec{l}) \leq |g|_\mathrm{Lip} \|\vec{\theta}_2 - \vec{\theta}_1\|_{\vec{H}_1^1} \|\vec{s}\|_{\vec{H}_1^1} \|\vec{l}\|_{\vec{L}_1^2}. \tag{5.13}$$

Due to the uniform boundedness of $\mathbb{K}^{-1}$ and $\mathbb{D}$, one can easily establish the following properties for all $\boldsymbol{u}, \boldsymbol{v}, \in \boldsymbol{V}_1^1$, $q \in L_1^2$, and $\vec{\theta}, \vec{\psi} \in \vec{H}^1$:

$$\big|a_1(\boldsymbol{u}, \boldsymbol{v})\big| \leq C_a \|\boldsymbol{u}\|_{\boldsymbol{V}_1^1} \|\boldsymbol{v}\|_{\boldsymbol{V}_1^1}, \tag{5.14a}$$

$$\big|a_2(\vec{\theta}, \vec{\psi})\big| \leq \hat{C}_a \|\vec{\theta}\|_{\vec{H}_1^1} \|\vec{\psi}\|_{\vec{H}_1^1}, \tag{5.14b}$$

$$\big|b(\boldsymbol{v}, q)\big| \leq C_b \|\boldsymbol{v}\|_{\boldsymbol{V}_1^1} \|q\|_{L_1^2}, \tag{5.14c}$$

$$\big|d_1(\vec{\theta}, \boldsymbol{v})\big| \leq C_F \|\vec{\theta}\|_{\vec{H}_1^1} \|\boldsymbol{v}\|_{\boldsymbol{V}_1^1}. \tag{5.14d}$$

Moreover, thanks to the axisymmetric version of the well-known Sobolev embeddings (see [20, 133]), we have that for $\hat{p} \geq 1$,

$$\|\boldsymbol{w}\|_{\boldsymbol{L}_1^{\hat{p}}} \leq C_{\hat{p}}^* \|\boldsymbol{w}\|_{\boldsymbol{V}_1^1} \quad \text{for all } \boldsymbol{w} \in \mathbf{V}_1^1, \tag{5.15}$$

where the constant $C_{\hat{p}}^* > 0$ depends only upon $|\Omega_{\mathrm{a}}|$ and $\hat{p}$. Also, for $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \boldsymbol{H}_1^1$ and $\vec{\theta}, \vec{\psi} \in \vec{H}_1^1$, Hölder's inequality and (5.15) with $\frac{1}{\hat{p}} + \frac{1}{\hat{p}^*} = \frac{1}{2}$ imply that (see [49])

$$\begin{aligned}
\big|c_1(\boldsymbol{w}; \boldsymbol{u}, \boldsymbol{v})\big| &\leq C_v \|\boldsymbol{w}\|_{\boldsymbol{H}_1^1} \|\boldsymbol{u}\|_{\boldsymbol{H}_1^1} \|\boldsymbol{v}\|_{\boldsymbol{H}_1^1}, \\
\big|c_2(\boldsymbol{w}; \vec{\theta}, \vec{\psi})\big| &\leq \bar{C}_v \|\boldsymbol{w}\|_{\boldsymbol{H}_1^1} \|\vec{\theta}\|_{\boldsymbol{H}_1^1} \|\vec{\psi}\|_{\vec{L}_1^3}, \\
\big|c_2(\boldsymbol{w}; \vec{\theta}, \vec{\psi})\big| &\leq \hat{C}_v \|\boldsymbol{w}\|_{\boldsymbol{H}_1^1} \|\vec{\theta}\|_{\vec{H}_1^1} \|\vec{\psi}\|_{\vec{H}_1^1}.
\end{aligned}$$

Next, Poincaré's inequality and the positive definiteness of $\mathbb{D}$ readily imply the following coercivities (see [32, Chapter IX]):

$$a_1(\boldsymbol{v}, \boldsymbol{v}) \geq \alpha_a \|\boldsymbol{v}\|_{\boldsymbol{V}_1^1}^2 \qquad \text{for all } \boldsymbol{v} \in \boldsymbol{V}_{\mathrm{in,wall}}^1(\Omega_{\mathrm{a}}), \tag{5.16}$$

$$a_2(\vec{\psi}, \vec{\psi}) \geq \hat{\alpha}_a \|\vec{\psi}\|_{\vec{H}_1^1}^2 \qquad \text{for all } \vec{\psi} \in \vec{H}_{1,\mathrm{in}}^1(\Omega_{\mathrm{a}}). \tag{5.17}$$

We then proceed to characterise the kernel of the bilinear form $b(\cdot, \cdot)$ as

$$\begin{aligned}
\boldsymbol{Z} &:= \big\{ \boldsymbol{v} \in \boldsymbol{V}_{1,\mathrm{in,wall}}^1(\Omega_{\mathrm{a}}) : b(\boldsymbol{v}, q) = 0 \text{ for all } q \in L_1^2 \big\} \\
&= \big\{ \boldsymbol{v} \in \boldsymbol{V}_{1,\mathrm{in,wall}}^1(\Omega_{\mathrm{a}}) : \mathrm{div}_{\mathrm{a}}\, \boldsymbol{v} = 0 \text{ a.e. in } \Omega_{\mathrm{a}} \big\},
\end{aligned}$$

and using integration by parts directly implies the relations (see [32, Section IX.2])

$$\begin{aligned}
&c_1(\boldsymbol{w}; \boldsymbol{v}, \boldsymbol{v}) = 0 \quad \text{and} \quad c_2(\boldsymbol{w}; \vec{\psi}, \vec{\psi}) = 0 \\
&\text{for all } \boldsymbol{w} \in \boldsymbol{Z},\ \boldsymbol{v} \in \boldsymbol{V}_{1,\mathrm{in,wall}}^1(\Omega_{\mathrm{a}}), \text{ and } \vec{\psi} \in \vec{H}_{1,\mathrm{in}}^1(\Omega_{\mathrm{a}}).
\end{aligned} \tag{5.18}$$

Note that for a given $\boldsymbol{w} \in \boldsymbol{Z}$, property (5.16) together with (5.18) readily lead to the ellipticity of the bilinear form

$$a_1(\cdot, \cdot) + c_1(\boldsymbol{w}, \cdot, \cdot) \colon \boldsymbol{V}_{1,\mathrm{in,wall}}^1(\Omega_{\mathrm{a}}) \times \boldsymbol{V}_{1,\mathrm{in,wall}}^1(\Omega_{\mathrm{a}}) \to \mathbb{R}.$$

Moreover, it is well known (i.e. [32, Proposition IX.1.1]) that an inf-sup condition holds for $b(\cdot, \cdot)$ in the following sense:

$$\sup_{\boldsymbol{v} \in \boldsymbol{V}_{1,\mathrm{in,wall}}^1(\Omega_{\mathrm{a}}) \setminus \{\boldsymbol{0}\}} \frac{b(\boldsymbol{v}, q)}{\|\boldsymbol{v}\|_{\boldsymbol{V}_1^1}} \geq \beta \|q\|_{L_1^2} \quad \text{for all } q \in L_1^2.$$

5.3 Well-posedness analysis of the continuous problem

This part of our analysis will be restricted to the case of no-slip velocity boundary conditions on the whole boundary. Then we introduce the spaces

$$H^1_{1,\diamond} := \{w \in H^1_1 : w = 0 \text{ on } \partial\Omega_{\mathrm{a}}\}, \quad V^1_{1,\diamond} := \{w \in V^1_1 : w = 0 \text{ on } \partial\Omega_{\mathrm{a}}\},$$

and $\boldsymbol{V}^1_{1,\diamond} := V^1_{1,\diamond} \times H^1_{1,\diamond}$.

From [111], we recall the weighted Sobolev inequality:

Lemma 5.1. *For all $v \in \boldsymbol{H}^1_1$ there holds*

$$\|\boldsymbol{v}\|^2_{\boldsymbol{L}^4_1} \leq \hat{C}\|\boldsymbol{v}\|_{\boldsymbol{L}^2_1}|\boldsymbol{v}|_{\boldsymbol{H}^1_1}.$$

We will also use the following lemma (for its proof in the axysimmetric case we refer the reader to [32, Chapter IX]):

Lemma 5.2. *If $(\boldsymbol{u}, p, \vec{\theta}, \vec{s}) \in \boldsymbol{V}^1_{1,\diamond} \times L^2_1 \times \vec{H}^1_{1,\diamond} \times \vec{L}^2_1$ solves (5.6), then $\boldsymbol{u} \in \boldsymbol{Z}$ is a solution of the following reduced problem:*

$$\begin{aligned}
&\text{For all } t \in (0, \mathcal{T}], \text{ find } (\boldsymbol{u}, \vec{\theta}, s) \in \boldsymbol{Z} \times \vec{H}^1_{1,\diamond} \times \vec{L}^2_1 \text{ such that}\\
&\big(\rho_{\mathrm{f}}\partial_t \boldsymbol{u}(t), \boldsymbol{v}\big)_{1,\Omega_{\mathrm{a}}} + a_1\big(\boldsymbol{u}(t), \boldsymbol{v}\big)\\
&\qquad + c_1\big(\boldsymbol{u}(t); \boldsymbol{u}(t), \boldsymbol{v}\big) = d_1(\vec{\theta}, \boldsymbol{v}) \quad \text{for all } \boldsymbol{v} \in \boldsymbol{V}^1_{1,\mathrm{in,wall}}(\Omega_{\mathrm{a}}), && (5.19\text{a})\\
&\big(\phi\partial_t \vec{\theta}(t), \vec{\psi}\big)_{1,\Omega_{\mathrm{a}}} + a_2\big(\vec{\theta}(t), \vec{\psi}\big)\\
&\qquad + c_2\big(\boldsymbol{u}(t); \vec{\theta}(t), \vec{\psi}\big) + d_2\big(\vec{s}(t); \vec{\theta}(t), \vec{\psi}\big) = 0 \quad \text{for all } \vec{\psi} \in \vec{H}^1_{1,\mathrm{in}}(\Omega_{\mathrm{a}}), && (5.19\text{b})\\
&\big(\partial_t \vec{s}(t), \vec{l}\big)_{1,\Omega_{\mathrm{a}}} + d_3\big(\vec{\theta}(t); \vec{s}(t), \vec{l}\big) - d_4\big(\vec{\theta}(t), \vec{l}\big) = 0 \quad \text{for all } \vec{l} \in \vec{L}^2_1. && (5.19\text{c})
\end{aligned}$$

Conversely, if $(\boldsymbol{u}, \vec{\theta}, \vec{s}) \in \boldsymbol{Z} \times \vec{H}^1_{1,\diamond} \times \vec{L}^2_1$ is a solution of (5.19), then there exists a pressure $p \in L^2_1$ such that $(\boldsymbol{u}, p, \vec{\theta}, \vec{s})$ is a solution of (5.6).

A similar problem of (5.6) but in Cartesian coordinates has been studied in [3]. The authors showed the existence of the solution by using the Galerkin method and applying the Cauchy-Lipschitz theorem. The proof of the existence of the solution of (5.6) can be showed by using the same method noting that $\boldsymbol{F}$ is a Lipcshitz-continuous function; and using equivalent imbeddings stated for weighted Sobolev spaces in [111] and weighted Poincaré like inequalities in [159, Section 4.3].

Theorem 5.1. *Assume that for $r \geq 4$,*

$$(\boldsymbol{u}, \vec{\theta}, s) \in L^2\big(0, \mathcal{T}; \boldsymbol{Z} \cap W^{1,r}_1(\Omega_{\mathrm{a}})\big) \times L^2(0, \mathcal{T}; \vec{H}^1_{1,\diamond}) \times L^2\big(0, \mathcal{T}; \vec{H}^1_1\big)$$

is a solution to problem (5.19). Then such solution is unique.

Proof. Throughout the proof, and for simplicity of the presentation, we assume that the model constants are scaled as $\phi, \rho_{\mathrm{b}}, \rho_{\mathrm{f}} = 1$. Let $(\boldsymbol{u}_1, \vec{\theta}_1, s_1)$ and $(\boldsymbol{u}_2, \vec{\theta}_2, s_2)$ be two solutions of (5.19). We denote

$$\boldsymbol{U} := \boldsymbol{u}_1 - \boldsymbol{u}_2, \quad \vec{\Theta} := \vec{\theta}_1 - \vec{\theta}_2, \quad \text{and} \quad \vec{S} := \vec{s}_1 - \vec{s}_2.$$

Now, from (5.19b), by adding and subtracting $c_2(\boldsymbol{u}_2, \vec{\theta}_1, \vec{\Theta})$ and $d_2(\vec{s}_2, \vec{\theta}_1, \vec{\Theta})$; and using properties (5.18) and (5.9) we obtain

$$\begin{aligned}
&(\partial_t \vec{\Theta}, \vec{\Theta})_{1,\Omega_{\mathrm{a}}} + a_2(\vec{\Theta}, \vec{\Theta}) \\
&= -c_2(\boldsymbol{U}; \vec{\theta}_1, \vec{\Theta}) - d_2(\vec{s}_2; \vec{\Theta}, \vec{\Theta}) - d_2(\vec{s}_1; \vec{\theta}_1, \vec{\Theta}) + d_2(\vec{s}_2, \vec{\theta}_1, \vec{\Theta}), \\
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\vec{\Theta}\|^2_{\vec{L}^2_1} + \alpha_2 |\vec{\Theta}|^2_{\vec{H}^1_1} \leq \|\boldsymbol{U}\|^2_{\vec{L}^4_1} |\vec{\theta}_1|_{\vec{H}^1_1} \|\vec{\Theta}\|_{\vec{L}^4} + |f|_{\mathrm{Lip}} \|\vec{S}\|_{\vec{L}^2_1} \|\vec{\theta}_1\|_{\vec{H}^1_1} \|\vec{\Theta}\|_{\vec{L}^4_1}.
\end{aligned}$$

By Lemma 5.1 and Young's inequality it follows that

$$\begin{aligned}
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\vec{\Theta}\|^2_{\vec{L}^2_1} + \alpha_2 |\vec{\Theta}|^2_{\vec{H}^1_1} \\
&\leq \frac{\hat{C}}{4}\left(\varepsilon_1 |\boldsymbol{U}|^2_{\boldsymbol{H}^1_1} + \frac{1}{\varepsilon_1}|\vec{\theta}_1|_{\vec{H}^1_1} \|\boldsymbol{U}\|^2_{\boldsymbol{L}^2_1} + \varepsilon_2 |\vec{\Theta}|^2_{\vec{H}^1_1} + \frac{1}{\varepsilon_2}|\vec{\theta}_1|^2_{\vec{H}^1_1} \|\vec{\Theta}\|^2_{\vec{L}^2_1}\right).
\end{aligned} \tag{5.20}$$

Now, selecting $\boldsymbol{v} = \boldsymbol{U}$ in (5.19a), adding and subtracting $c_1(\boldsymbol{u}_2; \boldsymbol{u}_1; \boldsymbol{U})$, we obtain

$$(\partial_t \boldsymbol{U}, \boldsymbol{U})_{1,\Omega_{\mathrm{a}}} + a_1(\boldsymbol{U}, \boldsymbol{U}) + c_1(\boldsymbol{u}_2; \boldsymbol{U}, \boldsymbol{U}) = -c_1(\boldsymbol{U}; \boldsymbol{u}_1, \boldsymbol{U}) + d_1(\vec{\Theta}, \boldsymbol{U}).$$

Employing properties (5.18) and (5.14d), we can readily see that

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{U}\|^2_{\boldsymbol{L}^2_1} + \nu \|\boldsymbol{\varepsilon}(\boldsymbol{U})\|^2_{\mathbb{L}^2_1} + \nu \|\boldsymbol{U}_r\|^2_{L^2_{-1}} \leq \|\boldsymbol{U}\|^2_{\boldsymbol{L}^4_1} |\boldsymbol{u}_1|_{\boldsymbol{H}^1_1} + C_{\boldsymbol{F}} |\vec{\Theta}|_{\vec{L}^2_1} \|\boldsymbol{U}\|_{L^2_1}.$$

Applying Lemma 5.1 and Young's inequality we conclude that

$$\begin{aligned}
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{U}\|^2_{\boldsymbol{L}^2_1} + \alpha_a \|\boldsymbol{U}\|^2_{\boldsymbol{V}^1_1} \\
&\leq \frac{\hat{C}\varepsilon_3}{2}|\boldsymbol{U}|^2_{\boldsymbol{V}^1_1} + \frac{\hat{C}}{2\varepsilon_3}\|\boldsymbol{U}\|^2_{L^2_1} |\boldsymbol{u}_1|_{\boldsymbol{H}^1_1} + \frac{C_{\boldsymbol{F}}}{2}\big(|\vec{\Theta}|^2_{\vec{L}^2_1} + \|\boldsymbol{U}\|^2_{\boldsymbol{L}^2_1}\big).
\end{aligned} \tag{5.21}$$

In the same manner, from (5.19c), after adding and subtracting $d_3(\vec{\theta}_2; \vec{s}_1, \vec{S})$, we infer

$$(\partial_t \vec{S}, \vec{S})_{1,\Omega_{\mathrm{a}}} + d_3(\vec{\theta}_2; \vec{S}, \vec{S}) = -d_3(\vec{\Theta}; \vec{s}_1, \vec{S}) - d_4(\vec{\Theta}, \vec{S}).$$

Using (5.10), (5.11), (5.12) and (5.13), we can assert that

$$\begin{aligned}
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\vec{S}\|_{\vec{L}^2_1} + g_1 \|\vec{S}\|_{\vec{L}^2_1} \\
&\quad \leq \frac{|g|_{\mathrm{Lip}}}{2}\big(\|\vec{\Theta}\|^2_{L^2_1} + \|\vec{s}_1\|^2_{\vec{H}^1_1} \|\vec{S}\|^2_{L^2_1}\big) + \frac{|g|_{\mathrm{Lip}} s^{\max}}{2}\big(\|\vec{\Theta}\|^2_{\vec{L}^2_1} + \|\vec{S}\|^2_{\vec{L}^2_1}\big),
\end{aligned} \tag{5.22}$$

and choosing $\varepsilon_1 = 2\nu\alpha_a/\hat{C}$, $\varepsilon_2 = 2\alpha_2/\hat{C}$ and $\varepsilon_3 = \nu\alpha_a/\hat{C}$, we obtain from (5.20), (5.21) and (5.22) that

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\big(\|\boldsymbol{U}\|^2_{\boldsymbol{L}^2_1} + \|\vec{\Theta}\|^2_{\vec{L}^2_1} + \|\vec{S}\|^2_{\vec{L}^2_1}\big) \\
&\leq C\big(|\boldsymbol{u}_1|^2_{\boldsymbol{H}^1_1} + |\vec{\theta}_1|^2_{\vec{H}^1_1} + \|\vec{\theta}_1\|^2_{\vec{L}^1_2} + \|\vec{s}_1\|^2_{\vec{H}^1_1} + 1\big)\big(\|\boldsymbol{U}\|^2_{\boldsymbol{L}^2_1} + \|\vec{\Theta}\|^2_{\vec{L}^2_1} + \|\vec{S}\|^2_{\vec{L}^2_1}\big).
\end{aligned}$$

We may now integrate from $\tau = 0$ to $\tau = t$ to infer the bound

$$\begin{aligned}
&\|\boldsymbol{U}\|^2_{\boldsymbol{L}^2_1} + \|\vec{\Theta}\|^2_{\vec{L}^2_1} + \|\vec{S}\|^2_{\vec{L}^2_1} \\
&\leq \int_0^t C\big(|\boldsymbol{u}_1|^2_{\boldsymbol{H}^1_1} + |\vec{\theta}_1|^2_{\vec{H}^1_1} + \|\vec{\theta}_1\|^2_{\vec{L}^1_2} + \|\vec{s}_1\|^2_{\vec{H}^1_1} + 1\big)\big(\|\boldsymbol{U}\|^2_{\boldsymbol{L}^2_1} + \|\vec{\Theta}\|^2_{\vec{L}^2_1} + \|\vec{S}\|^2_{\vec{L}^2_1}\big)\,\mathrm{d}\tau.
\end{aligned}$$

Applying Gronwall's lemma, we now conclude that $\boldsymbol{U} = \boldsymbol{0}$, $\vec{\Theta} = \vec{0}$ and $\vec{S} = \vec{0}$. □

5.4 Spatio-temporal discretisation

5.4.1 An axisymmetric $\boldsymbol{H}(\mathrm{div})$-conforming method

Let us denote by $\mathcal{T}_h$ a regular partition of Ω_a composed by triangular elements K of diameter h_K. First, we recall the definition of the two-dimensional BDM spaces (see e.g. [40]) locally on an element $K \in \mathcal{T}_h$, $\mathrm{BDM}_k(K) := (\mathcal{P}_k(K))^2$, where $\mathcal{P}_k(K)$ denotes the local space spanned by polynomials of degree up to k. In turn, related to the axisymmetric setting, as in [71] we define

$$\begin{aligned}\mathrm{BDM}_k^{\mathrm{axi}}(K) :=& \big\{\boldsymbol{v} \in \mathrm{BDM}_k(K) : \boldsymbol{v}\cdot\boldsymbol{n}_K|_{\Gamma^{\mathrm{sym}}} = 0\big\}\\ =& \big\{(v_r, v_z)^{\mathrm{T}} \in \mathrm{BDM}_k : v_r|_{\Gamma^{\mathrm{sym}}} = 0\big\},\end{aligned}$$

where the associated degrees of freedom are given by

$$\begin{aligned}\int_{\mathcal{E}_h} \boldsymbol{v}\cdot\boldsymbol{n}_K p r \,\mathrm{d}s, &\quad p \in R_k(\partial K) \quad \text{for } k \geq 0,\\ \int_K \boldsymbol{v}\cdot\nabla p r \,\mathrm{d}r\,\mathrm{d}z, &\quad p \in \mathcal{P}_{k-1}(K) \quad \text{for } k \geq 1,\\ \int_K \boldsymbol{v}\cdot\mathbf{curl}(b_K p) r \,\mathrm{d}r\,\mathrm{d}z, &\quad p \in \mathcal{P}_{k-2}(K) \quad \text{for } k \geq 2,\end{aligned}$$

where b_K denotes a bubble function on the element K and

$$R_k(\partial K) := \big\{\phi \in L^2(\partial K) : \phi|_e \in \mathcal{P}_k(e), e \in \mathcal{E}_h(K)\big\}.$$

Then, globally, for an integer k and a mesh $\mathcal{T}_h$ on Ω, we utilize the discrete spaces

$$\begin{aligned}\boldsymbol{\mathcal{H}}_h^k &:= \big\{\boldsymbol{v}_h \in \boldsymbol{H}(\mathrm{div}_\mathrm{a};\Omega_\mathrm{a}) : \boldsymbol{v}_h|_K \in \mathrm{BDM}_k^{\mathrm{axi}}(K) \text{ for all } K \in \mathcal{T}_h\big\},\\ \mathcal{Y}_h^k &:= \big\{q_h \in L_1^2(\Omega_\mathrm{a}) : q_h|_K \in \mathcal{P}_k(K) \text{ for all } K \in \mathcal{T}_h\big\},\\ \mathcal{M}_h^k &:= \big\{\psi_h \in C(\overline{\Omega_\mathrm{a}}) : \psi_h|_K \in \mathcal{P}_k(K) \text{ for all } K \in \mathcal{T}_h\big\},\end{aligned}$$

to define the following finite element subspaces for the approximation of the unknowns $\boldsymbol{u}, p, \vec{\theta}$ and $\vec{s}$, respectively, where the polynomial degree is $k \geq 1$:

$$\begin{aligned}\boldsymbol{V}_h &:= \boldsymbol{\mathcal{H}}_h^k \cap \boldsymbol{H}_0(\mathrm{div}_\mathrm{a};\Omega_\mathrm{a}), \quad \mathcal{Q}_h := \mathcal{Y}_h^{k-1},\\ \vec{\mathcal{M}}_{h,0} &:= \vec{\mathcal{M}}_h^k \cap \vec{H}_{1,\mathrm{in}}^1(\Omega), \quad \vec{\mathcal{S}}_h := \vec{\mathcal{Y}}_h^{k-1}.\end{aligned}$$

Let us recall that for axisymmetric cases the property $\mathrm{div}_\mathrm{a}\,\boldsymbol{V}_h \subseteq \mathcal{Q}_h$ is not preserved [72], and let us also recall from [71] the following discrete inf-sup condition for $b(\cdot,\cdot)$, where $\tilde{\beta}$ is independent of h:

$$\sup_{\boldsymbol{v}_h \in \boldsymbol{V}_h\setminus\{\boldsymbol{0}\}} \frac{b(\boldsymbol{v}_h, q_h)}{\|\boldsymbol{v}_h\|_{1,\mathcal{T}_h}} \geq \tilde{\beta}\|q_h\|_{L_{1,0}^2(\Omega_\mathrm{a})} \quad \text{for all } q_h \in \mathcal{Q}_h. \tag{5.23}$$

Associated with these finite-dimensional spaces, we state the following semi-discrete Galerkin for-

mulation for problem (5.1), (5.3):

$$
\begin{aligned}
&\text{For a fixed } t>0, \text{ find } (\boldsymbol{u}_h(t), p_h(t), \vec{\theta}_h(t), \vec{s}_h(t)) \in \boldsymbol{V}_h \times \mathcal{Q}_h \times \vec{\mathcal{M}}_{h,0} \times \vec{\mathcal{S}}_h \\
&\text{such that for all } (\boldsymbol{v}_h, q_h, \vec{\psi}_h, \vec{l}_h) \in \boldsymbol{V}_h \times \mathcal{Q}_h \times \vec{\mathcal{M}}_{h,0} \times \vec{\mathcal{S}}_h: \\
&\big(\rho_{\mathrm{f}} \partial_t \boldsymbol{u}_h(t), \boldsymbol{v}\big)_{1,\Omega_{\mathrm{a}}} + a_1^h\big(\boldsymbol{u}_h(t), \boldsymbol{v}_h\big) \\
&\qquad + c_1^h\big(\boldsymbol{u}_h(t); \boldsymbol{u}_h(t), \boldsymbol{v}_h\big) + b\big(\boldsymbol{v}_h, p_h(t)\big) = d_1\big(\vec{\theta}_h(t), \boldsymbol{v}_h\big), \\
&b\big(\boldsymbol{u}_h(t), q_h\big) = 0, \\
&\big(\phi \partial_t \vec{\theta}_h(t), \vec{\psi}\big)_{1,\Omega_{\mathrm{a}}} + a_2\big(\vec{\theta}_h(t), \vec{\psi}_h\big) \\
&\qquad + c_2^h\big(\boldsymbol{u}_h(t); \vec{\theta}_h(t), \vec{\psi}_h\big) = d_2\big(\vec{s}_h(t); \vec{\theta}_h(t), \vec{\psi}_h\big), \\
&\big(\partial_t \vec{s}_h(t), \vec{l}_h\big)_{1,\Omega_{\mathrm{a}}} + d_3\big(\vec{\theta}_h(t); \vec{s}_h(t), \vec{l}_h\big) = d_4\big(\vec{\theta}_h(t), \vec{l}_h\big).
\end{aligned}
\tag{5.24}
$$

Here the discrete versions of the trilinear forms $a_1(\cdot,\cdot)$, $c_1(\cdot;\cdot,\cdot)$ and $c_2(\cdot;\cdot,\cdot)$ are defined using a symmetric interior penalty, an upwind approach and a skew-symmetric form, respectively (see e.g. [46, 102, 109]):

$$
\begin{aligned}
a_1^h(\boldsymbol{u}, \boldsymbol{v}) &:= \int_{\Omega_{\mathrm{a}}} \Big(\mathbb{K}^{-1} \boldsymbol{u} \cdot \boldsymbol{v} + \nu \boldsymbol{\varepsilon}_h(\boldsymbol{u}) : \boldsymbol{\varepsilon}_h(\boldsymbol{v}) + \nu \frac{\boldsymbol{u}_r}{r} \frac{\boldsymbol{v}_r}{r} \Big) r \,\mathrm{d}r \,\mathrm{d}z \\
&\quad - \sum_{e \in \mathcal{E}_h} \int_e \Big(\{\!\{ \nu \boldsymbol{\varepsilon}_h(\boldsymbol{u}) \boldsymbol{n}_e \}\!\} \cdot [\![\boldsymbol{v}_{\boldsymbol{\tau}}]\!] - \{\!\{ \nu \boldsymbol{\varepsilon}_h(\boldsymbol{v}) \boldsymbol{n}_e \}\!\} \cdot [\![\boldsymbol{u}_{\boldsymbol{\tau}}]\!] \\
&\qquad\qquad + \frac{a_0}{h_e} \nu [\![\boldsymbol{u}_{\boldsymbol{\tau}}]\!] \cdot [\![\boldsymbol{v}_{\boldsymbol{\tau}}]\!] \Big) r \,\mathrm{d}s, \\
c_1^h(\boldsymbol{w}; \boldsymbol{u}, \boldsymbol{v}) &:= \frac{1}{2} \int_{\Omega_{\mathrm{a}}} \big((\boldsymbol{w} \cdot \nabla_h) \boldsymbol{u} \cdot \boldsymbol{v} - (\boldsymbol{w} \cdot \nabla_h) \boldsymbol{v} \cdot \boldsymbol{u} \big) r \,\mathrm{d}r \,\mathrm{d}z \\
&\quad + \sum_{e \in \mathcal{E}_h} \int_e \hat{\boldsymbol{w}}^{\mathrm{up}}(\boldsymbol{u}) \cdot \boldsymbol{v} r \,\mathrm{d}s, \\
c_2^h(\boldsymbol{u}_h; \vec{\theta}_h, \vec{\psi}_h) &:= \frac{1}{2} \left(\int_{\Omega_{\mathrm{a}}} (\boldsymbol{v} \cdot \nabla_h) \vec{\theta} \cdot \vec{\psi} r \,\mathrm{d}r \,\mathrm{d}z - \int_{\Omega_{\mathrm{a}}} (\boldsymbol{v} \cdot \nabla_h) \vec{\psi} \cdot \vec{\theta} r \,\mathrm{d}r \,\mathrm{d}z \right),
\end{aligned}
$$

where the fluxes are defined as

$$
\hat{\boldsymbol{w}}^{\mathrm{up}}(\boldsymbol{u}) := \frac{1}{2} \big(\boldsymbol{w} \cdot \boldsymbol{n}_K - |\boldsymbol{w} \cdot \boldsymbol{n}_K| \big) (\boldsymbol{u}^e - \boldsymbol{u}),
$$

and $\boldsymbol{u}^e$ denotes the trace of $\boldsymbol{u}$ taken from within the exterior of K.

We then proceed with the method of lines, and for the time discretisation we partition the interval $[0, \mathcal{T}]$ into N subintervals $[t_{n-1}, t_n]$ of length Δt. We will use an implicit, second-order backward differentiation formula (BDF2). Starting from the interpolates $\boldsymbol{u}_h^0$, $\vec{\theta}_h^0$ and $\vec{s}_h^0$ of the initial data on$\boldsymbol{V}_h$,

$\vec{\mathcal{M}}_{h,0}$ and $\vec{\mathcal{S}}_h$, respectively, we solve for $n = 1, \dots, N-1$ the nonlinear system

$$\begin{aligned}
&\left(\boldsymbol{u}_h^{n+1} - \frac{4}{3}\boldsymbol{u}_h^n + \frac{1}{3}\boldsymbol{u}_h^{n-1}, \boldsymbol{v}_h\right)_{1,\Omega_\mathrm{a}} \\
&= \frac{2}{3}\Delta t\big(d_1(\vec{\theta}_h^{n+1}, \boldsymbol{v}_h) - a_1^h(\boldsymbol{u}_h^{n+1}, \boldsymbol{v}_h) - c_1^h(\boldsymbol{u}_h^{n+1}; \boldsymbol{u}_h^{n+1}, \boldsymbol{v}_h) - b(\boldsymbol{v}_h, p_h^{n+1})\big), \\
&b(\boldsymbol{u}_h^{n+1}, q_h) = 0, \\
&\left(\vec{\theta}_h^{n+1} - \frac{4}{3}\vec{\theta}_h^n + \frac{1}{3}\vec{\theta}_h^{n-1}, \vec{\psi}_h\right)_{1,\Omega_\mathrm{a}} \\
&= \frac{2}{3}\Delta t\big(-d_2(s_h^{n+1}; \vec{\theta}_h^{n+1}, \vec{\psi}_h) - a_2(\vec{\theta}_h^{n+1}, \vec{\psi}_h) - c_2^h(\boldsymbol{u}_h^{n+1}; \vec{\theta}_h^{n+1}, \vec{\psi}_h)\big), \\
&\left(\vec{s}_h^{n+1} - \frac{4}{3}\vec{s}_h^n + \frac{1}{3}\vec{s}_h^{n-1}, \vec{l}_h\right)_{1,\Omega_\mathrm{a}} \\
&= \frac{2}{3}\Delta t\big(-d_3(\vec{\theta}_h^{n+1}; \vec{s}_h^{n+1}, \vec{l}_h) + d_4(\vec{\theta}_h^{n+1}, \vec{l}_h)\big)
\end{aligned} \tag{5.25}$$

for all $\boldsymbol{v}_h \in \boldsymbol{V}_h, q_h \in \mathcal{Q}_h, \vec{\psi}_h \in \vec{\mathcal{M}}_h$ and $\vec{s}_h \in \vec{\mathcal{S}}_h$.

Then, in a way analogous to the continuous case, we define the discrete kernel

$$\boldsymbol{Z}_h := \big\{\boldsymbol{v}_h \in \boldsymbol{V}_h : b(\boldsymbol{v}_h, q_h) = 0 \text{ for all } q_h \in \mathcal{Q}_h\big\},$$

however we cannot obtain a characterisation analogous to the discrete case.

5.4.2 Discrete stability properties

For the subsequent analysis, we introduce for $r \geq 0$ the broken $\boldsymbol{H}_\alpha^r(\mathcal{T}_h)$ space

$$\boldsymbol{H}_\alpha^r(\mathcal{T}_h) = \big\{\boldsymbol{v} \in \boldsymbol{L}_\alpha^2 : \boldsymbol{v}|_K \in \boldsymbol{H}_\alpha^r(K), K \in \mathcal{T}_h\big\},$$

as well as the following parameter- and mesh- dependent broken norms

$$\begin{aligned}
\|\boldsymbol{v}\|_{*,\mathcal{T}_h}^2 &:= \sum_{K\in\mathcal{T}_h} \|\boldsymbol{\varepsilon}_h(\boldsymbol{v})\|_{\mathbb{L}_1^2(K)}^2 + \sum_{K\in\mathcal{T}_h} \|v_r\|_{L_{-1}^2(K)} + \sum_{e\in\mathcal{E}_h} \frac{1}{h_e}\|[\![\boldsymbol{v}_\boldsymbol{\tau}]\!]\|_{\boldsymbol{L}_1^2(e)}^2, \\
\|\boldsymbol{v}\|_{1,\mathcal{T}_h}^2 &:= \|\boldsymbol{v}\|_{\boldsymbol{L}_1^2(\Omega_\mathrm{a})}^2 + \nu\|\boldsymbol{v}\|_{*,\mathcal{T}_h}^2 \quad \text{for all } \boldsymbol{v} \in \boldsymbol{H}_1^1(\mathcal{T}_h), \\
\|\boldsymbol{v}\|_{\mathcal{T}_h^2}^2 &:= \|\boldsymbol{v}\|_{1,\mathcal{T}_h}^2 + \sum_{K\in\mathcal{T}_h} h_K^2|\boldsymbol{v}|_{H_1^2(K)}^2 \quad \text{for all } \boldsymbol{v} \in \boldsymbol{H}_1^2(\mathcal{T}_h),
\end{aligned}$$

where the stronger norm $\|\cdot\|_{\mathcal{T}_h^2}^2$ is used to show continuity. It can be proven that this norm is equivalent to $\|\cdot\|_{1,\mathcal{T}_h}$ on $\boldsymbol{H}_1^1(\mathcal{T}_h)$ (see [67] and [19]). Finally, adapting the argument used in [102, Proposition 4.5] and relying on the equivalent weighted Sobolev embeddings in [111] we have the following discrete Sobolev embedding: for $r = 2, 4$ there exists a constant $C_\mathrm{emb} > 0$ such that

$$\|\boldsymbol{v}\|_{\boldsymbol{L}_1^r} \leq C_\mathrm{emb}\|\boldsymbol{v}\|_{1,\mathcal{T}_h} \qquad \text{for all } \boldsymbol{v} \in \boldsymbol{H}_1^1(\mathcal{T}_h). \tag{5.26}$$

Using these norms, we can establish continuity of the trilinear and bilinear forms involved, stated in the following lemma that can be proved following [137, Section 3.3.2], [71, Section 3] and [19, Section 4].

Lemma 5.3. *The following properties hold:*

$$\begin{aligned}
\left|a_1^h(\boldsymbol{u},\boldsymbol{v})\right| &\le C\|\boldsymbol{u}\|_{\mathcal{T}_h^2}\|\boldsymbol{v}\|_{1,\mathcal{T}_h} && \text{for all } \boldsymbol{u}\in\boldsymbol{H}_1^2(\mathcal{T}_h),\ \boldsymbol{v}\in\boldsymbol{V}_h,\\
\left|a_1^h(\boldsymbol{u},\boldsymbol{v})\right| &\le \tilde{C}_a\|\boldsymbol{u}\|_{1,\mathcal{T}_h}\|\boldsymbol{v}\|_{1,\mathcal{T}_h} && \text{for all } \boldsymbol{u},\boldsymbol{v}\in\boldsymbol{V}_h,\\
\left|b(\boldsymbol{v},q)\right| &\le \tilde{C}_b\|\boldsymbol{v}\|_{1,\mathcal{T}_h}\|q\|_{L_1^2(\Omega_{\mathrm{a}})} && \text{for all } \boldsymbol{v}\in\boldsymbol{H}_1^1(\mathcal{T}_h),\ q\in L_1^2(\Omega),
\end{aligned}$$

and for all $\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}\in\boldsymbol{H}_1^1(\mathcal{T}_h)$ *and* $\vec{\psi},\vec{\theta}\in[H_1^1(\Omega)]^m$*, there holds*

$$\left|d_1(\vec{\theta},\boldsymbol{v})\right| \le C_{\boldsymbol{F}}\|\vec{\theta}\|_{\vec{H}_1^1}\|\boldsymbol{v}\|_{1,\mathcal{T}_h}, \tag{5.27a}$$

$$\left|c_2^h(\boldsymbol{w};\vec{\theta},\vec{\psi})\right| \le \tilde{C}\|\boldsymbol{w}\|_{1,\mathcal{T}_h}\|\vec{\psi}\|_{\vec{H}_1^1}\|\vec{\theta}\|_{\vec{H}_1^1}. \tag{5.27b}$$

Note that while the coercivity of the form $a_2(\cdot,\cdot)$ in the discrete setting is readily implied by (5.17), there also holds (cf. [109, Lemma 3.2])

$$a_1^h(\boldsymbol{v},\boldsymbol{v}) \ge \tilde{\alpha}_a\|\boldsymbol{v}\|_{1,\mathcal{T}_h}^2 \quad \text{for all } \boldsymbol{v}\in\boldsymbol{V}_h, \tag{5.28}$$

provided that $a_0>0$ is sufficiently large and independent of the mesh size.

Let $\boldsymbol{w}\in\boldsymbol{H}_0(\mathrm{div}^0;\Omega)$, due to the skew-symmetric form of the operators c_1^h and c_2^h, and the positivity of the non-linear upwind term of c_1^h (see e.g. [138]), we can write

$$c_1^h(\boldsymbol{w};\boldsymbol{u},\boldsymbol{u}) \ge 0 \quad \text{for all } \boldsymbol{u}\in\boldsymbol{V}_h, \tag{5.29}$$

$$c_2^h(\boldsymbol{w};\vec{\psi}_h,\vec{\psi}_h) = 0 \quad \text{for all } \vec{\psi}_h\in\mathcal{M}_h, \tag{5.30}$$

as well as the following relation (which is based on (5.26) and follows by the same method as in [62, 102]):

$$\begin{aligned}
&\text{For any } \boldsymbol{w}_1,\boldsymbol{w}_2,\boldsymbol{u}\in\boldsymbol{H}_1^2(\mathcal{T}_h) \text{ there holds for all } \boldsymbol{v}\in\boldsymbol{V}_h\\
&\left|c_1^h(\boldsymbol{w}_1;\boldsymbol{u},\boldsymbol{v})\right| - \left|c_1^h(\boldsymbol{w}_2;\boldsymbol{u},\boldsymbol{v})\right| \le \tilde{C}_c\|\boldsymbol{w}_1-\boldsymbol{w}_2\|_{1,\mathcal{T}_h}\|\boldsymbol{v}\|_{1,\mathcal{T}_h}\|\boldsymbol{u}\|_{1,\mathcal{T}_h}.
\end{aligned} \tag{5.31}$$

5.4.3 Existence of discrete solutions

In what follows we will use the following algebraic relation: for any real numbers a^{n+1}, a^n, a^{n-1} and defining $\Lambda a^n := a^{n+1}-2a^n+a^{n-1}$, we have

$$\begin{aligned}
2(3a^{n+1}-4a^n+a^{n-1},a^n) = |a^{n+1}|^2 + |2a^{n+1}-a^n|^2 + |\Lambda a^n|^2\\
- |a^n|^2 - |2a^n-a^{n-1}|^2.
\end{aligned} \tag{5.32}$$

Theorem 5.2. *Let* $(\boldsymbol{u}_h^{n+1},p_h^{n+1},\vec{\theta}_h^{n+1},\vec{s}_h^{n+1})\in\boldsymbol{V}_h\times\mathcal{Q}_h\times\vec{\mathcal{M}}_{h,0}\times\mathcal{S}_h$ *be a solution of problem* (5.25)*.*

Then the following bounds are satisfied, where C_1, C_2 and C_3 are constants independent of h and Δt:

$$
\begin{aligned}
&\|\boldsymbol{u}_h^{n+1}\|^2_{\boldsymbol{L}_1^2} + \|2\boldsymbol{u}_h^{n+1} - \boldsymbol{u}_h^n\|^2_{\boldsymbol{L}_1^2} + \sum_{j=1}^n \|\Lambda \boldsymbol{u}_h^j\|^2_{\boldsymbol{L}_1^2} + \sum_{j=1}^n \Delta t \|\boldsymbol{u}_h^{j+1}\|^2_{1,\mathcal{T}_h} \\
&\le C_1\big(\|\vec{\theta}_h^1\|^2_{\vec{L}_1^2} + \|2\vec{\theta}_h^1 - \vec{\theta}_h^0\|^2_{\vec{L}_1^2} + \|\boldsymbol{u}_h^1\|^2_{\boldsymbol{L}_1^2} + \|2\boldsymbol{u}_h^1 - \boldsymbol{u}_h^0\|^2_{\boldsymbol{L}_1^2}\big), \\
&\|\vec{\theta}_h^{n+1}\|^2_{\vec{L}_1^2} + \|2\vec{\theta}_h^{n+1} - \vec{\theta}_h^n\|^2_{\vec{L}_1^2} + \sum_{j=1}^n \|\Lambda \vec{\theta}_h^j\|^2_{\vec{L}_1^2} + \sum_{j=1}^n \Delta t |\vec{\theta}_h^{j+1}|^2_{\vec{L}_1^2} \\
&\le C_2\big(\|\vec{\theta}_h^1\|^2_{\vec{L}_1^2} + \|2\vec{\theta}_h^1 - \vec{\theta}_h^0\|^2_{\vec{L}_1^2}\big), \\
&\|\vec{s}_h^{n+1}\|^2_{\vec{L}_1^2} + \|2\vec{s}_h^{n+1} - \vec{s}_h^n\|^2_{\vec{L}_1^2} + \sum_{j=1}^n \|\Lambda \vec{s}_h^j\|^2_{\vec{L}_1^2} \\
&\le C_3\big(\|\vec{\theta}_h^1\|^2_{\vec{L}_1^2} + \|2\vec{\theta}_h^1 - \vec{\theta}_h^0\|^2_{\vec{L}_1^2} + \|\vec{s}_h^1\|^2_{\vec{L}_1^2} + \|2\vec{s}_h^1 - \vec{s}_h^0\|^2_{\vec{L}_1^2} + n\Delta t C_d^2\big).
\end{aligned}
\tag{5.33}
$$

Proof. First we take $\vec{\psi}_h = 4\vec{\theta}_h^{n+1}$ in the third equation of (5.25) and use properties (5.7), (5.30) and relation (5.32) to deduce the inequality

$$
\begin{aligned}
&\|\vec{\theta}_h^{n+1}\|^2_{\vec{L}_1^2} + \|2\vec{\theta}_h^{n+1} - \vec{\theta}_h^n\|^2_{\vec{L}_1^2} + \|\Lambda\vec{\theta}_h^n\|^2_{\vec{L}_1^2} + 4\alpha_2 \Delta t |\vec{\theta}_h^{n+1}|^2_{H_1^1} \\
&\le \|\vec{\theta}_h^n\|^2_{\vec{L}_1^2} + \|2\vec{\theta}_h^n - \vec{\theta}_h^{n-1}\|^2_{\vec{L}_1^2}.
\end{aligned}
$$

Hence, summing over n, we get

$$
\begin{aligned}
&\|\vec{\theta}_h^{n+1}\|^2_{\vec{L}_1^2} + \|2\vec{\theta}_h^{n+1} - \vec{\theta}_h^n\|^2_{\vec{L}_1^2} + \sum_{j=1}^n \|\Lambda\vec{\theta}_h^j\|^2_{\vec{L}_1^2} + 4\alpha_2 \sum_{j=1}^n \Delta t |\vec{\theta}_h^{j+1}|^2_{H_1^1} \\
&\le \|\vec{\theta}_h^1\|^2_{\vec{L}_1^2} + \|2\vec{\theta}_h^1 - \vec{\theta}_h^0\|^2_{\vec{L}_1^2}.
\end{aligned}
\tag{5.34}
$$

Similarly, in the fourth equation of (5.25), we take $\vec{l}_h = 4\vec{s}_h^{n+1}$ and apply (5.12), (5.10) together with Young's inequality to get

$$
\begin{aligned}
&\|\vec{s}_h^{n+1}\|^2_{\vec{L}_1^2} + \|2\vec{s}_h^{n+1} - \vec{s}_h^n\|^2_{\vec{L}_1^2} + \|\Lambda\vec{s}_h^n\|^2_{\vec{L}_1^2} \\
&\le 4\Delta t C_d \|\vec{\theta}_h^{n+1}\|_{\vec{L}_1^2} + \|\vec{s}_h^n\|^2_{\vec{L}_1^2} + \|2\vec{s}_h^n - \vec{s}_h^{n-1}\|^2_{\vec{L}_1^2} \\
&\le 2\Delta t C_p |\vec{\theta}_h^{n+1}|^2_{\vec{H}_1^1} + 2C_d^2 \Delta t + \|\vec{s}_h^n\|^2_{\vec{L}_1^2} + \|2\vec{s}_h^n - \vec{s}_h^{n-1}\|^2_{\vec{L}_1^2}.
\end{aligned}
$$

Summing over n we therefore obtain

$$
\begin{aligned}
&\|\vec{s}_h^{n+1}\|^2_{\vec{L}_1^2} + \|2\vec{s}_h^{n+1} - \vec{s}_h^n\|^2_{\vec{L}_1^2} + \sum_{j=1}^n \|\Lambda\vec{s}_h^j\|^2_{\vec{L}_1^2} \\
&\le 2C_p \sum_{j=1}^n \Delta t |\vec{\theta}_h^{j+1}|^2_{\vec{H}_1^1} + 2n\Delta t C_d^2 + \|\vec{s}_h^1\|^2_{\vec{L}_1^2} + \|2\vec{s}_h^1 - \vec{s}_h^0\|^2_{\vec{L}_1^2}.
\end{aligned}
\tag{5.35}
$$

We get the second result of (5.33) by replacing (5.34) in (5.35). Finally we take $\boldsymbol{v}_h = 4\boldsymbol{u}_h^{n+1}$ and $q_h = 4p_h^{n+1}$ in the first and second equation of (5.25), respectively and apply (5.32), (5.27a), (5.28)

and (5.29) to deduce the estimate

$$\begin{aligned}&\|\boldsymbol{u}_h^{n+1}\|_{\boldsymbol{L}_1^2}^2+\|2\boldsymbol{u}_h^{n+1}-\boldsymbol{u}_h^n\|_{\boldsymbol{L}_1^2}^2+\|\Lambda\boldsymbol{u}_h^n\|_{\boldsymbol{L}_1^2}^2+4\Delta t\tilde{\alpha}_a\|\boldsymbol{u}_h^{n+1}\|_{1,\mathcal{T}_h}^2\\&\leq 4\Delta tC_{\boldsymbol{F}}\|\vec{\theta}_h^{n+1}\|_{\vec{L}_1^2}\|\boldsymbol{u}_h^{n+1}\|_{\boldsymbol{L}_1^2}+\|\boldsymbol{u}_h^n\|_{\boldsymbol{L}_1^2}^2+\|2\boldsymbol{u}_h^n-\boldsymbol{u}_h^{n-1}\|_{\boldsymbol{L}_1^2}^2.\end{aligned}$$

Now we use Young's inequality with $\varepsilon=\tilde{\alpha}_a$ to arrive at

$$\begin{aligned}&\|\boldsymbol{u}_h^{n+1}\|_{\boldsymbol{L}_1^2}^2+\|2\boldsymbol{u}_h^{n+1}-\boldsymbol{u}_h^n\|_{\boldsymbol{L}_1^2}^2+\|\Lambda\boldsymbol{u}_h^n\|_{\boldsymbol{L}_1^2}^2+\Delta t2\tilde{\alpha}_a\|\boldsymbol{u}_h^{n+1}\|_{1,\mathcal{T}_h}^2\\&\leq 2\frac{C_{\boldsymbol{F}}^2C_p}{\tilde{\alpha}_a}\Delta t|\vec{\theta}_h^{n+1}|_{\vec{H}_1^1}^2+\|\boldsymbol{u}_h^n\|_{\boldsymbol{L}_1^2}^2+\|2\boldsymbol{u}_h^n-\boldsymbol{u}_h^{n-1}\|_{\boldsymbol{L}_1^2}^2,\end{aligned}$$

and summing over n we can assert that

$$\begin{aligned}&\|\boldsymbol{u}_h^{n+1}\|_{\boldsymbol{L}_1^2}^2+\|2\boldsymbol{u}_h^{n+1}-\boldsymbol{u}_h^n\|_{\boldsymbol{L}_1^2}^2+\sum_{j=1}^n\|\Lambda\boldsymbol{u}_h^j\|_{\boldsymbol{L}_1^2}^2+2\tilde{\alpha}_a\sum_{j=1}^n\Delta t\|\boldsymbol{u}_h^{j+1}\|_{1,\mathcal{T}_h}^2\\&\leq\frac{C_{\boldsymbol{F}}^2C_p}{2}\sum_{j=1}^n\Delta t|\vec{\theta}_h^{j+1}|_{\vec{H}_1^1}^2+\|\boldsymbol{u}_h^1\|_{\boldsymbol{L}_1^2}^2+\|2\boldsymbol{u}_h^1-\boldsymbol{u}_h^0\|_{\boldsymbol{L}_1^2}^2.\end{aligned}\tag{5.36}$$

Finally we get the first result in (5.33) from the bounds (5.34) and (5.36). □

Theorem 5.3. *Assume that*

$$\frac{C_{\boldsymbol{F}}}{\tilde{\alpha}_a}\leq\frac{\alpha_2}{C_p}.\tag{5.37}$$

Then problem (5.25) *admits at least one solution*

$$(\boldsymbol{u}_h^{n+1},p_h^{n+1},\vec{\theta}_h^{n+1},\vec{s}_h^{n+1})\in\boldsymbol{V}_h\times\mathcal{Q}_h\times\vec{\mathcal{M}}_{h,0}\times\vec{\mathcal{S}}_h.$$

Proof. To simplify the proof we introduce the following constants:

$$\begin{aligned}C_u&:=C_1\big(\|\vec{\theta}_h^1\|_{\vec{L}_1^2}+\|2\vec{\theta}_h^1-\vec{\theta}_h^0\|_{\vec{L}_1^2}+\|\boldsymbol{u}_h^1\|_{\boldsymbol{L}_1^2}+\|2\boldsymbol{u}_h^1-\boldsymbol{u}_h^0\|_{\boldsymbol{L}_1^2}\big),\\C_\theta&:=C_2\big(\|\vec{\theta}_h^1\|_{\vec{L}_1^2}+\|2\vec{\theta}_h^1-\vec{\theta}_h^0\|_{\vec{L}_1^2}\big),\\C_s&:=C_3\big(\|\vec{\theta}_h^1\|_{\vec{L}_1^2}+\|2\vec{\theta}_h^1-\vec{\theta}_h^0\|_{\vec{L}_1^2}+\|\vec{s}_h^1\|_{\vec{L}_1^2}+\|2\vec{s}_h^1-\vec{s}_h^0\|_{\vec{L}_1^2}+n\Delta tC_d^2\big).\end{aligned}$$

We shall make use of Brouwer's fixed-point theorem in the form given by [82, Corollary 1.1, Chapter IV]:

Theorem 5.4 (Brouwer's fixed-point theorem)**.** *Let H be a finite-dimensional Hilbert space with scalar product denoted by $(,)_H$ and corresponding norm $\|\cdot\|_H$. Let $\Phi\colon H\to H$ be a continuous mapping for which there exists $\mu>0$ such that $(\Phi(u),u)_H\geq 0$ for all $u\in H$ with $\|u\|_H=\mu$. Then there exists an element $u\in H$ such that $\Phi(u)=0,\|u\|_H\leq\mu$.*

We proceed by induction on $n\geq 2$. We define the mapping

$$\Phi:\boldsymbol{V}_h\times\mathcal{Q}_h\times\vec{\mathcal{M}}_{h,0}\times\vec{\mathcal{S}}_h\to\boldsymbol{V}_h\times\mathcal{Q}_h\times\vec{\mathcal{M}}_{h,0}\times\vec{\mathcal{S}}_h$$

using the relation

$$
\begin{aligned}
&\big(\Phi(\boldsymbol{u}_h^{n+1}, p_h^{n+1}, \vec{\theta}_h^{n+1}, \vec{s}_h^{n+1}), (\boldsymbol{v}_h, q_h, \vec{\psi}_h, \vec{l}_h)\big)_{1,\Omega_\mathrm{a}} \\
&= \frac{1}{2\Delta t}\big(3\boldsymbol{u}_h^{n+1} - 4\boldsymbol{u}_h^{n} + \boldsymbol{u}_h^{n-1}, \boldsymbol{v}_h\big)_{1,\Omega_\mathrm{a}} + a_1^h\big(\boldsymbol{u}_h^{n+1}, \boldsymbol{v}_h\big) + c_1^h\big(\boldsymbol{u}_h^{n+1}; \boldsymbol{u}_h^{n+1}, \boldsymbol{v}_h\big) \\
&\quad + b(\boldsymbol{v}_h, p_h^{n+1}) - b(\boldsymbol{u}_h^{n+1}, q_h) - \big(\boldsymbol{F}\big(\vec{\theta}_h^{n+1}\big), \boldsymbol{v}_h\big)_{1,\Omega_\mathrm{a}} + \frac{1}{2\Delta t}\big(3\vec{\theta}_h^{n+1} - 4\vec{\theta}_h^{n} + \vec{\theta}_h^{n-1}, \vec{\psi}_h\big)_{1,\Omega_\mathrm{a}} \\
&\quad + a_2\big(\vec{\theta}_h^{n+1}, \vec{\psi}_h\big) + c_2^h\big(\boldsymbol{u}_h^{n+1}; \vec{\theta}_h^{n+1}, \vec{\psi}_h\big) + d_2\big(\vec{s}_h^{n+1}; \vec{\theta}_h^{n+1}, \vec{\psi}_h\big) \\
&\quad + \frac{1}{2\Delta t}\big(3\vec{s}_h^{n+1} - 4\vec{s}_h^{n} + \vec{s}_h^{n-1}, \vec{l}_h\big)_{1,\Omega_\mathrm{a}} + d_3\big(\vec{\theta}_h^{n+1}; \vec{s}_h^{n+1}, \vec{l}_h\big) - d_4\big(\vec{\theta}_h^{n+1}, \vec{l}_h\big).
\end{aligned}
$$

Note this map is well-defined and continuous on $\boldsymbol{V}_h \times \mathcal{Q}_h \times \vec{\mathcal{M}}_{h,0} \times \vec{\mathcal{S}}_h$. On the other hand, if we take

$$
(\boldsymbol{v}_h, q_h, \vec{\psi}_h, \vec{l}_h) = (\boldsymbol{u}_h^{n+1}, p_h^{n+1}, \vec{\theta}_h^{n+1}, \vec{s}_h^{n+1}),
$$

and employ (5.29), (5.30), (5.27a) and (5.28), we obtain

$$
\begin{aligned}
&\big(\Phi(\boldsymbol{u}_h^{n+1}, p_h^{n+1}, \vec{\theta}_h^{n+1}, \vec{s}_h^{n+1}), (\boldsymbol{u}_h^{n+1}, p_h^{n+1}, \vec{\theta}_h^{n+1}, \vec{s}_h^{n+1})\big)_{1,\Omega_\mathrm{a}} \\
&\geq -\frac{1}{2\Delta t}\|4\boldsymbol{u}_h^{n} - \boldsymbol{u}_h^{n-1}\|_{\boldsymbol{L}_1^2}\|\boldsymbol{u}_h^{n+1}\|_{\boldsymbol{L}_1^2} + \tilde{\alpha}_a\|\boldsymbol{u}_h^{n+1}\|_{1,\mathcal{T}_h}^2 - C_{\boldsymbol{F}}\|\theta_h^{n+1}\|_{\vec{L}_1^2}\|\boldsymbol{u}_h^{n+1}\|_{\boldsymbol{L}_1^2} \\
&\quad - \frac{1}{2\Delta t}\|4\vec{\theta}_h^{n} - \vec{\theta}_h^{n-1}\|_{\vec{L}_1^2}\|\vec{\theta}_h^{n+1}\|_{\vec{L}_1^2} + \alpha_2|\vec{\theta}^{n+1}|_{\vec{H}_1^1}^2 + \frac{3}{2\Delta t}\|\vec{s}_h^{n+1}\|_{L_2^1}^2 \\
&\quad - \frac{1}{2\Delta t}\|4\vec{s}_h^{n} - \vec{s}_h^{n-1}\|_{\vec{L}_1^2}\|\vec{s}_h^{n+1}\|_{\vec{L}_1^2} - C_d\|\vec{\theta}_h^{n+1}\|_{\vec{L}_1^2}.
\end{aligned}
$$

Next, using (5.33), inequality (5.37) and Young's inequality with constant $\varepsilon_1 = \tilde{\alpha}_a/C_{\boldsymbol{F}}$, we deduce that

$$
\begin{aligned}
&\big(\Phi(\boldsymbol{u}_h^{n+1}, p_h^{n+1}, \vec{\theta}_h^{n+1}, \vec{s}_h^{n+1}), (\boldsymbol{u}_h^{n+1}, p_h^{n+1}, \vec{\theta}_h^{n+1}, \vec{s}_h^{n+1})\big)_{1,\Omega_\mathrm{a}} \\
&\geq \frac{\tilde{\alpha}_a}{2}\|\boldsymbol{u}_h^{n+1}\|_{\boldsymbol{L}_1^2}^2 + \frac{3}{2\Delta t}\|\vec{s}_h^{n+1}\|_{\vec{L}_1^2}^2 + \frac{\alpha_2}{2Cp}\|\vec{\theta}_h^{n+1}\|_{\vec{L}_1^2}^2 - \frac{5}{2\Delta t}C_u\|\boldsymbol{u}_h^{n+1}\|_{\boldsymbol{L}_1^2(\Omega_\mathrm{a})} \\
&\quad - \left(\frac{5}{2\Delta t}C_\theta + C_d\right)\|\vec{\theta}_h^{n+1}\|_{\vec{L}_1^2} - \frac{5}{2\Delta t}C_s\|\vec{s}_h^{n+1}\|_{\vec{L}_1^2}.
\end{aligned}
$$

Then, setting

$$
C_R = \min\left\{\frac{\tilde{\alpha}_a}{2}, \frac{3}{2\Delta t}, \frac{\alpha_2}{2C_p}\right\}, \quad C_r = 2\max\left\{\frac{5}{2\Delta t}C_u, \frac{5}{2\Delta t}C_\theta + C_d, \frac{5}{2\Delta t}C_s\right\},
$$

we may apply the inequality $a + b \leq \sqrt{2}(a^2 + b^2)^{1/2}$, valid for all $a, b \in \mathbb{R}$, to obtain

$$
\begin{aligned}
&\big(\Phi(\boldsymbol{u}_h^{n+1}, p_h^{n+1}, \vec{\theta}_h^{n+1}, \vec{s}_h^{n+1}), (\boldsymbol{u}_h^{n+1}, p_h^{n+1}, \vec{\theta}_h^{n+1}, \vec{s}_h^{n+1})\big)_{1,\Omega_\mathrm{a}} \\
&\geq C_R\big(\|\boldsymbol{u}_h^{n+1}\|_{\boldsymbol{L}_1^2}^2 + \|\vec{\theta}_h^{n+1}\|_{\vec{L}_1^2}^2 + \|\vec{s}_h^{n+1}\|_{\vec{L}_1^2}^2\big) \\
&\quad - C_r\big(\|\boldsymbol{u}_h^{n+1}\|_{\boldsymbol{L}_1^2}^2 + \|\vec{\theta}_h^{n+1}\|_{\vec{L}_1^2}^2 + \|\vec{s}_h^{n+1}\|_{\vec{L}_1^2}^2\big)^{1/2}.
\end{aligned}
$$

Hence, the right-hand side is nonnegative on a sphere of radius $r := C_r/C_R$. Consequently, by Theorem 5.4, there exists a solution to the fixed-point problem

$$
\Phi\big(\boldsymbol{u}_h^{n+1}, p_h^{n+1}, \vec{\theta}_h^{n+1}, \vec{s}_h^{n+1}\big) = 0.
$$

□

As in the previous chapters, one cannot directly establish a discrete version of Theorem 5.1. In fact we were not able to control the discrete norms arising from (5.31), which would be necessary to establish a discrete counterpart of (5.21). However, even when uniqueness of the discrete counterpart remains an open problem, our non-exhaustive selection of numerical examples did not present any difficulties in this regard.

5.5 Error analysis

The following development follows the structure adopted in [4]. We start by recalling some interpolation results from [26] and [71].

Lemma 5.4. *Let $\mathcal{L}_h$ be the Lagrange interpolation operator $\mathcal{L}_h : C^0(\Omega_\mathrm{a}) \to V_h$, where V_h denotes the space of Lagrange finite elements of order k. We also consider its vectorial counterpart, keeping the same notation. Then for all l and for all p such that $1 \le l \le k+1$, $1 \le p \le +\infty$, $l > \frac{3}{p}$ or $p = 1, l = 3$ there exists a constant $C^* > 0$ independent of h, such that for all $v \in W_1^{l,p}(\Omega_\mathrm{a})$, the following inequalities hold;*

$$\|v - \mathcal{L}_h\, v\|_{L_1^p(\Omega_\mathrm{a})} \le C^* h^l |v|_{W_1^{l,p}(\Omega_\mathrm{a})}, \qquad |v - \mathcal{L}_h\, v|_{L_1^p(\Omega_\mathrm{a})} \le C^* h^{l-1} |v|_{W_1^{l,p}(\Omega_\mathrm{a})}.$$

Lemma 5.5. *Let Π_h be the $\mathrm{BDM}_k^{\mathrm{axi}}$ interpolation operator $\Pi_h : C^0(\Omega_\mathrm{a}) \to \boldsymbol{\mathcal{H}}_h^k$. Then for all $v \in H_1^{k+1}(\Omega_\mathrm{a})$, the following inequalities hold:*

$$\|v - \Pi_h\, v\|_{L_1^2(\Omega_\mathrm{a})} \le C^* h^{k+1} |v|_{H_1^{k+1}(\Omega_\mathrm{a})}, \qquad \|v - \Pi_h\, v\|_{1,\mathcal{T}_h} \le C^* h^k \|v\|_{H_1^{k+1}(\Omega_\mathrm{a})}.$$

Proof. The first result comes from [71, Corrollary A.7]. The proof of the second result comes much in the same way as in the Cartesian case, by making use of the equivalent weighted inverse inequalities and weighted approximation properties proved in [26], see [72, Section 3.1] and [19]. □

Lemma 5.6. *Let $\mathcal{I}_h$ denote the modified Clément interpolation operator*

$$\mathcal{I}_h : H_{0,1}^1(\Omega_\mathrm{a}) \to \mathcal{M}_h^k,$$

and the same notation is kept for its vectorial counterpart. Then for all l and for all p such that $1 \le l \le k+1$, $1 \le p \le +\infty$ there exists a constant $C^* > 0$ independently of h such that for any function $v \in W_1^{l,p}(\Omega_\mathrm{a})$,

$$\|v - \mathcal{I}_h\, v\|_{L_1^p(\Omega_\mathrm{a})} \le C^* h^l |v|_{W_1^{l,p}(\Omega_\mathrm{a})}.$$

Lemma 5.7. *Assume that $\boldsymbol{u} \in \boldsymbol{H}_1^2$ and $\vec{\theta} \in \vec{H}_1^1$. Then*

$$\begin{aligned}\big(\partial_t \boldsymbol{u}(t), \boldsymbol{v}\big)_{1,\Omega_\mathrm{a}} + a_1^h\big(\boldsymbol{u}(t), \boldsymbol{v}\big) + c_1^h\big(\boldsymbol{u}(t); \boldsymbol{u}(t), \boldsymbol{v}\big) + b(\boldsymbol{v}, p) - d_1\big(\vec{\theta}(t), \boldsymbol{v}\big) &= 0,\\ \big(\partial_t \vec{\theta}(t), \vec{\psi}\big)_{1,\Omega_\mathrm{a}} + a_2\big(\vec{\theta}(t), \vec{\psi}\big) + c_2^h\big(\vec{u}(t); \vec{\theta}(t), \vec{\psi}\big) + d_2\big(\vec{s}(t); \vec{\theta}(t), \vec{\psi}\big) &= 0\end{aligned}$$

for all $(\boldsymbol{v}, \vec{\psi}) \in \boldsymbol{V}_h \times \mathcal{M}_{h,0}$. A similar result also holds for the fourth equation in (5.24).

Proof. Since we assume $\boldsymbol{u} \in \boldsymbol{H}_1^2(\Omega_\mathrm{a})$, integration by parts yields the required result. See also [19]. □

Now we decompose the errors as follows:

$$\begin{aligned}
\boldsymbol{u}-\boldsymbol{u}_h &= E_{\boldsymbol{u}}+\xi_{\boldsymbol{u}} = (\boldsymbol{u}-\Pi_h\,\boldsymbol{u})+(\Pi_h\,\boldsymbol{u}-\boldsymbol{u}_h),\\
p-p_h &= E_p+\xi_p = (p-\mathcal{L}_h\,p)+(\mathcal{L}_h\,p-p_h),\\
\vec{\theta}-\vec{\theta}_h &= E_{\vec{\theta}}+\xi_{\vec{\theta}} = (\vec{\theta}-\mathcal{I}_h\,\vec{\theta})+(\mathcal{I}_h\,\vec{\theta}-\vec{\theta}_h),\\
\vec{s}-\vec{s}_h &= E_{\vec{s}}+\xi_{\vec{s}} = (\vec{s}-\mathcal{L}_h\,\vec{s})+(\mathcal{L}_h\,\vec{s}-\vec{s}_h).
\end{aligned}$$

Assuming that $\boldsymbol{u}_h^0=\Pi_h\,\boldsymbol{u}(0)$, $\vec{\theta}_h^0=\mathcal{I}_h\,\vec{\theta}(0)$ and $\vec{s}_h^0=\mathcal{L}_h\,\vec{s}(0)$, we will use also the notation $E_{\boldsymbol{u}}^n=\boldsymbol{u}(t_n)-\Pi_h\,\boldsymbol{u}(t_n)$ and $\xi_{\boldsymbol{u}}^n=\Pi_h\,\boldsymbol{u}(t_n)-\boldsymbol{u}_h^n$, and the corresponding notation for other variables. Since for the first time iteration of system (5.25) we adopt a backward Euler scheme, we require error estimates for this step.

Theorem 5.5. *Let us assume that*

$$\begin{aligned}
&\boldsymbol{u}\in L^\infty(0,\mathcal{T};H_1^3)\cap L^\infty(0,\mathcal{T};\boldsymbol{V}_{1,\diamond}^1(\Omega_{\rm a})), \quad \boldsymbol{u}'\in L^\infty(0,\mathcal{T};\boldsymbol{H}_1^1),\\
&\boldsymbol{u}''\in L^\infty(0,\mathcal{T};\boldsymbol{L}_1^2), \quad p\in L^\infty(0,\mathcal{T};H_1^2), \quad \vec{\theta}\in L^\infty(0,\mathcal{T};\vec{H}_{1,\diamond}^3(\Omega_{\rm a})),\\
&\vec{\theta}'\in L^\infty(0,\mathcal{T};\vec{H}_1^2), \quad \vec{\theta}''\in L^\infty(0,\mathcal{T};\vec{L}_1^2), \quad \vec{s}\in L^\infty(0,\mathcal{T};\vec{H}_1^3),\\
&\vec{s}'\in L^\infty(0,\mathcal{T};\vec{H}_1^2), \quad \vec{s}''\in L^\infty(0,\mathcal{T};\vec{H}_1^1),
\end{aligned}$$

and also that $\|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T};\boldsymbol{H}_1^1)}<M$ *for a sufficiently small constant* $M>0$ *(a precise condition for* M*, can be found in Theorem 5.6). Then there exist positive constants* C_u^1*,* C_θ^1*,* C_s^1*, independently of* h *and* Δt*, such that*

$$\begin{aligned}
&\|\xi_{\boldsymbol{u}}^1\|_{\boldsymbol{L}_1^2}^2+\frac{1}{2}\Delta t\tilde{\alpha}_a\|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h}^2\le C_u^1(h^{2k}+\Delta t^4),\\
&\frac{1}{4}\|\xi_{\vec{\theta}}^1\|_{\vec{L}_1^2}^2+\frac{1}{2}\Delta t\hat{\alpha}_a\|\xi_{\vec{\theta}}\|_{\vec{H}_1^1}^2\le C_\theta^1(h^{2k}+\Delta t^4),\\
&\frac{1}{2}\|\xi_{\vec{s}}^1\|_{\vec{L}_1^2}^2+\frac{1}{2}\Delta t g_1\|\xi_{\vec{s}}^1\|_{\vec{L}_1^2}^2\le C_s^1(h^{2k}+\Delta t^4).
\end{aligned}$$

Proof. Since these bounds are similar to those used in Theorems 5.6–5.8, we postpone some details until the proof of those theorems. First, based on the regularity assumptions for $\boldsymbol{u}$, for all $\boldsymbol{x}$ there exists $\gamma\in(0,1)$ such that

$$\boldsymbol{u}(0)=\boldsymbol{u}(\Delta t)-\Delta t\boldsymbol{u}'(\Delta t)+\frac{1}{2}\Delta t^2\boldsymbol{u}''(\Delta t\gamma),$$

where $\boldsymbol{u}$ satisfies the error inequality

$$\begin{aligned}
&\|\xi_{\boldsymbol{u}}^1\|_{\boldsymbol{L}_1^2}^2+\Delta t\tilde{\alpha}_a\|\xi_{\boldsymbol{u}}^1\|_{1,\mathcal{T}_h}^2\\
&\le -\big(\Pi_h\,\boldsymbol{u}(\Delta t)-\boldsymbol{u}(\Delta t)-(\boldsymbol{u}_h^0-\boldsymbol{u}(0)),\xi_{\boldsymbol{u}}^1\big)_{1,\Omega_{\rm a}}+\Delta t b\big(\mathcal{L}_h\,p(\Delta t)-p(\Delta t),\xi_{\boldsymbol{u}}^1\big)\\
&\quad+\Delta t a_1^h\big(\Pi_h\,\boldsymbol{u}(\Delta t),\xi_{\boldsymbol{u}}^1\big)-\Delta t\big(c_1^h(\boldsymbol{u}_h^1;\boldsymbol{u}_h^1,\xi_{\boldsymbol{u}}^1)-c_1^h(\boldsymbol{u}(\Delta t),\boldsymbol{u}(\Delta t),\xi_{\boldsymbol{u}}^1)\big)\\
&\quad-\Delta t d_1\big(\vec{\theta}_h^1-\vec{\theta}(\Delta t),\xi_{\boldsymbol{u}}^1\big)-\frac{\Delta t^2}{2}\big(\boldsymbol{u}''(\Delta t\gamma),\xi_{\boldsymbol{u}}^1\big),
\end{aligned}$$

which follows by choosing $\xi_{\boldsymbol{u}}^1$ as test function in the first equation of Lemma 5.7 and system (5.24), performing an Euler scheme step, subtracting both equations and adding $\pm a_1^h(\Pi_h\,\boldsymbol{u}(\Delta t),\xi_{\boldsymbol{u}}^1)$.

Now by applying the error approximation results from Lemmas 5.4 to 5.6, Young's inequality and the stability properties from Section 5.4.2, we get

$$\begin{aligned}
&\|\xi^1_{\boldsymbol{u}}\|^2_{\boldsymbol{L}^2_1} + \frac{1}{4}\Delta t\tilde{\alpha}_a\|\xi^1_{\boldsymbol{u}}\|^2_{1,\mathcal{T}_h} \\
&\leq Ch^{2k}\Delta t\big(\|\boldsymbol{u}(\Delta t)\|^2_{\boldsymbol{H}^{k+1}_1} + \|\boldsymbol{u}(0)\|^2_{\boldsymbol{H}^{k+1}_1} + \|\vec{\theta}(\Delta t)\|^2_{H^{k+1}_1} + \|p(\Delta t)\|^2_{H^k_1}\big) \\
&\quad + C\Delta t^4\|\boldsymbol{u}''\|^2_{L^\infty(0,\Delta t;\boldsymbol{L}^2_1)} + 48C^2_{\boldsymbol{F}}\Delta t\|\xi^1_{\vec{\theta}}\|^2_{\vec{L}^2_1}.
\end{aligned} \tag{5.38}$$

Next we follow the same steps to obtain for $\vec{\theta}$

$$\begin{aligned}
&\frac{1}{2}\|\xi^1_{\vec{\theta}}\|^2_{\vec{L}^2_1} + \frac{1}{2}\Delta t\hat{\alpha}_a\|\xi^1_{\vec{\theta}}\|^2_{\vec{H}^1_1} \\
&\leq C\Delta t h^{2k}\big(\|\boldsymbol{u}(\Delta t)\|^2_{\boldsymbol{H}^{k+1}_1} + \|\vec{\theta}(\Delta t)\|^2_{\vec{H}^{k+1}_1} + \|\vec{\theta}(0)\|^2_{\vec{H}^{k+1}_1}\big) \\
&\quad + C\Delta t^4\|T''\|^2_{L^\infty(0,\Delta t;L^2_1)} + \frac{3\tilde{C}C^*\Delta t}{2\hat{\alpha}_a}\|\xi_{\boldsymbol{u}}\|^2_{1,\mathcal{T}_h} + \frac{5\Delta t|f|^2_{\mathrm{Lip}}C^*}{\hat{\alpha}_a}\|\vec{\theta}(\Delta t)\|^2\|\xi^1_{\vec{s}}\|^2_{\vec{L}^2_1},
\end{aligned} \tag{5.39}$$

and analogously for $\vec{s}$

$$\begin{aligned}
&\frac{1}{2}\|\xi_{\vec{s}}\|^2_{\vec{L}^2_1} + \frac{1}{2}\Delta t g_1\|\xi_{\vec{s}}\|^2_{\vec{L}^2_1} \\
&\leq Ch^{2k}\Delta t^2\big(\|\vec{s}(\Delta t)\|^2_{\vec{H}^k_1} + \|\vec{s}(0)\|^2_{\vec{H}^k_1} + \|\vec{\theta}(\Delta t)\|^2_{\vec{H}^{k+1}_1}\big) \\
&\quad + C\Delta t^4\|\vec{s}''\|^2_{L^\infty(0,\Delta t;\vec{L}^2_1)} + \frac{5|g|^2_{\mathrm{Lip}}\Delta t}{2g_1}\big(1 + \|\vec{s}(\Delta t)\|^2_{\vec{H}^1_1}\big)\|\xi_{\vec{\theta}}\|^2_{\vec{H}^1_1}.
\end{aligned} \tag{5.40}$$

In this way, from (5.38) and (5.40) we have that

$$\begin{aligned}
&\frac{3\tilde{C}C^*\epsilon_2\Delta t}{2\hat{\alpha}_a}\|\xi_{\boldsymbol{u}}\|^2_{1,\mathcal{T}_h} \leq C(h^{2k} + \Delta t^4) + \frac{144\tilde{C}C^*C^2_{\boldsymbol{F}}\Delta t}{\tilde{\alpha}_a\hat{\alpha}}\|\xi^1_{\vec{\theta}}\|^2_{\vec{L}^2_1}, \\
&\frac{5\Delta t|f|^2_{\mathrm{Lip}}C^*}{\hat{\alpha}_a}\|\vec{\theta}(\Delta t)\|^2\|\xi^1_{\vec{s}}\|^2_{\vec{L}^2_1} \\
&\leq C(h^{2k} + \Delta t^4) + \frac{25\Delta t|f|^2_{\mathrm{Lip}}C^*|g|^2_{\mathrm{Lip}}}{\hat{\alpha}_a g_1^2}\big(1 + \|\vec{s}(\Delta t)\|^2_{\vec{H}^1_1}\big)\|\vec{\theta}(\Delta t)\|^2\|\xi_{\vec{\theta}}\|^2_{\vec{L}^2_1}.
\end{aligned}$$

We substitute these inequalities into (5.39) and consider Δt sufficiently small such that the terms multiplying $\|\xi_{\vec{\theta}}\|^2_{\vec{L}^2_1}$ can be absorbed into the left-hand side of the inequality to get

$$\frac{1}{4}\|\xi^1_{\vec{\theta}}\|^2_{\vec{L}^2_1} + \frac{1}{2}\Delta t\hat{\alpha}_a\|\xi_{\vec{\theta}}\|^2_{\vec{H}^1_1} \leq C^1_\theta(h^{2k} + \Delta t^4). \tag{5.41}$$

Finally we deduce the first and third desired estimates by directly substituting (5.41) on (5.38) and (5.40). □

Theorem 5.6. *Let $(\boldsymbol{u}, p, \vec{\theta}, \vec{s})$ be the solution of (5.4), (5.5) under the assumptions of Section 5.3, and $(\boldsymbol{u}_h, p_h, \vec{\theta}_h, \vec{s}_h)$ be the solution of (5.25). Suppose that*

$$\begin{aligned}
&\boldsymbol{u} \in L^\infty(0,\mathcal{T};\boldsymbol{H}^{k+1}_1) \cap L^\infty(0,\mathcal{T};\boldsymbol{V}^1_{1,\diamond}(\Omega_{\mathrm{a}})), \\
&\vec{\theta} \in L^\infty(0,\mathcal{T};\vec{H}^{k+1}_{1,\diamond}(\Omega_{\mathrm{a}})), \quad \boldsymbol{u}' \in L^\infty(0,\mathcal{T};\boldsymbol{H}^k_1), \quad \boldsymbol{u}^{(3)} \in L^2(0,\mathcal{T};\boldsymbol{L}^2_1)
\end{aligned}$$

and $\|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T};\boldsymbol{H}_1^1)} < M$ *for a sufficiently small constant* $M > 0$*. Then there exist positive constants* $C, \gamma_1 \geq 0$ *independent of* h *and* Δt *such that for all* $m+1 \leq N$*,*

$$\begin{aligned}
&\|\xi_{\boldsymbol{u}}^{m+1}\|_{\boldsymbol{L}_1^2}^2 + \|2\xi_{\boldsymbol{u}}^{m+1} - \xi_{\boldsymbol{u}}^{m}\|_{\boldsymbol{L}_1^2}^2 + \sum_{n=1}^{m} \|\Lambda \xi_{\boldsymbol{u}}^{n}\|_{\boldsymbol{L}_1^2}^2 + \sum_{n=1}^{m} \Delta t \tilde{\alpha}_a \|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2 \\
&\quad \leq C(\Delta t^4 + h^{2k}) + \sum_{n=1}^{m} \gamma_1 \Delta t \|\xi_{\vec{\theta}}^{n+1}\|_{\vec{L}_1^2}^2.
\end{aligned}$$

Proof. We choose as test function $\boldsymbol{v}_h = \xi_{\boldsymbol{u}}^{n+1}$ in the first equation of (5.25) and insert the terms

$$\begin{aligned}
&\pm \frac{1}{2\Delta t}\big(3\boldsymbol{u}(t_{n+1}) - 4\boldsymbol{u}(t_n) + \boldsymbol{u}(t_{n-1}), \xi_{\boldsymbol{u}}^{n+1}\big), \\
&\pm \frac{1}{2\Delta t}\big(3\Pi_h\, \boldsymbol{u}(t_{n+1}) - 4\Pi_h\, \boldsymbol{u}(t_n) + \Pi_h\, \boldsymbol{u}(t_{n-1}), \xi_{\boldsymbol{u}}^{n+1}\big), \quad \pm a_1^h\big(\Pi_h\, \boldsymbol{u}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1}\big).
\end{aligned}$$

Hence we get

$$\begin{aligned}
&\frac{1}{2\Delta t}\big(3\xi_{\boldsymbol{u}}^{n+1} - 4\xi_{\boldsymbol{u}}^{n} + \xi_{\boldsymbol{u}}^{n-1}, \xi_{\boldsymbol{u}}^{n+1}\big)_{1,\Omega_{\mathrm{a}}} + \frac{1}{2\Delta t}\big(3E_{\boldsymbol{u}}^{n+1} - 4E_{\boldsymbol{u}}^{n} + E_{\boldsymbol{u}}^{n-1}, \xi_{\boldsymbol{u}}^{n+1}\big)_{1,\Omega_{\mathrm{a}}} \\
&+ \frac{1}{2\Delta t}\big(3\boldsymbol{u}(t_{n+1}) - 4\boldsymbol{u}(t_n) + \boldsymbol{u}(t_{n-1}), \xi_{\boldsymbol{u}}^{n+1}\big)_{1,\Omega_{\mathrm{a}}} + a_1^h(\xi_{\boldsymbol{u}}^{n+1}, \xi_{\boldsymbol{u}}^{n+1}) \\
&+ a_1^h(\Pi_h\, \boldsymbol{u}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1}) + c_1^h(\boldsymbol{u}_h^{n+1}, \boldsymbol{u}_h^{n+1}, \xi_{\boldsymbol{u}}^{n+1}) + b(\xi_{\boldsymbol{u}}^{n+1}, p_h^{n+1}) \\
&= d_1(\vec{\theta}_h^{n+1}, \xi_{\boldsymbol{u}}^{n+1}).
\end{aligned} \tag{5.42}$$

Considering the first equation on Lemma 5.7 at $t = t_{n+1}$ with $\boldsymbol{v} = \xi_{\boldsymbol{u}}^{n+1}$, and after inserting the term

$$\pm \frac{1}{2\Delta t}\big(3\boldsymbol{u}(t_{n+1}) - 4\boldsymbol{u}(t_n) + \boldsymbol{u}(t_{n-1}), \xi_{\boldsymbol{u}}^{n+1}\big)_{1,\Omega_{\mathrm{a}}},$$

we readily deduce the identity

$$\begin{aligned}
&\frac{1}{2\Delta t}\big(3\boldsymbol{u}(t_{n+1}) - 4\boldsymbol{u}(t_n) + \boldsymbol{u}(t_{n-1}), \xi_{\boldsymbol{u}}^{n+1}\big)_{1,\Omega_{\mathrm{a}}} + a_1^h\big(\boldsymbol{u}(t_n), \xi_{\boldsymbol{u}}^{n+1}\big) \\
&+ c_1^h\big(\boldsymbol{u}(t_{n+1}), \boldsymbol{u}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1}\big) + b(\xi_{\boldsymbol{u}}^{n+1}, p(t_{n+1})) \\
&= d_1(\vec{\theta}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1}) \\
&- \left(\boldsymbol{u}'(t_{n+1}) - \frac{1}{2\Delta t}\big(3\boldsymbol{u}(t_{n+1}) - 4\boldsymbol{u}(t_n) + \boldsymbol{u}(t_{n-1})\big), \xi_{\boldsymbol{u}}^{n+1}\right)_{1,\Omega_{\mathrm{a}}}.
\end{aligned} \tag{5.43}$$

We can then subtract (5.43) from (5.42) and multiply both sides by $4\Delta t$ to obtain an identity $I_1 + I_2 + \cdots + I_8 = 0$, where

$$\begin{aligned}
&I_1 := 2\big(3\xi_{\boldsymbol{u}}^{n+1} - 4\xi_{\boldsymbol{u}}^{n} + \xi_{\boldsymbol{u}}^{n-1}, \xi_{\boldsymbol{u}}^{n+1}\big), \quad I_2 := 4\Delta t a_1^h(\xi_{\boldsymbol{u}}^{n+1}, \xi_{\boldsymbol{u}}^{n+1})_{1,\Omega_{\mathrm{a}}}, \\
&I_3 := 4\Delta t \left(\boldsymbol{u}'(t_{n+1}) - \frac{1}{2\Delta t}\big(3\boldsymbol{u}(t_{n+1}) - 4\boldsymbol{u}(t_n) + \boldsymbol{u}(t_{n-1})\big), \xi_{\boldsymbol{u}}^{n+1}\right)_{1,\Omega_{\mathrm{a}}}, \\
&I_4 := 2\big(3E_{\boldsymbol{u}}^{n+1} - 4E_{\boldsymbol{u}}^{n} + E_{\boldsymbol{u}}^{n-1}, \xi_{\boldsymbol{u}}^{n+1}\big), \quad I_5 := -4\Delta t d_1(\vec{\theta}_h^{n+1} - \vec{\theta}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1})_{1,\Omega_{\mathrm{a}}}, \\
&I_6 := 4\Delta t a_1^h\big(E_{\boldsymbol{u}}^{n+1}, \xi_{\boldsymbol{u}}^{n+1}\big), \\
&I_7 := 4\Delta t\big(c_1^h(\boldsymbol{u}_h^{n+1}, \boldsymbol{u}_h^{n+1}, \xi_{\boldsymbol{u}}^{n+1}) - c_1^h\big(\boldsymbol{u}(t_{n+1}), \boldsymbol{u}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1}\big)\big), \\
&I_8 := 4\Delta t b(\xi_{\boldsymbol{u}}^{n+1}, p_h^{n+1} - p(t_{n+1})).
\end{aligned}$$

For the first term, using (5.32) we can assert that

$$I_1 = \|\xi_{\boldsymbol{u}}^{n+1}\|_{\boldsymbol{L}_1^2}^2 + \|2\xi_{\boldsymbol{u}}^{n+1} - \xi_{\boldsymbol{u}}^{n}\|_{\boldsymbol{L}_1^2}^2 + \|\Lambda\xi_{\boldsymbol{u}}^{n+1}\|_{\boldsymbol{L}_1^2}^2 - \|\xi_{\boldsymbol{u}}^{n}\|_{\boldsymbol{L}_1^2}^2 - \|2\xi_{\boldsymbol{u}}^{n} - \xi_{\boldsymbol{u}}^{n-1}\|_{\boldsymbol{L}_1^2}^2.$$

Using the ellipticity stated in (5.28), we readily get

$$I_2 \geq 4\Delta t\tilde{\alpha}_a\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2.$$

By using Taylor's formula with integral remainder we have

$$\left|\boldsymbol{u}'(t_{n+1}) - \frac{3\boldsymbol{u}(t_{n+1}) - 4\boldsymbol{u}(t_n) + \boldsymbol{u}(t_{n-1})}{2\Delta t}\right| = \frac{\Delta t^{3/2}}{2\sqrt{3}}\|\boldsymbol{u}^{(3)}\|_{L^2(t^{n-1},t^{n+1};\boldsymbol{L}_1^2)},$$

then by combining Cauchy-Schwarz and Young's inequality, we obtain the bound

$$|I_3| \leq \frac{\Delta t^4}{24\varepsilon_1}\|\boldsymbol{u}^{(3)}\|_{L^2(t_{n-1},t_{n+1};\boldsymbol{L}_1^2)}^2 + \frac{\Delta t\varepsilon_1}{2}\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2.$$

Now we insert $\pm 4\Delta t E_{\boldsymbol{u}}'(t_{n+1})$ into the fourth term, which leads to

$$\begin{aligned} I_4 = &-4\Delta t(E_{\boldsymbol{u}}'(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1})_{1,\Omega_{\mathrm{a}}} \\ &+ \left(E_{\boldsymbol{u}}'(t_{n+1}) - \frac{3E_{\boldsymbol{u}}^{n+1} - 4E_{\boldsymbol{u}}^{n} + E_{\boldsymbol{u}}^{n-1}}{2\Delta t}, \xi_{\boldsymbol{u}}^{n+1}\right)_{1,\Omega_{\mathrm{a}}}. \end{aligned}$$

Proceeding as before and using Lemma 5.5 on the first term of I_4, we get

$$\begin{aligned} |I_4| \leq &\frac{C}{2\varepsilon_2}h^{2k}\|\boldsymbol{u}'\|_{\boldsymbol{L}^\infty(0,\mathcal{T};\boldsymbol{H}_1^k)}^2 + \frac{\Delta t\varepsilon_2}{2}\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2 \\ &+ \frac{\Delta t^4 C}{2\varepsilon_3}\|\boldsymbol{u}^{(3)}\|_{L^2(0,\mathcal{T};\boldsymbol{L}_1^2)}^2 + \frac{\Delta t\varepsilon_3}{2}\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2. \end{aligned}$$

Now by (5.27a), appealing to Lemma 5.6, and inserting $\pm 4\Delta t d_1(\mathcal{I}_h\,\vec{\theta}^{n+1}, \xi_{\boldsymbol{u}}^{n+1})$, we are left with

$$\begin{aligned} |I_5| &\leq 4\Delta t C_{\boldsymbol{F}}\|\xi_{\vec{\theta}}^{n+1} + E_{\vec{\theta}}^{n+1}\|_{\vec{L}_1^2}\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h} \\ &\leq \frac{16C_{\boldsymbol{F}}^2\Delta t}{2\varepsilon_4}\left(Ch^{2k}\|\vec{\theta}\|_{L^\infty(0,\mathcal{T};\vec{H}_1^{k+1})}^2 + \|\xi_{\vec{\theta}}^{n+1}\|_{\vec{L}_1^2}^2\right) + \frac{\Delta t\varepsilon_4}{2}\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2. \end{aligned}$$

And again by Lemmas 5.5 and 5.3 we immediately have

$$|I_6| \leq 4\Delta t\tilde{C}_a\|E_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h} \leq \frac{2\tilde{C}_a^2\Delta t h^{2k}}{\varepsilon_5}\|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T};\boldsymbol{H}_1^{k+1})}^2 + \frac{\Delta t\varepsilon_5}{2}\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2.$$

Adding and subtracting suitable terms within I_7 yields

$$I_7 = \tilde{I}_7 - 4\delta t c_1^h\big(\boldsymbol{u}_h^{n+1}, \xi_{\boldsymbol{u}}^{n+1}, \xi_{\boldsymbol{u}}^{n+1}\big),$$

where we define

$$\begin{aligned} \tilde{I}_7 := -4\Delta t\big(&c_1^h(\boldsymbol{u}(t_{n+1}), \Pi_h\,\boldsymbol{u}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1}) - c_1^h(\Pi_h\,\boldsymbol{u}(t_{n+1}), \Pi_h\,\boldsymbol{u}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1}) \\ &+ c_1^h(\Pi_h\,\boldsymbol{u}(t_{n+1}), \Pi_h\,\boldsymbol{u}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1}) - c_1^h(\Pi_h\,\boldsymbol{u}(t_{n+1}), \boldsymbol{u}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1} \\ &+ c_1^h(\Pi_h\,\boldsymbol{u}(t_{n+1}), \boldsymbol{u}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1}) - c_1^h(\boldsymbol{u}(t_{n+1})\boldsymbol{u}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1})\big). \end{aligned}$$

The bound (5.31) and Lemma 5.5 imply that

$$
\begin{aligned}
|\tilde{I}_7| &\le 4\Delta t\tilde{C}_c\big(\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2\|\Pi_h\,\boldsymbol{u}(t_{n+1})\|_{1,\mathcal{T}_h} + \|\Pi_h\,\boldsymbol{u}(t_{n+1})\|_{1,\mathcal{T}_h}\|E_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h} \\
&\quad + \|E_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}\|\boldsymbol{u}(t_{n+1})\|_{1,\mathcal{T}_h}\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}\big) \\
&\le 4\Delta t\bigg(\tilde{C}_cC^*\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2\|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T};\boldsymbol{H}_1^1)}^2 \\
&\quad + \frac{h^{2k}C\tilde{C}_c^2}{2\varepsilon_6}\|\boldsymbol{u}\|_{L^\infty(0,T;\boldsymbol{H}_1^1)}^2\|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T};\boldsymbol{H}_1^{k+1})}^2 + \frac{\varepsilon_6}{2}\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2 \\
&\quad + \frac{Ch^{2k}\tilde{C}_c^2}{2\varepsilon_7}\|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T};\boldsymbol{H}_1^{k+1})}^2\|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T};\boldsymbol{H}_1^1)}^2 + \frac{\varepsilon_7}{2}\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2\bigg) \\
&\le 4\Delta t\bigg(C^*\tilde{C}_cM\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2 + \frac{h^{2k}C}{2\varepsilon_6}\|\boldsymbol{u}\|_{L^\infty(0,T;\boldsymbol{H}_1^1)}^2\|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T};\boldsymbol{H}_1^{k+1})}^2 \\
&\qquad + \frac{\varepsilon_6}{2}\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2 + \frac{Ch^{2k}}{2\varepsilon_7}\|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T};\boldsymbol{H}_1^{k+1})}^2\|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T};\boldsymbol{H}_1^1)}^2 \\
&\qquad + \frac{\varepsilon_7}{2}\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2\bigg),
\end{aligned}
$$

where C^* is a positive constant coming from Lemma 5.5. Finally, using Lemmas 5.3 and 5.4 we obtain

$$
|I_8| \le \frac{8\Delta tC\tilde{C}_b^2h^{2k}}{\varepsilon_8}\|p\|_{L^\infty(0,\mathcal{T};H^k(\Omega_{\mathrm{a}}))}^2 + \frac{\Delta t\varepsilon_8}{2}\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2.
$$

Hence, by choosing $\varepsilon_i = \tilde{\alpha}_a/3$ for $i = \{1,2,3,4,5,8\}$, $\varepsilon_6 = \varepsilon_7 = 7\tilde{\alpha}_a/16$, collecting the above estimates, and summing over $1 \le n \le m$ for all $m+1 \le N$; we get

$$
\begin{aligned}
&\|\xi_{\boldsymbol{u}}^{m+1}\|_{\boldsymbol{L}_1^2}^2 + \|2\xi_{\boldsymbol{u}}^{m+1} - \xi_{\boldsymbol{u}}^m\|_{\boldsymbol{L}_1^2}^2 + \sum_{n=1}^m\|\Lambda\xi_{\boldsymbol{u}}^n\|_{\boldsymbol{L}_1^2}^2 - 3\|\xi_{\boldsymbol{u}}^1\|_{\boldsymbol{L}_1^2}^2 \\
&\quad + \sum_{n=1}^m\Delta t\tilde{\alpha}_a\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2 \le C(\Delta t^4 + h^{2k}) + \frac{24C_{\boldsymbol{F}}^2\Delta t}{\tilde{\alpha}_a}\sum_{n=1}^m\|\xi_{\vec{\theta}}^{n+1}\|_{\vec{L}_1^2}^2.
\end{aligned}
$$

where $4\tilde{C}_cC^*M \le \tilde{\alpha}_a/4$ and $\gamma_1 = 24C_{\boldsymbol{F}}^2/\tilde{\alpha}_a$. Finally, using Theorem 5.5, we get the desired result. □

Theorem 5.7. *Let $(\boldsymbol{u}, p, \vec{\theta}, \vec{s})$ be the solution of (5.4), (5.5) under the assumptions of Section 5.3 and $(\boldsymbol{u}_h, p_h, \vec{\theta}_h, \vec{s}_h)$ be the solution of (5.25). If*

$$
\begin{aligned}
&\boldsymbol{u} \in L^\infty(0,\mathcal{T};\boldsymbol{H}_1^{k+1}) \cap L^\infty(0,\mathcal{T};\boldsymbol{V}_{1,\diamond}^1(\Omega_{\mathrm{a}})), \quad \vec{\theta} \in L^\infty(0,\mathcal{T};\vec{H}_{1,\diamond}^{k+1}(\Omega_{\mathrm{a}})), \\
&\vec{\theta}' \in L^\infty(0,\mathcal{T};\vec{H}_1^k), \quad \vec{\theta}^{(3)} \in L^2(0,\mathcal{T};\vec{L}_1^2), \quad \vec{s} \in L^\infty(0,\mathcal{T};\vec{H}_1^k),
\end{aligned}
$$

then there exist constants $C, \gamma_s, \gamma_u > 0$, independent of h and Δt, such that for all $m+1 \le N$

$$
\begin{aligned}
&\|\xi_{\vec{\theta}}^{m+1}\|_{\vec{L}_1^2}^2 + \|2\xi_{\vec{\theta}}^{m+1} - \xi_{\vec{\theta}}^m\|_{\vec{L}_1^2}^2 + \sum_{n=1}^m\|\Lambda\xi_{\vec{\theta}}^{n+1}\|_{\vec{L}_1^2}^2 + \sum_{n=1}^m\Delta t\hat{\alpha}_a\|\xi_{\vec{\theta}}^{n+1}\|_{\vec{H}_1^1}^2 \\
&\qquad \le C(\Delta t^4 + h^{2k}) + \sum_{n=1}^m\gamma_s\Delta t\|\xi_{\vec{s}}^{n+1}\|_{\vec{L}_1^2}^2 + \sum_{n=1}^m\gamma_u\Delta t\|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2.
\end{aligned}
$$

Proof. Proceeding similarly as in the proof of Theorem 5.6, we choose as test function $\vec{\psi}_h = \xi_{\vec{\theta}}^{n+1}$ in the second equation of (5.25) and insert suitable additional terms to obtain the following identity, which is analogous to (5.42):

$$\begin{aligned}
&\frac{1}{2\Delta t}\big(3\xi_{\vec{\theta}}^{n+1} - 4\xi_{\vec{\theta}}^{n} + \xi_{\vec{\theta}}^{n-1}, \xi_{\vec{\theta}}^{n+1}\big)_{1,\Omega_\mathrm{a}} + \frac{1}{2\Delta t}\big(3E_{\vec{\theta}}^{n+1} - 4E_{\vec{\theta}}^{n} + E_{\vec{\theta}}^{n-1}, \xi_{\vec{\theta}}^{n+1}\big)_{1,\Omega_\mathrm{a}} \\
&+ \frac{1}{2\Delta t}\big(3\vec{\theta}(t_{n+1}) - 4\vec{\theta}(t_n) + \vec{\theta}(t_{n-1}), \xi_{\vec{\theta}}^{n+1}\big)_{1,\Omega_\mathrm{a}} \\
&+ a_2^h(\xi_{\vec{\theta}}^{n+1}, \xi_{\vec{\theta}}^{n+1}) + a_2^h(\mathcal{I}_h\,\vec{\theta}(t_{n+1}), \xi_{\vec{\theta}}^{n+1}) + c_2^h(\boldsymbol{u}_h^{n+1}, \vec{\theta}_h^{n+1}, \xi_{\vec{\theta}}^{n+1}) \\
&= -d_2(\vec{s}_h^{n+1}; \vec{\theta}_h^{n+1}, \xi_{\vec{\theta}}^{n+1}).
\end{aligned} \tag{5.44}$$

Starting from the second equation in Lemma 5.7, focusing on $t = t_{n+1}$, using $\vec{\psi} = \xi_{\vec{\theta}}^{n+1}$ and proceeding as in the derivation of (5.43), we obtain

$$\begin{aligned}
&\frac{1}{2\Delta t}\big(3\vec{\theta}(t_{n+1}) - 4\vec{\theta}(t_n) + \vec{\theta}(t_{n-1}), \xi_{\vec{\theta}}^{n+1}\big)_{1,\Omega_\mathrm{a}} + a_2^h\big(\vec{\theta}(t_{n+1}), \xi_{\vec{\theta}}^{n+1}\big) \\
&+ c_2^h\big(\boldsymbol{u}(t_{n+1}), \vec{\theta}(t_{n+1}), \xi_{\vec{\theta}}^{n+1}\big) \\
&= d_2\big(\vec{s}(t_{n+1}); \vec{\theta}(t_{n+1}), \xi_{\vec{\theta}}^{n+1}\big) \\
&\quad - \left(\vec{\theta}'(t_{n+1}) - \frac{3\vec{\theta}(t_{n+1}) - 4\vec{\theta}(t_n) + \vec{\theta}(t_{n-1})}{2\Delta t}, \xi_{\vec{\theta}}^{n+1}\right)_{1,\Omega_\mathrm{a}}.
\end{aligned} \tag{5.45}$$

Next we proceed to subtract (5.45) from (5.44), and to multiply both sides by $4\Delta t$. This leads to an identity $\hat{I}_1 + \hat{I}_2 + \cdots + \hat{I}_7 = 0$, where

$$\begin{aligned}
&\hat{I}_1 := 2\big(3\xi_{\vec{\theta}}^{n+1} - 4\xi_{\vec{\theta}}^{n} + \xi_{\vec{\theta}}^{n-1}, \xi_{\vec{\theta}}^{n+1}\big)_{1,\Omega_\mathrm{a}}, \quad \hat{I}_2 := 4\Delta t a_2^h\big(\xi_{\vec{\theta}}^{n+1}, \xi_{\vec{\theta}}^{n+1}\big), \\
&\hat{I}_3 := 4\Delta t\left(\vec{\theta}'(t_{n+1}) - \frac{3\vec{\theta}(t_{n+1}) - 4\vec{\theta}(t_n) + \vec{\theta}(t_{n-1})}{2\Delta t}, \xi_{\vec{\theta}}^{n+1}\right)_{1,\Omega_\mathrm{a}}, \\
&\hat{I}_4 := 2\big(3E_{\vec{\theta}}^{n+1} - 4E_{\vec{\theta}}^{n} + E_{\vec{\theta}}^{n-1}, \xi_{\vec{\theta}}^{n+1}\big)_{1,\Omega_\mathrm{a}}, \quad \hat{I}_5 := 4\Delta t a_2^h\big(E_{\vec{\theta}}^{n+1}, \xi_{\vec{\theta}}^{n+1}\big), \\
&\hat{I}_6 := 4\Delta t\big(c_2^h\big(\boldsymbol{u}_h^{n+1}, \vec{\theta}_h^{n+1}, \xi_{\boldsymbol{u}}^{n+1}\big) - c_1^h\big(\boldsymbol{u}(t_{n+1}), \vec{\theta}(t_{n+1}), \xi_{\boldsymbol{u}}^{n+1}\big)\big), \\
&\hat{I}_7 := 4\Delta t\big(d_2\big(\vec{s}_h^{n+1}, \vec{\theta}_h^{n+1}, \xi_{\vec{\theta}}^{n+1}\big) - d_2\big(\vec{s}(t_{n+1}); \vec{\theta}(t_{n+1}), \xi_{\vec{\theta}}^{n+1}\big)\big).
\end{aligned}$$

For the first, second, and third terms, we use (5.32), (5.17), and Taylor expansion together with Young's inequality, respectively, to obtain

$$\begin{aligned}
&\hat{I}_1 = \|\xi_{\vec{\theta}}^{n+1}\|_{\vec{L}_1^2}^2 + \|2\xi_{\vec{\theta}}^{n+1} - \xi_{\vec{\theta}}^{n}\|_{\vec{L}_1^2}^2 + \|\Lambda\xi_{\vec{\theta}}^{n+1}\|_{\vec{L}_1^2}^2 - \|\xi_{\vec{\theta}}^{n}\|_{\vec{L}_1^2}^2 - \|2\xi_{\vec{\theta}}^{n} - \xi_{\vec{\theta}}^{n-1}\|_{\vec{L}_1^2}^2, \\
&\hat{I}_2 \geq 4\Delta t\hat{\alpha}_a\|\xi_{\vec{\theta}}^{n+1}\|_{\vec{H}_1^1}^2, \\
&|\hat{I}_3| \leq \frac{\Delta t^4}{24\varepsilon_1}\|\vec{\theta}^{(3)}\|_{L^2(t_{n-1},t_{n+1};\vec{L}_1^2)}^2 + \frac{\Delta t\varepsilon_1}{2}\|\xi_{\vec{\theta}}^{n+1}\|_{\vec{H}_1^1}^2.
\end{aligned}$$

Inserting $\pm 4\Delta t E_{\vec{\theta}}'(t_{n+1})$ into $\hat{I}_4$ and using Lemma 5.6 leads to the bound

$$\begin{aligned}
|\hat{I}_4| \leq{}& \frac{C}{2\varepsilon_2}h^{2k}\|\vec{\theta}'\|_{L^\infty(0,\mathcal{T};\vec{H}_1^k)}^2 + \frac{\Delta t\varepsilon_2}{2}\|\xi_{\vec{\theta}}^{n+1}\|_{\vec{H}_1^1}^2 \\
&+ \frac{\Delta t^4 C}{2\varepsilon_3}\|\vec{\theta}^{(3)}\|_{L^2(0,\mathcal{T};\vec{L}_1^2)}^2 + \frac{\Delta t\varepsilon_3}{2}\|\xi_{\vec{\theta}}^{n+1}\|_{\vec{H}_1^1}^2.
\end{aligned}$$

Employing again Lemma 5.6 in combination with (5.14b) we have

$$|\hat{I}_5| \leq \frac{2\hat{C}_a^2 \Delta t h^{2k}}{\varepsilon_4} \|\vec{\theta}\|^2_{L^\infty(0,\mathcal{T};H_1^{k+1})} + \frac{\Delta t \varepsilon_4}{2} \|\xi_{\vec{\theta}}^{n+1}\|^2_{\vec{H}_1^1}.$$

In order to derive a bound for $\hat{I}_6$ we proceed as for the bound on I_7 in the proof of Theorem 5.6; namely adding and subtracting suitable terms in the definition of $\hat{I}_6$, defining $\tilde{I}_6$ in this case by

$$\hat{I}_6 = \tilde{I}_6 + 4\delta_t c_2^h\big(\boldsymbol{u}_h^{n+1}, \xi_{\vec{\theta}}^{n+1}, \xi_{\vec{\theta}}^{n+1}\big),$$

and applying (5.30), (5.27b) and Lemma 5.6 to the result, we get

$$\begin{aligned}
|\tilde{I}_6| \leq & 4\Delta t\Bigg(\frac{\tilde{C}^2 C^*}{2\varepsilon_5} \|\xi_{\boldsymbol{u}}^{n+1}\|^2_{1,\mathcal{T}_h} \|\vec{\theta}\|^2_{L^\infty(0,\mathcal{T};\vec{H}_1^1)} + \frac{1}{2\varepsilon_5} \|\xi_{\vec{\theta}}\|^2_{\vec{H}_1^1} \\
& + \frac{h^{2k} C \tilde{C}^2}{2\varepsilon_6} \|\boldsymbol{u}\|^2_{L^\infty(0,T;\boldsymbol{H}_1^1)} \|\vec{\theta}\|^2_{L^\infty(0,\mathcal{T};\vec{H}_1^{k+1})} + \frac{\varepsilon_6}{2} \|\xi_{\vec{\theta}}^{n+1}\|^2_{\vec{H}_1^1} \\
& + \frac{C h^{2k} \tilde{C}^2}{2\varepsilon_7} \|\boldsymbol{u}\|^2_{L^\infty(0,\mathcal{T};\boldsymbol{H}_1^{k+1})} \|\vec{\theta}\|^2_{L^\infty(0,\mathcal{T};\vec{H}_1^1)} + \frac{\varepsilon_7}{2} \|\xi_{\vec{\theta}}^{n+1}\|^2_{\vec{H}_1^1} \Bigg).
\end{aligned}$$

Next we add and subtract suitable terms in $\hat{I}_7$ to obtain

$$\begin{aligned}
\hat{I}_7 = & -4\Delta t\big(d_2\big(\vec{s}_h^{n+1}, \mathcal{I}_h\, \vec{\theta}(t_{n+1}), \xi_{\vec{\theta}}^{n+1}\big) - d_2\big(\mathcal{L}_h\, \vec{s}(t_{n+1}), \mathcal{I}_h\, \vec{\theta}(t_{n+1}), \xi_{\vec{\theta}}^{n+1}\big) \\
& + d_2\big(\mathcal{L}_h\, \vec{s}(t_{n+1}), \mathcal{I}_h\, \vec{\theta}(t_{n+1}), \xi_{\vec{\theta}}^{n+1}\big) - d_2\big(\mathcal{L}_h\, \vec{s}(t_{n+1}), \vec{\theta}(t_{n+1}), \xi_{\vec{\theta}}^{n+1}\big) \\
& + d_2\big(\mathcal{L}_h\, \vec{s}(t_{n+1}), \vec{\theta}(t_{n+1}), \xi_{\vec{\theta}}^{n+1}\big) - d_2\big(\vec{s}(t_{n+1}), \vec{\theta}(t_{n+1}), \xi_{\vec{\theta}}^{n+1}\big) \\
& + d_2\big(\vec{s}_h^{n+1}, \xi_{\vec{\theta}}^{n+1}, \xi_{\vec{\theta}}^{n+1}\big)\big).
\end{aligned}$$

After passing the last expression to the left-hand side and using (5.7), we can combine (5.8) and (5.9), to infer that the remaining terms in $\hat{I}_7$ (which we now denote as $\hat{I}_7^*$) are bounded as follows:

$$\begin{aligned}
|\hat{I}_7^*| \leq & \frac{8|f|^2_{\mathrm{Lip}} \Delta t}{\varepsilon_8} \|\xi_{\vec{s}}^{n+1}\|^2_{\vec{L}_1^2} \|\vec{\theta}\|^2_{L^\infty(0,\mathcal{T};\vec{H}_1^1)} + \frac{\varepsilon_8 \Delta t}{2} \|\xi_{\vec{\theta}}\|^2_{\vec{H}_1^1} \\
& + \frac{8 f_2^2 \Delta t h^{2k}}{\varepsilon_9} \|\vec{\theta}\|^2_{L^\infty(0,\mathcal{T};\vec{H}_1^{k+1})} + \frac{\Delta t \varepsilon_9}{2} \|\xi_{\vec{\theta}}\|^2_{\vec{H}_1^1} \\
& + \frac{8|f|^2_{\mathrm{Lip}} \Delta t h^{2k}}{\varepsilon_{10}} \|\vec{s}\|^2_{L^\infty(0,\mathcal{T};\vec{H}_1^k)} \|\vec{\theta}\|^2_{L^\infty(0,\mathcal{T};\vec{H}_1^1)} + \frac{\Delta t \varepsilon_{10}}{2} \|\xi_{\vec{\theta}}\|^2_{\vec{H}_1^1}.
\end{aligned}$$

In this manner, and after choosing $\varepsilon_i = 3\hat{\alpha}_a/7$ for $i \in \{1,2,3,4,8,9,10\}$ and $\varepsilon_5 = \varepsilon_6 = \varepsilon_7 = \hat{\alpha}_a/4$, we can collect the above estimates and sum over $1 \leq n \leq m$, for all $m+1 \leq N$, to get

$$\begin{aligned}
& \|\xi_{\vec{\theta}}^{m+1}\|^2_{\vec{L}_1^2} + \|2\xi_{\vec{\theta}}^{m+1} - \xi_{\vec{\theta}}^{m}\|^2_{\vec{L}_1^2} + \sum_{n=1}^{m} \|\Lambda \xi_{\vec{\theta}}^{n}\|^2_{\vec{L}_1^2} + \sum_{n=1}^{m} \Delta t \hat{\alpha}_a \|\xi_{\vec{\theta}}^{n+1}\|^2_{\vec{H}_1^1} - 3\|\xi_{\vec{\theta}}^{1}\|^2_{\vec{L}_1^2} \\
& \leq C(\Delta t^4 + h^{2k}) + \frac{56|f|^2_{\mathrm{Lip}} \Delta t}{3\hat{\alpha}_a} \|\vec{\theta}\|^2_{L^\infty(0,\mathcal{T};\vec{H}_1^1)} \sum_{n=1}^{m} \|\xi_{\vec{s}}^{n+1}\|^2_{\vec{L}_1^2} \\
& \quad + \frac{8\Delta t \tilde{C}^2 C^*}{\hat{\alpha}_a} \|\vec{\theta}\|^2_{L^\infty(0,\mathcal{T};\vec{H}_1^1)} \sum_{n=1}^{m} \|\xi_{\boldsymbol{u}}^{n+1}\|^2_{1,\mathcal{T}_h}.
\end{aligned}$$

Identifying the constants

$$\gamma_s = \frac{56|f|^2_{\mathrm{Lip}}}{3\hat{\alpha}_a}\|\vec{\theta}\|^2_{L^\infty(0,\mathcal{T};\vec{H}^1_1)}, \quad \gamma_u = \frac{8\tilde{C}^2C^*}{\hat{\alpha}_a}\|\vec{\theta}\|^2_{L^\infty(0,\mathcal{T};\vec{H}^1_1)}$$

we may conclude the proof. □

Theorem 5.8. *Let* $(\boldsymbol{u}, p, \vec{\theta}, \vec{s})$ *be the solution of* (5.4), (5.5) *under the assumptions of Section 5.3, and* $(\boldsymbol{u}_h, p_h, \vec{\theta}_h, \vec{s}_h)$ *be the solution of* (5.25). *If*

$$\boldsymbol{u} \in L^\infty(0,\mathcal{T};\boldsymbol{H}^{k+1}_1) \cap L^\infty(0,\mathcal{T};\boldsymbol{V}^1_{1,\diamond}), \quad \vec{\theta} \in L^\infty(0,\mathcal{T};\vec{H}^{k+1}_{1,\diamond}),$$
$$\vec{s} \in L^\infty(0,\mathcal{T};\vec{H}^k_1), \quad \vec{s}' \in L^\infty(0,\mathcal{T};\vec{H}^k_1), \quad \vec{s}^{(3)} \in L^2(0,\mathcal{T};\vec{L}^2_1),$$

then there exist constants $C, \gamma_2 > 0$ *that are independent of* h *and* Δt *such that for all* $m+1 \le N$

$$\|\xi^{m+1}_{\vec{s}}\|^2_{\vec{L}^2_1} + \|2\xi^{m+1}_{\vec{s}} - \xi^m_{\vec{s}}\|^2_{\vec{L}^2_1} + \sum_{n=1}^m \|\Lambda\xi^n_{\vec{s}}\|^2_{\vec{L}^2_1} + \sum_{n=1}^m \Delta t g_1 \|\xi^{n+1}_{\vec{s}}\|^2_{\vec{L}^2_1}$$
$$\le C(\Delta t^4 + h^{2k}) + \gamma_2 \sum_{n=1}^m \Delta t \|\xi^{n+1}_{\vec{\theta}}\|^2_{\vec{L}^2_1}.$$

Proof. We choose as test function $\vec{l}_h = \xi^{n+1}_{\vec{s}}$ in the third equation of (5.25) and add and substract suitable terms. Analogously to (5.43) and (5.44), we obtain

$$\begin{aligned}
&\frac{1}{2\Delta t}\big(3\xi^{n+1}_{\vec{s}} - 4\xi^n_{\vec{s}} + \xi^{n-1}_{\vec{s}}, \xi^{n+1}_{\vec{s}}\big)_{1,\Omega_{\mathrm{a}}} \\
&+ \frac{1}{2\Delta t}\big(3E^{n+1}_{\vec{s}} - 4E^n_{\vec{s}} + E^{n-1}_{\vec{s}} + 3\vec{s}(t_{n+1}) - 4\vec{s}(t_n) + \vec{s}(t_{n-1}), \xi^{n+1}_{\vec{s}}\big)_{1,\Omega_{\mathrm{a}}} \\
&+ d_3\big(\vec{\theta}^{n+1}_h, \xi^{n+1}_{\vec{s}}, \xi^{n+1}_{\vec{s}}\big) + d_3\big(\vec{\theta}^{n+1}_h; \mathcal{L}_h\,\vec{s}(t_{n+1}), \xi^{n+1}_{\vec{s}}\big) - d_4\big(\vec{\theta}^{n+1}_h, \xi^{n+1}_{\vec{s}}\big) = 0.
\end{aligned} \tag{5.46}$$

Now we consider (5.6d) at time $t = t_{n+1}$, using also $\vec{l} = \xi^{n+1}_{\vec{s}}$ as test function. Adding and subtracting a suitable term, we deduce the relation

$$\begin{aligned}
&\frac{1}{2\Delta t}\big(3\vec{s}(t_{n+1}) - 4\vec{s}(t_n) + \vec{s}(t_{n-1}), \xi^{n+1}_{\vec{s}}\big)_{1,\Omega_{\mathrm{a}}} + d_3\big(\vec{\theta}(t_{n+1}); \vec{s}(t_{n+1}), \xi^{n+1}_{\vec{s}}\big) \\
&= d_4\big(\vec{\theta}(t_{n+1}), \xi^{n+1}_{\vec{s}}\big) \\
&\quad - \left(\vec{s}'(t_{n+1}) - \frac{1}{2\Delta t}\big(3\vec{s}(t_{n+1}) - 4\vec{s}(t_n) + \vec{s}(t_{n-1})\big), \xi^{n+1}_{\vec{s}}\right)_{1,\Omega_{\mathrm{a}}}.
\end{aligned} \tag{5.47}$$

As in the two previous proofs, we subtract (5.47) from (5.46) and multiply both sides by $4\Delta t$ to obtain $\bar{I}_1 + \bar{I}_2 + \cdots + \bar{I}_6 = 0$, where

$$\begin{aligned}
&\bar{I}_1 := 2\big(3\xi^{n+1}_{\vec{s}} - 4\xi^n_{\vec{s}} + \xi^{n-1}_{\vec{s}}, \xi^{n+1}_{\vec{s}}\big)_{1,\Omega_{\mathrm{a}}}, \quad \bar{I}_2 := d_3(\vec{\theta}^{n+1}_h, \xi^{n+1}_{\vec{s}}, \xi^{n+1}_{\vec{s}}), \\
&\bar{I}_3 := 4\Delta t\left(\vec{s}'(t_{n+1}) - \frac{1}{2\Delta t}\big(3\vec{s}(t_{n+1}) - 4\vec{s}(t_n) + \vec{s}(t_{n-1})\big), \xi^{n+1}_{\vec{s}}\right)_{1,\Omega_{\mathrm{a}}}, \\
&\bar{I}_4 := 2\big(3E^{n+1}_{\vec{s}} - 4E^n_{\vec{s}} + E^{n-1}_{\vec{s}}, \xi^{n+1}_{\vec{s}}\big)_{1,\Omega_{\mathrm{a}}}, \\
&\bar{I}_5 := -4\Delta t\big(d_3(\vec{\theta}^{n+1}_h; \mathcal{L}_h\,\vec{s}(t_{n+1}), \xi^{n+1}_{\vec{s}}) - d_3(\vec{\theta}(t_{n+1}); \vec{s}(t_{n+1}), \xi^{n+1}_{\vec{s}})\big), \\
&\bar{I}_6 := -4\Delta t d_4\big(\vec{\theta}^{n+1}_h - \vec{\theta}(t_{n+1}), \xi^{n+1}_{\vec{s}}\big).
\end{aligned}$$

For the first, second, and third terms, we proceed to use (5.32), the ellipticity (5.10), and Taylor expansion to get

$$\begin{aligned}
\bar{I}_1 &= \|\xi_{\vec{s}}^{n+1}\|_{\vec{L}_1^2}^2 + \|2\xi_{\vec{s}}^{n+1} - \xi_{\vec{s}}^{n}\|_{\vec{L}_1^2}^2 + \|\Lambda\xi_{\vec{s}}^{n+1}\|_{\vec{L}_1^2}^2 - \|\xi_{\vec{s}}^{n}\|_{\vec{L}_1^2}^2 - \|2\xi_{\vec{s}}^{n} - \xi_{\vec{s}}^{n-1}\|_{\vec{L}_1^2}^2,\\
\bar{I}_2 &\geq 4\Delta t g_1 \|\xi_{\vec{s}}^{n+1}\|_{\vec{L}_1^2}^2,\\
|\bar{I}_3| &\leq \frac{\Delta t^4}{24\varepsilon_1}\|\vec{s}^{(3)}\|_{L^2(t_{n-1},t_{n+1};\vec{L}_1^2)}^2 + \frac{\Delta t\varepsilon_1}{2}\|\xi_{\vec{s}}^{n+1}\|_{\vec{H}_1^1}^2.
\end{aligned}$$

For the fourth term we include $\pm 4\Delta t E'_{\vec{s}}(t_{n+1})$ and use Taylor's formula and Lemma 5.6, which leads to

$$\begin{aligned}
|\bar{I}_4| \leq{}& \frac{C}{2\varepsilon_2}h^{2k}\|\vec{s}'\|_{L^\infty(0,\mathcal{T};\vec{H}_1^k)}^2 + \frac{\Delta t\varepsilon_2}{2}\|\xi_{\vec{s}}^{n+1}\|_{\vec{L}_1^2}^2\\
&+ \frac{\Delta t^4 C}{2\varepsilon_3}\|\vec{s}^{(3)}\|_{L^2(0,\mathcal{T};\vec{L}_1^2)}^2 + \frac{\Delta t\varepsilon_3}{2}\|\xi_{\vec{s}}^{n+1}\|_{\vec{L}_1^2}^2.
\end{aligned}$$

To handle $\bar{I}_5$, we add and subtract the terms

$$d_3\big(\vec{\theta}(t_{n+1});\vec{s}(t_{n+1}),\xi_{\vec{s}}^{n+1}\big) \quad \text{and} \quad d_3\big(\mathcal{I}_h\,\vec{\theta}(t_{n+1})M;\vec{s}(t_{n+1}),\xi_{\vec{s}}^{n+1}\big).$$

Then, owing to (5.11), (5.13), Lemma 5.4, and Young's inequality, we end up with

$$\begin{aligned}
|\bar{I}_5| \leq{}& \frac{Cg_2^2\Delta t h^{2k}}{\varepsilon_4}\|\vec{s}\|_{L^\infty(0,\mathcal{T};\vec{H}_1^k)}^2 + \frac{\varepsilon_4\Delta t}{2}\|\xi_{\vec{s}}\|_{\vec{L}_1^2}^2 + \frac{8|g|_{\mathrm{Lip}}^2\Delta t}{\varepsilon_5}\|\xi_{\vec{\theta}}^{n+1}\|_{\vec{H}_1^1}^2\|\vec{s}\|_{L^\infty(0,\mathcal{T};\vec{H}_1^1)}^2\\
&+ \frac{\varepsilon_5\Delta t}{2}\|\xi_{\vec{s}}\|_{\vec{L}_1^2}^2 + \frac{C|g|_{\mathrm{Lip}}^2 h^{2k}\Delta t}{\varepsilon_6}\|\vec{\theta}\|_{L^\infty(0,\mathcal{T};\vec{H}_1^{k+1})}^2\|\vec{s}\|_{L^\infty(0,\mathcal{T};\vec{H}_1^1)}^2\frac{\varepsilon_6\Delta t}{2}\|\xi_{\vec{s}}\|_{\vec{L}_1^2}^2.
\end{aligned}$$

Finally we insert $\pm 4\Delta t d_4(\mathcal{I}_h\,\vec{\theta}(t_{n+1}),\xi_{\vec{s}}^{n+1})$ in $\bar{I}_6$ and use Lemma 5.6 in order to deduce the bound

$$\begin{aligned}
|\bar{I}_6| &= \Big|4\Delta t\big(d_4\big(\vec{\theta}_h^{n+1} - \mathcal{I}_h\,\vec{\theta}(t_{n+1}),\xi_{\vec{s}}^{n+1}\big) + d_4\big(\mathcal{I}_h\,\vec{\theta}(t_{n+1}) - \vec{\theta}(t_{n+1}),\xi_{\vec{s}}^{n+1}\big)\big)\Big|\\
&\leq \frac{8|g|_{\mathrm{Lip}}^2\Delta t}{\varepsilon_7}\|\xi_{\vec{\theta}}^{n+1}\|_{\vec{L}_1^2}^2 + \frac{C|g|_{\mathrm{Lip}}^2 h^{2k}}{\varepsilon_8}\|\vec{\theta}\|_{L^\infty(0,\mathcal{T};\vec{H}_1^{k+1})}^2 + \frac{\varepsilon_7+\varepsilon_8}{2}\Delta t\|\xi_{\vec{s}}\|_{\vec{L}_1^2}^2.
\end{aligned}$$

It then suffices to take $\varepsilon_i = 3g_1/4$ for all $i \in \{1,\dots,10\}$ and to sum over $1 \leq n \leq m$, for all $m+1 \leq N$ in the above estimates, which, in combination with Theorem 5.6 implies that

$$\begin{aligned}
&\|\xi_{\vec{s}}^{m+1}\|_{\vec{L}_1^2}^2 + \|2\xi_{\vec{s}}^{m+1} - \xi_{\vec{s}}^{m}\|_{\vec{L}_1^2}^2 + \sum_{n=1}^{m}\|\Lambda\xi_{\vec{s}}^{n}\|_{\vec{L}_1^2}^2 + \sum_{n=1}^{m}\Delta t g_1\|\xi_{\vec{s}}^{n+1}\|_{\vec{L}_1^2}^2\\
&\leq C(\Delta t^4 + h^{2k}) + \frac{32|g|_{\mathrm{Lip}}^2}{3g_1}\Big(1 + \|\vec{s}\|_{L^\infty(0,\mathcal{T};\vec{H}_1^1)}^2\Big)\Delta t\sum_{n=1}^{m}\|\xi_{\vec{\theta}}^{n+1}\|_{\vec{L}_1^2}^2,
\end{aligned}$$

and the result follows by choosing

$$\gamma_2 = \frac{32|g|_{\mathrm{Lip}}^2}{3g_1}\big(1 + \|\vec{s}\|_{L^\infty(0,\mathcal{T};\vec{H}_1^1)}^2\big).$$

□

Theorem 5.9. *Under the same assumptions of Theorems 5.6 - 5.8, there exist positive constants $\hat{\gamma}_u$, $\hat{\gamma}_\theta$ and $\hat{\gamma}_s$ independent of Δt and h, such that for a sufficiently small Δt and all $m+1 \leq N$, the following inequalities hold:*

$$\left(\|\xi_{\boldsymbol{u}}^{m+1}\|^2_{\boldsymbol{L}_1^2} + \|2\xi_{\boldsymbol{u}}^{m+1} - \xi_{\boldsymbol{u}}^{m}\|^2_{\boldsymbol{L}_1^2} + \sum_{n=1}^{m}\big(\|\Lambda\xi_{\boldsymbol{u}}^{n}\|^2_{\boldsymbol{L}_1^2} + \Delta t\tilde{\alpha}_a\|\xi_{\boldsymbol{u}}^{n+1}\|^2_{1,\mathcal{T}_h}\big)\right)^{1/2}$$
$$\leq \hat{\gamma}_u(\Delta t^2 + h^k),$$
$$\left(\|\xi_{\vec{\theta}}^{m+1}\|^2_{\vec{L}_1^2} + \|2\xi_{\vec{\theta}}^{m+1} - \xi_{\vec{\theta}}^{m}\|^2_{\vec{L}_1^2} + \sum_{n=1}^{m}\big(\|\Lambda\xi_{\vec{\theta}}^{n}\|^2_{\vec{L}_1^2} + \Delta t\hat{\alpha}_a\|\xi_{\vec{\theta}}^{n+1}\|^2_{\vec{H}_1^1}\big)\right)^{1/2}$$
$$\leq \hat{\gamma}_\theta(\Delta t^2 + h^k),$$
$$\left(\|\xi_{\vec{s}}^{m+1}\|^2_{\vec{L}_1^2} + \|2\xi_{\vec{s}}^{m+1} - \xi_{\vec{s}}^{m}\|^2_{\vec{L}_1^2} + \sum_{n=1}^{m}\big(\|\Lambda\xi_{\vec{s}}^{n}\|^2_{\vec{L}_1^2} + \Delta t g_1\|\xi_{\vec{s}}^{n+1}\|^2_{\vec{L}_1^2}\big)\right)^{1/2}$$
$$\leq \hat{\gamma}_s(\Delta t^2 + h^k).$$

Proof. From Theorem 5.6 and 5.8 we have the estimates

$$\sum_{n=1}^{m}\gamma_u\Delta t\|\xi_{\boldsymbol{u}}^{n+1}\|^2_{1,\mathcal{T}_h} \leq C(\Delta t^4 + h^{2k}) + \frac{\gamma_1\gamma_u}{\tilde{\alpha}_a}\sum_{n=1}^{m}\Delta t\|\xi_{\vec{\theta}}^{n+1}\|^2_{\vec{L}_1^2},$$
$$\sum_{n=1}^{m}\gamma_s\Delta t\|\xi_{\vec{s}}^{n+1}\|^2_{\vec{L}_1^2} \leq C(\Delta t^4 + h^{2k}) + \frac{\gamma_s\gamma_2}{g_1}\sum_{n=1}^{m}\Delta t\|\xi_{\vec{\theta}}^{n+1}\|^2_{\vec{L}_1^2},$$

which, after substituting them back into Theorem5.7, yield

$$\|\xi_{\vec{\theta}}^{m+1}\|^2_{\vec{L}_1^2} + \|2\xi_{\vec{\theta}}^{m+1} - \xi_{\vec{\theta}}^{m}\|^2_{\vec{L}_1^2} + \sum_{n=1}^{m}\|\Lambda\xi_{\vec{\theta}}^{n}\|^2_{\vec{L}_1^2} + \sum_{n=1}^{m}\Delta t\hat{\alpha}_a\|\xi_{\vec{\theta}}^{n+1}\|^2_{\vec{H}_1^1}$$
$$\leq C(\Delta t^4 + h^{2k}) + \frac{\gamma_1\gamma_u g_1 + \gamma_s\gamma_2\tilde{\alpha}_a}{\tilde{\alpha}_a g_1}\sum_{n=1}^{m}\Delta t\|\xi_{\vec{\theta}}^{n+1}\|^2_{\vec{L}_1^2}.$$

For the last term on the right-hand side of this last bound we have

$$\|\xi_{\vec{\theta}}^{m+1}\|^2_{\vec{L}_1^2} \leq 2(\|\Lambda\xi_{\vec{\theta}}^{m}\|^2_{\vec{L}_1^2} + \|2\xi_{\vec{\theta}}^{m} - \xi_{\vec{\theta}}^{m-1}\|^2_{\vec{L}_1^2}),$$

and considering Δt sufficiently small and applying Gronwall's lemma, we readily infer the estimate

$$\|\xi_{\vec{\theta}}^{m+1}\|^2_{\vec{L}_1^2} + \|2\xi_{\vec{\theta}}^{m+1} - \xi_{\vec{\theta}}^{m}\|^2_{\vec{L}_1^2} + \sum_{n=1}^{m}\big(\|\Lambda\xi_{\vec{\theta}}^{n+1}\|^2_{\vec{L}_1^2} + \Delta t\hat{\alpha}_a\|\xi_{\vec{\theta}}^{n+1}\|^2_{\vec{H}_1^1}\big)$$
$$\leq C(\Delta t^4 + h^{2k}). \tag{5.48}$$

The first and third bounds follow by combining (5.48) and Theorems 5.6 and 5.8. □

Lemma 5.8. *Under the same assumptions of Theorem 5.9, we have*

$$\left(\sum_{n=1}^{m}\Delta t\|p(t_{n+1}) - p_h^{n+1}\|^2_{L_1^2}\right)^{1/2} \leq \hat{\gamma}_p(\Delta t^2 + h^k).$$

Proof. Owing to the inf-sup condition (5.23), there exists $\boldsymbol{w}_h \in \boldsymbol{Z}_h^{\perp}$ such that

$$b(\boldsymbol{w}_h, p(t_{n+1}) - p_h^{n+1}) = \|p(t_{n+1}) - p_h^{n+1}\|_{L_1^2}^2, \tag{5.49}$$

$$\|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} \leq \frac{1}{\tilde{\beta}} \|p(t_{n+1}) - p_h^{n+1}\|_{L_1^2}. \tag{5.50}$$

From (5.25) and Lemma 5.7, proceeding as in the proof of Theorem 5.6, we obtain

$$\begin{aligned}
&\Delta t b(\boldsymbol{w}_h, p(t_{n+1}) - p_h^{n+1}) \\
&= -\Delta t \left(\boldsymbol{u}'(t_{n+1}) - \frac{3\boldsymbol{u}_h^{n+1} - 4\boldsymbol{u}_h^n + \boldsymbol{u}_h^{n-1}}{2\Delta t}, \boldsymbol{w}_h \right)_{1,\Omega_{\mathrm{a}}} + \Delta t a_1^h(\boldsymbol{u}_h^{n+1} - \boldsymbol{u}(t_{n+1}), \boldsymbol{w}_h) \\
&\quad + \Delta t \big(c_1^h(\boldsymbol{u}_h^{n+1}; \boldsymbol{u}_h^{n+1}, \boldsymbol{w}_h) - c_1^h(\boldsymbol{u}(t_{n+1}); \boldsymbol{u}(t_{n+1}), \boldsymbol{w}_h) \big) \\
&\quad + \Delta t d_1(\vec{\theta}(t_{n+1}) - \vec{\theta}_h^{n+1}, \boldsymbol{w}_h) \\
&\leq \frac{\Delta t^2}{2\sqrt{3}} \|\boldsymbol{u}^{(3)}\|_{L^2(t_{n-1},t_{n+1},\boldsymbol{L}_1^2)} \sqrt{\Delta t} \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} + \tilde{C}_a C^* h^k \Delta t \|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T};\boldsymbol{H}_1^{k+1})} \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} \\
&\quad + \tilde{C}_a \Delta t \|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h} \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} + C^* \tilde{C}_c \Delta t \|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T};\boldsymbol{H}_1^1)} \|\xi_{\boldsymbol{u}}\|_{1,\mathcal{T}_h} \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} \\
&\quad + 2\Delta t C \tilde{C}_c h^k \|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T};\boldsymbol{H}_1^1)} \|\boldsymbol{u}\|_{L^\infty(0,\mathcal{T};\boldsymbol{H}_1^{k+1})} \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} \\
&\quad + C_{\boldsymbol{F}} \Delta t h^k C^* \|\vec{\theta}\|_{L^\infty(0,\mathcal{T};\vec{H}_1^{k+1})} \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h} + C_{\boldsymbol{F}} \Delta t \|\xi_{\vec{\theta}}\|_{\vec{L}_1^2} \|\boldsymbol{w}_h\|_{1,\mathcal{T}_h}.
\end{aligned}$$

Summing over $1 \leq n \leq m$ for all $m+1 \leq N$ and substituting back into equations (5.49) and (5.50), we obtain

$$\begin{aligned}
&\left(\sum_{n=1}^m \Delta t \|p(t_{n+1}) - p_h^{n+1}\|_{L_1^2}^2 \right)^{1/2} \\
&\leq \frac{C}{\tilde{\beta}} \left(\Delta t^2 + h^k + \left(\sum_{n=1}^m \Delta t \|\xi_{\vec{\theta}}^{n+1}\|_{\vec{L}_1^2}^2 \right)^{1/2} + \left(\sum_{n=1}^m \Delta t \|\xi_{\boldsymbol{u}}^{n+1}\|_{1,\mathcal{T}_h}^2 \right)^{1/2} \right),
\end{aligned}$$

and the desired result readily follows from Theorem 5.9. □

5.6 Numerical tests

5.6.1 Example 5.1: accuracy tests

In our first computational test we examine the convergence of the Galerkin method (5.24), taking as computational domain the square $\Omega = (0,1)^2$. We take the parameter values $\nu = 0.1$, $k^+(\boldsymbol{x}) = 1$, $\boldsymbol{g} = (0,-1)^T$, $\mathbb{K}^{-1} = \mathbb{I}$, $\mathbb{D} = 10^{-3}\mathbb{I}$, $D_s = 1$, $\rho_{\mathrm{f}} = \phi = 1$, $\rho_{\mathrm{b}} = 0.1$, $a_0 = 500 \times 10^k$, where k is the polynomial degree. Following the approach of manufactured solutions, we prescribe boundary data and additional external forces and adequate source terms so that the closed-form solutions to (5.1), (5.3) are given by the smooth functions

k	DoF	$\mathtt{e}_{\boldsymbol{u}}$	rate	$\mathtt{e}_p$	rate	$\mathtt{e}_{\vec{\theta}}$	rate	$\mathtt{e}_s$	rate
1	75	0.05435	–	0.57400	–	0.26530	–	0.11760	–
	259	0.02894	0.909	0.12480	2.201	0.13940	0.928	0.05934	0.986
	963	0.01466	0.981	0.05242	1.252	0.07039	0.986	0.02978	0.995
	3715	0.00736	0.995	0.02545	1.042	0.03537	0.993	0.01490	0.999
	14595	0.00368	0.998	0.01202	1.083	0.01792	0.981	0.00746	0.999
2	195	0.00537	–	0.77890	–	0.00071	–	0.05373	–
	715	0.00149	1.848	0.11910	2.710	0.00018	1.947	0.01480	1.860
	2739	0.00038	1.953	0.01749	2.767	4.619e-5	2.001	0.00378	1.970
	10723	9.074e-5	2.084	0.00249	2.813	1.154e-5	2.001	0.00095	1.992
	42435	2.328e-5	1.963	0.00052	2.256	2.909e-6	1.988	0.00024	1.998

Table 5.1: Example 5.1 (Spatial accuracy test): experimental errors and convergence rates for the approximate solutions $\boldsymbol{u}_h$, p_h, $\vec{\theta}_h$ and s_h. Values are displayed for schemes with first- and second-order in space (table produced by the author).

$$\boldsymbol{u}(r,z,t) = \begin{pmatrix} 0 \\ -\cos(r\pi/2)\exp(-t) \end{pmatrix}, \quad \vec{\theta}(r,z,t) = \begin{pmatrix} z^2r^2(3-2r)(1-\exp(-t)) \\ z^2r^2(3-2r)(1-\exp(-t)) \end{pmatrix},$$
$$p(r,z,t) = (r^3-2z^4)\sin(t), \quad \vec{s}(r,z,t) = \begin{pmatrix} 1-\exp(-z^2r^2(3-2r)(t+\exp(t))) \\ 1-\exp(-z^2r^2(3-2r)(t+\exp(t))) \end{pmatrix}.$$

As $\boldsymbol{u}$ is prescribed everywhere on $\partial\Omega_{\mathrm{a}}$, for sake of uniqueness we impose $p \in L^2_{0,1}(\Omega_{\mathrm{a}})$ through a Lagrange multiplier approach. Also note that the exact solutions satisfy the boundary conditions (5.5a), (5.5b), (5.5c) on the inlet, wall, and symmetry axis, respectively, whereas instead of (5.5d) we set

$$\boldsymbol{u} = \boldsymbol{u}_{\mathrm{out}}, \quad \mathbb{D}\nabla\vec{\theta}\cdot\boldsymbol{n} = \vec{0},$$

on the outlet $\Gamma_{\mathrm{a}}^{\mathrm{out}} \times (0,\mathcal{T}]$. The accuracy of the spatial semi-discretisation is tested by considering a sequence of uniformly refined meshes $\{\mathcal{T}_{h,l}\}_l$ of mesh size $h_l = 2^{-l}\sqrt{2}$, and fixing $\mathcal{T} = 0.005$ with $\Delta t = 0.001$. Relative errors in their natural norms, along with the corresponding convergence rates are computed as

$$\mathtt{e}_{\boldsymbol{u}} = \frac{\|\boldsymbol{u}-\boldsymbol{u}_h\|_{1,\mathcal{T}_h}}{\|\boldsymbol{u}\|_{1,\mathcal{T}_h}}, \quad \mathtt{e}_p = \frac{\|p-p_h\|_{L^2_1(\Omega_{\mathrm{a}})}}{\|p\|_{L^2_1(\Omega_{\mathrm{a}})}}, \quad \mathtt{e}_{\vec{\theta}} = \frac{\|\vec{\theta}-\vec{\theta}_h\|_{\vec{H}^1_1(\Omega_{\mathrm{a}})}}{\|\vec{\theta}\|_{\vec{H}^1_1(\Omega_{\mathrm{a}})}},$$
$$\mathtt{e}_{\vec{s}} = \frac{\|\vec{s}-\vec{s}_h\|_{\vec{H}^1_1(\Omega_{\mathrm{a}})}}{\|\vec{s}\|_{\vec{H}^1_1(\Omega_{\mathrm{a}})}}, \quad \mathtt{rate} = \log(e_{(\cdot)}/\tilde{e}_{(\cdot)})[\log(h/\tilde{h})]^{-1},$$

where $e, \tilde{e}$ denote errors generated on two consecutive meshes of sizes h and $\tilde{h}$, respectively. These quantities are listed in Table 5.1 for $k=0$ and $k=1$, and they indicate optimal error decay in the light of Theorem 5.9.

Δt	$\hat{\mathtt{e}}_{\boldsymbol{u}}$	rate	$\hat{\mathtt{e}}_p$	rate	$\hat{\mathtt{e}}_{\vec{\theta}}$	rate	$\hat{\mathtt{e}}_s$	rate
2.5	0.5496	–	0.5663	–	17.691	–	0.6738	–
1.25	0.1408	1.964	0.1177	2.266	3.2720	2.435	0.1673	2.009
0.625	0.0289	2.284	0.0258	2.188	0.6621	2.305	0.0409	2.032
0.3125	0.0066	2.119	0.0061	2.091	0.1519	2.124	0.0105	1.965
0.1562	0.0016	2.047	0.0015	1.976	0.0366	2.054	0.0027	1.934

Table 5.2: Example 5.1 (time accuracy test): experimental errors and convergence rates for the approximate solutions $\boldsymbol{u}_h$, p_h, $\vec{\theta}_h$ and s_h, computed for each refinement level (table produced by the author).

Regarding the convergence of the time advancing scheme, now we set $\mathcal{T} = 5$ and consider a sequence of uniform refined time partitions $\tau_l, l \in \{1, 2, 3, 4, 5\}$ where the time step is $5/2^l$. Absolute errors are computed as

$$\hat{\mathtt{e}}_{\boldsymbol{u}} = \left(\sum_{n=1}^{m} \Delta t \|\boldsymbol{u}(t_{n+1}) - \boldsymbol{u}_h^{n+1}\|_{1,\mathcal{T}_h}^2 \right)^{1/2}, \quad \hat{\mathtt{e}}_p = \left(\sum_{n=1}^{m} \Delta t \|p(t_{n+1}) - p_h^{n+1}\|_{L_1^2}^2 \right)^{1/2},$$
$$\hat{\mathtt{e}}_{\vec{\theta}} = \left(\sum_{n=1}^{m} \Delta t \|\vec{\theta}(t_{n+1}) - \vec{\theta}_h^{n+1}\|_{\vec{H}_1^1}^2 \right)^{1/2}, \quad \hat{\mathtt{e}}_{\vec{s}} = \left(\sum_{n=1}^{m} \Delta t \|\vec{s}(t_{n+1}) - \vec{s}_h^{n+1}\|_{\vec{L}_1^2}^2 \right)^{1/2},$$

and we readily observe from Table 5.2 that the method converges to the exact solution with the expected second-order rate.

5.6.2 Example 5.2: validation against experimental data

Now we define a different adimensionalisation of (5.1a)-(5.3d) that follows the recent model (tailored specifically for soil-based water filters for arsenic removal) proposed in [134]. This problem considers only one type of contaminant and only one type of adsorption. Defining as $L, v_i, \theta_0, s_{\max}$ the representative length of the column, the linear inflow rate, initial solids concentration, and maximum adsorption, respectively; we define dimensionless variables as

$$\bar{r} = \frac{r}{L}, \ \bar{z} = \frac{z}{L}, \ \bar{\boldsymbol{u}} = \frac{\boldsymbol{u}}{v_i}, \ \bar{\theta} = \frac{\theta}{\theta_0}, \ \bar{p} = \frac{L(p - p_{\text{atm}})}{\mu v_i}, \ \bar{s} = \frac{s}{s_{\max}}, \ \bar{t} = k^+ \theta_0 t,$$

and we also define the constants

$$\text{Re} = \frac{\rho_{\text{f}} v_i L}{\nu}, \quad \text{Pe} = \frac{v_i L}{D}, \quad \text{Da} = \frac{\kappa}{L^2}, \quad \alpha = \frac{\rho_{\text{b}} s_{\max}}{\theta_0}, \quad \beta = \frac{k^+ L^2 \theta_0}{D}. \tag{5.51}$$

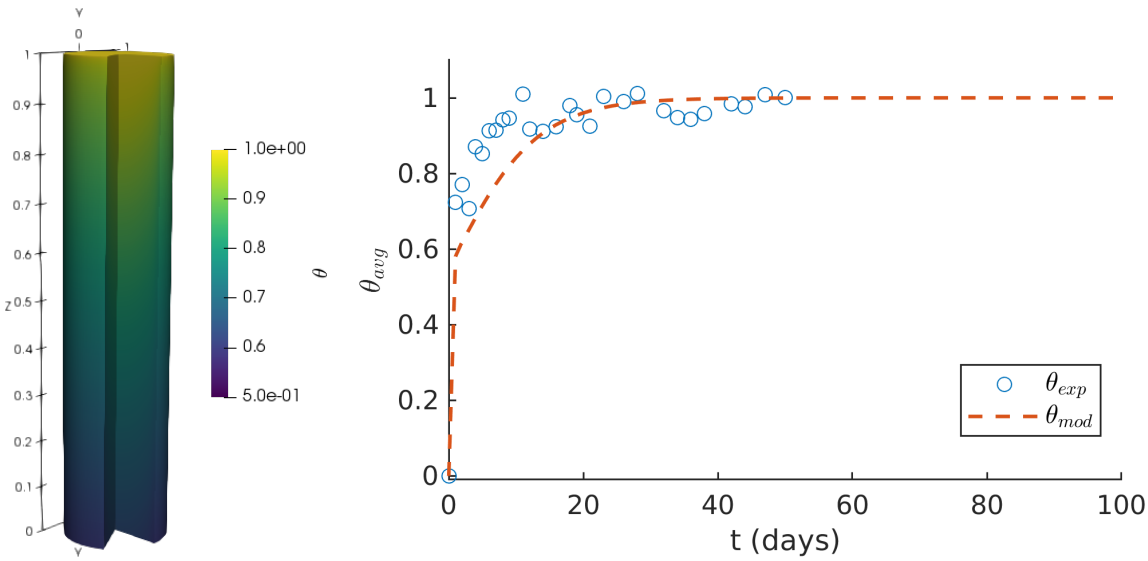


Figure 5.2: Example 5.2 (validation against experimental data): contaminant concentration after one day (left). Value of $\theta|_{\text{avg}}(t)$ (experimental observation from [134] and numerical simulation) using raw laterite as the adsorbent (right) (figure produced by the author).

Making abuse of notation, the problem defined in $\Omega_{\text{a}} \times (0, \mathcal{T}]$ adopts the form

$$\begin{aligned}
\frac{\beta\,\text{Re}}{\text{Pe}}\partial_t \boldsymbol{u} + \text{Re}\,\boldsymbol{u}\cdot\nabla\boldsymbol{u} + \frac{1}{\text{Da}}\boldsymbol{u} - \frac{1}{\phi}\mathbf{div}_{\text{a}}(\boldsymbol{\varepsilon}(\boldsymbol{u})) + \nabla p + \frac{1}{\phi}(\boldsymbol{u}_r/r^2)\boldsymbol{e}_1 &= \mathbf{0},\\
\text{div}_{\text{a}}\,\boldsymbol{u} &= 0,\\
\frac{\phi\beta}{\text{Pe}}\partial_t\theta - \frac{1}{\text{Pe}}\,\text{div}_{\text{a}}(\nabla\theta) + \boldsymbol{u}\cdot\nabla\theta &= -\frac{\alpha\beta}{\text{Pe}}\partial_t s,\\
\partial_t s &= \theta(1-s).
\end{aligned}$$

The setup consists of a lab-scale filter (a column of height 1 and radius $\bar{R} = 0.11$, already in dimensionless units) where one varies the feed flow rate, the arsenic concentration at the feed, and also the bed height. Gravitational effects are not considered, and the boundary and initial conditions are precisely as in (5.5a)-(5.5e). The configuration of the system implies that the non-dimensional constants from (5.51) assume the values

$$\text{Re} = 68.1, \quad \text{Pe} = 1.11\times 10^5, \quad \text{Da} = 8000, \quad \alpha = 248, \quad \beta = 136,$$

and the remaining parameter values are $\phi = 0.48$, $\boldsymbol{u}^{\text{in}}(r,z) = (0, \frac{1}{\bar{R}^2}(r-\bar{R})(r+\bar{R}))^{\mathsf{t}}$, $\theta^{\text{in}} = 1$. We employ a structured mesh of 8000 triangular elements and define a constant time step of $\Delta t = 0.15$ (adimensional time $t = 0.15 \approx 1$ day).

During the filtration process the soil-based bed reaches a point in time where it is no longer adequate for adsorption. This phenomenon can be observed in Figure 5.2 where we plot the evolution of the average concentration of the contaminant θ on the outlet, that is

$$\theta_{\text{avg}}(t) = \frac{2}{\bar{R}^2}\int_{\Gamma_{\text{a}}^{\text{out}}} \theta\, r\, \text{d}s.$$

We also compared the predictions of the model with experimental data, collected for a filter that uses raw laterite as an adsorbent medium, and to which an arsenic solution is injected in its upper part [134]. The qualitative results displayed on figure 5.2 seem to show an acceptable adjustment to the experimental data. This suggest that the model and the axisymmetric divergence-conforming scheme can be used effectively as a tool to study the behaviour of the filtration process under similar flow regimes.

5.6.3 Example 5.3: Two contaminants in two-layer filter

We model a filter with two contaminants and two layers. The domain has a R/L ratio of 0.22. While the inlet is the top wall, the outlet is the region $\{(z,r)|z=0 \text{ and } 0 \leq r \leq 0.25R\}$. For (5.4) we take (5.2) with $m=2$ and and we consider $\mu = 8.94\times10^{-4}\,\mathrm{Pa\,s}$, $v_i = 6.0\times10^{-3}\,\mathrm{m/s}$, $\rho_\mathrm{f} = 10^3\,\mathrm{Kg/m^3}$, $\theta_1^\mathrm{in} = 8.0\times10^{-5}\,\mathrm{Kg/m^3}$, $\theta_2^\mathrm{in} = 2.0\times10^{-5}\,\mathrm{Kg/m^3}$, $s_1^\mathrm{max} = 10^{-3}\,\mathrm{Kg/Kg}$, $s_2^\mathrm{max} = 10^{-2}\,\mathrm{Kg/Kg}$. In addition, the rheology of the grains is different in the top and bottom halves of the domain. More precisely, we have

$$
\begin{aligned}
&D_\mathrm{top} = 3.8\times10^{-11}\,\mathrm{m^2/s}, \quad D_\mathrm{bot} = 7.6\times10^{-12}\,\mathrm{m^2/s}, \quad \phi_\mathrm{top} = 0.32, \quad \phi_\mathrm{bot} = 0.28,\\
&\rho_\mathrm{b,top} = 1050\,\mathrm{Kg/m^3}, \quad \rho_\mathrm{b,bot} = 1100\,\mathrm{Kg/m^3}, \quad k^+_{1,\mathrm{top}} = 5.0\times10^{-3}\,\mathrm{m^3/(Kg\,s)},\\
&k^+_{2,\mathrm{top}} = 0\,\mathrm{m^3/(Kg\,s)}, \quad k^+_{1,\mathrm{bot}} = 2.5\times10^{-4}\,\mathrm{m^3/(Kg\,s)}, \quad k^+_{2,\mathrm{bot}} = 10^{-3}\,\mathrm{m^3/(Kg\,s)},
\end{aligned}
$$

and the permeability $\mathbb{K}(\boldsymbol{x}) = \kappa(\boldsymbol{x})\boldsymbol{I}$ has a log-uniform distribution in each layer that satisfies

$$
\begin{aligned}
1.57\times10^{-9}\,\mathrm{m^2} \leq \kappa_\mathrm{top}(\boldsymbol{x}) \leq 3.04\times10^{-6}\,\mathrm{m^2},\\
5.18\times10^{-10}\,\mathrm{m^2} \leq \kappa_\mathrm{bot}(\boldsymbol{x}) \leq 10^{-6}\,\mathrm{m^2}.
\end{aligned}
$$

Qualitative results for the concentration of the two contaminants at times $t = 10, 100$ and 300 are shown on Figure 5.3. As expected, most of the first contaminant is retained in the upper layer, whereas the second one passes the first layer to begin to be retained in the lower layer.

Now we change values to $s_1^\mathrm{max} = 10^{-7}\,\mathrm{Kg/Kg}$ and $s_2^\mathrm{max} = 10^{-6}\,\mathrm{Kg/Kg}$ and run the simulation for a longer time to assess how the swapping the order of layers and the geometry affect the contaminant removal, measured by $\theta_\mathrm{avg}(t)$. For the first two tests we use the same cylinder, altering only the order of the layers. As we can see from the top panels of Figure 5.4, reversing the order of the layers softens the transition towards saturation, but the most important behaviour is reached essentially at the same time in both cases. We also test with a truncated cone (see dimensions in the bottom left panel of Figure 5.4). The saturation is now achieved in a much shorter time, which could be explained by a combined effect of volume reduction (and therefore of adsorbent mass), and faster flow patterns that decrease the retention time and thus the adsorption of the system.

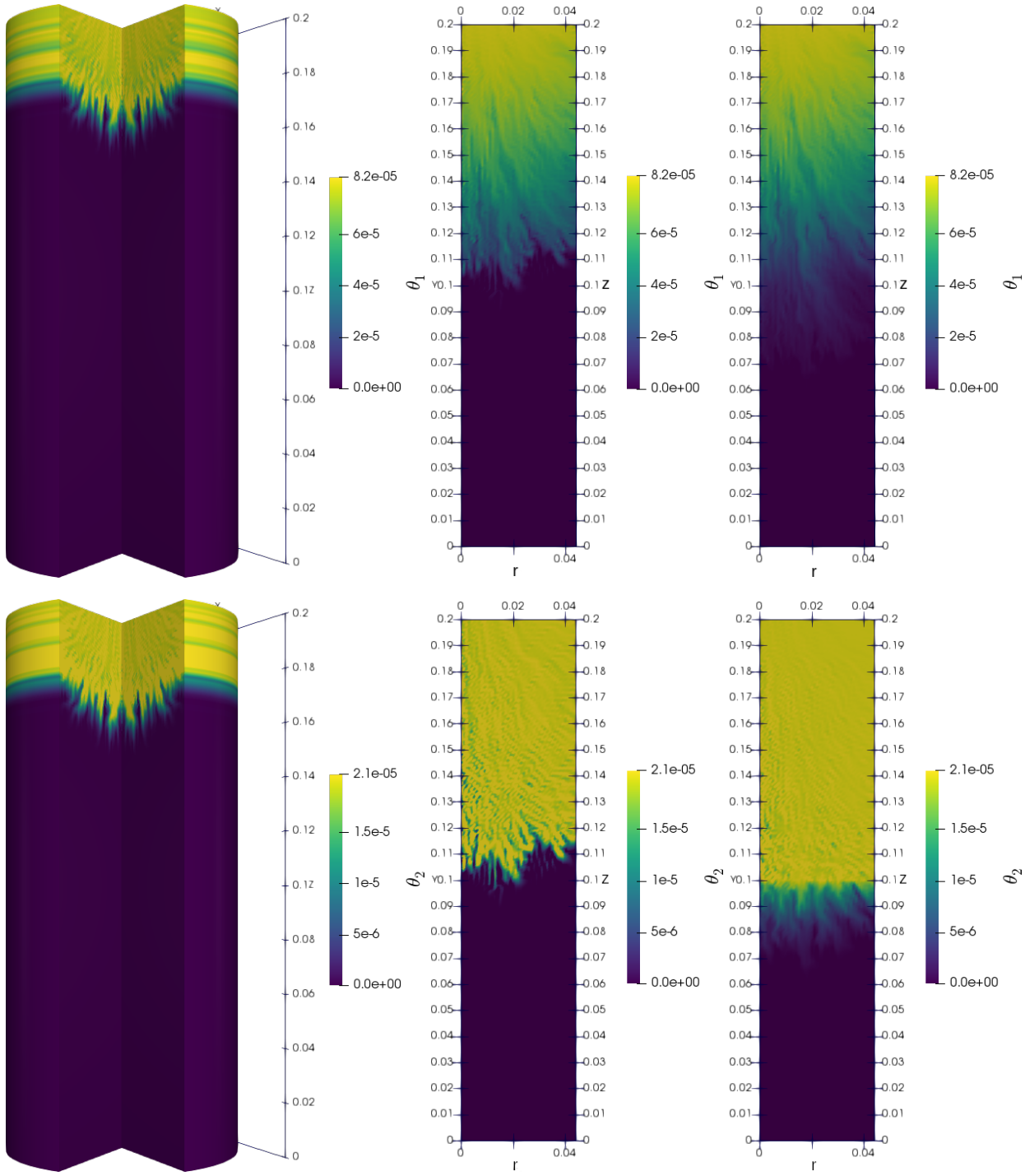


Figure 5.3: Example 5.3 (two contaminants in two-layer filter): concentration of contaminants at times $t = 10, 100, 300$ (figure produced by the author).

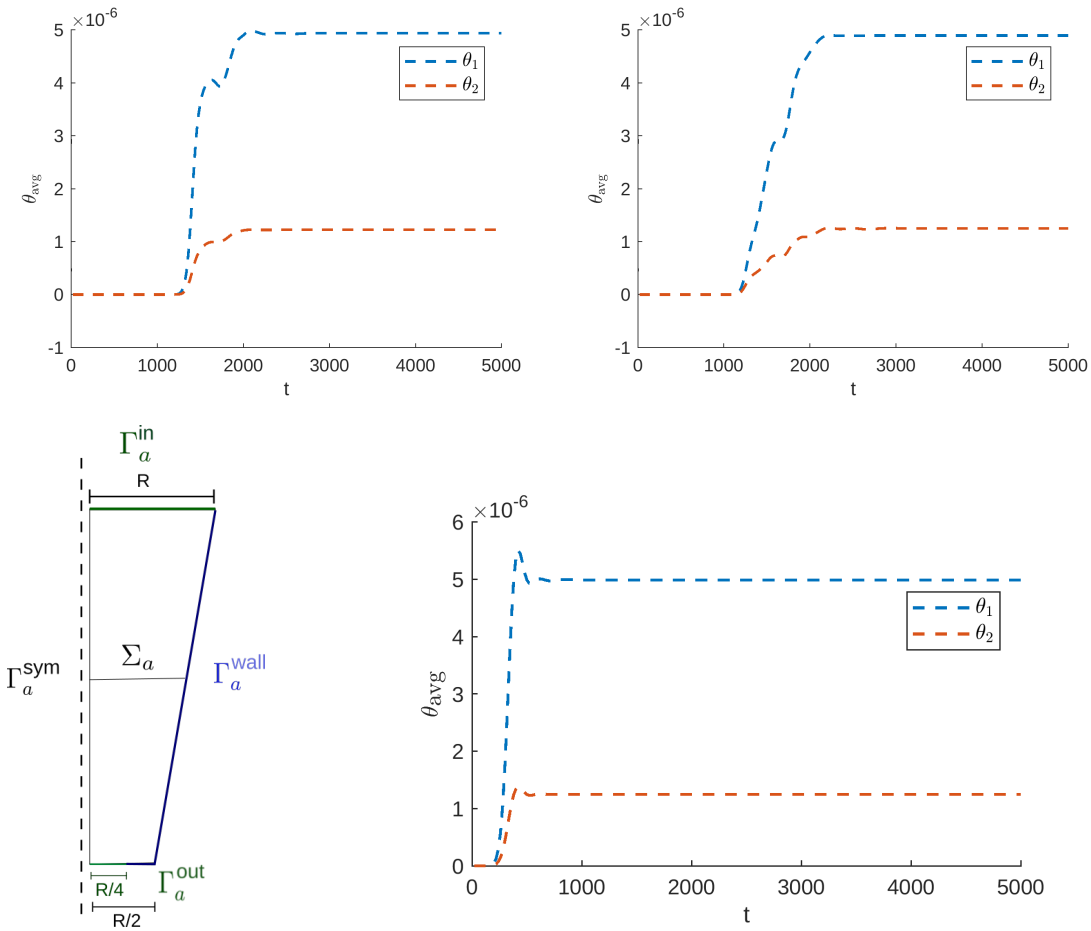


Figure 5.4: Example 5.3 (two contaminants in two-layer filter): concentration of contaminants $\theta_{\mathrm{avg},i}(t)$ using a cylinder and changing order of layers (top); and similar computation using a truncated cone (bottom) (figure produced by the author).

CHAPTER 6

Conclusions, current and future works

6.1 Conclusions

In this thesis, we have proposed and analysed PDE-based models for the coupling of flow equations and transport; we have proved their solvability using fixed-point theories and we have also proposed accurate, robust and reliable methods for the discretisation of these equations, with special emphasis in $\boldsymbol{H}(\mathrm{div})$-conforming formulations for the flow equations, whereas for the transport problem (resulting in a scalar or vectorial advection-diffusion equation) we have studied entropy stable schemes for stand-alone problems as well as finite element primal formulations when coupled with the flow equations. Furthermore, in each chapter, we have studied one or more modelling problems with engineering applications.

In **Chapter** 2 we studied entropy stable schemes for the numerical solution of initial value problems of nonlinear, possibly strongly degenerate systems of convection-diffusion equations proposed in [99]. As a new contribution, we demonstrated, firstly, that these schemes can naturally be extended to initial-boundary value problems with zero-flux boundary conditions in one space dimension, including an explicit bound on the growth of the total entropy. Secondly, it was shown that these assumptions are satisfied by certain diffusively corrected multiclass kinematic flow models of arbitrary size that describe traffic flow or the settling of dispersions and emulsions.

Numerical examples illustrate the behavior and accuracy of entropy stable schemes for these applications. They also confirm the theoretical bounds for entropy in both cases, zero-flux boundary conditions and periodic boundary conditions. Furthermore, the results of Examples 2.2, 2.3 and 2.4 demonstrate that entropy stable schemes have a competitive computational efficiency compared with other common numerical schemes, when used on diffusively corrected multiclass kinematic flow models, like the traffic model and the polydisperse sedimentation model presented here. In fact, Tables 2.2, 2.3 and 2.4, favor the ES scheme in terms of CPU time, although the CFL condition (2.39) was the same for all schemes with the same value of C_{CFL} in each case. A probable heuristic explanation of this observation lies in the difference of the computation of the numerical flux in each case, which involve for instance in the case of the KT scheme the calculation of slopes, evaluation of the minmod limiter function, and other operations not present in other schemes. Results were favorable for the ES schemes also for Examples 2.5 and 2.6, for which CPU times are not reported herein.

Although errors and errors rates are comparable with the other tested methods (Kurganov-Tadmor and component-wise Global Lax-Friedrichs) for coarser cell partitions, because of the differences on finer cell partitions we cannot entirely confirm that the methods converge to the same solution. This shortcoming is exacerbated by the lack of a well-posedness theory for (2.1) in the strongly degenerate case. It is therefore a topic requiring more careful study in future research.

In **Chapter** 3 we proposed a divergence-conforming finite element method for the double diffusive problem, considering temperature-dependent viscosity and possible cross-diffusion terms subject to the restriction of maintaining the coercivity of the diffusion operator. The formulation includes the Navier-Stokes/Brinkman flow description, which makes this model suitable for the study of flow in saturated porous media and interfaces between porous media and free flow. The numerical scheme is based on $\boldsymbol{H}(\mathrm{div})$-conforming BDM elements of order k for the velocity, discontinuous elements of order $k-1$ for the pressure, and Lagrangian finite elements of order k for temperature and the concentration of a solute. The main differences between the available well-posedness results and analysis of $\boldsymbol{H}(\mathrm{div})$-conforming methods for classical Boussinesq equations and the double-diffusive equations (3.1) are, of course, caused by the vector-valued nature of the quantities (the components of $\vec{m}$) that diffuse in (3.1) while in the classical Boussinesq formulation there is only one scalar diffusive quantity (for instance, solely temperature). Some of the arguments related to the well-posedness analysis of the continuous problem, in particular those related to handling non-homogenous Dirichlet data by a lifting argument [125, 137], carry over almost verbatim from the scalar to the vectorial case. However, the bilinear form associated with the term $-\mathbf{div}(\mathbb{D}\nabla\vec{m})$ must be coercive so that stability is ensured. This requirement, in turn, imposes restrictions on the choice of the diffusion matrix $\mathbb{D}$; this matrix must be positive definite (though not necessarily symmetric). These properties are essential for the proof of existence of a discrete solution.

Regarding our computational tests, it is worth mentioning, that the results for different thermal Rayleigh values computed in Example 3.2 and summarised on the right panel of Table 3.4, were quite close to the results published by [59, 86]. For $\mathrm{Ra} \leq 1000$, the values of Nu and Sh were within a relative error of 3%, while for the last value $\mathrm{Ra} = 2000$, within 6%, which confirms that the proposed scheme gives results comparable to other known methods for a large range of parameter values, with the benefit that this formulation produces exactly divergence-free velocity approximations, which are of particular importance in ensuring that solutions to the flow equations remain locally conservative as well as energy stable (see e.g. [62]). Moreover, the error estimates of velocity can be derived in a pressure-robust manner (see [100]) which can be seen on the results presented in our accuracy test (see table 3.4). Another consequence of local conservation is that the coupled systems (in the present case, of temperature and reactive concentrations) can be written, at the discrete level, in exact divergence form. It is also interesting to note, that even if solvability of the coupled problem cannot be guaranteed if $\mathbb{D}$ is not positive definite, the convergence of the Newton iterations in Example 3.2 was observed for a broad range of parameters ($\mathrm{Sr}, \mathrm{Pr} \in [10^{-3}, 10^{3}]$, $N \in [1, 10]$, $\mathrm{Da} \in [10^{-7}, 1]$, $\mathrm{Ra} \in [100, 2000]$). This suggests that it may be interesting to study whether the stability analysis can be improved, to include a wider spectrum of values for the diffusion matrix.

In **Chapter** 4 we have advanced a model for the process of clarification and thickener in cylindrical units in the presence of a spinning rake structure. The model is intrinsically 3D, it incorporates a detailed flow-sedimentation coupling in the settling mixture and it considers a simplified, one-way

coupling that only imposes the velocity of the rotating arm which affects both the transport of solid particles and the revolving flow near the bottom of the tank. The novelty of the treatment consisted in the inclusion of terms that account for the influence of the rake motion on the momentum balance and the removal of solids. We also adapted techniques of the immersed boundary finite element method (see e.g. [34]) for the analysis and numerical approximation of those terms. This addition constitutes an important generalisation over existing models for sedimentation-consolidation processes reviewed in e.g. [42]. The numerical method we used is based on H(div)-conforming finite element methods for the flow and classical Lagrange elements for the solids concentration. A monolithic Newton scheme with exact Jacobian was employed in all cases, and we generated several tests to confirm the accuracy of the method and analysed several cases relevant to the process of clarification. Based on the results of this set of simulations, we could identify the solids removal coefficient α as the most sensitive factor on the outputs of overflow concentration and overflow flow rate. On the other hand, the combined contributions from drag and density did not seem to have a large effect on these markers. A further inspection revealed that the effects are not necessarily localised but they differ over the height of the device. Although several interesting extensions regarding the rake mechanics were left to be explored (some of which will be mentioned in the section of future works), we hope that this study helps in gaining a fuller understanding of the operating conditions in clarifier units.

In **Chapter** 5 we have advanced a second-order divergence-conforming discretisation for the system of partial differential equations modeling soil-based water filtering devices. Specifically, we introduced an axisymmetric $\boldsymbol{H}(\mathrm{div})$-conforming method based on two-dimensional BDM spaces [40] combined with an implicit, second-order backward differentiation formula for time discretisation. Based on discrete stability properties, we proved that the discrete problem has at least one solution. Furthermore, we derived an optimal a priori error estimate for the numerical scheme, where the main difficulty is the fully discrete analysis verifying that each of the terms is bounded optimally in the corresponding weighted spaces. Results of our accuracy test in the first numerical example, confirm that the method converges to the exact solution with the expected second-order rate. We also compared the predictions of the model with experimental data, collected for a filter that uses raw laterite as an adsorbent medium, and to which an arsenic solution is injected in its upper part [134]. The qualitative results displayed on figure 5.2 seem to show an acceptable adjustment to the experimental data. This suggest that the model and the axisymmetric divergence-conforming scheme can be used effectively as a tool to study the behaviour of the filtration process under similar flow regimes.

6.2 Ongoing research

Our current and ongoing investigation is an extension of the models presented through this thesis, to fluid-structure interactions with biological applications. Firstly, we intend to develop a model to quantify the effect of platelet count, shear rate and injury size on the initiation of blood coagulation under venous flow conditions.

Blood coagulation is a complex process that leads to thrombus formation inside blood vessels. It is initiated by the damage of the endothelial tissue at the internal surface of blood vessel walls. As a result, tissue factor (TF), that is normally isolated from blood plasma, is bared and forms a complex with factor VII. This complex activates factors IX and X that initiate the coagulation cascade. The

formation of thrombin accelerates the reaction of conversion of fibrinogen into fibrin. The latter forms fibrin polymer which constitutes the clot together with platelets. Red blood cells are trapped inside the clot and further reinforce it. Once initiated by activated factors IX and X, thrombin production is self-sustained through the positive feedback loops of the coagulation cascade [39]. Further, blood vessels are embedded in a biological tissue and, during the flow of blood, pressure is applied to the internal surfaces producing deformation of the vessel walls.

In recent years mathematical modelling has provided an important tool for the qualitative understanding of the underlying mechanisms behind thrombus growth [39] and several approaches have been proposed for the mathematical modelling of blood coagulation (see [13, 39, 95] and references therein). However, quantifying the effects of these mechanisms has been far more complicated, not only for the complexity of the coagulation factors interaction, but also for the disparities in the values of blood coagulation kinetic constants reported in literature, and the difficulties in performing *in vitro* experiments of thrombus growth, needed for the validation of computational models [79].

We devote our current work to extend the mathematical model of clot growth dynamics proposed in [39], by taking into account fluid-structure interactions. A classical approach to this problem has been the generation of a single mesh with a fitted interface between fluid and vessel wall. Usually, for the structure problem a Lagrangian formulation is considered, while for the fluid one an arbitrary Lagrangian Eulerian (ALE) approach is employed to account for the movement of the interface [128]. A different approach considers unfitted meshes, where the fluid mesh is fixed and the structure one is free to move independently. Within these methods we could mention the immersed boundary method [140], and fictitious domain methods. For our problem, we propose the use of the immersed boundary method with Lagrange multiplier introduced in [37] as an adaptation of the original method to a finite elements version.

The base model consists of advection-diffusion-reaction equations describing the spatio-temporal distributions of blood coagulation factors and platelet subtypes during thrombus development, coupled

with the Navier-Stokes equations to describe the dynamics of blood flow in the vessel,

$$
\begin{aligned}
\frac{\partial P}{\partial t} + \operatorname{div}(\boldsymbol{u}P - D\nabla P) &= (k_1\phi_c + k_2 B_a + k_3 T + k_4 T^2 + k_5 T^3)P, \\
\frac{\partial T}{\partial t} + \operatorname{div}(\boldsymbol{u}T - D\nabla T) &= (k_1\phi_c + k_2 B_a + k_3 T + k_4 T^2 + k_5 T^3)P - k_6 AT, \\
\frac{\partial B_a}{\partial t} + \operatorname{div}(\boldsymbol{u}B_a\ D\nabla B_a) &= k_7\phi_c(B^0 - B_a) + k_8 T(B^0 - B_a) - k_9 AB_a, \\
\frac{\partial A}{\partial t} + \operatorname{div}(\boldsymbol{u}A - D\nabla A) &= -k_6 AT - k_9 AB_a, \\
\frac{\partial F_g}{\partial t} + \operatorname{div}(\boldsymbol{u}F_g - D\nabla F_g) &= -\frac{k_{10} T F_g}{K_{10} + F_g} - k_{11} F, \\
\frac{\partial F_p}{\partial t} &= k_{11} F, \\
\frac{\partial \phi_f}{\partial t} + \operatorname{div}(k(\phi_c + \phi_f))(\boldsymbol{u}\phi_f - D_p\nabla\phi_f) &= -k_{12} T\phi_f - k_{13}\phi_f\phi_c, \\
\frac{\partial \phi_c}{\partial t} + \operatorname{div}(k(\phi_c + \phi_f))(\boldsymbol{u}\phi_c - D_p\nabla\phi_c) &= k_{12} T\phi_f + k_{13}\phi_f\phi_c, \\
\rho_f \frac{\partial \boldsymbol{u}}{\partial t} + \operatorname{div}(\rho\boldsymbol{u}\boldsymbol{u}^T - \mu\nabla\boldsymbol{u} + Ip) &= -\frac{\mu}{K_f}\boldsymbol{u}, \\
\operatorname{div}(\boldsymbol{u}) &= 0.
\end{aligned}
$$

Here $\boldsymbol{u}$, is the flow velocity, p is the pressure, ρ_f the density of the blood, μ is the dynamic viscosity, D is the diffusion coefficient taken the same for all clotting factors, P is the concentration of prothrombin, T is the concentration of thrombin, B_a represents the concentration of clotting factors FIXa and FXa involved in the initiation phase. A, F_g, F, F_p are the concentrations of antithrombin, fibrinogen, fibrin and fibrin polymer, respectively. The density of platelets in flow is denoted by ϕ_f and we use ϕ_c for density in the clot. Following [79], the effective diffusion coefficient for platelets $k(\phi_c + \phi_f)$ and the hydraulic permeability of the clot K_f are given by

$$
\begin{aligned}
k(\phi_c + \phi_f) &= \tanh\left(\pi\left(1 - \frac{\phi_c + \phi_f}{\phi_{\max}}\right)\right), \\
\frac{1}{K_f} &= \frac{16}{\alpha^2}\tilde{F}_p^{3/2}(1 + 56\tilde{F}_p^3)\left(\frac{\phi_{\max} + \phi_c}{\phi_{\max} - \phi_c}\right),
\end{aligned}
$$

where the normalized concentration of fibrin polymer in the clot, $\tilde{F}_p$ is given by,

$$
\tilde{F}_p = \max\left(0.001, \frac{F_p}{0.015\log(F_p) + 0.13}\right).
$$

We consider as a domain a segment of a vein whose two-dimensional approximation is schematically represented in the figure 6.1. Furthermore, the initial and boundary conditions are given by,

$$
\begin{aligned}
&p = p_{\text{in}} && \text{on } \Gamma_{\text{in}}, \\
&p = 0 && \text{on } \Gamma_{\text{out}}, \\
&\boldsymbol{u} = \boldsymbol{0} && \text{on } \Gamma_{\text{wall}} \cup \Gamma_{\text{damaged}}, \\
&P = P_0, A = A_0, F_g = F_g^0, \phi_f = \phi_f^0 && \text{on } \Gamma_{\text{in}}, \\
&\phi_c = \phi_c^0 && \text{on } \Gamma_{\text{damaged}},
\end{aligned}
$$

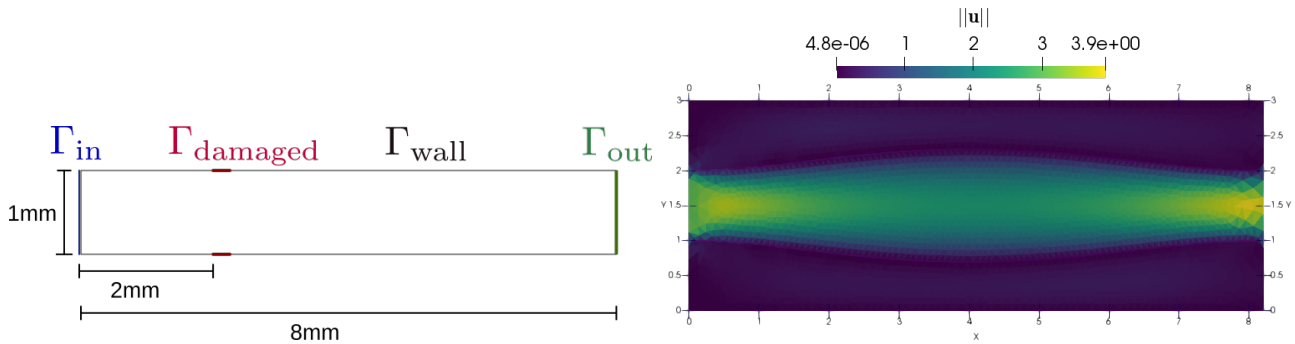


Figure 6.1: Schematic view of the blood vessel 2D approximation (left), snapshot of the velocity norm for a fluid-structure simulation (right) (figure produced by the author).

Weak formulation: The weak formulation of the problem reads as follows: for $t \in (0, \mathcal{T})$, find $F_p(t) \in L^2(\Omega)$ and $P(t)$, $T(t)$, $B_a(t)$, $A(t)$, $F_g(t)$, $\phi_f(t)$, $\phi_c(t) \in H^1(\Omega)$ such that

$$\begin{aligned}
(\partial_t P, m_P)_\Omega + c_2(\boldsymbol{u}; P, m_P) + a_2(P, m_P) &= ((k_1\phi_c + k_2 B_a + k_3 T + k_4 T^2 + k_5 T^3)P, m_P)_\Omega, \\
(\partial_t T, m_T)_\Omega + c_2(\boldsymbol{u}; T, m_T) + a_2(T, m_T) &= ((k_1\phi_c + k_2 B_a + k_3 T + k_4 T^2 + k_5 T^3)P \\
&\quad - k_6 AT, m_T)_\Omega, \\
(\partial_t B_a, m_B)_\Omega + c_2(\boldsymbol{u}; B_a, m_B) + a_2(B_a, m_B) &= (k_7\phi_c(B^0 - B_a) + k_8 T(B^0 - B_a) - k_9 AB_a, m_B)_\Omega, \\
(\partial_t A, m_A)_\Omega + c_2(\boldsymbol{u}; A, m_A) + a_2(A, m_A) &= (-k_6 AT - k_9 AB_a, m_A), \\
(\partial_t F_g, m_g)_\Omega + c_2(\boldsymbol{u}; F_g, m_g) + a_2(F_g, m_g) &= \left(-\frac{k_{10} T F_g}{K_{10} + F_g} - k_{11} F, m_g\right)_\Omega, \\
(\partial_t F_p, m_p)_\Omega &= (k_{11} F, m_p)_\Omega, \\
(\partial_t \phi_f, m_f)_\Omega + c_3(\boldsymbol{u}, \boldsymbol{\varphi}; \phi_f, m_f) + a_3(\phi_f, m_f) &= (-k_{12} T\phi_f - k_{13}\phi_f\phi_c, m_f), \\
(\partial_t \phi_c, m_c)_\Omega + c_3(\boldsymbol{u}, \boldsymbol{\varphi}; \phi_c, m_c) + a_3(\phi_c, m_c) &= (k_{12} T\phi_f + k_{13}\phi_f\phi_c, m_c),
\end{aligned} \tag{6.1}$$

for all $m_P \in H^1(\Omega)$, $m_T \in H^1(\Omega)$, $m_B \in H^1(\Omega)$, $m_A \in H^1(\Omega)$, $m_g \in H^1(\Omega)$, $m_p \in L^2(\Omega)$, $m_f \in H^1(\Omega)$ and $m_c \in H^1(\Omega)$. Here the variational forms c_2, a_2, c_3 and a_3 are defined as follows,

$$a_2(m_1, m_2) = (D\nabla m_1, \nabla m_2)_\Omega \quad c_2(\boldsymbol{u}, m_1, m_2) = (\boldsymbol{u} \cdot \nabla m_1, m_2)_\Omega,$$
$$a_3(\phi_1, \phi_2) = (k(\phi_c + \phi_f) D_p \nabla\phi_1, \nabla\phi_2)_\Omega, \quad c_3(\boldsymbol{u}, \boldsymbol{\varphi}; \phi_1, \phi_2) = (\boldsymbol{u} \cdot \nabla(k(\boldsymbol{\varphi})\phi_1), \phi_2)_\Omega$$

Fluid-structure interaction: Now for the fluid-structure equations, we follow the approach presented in [37]. The immersed boundary method is both a mathematical formulation and a numerical method for fluid-structure interactions, in which immerse compressible visco-elastic bodies interact with and incompressible fluid [36]. The original immersed boundary method was developed by Peskin [140] for the computer simulation of fluid-structure interactions, with special focus in biological fluid dynamics. It links the Lagrangian and Eulerian frameworks and therefore allows to exploit the strengths of both formulations. Although, this method relies on the finite differences method, and a Dirac delta distribution to pass information between frameworks, in [34], a suitable modification

using finite elements was proposed. A posterior modification by the same authors added more flexibility to the scheme through the use of a Lagrange multiplier [33]. The theoretical properties of the immersed boundary method with Lagrange multiplier, between other benefits, show unconditional stability for semi-implicit time discretisations and inf-sup stability for the global saddle point problem under suitable mesh conditions [37].

In this approach, the Navier-Stokes equations are considered everywhere and the presence of the solid structure is taken into account by means of a source term which depends on the unknown position of the structure. These equations are coupled with the condition that the structure moves at the same velocity of the underlying fluid, that is

$$\partial_t \boldsymbol{X}(\boldsymbol{s},t) = \boldsymbol{u}\left(\boldsymbol{X}(\boldsymbol{s},t)\right).$$

Furthermore, a suitable Lagrange multiplier is introduced in the ordinary differential equation that governs the evolution of the solid, thus giving more flexibility to the resulting numerical scheme. More precisely, the equation now reads

$$\gamma_1(\mu, \boldsymbol{u}(\boldsymbol{X}(\cdot,t),t)) - \gamma_2(\mu, \partial_t \boldsymbol{X}(\cdot,t)) = 0 \qquad \text{for all } \mu \in \Lambda,$$

where $\gamma_1(\cdot,\cdot)$ and $\gamma_2(\cdot,\cdot)$ are bilinear forms such that $\gamma_1(\mu, \boldsymbol{v}(\boldsymbol{X})) - \gamma_2(\mu, \boldsymbol{Y}) = 0$ for all $\mu \in \Lambda$ implies $\boldsymbol{v}(\boldsymbol{X}) = \boldsymbol{Y}$.

Let $\Omega \subset \mathbb{R}^d, d = 2,3$, be a bounded domain with Lipschitz continuous boundary. We assume Ω is subdivided into two time-dependant subregions $\Omega_f(t)$ and $\Omega_s(t)$, which correspond to the fluid and the solid material, respectively. We assume, then that the solid domain Ω_s is a image of the reference domain $\mathcal{B} \subset \mathbb{R}^2$. The mapping $\boldsymbol{X} : \mathcal{B} \to \mathbb{R}^2$ associates to each point $\boldsymbol{s} \in \mathcal{B}$ its image $\boldsymbol{x} = \boldsymbol{X}(\boldsymbol{s},t)$ at time t. We denote by λ and μ the Lamé constants. The problem considered is the following: given an initial velocity $\boldsymbol{u}_0 \in \boldsymbol{H}_0^1$, and initial body position $\boldsymbol{X}_0 \in \boldsymbol{W}^{1,\infty}(\mathcal{B})$, find velocity and pressure $(\boldsymbol{u}(t), p(t)) \in (\boldsymbol{H}_0^1(\Omega) \times L_0^2(\Omega))$, body position $\boldsymbol{X}(t) \in \boldsymbol{H}^1(\mathcal{B})$, and a Lagrange multiplier $\chi(t) \in \Lambda$ such that for almost every $t \in]0, \mathcal{T}[$ holds

$$\begin{aligned}
\rho_f(\partial_t \boldsymbol{u}, \boldsymbol{v})_\Omega + a_1(\boldsymbol{u}, \boldsymbol{v}) + c_1(\boldsymbol{u}; \boldsymbol{u}, \boldsymbol{v}) + b(\boldsymbol{v}, p) + d_X(\chi, \boldsymbol{v}(\boldsymbol{X})) &= 0,\\
b(\boldsymbol{u}, q) &= 0,\\
(\partial_{tt}\boldsymbol{X}, \boldsymbol{Y})_\mathcal{B} + a_X(\boldsymbol{X}, \boldsymbol{Y}) - d_X(\chi, \boldsymbol{Y}) &= 0,\\
d_X(\gamma, \boldsymbol{u}(\boldsymbol{X}) - \partial_t \boldsymbol{X}) &= 0,
\end{aligned} \tag{6.2}$$

for all $\boldsymbol{v} \in \boldsymbol{H}_0^1(\Omega)$, $q \in L_0^2(\Omega)$, $\boldsymbol{Y} \in H^1(\mathcal{B})$, $\gamma \in \Lambda$. The variational forms are given by

$$a_1(\boldsymbol{u}, \boldsymbol{v}) := \frac{\mu}{K_f}(\boldsymbol{u}, \boldsymbol{v})_\Omega + (\mu \nabla(\boldsymbol{u}), \nabla(\boldsymbol{v}))_\Omega, \quad b(\boldsymbol{v}, q) := (q, \operatorname{div} \boldsymbol{v})_\Omega,$$

$$c_1(\boldsymbol{w}; \boldsymbol{u}, \boldsymbol{v}) := \big((\boldsymbol{w} \cdot \nabla)\boldsymbol{u}, \boldsymbol{v}\big)_\Omega,$$

$$a_X(\boldsymbol{X}, \boldsymbol{Y}) = \Big(\frac{\mu}{2}\nabla \boldsymbol{X} : \nabla \boldsymbol{Y} + \lambda \operatorname{div}(\boldsymbol{X}) \operatorname{div}(\boldsymbol{Y})\Big)_\mathcal{B}, \quad d_X(\chi, \boldsymbol{u}) = (\eta_1 \nabla \chi : \nabla \boldsymbol{u} + \eta_2 \chi \cdot \boldsymbol{u})_\mathcal{B}$$

If we consider both η_1 and η_2 positive constants different from zero, $\boldsymbol{H}^1(\mathcal{B})$ can be used as the space Λ.

Finite element discretisation: Let $V_h \subset \boldsymbol{H}_0^1(\Omega)$ and $Q_h \subset L_0^2(\Omega)$ be finite element spaces which satisfy the usual discrete ellipticity on the kernel and the discrete inf-sup condition. Moreover,

we consider the finite-dimensional subspaces $\vec{M}_h \subset [H^1(\Omega)]^7 \times L^2(\Omega)$, $\boldsymbol{S}_h \subset \boldsymbol{H}^1(\mathcal{B})$ and $\Lambda_h \subset \Lambda$. Then the finite element counterpart of the problem (6.1) reads as follows: given an initial velocity $\boldsymbol{u}_{0h} \in \boldsymbol{V}_h$, and initial body position $\boldsymbol{X}_0 \in \boldsymbol{W}^{1,\infty}(\mathcal{B})$, find velocity, pressure $(\boldsymbol{u}_h(t), p_h(t)) \in \boldsymbol{V}_h \times Q_h)$, blood coagulation factors $(P_h(t), T_h(t), B_{a,h}(t), A_h(t), F_{g,h}(t), \phi_{f,h}(t), \phi_{c,h}(t), F_{p,h}(t)) \in \vec{M}_h)$, body position $\boldsymbol{X}_h(t) \in \boldsymbol{S}_h(\mathcal{B})$, and a Lagrange multiplier $\chi_h(t) \in \Lambda_h$ such that for almost every $t \in]0, \mathcal{T}[$ holds

$$
\begin{aligned}
\rho_f(\partial_t \boldsymbol{u}_h, \boldsymbol{v})_\Omega + a_1(\boldsymbol{u}_h, \boldsymbol{v}) + c_1(\boldsymbol{u}_h; \boldsymbol{u}_h, \boldsymbol{v}) + b(\boldsymbol{v}, p_h) &= -d_X(\chi_h, \boldsymbol{v}(\boldsymbol{X}_h)),\\
b(\boldsymbol{u}_h, q) &= 0,\\
(\partial_{tt}\boldsymbol{X}_h, \boldsymbol{Y})_\mathcal{B} + a_X(\boldsymbol{X}_h, \boldsymbol{Y}) &= d_X(\chi_h, \boldsymbol{Y}),\\
d_X(\gamma, \boldsymbol{u}_h(\boldsymbol{X}_h) - \partial_t \boldsymbol{X}_h) &= 0,\\
(\partial_t P_h, m_P)_\Omega + c_2(\boldsymbol{u}_h; P_h, m_P) + a_2(P_h, m_P) &= ((k_1\phi_{c,h} + k_2 B_{a,h} + k_3 T + k_4 T_h^2\\
&\quad + k_5 T_h^3)P_h, m_P)_\Omega,\\
(\partial_t T_h, m_T)_\Omega + c_2(\boldsymbol{u}_h; T_h, m_T) + a_2(T_h, m_T) &= ((k_1\phi_{c,h} + k_2 B_{a,h} + k_3 T_h + k_4 T_h^2 + k_5 T_h^3)P_h\\
&\quad - k_6 A_h T_h, m_T)_\Omega,\\
(\partial_t B_{a,h}, m_B)_\Omega + c_2(\boldsymbol{u}_h; B_{a,h}, m_B) + a_2(B_{a,h}, m_B) &= (k_7\phi_{c,h}(B^0 - B_{a,h}) + k_8 T_h(B^0 - B_{a,h})\\
&\quad - k_9 A_h B_{a,h}, m_B)_\Omega,\\
(\partial_t A_h, m_A)_\Omega + c_2(\boldsymbol{u}_h; A_h, m_A) + a_2(A_h, m_A) &= (-k_6 A_h T_h - k_9 A_h B_{a,h}, m_A),\\
(\partial_t F_{g,h}, m_g)_\Omega + c_2(\boldsymbol{u}_h; F_{g,h}, m_g) + a_2(F_{g,h}, m_g) &= (-\frac{k_{10} T_h F_{g,h}}{K_{10} + F_{g,h}} - k_{11} F_h, m_g)_\Omega,\\
(\partial_t F_{p,h}, m_p)_\Omega &= (k_{11} F_h, m_p)_\Omega,\\
(\partial_t \phi_{f,h}, m_f)_\Omega + c_3(\boldsymbol{u}_h, \boldsymbol{\varphi}_h; \phi_{f,h}, m_f) + a_3(\phi_{f,h}, m_f) &= (-k_{12} T_h \phi_{f,h} - k_{13}\phi_{f,h}\phi_{c,h}, m_f),\\
(\partial_t \phi_{c,h}, m_c)_\Omega + c_3(\boldsymbol{u}_h, \boldsymbol{\varphi}_h; \phi_{c,h}, m_c) + a_3(\phi_{c,h}, m_c) &= (k_{12} T_h \phi_{f,h} + k_{13}\phi_{f,h}\phi_{c,h}, m_c),
\end{aligned}
$$

for all $\boldsymbol{v} \in \boldsymbol{V}_h$, $q \in Q_h$, $(m_P, mT, m_B, m_A, m_g, m_f, m_c, m_p) \in \vec{M}_h$, $\boldsymbol{Y} \in \boldsymbol{S}_h$, $\gamma \in \Lambda_h$.

To solve numerically the fully coupled problem it is necessary to introduce an appropriate time discretisation. The simplest choice if one wants to maintain stability in the solution would be the use of an implicit technique, however the Navier-Stokes equations are strongly coupled through the source term with the structure elasticity system which implies the resolution of a fully nonlinear coupled system of equations at each time step. Currently, we employ a first order semi-implict time discretisation to solve the nonlinear system but the exploration of other semi-implicit modifications could be of practical interest. Let us subdivide the time interval $[0, \mathcal{T}]$ into N equal parts with time step Δt, for $n = 1, ...N$, then the relevant part of the discrete in space and time system reads,

$$
\begin{aligned}
\rho_f\left(\frac{\boldsymbol{u}_h^{n+1} - \boldsymbol{u}_h^n}{\Delta t}, \boldsymbol{v}\right)_\Omega + a_1(\boldsymbol{u}_h^{n+1}, \boldsymbol{v}) + c_1(\boldsymbol{u}_h^n; \boldsymbol{u}_h^{n+1}, \boldsymbol{v}) + b(\boldsymbol{v}, p_h^{n+1}) &= -d_X(\chi_h^{n+1}, \boldsymbol{v}(\boldsymbol{X}_h^n)),\\
b(\boldsymbol{u}_h^{n+1}, q) &= 0,\\
\left(\frac{\boldsymbol{X}_h^{n+1} - 2\boldsymbol{X}_h^n + \boldsymbol{X}_h^{n-1}}{\Delta t^2}, \boldsymbol{Y}\right)_\mathcal{B} + a_X(\boldsymbol{X}_h^{n+1}, \boldsymbol{Y}) &= d_X(\chi_h^{n+1}, \boldsymbol{Y}),\\
d_X\left(\gamma, \boldsymbol{u}_h^{n+1}(\boldsymbol{X}_h^n) - \frac{\boldsymbol{X}_h^{n+1} - \boldsymbol{X}_h^n}{\Delta t}\right) &= 0.
\end{aligned}
$$

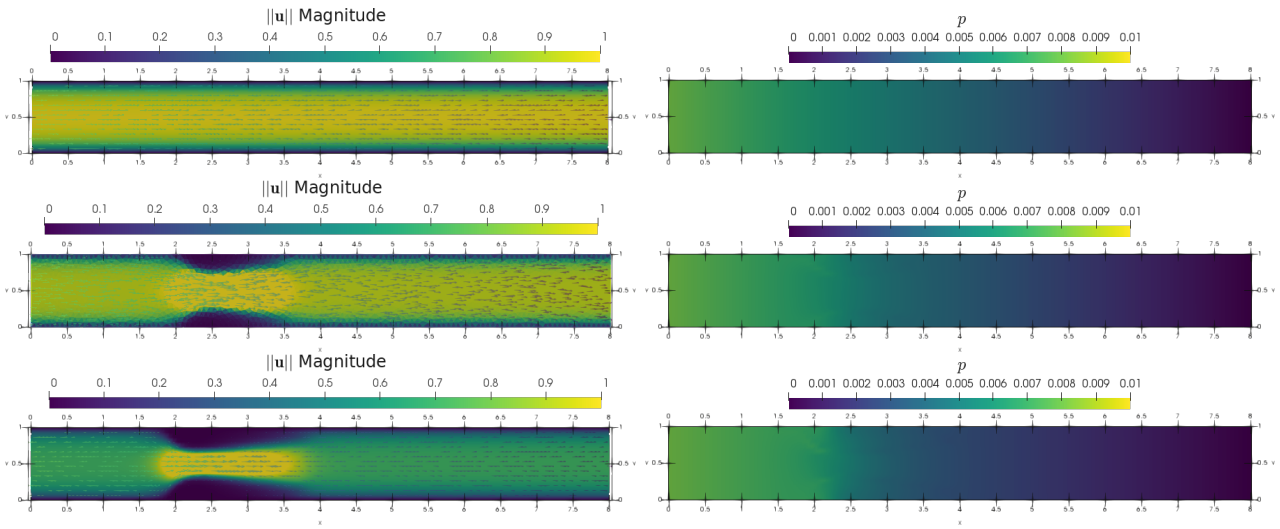


Figure 6.2: Snapshots of velocity and pressure at times $t = 0, 60, 80$ s, with rigid walls (figure produced by the author).

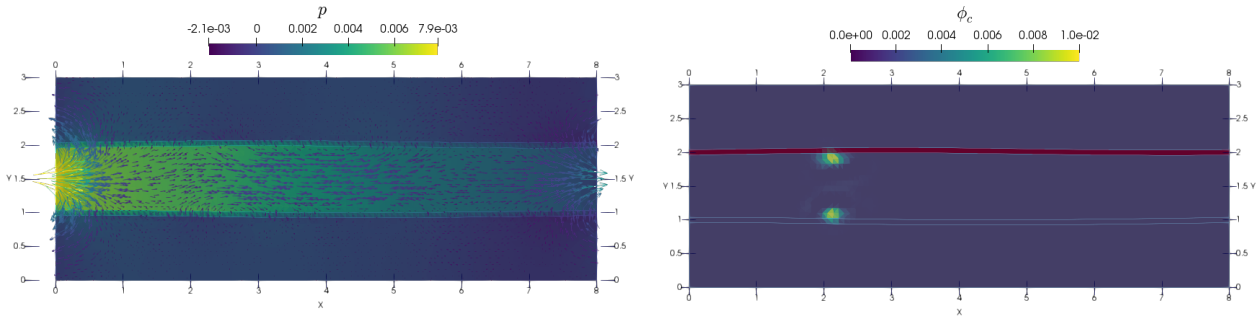


Figure 6.3: Snapshots of pressure with velocity stream lines (left) and platelets concentration ϕ_c (right) at times $t = 0.4$ s for the fluid-structure problem (figure produced by the author).

We have performed some tests with a two-dimensional reduction of the blood vessel (see figure 6.1), assuming rigid walls or with the complete fluid-structure model, to determine if the expected qualitative effects are capture by the model. Results assuming rigid blood vessel walls and BDF2 time discretisation are shown in figure 6.2 and seem to be in qualitative agreement with other computational studies of blood vessel occlusions (i.e. [28]). Preliminary results for the fluid structure model, using the semi-implicit time discretisaton and a stable mini-element pair, show a reasonable deformation for different pressure values when tested on the fluid-structure sub-problem (6.2) (an example is shown on figure 6.1 right), and at initial stages of the complete model with blood coagulation factors reactions (see figure 6.3 left). However, more work is still necessary to obtain a stable and performant scheme for longer time frames, and a more realistic three dimensional model.

6.3 Future works

The general framework addressed in this thesis is very extensive and has a large number of potential applications at theoretical and engineering level, which means that the research in this subject is far from being complete. Indeed, this work has motivated several ongoing and future projects, some of which we briefly described below.

1. To extend the domain of application and develop formulas for estimating parameters associated with entropy stable schemes: We acknowledge that the current form of the entropy stable schemes make them difficult to apply to more general real life problems. Their main limitation is the requirement of a diffusion matrix $\boldsymbol{K}(\boldsymbol{u})$ such that the product $\boldsymbol{K}\eta^{-1}_{\boldsymbol{u},\boldsymbol{u}}$ is positive definite, and the difficulty to obtain stable numerical fluxes from relation (2.28). It would be interesting to investigate alternatives to relax these restrictions. A further inconvenience for the practical use of the schemes is the necessity to fine-tune the extra viscosity parameter α so that oscillations of the numerical solution are avoided but the numerical solutions do not become more smeared than necessary. Clearly, it would be desirable to have a formula at hand that directly generates a value of this parameter from the coefficient functions of (2.1) and $\boldsymbol{u}_0$, avoiding the trial-and-error procedure underlying the final choices of α in our numerical examples. A first step in this direction could be to use the CFL constant to derive an upper bound for the value of α, following the lines of Lemma 2.1 in [114].

2. To extend the model for clarifiers by taking into account the mechanical properties of the rake, rheological models and solid-flux theory: With respect to the model for clarifiers introduced in **Chapter** 4, several interesting extensions were left to be explored. Regarding modelling aspects, we mention that the present approach is likely to be more suitable for the application to clarifiers in wastewater treatment, since for that application the rake can be moved more easily through compacted sludge. In contrast, the sediments formed by the settling of mineral suspension exhibit major resistance to the motion of the rake, and the torque that needs to be applied (that is, the cost of energy) [147] and the precise conditions under which the rake could brake are of utmost importance (a rake being stuck or broken represents a major shutdown of the industrial process) [156]. Our model currently does not resolve the stresses generated in the structure, which is a natural next step. While the approach (4.4) is a rough approximation of the experimental and numerical observation that "rake blades typically suck material behind them as they move as well as pushing material in front of them" [167, p. 102] one could also easily extend the present development to the case of more adequate rheological models for the suspension [56], partly including the effect of shear [84,89] and changes in flock structure [61,90,121]. In addition, for the flow regimes we have studied here, turbulent effects have little relevance but in some industrial settings this is crucial to resolve the separation of clear fluid and solid particles [65,107]. Model reduction and the consistent connection with solid-flux theory should also be considered eventually [68].

3. To study fluid-structure interactions in order to address biological applications of interest: We are interested in extending this work, to address models with biological applications such as the clotting on blood vessels [39] or calcium signalling in embryonic epithelial cells [101]. These

models involve the study of partial differential equations describing poroelastic fluid-structure interactions coupled with systems of transport equations. Although we are open to other alternatives, for the fluid-structure aspect of the model, for now we are working with the immersed boundary method approach [35], that was already briefly reviewed when working on **Chapter** 4. On one side, calcium signalling is one of the most important mechanisms of information propagation in the body, playing an important role as a second messenger in several processes such as embryogenesis, heart function, blood clotting, muscle contraction and diseases of the muscular and nervous systems [101]. In the specific case of embryogenesis, calcium elevation leads to contractions in embryonic cells. Our objective is to develop a multi-dimensional version of the mechanochemical model introduced in [101] describing the interplay of calcium signalling with the mechanics of embryonic epithelial tissue during apical constriction. The model couples the reaction system or Atri's model [21] with a fluid-structure interaction system where we assume the cell compound is surrounded by a viscous fluid and the tissue behaves as a neo-Hookean material. On the side of the coagulation model, we have initial results with a two-dimensional fixed walls model and the next steps include working with the fluid-structure poroelastic model and the full three-dimensional model.

4. Improve the numerical methods: There are a number of improvements we can add in terms of the numerical methods presented throughout this work. For instance, to concentrate on the design of partitioned solvers and efficient preconditioners needed for costly 3D computations with long time horizons [22]. We could also incorporate mixed formulations as an alternative for handling non-homogeneous Dirichlet boundary conditions in the transport equations, removing, for instance, the need to introduce liftings and their consequent numerical approximation, as discussed in the third chapter. Even more, it would allow us to directly obtain other quantities of interest for some engineering applications such as vorticity or stress [17]. We are also interested in study space adaptivity through residual-based a posteriori error indicators [11], and employ more advanced flux reconstruction techniques useful in the regimes of convection-dominated flows and degenerate diffusion of solids due to compression effects [45].

Bibliography

[1] A. Abeynaike, A. Sederman, Y. Khan, M. Johns, J. Davidson, and M. Mackley, *The experimental measurement and modelling of sedimentation and creaming for glycerol/biodiesel droplet dispersions*, Chemical Engineering Science, 79 (2012), pp. 125–137.

[2] A. Agouzal and K. Allali, *Numerical analysis of reaction front propagation model under Boussinesq approximation*, Mathematical Methods in the Applied Sciences, 26 (2003), pp. 1529–1572.

[3] R. Agroum, C. Bernardi, and J. Satouri, *Spectral discretization of the time-dependent Navier-Stokes problem coupled with the heat equation*, Applied Mathematics and Computation, 268 (2015), pp. 59–82.

[4] R. Aldbaissy, F. Hecht, G. Mansour, and T. Sayah, *A full discretisation of the time-dependent Boussinesq (buoyancy) model with nonlinear viscosity*, Calcolo. A Quarterly on Numerical Analysis and Theory of Computation, 55 (2018), pp. Art. 44, 49.

[5] K. Allali, *A priori and a posteriori error estimates for Boussinesq equations*, International Journal of Numerical Analysis and Modeling, 2 (2005), pp. 179–196.

[6] A. Allendes, G. R. Barrenechea, and C. Naranjo, *A divergence-free low-order stabilized finite element method for a generalized steady state Boussinesq problem*, Computer Methods in Applied Mechanics and Engineering, 340 (2018), pp. 90–120.

[7] M. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. Rognes, and G. Wells, *The fenics project version 1.5*, Archive of Numerical Software, 3 (2015), pp. 9–23.

[8] M. Alvarez, G. N. Gatica, B. Gomez-Vargas, and R. Ruiz-Baier, *New mixed finite element methods for natural convection with phase-change in porous media*, Journal of Scientific Computing, 80 (2019), pp. 141–174.

[9] M. Alvarez, 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.

[10] ——, *A mixed-primal finite element approximation of a sedimentation-consolidation system*, Mathematical Models and Methods in Applied Sciences, 26 (2016), pp. 867–900.

[11] ——, *A posteriori error estimation for an augmented mixed-primal method applied to sedimentation-consolidation systems*, Journal of Computational Physics, 367 (2018), pp. 322–346.

[12] M. Amara, D. Capatina-Papaghiuc, B. Denel, and P. Terpolilli, *Mixed finite element approximation for a coupled petroleum reservoir model*, ESAIM: Mathematical Modelling and Numerical Analysis, 39 (2005), pp. 349–376.

[13] M. Anand, K. Rajagopal, and K. Rajagopal, *A model for the formation, growth, and lysis of clots in quiescent plasma. a comparison between the effects of antithrombin III deficiency and protein c deficiency*, Journal of Theoretical Biology, 253 (2008), pp. 725–738.

[14] V. Anaya, M. Bendahmane, D. Mora, and R. Ruiz-Baier, *On a vorticity-based formulation for reaction-diffusion-Brinkman systems*, Networks and Heterogeneous Media, 13 (2018), pp. 69–94.

[15] V. Anaya, D. Mora, C. Reales, and R. Ruiz-Baier, *Stabilized mixed approximation of axisymmetric Brinkman flows*, ESAIM: Mathematical Modelling and Numerical Analysis, 49 (2015), pp. 855–874.

[16] ——, *Mixed methods for a stream-function–vorticity formulation of the axisymmetric Brinkman equations*, Journal of Scientific Computing, 71 (2017), pp. 348–364.

[17] V. Anaya, D. Mora, C. Reales, and R. Ruiz-Baier, *Vorticity-pressure formulations for the brinkman-darcy coupled problem*, Numerical Methods for Partial Differential Equations, 35 (2019), pp. 528–544.

[18] S. M. Aouadi, C. Bernardi, and J. Satouri, *Mortar spectral element discretization of the Stokes problem in axisymmetric domains*, Numerical Methods for Partial Differential Equations., 30 (2014), pp. 44–73.

[19] D. N. Arnold, F. Brezzi, B. Cockburn, and L. D. Marini, *Unified analysis of discontinuous Galerkin methods for elliptic problems*, SIAM Journal on Numerical Analysis, 39 (2001/02), pp. 1749–1779.

[20] F. Assous, P. Ciarlet, Jr., and S. Labrunie, *Theoretical tools to solve the axisymmetric Maxwell equations*, Mathematical Methods in the Applied Sciences, 25 (2002), pp. 49–78.

[21] A. Atri, J. Amundson, D. Clapham, and J. Sneyd, *A single-pool model for intracellular calcium oscillations and waves in the xenopus laevis oocyte*, Biophysical Journal, 65 (1993), pp. 1727–1739.

[22] S. Badia and F. Verdugo, *Robust and scalable domain decomposition solvers for unfitted finite element methods*, Journal of Computational and Applied Mathematics, 344 (2018), pp. 740–759.

[23] G. Baird, R. Bürger, P. E. Méndez, and R. Ruiz-Baier, *Second-order schemes for axisymmetric navier-stokes-brinkman and transport equations modelling water filters.*, Centro de Investigación en Ingeniería Matemática (CI^{2}MA), Universidad de Concepción, Chile, Preprint 2019-23 (2019).

[24] C. S. Balla and K. Naikoti, *Soret and dufour effects on free convective heat and solute transfer in fluid saturated inclined porous cavity*, Engineering Science and Technology, 18 (2015), pp. 543–554.

[25] P. Bastian and B. Rivière, *Superconvergence and $H(\mathrm{div})$ projection for discontinuous Galerkin methods*, International Journal for Numerical Methods in Fluids, 42 (2003), pp. 1043–1057.

[26] Z. Belhachmi, C. Bernardi, and S. Deparis, *Weighted Clément operator and application to the finite element discretization of the axisymmetric Stokes problem*, Numerische Mathematik, 105 (2006), pp. 217–247.

[27] Z. Belhachmi, C. Bernardi, S. Deparis, and F. Hecht, *An efficient discretization of the Navier-Stokes equations in an axisymmetric domain. I. The discrete problem and its numerical analysis*, Journal of Scientific Computing, 27 (2006), pp. 97–110.

[28] A. V. Belyaev, M. A. Panteleev, and F. I. Ataullakhanov, *Threshold of microvascular occlusion: Injury size defines the thrombosis scenario*, Biophysical Journal, 109 (2015), pp. 450–456.

[29] S. Benzoni-Gavage and R. M. Colombo, *An n-populations model for traffic flow*, European Journal of Applied Mathematics, 14 (2002), pp. 587–612.

[30] S. Benzoni-Gavage, R. M. Colombo, and P. Gwiazda, *Measure valued solutions to conservation laws motivated by traffic modelling*, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 462 (2006), pp. 1791–1803.

[31] C. Bernardi and N. Chorfi, *Spectral discretization of the vorticity, velocity, and pressure formulation of the Stokes problem*, SIAM Journal on Numerical Analysis, 44 (2006), pp. 826–850.

[32] C. Bernardi, M. Dauge, and Y. Maday, *Spectral methods for axisymmetric domains*, vol. 3 of Series in Applied Mathematics (Paris), Gauthier-Villars, Éditions Scientifiques et Médicales Elsevier, Paris; North-Holland, Amsterdam, 1999.

[33] D. Boffi, N. Cavallini, and L. Gastaldi, *The finite element immersed boundary method with distributed Lagrange multiplier*, SIAM Journal on Numerical Analysis, 53 (2015), pp. 2584–2604.

[34] D. Boffi and L. Gastaldi, *A finite element approach for the immersed boundary method*, vol. 81, 2003, pp. 491–501.

[35] ——, *Discrete models for fluid-structure interactions: the finite element immersed boundary method*, Discrete and Continuous Dynamical Systems. Series S, 9 (2016), pp. 89–107.

[36] D. Boffi, L. Gastaldi, L. Heltai, and C. S. Peskin, *On the hyper-elastic formulation of the immersed boundary method*, Computer Methods in Applied Mechanics and Engineering, 197 (2008), pp. 2210–2231.

[37] D. Boffi, F. Hecht, and O. Pironneau, *Distributed Lagrange multiplier for fluid-structure interactions*, in Numerical methods for PDEs, vol. 15 of SEMA SIMAI Springer Ser., Springer, Cham, 2018, pp. 129–145.

[38] S. Boscarino, R. Bürger, P. Mulet, G. Russo, and L. M. Villada, *Linearly implicit IMEX runge–kutta methods for a class of degenerate convection-diffusion problems*, SIAM Journal on Scientific Computing, 37 (2015), pp. B305–B331.

[39] A. Bouchnita, T. Galochkina, P. Kurbatova, P. Nony, and V. Volpert, *Conditions of microvessel occlusion for blood coagulation in flow*, International Journal for Numerical Methods in Biomedical Engineering, 33 (2017), p. e2850.

[40] F. Brezzi, J. Douglas, Jr., and L. D. Marini, *Two families of mixed finite elements for second order elliptic problems*, Numerische Mathematik, 47 (1985), pp. 217–235.

[41] F. Brezzi and M. Fortin, *Mixed and hybrid finite element methods*, vol. 15 of Springer Series in Computational Mathematics, Springer-Verlag, New York, 1991.

[42] R. Bürger, J. Careaga, S. Diehl, C. Mejías, and R. Ruiz-Baier, *Convection-diffusion-reaction and transport-flow problems motivated by models of sedimentation: some recent advances*, in Proceedings of the International Congress of Mathematicians, B. Sirakov, P. de Souza, and M. Viana, eds., vol. IV: Invited Lectures, World Scientific, Singapore, 2018, pp. 3489–3514.

[43] R. Bürger, S. Diehl, M. C. Martí, P. Mulet, I. Nopens, E. Torfs, and P. A. Vanrolleghem, *Numerical solution of a multi-class model for batch settling in water resource recovery facilities*, Applied Mathematical Modelling, 49 (2017), pp. 415–436.

[44] R. Bürger, R. Donat, P. Mulet, and C. A. Vega, *On the implementation of WENO schemes for a class of polydisperse sedimentation models*, Journal of Computational Physics, 230 (2011), pp. 2322–2344.

[45] R. Bürger, K.-K. Fjelde, K. Höfler, and K. H. Karlsen, *Central difference solutions of the kinematic model of settling of polydisperse suspensions and three-dimensional particle-scale simulations*, Journal of Engineering Mathematics, 41 (2001), pp. 167–187.

[46] R. Bürger, S. K. Kenettinkara, R. Ruiz-Baier, and H. Torres, *Coupling of discontinuous Galerkin schemes for viscous flow in porous media with adsorption*, SIAM Journal on Scientific Computing, 40 (2018), pp. B637–B662.

[47] 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.

[48] R. Bürger, P. E. Méndez, and C. Parés, *On entropy stable schemes for degenerate parabolic multispecies kinematic flow models*, Numerical Methods for Partial Differential Equations, 35 (2019), pp. 1847–1872.

[49] R. Bürger, P. E. Méndez, and R. Ruiz-Baier, *On $H(\mathrm{div})$-conforming methods for double-diffusion equations in porous media*, SIAM Journal on Numerical Analysis, 57 (2019), pp. 1318–1343.

[50] ——, *A second-order $H(\mathrm{div})$-conforming scheme for the simulation of sedimentation and flow in circular clarifiers with a rotating rake*, Centro de Investigación en Ingeniería Matemática (CI^{2}MA), Universidad de Concepción, Chile, Preprint 2019-39 (2019).

[51] R. Bürger, P. Mulet, and L. Rubio, *Implicit-explicit methods for the efficient simulation of the settling of dispersions of droplets and colloidal particles*, Advances in Applied Mathematics and Mechanics, 10 (2018), pp. 445–467.

[52] R. Bürger, P. Mulet, L. Rubio, and M. Sepúlveda, *Linearly implicit-explicit schemes for the equilibrium dispersive model of chromatography*, Applied Mathematics and Computation, 317 (2018), pp. 172–186.

[53] R. Bürger, P. Mulet, and L. M. Villada, *A diffusively corrected multiclass lighthill-whitham-richards traffic model with anticipation lengths and reaction times*, Advances in Applied Mathematics and Mechanics, 5 (2013), pp. 728–758.

[54] ——, *Regularized nonlinear solvers for IMEX methods applied to diffusively corrected multi-species kinematic flow models*, SIAM Journal on Scientific Computing, 35 (2013), pp. B751–B777.

[55] 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.

[56] R. Bürger, W. L. Wendland, and F. Concha, *Model equations for gravitational sedimentation-consolidation processes*, ZAMM. Zeitschrift für Angewandte Mathematik und Mechanik. Journal of Applied Mathematics and Mechanics, 80 (2000), pp. 79–92.

[57] J. Carrillo, *Entropy solutions for nonlinear degenerate problems*, Archive for Rational Mechanics and Analysis, 147 (1999), pp. 269–361.

[58] M. J. Castro, U. S. Fjordholm, S. Mishra, and C. Parés, *Entropy Conservative and Entropy Stable Schemes for Nonconservative Hyperbolic Systems*, SIAM Journal on Numerical Analysis, 51 (2013), pp. 1371–1391.

[59] A. Çibik and S. Kaya, *Finite element analysis of a projection-based stabilization method for the Darcy-Brinkman equations in double-diffusive convection*, Applied Numerical Mathematics, 64 (2013), pp. 35–49.

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

[61] G. M. Channell and C. F. Zukoski, *Shear and compressive rheology of aggregated alumina suspensions*, AIChE Journal, 43 (1997), pp. 1700–1708.

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

[63] E. Colmenares, G. N. Gatica, and R. Oyarzúa, *Fixed point strategies for mixed variational formulations of the stationary Boussinesq problem*, Comptes Rendus Mathématique. Académie des Sciences. Paris, 354 (2016), pp. 57–62.

[64] X. Cui, *The regularity criterion for weak solutions to the n-dimensional Boussinesq system*, Boundary Value Problems, (2017), pp. Paper No. 44, 12.

[65] S. Das, H. Bai, C. Wu, J.-H. Kao, B. Barney, M. Kidd, and M. Kuettel, *Improving the performance of industrial clarifiers using three-dimensional computational fluid dynamics*, Engineering Applications of Computational Fluid Mechanics, 10 (2016), pp. 130–144.

[66] R. de Aguiar, B. Climent-Ezquerra, M. A. Rojas-Medar, and M. D. Rojas-Medar, *On the convergence of Galerkin spectral methods for a bioconvective flow*, Journal of Mathematical Fluid Mechanics, 19 (2017), pp. 91–104.

[67] D. A. Di Pietro and A. Ern, *Mathematical aspects of discontinuous Galerkin methods*, vol. 69 of Mathématiques & Applications (Berlin) [Mathematics & Applications], Springer, Heidelberg, 2012.

[68] S. Diehl, *The solids-flux theory - confirmation and extension by using partial differential equations*, Water Research, 42 (2008), pp. 4976–4988.

[69] M. Discacciati, D. Hacker, A. Quarteroni, S. Quinodoz, S. Tissot, and F. M. Wurm, *Numerical simulation of orbitally shaken viscous fluids with free surface*, International Journal for Numerical Methods in Fluids, 71 (2013), pp. 294–315.

[70] R. Donat, F. Guerrero, and P. Mulet, *Implicit-explicit WENO scheme for the equilibrium dispersive model of chromatography*, Applied Numerical Mathematics, 123 (2018), pp. 22–42.

[71] V. J. Ervin, *Approximation of axisymmetric Darcy flow using mixed finite element methods*, SIAM Journal on Numerical Analysis, 51 (2013), pp. 1421–1442.

[72] ——, *Approximation of coupled Stokes-Darcy flow in an axisymmetric domain*, Computer Methods in Applied Mechanics and Engineering, 258 (2013), pp. 96–108.

[73] S. Evje and K. H. Karlsen, *Monotone difference approximations of BV solutions to degenerate convection-diffusion equations*, SIAM Journal on Numerical Analysis, 37 (2000), pp. 1838–1860.

[74] R. S. Falk and M. Neilan, *Stokes complexes and the construction of stable finite elements with pointwise mass conservation*, SIAM Journal on Numerical Analysis, 51 (2013), pp. 1308–1326.

[75] U. S. Fjordholm and S. Mishra, *Accurate numerical discretizations of non-conservative hyperbolic systems*, ESAIM: Mathematical Modelling and Numerical Analysis, 46 (2011), pp. 187–206.

[76] U. S. Fjordholm, S. Mishra, and E. Tadmor, *Well-balanced and energy stable schemes for the shallow water equations with discontinuous topography*, Journal of Computational Physics, 230 (2011), pp. 5587–5609.

[77] ——, *Arbitrarily high-order accurate entropy stable essentially nonoscillatory schemes for systems of conservation laws*, SIAM Journal on Numerical Analysis, 50 (2012), pp. 544–573.

[78] U. S. Fjordholm, S. Mishra, and E. Tadmor, *Arbitrarily high-order accurate entropy stable essentially nonoscillatory schemes for systems of conservation laws*, SIAM Journal on Numerical Analysis, 50 (2012), pp. 544–573.

[79] A. L. Fogelson and K. B. Neeves, *Fluid mechanics of blood clot formation*, Annual Review of Fluid Mechanics, 47 (2015), pp. 377–403.

[80] G. P. Galdi and A. L. Silvestre, *On the motion of a rigid body in a Navier-Stokes liquid under the action of a time-periodic force*, Indiana University Mathematics Journal, 58 (2009), pp. 2805–2842.

[81] C. Geuzaine and J.-F. Remacle, *Gmsh: A 3-d finite element mesh generator with built-in pre- and post-processing facilities*, International Journal for Numerical Methods in Engineering, 79 (2009), pp. 1309–1331.

[82] V. Girault and P.-A. Raviart, *Finite element methods for Navier-Stokes equations*, vol. 5 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1986.

[83] V. Girault, B. Rivière, and M. F. Wheeler, *A discontinuous Galerkin method with nonoverlapping domain decomposition for the Stokes and Navier-Stokes problems*, Mathematics of Computation, 74 (2005), pp. 53–84.

[84] B. R. Gladman, M. Rudman, and P. J. Scales, *The effect of shear on gravity thickening: Pilot scale modelling*, Chemical Engineering Science, 65 (2010), pp. 4293–4301.

[85] S. Gottlieb, C.-W. Shu, and E. Tadmor, *Strong stability-preserving high-order time discretization methods*, SIAM Review, 43 (2001), pp. 89–112.

[86] B. Goyeau, J.-P. Songbe, and D. Gobin, *Numerical study of double-diffusive natural convection in a porous cavity using the darcy-brinkman formulation*, International Journal of Heat and Mass Transfer, 39 (1996), pp. 1363–1378.

[87] B. D. Greenshields, *A study of traffic capacity*, in Proceedings of the highway research board, vol. 14, 1935, pp. 448–477.

[88] J. L. Guo and P. N. Kaloni, *Double-diffusive convection in a porous medium, nonlinear stability, and the Brinkman effect*, Studies in Applied Mathematics, 94 (1995), pp. 341–358.

[89] K. Gustavsson and J. Oppelstrup, *Consolidation of concentrated suspensions – numerical simulations using a two-phase fluid model*, Computing and Visualization in Science, 3 (2000), pp. 39–45.

[90] K. Gustavsson, J. Oppelstrup, and J. Eiken, *Consolidation of concentrated suspensions – shear and irreversible floc structure rearrangement*, Computing and Visualization in Science, 4 (2001), pp. 61–66.

[91] J. Guzman and M. Neilan, *Conforming and divergence-free stokes elements in three dimensions*, IMA Journal of Numerical Analysis, 34 (2013), pp. 1489–1508.

[92] P. Hansbo and M. G. Larson, *Discontinuous Galerkin methods for incompressible and nearly incompressible elasticity by Nitsche's method*, Computer Methods in Applied Mechanics and Engineering, 191 (2002), pp. 1895–1908.

[93] J. C. Heinrich, *A finite element model for double diffusive convection*, International Journal for Numerical Methods in Engineering, 20 (1984), pp. 447–464.

[94] A. Hiltebrand and S. Mishra, *Entropy stable shock capturing space–time discontinuous Galerkin schemes for systems of conservation laws*, Numerische Mathematik, 126 (2014), pp. 103–151.

[95] M. F. Hockin, K. C. Jones, S. J. Everse, and K. G. Mann, *A model for the stoichiometric regulation of blood coagulation*, Journal of Biological Chemistry, 277 (2002), pp. 18322–18333.

[96] T. L. Horváth and S. Rhebergen, *A locally conservative and energy-stable finite-element method for the Navier-Stokes problem on time-dependent domains*, International Journal for Numerical Methods in Fluids, 89 (2019), pp. 519–532.

[97] T. Hughes, L. Franca, and M. Mallet, *A new finite element formulation for computational fluid dynamics: I. symmetric forms of the compressible euler and navier-stokes equations and the second law of thermodynamics*, Computer Methods in Applied Mechanics and Engineering, 54 (1986), pp. 223–234.

[98] F. Ismail and P. L. Roe, *Affordable, entropy-consistent Euler flux functions. II. Entropy production at shocks*, Journal of Computational Physics, 228 (2009), pp. 5410–5436.

[99] S. Jerez and C. Parés, *Entropy Stable Schemes for Degenerate Convection-Diffusion Equations*, SIAM Journal on Numerical Analysis, 55 (2017), pp. 240–264.

[100] V. John, A. Linke, C. Merdon, M. Neilan, and L. G. Rebholz, *On the divergence constraint in mixed finite element methods for incompressible flows*, SIAM Review, 59 (2017), pp. 492–544.

[101] K. Kaouri, P. K. Maini, P. A. Skourides, N. Christodoulou, and S. J. Chapman, *A simple mechanochemical model for calcium signalling in embryonic epithelial cells*, Journal of Mathematical Biology, 78 (2019), pp. 2059–2092.

[102] O. A. Karakashian and W. N. Jureidini, *A nonconforming finite element method for the stationary Navier-Stokes equations*, SIAM Journal on Numerical Analysis, 35 (1998), pp. 93–120.

[103] K. H. Karlsen, *Upwind difference approximations for degenerate parabolic convection-diffusion equations with a discontinuous coefficient*, IMA Journal of Numerical Analysis, 22 (2002), pp. 623–664.

[104] K. H. Karlsen, U. Koley, and N. H. Risebro, *An error estimate for the finite difference approximation to degenerate convection–diffusion equations*, Numerische Mathematik, 121 (2011), pp. 367–395.

[105] K. H. Karlsen and N. H. Risebro, *Convergence of finite difference schemes for viscous and inviscid conservation laws with rough coefficients*, ESAIM: Mathematical Modelling and Numerical Analysis, 35 (2001), pp. 239–269.

[106] ——, *On the uniqueness and stability of entropy solutions of nonlinear degenerate parabolic equations with rough coefficients*, Discrete and Continuous Dynamical Systems, 9 (2003), pp. 1081–1104.

[107] D. Kleine and B. D. Reddy, *Finite element analysis of flows in secondary settling tanks*, International Journal for Numerical Methods in Engineering, 64 (2005), pp. 849–876.

[108] D. Kolomenskiy and K. Schneider, *A Fourier spectral method for the Navier-Stokes equations with volume penalization for moving solid obstacles*, Journal of Computational Physics, 228 (2009), pp. 5687–5709.

[109] J. Könnö and R. Stenberg, *$H(\mathrm{div})$-conforming finite elements for the Brinkman problem*, Mathematical Models and Methods in Applied Sciences, 21 (2011), pp. 2227–2248.

[110] S. N. Kružkov, *First order quasilinear equations in several independent variables*, Mathematics of the USSR-Sbornik, 10 (1970), pp. 217–243.

[111] A. Kufner, *Weighted Sobolev spaces*, vol. 31 of Teubner-Texte zur Mathematik [Teubner Texts in Mathematics], BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1980.

[112] I. Kukavica, F. Wang, and M. Ziane, *Persistence of regularity for solutions of the Boussinesq equations in Sobolev spaces*, Advances in Differential Equations, 21 (2016), pp. 85–108.

[113] A. Kurganov and C.-T. Lin, *On the reduction of numerical dissipation in central-upwind schemes*, Communications in Computational Physics, 2 (2007), pp. 141–163.

[114] A. Kurganov and Y. Liu, *New adaptive artificial viscosity method for hyperbolic systems of conservation laws*, Journal of Computational Physics, 231 (2012), pp. 8114–8132.

[115] A. Kurganov, S. Noelle, and G. Petrova, *Semidiscrete central-upwind schemes for hyperbolic conservation laws and hamilton–jacobi equations*, SIAM Journal on Scientific Computing, 23 (2001), pp. 707–740.

[116] A. Kurganov and E. Tadmor, *New high-resolution central schemes for nonlinear conservation laws and convection–diffusion equations*, Journal of Computational Physics, 160 (2000), pp. 241–282.

[117] G. J. Kynch, *A theory of sedimentation*, Transactions of the Faraday Society, 48 (1952), pp. 166–176.

[118] P. D. Lax, *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*, Society for Industrial and Applied Mathematics, Jan. 1973.

[119] H. G. Lee and J. Kim, *Numerical investigation of falling bacterial plumes caused by bioconvection in a three-dimensional chamber*, European Journal of Mechanics. B. Fluids, 52 (2015), pp. 120–130.

[120] P. Lenarda, M. Paggi, and R. Ruiz-Baier, *Partitioned coupling of advection-diffusion-reaction systems and Brinkman flows*, Journal of Computational Physics, 344 (2017), pp. 281–302.

[121] D. Lester, M. Rudman, and P. Scales, *Macroscopic dynamics of flocculated colloidal suspensions*, Chemical Engineering Science, 65 (2010), pp. 6362–6378.

[122] R. J. LeVeque, *Numerical Methods for Conservation Laws*, Birkhäuser Basel, 1992.

[123] M. J. Lighthill and G. B. Whitham, *On kinematic waves. II. A theory of traffic flow on long crowded roads*, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 229 (1955), pp. 317–345.

[124] C. Lin and L. E. Payne, *Continuous dependence on the Soret coefficient for double diffusive convection in Darcy flow*, Journal of Mathematical Analysis and Applications, 342 (2008), pp. 311–325.

[125] S. A. Lorca and J. L. Boldrini, *Stationary solutions for generalized Boussinesq models*, Journal of Differential Equations, 124 (1996), pp. 389–406.

[126] ——, *The initial value problem for a generalized Boussinesq model*, Nonlinear Analysis. Theory, Methods & Applications, 36 (1999), pp. 457–480.

[127] R. B. Lowrie and J. E. Morel, *Discontinuous Galerkin for hyperbolic systems with stiff relaxation*, in Discontinuous Galerkin methods (Newport, RI, 1999), vol. 11 of Lect. Notes Comput. Sci. Eng., Springer, Berlin, 2000, pp. 385–390.

[128] A. Lozovskiy, M. A. Olshanskii, V. Salamatova, and Y. V. Vassilevski, *An unconditionally stable semi-implicit FSI finite element method*, Computer Methods in Applied Mechanics and Engineering, 297 (2015), pp. 437–454.

[129] E. Magyari and A. Postelnicu, *Double-diffusive natural convection flows with thermosolutal symmetry in porous media in the presence of the Soret-Dufour effects*, Transport in Porous Media, 88 (2011), pp. 149–167.

[130] A. Maiti, J. K. Basu, and S. De, *Development of a treated laterite for arsenic adsorption: Effects of treatment parameters*, Industrial & Engineering Chemistry Research, 49 (2010), pp. 4873–4886.

[131] A. Maiti, S. DasGupta, J. K. Basu, and S. De, *Batch and column study: Adsorption of arsenate using untreated laterite as adsorbent*, Industrial & Engineering Chemistry Research, 47 (2008), pp. 1620–1629.

[132] A. Maiti, H. Sharma, J. K. Basu, and S. De, *Modeling of arsenic adsorption kinetics of synthetic and contaminated groundwater on natural laterite*, Journal of Hazardous Materials, 172 (2009), pp. 928–934.

[133] Mercier, B. and Raugel, G., *Resolution d'un problème aux limites dans un ouvert axisymétrique par éléments finis en r, z et séries de fourier en t*, RAIRO. Analyse numérique, 16 (1982), pp. 405–461.

[134] R. Mondal, S. Mondal, K. V. Kurada, S. Bhattacharjee, S. Sengupta, M. Mondal, S. Karmakar, S. De, and I. M. Griffiths, *Modelling the transport and adsorption dynamics of arsenic in a soil bed filter*, Chemical Engineering Science, 210 (2019), p. 115205.

[135] K. D. Nikitin, M. A. Olshanskii, K. M. Terekhov, Y. V. Vassilevski, and R. M. Yanbarisov, *An adaptive numerical method for free surface flows passing rigidly mounted obstacles*, Computers & Fluids, 148 (2017), pp. 56–68.

[136] P. Nithiarasu, K. N. Seetharamu, and T. Sundararajan, *Double-diffusive natural convection in an enclosure filled with fluid-saturated porous medium: A generalized non-darcy approach*, Numerical Heat Transfer, Part A: Applications, 30 (1996), pp. 413–426.

[137] 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.

[138] R. Oyarzúa and M. Serón, *A divergence-conforming dg-mixed finite element method for the stationary boussinesq problem*, CI^{2}MA, Universidad de Concepción, Chile. [Preprint available from http://www.ci2ma.udec.cl], (2018).

[139] P. R. Patil and C. P. Parvathy, *Thermohaline convection with cross-diffusion in an anisotropic porous medium*, Indian Academy of Sciences. Proceedings. Mathematical Sciences, 99 (1989), pp. 93–101.

[140] C. S. Peskin, *The immersed boundary method*, Acta Numerica, 11 (2002), p. 479–517.

[141] C. Rana, M. Mishra, and A. D. Wit, *Effect of anti-langmuir adsorption on spreading in porous media*, EPL (Europhysics Letters), 124 (2019), p. 64003.

[142] S. Rhebergen and G. N. Wells, *Analysis of a hybridized/interface stabilized finite element method for the Stokes equations*, SIAM Journal on Numerical Analysis, 55 (2017), pp. 1982–2003.

[143] ——, *A hybridizable discontinuous Galerkin method for the Navier-Stokes equations with pointwise divergence-free velocity field*, Journal of Scientific Computing, 76 (2018), pp. 1484–1501.

[144] P. I. Richards, *Shock waves on the highway*, Operations Research, 4 (1956), pp. 42–51.

[145] J. Richardson and W. Zaki, *Sedimentation and fluidisation: Part i*, Transactions of the Institution of Chemical Engineers, 32 (1954), pp. 35–53.

[146] F. Rosso and G. Sona, *Gravity-driven separation of oil-water dispersions*, Advances in Mathematical Sciences and Applications., 11 (2001), pp. 127–151.

[147] M. Rudman, K. Simic, D. Paterson, P. Strode, A. Brent, and I. Šutalo, *Raking in gravity thickeners*, International Journal of Mineral Processing, 86 (2008), pp. 114–130.

[148] R. Ruiz-Baier and I. Lunati, *Mixed finite element – discontinuous finite volume element discretization of a general class of multicontinuum models*, Journal of Computational Physics, 322 (2016), pp. 666–688.

[149] J. N. Shadid, R. S. Tuminaro, and H. F. Walker, *An inexact Newton method for fully coupled solution of the Navier-Stokes equations with heat and mass transport*, Journal of Computational Physics, 137 (1997), pp. 155–185.

[150] Q. Shao, M. Fahs, A. Younes, A. Makradi, and T. Mara, *A new benchmark reference solution for double-diffusive convection in a heterogeneous porous medium*, Numerical Heat Transfer, Part B: Fundamentals, 70 (2016), pp. 373–392.

[151] K. Stamatakis and C. Tien, *Batch sedimentation calculations — the effect of compressible sediment*, Powder Technology, 72 (1992), pp. 227–240.

[152] L. Szalai, P. Krebs, and W. Rodi, *Simulation of flow in circular clarifiers with and without swirl*, Journal of Hydraulic Engineering, 120 (1994), pp. 4–21.

[153] E. Tadmor, *The numerical viscosity of entropy stable schemes for systems of conservation laws. I*, Mathematics of Computation, 49 (1987), pp. 91–103.

[154] L. Tang, D. Liu, F.-Y. Zhao, and G.-F. Tang, *Combined heat and moisture convective transport in a partial enclosure with multiple free ports*, Applied Thermal Engineering, 30 (2010), pp. 977–990.

[155] L. Q. Tang and T. T. H. Tsang, *A least-squares finite element method for doubly-diffusive convection*, International Journal of Computational Fluid Dynamics, 3 (1994), pp. 1–17.

[156] S. Teerikoski, *Optimal control of clarifier-thickeners*, Master's thesis, Uppsala University, Division of Systems and Control, 2017.

[157] R. Temam, *Navier-Stokes equations. Theory and numerical analysis*, North-Holland Publishing Co., Amsterdam-New York-Oxford, 1977. Studies in Mathematics and its Applications, Vol. 2.

[158] J. Towers, K. Karlsen, and R. Bürger, *On some difference schemes and entropy conditions for a class of multi-species kinematic flow models with discontinuous flux*, Networks and Heterogeneous Media, 5 (2010), pp. 461–485.

[159] B. O. Turesson, *Nonlinear potential theory and weighted Sobolev spaces*, vol. 1736 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 2000.

[160] J.-M. Vanson, A. Boutin, M. Klotz, and F.-X. Coudert, *Transport and adsorption under liquid flow: the role of pore geometry*, Soft Matter, 13 (2017), pp. 875–885.

[161] C. Y. H. Wong, P. H. Trinh, and S. J. Chapman, *Shear-induced instabilities of flows through submerged vegetation.* arXiv:1904.11981, 2019.

[162] G. Wong and S. Wong, *A multi-class traffic flow model – an extension of LWR model with heterogeneous drivers*, Transportation Research Part A: Policy and Practice, 36 (2002), pp. 827–841.

[163] J. Woodfield, M. Alvarez, B. Gómez-Vargas, and R. Ruiz-Baier, *Stability and finite element approximation of phase change models for natural convection in porous media*, Journal of Computational and Applied Mathematics, 360 (2019), pp. 117–137.

[164] K. Yang, P. Sun, L. Wang, J. Xu, and L. Zhang, *Modeling and simulations for fluid and rotating structure interactions*, Computer Methods in Applied Mechanics and Engineering, 311 (2016), pp. 788–814.

[165] N. Zabaras and D. Samanta, *A stabilized volume-averaging finite element method for flow in porous media and binary alloy solidification processes*, International Journal for Numerical Methods in Engineering, 60 (2004), pp. 1103–1138.

[166] Y. Zhuang, H. Yu, and Q. Zhu, *A thermal non-equilibrium model for 3d double diffusive convection of power-law fluids with chemical reaction in the porous medium*, International Journal of Heat and Mass Transfer, 115 (2017), pp. 670–694.

[167] I. Šutalo, D. Paterson, and M. Rudman, *Flow visualisation and computational prediction in thickener rake models*, Minerals Engineering, 16 (2003), pp. 93–102.