8-2017

Time-dependent Stokes-Darcy Flow with Deposition

Javier Ruiz Ramirez
Clemson University, JAVIER@CLEMSON.EDU

Time-dependent Stokes-Darcy flow with deposition

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Mathematical Sciences

---

by
Javier Ruiz Ramírez
August 2017

Accepted by:
Dr. Vincent Ervin, Committee Chair
Dr. Timo Heister
Dr. Lea Jenkins
Dr. Hyesuk Lee

Abstract

This thesis investigates two nonlinear systems of time-dependent partial differential equations that model a filtration process. Existence and uniqueness results for the governing equations is established. For each system, a finite element scheme capable of approximating the solutions is investigated. Accompanying numerical experiments corroborate the analytical findings. Finally, an optimization application concerning the design of a filter is discussed and supported with a numerical study.

Dedication

I dedicate this thesis to my parents, Javier and Patricia.

Acknowledgments

First and foremost, I would like to thank my adviser Prof. Vince Ervin for his help and example throughout this chapter of my academic life. As a person and mathematician, he is someone I will always look up to.

Second, I would like to thank the committee members, Dr. Hyesuk Lee, Dr. Lea Jenkins, and Dr. Timo Heister for their helpful suggestions and remarks that contributed to the completion of this thesis.

Third, I am also thankful to my teachers, Dr. Adams Warren, Dr. Michael Burr, Dr. Jim Brown, Dr. Qingshan Chen, Dr. Colin Gallagher, Dr. Taufiquar Khan, Dr. Peter Kiessler, Dr. Shitao Liu, Dr. Robert Lund, Dr. Mishko Mitkovski, Dr. Leo Rebholz, Dr. Martin Schmoll, Dr. Margaret Wiecek, Dr. Jeong-Rock Yoon, Dr. William Moss, Dr. Daniel Warner, Dr. Pietro Belotti, and Dr. Douglas Hirt for the knowledge they shared.

Forth, I am grateful for all the help I received from the Mathematics Department. In particular from Dr. Chris Cox, April Haynes, Kevin Hedetniemi, Dr. Kevin James, Dr. Calvin Williams, Dr. Judith McKnew, and Randy Davidson.

Fifth, I would like to thank Jaman, Rashed, Rafiul, Dilhani, Drew, Chris, Fawwaz, Fatih, and Samet for the time we spent talking about math and other topics.

Sixth, I would like to acknowledge Clemson University for generous allotment of computer time on the Palmetto cluster.

Seventh, I am very indebted to Dr. Jorge Eduardo Macías-Díaz, without whose support and teachings I would have never had access to an institution like Clemson University.

Last but not least, I owe everything to my parents, that have sacrificed so much to see me lead a honest and respectable life.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Fluid flow through porous media is a ubiquitous process in our everyday lives. From the routine activities such as: The preparation of espresso coffee in the morning [39], the water we drink from the faucet [76], and the car we drive to work [66]. To the less obvious but not less important such as: The absorption of nutrients in the small intestine [58], the cleansing of blood in the kidneys [51], and the prevention of postoperative infections [59]. All these phenomena rely on the separation of some solid from a fluid by means of a medium that is permeable to the fluid but (mostly) impermeable to the solid. This mechanism of operation is what defines filtration.

This thesis presents a filtration model that aims to be applicable to the purification industry. To understand the underlying equations, Chapter 2 is devoted to the derivation of the fundamental flow equations the we consider. Namely, the Stokes and Darcy equations. Moreover, we discuss some relevant physical parameters such as the viscosity, the porosity, and the permeability.

Subsequently, in Chapter 3, we introduce and justify the core ideas of this work. The Darcy problem with deposition. Therein, the existence and uniqueness of the modeling equations is established.

Next, in Chapter 4, a finite element scheme for the Darcy problem with deposition is investigated. It is shown that it possesses a unique solution and optimal convergence properties. Numerical examples that corroborate the analytical findings and explore the performance of different filters, are also included.

Chapter 5 is the major contribution of this thesis. It generalizes the system of equations discussed in Chapter 3 by coupling the Stokes equations to the Darcy problem with deposition. Chapter 5 starts with a motivation for the model and a short literature review. Later, the notation and main assumptions are stated for ease of reference. The well-posedness of the system of equations with appropriate boundary conditions follows. Finally, the chapter closes showing some nonnegativity and boundedness results related to the porosity.

Afterwards, Chapter 6 introduces a finite element scheme for the coupled problem. We show that the approximation scheme is capable of preserving the nonnegativity and boundedness of the porosity in the continuous setting. Furthermore, the scheme can be fully decoupled, allowing for the parallel solution of the Stokes and Darcy subproblems. The chapter concludes with numerical experiments that are in agreement with the expected convergence rates.

The penultimate chapter of this thesis, Chapter 7, is concerned with the design of a filter through optimization. Specifically, we consider a multicriteria objective func-

tion that aims to maximize the lifetime of the filter and the amount of particulate captured throughout its lifetime. The decision variable is the initial porosity profile, and the constraints correspond to the system of equations describing the coupled Stokes-Darcy problem with deposition. Numerical experiments in 2D and 3D exploring linear and parabolic profiles are given at the end of this chapter.

Finally, we conclude this thesis in Chapter 8 summarizing the obtained results and discussing future avenues of research.

Chapter 2

Derivations

2.1 Stokes equations

The objective of this section is to derive a set of equations suitable to model a highly viscous or creeping flow. The main ingredient for the subsequent derivations is the continuity equation. This expression describes a local conservation law. Before we state this law, we introduce a few definitions.

Definition 2.1. *Let $\Omega \subset \mathbb{R}^d$ be a fixed bounded domain. Let Q denote a scalar or vector property that is being transported by a velocity vector field $\mathbf{u} : \Omega \to \mathbb{R}^d$. Furthermore, assume that the density of Q, i.e., amount of Q per unit volume at a given point in the flow is given by $\mathcal{Q}$. Then, we say that the flux of Q at $\mathbf{x} \in \mathbb{R}^d$ is $\mathbf{q}(\mathbf{x}) = \mathcal{Q}(\mathbf{x})\,\mathbf{u}(\mathbf{x})$. Note that the units of $\mathbf{q}$ are quantity of Q per unit area per unit time.*

In order to state a conservation law, we need to understand how the amount of a certain property can vary through time. We make this precise in the next definition.

Definition 2.2. *The rate of change of Q in Ω is defined as the sum of all the sources*

that generate Q in Ω, minus all the sinks that consume Q in Ω, plus a balance of incoming minus outgoing fluxes of Q across the boundary of Ω.

The mathematical statement corresponding to Definition 2.2, is given by the following equation.

Definition 2.3 (Continuity equation)**.** *Let f denote the net rate of consumption or generation of Q in Ω per unit volume and let $\mathbf{n}$ denote the outward unit normal vector to $\partial\Omega$. Then,*

$$\frac{\partial}{\partial t}\int_{\Omega} \mathcal{Q}\, d\Omega = \int_{\Omega} f\, d\Omega - \int_{\partial\Omega} \mathbf{q}\cdot\mathbf{n}\, dS.$$

Equivalently, assuming $\mathcal{Q}$ has enough regularity, we can exchange the integration and differentiation operator, apply the divergence theorem to the boundary integral [14], and take the limit as the volume of Ω tends to zero, to obtain the differential form

$$\frac{\partial \mathcal{Q}}{\partial t} + \nabla\cdot\mathbf{q} = f. \tag{2.1}$$

Letting $\mathcal{Q}$ in (2.1) denote mass per unit volume, i.e., density (ρ) and assuming there are no sinks or sources, yields the following expression for the conservation of mass.

Definition 2.4 (Conservation of mass)**.**

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(\rho\,\mathbf{u}) = 0 \text{ in } \Omega. \tag{2.2}$$

Remark 2.1. *In this work we assume that ρ is constant in space and time throughout*

Ω. *Consequently, from* (2.2)*, it follows that*

$$\nabla \cdot \mathbf{u} = 0 \text{ in } \Omega. \tag{2.3}$$

Whenever we say that that a fluid is incompressible, we mean that it satisfies (2.3)*.*

Now letting $\mathcal{Q}$ in (2.1) represent momentum per unit volume ($\rho\mathbf{u}$) and $\mathbf{h}$ the sources and sinks, yields the following expression for the conservation of momentum.

Definition 2.5 (Conservation of momentum)**.**

$$\frac{\partial(\rho\,\mathbf{u})}{\partial t} + \nabla \cdot (\rho\,\mathbf{u} \otimes \mathbf{u}) = \mathbf{h} \text{ in } \Omega, \tag{2.4}$$

where $\mathbf{u} \otimes \mathbf{u}$ *is a square matrix, whose entry* i, j *corresponds to the product* $u_i u_j$ *with* u_i *indicating the* i*-th component of the vector* $\mathbf{u}$*.*

Expanding the term $\nabla \cdot (\rho\,\mathbf{u} \otimes \mathbf{u})$ in (2.4) and using the conservation law (2.2), we obtain Newton's second law (the rate of change of momentum is directly proportional to the acting force)

$$\rho \frac{D\mathbf{u}}{Dt} = \mathbf{h}, \tag{2.5}$$

where the operator $\frac{D(\,\cdot\,)}{Dt}$ is the convective or substantial derivative

$$\frac{D(\,\cdot\,)}{Dt} = \frac{\partial(\,\cdot\,)}{\partial t} + \mathbf{u} \cdot \nabla(\,\cdot\,). \tag{2.6}$$

Before we characterize the force density $\mathbf{h}$ as a sum of surface and body forces, we introduce an important law in rheology [17].

Definition 2.6 (Newton's law of viscosity)**.** *Let u_i denote the i-th component of the velocity vector $\mathbf{u}$. Assume a layer or element of fluid is moving in the j-direction and rubs against a plane $\mathcal{P}$ normal to the i-direction. As this layer moves, a force per unit area in the j-direction is experienced on $\mathcal{P}$. This quantity is called a stress and is denoted by τ_{ij}. If for a given fluid there exists a proportionality constant μ (the dynamic viscosity) such that the expression*

$$\tau_{ij} = -\mu \left(\frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j} \right), \tag{2.7}$$

is satisfied, we say that the fluid is Newtonian.

Define $\mathbb{D}(\mathbf{u}) = \frac{1}{2}(\nabla \mathbf{u} + (\nabla \mathbf{u})^T)$. Then, the tensor $\boldsymbol{\tau} = (\tau_{ij}) = -2\mu\, \mathbb{D}(\mathbf{u})$ is called the viscous or deformation tensor.

We state a few remarks related to Newton's law of viscosity.

Remark 2.2. *The viscosity μ has SI units* Pa s *(pascal second), but sometimes is measured in centipoise (cP), where 1 cP is* 10^{-3} Pa s.

Remark 2.3. *Consider two static layers of fluid. If we put into motion the lower one, we would observe that due to the stress between these surfaces, the upper layer would also start to move. Thus, τ_{ij} can also be understood as the flux of j-momentum in the positive i-direction.*

Remark 2.4. *Let $a, b \in \mathbb{R}$ with $a < b$. The negative sign in (2.7) is due to the convention that τ_{ij} is the force per unit area in the j-direction exerted by the fluid located at a lower layer (a plane of the form $i = a$) on the fluid located at an upper layer (a plane of the form $i = b$). For example, in the setting described in Remark*

2.3, the velocity gradient $\frac{du_j}{dx_i}$ is negative. Thus, in view of the adopted convention the corresponding stress τ_{ij} is positive.

The object that captures the effect of surface forces is the stress tensor $\mathbb{T}$, which is a function of the velocity ($\mathbf{u}$) and the pressure (p). For incompressible Newtonian fluids, the type of fluids that we consider in this work, $\mathbb{T}$ is given by

$$\mathbb{T} = -2\mu\,\mathbb{D}(\mathbf{u}) + p\mathbb{I}, \tag{2.8}$$

where $\mathbb{I}$ is the identity tensor.

Notation 2.1. *If we relate the coordinates (x, y, z) to the numbers $(1, 2, 3)$, the notation $\mathbb{T}_{yx}$ indicates the $(2, 1)$-entry of the tensor $\mathbb{T}$. Moreover, the expression $\mathbb{T}_x$ denotes the column vector composed of the elements in the first column of $\mathbb{T}$. Any other combination of x, y, z is treated similarly.*

In view of Definition 2.6, we can characterize the force acting on a surface due to the stress tensor $\mathbb{T}$ through the following definition.

Definition 2.7. *Let $\mathbf{n}$ be a normal vector at a point q on a surface S. Then, the stress or traction vector $\mathbf{T}$ at q, which describes the force per unit area due to the friction of the fluid with S and the pressure is given by $\mathbf{T} = \mathbf{n} \cdot \mathbb{T} = \mathbb{T} \cdot \mathbf{n}$.*

Remark 2.5. *From Remark 2.3 it follows that the tensor $\mathbb{T}$ in (2.8) describes the combined effects of momentum flux due to friction between layers of fluid and the pressure to which it is subjected. Moreover, note that since the pressure only acts normal to a surface, the pressure tensor necessarily has to be of the form $p\mathbb{I}$.*

Now we return to the characterization of the force density function $\mathbf{h}$ in (2.5). Consider a fixed infinitesimal hexahedral element $\mathcal{C}$ with one vertex at the position

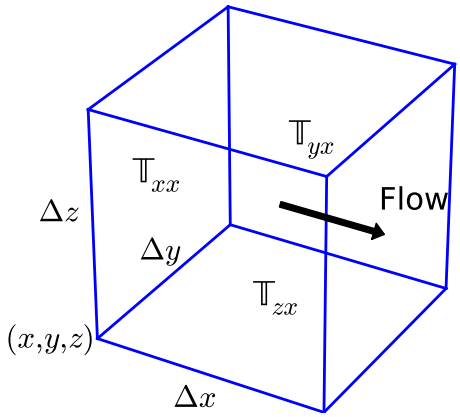


Figure 2.1: Balance of x-momentum across the faces of an infinitesimal cube. As the fluid flows in the x-direction, the pressure and the friction between the fluid particles generate stresses $\mathbb{T}_{xx}$, $\mathbb{T}_{yx}$ and $\mathbb{T}_{zx}$ on the x-, y-, and z-planes, respectively.

(x, y, z) and another one at $(x + \Delta x, y + \Delta y, z + \Delta z)$. We compute the contributions on $\mathcal{C}$ due to surface and body forces. A graphical depiction of $\mathcal{C}$ is given in Figure 2.1. We begin with the surface forces. First consider the balance of x-momentum across the faces of $\mathcal{C}$. Following the convention of Definition 2.2, i.e., incoming momentum flux minus outgoing momentum flux, we obtain

$$\begin{aligned}\text{Net flux} = &\left(\mathbb{T}_{xx}|_x - \mathbb{T}_{xx}|_{x+\Delta x}\right)\Delta y\,\Delta z + \left(\mathbb{T}_{yx}|_y - \mathbb{T}_{yx}|_{y+\Delta y}\right)\Delta x\,\Delta z \\ &+ \left(\mathbb{T}_{zx}|_z - \mathbb{T}_{zx}|_{z+\Delta z}\right)\Delta x\,\Delta y,\end{aligned}$$

which is equivalent to

$$\left(\frac{\mathbb{T}_{xx}|_x - \mathbb{T}_{xx}|_{x+\Delta x}}{\Delta x} + \frac{\mathbb{T}_{yx}|_y - \mathbb{T}_{yx}|_{y+\Delta y}}{\Delta y} + \frac{\mathbb{T}_{zx}|_z - \mathbb{T}_{zx}|_{z+\Delta z}}{\Delta z}\right)\Delta x\,\Delta y\,\Delta z. \tag{2.9}$$

Dividing (2.9) by $\Delta x\,\Delta y\,\Delta z$ and taking the limit of the volume of $\mathcal{C}$ to zero, yields the net x-momentum density contribution $-\nabla \cdot \mathbb{T}_x$. Proceeding analogously with the balance of y- and z-momentum, we obtain the function $-\nabla \cdot \mathbb{T}$, which should be

interpreted as a vector whose i-th component is the divergence of the i-th column of $\mathbb{T}$. Finally, letting $\rho\mathbf{f}$ denote the density of body forces in Ω, e.g., gravity times density, we can rewrite $\mathbf{h}$ in (2.5) to obtain

$$\mathbf{h} = -\nabla \cdot \mathbb{T} + \rho\mathbf{f} \text{ in } \Omega. \tag{2.10}$$

Expanding the definition of the substantial derivative (2.6) in (2.5) and replacing $\mathbf{h}$ with (2.10), yields the Cauchy momentum equation.

Definition 2.8 (Cauchy momentum equation)**.**

$$\rho\left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u}\right) + \nabla \cdot \mathbb{T} = \rho\mathbf{f} \text{ in } \Omega. \tag{2.11}$$

Furthermore, if we assume that the medium is isotropic, i.e., the viscosity and the density do not change in space, make use of the incompressibility condition (2.3), and expression (2.8) for the stress tensor, we obtain the Navier-Stokes equations.

Definition 2.9 (Navier-Stokes equations)**.**

$$\begin{aligned} \rho\left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u}\right) - \mu\Delta\mathbf{u} + \nabla p &= \rho\mathbf{f} \text{ in } \Omega, \\ \nabla \cdot \mathbf{u} &= 0 \text{ in } \Omega. \end{aligned} \tag{2.12}$$

What remains is to show that Navier-Stokes equations simplify when the viscous effects dominate the convective effects. For that purpose, we normalize (2.12) by introducing the dimensionless variables and operators

$$\widetilde{\mathbf{u}} = \frac{\mathbf{u}}{u_0},\ \widetilde{\mathbf{x}} = \frac{\mathbf{x}}{l_0},\ \widetilde{t} = \frac{t\,u_0}{l_0},\ \widetilde{p} = \frac{p}{\mu\,u_0\,l_0},\ \widetilde{\nabla} = l_0\,\nabla,\ \widetilde{\Delta} = l_0^2\,\Delta, \tag{2.13}$$

where u_0 and l_0 are scaling factors whose magnitude typically matches the average or maximum speed of the flow and the dimension of the domain, respectively. Substituting (2.13) into (2.12) and omitting the function $\mathbf{f}$, yields

$$\frac{\partial \widetilde{\mathbf{u}}}{\partial \widetilde{t}} + \widetilde{\mathbf{u}} \cdot \widetilde{\nabla} \widetilde{\mathbf{u}} - \frac{\mu}{\rho\, l_0\, u_0} \widetilde{\Delta} \widetilde{\mathbf{u}} + \frac{\mu}{\rho\, l_0\, u_0} \widetilde{\nabla} \widetilde{p} = \mathbf{0} \text{ in } \Omega,$$
$$\widetilde{\nabla} \cdot \widetilde{\mathbf{u}} = 0 \text{ in } \Omega. \tag{2.14}$$

Before we simplify (2.12), we introduce an important parameter that characterizes a flow.

Definition 2.10. *The dimensionless group*

$$\mathrm{Re} = \frac{\rho\, u_0\, l_0}{\mu}$$

is called the Reynolds number and describes the ratio between the inertial forces and the viscous forces.

In view of Definition 2.10 and (2.14), it follows that when $\mathrm{Re} \ll 1$ the time derivative and the nonlinear term in (2.12) are negligible with respect to the other terms. Hence, (2.12) simplifies to the Stokes equations.

Definition 2.11 (Stokes equations)**.**

$$-\mu \Delta \mathbf{u} + \nabla p = \rho \mathbf{f} \text{ in } \Omega,$$
$$\nabla \cdot \mathbf{u} = 0 \text{ in } \Omega. \tag{2.15}$$

Remark 2.6. *If one applies the same dimensional analysis to the Cauchy momentum equation* (2.11)*, one obtains an expression more general than* (2.15) *that requires the*

specification of the stress tensor. The system

$$\nabla \cdot \mathbb{T} = \rho \mathbf{f} \text{ in } \Omega,$$
$$\nabla \cdot \mathbf{u} = 0 \text{ in } \Omega, \tag{2.16}$$

is also called the Stokes equations.

In the next chapter we introduce and derive a set of equations that model the flow of a fluid in a porous, homogeneous medium.

2.2 Darcy equations

The theory of fluid flow through porous media is extensive [53, 11, 52, 2, 55, 70, 32, 33]. In this section we discuss the standard model in porous media flow, the Darcy equations

$$\mu \, \kappa^{-1} \, \mathbf{U} + \nabla p = \mathbf{f} \text{ in } \Omega, \tag{2.17}$$
$$\nabla \cdot \mathbf{U} = 0 \text{ in } \Omega, \tag{2.18}$$

subject to suitable boundary conditions. The vector function $\mathbf{U}$ denotes an averaged fluid velocity, $\mathbf{f}$ represents an external force density, μ is the dynamic viscosity of the fluid, κ is a constant that depends on the permeability of the medium and p is the pressure that the fluid experiences throughout the bounded domain Ω. The Darcy equations are named after the French engineer Henry Philibert Gaspard Darcy who studied laminar flow through porous media and whose research resulted in the equations that bear his name.

2.2.1 Darcy's experiment

We now describe the experiment that Darcy conducted in the city of Dijon, France, in 1855 [29]. A long vertical cylinder of cross section A, total length Δh and covered at the bottom with a net, is filled with some granulose material up to some height L. Then, an impermeable barrier is inserted and the cylinder is filled with water. Subsequently, the barrier is removed, the volume of water in the cylinder is kept constant through a separate mechanism and the volumetric flow rate Q at the bottom is measured. Under the aforementioned conditions Darcy found that

$$Q = K A \, \Delta h \, / L, \tag{2.19}$$

where K is a constant dependent on the porous medium which is called hydraulic conductivity in hydrogeology. Dividing both sides of (2.19) by A, one obtains the filtration speed $U = Q/A$, which should not be confused with the position dependent velocity throughout the porous medium. Further experimentation revealed that K varies inversely with the dynamic viscosity of the fluid under study and that for a randomly oriented cylinder the flow is mostly driven by two factors: (1) The weight per unit length of a column of fluid ($\rho \, \mathbf{g} \, A$), where $\mathbf{g}$ denotes the gravity vector, and (2) the pressure gradient (∇p). Thus, Darcy's law takes the general form

$$\mathbf{U} = -\frac{\kappa}{\mu} \left(\nabla p - \rho \, \mathbf{g} \right),$$

which is (2.17) with $\mathbf{f} = \rho \, \mathbf{g}$.

Now we briefly discuss some aspects related to the Darcy equations.

2.2.2 Porosity

Materials are considered porous when they have a percentage of voids or interconnected channels through which a fluid may flow. The porosity η, is a nondimensional quantity that results from the ratio of the volume of voids to the total volume [29]. For example, for sand and sandstone $\eta \approx 0.3$, for rocks like limestone and dolomite $\eta \in (0.01, 0.05)$, and for clays like unconsolidated marine red, $\eta \approx 0.9$ [71]. These examples fall inside the category of granulated rocks and the porosity is classified as interstitial porosity. In the case where the medium has a network of interconnected fractures not sealed by some deposit, one talks of fracture porosity as opposed to interstitial porosity.

2.2.3 Permeability

The units of permeability are length squared. The permeability is commonly expressed in darcys, where 1 darcy equals $0.987 \cdot 10^{-12}$ m^2. As a working definition, we have that for a flow of $Q = 1$ cm^3/s, through a cross section area $A = 1$ cm^2, for a fluid with viscosity $\mu = 1$ cP, and a pressure gradient of 760 mmHg/cm, the corresponding permeability of the medium, κ, is 1 darcy [29]. In the case of a nonisothermal flow, the viscosity is a function of temperature. For example, water at 20 °C has a viscosity $\mu \approx 1.002$ cP and a density $\rho \approx 1000$ Kg/m^3, which results in a speed of $0.966 \cdot 10^{-8}$ m/s for a medium with $\kappa = 1$ milidarcy. As temperature increases, the viscosity of water decreases and therefore the volumetric flow rate increases. Up to this point, we have assumed that the permeability κ is isotropic throughout the material, i.e., κ is independent of the orientation of space, which is in general not true. As an example, permeability anisotropy can be observed in sandstone, where overgrowth of thin layers of shale and quartz block most of the vertical flow, making horizontal

permeability much higher than the vertical permeability [7]. In order to take into account the dependency of permeability on orientation, κ is taken to be a symmetric tensor of rank 2, i.e.,

$$\kappa = \begin{pmatrix} \kappa_{xx} & \kappa_{xy} & \kappa_{xz} \\ \kappa_{yx} & \kappa_{yy} & \kappa_{yz} \\ \kappa_{zx} & \kappa_{zy} & \kappa_{zz} \end{pmatrix}, \text{ where } \kappa_{ij} = \kappa_{ji}.$$

There are a number of postulated models for the relationship between κ and η. The most commonly cited under isotropic conditions is the Blake-Kozeny model [15]

$$\kappa(\eta) \;=\; \frac{D_p^2\,\eta^3}{150\,(1-\eta)^2}\,, \tag{2.20}$$

where D_p represents a material constant – the diameter of the particles comprising the porous medium.

2.2.4 Microscopic vs. Macroscopic scale

Above, the Darcy law (2.17) was derived from experimental observations and therein we obtained the filtration velocity $\mathbf{U}$. This is a macroscopic quantity which is a statistical average of the microscopic velocity $\mathbf{u}$, which is the real fluid velocity in the pores of the medium. To indicate this relationship, we use the notation $\mathbf{U} = \langle \mathbf{u} \rangle$, where $\langle \cdot \rangle$ denotes an averaging procedure. The fact that the microscopic velocity $\mathbf{u}$ satisfies (2.18) follows from the conservation of mass and Remark 2.1. Thus, it remains to show that the macroscopic variable $\mathbf{U}$ satisfies the same condition. We follow the ideas described in [29]. In order to define a macroscopic variable, we need to integrate its microscopic equivalent in space multiplied by a suitable nonnegative weighting function. For that purpose, we give two definitions.

Definition 2.12. *We say that $w \in L^2(\Omega)$ is a weighting function if $w \geq 0$ a.e. and $\|w\|_{L^1(\Omega)} = 1$.*

We now formally define the $\langle \cdot \rangle$ operator.

Definition 2.13. *For $f \in L^2(\Omega)$, $\widetilde{f}$ its zero extension to $L^2(\mathbb{R}^d)$ and w a given weighting function, $\langle f \rangle$ is given by*

$$\langle f \rangle (\mathbf{x}) = \int_\Omega \widetilde{f}(\mathbf{x}+\mathbf{y}) w(\mathbf{y})\, d\mathbf{y}.$$

For the case where the function to average is a vector function, $\langle \cdot \rangle$ is defined componentwise. We proceed to prove (2.18).

Lemma 2.1. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain and let w be a weighting function. Let $\mathbf{u} \in L^2(\Omega)^d$ with $\dfrac{\partial u_i}{\partial x_i} \in L^1(\Omega)$ for $i = 1, \ldots, d$. Furthermore, assume $\nabla \cdot \mathbf{u} = 0$ almost everywhere. Then, the incompressibility condition $\nabla \cdot \langle \mathbf{u} \rangle = 0$ is also satisfied.*

Proof. From $\nabla \cdot \mathbf{u} = 0$, it follows that for $\mathbf{x} \in \Omega$

$$\int_\Omega \nabla_{\mathbf{x}} \cdot \widetilde{\mathbf{u}}(\mathbf{x}+\mathbf{y})\, w(\mathbf{y})\, d\mathbf{y} = 0.$$

Using the fact that Ω is fixed, we interchange the order of integration and differentiation (see [13]) to obtain

$$\nabla \cdot \langle \mathbf{u} \rangle (\mathbf{x}) = \nabla_{\mathbf{x}} \cdot \int_\Omega \widetilde{\mathbf{u}}(\mathbf{x}+\mathbf{y})\, w(\mathbf{y})\, d\mathbf{y} = 0,$$

proving the claim. □

For the rest of this document, whenever we refer to the Darcy velocity $\langle \mathbf{u} \rangle$, we use the notation $\mathbf{u}$.

In the next chapter we consider an extension of system (2.17)-(2.18) by including an equation which describes the temporal evolution of the porosity throughout the domain.

Chapter 3

Darcy flow with deposition

3.1 Introduction

In this chapter we consider a filtration process which can be modeled as a fluid flowing through a porous medium. We make the simplifying assumption that the rate of particulate deposition in the filter is only dependent on the porosity and the magnitude of the fluid velocity at that point. Of interest are the modeling equations

$$\mu\,\kappa^{-1}(\eta)\,\mathbf{u} + \nabla p = \mathbf{f} \text{ in } \Omega, \tag{3.1}$$

$$\nabla \cdot \mathbf{u} = 0 \text{ in } \Omega, \tag{3.2}$$

$$\frac{\partial \eta}{\partial t} + \mathrm{dep}(\mathbf{u}, \eta) = 0 \text{ in } \Omega, \tag{3.3}$$

subject to suitable boundary and initial conditions. In (3.1)-(3.2) $\mathbf{u}$ and p denote the velocity and pressure of the fluid, respectively, μ the dynamic viscosity and η and $\kappa(\eta)$ represent the porosity and permeability throughout the filter (Ω), respectively. The function $\mathrm{dep}(\,\cdot\,,\,\cdot\,)$, which we assume depends upon $\mathbf{u}$ and η, characterizes the rate of change of porosity through time. (A discussion of the model follows in Section

3.2).

The lack of regularity of the fluid velocity, $\mathbf{u} \in \mathbf{H}_{\text{div}}(\Omega)$, leads to an open question of the existence of a solution to (3.1)-(3.2). In order to obtain a modeling system of equations for which a solution can be shown to exist, we replace η in (3.1) by a *smoothed* porosity, η^s. The approach of regularizing the model with the introduction of η^s is, in part, motivated by the Darcy fluid flow equations, which can be derived by *averaging*, e.g., volume averaging [74], homogenization [3], mixture theory [65]. In [38], the case of (steady-state) generalized Newtonian fluid flow through a porous medium is considered using equations (3.1) and (3.2) with $\mu\,\kappa^{-1}(\eta)$ replaced by $\beta(|\mathbf{u}|) = \mu(|\mathbf{u}|)\,\kappa^{-1}$. Under the general assumptions that $\beta(\,\cdot\,)$ is a positive, bounded, Lipschitz continuous function, bounded away from zero, and with $\beta(|\mathbf{u}|)$ replaced with $\beta(|\mathbf{u}^s|)$, existence of a solution was established. Two smoothing operators for $\mathbf{u}$ were presented. One was a local averaging operator, whereby $\mathbf{u}^s(\mathbf{x})$ is obtained by averaging $\mathbf{u}$ in a neighborhood of $\mathbf{x}$. The second smoothing operator, which is nonlocal, computes $\mathbf{u}^s(\mathbf{x})$ using a differential filter applied to $\mathbf{u}$. That is, $\mathbf{u}^s$ is given by the solution to an elliptic differential equation whose right hand side is $\mathbf{u}$. For establishing the existence of a solution the key property that the smoothing operator needs to satisfy is that it transforms a weakly convergent sequence in $L^2(\Omega)$ into a sequence which converges strongly in $L^\infty(\Omega)$.

A similar model to (3.1)-(3.3) arises in the study of single-phase, miscible displacement of one fluid by another in a porous medium. For this problem η would denote a fluid concentration, and the hyperbolic deposition equation (3.3) is replaced by a parabolic transport equation. Existence and uniqueness for this problem has been investigated and established by Feng [41] and Chen and Ewing [27]. Because of the connection of this model to oil extraction, numerical approximation schemes for this problem have been well established. A summary of these methods is discussed

in the recent papers by Bartels, Jensen and Müller [12], and Riviére and Walkington [67].

A steady-state nonlinear Darcy fluid flow problem, with a permeability dependent on the pressure was investigated by Azaïez, Ben Belgacem, Bernardi, and Chorfi [8], and Girault, Murat, and Salgado [46]. For the permeability function Lipschitz continuous, and bounded above and below, existence of a solution $(\mathbf{u},\, p) \in L^2(\Omega)\times H^1(\Omega)$ was established. Important in handling the nonlinear permeability function, in establishing existence of a solution, was the property that $p \in H^1(\Omega)$. In [8] the authors also investigated a spectral numerical approximation scheme for the nonlinear Darcy problem, assuming an axisymmetric domain Ω. A convergence analysis for the finite element discretization of this problem was given in [46].

Following a discussion of the model in Section 3.2, existence of a solution to the modeling equations is established in Section 3.3. An approximation scheme for the filtration model is presented in Section 4.1, and an a priori error estimate is derived. A numerical simulation of the filtration process is presented in Section 4.2.

3.2 Discussion of Filtration Model

In this section we discuss the modeling equations we investigate for the filtration process. We assume that the process can be modeled as fluid flowing through a porous medium with a varying permeability. Additionally we assume that the process has a fixed time horizon, T. (For example, for industrial filters the most practical time to change filters is during scheduled maintenance periods. Drivers are encouraged to change the oil filters in their cars every 3000 miles or every three months, whichever comes first.) We use the following parameters/variables to model the process.
$\Omega \subset \mathbb{R}^d$ $(d = 2,\, 3)$ – the region occupied by the filter,

$\mathbf{u}$ – the fluid velocity,

p – the fluid pressure,

η – the porosity of the medium, $0 < \eta < 1$,

κ – the permeability of the medium, $0 < \kappa < \infty$,

μ – the fluid viscosity,

$\mathbf{n}$ – the unit outer normal to Ω.

We model the fluid flow using the Darcy fluid flow equations:

$$\mu\,\kappa^{-1}(\eta)\,\mathbf{u} + \nabla p = \mathbf{0} \text{ in } \Omega \times (0,T), \tag{3.4}$$

$$\nabla \cdot \mathbf{u} = 0 \text{ in } \Omega \times (0,T). \tag{3.5}$$

Assumption 3.1. *The particulate is sufficiently sparsely distributed in the fluid that the conservation of mass equation (3.5) is still a valid approximation for the model.*

Relationship between the permeability κ and the porosity η

As $\eta \to 0$ the "porous" medium transitions to a "solid" medium, in which case the permeability, $\kappa \to 0$. As $\eta \to 1$ the medium's resistance to the flow goes to zero, i.e., its permeability goes to infinity, and the modeling equations are no longer appropriate to describe the fluid flow.

Remark 3.1. *The permeability of granite is approximately* $10^{-3} - 10^{-4}$ *millidarcy. In a filtering process the permeability throughout* Ω *will always be greater than that of granite. So, it is reasonable to assume that* $\kappa(\eta)$ *is bounded from below. At the beginning of the filtering process there is a prescribed permeability (porosity) throughout* Ω*. As the filtering process decreases the permeability (porosity) throughout* Ω*, it is also reasonable to assume that* $\kappa(\eta)$ *is bounded from above.*

3.2.1 Modeling the deposition

The deposition on the particulate in the filter is modeled by an equation describing the change of porosity. We assume that $\partial\eta(\mathbf{x},t)/\partial t$ is a function of the magnitude of the velocity and the porosity, i.e.,

$$\frac{\partial \eta}{\partial t} + \mathrm{dep}(|\mathbf{u}|, \eta) = 0 \text{ in } \Omega \times (0,T). \tag{3.6}$$

As a first approximation, we assume that the deposition function $\mathrm{dep}(\cdot,\cdot)$ is a separable function of $|\mathbf{u}|$ and η,

$$\mathrm{dep}(|\mathbf{u}|, \eta) = g(|\mathbf{u}|)\, h(\eta) \text{ in } \Omega \times (0,T). \tag{3.7}$$

The function $g(\cdot)$

We assume that if the fluid is flowing too quickly there is little opportunity for the particles within the fluid to be captured by the filter. Therefore, beyond a critical value for $|\mathbf{u}|$, say s_c, $g(|\mathbf{u}|)$ is a monotonically decreasing function of $|\mathbf{u}|$. If $|\mathbf{u}|$ is very small then, given that we are modeling a sparsely distributed particulate in the fluid, the rate of deposition will also be very small due to the amount of particulate passing through the filter. In consideration of the about two situations, we postulate that $g(|\mathbf{u}|)$ is a Lipschitz continuous, non-negative, unimodal function, with maximum value occurring at $|\mathbf{u}| = s_c$.

The function $h(\cdot)$

We assume that as the porosity decreases the rate at which deposition occurs also decreases. This corresponds to the situation that as the deposition occurs (i.e., the porosity decreases) there is less of the filter available for the particulate to adhere to.

Based on this, we assume that $h(\eta)$ is a continuous, non-negative, increasing function. Two simple models for $h(\eta)$ are:

$$h(\eta) = \eta \implies \eta(\mathbf{x}, t) = \eta_0(\mathbf{x}) \exp(-\int_0^t g(|\mathbf{u}(\mathbf{x}, s)|)\, ds), \tag{3.8}$$

$$h(\eta) = \eta^r \ (r > 1) \implies \eta(\mathbf{x}, t) = \frac{\eta_0(\mathbf{x})}{\left(1 + (r-1)\,\eta_0(\mathbf{x})^{r-1} \int_0^t g(|\mathbf{u}(\mathbf{x}, s)|)\, ds\right)^{1/(r-1)}}, \tag{3.9}$$

where $\eta_0(\mathbf{x})$ denotes the initial porosity distribution throughout the filter.

3.2.2 Boundary conditions

We assume that the boundary of the filter, $\partial\Omega$, is made up of three parts: an inflow region, Γ_{in}, an outflow region, Γ_{out} and the "walls of the filter," Γ, i.e., $\partial\Omega = \overline{\Gamma}_{\text{in}} \cup \overline{\Gamma}_{\text{out}} \cup \overline{\Gamma}$. For well-posedness, equations (3.4)-(3.5) require a scalar boundary condition (typically $\mathbf{u} \cdot \mathbf{n}$ or p) be specified on $\partial\Omega$.

Inflow boundary condition

Two physically reasonable boundary conditions to consider on Γ_{in} are

$$-\mathbf{u} \cdot \mathbf{n} = g_{\text{in}} \text{ and} \tag{3.10}$$

$$p = p_{\text{in}}. \tag{3.11}$$

Condition (3.10) specifies the flow profile at the inflow boundary, whereas (3.11) corresponds to a specified pressure along the inflow.

Outflow boundary condition

The fluid outflow profile will be affected by the deposition occurring in the filter. Therefore specifying an outflow profile is not reasonable for this problem. Rather, at the outflow boundary we assume a specified pressure

$$p = p_{\text{out}} \text{ on } \Gamma_{\text{out}}. \tag{3.12}$$

Wall boundary condition

Along the walls of the filter we assume a no penetration condition, specifically

$$\mathbf{u} \cdot \mathbf{n} = 0 \text{ on } \Gamma. \tag{3.13}$$

For the mathematical analysis of this problem it is convenient to have homogeneous boundary conditions. This is achieved by introducing a suitable change of variables. For example, in case the specified boundary conditions are

$$-\mathbf{u} \cdot \mathbf{n} = g_{\text{in}} \text{ on } \Gamma_{\text{in}}, \quad \mathbf{u} \cdot \mathbf{n} = 0 \text{ on } \Gamma, \quad p = p_{\text{out}} \text{ on } \Gamma_{\text{out}},$$

we introduce functions $\mathbf{u}_b(\mathbf{x}, t)$ (see [44]) and $p_b(\mathbf{x}, t)$ defined on $\overline{\Omega}$ satisfying

$$\nabla \cdot \mathbf{u}_b = 0\,, \text{ in } \Omega \times (0, T]\,, \tag{3.14}$$

$$\mathbf{u}_b \cdot \mathbf{n} = -g_{\text{in}}\,, \text{ on } \Gamma_{\text{in}} \times (0, T]\,, \tag{3.15}$$

$$\mathbf{u}_b \cdot \mathbf{n} = \frac{1}{|\Gamma_{\text{out}}|} \int_{\Gamma_{\text{in}}} g_{\text{in}}(s)\, ds\,, \text{ on } \Gamma_{\text{out}} \times (0, T]\,, \tag{3.16}$$

$$\mathbf{u}_b \cdot \mathbf{t}_i = 0\,, \text{ on } \Gamma_{\text{in}} \times (0, T]\,, \tag{3.17}$$

$$\mathbf{u}_b = \mathbf{0}\,, \text{ on } \partial\Omega \backslash (\Gamma_{\text{in}} \cup \Gamma_{\text{out}}) \times (0, T]\,, \tag{3.18}$$

where $t_i,\ i = 1, \ldots, (d-1)$ denotes an orthogonal set of tangent vectors on Γ_{in}, and $|\Gamma_{\text{out}}|$ the measure of Γ_{out} with respect to ds, and

$$-\Delta p_b = 0\,,\ \text{in } \Omega \times (0,T]\,, \tag{3.19}$$

$$p_b = p_{\text{out}}\,,\ \text{on } \Gamma_{\text{out}} \times (0,T]\,, \tag{3.20}$$

$$\frac{\partial p_b}{\partial \mathbf{n}} = 0\,,\ \text{on } \partial\Omega \backslash \Gamma_{\text{out}} \times (0,T]\,. \tag{3.21}$$

(In case the pressure is specified on the inflow boundary Γ_{in} then $\mathbf{u}_b = \mathbf{0}$ and the definition of p_b is appropriately modified.)

With the change of variables: $\mathbf{u} = \mathbf{u}_a + \mathbf{u}_b$ and $p = p_a + p_b$ we obtain the following system of modeling equations for the filtration process:

$$\mu\,\kappa^{-1}(\eta)\,\mathbf{u}_a + \mu\,\kappa^{-1}(\eta)\,\mathbf{u}_b + \nabla p_a = -\nabla p_b\,,\ \text{in } \Omega \times (0,T]\,, \tag{3.22}$$

$$\nabla \cdot \mathbf{u}_a = 0\,,\ \text{in } \Omega \times (0,T]\,, \tag{3.23}$$

$$\frac{\partial \eta}{\partial t} + g(|\mathbf{u}_a + \mathbf{u}_b|)\,h(\eta) = 0\,,\ \text{in } \Omega \times (0,T]\,, \tag{3.24}$$

$$\mathbf{u}_a \cdot \mathbf{n} = 0\,,\ \text{on } \Gamma_{\text{in}} \cup \Gamma \times (0,T]\,, \tag{3.25}$$

$$p_a = 0\,,\ \text{on } \Gamma_{\text{out}} \times (0,T]\,, \tag{3.26}$$

$$\eta(\mathbf{x}, 0) = \eta_0(\mathbf{x}) \text{ in } \Omega\,. \tag{3.27}$$

To simplify the notation, we let $\mathbf{b} = \mathbf{u}_b$, $\mathbf{f} = -\nabla p_b$, $\beta(\eta) = \mu\,\kappa^{-1}(\eta)$, and drop the a subscript on $\mathbf{u}_a$ and p_a to obtain

$$\beta(\eta)\mathbf{u} + \beta(\eta)\mathbf{b} + \nabla p = \mathbf{f}\,,\ \text{in } \Omega \times (0,T]\,, \tag{3.28}$$

$$\nabla \cdot \mathbf{u} = 0\,,\ \text{in } \Omega \times (0,T]\,, \tag{3.29}$$

$$\frac{\partial \eta}{\partial t} + g(|\mathbf{u} + \mathbf{b}|)\,h(\eta) = 0\,,\ \text{in } \Omega \times (0,T]\,, \tag{3.30}$$

$$\mathbf{u} \cdot \mathbf{n} = 0\,,\ \text{on } \Gamma_{\text{in}} \cup \Gamma \times (0,T]\,, \tag{3.31}$$

$$p = 0\,,\ \text{on } \Gamma_{\text{out}} \times (0,T]\,, \tag{3.32}$$

$$\eta(\mathbf{x},0) = \eta_0(\mathbf{x}) \text{ in } \Omega\,. \tag{3.33}$$

Remark 3.2. *$\beta(\eta)$ is implicitly a function of $\mathbf{x}$ through the dependence of η on $\mathbf{x}$.*

In the next section we show that, under suitable assumptions on $\beta(\,\cdot\,)$, there exists a solution to (3.28)-(3.33).

3.3 Existence and Uniqueness

In this section we investigate the existence and uniqueness of solutions to the nonlinear system equations (3.28)-(3.33). We assume that $\Omega \subset \mathbb{R}^d$, $d = 2$ or 3, is a convex polyhedral domain and for vectors in $\mathbb{R}^d$, $|\,\cdot\,|$ denotes the Euclidean norm.

Weak formulation of (3.28)-(3.33)

Define the Hilbert spaces

$$\mathbf{H}_{\text{div}}(\Omega) = \left\{ \mathbf{v} \in L^2(\Omega) \,|\, \nabla \cdot \mathbf{v} \in L^2(\Omega) \right\},$$
$$X = \left\{ \mathbf{v} \in \mathbf{H}_{\text{div}}(\Omega) \,|\, \mathbf{v} \cdot \mathbf{n} = 0 \text{ on } \Gamma_{\text{in}} \cup \Gamma \right\}.$$

Define the bilinear form $b(\cdot, \cdot) : X \times L^2(\Omega) \to \mathbb{R}$ and the div-free subspace, Z, of X as

$$\begin{aligned} b(\mathbf{v}\,,\, q) &= \int_\Omega q \nabla \cdot \mathbf{v} \, d\Omega \,, \\ Z &= \left\{ \mathbf{v} \in X \,|\, b(\mathbf{v}\,,\, q) \,=\, 0 \,, \forall q \in L^2(\Omega) \right\}. \end{aligned}$$

We use

$$(f\,,\, g) \,=\, \int_\Omega f \cdot g \, d\Omega \,, \quad \text{and} \quad \|f\| \,=\, (f\,,\, f)^{1/2}$$

to denote the L^2 inner product and the L^2 norm over Ω, respectively, for both scalar and vector valued functions.

Additionally, we introduce the norm

$$\|\mathbf{v}\|_X = \left(\int_\Omega \left(\nabla \cdot \mathbf{v} \, \nabla \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{v} \right) d\Omega \right)^{1/2}. \tag{3.34}$$

Remark 3.3. *For $\mathbf{v} \in \mathbf{H}_{\text{div}}(\Omega)$ it follows that $\mathbf{v} \cdot \mathbf{n} \in H^{-1/2}(\partial\Omega)$. For the interpretation of the condition $\mathbf{v} \cdot \mathbf{n} \,=\, 0$ on $\Gamma_{\text{in}} \cup \Gamma$ see [45, 72].*

We make the following assumptions on $\beta(\,\cdot\,)$, $g(\,\cdot\,)$ and $h(\,\cdot\,)$.

Assumptions on $\beta(\,\cdot\,)$

Aβ1: $\beta(\,\cdot\,) : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$,

Aβ2: $0 \,<\, \beta_{\min} \,\leq\, \beta(s) \,\leq\, \beta_{\max}, \;\forall s \in \mathbb{R}^+$,

Aβ3: β is Lipschitz continuous, $|\beta(s_1) - \beta(s_2)| \leq \beta_{\text{Lip}} |s_1 - s_2|$.

Assumptions on $g(\cdot)$

Ag1: $g(\cdot) : \mathbb{R}^+ \cup \{0\} \longrightarrow \mathbb{R}^+ \cup \{0\}$,

Ag2: $g(s) \leq g_{\max}, \ \forall s \in \mathbb{R}^+ \cup \{0\}$,

Ag3: g is Lipschitz continuous, $|g(s_1) - g(s_2)| \leq g_{\text{Lip}} |s_1 - s_2|$.

Assumptions on $h(\cdot)$

Ah1: $h(\cdot) : \mathbb{R}^+ \longrightarrow \mathbb{R}^+ \cup \{0\}$,

Ah2: $h(s) \leq h_{\max}, \ \forall s \in \mathbb{R}^+$,

Ah3: h is Lipschitz continuous, $|h(s_1) - h(s_2)| \leq h_{\text{Lip}} |s_1 - s_2|$.

Remark 3.4. *The assumptions on β, g, and h are consistent with the discussion in Section 3.2.1. Note that $g_{\max}$, $h_{\max}$ should be given to guarantee positivity of the porosity ($0 < \eta$).*

Assumptions on η^s

Aη^s1: For $\eta(\cdot, t) \in L^2(\Omega)$, there exists a constant C_s, independent of t, such that $\|\eta^s(t)\|_{L^\infty(\Omega)} \leq C_s \|\eta(t)\|_{L^2(\Omega)}$,

Aη^s2: The mapping $\eta(\cdot, t) \mapsto \eta^s(\cdot, t)$ is linear.

Two smoothers which satisfy ***Aη^s1*** and ***Aη^s2*** are discussed in [38]. One is a local averaging operator and the other a differential smoothing operator.

We assume that $\mathbf{b}$ and $\mathbf{f}$ are continuous with respect to t on the interval $(0, T)$ and have a continuous extension to the interval $(0 - \delta, T)$ for some $\delta > 0$, i.e., $\mathbf{b}, \mathbf{f} \in C^0(0^-, T; L^2(\Omega))$.

We restate (3.28)-(3.33) as: *Given* $\eta_0(\mathbf{x}) \in L^2(\Omega)$, $\mathbf{b}, \mathbf{f} \in C^0(0^-, T; L^2(\Omega)) \cap$

$L^\infty(0,T;\, L^2(\Omega))$, *find* $(\mathbf{u},\, p) \in L^2(0,T;\, X) \times L^2(0,T;\, L^2(\Omega))$, $\eta \in H^1\ (0,T;\, L^2(\Omega))$ *such that for a.e.* t, $0 < t < T$,

$$
\begin{aligned}
(\beta(\eta^s)\mathbf{u}\,,\, \mathbf{v}) + (\beta(\eta^s)\mathbf{b}\,,\, \mathbf{v}) \ -\ b(\mathbf{v}\,,\, p) = &\qquad (\mathbf{f}\,,\, \mathbf{v})\ ,\ \forall \mathbf{v} \in X\,, &(3.35)\\
b(\mathbf{u}\,,\, q) = &\qquad 0\,,\ \forall q \in L^2(\Omega), &(3.36)\\
(\frac{\partial \eta}{\partial t}\,,\, \xi) + (g(|\mathbf{u} + \mathbf{b}|)\, h(\eta)\,,\, \xi) = &\qquad 0\,, \forall \xi \in L^2(\Omega)\,. &(3.37)
\end{aligned}
$$

For the spaces X and $L^2(\Omega)$ we have the following inf-sup condition

$$
\inf_{q \in L^2(\Omega)} \sup_{\mathbf{v} \in X} \frac{b(\mathbf{v}\,,\, q)}{\|q\|\, \|\mathbf{v}\|_X} \ \geq\ c_0 \ >\ 0\,. \tag{3.38}
$$

In view of the inf-sup condition we can restate (3.35)-(3.37) as: *Given* $\eta_0(\mathbf{x}) \in L^2(\Omega)$, $\mathbf{b}$, $\mathbf{f} \in C^0(0^-, T;\, L^2(\Omega)) \cap L^\infty(0,T;\, L^2(\Omega))$, *find* $\mathbf{u} \in L^2(0,T;\, Z)$, $\eta \in H^1(0,T;\, L^2(\Omega))$ *such that for a.e.* t, $0 < t < T$,

$$
\begin{aligned}
(\beta(\eta^s)\mathbf{u}\,,\, \mathbf{v}) + (\beta(\eta^s)\mathbf{b}\,,\, \mathbf{v}) \ &=\ (\mathbf{f}\,,\, \mathbf{v})\ ,\ \forall \mathbf{v} \in Z\,, &(3.39)\\
(\frac{\partial \eta}{\partial t}\,,\, \xi) + (g(|\mathbf{u} + \mathbf{b}|)\, h(\eta)\,,\, \xi) \ &=\ 0\,, \forall \xi \in L^2(\Omega)\,. &(3.40)
\end{aligned}
$$

Introduce the bounded Darcy projection operator: $\mathcal{P}_\eta : L^2(\Omega) \to Z$ defined by $\mathbf{z} = \mathcal{P}_\eta(\mathbf{r})$ where,

$$
\begin{aligned}
(\beta(\eta^s)\mathbf{z}\,,\, \mathbf{v}) \ -\ b(\mathbf{v}\,,\, \lambda) \ &=\ (\mathbf{r}\,,\, \mathbf{v})\,,\ \forall \mathbf{v} \in X\,,\\
b(\mathbf{z}\,,\, q)\,, \ &=\ 0\ ,\ \forall q \in L^2(\Omega)\,.
\end{aligned}
$$

That $\mathcal{P}_\eta$ is well defined follows from ***A*β1**-***A*β3**. Note that $\mathbf{u}$ in (3.39) may be

written as

$$\mathbf{u} = \mathcal{P}_\eta(\mathbf{f} - \beta(\eta^s)\mathbf{b}). \tag{3.41}$$

Additionally, from (3.39) with the choice $\mathbf{v} = \mathbf{u}$, it is straight forward to see that

$$\begin{aligned} \|\mathbf{u}(t)\| = \|\mathbf{u}(t)\|_X &\leq \frac{1}{\beta_{\min}}(\|\mathbf{f}(t)\| + \beta_{\max}\|\mathbf{b}(t)\|) \\ &\leq \frac{1}{\beta_{\min}}(\|\mathbf{f}\|_{L^\infty(0,T;\, L^2(\Omega))} + \beta_{\max}\|\mathbf{b}\|_{L^\infty(0,T;\, L^2(\Omega))})\,. \end{aligned} \tag{3.42}$$

Using $\mathcal{P}_\eta$ we can restate (3.39), (3.40) as:
Given $\eta_0(\mathbf{x}) \in L^2(\Omega)$, $\mathbf{b}, \mathbf{f} \in C^0(0^-, T;\, L^2(\Omega)) \cap L^\infty(0, T;\, L^2(\Omega))$, *find* $\eta \in H^1(0, T;\, L^2(\Omega))$ *such that for a.e.* t, $0 < t < T$,

$$(\frac{\partial \eta}{\partial t}\,,\, \xi) + (g(|\mathcal{P}_\eta(\mathbf{f} - \beta(\eta^s)\mathbf{b}) + \mathbf{b}|)\, h(\eta)\,,\, \xi) = 0\,, \forall \xi \in L^2(\Omega)\,. \tag{3.43}$$

We recall the Picard-Lindelöf theorem. (Also know as the Cauchy-Lipschitz theorem).

Theorem 3.1 ([49], Theorem I.3.1)**.** *Let I denote a domain in $\mathbb{R}$ containing the point t_0, Y a Banach space and $f : \mathbb{R} \times Y \longrightarrow Y$. Suppose that f is continuous in its first variable and locally Lipschitz continuous in its second variable. Then, there exists $\varepsilon > 0$ such that the initial value problem*

$$u' = f(t, u)\,, \tag{3.44}$$

$$u(t_0) = u_0\,, \tag{3.45}$$

has a unique solution in $C^0(t_0 - \varepsilon, t_0 + \varepsilon; Y)$.

Let $\mathcal{F}(t, \eta) = g(|\mathcal{P}_\eta(\mathbf{f} - \beta(\eta^s)\mathbf{b}) + \mathbf{b}|)\, h(\eta)$. The continuity of $\mathcal{F}$ with respect

to t follows from the continuity of $\mathbf{f}$ and $\mathbf{b}$ with respect to t, the boundedness of $\mathcal{P}_\eta$, and the continuity of g. To investigate the local Lipschitz continuity of $\mathcal{F}$ with respect to η consider the following.

$$\begin{aligned}
\|\mathcal{F}(t,\eta_1) \;-\; \mathcal{F}(t,\eta_2)\| &= \|g(|\mathcal{P}_{\eta_1}(\mathbf{f} - \beta(\eta_1^s)\mathbf{b}) + \mathbf{b}|)\, h(\eta_1) \\
&\qquad -g(|\mathcal{P}_{\eta_2}(\mathbf{f} - \beta(\eta_2^s)\mathbf{b}) + \mathbf{b}|)\, h(\eta_2)\| \qquad (3.46)\\
&\le \|g(|\mathcal{P}_{\eta_1}(\mathbf{f} - \beta(\eta_1^s)\mathbf{b}) + \mathbf{b}|)\,(h(\eta_1) - h(\eta_2))\| \\
&\quad +\|(g(|\mathcal{P}_{\eta_1}(\mathbf{f} - \beta(\eta_1^s)\mathbf{b}) + \mathbf{b}|) - g(|\mathcal{P}_{\eta_2}(\mathbf{f} - \beta(\eta_2^s)\mathbf{b}) + \mathbf{b}|)\, h(\eta_2)\| \\
&\le g_{\max}\, \|h(\eta_1) - h(\eta_2)\| \\
&\quad + g_{\mathrm{Lip}}\|(|\mathcal{P}_{\eta_1}(\mathbf{f} - \beta(\eta_1^s)\mathbf{b}) + \mathbf{b}| - |\mathcal{P}_{\eta_2}(\mathbf{f} - \beta(\eta_2^s)\mathbf{b}) + \mathbf{b}|)\, h(\eta_2)\| \\
&\le g_{\max}\, h_{\mathrm{Lip}}\, \|\eta_1 - \eta_2\| \\
&\quad + g_{\mathrm{Lip}}\| \left(\mathcal{P}_{\eta_1}(\mathbf{f} - \beta(\eta_1^s)\mathbf{b}) - \mathcal{P}_{\eta_2}(\mathbf{f} - \beta(\eta_2^s)\mathbf{b})\right)\, h(\eta_2)\| \\
&\le g_{\max}\, h_{\mathrm{Lip}}\, \|\eta_1 - \eta_2\| \\
&\quad + g_{\mathrm{Lip}}\, h_{\max}\|\mathcal{P}_{\eta_1}(\mathbf{f} - \beta(\eta_1^s)\mathbf{b}) - \mathcal{P}_{\eta_2}(\mathbf{f} - \beta(\eta_2^s)\mathbf{b})\|\,. \qquad (3.47)
\end{aligned}$$

Now, with $\mathbf{u}_1 = \mathcal{P}_{\eta_1}(\mathbf{f} - \beta(\eta_1^s)\mathbf{b}) \in Z$ and $\mathbf{u}_2 = \mathcal{P}_{\eta_2}(\mathbf{f} - \beta(\eta_2^s)\mathbf{b}) \in Z$ we have that

$$(\beta(\eta_1^s)\mathbf{u}_1\,,\,\mathbf{v}) \;=\; (\mathbf{f}\,,\mathbf{v}) \;-\; (\beta(\eta_1^s)\mathbf{b}\,,\,\mathbf{v})\;,\; \forall \mathbf{v} \in Z, \qquad (3.48)$$

$$\text{and}\quad (\beta(\eta_2^s)\mathbf{u}_2\,,\,\mathbf{v}) \;=\; (\mathbf{f}\,,\mathbf{v}) \;-\; (\beta(\eta_2^s)\mathbf{b}\,,\,\mathbf{v})\;,\; \forall \mathbf{v} \in Z. \qquad (3.49)$$

Subtracting (3.49) from (3.48), and with the choice $\mathbf{v} = \mathbf{u}_1 - \mathbf{u}_2$, yields

$$\begin{aligned}
(\beta(\eta_1^s)(\mathbf{u}_1 - \mathbf{u}_2)\,,\,(\mathbf{u}_1 - \mathbf{u}_2)) &= ((\beta(\eta_2^s) - \beta(\eta_1^s))\mathbf{u}_2\,,\,(\mathbf{u}_1 - \mathbf{u}_2)) \\
&\quad + ((\beta(\eta_2^s) - \beta(\eta_1^s))\mathbf{b}\,,\,(\mathbf{u}_1 - \mathbf{u}_2))\,.
\end{aligned}$$

Hence, with (3.42),

$$\begin{aligned}
\beta_{\min}\|\mathbf{u}_1 - \mathbf{u}_2\| &\leq \beta_{\mathrm{Lip}}\| \, |\eta_2^s - \eta_1^s| \, \mathbf{u}_2\| + \beta_{\mathrm{Lip}}\| \, |\eta_2^s - \eta_1^s| \, \mathbf{b}\| \\
&\leq \beta_{\mathrm{Lip}} \left(\|\mathbf{u}_2\| + \|\mathbf{b}\|\right) \, \|\eta_2^s - \eta_1^s\|_\infty \\
&\leq \beta_{\mathrm{Lip}} \left(\frac{1}{\beta_{\min}}\|\mathbf{f}\| + \frac{\beta_{\max}}{\beta_{\min}}\|\mathbf{b}\| + \|\mathbf{b}\| \right) \, C_s \, \|\eta_2 - \eta_1\| \, .
\end{aligned}$$

Thus we obtain

$$\begin{aligned}
\|\mathcal{P}_{\eta_1}(\mathbf{f} \, - \, \beta(\eta_1^s)\mathbf{b}) - \mathcal{P}_{\eta_2}(\mathbf{f} \, - \, \beta(\eta_2^s)\mathbf{b})\| &= \|\mathbf{u}_1 - \mathbf{u}_2\| \\
&\leq \, C_s \, \frac{\beta_{\mathrm{Lip}}}{\beta_{\min}^2} \left(\|\mathbf{f}\| + (\beta_{\max} + \beta_{\min})\|\mathbf{b}\|\right) \, \|\eta_2 \, - \, \eta_1\|. \qquad (3.50)
\end{aligned}$$

Combining (3.47) and (3.50), we have

$$\|\mathcal{F}(t, \eta_1) \, - \, \mathcal{F}(t, \eta_2)\| \, \leq \, \mathcal{F}_{\mathrm{Lip}} \, \|\eta_2 \, - \, \eta_1\|,$$

where $\mathcal{F}_{\mathrm{Lip}} = \mathcal{F}_1 + \mathcal{F}_2$ with $\mathcal{F}_1 = g_{\max} \, h_{\mathrm{Lip}}$ and

$$\mathcal{F}_2 = C_s \, \frac{\beta_{\mathrm{Lip}}}{\beta_{\min}^2} \, g_{\mathrm{Lip}} \, h_{\max} \left(\|\mathbf{f}\|_{L^\infty(0,T;\, L^2(\Omega))} + (\beta_{\max} + \beta_{\min})\|\mathbf{b}\|_{L^\infty(0,T;\, L^2(\Omega))} \right) .$$

Then, from Theorem 3.1, we have that there exists $\varepsilon > 0$ such that there exists a unique solution $\eta \in C^0(0, \varepsilon;\, L^2(\Omega))$ to (3.43).

Regarding the additional regularity of η, formally taking ξ equal to $\partial\eta/\partial t$ in (3.43) we have

$$\begin{aligned}
\|\frac{\partial \eta}{\partial t}\|^2 &\leq \|g(|\mathcal{P}_\eta(\mathbf{f} \, - \, \beta(\eta^s)\mathbf{b}) \, + \, \mathbf{b}|) \, h(\eta)\| \, \|\frac{\partial \eta}{\partial t}\| \\
\Rightarrow \quad \|\frac{\partial \eta}{\partial t}\| &\leq g_{\max} \, h_{\max} \, |\Omega| \, . \qquad (3.51)
\end{aligned}$$

Using the established fact that $\eta \in C^0(0,\varepsilon;\, L^2(\Omega))$, it then follows that $\eta \in H^1(0,\varepsilon;\, L^2(\Omega))$. In order to establish this result rigorously, we consider a Galerkin approximation of (3.43) in which the approximation of $\partial\eta/\partial t$ does indeed lie in $L^2(0,\varepsilon;\, L^2(\Omega))$ and then taking the limit.

Next, note that $\|\mathcal{F}(t,\eta)\| \leq g_{\max}\, h_{\max}\, |\Omega|^{1/2}$. Hence both $\mathcal{F}(t,\eta)$ and its Lipschitz constant with respect to η are bounded independent of t and η. Then, from the proof of Theorem 3.1 (see [49]), ε may be chosen such that $0 < \varepsilon < 1/\mathcal{F}_{\text{Lip}}$. As $\varepsilon > 0$ can be chosen independent of t and η, the solution can be extended to $0 < t < T$.

We summarize the above discussion in the following theorem.

Theorem 3.2. *Let $\eta_0(\mathbf{x}) \in L^2(\Omega)$ and $\mathbf{b}, \mathbf{f} \in C^0(0^-, T;\, L^2(\Omega)) \cap L^\infty(0,T;\, L^2(\Omega))$ be given. Then, for $\beta(\,\cdot\,)$ satisfying $\boldsymbol{A\beta 1}$–$\boldsymbol{A\beta 3}$, $g(\,\cdot\,)$ satisfying $\boldsymbol{Ag1}$–$\boldsymbol{Ag3}$, $h(\,\cdot\,)$ satisfying $\boldsymbol{Ah1}$–$\boldsymbol{Ah3}$, $\eta(\,\cdot\,)$ satisfying $\boldsymbol{A\eta^s 1}$–$\boldsymbol{A\eta^s 2}$, there exists a unique solution $(\mathbf{u},\, p) \in L^2(0,T;\, X) \times L^2(0,T;\, L^2(\Omega))$, $\eta \in H^1(0,T;\, L^2(\Omega))$ satisfying (3.35)-(3.37), for a.e. t, $0 < t < T$.*

Remark 3.5. *Important in establishing the existence and uniqueness of the solution is the assumption that $\beta(\mathbf{x},t) \geq \beta_{\min}$. After a finite amount of time, assuming that the mathematical equations correctly model the physical problem, the filter starts to clog and this assumption is violated.*

In the next chapter we explore the numerical approximation of the solution to (3.35)-(3.37).

Chapter 4

Approximation of Darcy flow with deposition

4.1 Finite element approximation

In this section we investigate the finite element approximation to:

Given $\eta_0(\mathbf{x}) \in L^2(\Omega)$, $\mathbf{b}, \mathbf{f} \in C^0(0^-, T; L^2(\Omega)) \cap L^\infty(0, T; L^2(\Omega))$, *find* $(\mathbf{u}, p) \in L^2(0, T; X) \times L^2(0, T; L^2(\Omega))$, $\eta \in H^1(0, T; L^2(\Omega))$ *such that for a.e.* t, $0 < t < T$,

$$(\beta(\eta^s)\mathbf{u}\,,\,\mathbf{v}) \;+\; (\beta(\eta^s)\mathbf{b}\,,\,\mathbf{v}) \;-\; b(\mathbf{v}\,,\,p) \;=\; (\mathbf{f}\,,\,\mathbf{v})\;,\;\forall \mathbf{v} \in X\,, \tag{4.1}$$

$$b(\mathbf{u}\,,\,q) \;=\; 0\,,\;\forall q \in L^2(\Omega), \tag{4.2}$$

$$(\frac{\partial \eta}{\partial t}\,,\,\xi) \;+\; (g(|\mathbf{u} + \mathbf{b}|)\,\eta\,,\,\xi) \;=\; 0\,,\forall \xi \in L^2(\Omega)\,. \tag{4.3}$$

Note that with regard to the general modeling equations (3.28)-(3.33), here we have chosen $h(\eta) = \eta$.

As before, let $\Omega \subset \mathbb{R}^d$ denote a convex polygonal or polyhedral domain and let $\mathcal{T}_h$ be

a triangulation of Ω made either of triangles or quadrilaterals in $\mathbb{R}^2$ or tetrahedra or hexahedra in $\mathbb{R}^3$. Thus, the computational domain is defined by $\Omega = \bigcup_{K\in\mathcal{T}_h} K$. We assume that there exist constants c_1, c_2 such that $c_1 h \le h_K \le c_2 \rho_K$, where h_K is the diameter of the cell K, ρ_K is the diameter of the biggest neighborhood included in K, and $h = \max_{K\in\mathcal{T}_h} h_k$. For $k \in \mathbb{N}$, let

$$\mathbb{P}_k^T = \text{span}\{x_1^{\alpha_1} x_2^{\alpha_2} \dots x_d^{\alpha_d} \,:\, 0 \le \alpha_1 + \alpha_2 + \cdots + \alpha_d \le k\}, \text{ and} \tag{4.4}$$

$$\mathbb{P}_k^Q = \text{span}\{x_1^{\alpha_1} x_2^{\alpha_2} \dots x_d^{\alpha_d} \,:\, 0 \le \alpha_1,\, \alpha_2,\, \dots,\, \alpha_d \le k\}. \tag{4.5}$$

For K a triangle/tetrahedron in $\mathbb{R}^d$ we let $P_k(K) = \{f \,:\, f|_K \in \mathbb{P}_k^T\}$. For K a quadrilateral/hexahedron in $\mathbb{R}^d$ we let $P_k(K) = \{f \,:\, f|_K \in \mathbb{P}_k^Q\}$. $RT_k(\mathcal{T}_h)$ is used the denote the Raviart-Thomas space of order k [18]. We use the following finite element spaces:

$$\begin{aligned}
X_h &= RT_k(\mathcal{T}_h) \cap X, \quad Q_h = \left\{q \in L^2(\Omega) \,:\, q|_K \in P_k(K),\, \forall K \in \mathcal{T}_h\right\},\\
Z_h &= \left\{\mathbf{v} \in X_h \,:\, (q, \mathbf{v}) = 0,\, \forall q \in Q_h\right\},\\
R_h &= \left\{r \in L^2(\Omega) \,:\, r|_K \in P_m(K),\, \forall K \in \mathcal{T}_h\right\},\\
R_h^s &= \left\{r \in C^0(\Omega) \,:\, r|_K \in P_{\max\{1,m\}}(K),\, \forall K \in \mathcal{T}_h\right\}.
\end{aligned}$$

Note that as $\nabla\cdot X_h = Q_h$, for $\mathbf{v} \in Z_h$ we have that $\|\nabla\cdot\mathbf{v}\| = 0$, thus $\|\mathbf{v}\|_X = \|\mathbf{v}\|$ (see (3.34)). For N given, let $\Delta t = T/N$, and $t_n = n\Delta t$, $n = 0, 1, \dots, N$. Additionally, define

$$d_t f^n = \frac{f^n - f^{n-1}}{\Delta t}, \quad \overline{f}^n = \frac{f^n + f^{n-1}}{2}, \quad \widetilde{f}^n = f^{n-1} + \frac{1}{2} f^{n-2} - \frac{1}{2} f^{n-3}.$$

The following norms are used in the analysis

$$\|v\|_\infty = \|v(t)\|_{L^\infty(\Omega)},\ |||v|||_k = \left(\sum_{n=0}^{N} \|v(t_n)\|_k^2 \Delta t\right)^{1/2},\ |||v|||_\infty = \sup_{0\le n\le N} \|v(t_n)\|_\infty.$$

For the *a priori* error estimates presented below the solution $(\mathbf{u}, p, \eta)$ to (4.1)-(4.3) is required to be *sufficiently regular*. The regularity assumptions are, for some $\delta > 0$,

$$\begin{aligned}
&\mathbf{u} \in L^\infty(0,T;L^\infty(\Omega)) \cap L^\infty(0,T;H^{k+1}(\Omega)), \quad \mathbf{u}_t \in L^\infty(0,\delta;L^2(\Omega)),\\
&\mathbf{u}_{tt} \in L^2(0,T;L^2(\Omega)), \quad p \in L^\infty(0,T;H^{k+1}(\Omega)),\\
&\eta \in L^\infty(0,T;L^\infty(\Omega)) \cap L^\infty(0,T;H^{m+1}(\Omega)), \quad \eta_t \in L^\infty(0,T;H^{m+1}(\Omega)),\\
&\eta_{tt} \in L^2(0,T;L^2(\Omega)) \cap L^\infty(0,\delta;L^2(\Omega)), \quad \eta_{ttt} \in L^2(0,T;L^2(\Omega)).
\end{aligned} \tag{4.6}$$

Throughout, we use C to denote a generic nonnegative constant, independent of the mesh parameter h and time step Δt, whose actual value may change from line to line in the analysis.

Initialization of the Approximation Scheme

The approximation scheme described and analyzed below is a three-level scheme. To initialize the procedure suitable approximations are required for $\mathbf{u}_h^1$, $\mathbf{u}_h^2$ and η_h^2. Here we state our assumptions on these initial approximates. (An initialization procedure is presented in Appendix A.2)

$$\|\mathbf{u}^n - \mathbf{u}_h^n\|_X^2 \ +\ \|\eta^n - \eta_h^n\|^2 \ \le\ C(\Delta t)^4 \ +\ C\left(h^{2k+2} \ +\ h^{2m+2}\right),\ \text{ for } n = 0,1,2. \tag{4.7}$$

Approximation Scheme

The approximation scheme we investigate is: *Given* $\eta_0 \in R_h$, *for* $n = 3,\dots,N$,

determine $(\mathbf{u}_h^n, p_h^n, \eta_h^n) \in X_h \times Q_h \times R_h$ *satisfying*

$$(\beta(\eta_h^{n,s})\mathbf{u}_h^n + \beta(\eta_h^{n,s})\mathbf{b}^n, \mathbf{v}) - (p_h^n, \nabla \cdot \mathbf{v}) = (\mathbf{f}^n, \mathbf{v}) \; \forall \mathbf{v} \in X_h, \tag{4.8}$$

$$(q, \nabla \cdot \mathbf{u}_h^n) = 0 \; \forall q \in Q_h, \tag{4.9}$$

$$(d_t \eta_h^n, r) + (g(|\widetilde{\mathbf{u}}_h^n + \mathbf{b}^{n-1/2}|)\overline{\eta}_h^n, r) = 0 \; \forall r \in R_h, \tag{4.10}$$

where $\mathbf{b}^{n-1/2} = \mathbf{b}(\frac{t^n + t^{n-1}}{2})$.

Regarding $\eta_h^{n,s}$, note that applying a smoother, $\mathcal{S}$, to a function $\eta_h^n \in R_h$ (typically) does not result in $\eta_h^{n,s} \in R_h^s$. Therefore, we let $\mathcal{S}(\eta_h^n) \in H^{m+1}(\Omega) \cap C^0(\Omega)$ denote the result of the smoother applied to η_h^n, and define

$$\eta_h^{n,s}(x) \;=\; I_h \mathcal{S}(\eta_h^n)(x)\,, \tag{4.11}$$

where $I_h : C^0(\Omega) \longrightarrow R_h^s$ denotes an interpolation operator.

We assume that the smoothed porosity $\mathcal{S}(\eta_h^n)$ is sufficiently regular such that there exists a constant dependent on $\mathcal{S}(\,\cdot\,)$, $C_{\mathcal{S}}$, such that

$$\|\mathcal{S}(\eta_h^n) \,-\, I_h \mathcal{S}(\eta_h^n)\|_{L^\infty(\Omega)} \;=\; \|\mathcal{S}(\eta_h^n) \,-\, \eta_h^{n,s}\|_{L^\infty(\Omega)} \;\leq\; C_{\mathcal{S}} h^{m+1}\,. \tag{4.12}$$

The precise dependence of $C_{\mathcal{S}}$ on $\mathcal{S}(\,\cdot\,)$ will depend on the particular smoother used.

The computability of the approximation scheme (4.8)-(4.10) is established in the following lemma.

Lemma 4.1. *There exists a unique solution* $(\mathbf{u}_h^n, p_h^n, \eta_h^n) \in (X_h, Q_h, R_h)$ *satisfying (4.8)-(4.10).*

Proof. For $\{\phi_j\}_{j=1}^{N_R}$ a basis for R_h, and $\eta_h^n = \sum_{j=1}^{N_R} c_j \phi_j$, equation (4.10) is equivalent

to $A\mathbf{c} \;=\; \mathbf{d}$, where $\mathbf{c} = [c_1, c_2, \ldots, c_{N_R}]^T$, and for $i, j = 1, 2, \ldots, N_R$,

$$A_{ij} = \left(\left(\frac{1}{\Delta t} + \frac{1}{2} g(|\widetilde{\mathbf{u}}_h^n + \mathbf{b}^{n-1/2}|) \right) \phi_j, \phi_i \right), \quad \text{and}$$
$$\mathbf{d}_i = \int_\Omega \left(\frac{1}{\Delta t} - \frac{1}{2} g(|\widetilde{\mathbf{u}}_h^n + \mathbf{b}^{n-1/2}|) \right) \eta_h^{n-1} \, \phi_i \, d\Omega.$$

Note that as $A\mathbf{c} = \mathbf{0} \;\; \Rightarrow \;\; \mathbf{c}^T A \mathbf{c} = 0$

$$\Leftrightarrow \;\; \int_\Omega \left(\frac{1}{\Delta t} + \frac{1}{2} g(|\widetilde{\mathbf{u}}_h^n + \mathbf{b}^{n-1/2}|) \right) \eta_h^n \, \eta_h^n d\Omega \;=\; 0 \Rightarrow \eta_h^n \;=\; 0$$
$$\Rightarrow \mathbf{c} \;=\; 0\,,$$

it follows that the (square) linear system (4.10) has a unique solution for η_h^n.

Given η_h^n and assumption $\boldsymbol{A\beta 2}$, the existence and uniqueness of $(\mathbf{u}_h^n, p_h^n) \in X_h \times Q_h$ is well known from the approximation theory of the Darcy equations. $\square$

Theorem 4.1. *For $(\mathbf{u}, p, \eta)$ satisfying (4.1)-(4.3) and (4.6) and $(\mathbf{u}_h^n, p_h^n, \eta_h^n)$ satisfying (4.8)-(4.10), and assuming that $C_{\mathcal{S}(\eta_h^n)}$ given in (4.12) is bounded by $C_{\mathcal{S}} \|\eta^n\|_{m+1}$, we have that for Δt sufficiently small there exists $C > 0$ independent of h and Δt, such that for $n = 1, 2, \ldots, N$,*

$$\|\mathbf{u}^n - \mathbf{u}_h^n\|_X \;+\; \|p^n - p_h^n\| \;+\; \|\eta^n - \eta_h^n\| \;\leq\; C\left((\Delta t)^2 \;+\; h^{k+1} \;+\; h^{m+1}\right). \quad (4.13)$$

Outline of the proof. The complete details of the proof are given in Appendix A.1. Here we briefly summarize the key steps in the proof.

Step 0. Notation.

For $\mathcal{U}^n, \tau^n$ in Z_h and R_h, respectively, let

$$\begin{aligned} \mathbf{\Lambda}^n &= \mathbf{u}^n - \mathcal{U}^n, & \mathbf{E}^n &= \mathcal{U}^n - \mathbf{u}_h^n, \\ \psi^n &= \eta^n - \tau^n, & F^n &= \tau^n - \eta_h^n, \\ \varepsilon_{\mathbf{u}} &= \mathbf{u}^n - \mathbf{u}_h^n, & \varepsilon_\eta &= \eta^n - \eta_h^n. \end{aligned} \tag{4.14}$$

Step 1. Derive an estimate for $\|\mathbf{E}^n\|$.

Beginning with (4.8) and (3.35) we obtain

$$\beta_{\min}\|\mathbf{E}^n\| \le \beta_{\text{Lip}}\|\mathbf{u}^n + \mathbf{b}^n\|_\infty \|\eta_h^{n,s} - \eta^{n,s}\| + \beta_{\max}\|\mathbf{\Lambda}^n\|. \tag{4.15}$$

Step 2. Estimate $\|\eta_h^{n,s} - \eta^{n,s}\|$.

With the triangle inequality,

$$\begin{aligned} \|\eta_h^{n,s} - \eta^{n,s}\| &\le \|\eta_h^{n,s} - \mathcal{S}(\eta_h^n)\| + \|\mathcal{S}(\eta_h^n) - \eta^{n,s}\| \\ &\le \ \|\eta_h^{n,s} - \mathcal{S}(\eta_h^n)\| + |\Omega|^{1/2} C_s \|\eta_h^n - \eta^n\| \\ &\le \|\eta_h^{n,s} - \mathcal{S}(\eta_h^n)\| + |\Omega|^{1/2} C_s (\|\psi^n\| + \|F^n\|). \end{aligned} \tag{4.16}$$

Step 3. Derive an estimate for $\|F^n\|^2 - \|F^{n-1}\|^2$.

Beginning with (4.10) and (3.37) we obtain

$$\begin{aligned} \|F^n\|^2 - \|F^{n-1}\|^2 \le{}& \Delta t \|d_t \psi^n\|^2 + 2\Delta t g_{\text{Lip}}^2 \|\overline{\eta}^n\|_\infty^2 (\|\widetilde{\mathbf{E}}^n\|^2 + \|\widetilde{\mathbf{\Lambda}}^n\|^2) \\ &+ 2\Delta t g_{\max}^2 \|\overline{\psi}^n\|^2 + \Delta t (6 + 2g_{\max}^2) \|\overline{F}^n\|^2 + \Delta t R^n(\mathbf{u}, \eta), \end{aligned} \tag{4.17}$$

where

$$R^n(\mathbf{u},\eta) = \|d_t\eta^n - \frac{\partial \eta^{n-1/2}}{\partial t}\|^2 + g_{\mathrm{Lip}}^2 \|\overline{\eta}^n\|_\infty^2 \|\widetilde{\mathbf{u}}^n - \mathbf{u}^{n-1/2}\|^2 + g_{\max}^2 \|\overline{\eta}^n - \eta^{n-1/2}\|^2. \tag{4.18}$$

Step 4. Derive a bound for $\|F^\ell\|^2$.

Summing (4.17) from $n = 3$ to $n = l$, and using the discrete Gronwall's lemma [50, 60], we obtain, with constants C, $C_{\mathcal{S}}$, w_1, w_2, $K(w_3)$, w_4, w_5,

$$\begin{aligned}\|F^\ell\|^2 \le K(w_3)\Big(& w_1 C h^{2k+2} |||\mathbf{u}|||_{k+1}^2 + (w_2 C + w_4 C_{\mathcal{S}}^2) h^{2m+2} |||\eta|||_{m+1}^2 \\ & + C h^{2m+2} \int_{t_2}^{t_\ell} \|\eta_t\|_{m+1}^2 \, dt + (\Delta t)^4 \, w_5 \int_0^{t_{\ell-2}} \|\mathbf{u}_{tt}\|^2 \, dt + \|F^2\|^2 \\ & + (\Delta t)^4 \, \frac{g_{\max}^2}{48} \int_{t_2}^{t_\ell} \|\eta_{tt}\|^2 \, dt + \frac{(\Delta t)^4}{1280} \int_{t_2}^{t_\ell} \|\eta_{ttt}\|^2 \, dt \Big).\end{aligned} \tag{4.19}$$

Step 5. Derive a bound for $\|\mathbf{E}^\ell\|^2$.

Combining (4.19) with (4.16) and (4.15), we obtain a bound for $\|\mathbf{E}^\ell\|^2$.

Step 6. Error bounds for $\|\mathbf{u}^\ell - \mathbf{u}_h^\ell\|^2$ and $\|\eta^\ell - \eta_h^\ell\|^2$.

The error bounds for $\|\mathbf{u}^\ell - \mathbf{u}_h^\ell\|^2$ and $\|\eta^\ell - \eta_h^\ell\|^2$ then follow from the bounds for $\|\mathbf{E}^\ell\|^2$, $\|F^\ell\|^2$, assumption (4.7), and using

$$\|\mathbf{u}^\ell - \mathbf{u}_h^\ell\|^2 \le 2(\|\mathbf{E}^\ell\|^2 + \|\mathbf{\Lambda}^\ell\|^2), \ \|\eta^\ell - \eta_h^\ell\|^2 \le 2(\|F^\ell\|^2 + \|\psi^\ell\|^2).$$

Step 7. Error bound for $\|p^n - p_h^n\|$.

The error bound for $\|p^n - p_h^n\|$ is obtained by using the discrete inf-sup condition

$$0 < \gamma < \inf_{q_h \in Q_h} \sup_{\mathbf{v}_h \in X_h} \frac{(q_h, \nabla \cdot \mathbf{v}_h)}{\|q_h\| \, \|\mathbf{v}_h\|_X}$$

to show that for an arbitrary element $\mathcal{Q}^n \in Q_h$,

$$\begin{aligned}\gamma \, \|p_h^n - \mathcal{Q}^n\| &\leq \sup_{\mathbf{v}_h \in X_h} \frac{(p_h^n - \mathcal{Q}^n, \nabla \cdot \mathbf{v}_h)}{\|\mathbf{v}_h\|_X} \\ &\leq \beta_{\max} \|\mathbf{u}_h^n - \mathbf{u}^n\| + \beta_{\mathrm{Lip}} \|\mathbf{u}^n + \mathbf{b}^n\|_\infty \|\eta_h^{n,s} - \eta^{n,s}\| + \|p^n - \mathcal{Q}^n\|.\end{aligned}$$

The estimate in (4.13) then follows from interpolation properties of Q_h and the triangle inequality,

$$\|p^n - p_h^n\| \leq \|p^n - \mathcal{Q}^n\| + \|p_h^n - \mathcal{Q}^n\|.$$

□

4.2 Numerical examples

In this section we present four numerical examples to illustrate the numerical approximation scheme (4.8)-(4.10). Example 1 and Example 2, for which we have an exact solution, investigate the derived a priori error estimate for the approximation (4.13), and the dependence of the approximation on the smoother. Example 3 and Example 4 use the numerical approximation scheme to investigate the performance of several filters.

For the numerical implementation of the scheme (4.8)-(4.10), the C++ finite element library `deal.II` [10] was used. In the 2D setting (Example 1 and Example 2) the domain is partitioned into quadrilaterals. In the 3D setting (Example 3 and Example 4) the domain is partitioned into hexahedra. We let $\mathrm{disc}P_k = \{f : f|_K \in \mathbb{P}_k^Q, \forall K \in \mathcal{T}_h\}$, and $\mathrm{cont}P_k = \{f \in C^0(\Omega) : f|_K \in \mathbb{P}_k^Q, \forall K \in \mathcal{T}_h\}$ (see (4.5)).

Example 1 and Example 2

We consider $\Omega = (-1, 1) \times (0, 1)$ and approximate (3.35)-(3.37) for $t \in (0, T]$ with $T = 0.5$. The true solution for the velocity and pressure is given by

$$\mathbf{u} = \begin{bmatrix} txy - t^2y^2 \\ tx + t^2x^2 - ty^2/2 \end{bmatrix}, \; p(x, y) = 2t^2x - ty^2.$$

The function g in the definition of the deposition function is set to $g(|\mathbf{u}|) = |\mathbf{u}|^2 + 1$. Assuming that the error in the numerical approximations is of order $\mathcal{O}(\Delta t^2 + h^{k+1})$ (see (4.13)), we chose $(\Delta t)^2 \propto h^{k+1}$.

For a function f and its approximations, $f^n_{h_1}$, $f^n_{h_2}$, computed on partitions of Ω with mesh parameters h_1 and h_2, we define the numerical convergence rate $r_{\|\cdot\|}$ as:

$$r_{\|\cdot\|} := \frac{\log(\|f(N\Delta t) - f^N_{h_1}\|/\|f(N\Delta t) - f^N_{h_2}\|)}{\log(h_1/h_2)}.$$

The quantity $r_{|||\cdot|||}$ is defined similarly.

Two different smoothers are investigated. For the first one, we compute the smoothed porosity η^s using a local averaging procedure. Specifically,

$$\eta^s(\mathbf{x}) = \frac{1}{|V(\mathbf{x})|} \int_{V(\mathbf{x})} \eta(\mathbf{x}) \, d\Omega,$$

where $|V(\mathbf{x})| = \delta$ denotes the area (volume) of the averaging domain $V(x)$.

The second smoother is the differential smoother

$$-\delta \Delta \eta^s + \eta^s = \eta \text{ in } \Omega$$

$$\eta^s = \eta \text{ on } \partial\Omega.$$

For both smoothers we set $\delta = 0.05$.

Example 1.

In this example we define $\eta(x, y) = 0.8 - 0.5t^2xy$ and $\beta(\eta) = \eta^2 + 0.1$. Computations using the local averaging smoother, with $(\mathbf{u}_h, p_h, \eta_h, \eta_h^s) \in (RT_0, \text{disc}P_0, \text{disc}P_0, \text{cont}P_1)$, and $(\mathbf{u}_h, p_h, \eta_h, \eta_h^s) \in (RT_1, \text{disc}P_1, \text{disc}P_1, \text{cont}P_1)$, in the $\|\cdot\|$ norm are presented in Table 4.1, and Table 4.2, respectively. Similar results are also obtained using the $|||\cdot|||$ norm, and when using the differential smoother.

The results indicate that the numerical convergence rates are consistent with those predicted by Theorem 4.1.

h	Δt	$\|\mathbf{u} - \mathbf{u}_h\|_X$	$r_{\|\cdot\|}$	$\|p - p_h\|$	$r_{\|\cdot\|}$	$\|\eta - \eta_h\|$	$r_{\|\cdot\|}$
1/2	2^{-3}	1.241E-01	1.22	1.176E-01	1.31	1.387E-03	0.86
1/4	$2^{-7/2}$	5.346E-02	0.78	4.759E-02	0.69	7.664E-04	0.49
1/8	2^{-4}	3.124E-02	1.05	2.946E-02	1.07	5.474E-04	1.08
1/16	$2^{-9/2}$	1.507E-02	0.95	1.403E-02	0.93	2.583E-04	0.81
1/32	2^{-5}	7.814E-03	1.05	7.366E-03	1.07	1.473E-04	1.19
1/64	$2^{-11/2}$	3.770E-03	-	3.507E-03	-	6.453E-05	-
Predicted convergence rate (see (4.13)): 1							

Table 4.1: Example 1: Convergence rates for $(\mathbf{u}_h, p_h, \eta_h, \eta_h^s) \in (RT_0, \text{disc}P_0, \text{disc}P_0, \text{cont}P_1)$. The norm is computed at the final time $t = T$.

h	Δt	$\|\mathbf{u} - \mathbf{u}_h\|_X$	$r_{\|\cdot\|}$	$\|p - p_h\|$	$r_{\|\cdot\|}$	$\|\eta - \eta_h\|$	$r_{\|\cdot\|}$
1	2^{-3}	1.529E-02	1.55	9.882E-03	1.78	8.720E-04	0.99
1/2	2^{-4}	5.209E-03	1.80	2.882E-03	1.90	4.379E-04	1.78
1/4	2^{-5}	1.495E-03	1.91	7.721E-04	1.95	1.279E-04	1.91
1/8	2^{-6}	3.990E-04	1.95	1.995E-04	1.98	3.402E-05	1.96
1/16	2^{-7}	1.030E-04	-	5.067E-05	-	8.743E-06	-
Predicted convergence rate (see (4.13)): 2							

Table 4.2: Example 1: Convergence rates for $(\mathbf{u}_h, p_h, \eta_h, \eta_h^s) \in (RT_1, \text{disc}P_1, \text{disc}P_1, \text{cont}P_1)$. The norm is computed at the final time $t = T$.

Example 2.

In this example we demonstrate the importance of smoothing the porosity input to the $\beta(\,\cdot\,)$ function. Using a perturbed porosity field $\eta = 0.8 - 0.5t^2xy + 0.03125\sin(169x)\cos(169y)$ (see Figure 4.1), and letting $\beta(\,\cdot\,)$ given by (see Figure 4.2)

$$\beta(s) = \begin{cases} 9.1, & 0 \le s < 0.1 \\ -8.5s + 9.95, & 0.1 \le s < 0.3 \\ 7.4, & 0.3 \le s < 0.4 \\ -8.5s + 10.8, & 0.4 \le s < 0.6 \\ 5.7, & 0.6 \le s < 0.7 \\ -18.5s + 18.65, & 0.7 \le s < 0.9 \\ 2.0, & 0.9 \le s \le 1.0, \end{cases}$$

we study the convergence rates under two different conditions. First, without smoothing the porosity input to the $\beta(\,\cdot\,)$ function (see Table 4.3), and then with a smoothed porosity input to $\beta(\,\cdot\,)$ (see Table 4.4).

In this example the differential smoother is applied with $\delta = 0.05$.

$(\mathbf{u}_h,\, p_h,\, \eta_h) \in (RT_0,\, \mathrm{disc}P_0,\, \mathrm{disc}P_0)$							
h	Δt	$\|\mathbf{u} - \mathbf{u}_h\|_X$	$r_{\|\cdot\|}$	$\|p - p_h\|$	$r_{\|\cdot\|}$	$\|\eta - \eta_h\|$	$r_{\|\cdot\|}$
1/2	2^{-3}	1.674E-01	1.31	1.756E-01	1.43	1.976E-03	1.08
1/4	$2^{-7/2}$	6.742E-02	0.70	6.534E-02	0.84	9.324E-04	-0.79
1/8	2^{-4}	4.153E-02	1.06	3.640E-02	1.15	1.612E-03	0.66
1/16	$2^{-9/2}$	1.994E-02	0.31	1.636E-02	0.99	1.017E-03	-0.10
1/32	2^{-5}	1.611E-02	-	8.210E-03	-	1.091E-03	-
Convergence is not guaranteed by the theory.							

Table 4.3: Example 2: Convergence rates without smoothing the porosity. The norm is computed at the final time $t = T$.

The results in Table 4.3 indicate that without smoothing, the convergence rate

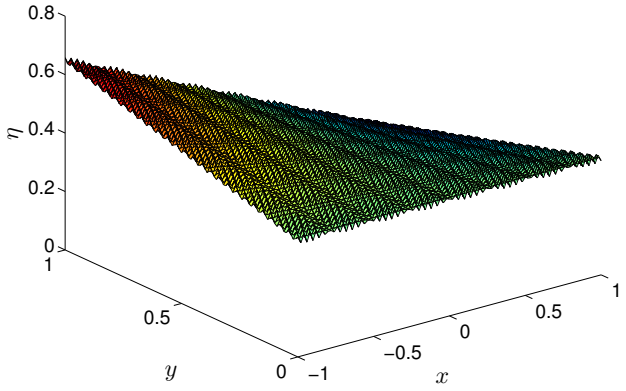


Figure 4.1: Porosity field for Ex. 2 at time $t = 0.5$.

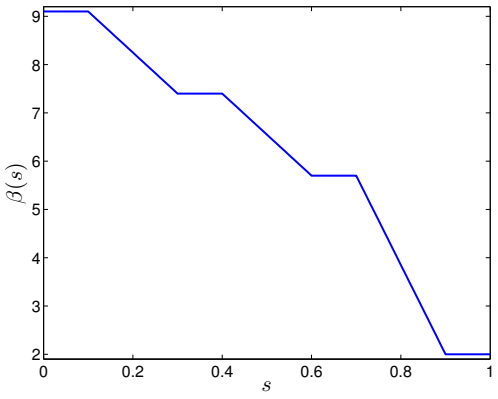


Figure 4.2: Plot of $\beta(s)$, Ex.2.

$(\mathbf{u}_h,\, p_h,\, \eta_h,\, \eta_h^s) \in (RT_0,\, \mathrm{disc}P_0,\, \mathrm{disc}P_0,\, \mathrm{cont}P_1)$							
h	Δt	$\|\mathbf{u}-\mathbf{u}_h\|_X$	$r_{\|\cdot\|}$	$\|p-p_h\|$	$r_{\|\cdot\|}$	$\|\eta-\eta_h\|$	$r_{\|\cdot\|}$
1/2	2^{-3}	1.671E-01	1.30	1.759E-01	1.42	2.202E-03	1.46
1/4	$2^{-7/2}$	6.786E-02	0.88	6.562E-02	0.85	8.013E-04	0.32
1/8	2^{-4}	3.691E-02	1.08	3.638E-02	1.15	6.410E-04	0.24
1/16	$2^{-9/2}$	1.746E-02	0.96	1.640E-02	0.99	5.432E-04	0.74
1/32	2^{-5}	8.980E-03	-	8.235E-03	-	3.259E-04	-
Predicted convergence rate: 1							

Table 4.4: Example 2: Convergence rates when the porosity is smoothed. The norm is computed at the final time $t = T$.

of the velocity drops drastically and the porosity does not converge. In contrast, using the smoothed porosity as input to $\beta(\,\cdot\,)$, the obtained approximations are convergent (see Table 4.4).

Example 3 and Example 4.

We consider $\Omega = (-1,1) \times (0,1) \times (0,1)$ and approximate (4.1)-(4.3) using (4.8)-(4.10), for $t \in (0,1]$. No flux boundary conditions, $\mathbf{u} \cdot \mathbf{n} = 0$, are imposed on the walls $x = -1$, $x = 1$, $y = -1$, $y = 1$, and a zero pressure condition, $p = 0$, on the

outflow boundary $z = -1$. Four filters, labelled I-IV, with different initial porosity profiles are investigated. All filters share the same initial non void space, $\nu(0)$, where

$$\nu(t) = \int_{\Omega} (1 - \eta(\mathbf{x}, t))\, d\Omega.$$

The chosen computational parameters are $\Delta t = 2^{-5}$ and $h = 0.1$. The initial porosity profiles in Filters I-IV (see Figures 4.3 - 4.6) are:

I: The porosity is uniformly distributed throughout the domain Ω.

II: The porosity increases radially in a continuous fashion.

III: The porosity decreases radially in a continuous fashion.

IV: The porosity decreases continuously in the positive z-direction.

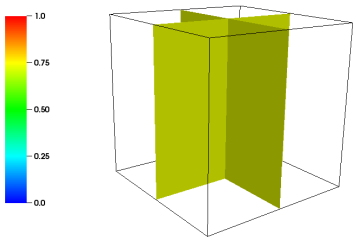


Figure 4.3: Filter I: Initial porosity field.

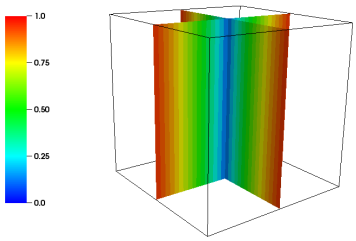


Figure 4.4: Filter II: Initial porosity field.

Example 3.

In this example we specify the inflow velocity for the fluid. Namely, $\mathbf{u} \cdot \mathbf{n} = -f$ on $z = 1$, where

$$f(\mathbf{x}, t) = (1 - x^2)(1 - y^2) \min\{1, 4t\}.$$

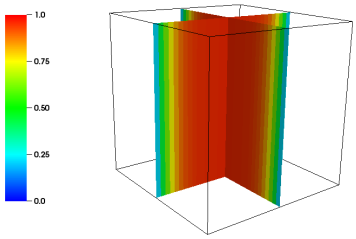


Figure 4.5: Filter III: Initial porosity field.

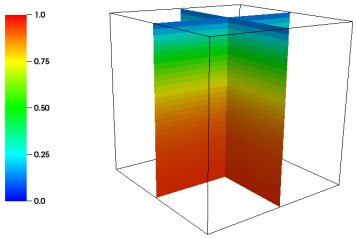


Figure 4.6: Filter IV: Initial porosity field.

The non void space and maximum pressure within each filter is measured at the time $T = 1$. The results are presented in Table 4.5. Plots of the final porosity fields for each filter are presented in Figure 4.7 - Figure 4.10.

Filter	**Nonvoid space** $\nu(1)$	**Max. pressure**
I	5.34	7.40
II	5.41	11.00
III	5.50	7.03
IV	5.33	14.07

Table 4.5: Example 3: Non void space $\nu(t)$, and maximum pressure within each filter at $T = 1$.

In terms of the particulate deposited, all the filters' performances are similar with a difference between filters of less than 2.5%. However, there is a significant difference in terms of the maximum pressure $p_{\max}$ attained within each filter at $T = 1$. The maximum value $p_{\max} = 14.07$ is obtained by Filter IV and the minimum value $p_{\max} = 7.03$ is obtained by Filter III.

Example 4.

In this example the inflow pressure, p_{in}, is specified. The non void space within the filter and total amount of fluid that went through the filter is measured at the

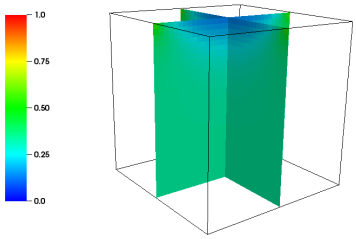


Figure 4.7: Filter I: Porosity field at $T = 1$.

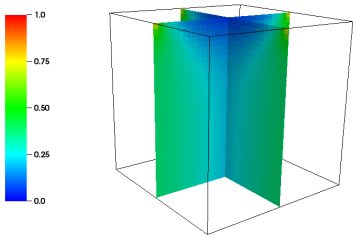


Figure 4.8: Filter II: Porosity field at $T = 1$.

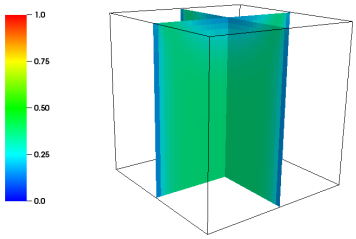


Figure 4.9: Filter III: Porosity field at $T = 1$.

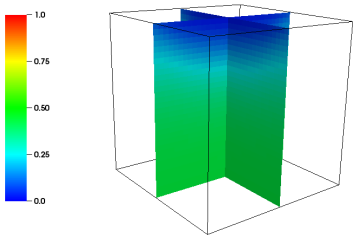


Figure 4.10: Filter IV: Porosity field at $T = 1$.

time $T = 1$. The inflow pressure is given by

$$p(\mathbf{x}, t) = 10 \min\{1, 4t\}.$$

The results are summarized in Table 4.6.

Filter	**Non void space** $\nu(1)$	**Total flow**
I	6.61	5.74
II	6.64	5.65
III	6.66	5.56
IV	6.34	5.02

Table 4.6: Example 4: Non void space at $T = 1$, and total flow from $t = 0$ to $t = T$.

Similar to Example 3, the deposition within each of the filters is comparable,

differing by less than 6%. The total flow, however, differs by more than 10% with the highest total flow of 5.74 corresponding to Filter I and the lowest total flow of 5.02 corresponding to Filter IV.

In the next chapter we couple the Stokes equations (2.15) to the Darcy with deposition equations (3.1)-(3.3) and prove the existence and uniqueness of a weak solution to the resulting system of equations.

Chapter 5

Coupled Stokes-Darcy flow with deposition

5.1 Introduction

In this chapter we are interested in an application related to purification that builds upon the theory developed in the previous chapter. A concrete description of the problem we wish to study follows. Some fluid (sewage or industrial wastewaters) coming from an originally stagnant body of water, such as an activated sludge tank, advects some pollutant in the form of particulate at a low Reynolds regime inside a processing plant. Therein, the first unitary operation that is carried out to purify the liquid is filtration. The liquid is forced to pass through sand beds by means of a pressure gradient and is later recollected for further processing. As the time progresses, the filter, i.e., the sand bed, clogs and consequently less fluid is allowed to pass until eventually a certain efficiency threshold is attained, which requires the filter to be replaced or washed, interrupting the whole operation.

In this description it is evident that we need a model to describe the nature of the

filtration process in the porous region $\Omega_1 \subset \mathbb{R}^d$. A natural choice is to use the Darcy equations, whose mixed formulation reads:

$$\beta(\eta)\,\mathbf{u}_1 + \nabla p_1 = \mathbf{f}_1 \text{ in } \Omega_1,$$
$$\nabla \cdot \mathbf{u}_1 = 0 \text{ in } \Omega_1,$$

where $\mathbf{u}_1$ and p_1 represent the Darcy velocity and pressure, respectively, η is the porosity, $\mathbf{f}_1$ is an external force, and β denotes the (scaled) inverse permeability tensor which depends on η and the dynamic viscosity in Ω_1, μ_1, which we assume constant. However, similar to what happens in slow sand filters where a thin layer grows in the surface of the filter as time progresses, our equations should also take into account the temporal evolution of the properties of the filter such as porosity. For that matter, one needs an equation describing the deposition of the particulate in the filter, which couples the Darcy flow with the deposition:

$$\frac{\partial \eta}{\partial t} + \mathrm{dep}(\mathbf{u}_1, \eta) = 0 \text{ in } \Omega_1.$$

Assuming the deposition term separates into two functions, one depending on the magnitude of the velocity and the other on the porosity, we obtain

$$\frac{\partial \eta}{\partial t} + g(|\mathbf{u}_1|)h(\eta) = 0 \text{ in } \Omega_1.$$

An analysis for the system of equations

$$\begin{aligned} \beta(\eta^s)\mathbf{u}_1 + \nabla p_1 &= \mathbf{f}_1 \text{ in } \Omega_1, \\ \nabla \cdot \mathbf{u}_1 &= 0 \text{ in } \Omega_1, \\ \frac{\partial \eta}{\partial t} + g(|\mathbf{u}_1|)h(\eta) &= 0 \text{ in } \Omega_1, \end{aligned} \tag{5.1}$$

with β, g and h satisfying suitable regularity and boundedness conditions, and η^s a *smoothed* porosity, is thoroughly discussed in Chapter 3. The reason for using a smoothed or averaged porosity is twofold: It confers additional regularity critical to the analysis and is consistent with the derivation of the Darcy velocity, which results from an averaging or homogenization procedure [4]. In this work we continue this practice and replace $\beta(\eta)$ with $\beta(\eta^s)$.

The previously introduced system of equations can be used to model the flow for the filtration part of the problem. Nonetheless, there remains to consider the flow from the storage unit (denoted here by $\Omega_2 \subset \mathbb{R}^d$) to the filtration unit Ω_1. Based on the theory derived in Section 2.1, for small Reynolds numbers, i.e., viscous terms dominating inertial terms, the Stokes equations (2.16) are applicable and thus, we use them to describe the incoming flow to the filter. The modeling equations inside the transition domain that connects the storage unit to the filtration unit are:

$$\begin{aligned} \nabla \cdot \mathbb{T}(\mathbf{u}_2, p_2) &= \mathbf{f}_2 \text{ in } \Omega_2, \\ \nabla \cdot \mathbf{u}_2 &= 0 \text{ in } \Omega_2, \end{aligned} \tag{5.2}$$

where $\mathbf{u}_2$ is the Stokes velocity, p_2 is the Stokes pressure, $\mathbb{T}$ is the stress tensor given in (2.8) and $\mathbf{f}_2$ is a forcing term.

Once it has been established that we have to deal with two different flow regimes, in an

attempt to greatly simplify the model, one could assume that these two flow conditions can be treated independently. First solve Stokes, and then, use the resulting outflow conditions as the inflow for the Darcy model. Though simple, this idea is incorrect as it suggests that the Stokes domain influences the Darcy domain without being itself affected by changes in the latter. For example, as the particulate deposits in a specific region in Ω_1, the porosity of the medium diminishes and consequently the upstream Stokes fluid velocity changes. Moreover, without the careful choice of adequate interface constraints relating the velocity and the pressure of each domain, the physical significance of the model would be lost. Thus, the crucial and also most challenging part to study is the interface Γ connecting Ω_2 with Ω_1, i.e., $\Gamma = \overline{\Omega}_1 \cap \overline{\Omega}_2$. Let $\mathbf{n}$ denote an outward unit normal vector and define $\mathbf{n}|_\Gamma$ to point from the Stokes boundary into the Darcy boundary. Systems (5.1) and (5.2) are coupled through the equations on the interface

$$(\mathbf{u}_1 - \mathbf{u}_2) \cdot \mathbf{n} = 0 \text{ on } \Gamma, \tag{5.3}$$

$$\mathbf{n} \cdot \mathbb{T} \cdot \mathbf{n} = p_1 \text{ on } \Gamma, \tag{5.4}$$

which describe the continuity of the normal components of the velocity and the balance of the normal forces, respectively. It is also necessary to specify the tangential component of the traction vector to correctly describe the physics of the problem. One alternative is to use the Beavers-Joseph condition which relates the tangential stress on the Stokes domain to the difference between the Stokes and Darcy velocity [64]. Mathematically,

$$P_{\mathbf{t}}(\mathbb{T} \cdot \mathbf{n}) = \Psi(\eta)\, P_{\mathbf{t}}(\mathbf{u}_2 - \mathbf{u}_1) \text{ on } \Gamma, \tag{5.5}$$

where $P_{\mathbf{t}}(\mathbf{v}(p))$ is the projection of $\mathbf{v}$ onto the tangent plane at the point p on Γ. The projection operator $P_{\mathbf{t}} : \mathbb{R}^d \to \mathbb{R}^d$ is defined in the usual way

$$P_{\mathbf{t}}(\mathbf{v}) = \sum_{m=1}^{d-1} (\mathbf{v} \cdot \mathbf{t}_m)\mathbf{t}_m,$$

where $\{\mathbf{t}_i\}_{i=1}^{d-1}$ is a local orthonormal basis for the tangent space at p. The proportionality function $\Psi(\,\cdot\,)$ is given by

$$\Psi(\eta) = \frac{\alpha\, \mu_2 \sqrt{d}}{\rho_2 \sqrt{\text{trace}(\beta^{-1}(\eta))}}, \tag{5.6}$$

where $\alpha \in \mathbb{R}^+$ is a constant determined experimentally, and ρ_2, μ_2 are the density and the dynamic viscosity of the fluid in Ω_2, respectively. In this work, however, we use the Beavers-Joseph-Saffmann condition

$$P_{\mathbf{t}}(\mathbb{T} \cdot \mathbf{n}) = \Psi(\eta)\, P_{\mathbf{t}}(\mathbf{u}_2) \text{ on } \Gamma, \tag{5.7}$$

which relates the tangential stress to only the tangential component of the Stokes velocity. The use of (5.7) for our problem is justified based on experimentation and its derivation through statistical methods [69]. The mathematical correctness of (5.7) was established under the conditions described in [16] and the limiting case when the pore size tends to zero [54].

Remark 5.1. *An additional coupling between the Stokes and Darcy regions arises from the fact that the porosity η appears in both* (5.1) *and* (5.7)*. This will play an important role when we prove the existence of a weak solution to system* (5.1)-(5.4)*,* (5.7)*. Furthermore, since $\Psi(\,\cdot\,)$ is a function of $\beta(\,\cdot\,)$ we replace $\Psi(\eta)$ with $\Psi(\eta^s)$.*

5.2 Literature review

The idea of coupling two different flow regimes has received considerable attention in the last 50 years. We can trace back the seed that later flourished into the richness of articles on the subject to the seminal paper [16], which introduced and experimentally confirmed the Beavers-Joseph condition. Evidence of the impact of this work are the more than 2000 references that cite it. A couple of years later, the paper [69] by Saffman verified through statistical means the validity of a simplified version of the Beavers-Joseph condition that results from neglecting the Darcy velocity from the term describing the velocity difference between the Stokes and Darcy domain. However, it was not until the year 2000 that the Beavers-Joseph-Saffman condition (5.7) was proven mathematically correct using boundary layer theory and homogenization theory [54]. Subsequently, in 2002, the work by Discacciati, Miglio and Quarteroni [31] considered the steady-state Darcy-Stokes model with a no-slip condition ($\alpha = 0$) on the interface using the pressure formulation for the Darcy domain. In the same year, the paper by Layton, Schieweck and Yotov [63] approached the same problem using the mixed formulation of Darcy equations and this time incorporating the Beavers-Joseph-Saffman condition. Therein, a thorough analysis of the well-posedness of the continuous and discrete problems is given. Five years later, motivated by a problem in a porous domain with several small cavities, Arbogast and Brunson in [6] introduced a finite element space whose elements adjacent to the interface Γ are modified in order to take into account the discontinuous tangential velocity. Also in 2007, the techniques of Galvis and Sarkis [45] allowed for the extension of the inf-sup condition discussed in [63] to the larger and physically correct space $H^{1/2}(\Gamma)$, replacing the smaller $H_{00}^{1/2}(\Gamma)$. A year later, Kaper, Mardal and Winther [57] employed a non-conforming discretization where the same finite

element space is employed in the porous and fluid domain. The elements have a continuous normal component over each edge, a weakly continuous tangential component over edges interior to each domain and a discontinuous tangential component over the interface. In the same period of time, Urquiza, N'Dri, Garon and Delfour [73] studied the coupled Stokes-Darcy system using the primal mixed formulation of the Darcy equations, i.e., integrating by parts the incompressibility condition instead of the gradient of the pressure. Furthermore, the authors include an stabilization term in the Darcy formulation which results in an elliptic bilinear form, making the consideration of an inf-sup condition unnecessary. An important generalization in the direction of nonlinear PDE's is the paper [36] by Ervin, Jenkins and Sun that considered a non-Newtonian fluid in both domains with the viscosity a function of the deformation tensor. Their work exhibits existence and uniqueness using the strong monotonicity of a nonlinear form and provides an a priori error estimate for the given finite element approximation. This paper was later followed by [37], where the same authors now use a mortar space to recast the coupled nonlinear Darcy-Stokes system as two reduced problems on the interface whose sole unknown is the interfacial pressure on the Darcy domain. In 2009, Rui and Zhang [68] and a year later, Feng, Qi, Zhu and Ju [40] added an stabilization term to the Stokes-Darcy weak formulation that allowed them to use the Crouzeix-Raviart element for both the porous and fluid domain. This resulted in the conservation of mass in each element, a highly desirable property. Then, in the same year, Badea et al. [9] replaced the Stokes momentum equation with the nonlinear Navier-Stokes and introduced a Poincaré-Steklov operator to reformulate the coupled problem as an interface equation with the normal flux as the unknown. Another work that aimed for pointwise conservation of mass is [56] where Kanschat and Rivière used a divergence-conforming velocity space with a discontinuous Galerkin method for the Stokes domain. Also relevant, is the work by

Ervin, Jenkins and Lee [35] that formulates the Stokes-Darcy system as a constrained optimal control problem, where the inflow, outflow and continuity of flux across the interface are enforced through the minimization of a functional. Particularly interesting is their least squares computational algorithm which gives optimal convergence rates.

Moving now to time dependent problems, in [24], Çeşmelioğlu and Rivière considered the evolutionary Navier-Stokes equations coupled with the Darcy equations in its pressure formulation and assume a Beavers-Joseph-Saffman law. Therein, the authors exhibit existence of a weak solution using the Galerkin method in an increasing sequence of finite dimensional Hilbert spaces. The solution is later shown to be strong by means of Sobolev embedding theorems. Moreover, a Crank-Nicolson scheme is proposed and second order convergence in time is established. Then, almost a decade after [63], the first work that considered the Darcy-Stokes problem using the more general Beavers-Joseph condition in both the steady-state and time-dependent setting was [23], where Cao, Gunzburger, Hua and Wang proved existence and uniqueness under the assumptions of a scaled Darcy law and a small parameter α. A follow-up to [23] is [22], where Cao et al. investigated a scaled Darcy law and the corresponding numerical solutions using a backward Euler scheme.

Once relevant analytical questions such as well-posedness were successfully answered, the scientific community started turning its attention towards the question of generating accurate and stable approximation schemes. In this line, the work [25] by W. Chen, P. Chen, Gunzburger and Yan proves that for the Taylor-Hood finite element pair in a particular refinement of a uniform isosceles right-triangle mesh, one can obtain a superconvergence order of 2.5 in the H^1 norm after applying a postprocessing interpolation operator. The next year, Cui and Yan [28] proved *a posteriori* upper and lower bounds for the error estimates of the Stokes-Darcy equations. These type

of estimates are particularly useful in providing reliable indicators for adaptive mesh refinement. Later, in 2012, Layton, Tran and Xiong [62] introduced four methods that decouple the Darcy-Stokes system and provided stability conditions that are functions of the physical parameters and the mesh size. In the same year, Layton and Trenchea exploited the skew symmetric coupling that arises by choosing the pressure formulation for the Darcy equations and analyzed two second order schemes: Crank-Nicolson Leap-Frog and BDF2-AB2, both members of the family of implicit-explicit (IMEX) methods. Also members of the IMEX family are the two schemes proposed by Chen, Gunzburger, Sun and Wang in [26]. Using a combination of a second order backward differentiation formula and a second order Gear's extrapolation for the first method and second order Adams-Bashforth and Adams-Moulton formulas for the second method, they achieved the much-desired unconditional stability. An alternative to introducing a stabilization term as it was done in [26], can be found in [61], where Layton, Tran and Trenchea show long time stability under a mild step restriction that depends upon the physical parameters. Finally, we would like to mention the work by Cao, Gunzburger, He and Wang [21] that considered Robin boundary conditions and proved, for a backward-Euler type scheme, unconditional stability and optimal convergence.

In the next section we restate the modeling equations and introduce the nomenclature and appropriate spaces in order to carry out the analysis.

5.3 Notation and assumptions

Throughout this manuscript, the symbol C indicates a generic constant independent of the discretization parameters, whose value may change from line to line.

The symbol μ denotes the dynamic viscosity in the fluid domain Ω_2. We let $\mathbf{u}$, p and η denote the velocity, pressure and porosity throughout Ω, respectively and use a subscript $i = 1, 2$ to indicate if a variable corresponds to the Darcy ($i = 1$) or Stokes ($i = 2$) domain. For example, $\mathbf{u}_1$ refers to the velocity on the Darcy domain. Moreover, for a generic function f_i supported on Ω_i, we extend f_i to the whole Ω by setting $f_i \equiv 0$ on $\Omega \setminus \Omega_i$. We omit the subscript whenever it is clear over which region the function is evaluated. We partition the boundary of $\Omega = \Omega_1 \cup \Omega_2$ into three disjoint pieces: The interface or connecting boundary $\Gamma = \partial\Omega_1 \cap \partial\Omega_2$, the Darcy boundary $\Gamma_1 = \partial\Omega_1 \setminus \Gamma$ and the Stokes boundary $\Gamma_2 = \partial\Omega_2 \setminus \Gamma$.

Assumption 5.1. *For the mathematical analysis we assume* $\mathbf{u} = 0$ on Γ_2 *and* $\mathbf{u} \cdot \mathbf{n} = 0$ on Γ_1.

We use the notation

$$(f, g)_U = \int_U f(\mathbf{x}) \cdot g(\mathbf{x})\, d\mathbf{x}, \quad \|f\|_U^2 = (f, f)_U,$$

to denote the L^2 inner product and norm on $U \subset \Omega$, where the dot product is replaced with the Frobenius product in the case of tensors, and

$$\langle f, g \rangle_\Gamma = \int_\Gamma f \cdot g\, dS$$

to indicate either a surface integral along the interface Γ, or the duality pairing between f and g. We omit the subscript whenever it is clear from the context over which region we compute the integral. The function $|\cdot|$ represents the Euclidean norm for vectors, the Frobenius norm for tensors and the Lebesgue measure for sets. The relevant function spaces in the following derivations are: The Darcy and Stokes

velocity spaces

$$\mathbf{X}_1 = \{\mathbf{v}_1 \in \mathbf{H}_{\text{div}}(\Omega_1) \,|\, \mathbf{v}_1 \cdot \mathbf{n} = 0 \text{ on } \Gamma_1\}, \quad \mathbf{X}_2 = \left\{\mathbf{v}_2 \in H^1(\Omega_2)^d \,|\, \mathbf{v}_2 = 0 \text{ on } \Gamma_2\right\},$$

the Darcy and Stokes pressure spaces, $Q_1 = L^2(\Omega_1)$ and $Q_2 = L^2(\Omega_2)$, the porosity space $L^2(\Omega_1)$, the space of Lagrange multipliers $\Lambda = H^{1/2}(\Gamma)$, and the space of continuous functions on Ω_1, $C^0(\Omega_1)$. The corresponding norms in the velocity spaces are, for $\mathbf{u}_1 \in \mathbf{X}_1$ and $\mathbf{u}_2 \in \mathbf{X}_2$:

$$\|\mathbf{u}_1\|_{\mathbf{X}_1} = \left(\|\mathbf{u}_1\|^2 + \|\nabla \cdot \mathbf{u}_1\|^2\right)^{1/2}, \quad \|\mathbf{u}_2\|_{\mathbf{X}_2} = \sqrt{2\mu}\,\|\mathbb{D}(\mathbf{u}_2)\|.$$

Furthermore, we introduce the spaces $Q = \left\{q \in L^2(\Omega) \,\middle|\, (q,1)_\Omega = 0\right\}$,

$$\mathbf{X} = \left\{\mathbf{v} \in L^2(\Omega)^d \,\middle|\, \mathbf{v}_1 \in \mathbf{X}_1 \text{ and } \mathbf{v}_2 \in \mathbf{X}_2\right\},$$

its continuous dual $\mathbf{X}'$, and the norm

$$\|\mathbf{u}\|_{\mathbf{X}} = \left(\|\mathbf{u}\|^2_{\mathbf{X}_1} + \|\mathbf{u}\|^2_{\mathbf{X}_2}\right)^{1/2}.$$

Some standard results that we use in the analysis are:

Lemma 5.1 (Korn's first inequality (see [20]))**.** *Let U be a polyhedral domain in $\mathbb{R}^d$ ($d = 2$ or 3), and let $\Gamma \subset \partial U \subset \mathbb{R}^{d-1}$ be such that $|\Gamma| > 0$. Then, there exists a positive constant $C = C(U, \Gamma)$ such that*

$$\|\nabla \mathbf{u}\|_{L^2(U)} \leq C\,\|\mathbb{D}(\mathbf{u})\|_{L^2(U)} \quad \forall \mathbf{u} \in H^1(U)^d,\ \mathbf{u}|_\Gamma = \mathbf{0}.$$

Lemma 5.2 (Poincaré-Friedrichs inequality (see [19]))**.** *Let U be a bounded domain*

in $\mathbb{R}^d$ *and let* $\Gamma \subset \partial U \subset \mathbb{R}^{d-1}$ *be such that* $|\Gamma| > 0$. *Then, there exists a positive constant* $C = C(U, \Gamma)$ *such that*

$$\|\mathbf{u}\|_{L^2(U)} \leq C \|\nabla \mathbf{u}\|^2_{L^2(U)} \quad \forall \mathbf{u} \in H^1(U),\ \mathbf{u}|_\Gamma = \mathbf{0}.$$

Lemma 5.3. *There exists a positive constant* C_K *such that*

$$\|\mathbf{u}\|_{H^1(\Omega_2)} \leq C_K \|\mathbf{u}\|_{\mathbf{X}_2} \quad \forall \mathbf{u} \in \mathbf{X}_2.$$

Proof. This is a direct consequence of Lemma 5.1 and Lemma 5.2. □

Theorem 5.1 (Trace theorem (see [30]))**.** *Let* U *be a bounded simply connected Lipschitz domain and let* $s \in (1/2, 3/2)$. *The map* $\mathcal{T} : H^s(U) \to H^{s-1/2}(\partial U)$ *that sends a function to its trace is a surjective bounded linear operator. In particular, in view of Lemma 5.3, there exists a positive constant* $C_\mathcal{T}$ *such that*

$$\|\mathcal{T}\mathbf{u}\|_{L^2(\Gamma)} \leq C_\mathcal{T} \|\mathbf{u}\|_{\mathbf{X}_2} \quad \forall \mathbf{u} \in \mathbf{X}_2.$$

The following definition introduces the concept of a smoothing operator in Ω_1.

Definition 5.1. *We say that the linear operator* $\mathcal{S} : L^2(\Omega_1) \to C^0(\Omega_1)$ *is smoothing if:*

- ***S*1***: There exists a constant* $C_s = C_s(\Omega_1)$ *such that* $\|\mathcal{S}(u)\|_{L^\infty(\Omega_1)} \leq C_s \|u\|_{L^2(\Omega_1)}$ *for all* $u \in L^2(\Omega_1)$, *and*
- ***S*2***: if* $\{u_n\}_{n=1}^\infty \subset L^2(\Omega_1)$ *such that* $u_n \rightharpoonup u$ *in* $L^2(\Omega_1)$, *then* $\mathcal{S}(u_n) \to \mathcal{S}(u)$ *in* $L^\infty(\Omega_1)$, *i.e.,* $\mathcal{S}$ *transforms a weakly convergent sequence in* $L^2(\Omega_1)$ *into a strongly convergent sequence in* $L^\infty(\Omega_1)$.

Examples of smoothing operators can be found in [38]. For the rest of the article we adopt the convention of indicating the action of $\mathcal{S}$ on η as η^s.

The full system is now restated for ease of reference. For $T \in \mathbb{R}^+$ denoting the time horizon of the filtration process, we consider

$$\begin{aligned}
\beta(\eta^s)\mathbf{u} + \nabla p &= \mathbf{f} \text{ in } \Omega_1 \times (0,T), \\
\nabla \cdot \mathbf{u} &= 0 \text{ in } \Omega_1 \times (0,T), && (5.8)\\
\frac{\partial \eta}{\partial t} + g(|\mathbf{u}|)\, h(\eta) &= 0 \text{ in } \Omega_1 \times (0,T), && (5.9)\\
\eta &= \eta_0 \text{ in } \Omega_1 \times \{0\}, \\
\nabla \cdot \mathbb{T}(\mathbf{u}, p) &= \mathbf{f} \text{ in } \Omega_2 \times (0,T), \\
\nabla \cdot \mathbf{u} &= 0 \text{ in } \Omega_2 \times (0,T), && (5.10)
\end{aligned}$$

with interface conditions

$$(\mathbf{u}_2 - \mathbf{u}_1) \cdot \mathbf{n} = 0 \text{ on } \Gamma \times (0,T), \tag{5.11a}$$

$$\mathbf{n} \cdot \mathbb{T}(\mathbf{u}_2, p_2) \cdot \mathbf{n} = p_1 \text{ on } \Gamma \times (0,T), \tag{5.11b}$$

$$P_{\mathbf{t}}(\mathbb{T}(\mathbf{u}_2, p_2) \cdot \mathbf{n}) = \Psi(\eta)\, P_{\mathbf{t}}(\mathbf{u}_2) \text{ on } \Gamma \times (0,T), \tag{5.11c}$$

and boundary data

$$\mathbf{u} \cdot \mathbf{n} = 0 \text{ on } \Gamma_1 \times (0,T), \tag{5.12a}$$

$$\mathbf{u} = 0 \text{ on } \Gamma_2 \times (0,T). \tag{5.12b}$$

We make the following assumptions:

A1: The functions

$$\beta(\,\cdot\,):\mathbb{R}^+\to\mathbb{R}^+,\quad g(\,\cdot\,):\mathbb{R}^+\cup\{0\}\to\mathbb{R}^+\cup\{0\},$$
$$\Psi(\,\cdot\,):\mathbb{R}^+\to\mathbb{R}^+,\quad h(\,\cdot\,):\mathbb{R}^+\cup\{0\}\to\mathbb{R}^+\cup\{0\},$$

satisfy the following bounds:

$$0<\beta_{\min}\le\beta(\,\cdot\,)\le\beta_{\max},\quad g(\,\cdot\,)\le g_{\max},$$
$$0<\Psi_{\min}\le\Psi(\,\cdot\,)\le\Psi_{\max},\quad h(\,\cdot\,)\le h_{\max}.$$

A2: $\beta(\,\cdot\,)$, $g(\,\cdot\,)$, $h(\,\cdot\,)$, $\Psi(\,\cdot\,)$ are Lipschitz continuous with Lipschitz constants $\beta_{\rm Lip}$, $g_{\rm Lip}$, $h_{\rm Lip}$ and $\Psi_{\rm Lip}$, respectively.

A3: $\eta_0\in L^\infty(\Omega_1)$ and $\eta_0(\mathbf{x})\ge 0$ for *a.e.* $\mathbf{x}$ in Ω_1.

A4: $\mathbf{f}\in C^0(0^-,T;\,L^2(\Omega))$ defined as $C^0(0-\delta,T;\,L^2(\Omega))\cap L^\infty(0,T;L^2(\Omega))$ for some $\delta>0$.

Remark 5.2. *Note that since η^s is the result of $\mathcal{S}(\eta)$, the Beavers-Joseph-Saffman condition* (5.11c) *is well-defined in view that $\eta^s\in C^0(\Omega_1)$.*

Additional notation that we need in the subsequent analysis is: For $\mathbf{u},\mathbf{v}\in\mathbf{X}$, $q\in Q$, $\eta\in L^2(\Omega_1)$ and $\nu\in\Lambda$, define

$$a_1(\eta;\mathbf{u},\mathbf{v})=\left(\beta(\eta^s)\,\mathbf{u},\mathbf{v}\right)_{\Omega_1},\quad a_2(\mathbf{u},\mathbf{v})=\Big(2\mu\,\mathbb{D}(\mathbf{u}),\mathbb{D}(\mathbf{v})\Big)_{\Omega_2},$$
$$b(\mathbf{v},q)=-(\nabla\cdot\mathbf{v},q)_\Omega,\quad c(\mathbf{v},\nu)=\langle(\mathbf{v}_2-\mathbf{v}_1)\cdot\mathbf{n},\nu\rangle_\Gamma,\quad \ell(\mathbf{v})=(\mathbf{f},\mathbf{v})_\Omega,$$
$$d(\eta;\mathbf{u},\mathbf{v})=\langle\Psi(\eta^s)\,P_{\mathbf{t}}(\mathbf{u}_2),P_{\mathbf{t}}(\mathbf{v}_2)\rangle_\Gamma. \tag{5.13}$$

Remark 5.3. *The definition of the bilinear form $a_2(\cdot\,,\cdot\,)$ is natural in view of the definition of the stress tensor*

$$\mathbb{T} = -2\mu\,\mathbb{D}(\mathbf{u}) + p\mathbb{I}.$$

Remark 5.4. *Owing to assumption* ***A4****, the operator $\ell : \mathbf{X} \to \mathbb{R}$ introduced in* (5.13) *is a continuous linear functional. Thus, $\ell \in \mathbf{X}'$.*

5.4 Existence and uniqueness

Multiplying (5.8)-(5.10) by the corresponding test functions and incorporating conditions (5.11)-(5.12), the resulting weak form is: *Given $\eta_0 \in L^2(\Omega_1)$ and $\mathbf{f} \in C^0(0^-, T;\, L^2(\Omega))$, find $\mathbf{u} \in L^2(0,T;\mathbf{X})$, $p \in L^2(0,T;Q)$, $\lambda \in L^2(0,T;\Lambda)$ and $\eta \in H^1(0,T;\, L^2(\Omega_1))$, satisfying $\eta(\cdot,0) = \eta_0$ a.e. in Ω_1, and for a.e. $t \in (0,T)$*

$$a_1(\eta;\mathbf{u},\mathbf{v}) + a_2(\mathbf{u},\mathbf{v}) + b(\mathbf{v},p) + c(\mathbf{v},\lambda) + d(\eta;\mathbf{u},\mathbf{v}) = \ell(\mathbf{v})\ \forall \mathbf{v} \in \mathbf{X}, \tag{5.14}$$

$$b(\mathbf{u},q) = 0\ \forall q \in Q, \tag{5.15}$$

$$c(\mathbf{u},\nu) = 0\ \forall \nu \in \Lambda, \tag{5.16}$$

$$(\frac{\partial \eta}{\partial t}, \xi)_{\Omega_1} + \Big(g(|\mathbf{u}|)\, h(\eta), \xi\Big)_{\Omega_1} = 0\ \forall \xi \in L^2(\Omega_1). \tag{5.17}$$

To simplify the analysis, we introduce the space

$$\mathbf{V} = \left\{ \mathbf{v} \in \mathbf{X} \,\middle|\, c(\mathbf{v},\nu) = 0\ \forall \nu \in \Lambda,\ \text{ and }\ b(\mathbf{v},q) = 0\ \forall q \in Q \right\}.$$

In order to restrict the analysis to $\mathbf{V}$, we require the following inf-sup condition.

Lemma 5.4. *There exists a positive constant γ such that*

$$\gamma < \inf_{\mathbf{0}\neq(q,\nu)\in Q\times\Lambda} \sup_{\mathbf{0}\neq\mathbf{v}\in\mathbf{X}} \frac{b(\mathbf{v},q)+c(\mathbf{v},\nu)}{\|\mathbf{v}\|_{\mathbf{X}}\ \|(q,\nu)\|_{Q\times\Lambda}},$$

where $\|(q,\nu)\|^2_{Q\times\Lambda} = \|q\|^2_{L^2(\Omega)} + \|\nu\|^2_{\Lambda}$.

Proof. The result follows from Proposition 4.7 and Remark 4.8 in [45]. □

In view of the definition of $\mathbf{V}$ and Lemma 5.4, assuming $\eta \in L^2(\Omega_1)$ and $\mathbf{f} \in C^0(0^-, T;\ L^2(\Omega))$ are given, system (5.14)-(5.16) is equivalent to: Find $\mathbf{u} \in L^2(0,T;\mathbf{V})$ satisfying

$$a_1(\eta;\mathbf{u},\mathbf{v}) + a_2(\mathbf{u},\mathbf{v}) + d(\eta;\mathbf{u},\mathbf{v}) = \ell(\mathbf{v}) \quad \forall \mathbf{v}\in\mathbf{V}. \tag{5.18}$$

First we show that for $\eta \in L^2(\Omega_1)$ given, (5.18) is well-posed.

Lemma 5.5. *Let $\eta \in L^2(\Omega_1)$ be given. Let $\mathbf{u},\mathbf{v}\in\mathbf{V}$ and define $a : L^2(\Omega_1)\times\mathbf{V}\times\mathbf{V} \to \mathbb{R}$ by*

$$a(\eta;\mathbf{u},\mathbf{v}) = a_1(\eta;\mathbf{u},\mathbf{v}) + a_2(\mathbf{u},\mathbf{v}) + d(\eta;\mathbf{u},\mathbf{v}). \tag{5.19}$$

Then, the problem: Find $\mathbf{u}\in\mathbf{V}$ satisfying for all $\mathbf{v}\in\mathbf{V}$

$$a(\eta;\mathbf{u},\mathbf{v}) = \ell(\mathbf{v}), \tag{5.20}$$

has a unique solution. We call $(\eta,\mathbf{u})$ a solution pair to (5.20).

Proof. In view of assumption ***A1*** and the continuity of the trace map, it follows that

$$\begin{aligned} a(\eta; \mathbf{u}, \mathbf{v}) &\leq \beta_{\max} \|\mathbf{u}\|_{\mathbf{X}_1} \|\mathbf{v}\|_{\mathbf{X}_1} + 2\mu \|\mathbf{u}\|_{\mathbf{X}_2} \|\mathbf{v}\|_{\mathbf{X}_2} + \Psi_{\max} C_{\mathcal{T}}^2 \|\mathbf{u}\|_{\mathbf{X}_2} \|\mathbf{v}\|_{\mathbf{X}_2} \\ &\leq \max\left\{\beta_{\max},\, 2\mu,\, \Psi_{\max} C_{\mathcal{T}}^2\right\} \|\mathbf{u}\|_{\mathbf{X}} \|\mathbf{v}\|_{\mathbf{X}}. \end{aligned} \tag{5.21}$$

Now observe that owing to ***A1*** and the nonnegativity of $d(\eta, \mathbf{u}, \mathbf{u})$,

$$a(\eta; \mathbf{u}, \mathbf{u}) \geq \beta_{\min} \|\mathbf{u}\|_{\mathbf{X}_1}^2 + \|\mathbf{u}\|_{\mathbf{X}_2}^2 \geq \min\left\{\beta_{\min},\, 1\right\} \|\mathbf{u}\|_{\mathbf{X}}^2. \tag{5.22}$$

Finally, note that in view of Lemma 5.3

$$\begin{aligned} \ell(\mathbf{v}) = (\mathbf{f}, \mathbf{v})_{\Omega_1} + (\mathbf{f}, \mathbf{v})_{\Omega_2} &\leq \|\mathbf{f}\|_{L^2(\Omega_1)} \|\mathbf{v}\|_{L^2(\Omega_1)} + \|\mathbf{f}\|_{H^{-1}(\Omega_2)} \|\mathbf{v}\|_{H^1(\Omega_2)} \\ &\leq \|\mathbf{f}\|_{L^2(\Omega_1)} \|\mathbf{v}\|_{L^2(\Omega_1)} + \|\mathbf{f}\|_{L^2(\Omega_2)} C_K \|\mathbf{v}\|_{\mathbf{X}_2} \leq \max\left\{1,\, C_K\right\} \|\mathbf{f}\| \|\mathbf{v}\|_{\mathbf{X}}. \end{aligned} \tag{5.23}$$

Consequently, from (5.21), (5.22) and (5.23), the existence of a unique solution to (5.20) follows by the Lax-Milgram lemma. □

The next corollary provides an estimate for the norm of the solution to problem (5.20).

Corollary 5.1. *Define*

$$C_\beta = \frac{1}{\min\left\{\beta_{\min},\, 1\right\}}, \quad C_b = \max\left\{1,\, C_K\right\} C_\beta, \quad C_{\mathbf{f}} = C_b \|\mathbf{f}\|_{L^\infty(0,T;L^2(\Omega))},$$

and let $\mathbf{u} \in \mathbf{V}$ *be the solution to the problem stated in Lemma 5.5. Then,*

$$\|\mathbf{u}\|_{\mathbf{X}}(t) \leq C_{\mathbf{f}}. \tag{5.24}$$

Proof. This is a direct consequence of (5.22) and (5.23). □

The following result is related to the continuity of the solution $\mathbf{u}$ as a function of the porosity η.

Lemma 5.6. *Let* $(\eta^1, \mathbf{u}^1)$, $(\eta^2, \mathbf{u}^2) \in L^2(\Omega_1) \times \mathbf{V}$ *be solution pairs to problem* (5.20). *Then,*

$$\|\mathbf{u}^2 - \mathbf{u}^1\|^2_{\mathbf{X}}(t) \leq C_\beta \, C^2_{\mathbf{f}} \left(\Psi^2_{\mathrm{Lip}} \, C^4_{\mathcal{T}} + \frac{\beta^2_{\mathrm{Lip}}}{\beta_{\min}} \right) \|\mathcal{S}(\eta^2 - \eta^1)\|^2_{L^\infty(\Omega_1)}(t). \tag{5.25}$$

Proof. For clarity of exposition, we suppress the dependence of the functions on t. First note that

$$\begin{aligned} a_1(\eta^1; \mathbf{u}^1, \mathbf{v}) - a_1(\eta^2; \mathbf{u}^2, \mathbf{v}) = \big(\beta(\eta^{1,s})\,(\mathbf{u}^1 - \mathbf{u}^2), \mathbf{v}\big)_{\Omega_1} \\ + \big(\big(\beta(\eta^{1,s}) - \beta(\eta^{2,s})\big)\ \mathbf{u}^2, \mathbf{v}\big)_{\Omega_1}. \end{aligned} \tag{5.26}$$

Similarly for $d(\,\cdot\,,\,\cdot\,,\,\cdot\,)$ and $a_2(\,\cdot\,,\,\cdot\,)$,

$$\begin{aligned} d(\eta^1; \mathbf{u}^1, \mathbf{v}) - d(\eta^2; \mathbf{u}^2, \mathbf{v}) = \big\langle \Psi(\eta^{1,s})\, P_{\mathbf{t}}(\mathbf{u}^1_2 - \mathbf{u}^2_2), P_{\mathbf{t}}(\mathbf{v}_2) \big\rangle_\Gamma \\ + \big\langle \big(\Psi(\eta^{1,s}) - \Psi(\eta^{2,s})\big)\ P_{\mathbf{t}}(\mathbf{u}^2_2), P_{\mathbf{t}}(\mathbf{v}_2) \big\rangle_\Gamma, \end{aligned} \tag{5.27}$$

$$a_2(\mathbf{u}^1, \mathbf{v}) - a_2(\mathbf{u}^2, \mathbf{v}) = \big(2\mu\, \mathbb{D}(\mathbf{u}^1 - \mathbf{u}^2), \mathbb{D}(\mathbf{v})\big)_{\Omega_2}. \tag{5.28}$$

Now observe that

$$\begin{aligned} a(\eta^1; \mathbf{u}^1, \mathbf{v}) - a(\eta^2; \mathbf{u}^2, \mathbf{v}) = a_1(\eta^1; \mathbf{u}^1, \mathbf{v}) - a_1(\eta^2; \mathbf{u}^2, \mathbf{v}) \\ + a_2(\mathbf{u}^1, \mathbf{v}) - a_2(\mathbf{u}^2, \mathbf{v}) + d(\eta^1; \mathbf{u}^1, \mathbf{v}) - d(\eta^2; \mathbf{u}^2, \mathbf{v}) = 0. \end{aligned} \tag{5.29}$$

Adding (5.26) , (5.27) and (5.28), and using (5.29), yields

$$\begin{aligned}
&\big(\beta(\eta^{1,s})\,(\mathbf{u}^2-\mathbf{u}^1),\mathbf{v}\big)_{\Omega_1}+\big\langle\Psi(\eta^{1,s})\,P_{\mathbf{t}}(\mathbf{u}_2^2-\mathbf{u}_2^1),P_{\mathbf{t}}(\mathbf{v}_2)\big\rangle_{\Gamma}\\
&+\big(2\mu\,\mathbb{D}(\mathbf{u}^2-\mathbf{u}^1),\mathbb{D}(\mathbf{v})\big)_{\Omega_2}=\big\langle\big(\Psi(\eta^{1,s})-\Psi(\eta^{2,s})\big)\;P_{\mathbf{t}}(\mathbf{u}_2^2),P_{\mathbf{t}}(\mathbf{v}_2)\big\rangle_{\Gamma}\\
&+\big(\big(\beta(\eta^{1,s})-\beta(\eta^{2,s})\big)\;\mathbf{u}^2,\mathbf{v}\big)_{\Omega_1}. \qquad (5.30)
\end{aligned}$$

Setting $\mathbf{v}=\mathbf{u}_2-\mathbf{u}_1$ in (5.30) and using assumptions $\boldsymbol{A1}$, $\boldsymbol{A2}$, and the trace theorem, we obtain

$$\begin{aligned}
&\beta_{\min}\,\|\mathbf{u}^2-\mathbf{u}^1\|^2_{\mathbf{X}_1}+\|\mathbf{u}^2-\mathbf{u}^1\|^2_{\mathbf{X}_2}\\
&\qquad\le\Psi_{\mathrm{Lip}}\,\|\eta^{1,s}-\eta^{2,s}\|_{L^\infty(\Omega_1)}\,C_{\mathcal{T}}^2\,\|\mathbf{u}^2\|_{\mathbf{X}_2}\,\|\mathbf{u}^2-\mathbf{u}^1\|_{\mathbf{X}_2}\\
&\qquad+\beta_{\mathrm{Lip}}\,\|\eta^{1,s}-\eta^{2,s}\|_{L^\infty(\Omega_1)}\,\|\mathbf{u}^2\|_{\mathbf{X}_1}\,\|\mathbf{u}^2-\mathbf{u}^1\|_{\mathbf{X}_1}. \qquad (5.31)
\end{aligned}$$

Applying Corollary 5.1 to bound $\|\mathbf{u}\|_{\mathbf{X}_i}$ for $i=1,2$ and Young's inequality in (5.31), yields

$$\begin{aligned}
&\beta_{\min}\,\|\mathbf{u}^2-\mathbf{u}^1\|^2_{\mathbf{X}_1}+\|\mathbf{u}^2-\mathbf{u}^1\|^2_{\mathbf{X}_2}\le\frac{1}{4\varepsilon_2}\Psi^2_{\mathrm{Lip}}\,\|\eta^{1,s}-\eta^{2,s}\|^2_{L^\infty(\Omega_1)}\,C_{\mathcal{T}}^4\,C_{\mathbf{f}}^2\\
&\quad+\varepsilon_2\,\|\mathbf{u}^2-\mathbf{u}^1\|^2_{\mathbf{X}_2}+\frac{1}{4\varepsilon_1}\beta^2_{\mathrm{Lip}}\,\|\eta^{1,s}-\eta^{2,s}\|^2_{L^\infty(\Omega_1)}\,C_{\mathbf{f}}^2+\varepsilon_1\|\mathbf{u}^2-\mathbf{u}^1\|^2_{\mathbf{X}_1}. \qquad (5.32)
\end{aligned}$$

Finally, setting $\varepsilon_1=\frac{\beta_{\min}}{2}$ and $\varepsilon_2=1/2$ in (5.32), we obtain

$$\beta_{\min}\|\mathbf{u}^2-\mathbf{u}^1\|^2_{\mathbf{X}_1}+\|\mathbf{u}^2-\mathbf{u}^1\|^2_{\mathbf{X}_2}\le\|\eta^{1,s}-\eta^{2,s}\|^2_{L^\infty(\Omega_1)}C_{\mathbf{f}}^2\left(\Psi^2_{\mathrm{Lip}}\,C_{\mathcal{T}}^4+\frac{\beta^2_{\mathrm{Lip}}}{\beta_{\min}}\right). \qquad (5.33)$$

Estimate (5.25) follows from (5.33). □

Corollary 5.2. *Let* $(\eta_1,\mathbf{u}_1)$, $(\eta_2,\mathbf{u}_2)\in L^2(\Omega_1)\times\mathbf{V}$ *be solution pairs to problem* (5.20).

Then, there exists a positive constant C_{Lip}, independent of t, such that

$$\|\mathbf{u}_2 - \mathbf{u}_1\|_{\mathbf{X}}(t) \leq C_{\text{Lip}} \|\eta_1 - \eta_2\|_{L^2(\Omega_1)}(t). \tag{5.34}$$

Proof. In view of Lemma 5.6, and property $\boldsymbol{S1}$ of the smoother, estimate (5.34) follows, where C_{Lip} is the square root of the product of the constants arising in (5.25) and C_s^2. □

The next lemma shows that for a given porosity η, the solution $\mathbf{u}$ to (5.20) depends continuously on the forcing term $\mathbf{f}$.

Lemma 5.7. *Let $\eta \in L^2(\Omega_1)$ be given and let $\mathbf{u}^1, \mathbf{u}^2$ be the solutions to* (5.20) *corresponding to the linear functionals ℓ^1 and ℓ^2, respectively. Then,*

$$\|\mathbf{u}^1 - \mathbf{u}^2\|_{\mathbf{X}}(t) \leq C_\beta \|\ell^1 - \ell^2\|_{\mathbf{X}'}(t).$$

Proof. Consider the problems

$$a(\eta, \mathbf{u}^1, \mathbf{v}) = \ell^1(\mathbf{v}) \ \forall \mathbf{v} \in \mathbf{V}, \tag{5.35}$$

$$a(\eta, \mathbf{u}^2, \mathbf{v}) = \ell^2(\mathbf{v}) \ \forall \mathbf{v} \in \mathbf{V}. \tag{5.36}$$

Subtracting (5.35) from (5.36) and proceeding in a similar manner as in Lemma 5.6 (see (5.31)), we obtain the bound

$$\beta_{\min} \|\mathbf{u}^2 - \mathbf{u}^1\|_{\mathbf{X}_1}^2 + \|\mathbf{u}^2 - \mathbf{u}^1\|_{\mathbf{X}_2}^2 \leq \|\ell^2 - \ell^1\|_{\mathbf{X}'} \|\mathbf{u}^2 - \mathbf{u}^1\|_{\mathbf{X}}. \tag{5.37}$$

From (5.37), the result follows. □

The next corollary is a generalization of Corollary 5.2 and Lemma 5.7.

Corollary 5.3. *Let* $(\eta^1, \mathbf{u}^1)$, $(\eta^2, \mathbf{u}^2) \in L^2(\Omega_1) \times \mathbf{V}$ *be solution pairs to*

$$a(\eta^1, \mathbf{u}^1, \mathbf{v}) = \ell^1(\mathbf{v}) \ \forall \mathbf{v} \in \mathbf{V},$$
$$a(\eta^2, \mathbf{u}^2, \mathbf{v}) = \ell^2(\mathbf{v}) \ \forall \mathbf{v} \in \mathbf{V}.$$

Then,

$$\|\mathbf{u}^2 - \mathbf{u}^1\|^2_{\mathbf{X}}(t) \leq 2C_{\text{Lip}} \|\eta^2 - \eta^1\|^2_{L^2(\Omega_1)}(t) + 4C_\beta \|\ell^2 - \ell^1\|^2_{\mathbf{X}'}(t). \tag{5.38}$$

Proof. Following the same steps that lead to (5.32) and (5.37), and applying Young's inequality, yields

$$\begin{aligned}\beta_{\min} \|\mathbf{u}^2 - \mathbf{u}^1\|^2_{\mathbf{X}_1} + \|\mathbf{u}^2 - \mathbf{u}^1\|^2_{\mathbf{X}_2} \leq{}& \frac{1}{4\varepsilon_2} \Psi^2_{\text{Lip}} \|\mathcal{S}(\eta^1 - \eta^2)\|^2_{L^\infty(\Omega_1)} C^4_{\mathcal{T}} C^2_{\mathbf{f}} \\ &+ \varepsilon_2 \|\mathbf{u}^2 - \mathbf{u}^1\|^2_{\mathbf{X}_2} + \frac{1}{4\varepsilon_1} \beta^2_{\text{Lip}} \|\mathcal{S}(\eta^1 - \eta^2)\|^2_{L^\infty(\Omega_1)} C^2_{\mathbf{f}} + \varepsilon_1 \|\mathbf{u}^2 - \mathbf{u}^1\|^2_{\mathbf{X}_1} \\ &+ \varepsilon_3 \|\ell^2 - \ell^1\|^2_{\mathbf{X}'} + \frac{1}{4\varepsilon_3} \|\mathbf{u}^2 - \mathbf{u}^1\|_{\mathbf{X}}. \end{aligned} \tag{5.39}$$

Setting $\varepsilon_1 = \frac{\beta_{\min}}{2}$, $\varepsilon_2 = 1/2$, $\varepsilon_3 = C_\beta$ in (5.39) and replacing the $L^\infty(\Omega)$ norm of the smoothed variables with the $L^2(\Omega_1)$ norm of the original variables as described in the proof of Corollary 5.2, we obtain

$$\begin{aligned}\beta_{\min} \|\mathbf{u}^2 - \mathbf{u}^1\|^2_{\mathbf{X}_1} + \|\mathbf{u}^2 - \mathbf{u}^1\|^2_{\mathbf{X}_2} \leq{}& C_{\text{Lip}} \|\eta^2 - \eta^1\|^2_{L^2(\Omega_1)}(t) + 2C_\beta \|\ell^2 - \ell^1\|^2_{\mathbf{X}'} \\ &+ \frac{C_\beta^{-1}}{2} \|\mathbf{u}^2 - \mathbf{u}^1\|_{\mathbf{X}}. \end{aligned} \tag{5.40}$$

Finally, making use of the bound

$$C_\beta^{-1} \|\mathbf{u}^2 - \mathbf{u}^1\|^2_{\mathbf{X}} \leq \beta_{\min} \|\mathbf{u}^2 - \mathbf{u}^1\|^2_{\mathbf{X}_1} + \|\mathbf{u}^2 - \mathbf{u}^1\|^2_{\mathbf{X}_2}$$

in (5.40), estimate (5.38) follows. □

We proceed to define two operators. One outputs the Darcy velocity as a function of the porosity η and the other describes the deposition function in terms of η.

Definition 5.2. *Let* $\mathcal{P} : L^2(\Omega_1) \times (0,T) \to \mathbf{V}$ *be given by*

$$\mathcal{P}(\eta, t) = \mathbf{u}(t),$$

where $\mathbf{u}(t)$ *is defined through the solution pair* $(\eta, \mathbf{u})$ *of the problem introduced in* (5.20)*, and define* $\mathcal{F} : L^2(\Omega_1) \times (0,T) \to L^2(\Omega_1)$ *by*

$$\mathcal{F}(\eta, t) = g\left(|\mathcal{P}(\eta, t)|\right)\, h(\eta).$$

Remark 5.5. *The time dependency of* $\mathcal{P}$ *is due the forcing term* $\mathbf{f}(t)$*.*

The following properties of $\mathcal{P}$ and $\mathcal{F}$ are used in the main result of this section.

Lemma 5.8. *The operator* $\mathcal{P}(\,\cdot\,, t)$ *is Lipschitz continuous for every* $t \in (0,T)$*.*

Proof. This is a direct consequence of Corollary 5.2. □

Lemma 5.9. *The operator* $\mathcal{P}(\eta,\,\cdot\,)$ *is continuous for every* $\eta \in L^2(\Omega_1)$*.*

Proof. Let $t \in (0,T)$, $\varepsilon > 0$ and $\eta \in L^2(\Omega_1)$ be given. With reference to Lemma 5.7, define the linear functionals $\ell^1(\mathbf{v}) = (\mathbf{f}(t), \mathbf{v})_\Omega$ and $\ell^2(\mathbf{v}) = (\mathbf{f}(t+h), \mathbf{v})_\Omega$ for some $h \in \mathbb{R}$, and let $\mathbf{u}^1 = \mathcal{P}(\eta, t)$, $\mathbf{u}^2 = \mathcal{P}(\eta, t+h)$ be the corresponding solutions, respectively. Then, by Lemma 5.7,

$$\|\mathbf{u}^1 - \mathbf{u}^2\|_{\mathbf{X}} \leq C_\beta \, \|\mathbf{f}(t) - \mathbf{f}(t+h)\|. \tag{5.41}$$

Owing to assumption ***A4***, we can find an open ball $B \subset \mathbb{R}$ centered at zero of radius $\delta > 0$ such that for all $h \in B$,

$$\|\mathbf{f}(t) - \mathbf{f}(t+h)\| \leq \varepsilon. \tag{5.42}$$

Hence, combining (5.41) and (5.42) the result follows. □

Lemma 5.10. *The operator $\mathcal{F}(\,\cdot\,, t)$ is Lipschitz continuous for every $t \in \mathbb{R}^+$.*

Proof. In view that the composition of Lipschitz continuous functions is Lipschitz and the product of bounded Lipschitz continuous functions is Lipschitz, assumptions ***A1*** and ***A2*** together with Lemma 5.8 imply the assertion. □

Lemma 5.11. *The operator $\mathcal{F}(\eta,\, \cdot\,)$ is continuous for every $\eta \in L^2(\Omega_1)$.*

Proof. From assumption ***A2*** and Lemma 5.9, the function $g\left(|\mathcal{P}(\eta, \cdot)|\right)$ is continuous. Hence, $\mathcal{F}(\eta,\, \cdot\,)$ is continuous. □

Remark 5.6. *In view of **A1**, the operator $\mathcal{F}(\cdot, \cdot)$ is uniformly bounded in $L^2(\Omega_1) \times (0, T)$ by the constant $C_{\mathcal{F}} = g_{\max}\, h_{\max}$. Thus,*

$$\|\mathcal{F}(\cdot, \cdot)\|_{L^2(\Omega_1)} \leq C_{\mathcal{F}}\, |\Omega_1|^{1/2}.$$

Remark 5.7. *With an additional assumption on $\mathbf{f}$, we can strengthen Lemma 5.9 to obtain Lipschitz continuity on the variable t and, additionally, Lipschitz continuity on the whole domain $L^2(\Omega_1) \times (0, T)$. This idea is explored in Section 5.5.*

Before introducing the main theorem of this section, we restate the Picard-Lindelöf theorem.

Theorem 5.2 ([49], Theorem I.3.1)**.** *Let I denote a domain in $\mathbb{R}$ containing the point t_0, Y a Banach space and $f : Y \times \mathbb{R} \to Y$. Suppose that f is locally Lipschitz continuous in its first variable and continuous in its second variable. Then, there exists $\varepsilon > 0$ such that the initial value problem*

$$u' = f(u, t),$$
$$u(t_0) = u_0,$$

has a unique solution in $C^0(t_0 - \varepsilon, t_0 + \varepsilon; Y)$.

Theorem 5.3. *Under assumptions **A1**-**A4** and **S1**-**S2**, there exists a unique solution $\mathbf{u} \in L^2(0, T; \mathbf{V})$, $p \in L^2(0, T; Q)$, $\lambda \in L^2(0, T; \Lambda)$ and $\eta \in H^1(0, T;\, L^2(\Omega_1))$ satisfying* (5.14)-(5.17) *for a.e. $t \in (0, T)$.*

Proof. First, we focus on computing the porosity. In view of (5.17) and using Definition 5.2, we consider the problem: Find $\eta \in C^0(0, T)$ such that

$$\frac{\partial \eta}{\partial t} = -g(|\mathcal{P}(\eta, t)|)\, h(\eta) = -\mathcal{F}(\eta, t)\ \forall t \in (0, T). \tag{5.43}$$

Let $t \in (0, T)$ be given. Owing to Theorem 5.2, Lemma 5.10 and Lemma 5.11, there exists $\varepsilon > 0$ and an interval $(t - \varepsilon, t + \varepsilon)$ where the existence of a unique η is guaranteed. From Corollary 5.2 and Remark 5.6, the Lipschitz constant C_{Lip} and the bound for $\|\mathcal{F}(\,\cdot\,,\,\cdot\,)\|_{L^2(\Omega_1)}$ are independent of η_0 and t. Hence, one can extend the solution to the whole interval $(0, T)$. Next, we use η in Lemma 5.5 to obtain the velocity $\mathbf{u}$. Finally, owing to the inf-sup condition in Lemma 5.4, the existence and uniqueness of p and λ follow. □

A simple consequence of Theorem 5.3 is the next corollary, which upgrades the regularity of η.

Corollary 5.4. *The porosity function η given by Theorem 5.3 is Lipschitz continuous on $(0,T)$.*

Proof. In view of Lemma 5.10 and Lemma 5.11, $\mathcal{F}(\,\cdot\,,\,\cdot\,)$ is continuous. Furthermore, owing to Theorem 5.3, the obtained solution η is continuous on $(0,T)$. Thus, from (5.43), it follows that $\dfrac{\partial \eta}{\partial t}$ is continuous. Consequently η is $C^1(0,T)$ and therefore Lipschitz continuous on the same interval. □

Notation 5.1. *We denote the Lipschitz constant of η in Corollary 5.4 by η_{Lip}.*

In the next section we aim to extend the regularity of η and $\mathbf{u}$ in Theorem 5.3 by upgrading the regularity of $\mathbf{f}$.

5.5 Additional regularity

The key ingredient in the subsequent derivations is to assume that the function $\mathbf{f}$ in (5.8), (5.10) is Lipschitz continuous. At the end of this section we conclude that $\mathbf{u}$ is Lipschitz continuous and $\eta \in H^2(0,T;L^2(\Omega_1))$.

Notation 5.2. *We denote the Lipschitz constant of $\mathbf{f}$ by $\mathbf{f}_{\mathrm{Lip}}$.*

Lemma 5.12. *Assume $\mathbf{f} : (0,T) \to L^2(\Omega)$ is Lipschitz continuous. Then, the operator $\mathcal{P}(\,\cdot\,,\,\cdot\,)$ is Lipschitz continuous on $L^2(\Omega_1) \times (0,T)$.*

Proof. Let η^1, $\eta^2 \in L^2(\Omega_1)$ be given. Using similar notation to the one introduced in the proof of Lemma 5.9, and owing to Corollary 5.3, we obtain the bound

$$\|\mathbf{u}^2 - \mathbf{u}^1\|_{\mathbf{X}}^2 \leq 2C_{\mathrm{Lip}} \|\eta^2 - \eta^1\|_{L^2(\Omega_1)}^2 + 4C_\beta \|\mathbf{f}(t) - \mathbf{f}(t+h)\|^2, \tag{5.44}$$

where $\mathbf{u}^1 = \mathcal{P}(\eta^1, t)$, $\mathbf{u}^2 = \mathcal{P}(\eta^2, t+h)$. Define $t_1 = t$ and $t_2 = t+h$. Making use of the Lipschitz continuity of $\mathbf{f}$ in (5.44), yields

$$\|\mathbf{u}^2 - \mathbf{u}^1\|_{\mathbf{X}}^2 \leq 2C_{\text{Lip}} \|\eta^2 - \eta^1\|_{L^2(\Omega_1)}^2 + 4C_\beta \, \mathbf{f}_{\text{Lip}}^2 \, |t_2 - t_1|^2 \, |\Omega| \,. \tag{5.45}$$

Hence, the Lipschitz continuity of $\mathcal{P}(\,\cdot\,,\,\cdot\,)$ follows from (5.45). □

Similar to Lemma 5.12, we can upgrade the regularity of $\mathcal{F}$ by means of the additional regularity of $\mathbf{f}$.

Corollary 5.5. *Assume* $\mathbf{f} : (0,T) \to L^2(\Omega)$ *is Lipschitz continuous. Then, the operator* $\mathcal{F}(\,\cdot\,,\,\cdot\,)$ *is Lipschitz continuous on* $L^2(\Omega_1) \times (0,T)$.

Proof. This is a direct consequence of Lemma 5.12 and the arguments given in the proof of Lemma 5.10. □

To close this section, we prove two results that improve the regularity of $\mathbf{u}$ and $\dfrac{\partial \eta}{\partial t}$.

Corollary 5.6. *Assume* $\mathbf{f} : (0,T) \to L^2(\Omega)$ *is Lipschitz continuous. Then, the velocity* $\mathbf{u}$ *given by Theorem 5.3 is Lipschitz continuous on* $(0,T)$.

Proof. Using the same notation introduced in Lemma 5.12, let $\mathbf{u}^1 = \mathbf{u}(t_1)$, $\mathbf{u}^2 = \mathbf{u}(t_2)$ and $\eta^1 = \eta(t_1)$, $\eta^2 = \eta(t_2)$. Then, owing to the Lipschitz continuity of $\mathbf{f}$ and (5.45), we obtain

$$\|\mathbf{u}(t_2) - \mathbf{u}(t_1)\|_{\mathbf{X}}^2 \leq 2C_{\text{Lip}} \|\eta(t_2) - \eta(t_2)\|_{L^2(\Omega_1)}^2 + 4C_\beta \, \mathbf{f}_{\text{Lip}}^2 \, |t_2 - t_1|^2 \, |\Omega| \,. \tag{5.46}$$

Finally, a direct application of Corollary 5.4 in (5.46) yields

$$\|\mathbf{u}(t_2) - \mathbf{u}(t_1)\|_{\mathbf{X}}^2 \leq 2C_{\text{Lip}} \, \eta_{\text{Lip}}^2 \, |t_2 - t_1|^2 \, |\Omega_1| + 4C_\beta \, \mathbf{f}_{\text{Lip}}^2 \, |t_2 - t_1|^2 \, |\Omega| \,,$$

proving the claim. □

Corollary 5.7. *Assume* $\mathbf{f} : (0,T) \to L^2(\Omega)$ *is Lipschitz continuous. Then, the function* $\dfrac{\partial \eta}{\partial t}$ *given in* (5.43) *is Lipschitz continuous on* $(0,T)$. *In particular* $\dfrac{\partial^2 \eta}{\partial t^2} \in L^\infty(0,T;L^2(\Omega_1))$.

Proof. Combining Corollary 5.5 and Corollary 5.4, it follows that the right hand side of (5.43) is Lipschitz continuous. Hence, $\dfrac{\partial \eta}{\partial t}$ is Lipschitz and consequently, by Rademacher's theorem, differentiable almost everywhere. Moreover, the Lipschitz continuity of $\dfrac{\partial \eta}{\partial t} : (0,T) \to L^2(\Omega_1)$ implies that $\|\dfrac{\partial^2 \eta}{\partial t^2}\|_{L^2(\Omega_1)}(t)$ is bounded in $(0,T)$. Hence $\dfrac{\partial^2 \eta}{\partial t^2} \in L^\infty(0,T;L^2(\Omega_1))$, concluding the proof. □

We summarize the last two propositions in the following remark.

Remark 5.8. *Under the additional regularity assumption*

A5*: The forcing term* $\mathbf{f} : (0,T) \to L^2(\Omega)$ *is Lipschitz continuous,*

it follows that $\mathbf{u} : (0,T) \to \mathbf{X}$ *is Lipschitz continuous and* $\dfrac{\partial^2 \eta}{\partial t^2} \in L^\infty(0,T;\, L^2(\Omega_1))$. *In particular* $\eta \in H^2(0,T;L^2(\Omega_1))$.

The next section further extends the properties of η and shows that η is a nonnegative bounded function. This is relevant in view that, physically, the porosity is always between zero and one.

5.6 Nonnegativity of the porosity

In this section we show that the porosity function η remains bounded during the filtration process, and, more importantly, that it is nonnegative a.e. in Ω_1. A brief outline of how this is achieved follows. First, we introduce some notation and

discretize in time the deposition equation (5.43). Subsequently, we show that the discretized problem is well-posed and exhibit the boundedness and nonnegativity of the discrete porosity function. Then, we proceed to construct sequences of functions that approximate the continuous porosity and show that they converge to a common limit. Finally, we close this section by proving that under assumption ***A5***, the discrete porosity converges to the continuous porosity as the time step goes to zero.

First, we discretize the interval $[0, T]$ into $M + 1$ uniformly spaced times $t_k = k\,\Delta t$, $k = 0, 1, \ldots, M$, where $\Delta t = T/M$ and consider the following problem.

Lemma 5.13. *Define* $\mathbf{u}_k = \mathbf{u}(t_k) \in \mathbf{X}$*, where* $\mathbf{u}$ *is the solution obtained in Theorem 5.3. Then, the problem: Given* $\eta_0 \in L^2(\Omega_1)$*, find* $\eta_k \in L^2(\Omega_1)$*, for* $k = 1, \ldots, M$ *such that for all* $\xi \in L^2(\Omega_1)$

$$\left(\frac{\eta_k - \eta_{k-1}}{\Delta t}, \xi\right)_{\Omega_1} + \Big(g(|\mathbf{u}_k|)h(\eta_k), \xi\Big)_{\Omega_1} = 0 \tag{5.47}$$

has a unique solution, provided $\Delta t < g_{\max}\, h_{\mathrm{Lip}}$.

Proof. Assume $\eta_0, \ldots, \eta_{k-1}$ have already been computed. Define the operator $A : L^2(\Omega_1) \to L^2(\Omega_1)$ by $y = Ax$, where y satisfies

$$\left(\frac{y - \eta_{k-1}}{\Delta t}, \xi\right)_{\Omega_1} + \Big(g(|\mathbf{u}_k|)h(x), \xi\Big)_{\Omega_1} = 0 \quad \forall \xi \in L^2(\Omega_1). \tag{5.48}$$

Let $x_1, x_2 \in L^2(\Omega_1)$ and define $y_1 = Ax_1$, $y_2 = Ax_2$. Then, from (5.48), it follows that

$$(y_1 - y_2, \xi)_{\Omega_1} = -\Delta t \Big(g(|\mathbf{u}_k|)\ (h(x_1) - h(x_2))\,, \xi\Big)_{\Omega_1} \quad \forall \xi \in L^2(\Omega_1). \tag{5.49}$$

Thus, setting $\xi = y_1 - y_2$ in (5.49), using Cauchy-Schwarz and assumptions ***A1*** and

A2, yields

$$\|y_1 - y_2\|_{L^2(\Omega_1)} \leq \Delta t \, g_{\max} \, h_{\text{Lip}} \, \|x_1 - x_2\|_{L^2(\Omega_1)},$$

implying that A is a contraction in $L^2(\Omega_1)$. Consequently, owing to Banach's fixed point theorem, it follows that A has a unique fixed point, proving the existence of a unique solution to (5.47). The result follows by induction. □

Definition 5.3. *We use the notation $\boldsymbol{\eta}^M$ to denote the tuple $(\eta_0, \eta_1, \ldots, \eta_M)$, where the η_k, $k = 1, \ldots, M$ are defined through Lemma 5.13.*

Lemma 5.14. *Let $m \in \mathbb{Z}^+$ be given with $m \leq M$. The solution η_k given in Lemma 5.13 satisfies the estimate*

$$\|\eta_m\|^2_{L^2(\Omega_1)} + \sum_{k=1}^{m} \|\eta_k - \eta_{k-1}\|^2_{L^2(\Omega_1)} \leq \exp\left(4T \, g_{\max} \, h_{\max} \, |\Omega_1|\right) \|\eta_0\|^2_{L^2(\Omega_1)}, \tag{5.50}$$

provided

$$\Delta t < \frac{1}{4 \, g_{\max} \, h_{\max} \, |\Omega_1|}.$$

Proof. Let $\xi = \eta_k$ in (5.47) and use assumption ***A1***, to obtain

$$\|\eta_k\|^2_{L^2(\Omega_1)} - \|\eta_{k-1}\|^2_{L^2(\Omega_1)} + \|\eta_k - \eta_{k-1}\|^2_{L^2(\Omega_1)} \leq 2\Delta t \, g_{\max} \, h_{\max} \, |\Omega_1| \, \|\eta_k\|_{L^2(\Omega_1)}. \tag{5.51}$$

Summing (5.51) from $k = 1$ to $k = m$, yields

$$\|\eta_m\|^2_{L^2(\Omega_1)} + \sum_{k=1}^{m} \|\eta_k - \eta_{k-1}\|^2_{L^2(\Omega_1)} \leq \|\eta_0\|^2_{L^2(\Omega_1)} + C \, \Delta t \, \sum_{k=1}^{m} \|\eta_k\|_{L^2(\Omega_1)}, \tag{5.52}$$

where $C = 2\, g_{\max}\, h_{\max}\, |\Omega_1|$. Finally, from a discrete version of Gronwall's lemma (see [60] pg. 167), we obtain

$$\|\eta_m\|^2_{L^2(\Omega_1)} + \sum_{k=1}^{m} \|\eta_k - \eta_{k-1}\|^2_{L^2(\Omega_1)} \leq \exp\left(\frac{m\, C\, \Delta t}{1 - C\, \Delta t}\right) \|\eta_0\|^2_{L^2(\Omega_1)}.$$

Bounding $m\, \Delta t$ with $M\, \Delta t = T$, and observing that $1 - C\, \Delta t < 1/2$, estimate (5.50) now follows. □

Lemma 5.15. *Let $\eta_{k-1} \in L^2(\Omega_1)$ be given with $\eta_{k-1} \geq 0$ a.e. in Ω_1. Then, the solution η_k given in Lemma 5.13 is nonnegative a.e. in Ω_1.*

Proof. Define $\eta_k^- = \max\{0, -\eta_k\}$, i.e., the negative part of η_k. Let $\xi = -\eta_k^-$ in (5.47) and define $h(\,\cdot\,)$ to be zero for negative arguments. Thus, we obtain

$$\|\eta_k^-\|^2_{L^2(\Omega_1)} = -\left(\eta_{k-1}, \eta_k^-\right)_{\Omega_1} + \Delta t \left(g(|\mathbf{u}_k|)h(\eta_k), \eta_k^-\right)_{\Omega_1} = -\left(\eta_{k-1}, \eta_k^-\right)_{\Omega_1} \leq 0.$$

Consequently, η_k^- is zero a.e. in Ω_1, implying the nonnegativity of η_k. □

Lemma 5.16. *Let $\eta_{k-1} \in L^2(\Omega_1)$ be given. Then, the solution η_k given in Lemma 5.13 satisfies $\eta_k \leq \eta_{k-1}$ a.e. in Ω_1. Equivalently, η_k is monotonically decreasing on k.*

Proof. Define $(\eta_k - \eta_{k-1})^+ = \max\{0, \eta_k - \eta_{k-1}\}$. Setting $\xi = (\eta_k - \eta_{k-1})^+$ in (5.47), yields

$$\begin{aligned}\|(\eta_k - \eta_{k-1})^+\|^2_{L^2(\Omega_1)} &= \left(\eta_k - \eta_{k-1}, (\eta_k - \eta_{k-1})^+\right)_{\Omega_1} \\ &= -\Delta t \left(g(|\mathbf{u}_k|)h(\eta_k), (\eta_k - \eta_{k-1})^+\right)_{\Omega_1} \leq 0.\end{aligned}$$

Therefore, $(\eta_k - \eta_{k-1})^+$ is zero a.e. in Ω_1, implying that $\eta_k \leq \eta_{k-1}$ a.e. in Ω_1. □

Corollary 5.8. *Let $\eta_{k-1} \in L^2(\Omega_1)$ be given with $0 \leq \eta_{k-1} \leq \eta_0$ a.e. in Ω_1. Then, the solution η_k given in Lemma 5.13 is also bounded above by η_0 a.e. in Ω_1.*

Proof. This is a direct consequence of Lemma 5.16. □

Remark 5.9. *In view of assumption **A3**, Lemma 5.15 and Corollary 5.8, it follows that $\eta_k \in L^\infty(\Omega_1)$ for $k = 0, \ldots, M$.*

Now that we have established some relevant properties of the functions in the vector $\boldsymbol{\eta}^M$, it remains to exhibit that the continuous solution η possesses the same properties. We achieve this by constructing a sequence of interpolants using $\boldsymbol{\eta}^M$ and showing that they converge to η.

Definition 5.4. *Let $\Delta t > 0$ and $\eta_0 \in L^2(\Omega_1)$ be given. Then, the piecewise constant and piecewise linear interpolants $\eta^*_{\Delta t}, \eta^{**}_{\Delta t} : [0, T] \to L^2(\Omega_1)$ are given by*

$$\eta^*_{\Delta t}(t) = \begin{cases} \eta_0, & t = 0, \\ \eta_k, & (k-1)\,\Delta t < t \leq k\,\Delta t \end{cases}$$
$$\eta^{**}_{\Delta t}(t) = \left(\frac{\eta_k - \eta_{k-1}}{\Delta t}\right)(t - (k-1)\Delta t) + \eta_{k-1}, \quad (k-1)\,\Delta t \leq t \leq k\,\Delta t,$$

for $k = 1, \ldots, M$. Similarly, we define $\mathbf{u}^*_{\Delta t} : [0, T] \to \mathbf{X}$ by

$$\mathbf{u}^*_{\Delta t}(t) = \begin{cases} \mathbf{u}(0), & t = 0, \\ \mathbf{u}(t_k), & (k-1)\,\Delta t < t \leq k\,\Delta t \end{cases}$$

for $k = 1, \ldots, M$, where $\mathbf{u}$ is the solution found in Theorem 5.3.

Remark 5.10. *In view of Definition 5.4 and Lemma 5.13, $\eta^{**}_{\Delta t}$, $\eta^*_{\Delta t}$ and $\mathbf{u}^*_{\Delta t}$ satisfy*

for all $\xi \in L^2(\Omega_1)$ *and for all* $t \in (0,T)$

$$\left(\frac{\partial \eta^{**}_{\Delta t}}{\partial t}, \xi\right)_{\Omega_1} + \left(g(|\mathbf{u}^*_{\Delta t}|)h(\eta^*_{\Delta t}), \xi\right)_{\Omega_1} = 0. \tag{5.53}$$

The following lemma establishes norm estimates that we use to extract weakly convergent subsequences.

Lemma 5.17. *There exists a positive constant* C*, independent of* Δt*, such that:*

$$\|\frac{\partial \eta^{**}_{\Delta t}}{\partial t}\|_{L^2(0,T;L^2(\Omega_1))} \leq C, \tag{5.54}$$

$$\|\eta^*_{\Delta t}\|_{L^\infty((0,T)\times\Omega_1)}, \|\eta^{**}_{\Delta t}\|_{L^\infty((0,T)\times\Omega_1)} \leq C, \tag{5.55}$$

$$\|\eta^*_{\Delta t} - \eta^{**}_{\Delta t}\|_{L^2(0,T;L^2(\Omega_1))} \leq C\sqrt{\Delta t}, \tag{5.56}$$

$$\|\eta^{**}_{\Delta t}\|_{H^1(0,T;L^2(\Omega_1))} \leq C. \tag{5.57}$$

Proof. First note that

$$\begin{aligned}\|\frac{\partial \eta^{**}_{\Delta t}}{\partial t}\|^2_{L^2(0,T;L^2(\Omega_1))} &= \sum_{k=1}^{M} \int_{t_{k-1}}^{t_k} \|\frac{\partial \eta^{**}_{\Delta t}}{\partial t}(s)\|^2_{L^2(\Omega_1)}\, ds \\ &= \sum_{k=1}^{M} \int_{t_{k-1}}^{t_k} \|\frac{\eta_k - \eta_{k-1}}{\Delta t}\|^2_{L^2(\Omega_1)}\, ds = \frac{1}{\Delta t}\sum_{k=1}^{M} \|\eta_k - \eta_{k-1}\|^2_{L^2(\Omega_1)}.\end{aligned} \tag{5.58}$$

Setting $\xi = \eta_k - \eta_{k-1}$ in (5.47), we obtain

$$\frac{1}{\Delta t}\|\eta_k - \eta_{k-1}\|_{L^2(\Omega_1)} \leq \|g(|\mathbf{u}_k|)h(\eta_k)\|_{L^2(\Omega_1)} \leq |\Omega_1|^{1/2}\, g_{\max}\, h_{\max}. \tag{5.59}$$

Thus, squaring (5.59), multiplying by Δt and summing from $k = 1$ to $k = M$, yields

$$\frac{1}{\Delta t}\sum_{k=1}^{M} \|\eta_k - \eta_{k-1}\|^2_{L^2(\Omega_1)} \leq |\Omega_1|\, g^2_{\max}\, h^2_{\max}\, T. \tag{5.60}$$

Combining (5.58) and (5.60), (5.54) follows. Estimate (5.55) is a direct consequence of Definition 5.4 and Remark 5.9. Also owing to Definition 5.4 and Lemma 5.14,

$$\begin{aligned}\|\eta^*_{\Delta t} - \eta^{**}_{\Delta t}\|^2_{L^2(0,T;L^2(\Omega_1))} &= \sum_{k=1}^{M} \int_{(k-1)\,\Delta t}^{k\,\Delta t} \|\eta^*_{\Delta t}(t) - \eta^{**}_{\Delta t}(t)\|^2_{L^2(\Omega_1)}\, dt \\ &= \sum_{k=1}^{M} \|\eta_k - \eta_{k-1}\|^2_{L^2(\Omega_1)} \int_{(k-1)\,\Delta t}^{k\,\Delta t} \left(\frac{t - k\Delta t}{\Delta t}\right)^2 dt \leq \frac{\Delta t}{3} C,\end{aligned}$$

where C is given by (5.50). Hence (5.56) holds. Finally, (5.57) is a direct consequence of (5.54) and (5.55). □

With the aim of letting $\Delta t \to 0$, we consider the sequence of functions $\{\eta^{**}_n\}_{n=1}^{\infty}$ and $\{\eta^*_n\}_{n=1}^{\infty}$, where $\eta^{**}_n = \eta^{**}_{\Delta t}$, $\eta^*_n = \eta^*_{\Delta t}$ and $\Delta t = 1/n$. The subsequent analysis uses the concept of weak-$*$ convergence.

Definition 5.5. *Let $U \subset \mathbb{R}^n$ be a bounded domain. A sequence $\{f_n\}_{n=1}^{\infty} \subset L^\infty(U)$ converges weak-$*$ to f in $L^\infty(U)$ if*

$$\langle f_n, \phi\rangle \to \langle f, \phi\rangle \quad \forall \phi \in L^1(U).$$

The next result is a corollary of Lemma 5.17.

Corollary 5.9. *Assume $\eta^*_n \to \eta^1$ weak-$*$ in $L^\infty((0,T)\times\Omega_1)$ and $\eta^{**}_n \to \eta^2$ weak-$*$ in $L^\infty((0,T)\times\Omega_1)$. Then, $\eta^1 = \eta^2$ almost everywhere.*

Proof. Let $\phi \in L^2(0,T;L^2(\Omega_1)) \subset L^1((0,T)\times\Omega_1)$. Then,

$$\left\langle \eta^1 - \eta^2, \phi\right\rangle = \left\langle \eta^1 - \eta^*_n, \phi\right\rangle + \left\langle \eta^*_n - \eta^{**}_n, \phi\right\rangle + \left\langle \eta^{**}_n - \eta^2, \phi\right\rangle \tag{5.61}$$

Taking the limit $n \to \infty$ in (5.61), using the weak-$*$ convergence of η^*_n, η^{**}_n, and

Cauchy-Schwarz, we obtain

$$\left|\left\langle \eta^1 - \eta^2, \phi \right\rangle\right| \leq \lim_{n\to\infty} \|\eta_n^* - \eta_n^{**}\|_{L^2(0,T;L^2(\Omega_1))} \|\phi\|_{L^2(0,T;L^2(\Omega_1))}. \tag{5.62}$$

Owing to (5.56) and (5.62), it follows that $\langle \eta^1 - \eta^2, \phi \rangle = 0$ for all $\phi \in L^2(0,T;L^2(\Omega_1))$. Hence $\|\eta^1 - \eta^2\|_{L^2(0,T;L^2(\Omega_1))} = 0$, implying that $\eta^1 = \eta^2$ almost everywhere. □

We state two lemmas. One relates continuity and weak convergence and the other describes a continuous embedding.

Lemma 5.18. *Let X and Y be normed vector spaces with X' and Y' its corresponding duals. Let $T : X \to Y$ be a bounded linear operator and let $\{x_n\}_{n=1}^{\infty}$ be a sequence in X such that $x_n \rightharpoonup x$. Then, $T(x_n) \rightharpoonup T(x)$.*

Proof. Let $y^* \in Y'$. Define the bounded linear functional $f \in X'$ by $f = y^* \circ T$. Owing to the fact that $x_n \rightharpoonup x$, it follows that $f(x_n) \to f(x)$, i.e., $y^*(T(x_n)) \to y^*(T(x))$. Observing that y^* is arbitrary, the proposition follows. □

Lemma 5.19. *Let $U \subset \mathbb{R}^n$ be a domain and assume $n < 2m$ for $m \in \mathbb{Z}^+$. Then, the Hilbert space $H^m(U)$ is continuously embedded into $C^0(U)$. In particular, $H^1(0,T)$ is continuously embedded into $L^\infty(0,T)$.*

Proof. This is a direct consequence of the Sobolev embedding theorem. For a proof, see Case C of Theorem 5.4 in [1]. □

We are now in position to show that η_n^* and η_n^{**} weak-$*$ converge to a common limit.

Lemma 5.20. *There exists a subsequence of $\left\{\boldsymbol{\eta}^M\right\}_{M\geq 1}$ and a function $\eta^* \in L^\infty((0,T)$*

$\times\Omega_1)$, *such that*

$$\eta_n^* \to \eta^* \text{ weak-}* \text{ in } L^\infty((0,T)\times\Omega_1), \tag{5.63}$$

$$\eta_n^{**} \to \eta^* \text{ weak-}* \text{ in } L^\infty((0,T)\times\Omega_1), \tag{5.64}$$

$$\eta_n^{**} \to \eta^* \text{ weakly in } H^1(0,T;L^2(\Omega_1)), \tag{5.65}$$

$$\frac{\partial \eta_n^{**}}{\partial t} \to \frac{\partial \eta^*}{\partial t} \text{ weakly in } L^2(0,T;L^2(\Omega_1)). \tag{5.66}$$

Proof. For the sake of simplicity, all the subsequences that we derive in the analysis are labeled as the original sequences. In view of (5.55), it follows from the Banach-Alaoglu theorem the existence of a subsequence of $\{\boldsymbol{\eta}^M\}$ and a function $\eta^* \in L^\infty((0,T)\times\Omega_1)$ such that (5.63) is satisfied. By the same token, there exists a subsequence of η_n^{**} and a function $\widetilde{\eta} \in L^\infty((0,T)\times\Omega_1)$ such that $\eta_n^{**} \to \widetilde{\eta}$ weak-$*$ in $L^\infty((0,T)\times\Omega_1)$. From Corollary 5.9 it follows that $\widetilde{\eta} = \eta^*$, establishing (5.64). Now observe that (5.57) and the Banach-Alaoglu theorem yield a further subsequence and a function $\widehat{\eta} \in H^1(0,T;L^2(\Omega_1))$, such that

$$\eta_n^{**} \to \widehat{\eta} \text{ weakly in } H^1(0,T;L^2(\Omega_1)). \tag{5.67}$$

From the continuous embedding $H^1(0,T) \hookrightarrow L^\infty(0,T)$ (see Lemma 5.19), it follows that

$$\eta_n^{**} \to \widehat{\eta} \text{ weak-}* \text{ in } L^\infty(0,T;L^2(\Omega_1)). \tag{5.68}$$

Moreover, owing to (5.64) and the fact that $L^\infty((0,T)\times\Omega_1) \subset L^\infty(0,T;L^2(\Omega_1))$, we

obtain

$$\eta_n^{**} \to \eta^* \text{ weak-}* \text{ in } L^\infty(0,T;L^2(\Omega_1)). \tag{5.69}$$

Thus, in view of (5.68), (5.69) and the uniqueness of weak-$*$ limits, it follows that $\widehat{\eta} = \eta^*$. This establishes (5.65). Finally, in view of Lemma 5.18 and the fact that the time derivative is a bounded linear operator from $H^1(0,T)$ to $L^2(0,T)$, expression (5.66) follows from (5.65). □

The following lemma gives an error estimate in the L^2 norm for a first order approximation of the time derivative. We use this result in the next proposition, where we establish that the L^2 difference between the continuous function η and its discrete analog η_k is proportional to Δt.

Lemma 5.21. *Let $f \in H^2(0,T;L^2(\Omega))$ and let $\Delta t > 0$ be given. Then,*

$$\|\frac{\partial f}{\partial t}(t_k) - \frac{f(t_k) - f(t_{k-1})}{\Delta t}\|^2_{L^2(\Omega)} \leq \frac{\Delta t}{3}\int_{t_{k-1}}^{t_k} \|\frac{\partial^2 f}{\partial t^2}(s)\|^2_{L^2(\Omega)}\, ds,$$

for $(t_{k-1}, t_k) \subset (0,T)$.

Proof. By Taylor's theorem with the integral form of the remainder, we obtain

$$f(t_{k-1}) = f(t_k) - \Delta t\frac{\partial f}{\partial t}(t_k) + \int_{t_{k-1}}^{t_k}(s - t_{k-1})\frac{\partial^2 f}{\partial t^2}(s)\, ds.$$

Thus,

$$\begin{aligned}\|\frac{f(t_k) - f(t_{k-1})}{\Delta t} - \frac{\partial f}{\partial t}(t_k)\|^2 &= \frac{1}{\Delta t^2}\int_\Omega \left(\int_{t_{k-1}}^{t_k}(s - t_{k-1})\frac{\partial^2 f}{\partial t^2}(s)\, ds\right)^2 d\Omega \\ &\leq \frac{1}{\Delta t^2}\int_\Omega \left(\int_{t_{k-1}}^{t_k}(s - t_{k-1})^2\, ds \int_{t_{k-1}}^{t_k}\left(\frac{\partial^2 f}{\partial t^2}(s)\right)^2 ds\right) d\Omega. \end{aligned} \tag{5.70}$$

The result now follows from (5.70) by interchanging the order of integration. □

Now we show that the discrete solution η_k introduced in Lemma 5.13 converges to the continuous one given by Theorem 5.3.

Lemma 5.22. *Assume* $\dfrac{\partial^2 \eta}{\partial t^2} \in L^2(0,T;L^2(\Omega_1))$. *Let* $M, n \in \mathbb{Z}^+$ *and* $\Delta t \in \mathbb{R}^+$ *be such that* $\Delta t = T/M = 1/n$ *and* $\Delta t < \frac{1}{2}(1 + 2\, g_{\max}\, h_{\text{Lip}})$. *Define* $e_k = \eta(t_k) - \eta_k$ *for* $k = 0, \ldots, M$, *where* $\eta(t)$ *is the solution found in Theorem 5.3 and* η_k *is an element of* $\boldsymbol{\eta}^M$ *(see Definition 5.3). Then,*

$$\|e_k\|_{L^2(\Omega_1)} \leq \mathfrak{C}_1\, \Delta t, \ \textit{where} \tag{5.71}$$
$$\mathfrak{C}_1^2 = \exp\Big(2\,T\,(1 + 2\, g_{\max}\, h_{\text{Lip}})\Big)\left(\frac{1}{3}\|\frac{\partial^2 \eta}{\partial t^2}\|^2_{L^2(0,T;L^2(\Omega_1))}\right),$$

and

$$\|\eta - \eta_n^*\|_{L^2(0,T;\Omega_1)} \leq \mathfrak{C}_2\, \Delta t \quad \text{for} \quad \mathfrak{C}_2^2 = 2\|\eta'\|^2_{L^2(0,T;\Omega_1)} + 2T\, \mathfrak{C}_1^2. \tag{5.72}$$

Proof. From (5.17), we obtain

$$\begin{aligned}\left(\frac{\eta(t_k) - \eta(t_{k-1})}{\Delta t}, \xi\right)_{\Omega_1} &+ \Big(g(|\mathbf{u}(t_k)|)\, h(\eta(t_k)), \xi\Big)_{\Omega_1} \\ &= \left(\frac{\eta(t_k) - \eta(t_{k-1})}{\Delta t} - \frac{\partial \eta}{\partial t}(t_k), \xi\right)_{\Omega_1}\end{aligned} \tag{5.73}$$

Subtracting (5.47) from (5.73), yields

$$\begin{aligned}\left(\frac{e_k - e_{k-1}}{\Delta t}, \xi\right)_{\Omega_1} &+ \Big(g(|\mathbf{u}(t_k)|)\, \big(h(\eta(t_k)) - h(\eta_k)\big), \xi\Big)_{\Omega_1} \\ &= \left(\frac{\eta(t_k) - \eta(t_{k-1})}{\Delta t} - \frac{\partial \eta}{\partial t}(t_k), \xi\right)_{\Omega_1}.\end{aligned} \tag{5.74}$$

Setting $\xi = e_k$ in (5.74), and using assumptions $\boldsymbol{A1}$ and $\boldsymbol{A2}$, we obtain

$$\frac{1}{2\Delta t}\big(\|e_k\|^2_{L^2(\Omega_1)} - \|e_{k-1}\|^2_{L^2(\Omega_1)} + \|e_k - e_{k-1}\|^2_{L^2(\Omega_1)}\big) \\ \leq g_{\max}\, h_{\mathrm{Lip}}\, \|e_k\|^2_{L^2(\Omega_1)} + \left(\frac{\eta(t_k) - \eta(t_{k-1})}{\Delta t} - \frac{\partial \eta}{\partial t}(t_k), e_k\right)_{\Omega_1}. \quad (5.75)$$

Applying Young's inequality and Lemma 5.21 to (5.75), yields

$$\|e_k\|^2_{L^2(\Omega_1)} - \|e_{k-1}\|^2_{L^2(\Omega_1)} + \|e_k - e_{k-1}\|^2_{L^2(\Omega_1)} \\ \leq \Delta t\,(1 + 2\, g_{\max}\, h_{\mathrm{Lip}})\, \|e_k\|^2_{L^2(\Omega_1)} + \frac{(\Delta t)^2}{3} \int_{t_{k-1}}^{t_k} \|\frac{\partial^2 \eta}{\partial t^2}(s)\|^2_{L^2(\Omega_1)}\, dt. \quad (5.76)$$

Let $m \in \mathbb{Z}^+$, with $m \leq M$. Summing (5.76) from $k = 1$ to $k = m$, and using a discrete version of Gronwall's lemma (see [60] pg. 167), we obtain

$$\|e_m\|^2_{L^2(\Omega_1)} + \sum_{k=1}^{m} \|e_k - e_{k-1}\|^2_{L^2(\Omega_1)} \\ \leq \exp\Big(2\, T\,(1 + 2\, g_{\max}\, h_{\mathrm{Lip}})\Big) \left(\frac{(\Delta t)^2}{3} \|\frac{\partial^2 \eta}{\partial t^2}\|^2_{L^2(0,T;L^2(\Omega_1))}\right). \quad (5.77)$$

Statement (5.71) follows from (5.77).

To prove (5.72), recall that η_n^* is a piecewise constant interpolant such that $\eta_n^*(s) = \eta_k$ for $s \in ((k-1)\Delta t, k\Delta t]$. Hence, owing to (5.71),

$$\int_0^T \|\eta - \eta_n^*\|^2_{L^2(\Omega_1)}(t)\, dt = \sum_{k=1}^{M} \int_{(k-1)\Delta t}^{k\Delta t} \|\eta(t) - \eta_k\|^2_{L^2(\Omega_1)}\, dt \\ \leq 2 \sum_{k=1}^{M} \int_{(k-1)\Delta t}^{k\Delta t} \|\eta(t) - \eta(t_k)\|^2_{L^2(\Omega_1)} + \|\eta(t_k) - \eta_k\|^2_{L^2(\Omega_1)}\, dt \\ \leq 2 \sum_{k=1}^{M} \int_{(k-1)\Delta t}^{k\Delta t} \|\eta(t) - \eta(t_k)\|^2_{L^2(\Omega_1)} + \mathfrak{C}_1^2\, (\Delta t)^2\, dt. \quad (5.78)$$

Now note that for $t \in ((k-1)\Delta t, k\Delta t]$ and Cauchy-Schwarz

$$\begin{aligned}\|\eta(t_k) - \eta(t)\|^2_{L^2(\Omega_1)} &= \int_{\Omega_1} \left(\int_t^{t_k} \eta'(s)\, ds \right)^2 d\Omega_1 \\ &\leq \int_{\Omega_1} \left(\int_t^{t_k} (\eta'(s))^2\, ds \int_t^{t_k} 1\, ds \right) d\Omega_1 \leq \Delta t \int_{t_{k-1}}^{t_k} \|\eta'\|^2_{L^2(\Omega_1)}\, ds. \end{aligned} \tag{5.79}$$

Thus, substituting (5.79) in (5.78), we obtain

$$\begin{aligned}\int_0^T \|\eta - \eta_n^*\|^2_{L^2(\Omega_1)}(t)\, dt &\leq 2 \sum_{k=1}^{M} \int_{(k-1)\Delta t}^{k\Delta t} \left(\Delta t \int_{t_{k-1}}^{t_k} \|\eta'(s)\|^2_{L^2(\Omega_1)}\, ds + \mathfrak{C}_1^2\, (\Delta t)^2 \right) dt \\ &= 2(\Delta t)^2 \int_0^T \|\eta'(s)\|^2_{L^2(\Omega_1)}\, ds + 2T\, \mathfrak{C}_1^2\, (\Delta t)^2 \\ &= (\Delta t)^2 \left(2\|\eta'\|^2_{L^2(0,T;\Omega_1)} + 2T\, \mathfrak{C}_1^2 \right). \end{aligned} \tag{5.80}$$

Statement (5.72) follows from (5.80). □

We now state the main result of this section.

Lemma 5.23. *Let assumption **A5** hold. Then, $\eta = \eta^*$, where η is the solution found in Theorem 5.3 and η^* is the weak limit introduced in Lemma 5.20.*

Proof. Remark 5.8 readily yields $\dfrac{\partial^2 \eta}{\partial t^2} \in L^2(0,T;L^2(\Omega_1))$. Let $\phi \in L^2(0,T;\Omega_1) \subset L^1(0,T;\Omega_1)$. Then, owing to Cauchy-Schwarz,

$$\begin{aligned}\langle \eta - \eta^*, \phi \rangle &= \langle \eta - \eta_n^*, \phi \rangle + \langle \eta_n^* - \eta^*, \phi \rangle \\ &\leq \|\eta - \eta_n^*\|_{L^2(0,T;\Omega_1)} \|\phi\|_{L^2(0,T;\Omega_1)} + \langle \eta_n^* - \eta^*, \phi \rangle\,. \end{aligned} \tag{5.81}$$

Hence, in view of the second statement of Lemma 5.22, the weak-$*$ convergence of η_n^*

to η^* (see (5.63)), and the fact that $1/n = \Delta t$, it follows from (5.81) that

$$\langle \eta - \eta^*, \phi \rangle \leq \lim_{n\to\infty} \|\eta - \eta_n^*\|_{L^2(0,T;\Omega_1)} \|\phi\|_{L^2(0,T;\Omega_1)} + \lim_{n\to\infty} \langle \eta_n^* - \eta^*, \phi \rangle = 0. \tag{5.82}$$

We conclude that $\eta = \eta^*$ almost everywhere in $(0,T) \times \Omega_1$. □

We summarize the results of the last two sections in the next theorem.

Theorem 5.4. *Let the assumptions* ***A*1-*A*5** *and* ***S*1-*S*2** *hold. Let $\eta_0 \in L^2(\Omega_1)$ be given with η_0 nonnegative and bounded a.e. in Ω_1. Then, there exist unique solutions $\mathbf{u} \in H^1(0,T;\mathbf{X})$, $p \in H^1(0,T;Q)$, $\lambda \in H^1(0,T;\Lambda)$ and $\eta \in H^2(0,T;L^2(\Omega_1))$, that satisfy the system* (5.14)-(5.17) *for a.e. $t \in (0,T)$ with $\eta(0) = \eta_0$. Moreover, $0 \leq \eta(\mathbf{x},t) \leq \eta_0(\mathbf{x})$ for a.e. $(t,\mathbf{x}) \in (0,T) \times \Omega_1$.*

Proof. The proposition is an immediate consequence of Remark 5.8, Theorem 5.3 and Lemma 5.23. The additional regularity in the time variable for p and λ follows from the regularity of $\mathbf{u}$ and the inf-sup condition given in Lemma 5.4. □

The following chapter introduces the finite element approximation of the coupled Stokes-Darcy problem with deposition and explores the convergence properties of the given numerical scheme though some numerical examples.

Chapter 6

Approximation of coupled Stokes-Darcy flow with deposition

6.1 Finite element approximation

In this section we investigate the finite element approximation to:

Given $\eta_0 \in L^2(\Omega_1)$ *and* $\mathbf{f} \in C^0(0^-, T; L^2(\Omega))$, *find* $\mathbf{u} \in L^2(0,T;\mathbf{X})$, $p \in L^2(0,T;Q)$, $\lambda \in L^2(0,T;\Lambda)$ *and* $\eta \in H^1(0,T;L^2(\Omega_1))$, *satisfying* $\eta(\cdot,0) = \eta_0$ *a.e. in* Ω_1, *and for a.e.* $t \in (0,T)$

$$a_1(\eta;\mathbf{u},\mathbf{v}) + a_2(\mathbf{u},\mathbf{v}) + b(\mathbf{v},p) + c(\mathbf{v},\lambda) + d(\eta;\mathbf{u},\mathbf{v}) = \ell(\mathbf{v}) \ \forall \mathbf{v} \in \mathbf{X}, \tag{6.1}$$

$$b(\mathbf{u},q) = 0 \ \forall q \in Q, \tag{6.2}$$

$$c(\mathbf{u},\nu) = 0 \ \forall \nu \in \Lambda, \tag{6.3}$$

$$(\frac{\partial \eta}{\partial t}, \xi)_{\Omega_1} + \Big(g(|\mathbf{u}|)\,\eta, \xi\Big)_{\Omega_1} = 0 \ \forall \xi \in L^2(\Omega_1). \tag{6.4}$$

Remark 6.1. *Note that with regard to the general modeling equation* (5.9), *we have chosen in* (6.4) $h(\eta) = \eta$.

Remark 6.2. *With minor modifications, the following discussion can be extended to include partitions composed of triangles in 2D or tetrahedra in 3D.*

Let $\Omega_1, \Omega_2 \subset \mathbb{R}^d$, $d \in \{2, 3\}$, denote convex domains where Ω_1 denotes the Darcy or porous domain, and Ω_2 denotes the Stokes or fluid domain. Moreover, let $\Gamma = \overline{\Omega}_1 \cap \overline{\Omega}_2 \subset \mathbb{R}^{d-1}$ be the interface that connects them. We decompose Ω_i, $i = 1, 2$, into either quadrilaterals in 2D or hexahedra in 3D to obtain a partition $\mathcal{T}_h^i$, where h is defined below. Hence, the computational domain is $\Omega = \Omega_1 \cup \Omega_2 \cup \Gamma$, with $\Omega_i = \bigcup_{K \in \mathcal{T}_h^i} K$. We assume that for each partition there exist constants c_1, c_2 such that $c_1 h \leq h_K \leq c_2 \rho_K$, where h_K is the diameter of the cell K, ρ_K is the diameter of the largest sphere included in K, and $h = \max_{K \in \mathcal{T}_h^i} h_k$. Moreover, we assume that the partitions match along Γ, i.e., if F is a face on Γ of a cell in the Stokes domain, then it is also a face of a cell in the Darcy domain. For $k \in \mathbb{N}$, K a cell of a triangulation, and $\mathbf{F}_K : \widehat{K} \to K$ a $C^1(\widehat{K})$ mapping from the reference cell (unit square in 2D, unit cube in 3D) $\widehat{K}$ to the cell K, define the $(k+1)^d$-dimensional polynomial space

$$\mathbb{P}_k^Q = \text{span}\{x_1^{\alpha_1} x_2^{\alpha_2} \dots x_d^{\alpha_d} : 0 \leq \alpha_1, \alpha_2, \dots, \alpha_d \leq k\},$$

and let $P_k(K) = \{f : f|_K = \widehat{f} \circ \mathbf{F}_K^{-1}, \text{ where } \widehat{f} \in \mathbb{P}_k^Q\}$. Moreover, for F a face of K, the space $P_k(F)$ is defined similarly. The notation $RT_k(\mathcal{T}_h^1)$ is used the denote the Raviart-Thomas space of order k on Ω_1 [18]. We use the following finite element

spaces:

$$\begin{aligned}
\mathbf{X}_{1,h} &= \mathbf{X}_1 \cap RT_k(\mathcal{T}_h^1), \quad Q_{i,h} = \left\{q \in L^2(\Omega_i) \,:\, q|_K \in P_k(K),\, \forall K \in \mathcal{T}_h^i\right\},\\
\mathbf{X}_{2,h} &= \mathbf{X}_2 \cap P_{k+1}(K)^d, \quad Z_h = \left\{\mathbf{v} \in \mathbf{X}_{1,h} \,:\, (q, \nabla\cdot\mathbf{v})_{\Omega_1} = 0,\, \forall q \in Q_{1,h}\right\},\\
\mathbf{X}_h &= \mathbf{X}_{1,h} \times \mathbf{X}_{2,h}, \quad Q_{i,h}^0 = \left\{q \in Q_{i,h} \,:\, (q,1)_{\Omega_i} = 0\right\},\\
Q_h &= \left\{(q_1,q_2) \in Q_{1,h} \times Q_{2,h} \,:\, (q_1,1)_{\Omega_1} + (q_2,1)_{\Omega_2} = 0\right\},\\
\Lambda_h &= \left\{f \in \Lambda \,:\, f|_F \in P_{k+1}(F),\, \forall F \in \Gamma\right\},\\
R_h &= \left\{r \in L^2(\Omega_1) \,:\, r|_K \in P_m(K),\, \forall K \in \mathcal{T}_h^1\right\},\\
R_{h,0} &= \left\{r \in L^2(\Omega_1) \,:\, r|_K \in P_0(K),\, \forall K \in \mathcal{T}_h^1\right\},\\
R_h^s &= \left\{r \in C^0(\Omega_1) \,:\, r|_K \in P_{\max\{1,m\}}(K),\, \forall K \in \mathcal{T}_h^1\right\}. \qquad (6.5)
\end{aligned}$$

Notation 6.1. *For F_i a face on Γ, let $W_{h,i}$ be a local partition of F_i, i.e., $F_i = \cup_{S\in W_{h,i}} S$. Note that in 2D, each of the F_i is divided into line segments while in 3D it is divided into quadrilaterals. The collection of all these partitions is denoted by W_h. Furthermore, define the space*

$$\Lambda_{h,0} = \left\{f \in L^2(\Gamma) \,:\, f \in P_0(S),\, \forall S \in W_h\right\}.$$

Notation 6.2. *For U denoting either $\mathcal{T}_h^1$, $\mathcal{T}_h^2$, Γ, or W_h, we define the space of continuous nodal Lagrange elements*

$$Q_k(U) = \left\{f \in C^0(U) \,:\, f \in P_k(K),\, \forall K \in U\right\},$$

and the space of discontinuous nodal Lagrange elements

$$\mathrm{disc}Q_k(U) = \left\{f \in L^2(U) \,:\, f \in P_k(K),\, \forall K \in U\right\}.$$

We omit the dependency on U whenever it is clear from the context.

Remark 6.3. *Note that as $\nabla\cdot\mathbf{X}_{1,h} = Q_{1,h}$, for $\mathbf{v}\in Z_h$ we have that $\|\nabla\cdot\mathbf{v}\|_{L^2(\Omega_1)} = 0$. Thus, $\|\mathbf{v}\|_{\mathbf{X}_1} = \|\mathbf{v}\|_{L^2(\Omega_1)}$.*

For N given, let $\Delta t = T/N$, and $t_n = n\Delta t$, $n = 0, 1, \ldots, N$. Additionally, define

$$d_t f^n = \frac{f^n - f^{n-1}}{\Delta t}, \quad \overline{f}^n = \frac{f^n + f^{n-1}}{2}, \quad \widetilde{f}^n = f^{n-1} + \frac{1}{2}f^{n-2} - \frac{1}{2}f^{n-3}.$$

For the *a priori* error estimates presented in the next section, the solution $(\mathbf{u}, p, \eta)$ to (6.1)-(6.4) is required to be sufficiently regular. The regularity assumptions are, for some $\delta > 0$,

$$\begin{aligned}
&\mathbf{u}\in L^\infty(0,T;L^\infty(\Omega))\cap L^\infty(0,T;H^{k+1}(\Omega)), \quad \mathbf{u}_t\in L^\infty(0,\delta;L^2(\Omega)),\\
&\mathbf{u}_{tt}\in L^2(0,T;L^2(\Omega)), \quad p\in L^\infty(0,T;H^{k+1}(\Omega)),\\
&\eta\in L^\infty(0,T;L^\infty(\Omega_1))\cap L^\infty(0,T;H^{m+1}(\Omega_1)), \quad \eta_t\in L^\infty(0,T;H^{m+1}(\Omega_1)),\\
&\eta_{tt}\in L^2(0,T;L^2(\Omega_1))\cap L^\infty(0,\delta;L^2(\Omega_1)), \quad \eta_{ttt}\in L^2(0,T;L^2(\Omega_1)).
\end{aligned} \tag{6.6}$$

Throughout, we use C to denote a generic nonnegative constant, independent of the mesh parameter h and time step Δt, whose actual value may change from line to line in the analysis.

Initialization of the approximation scheme

The approximation scheme described and analyzed below is a three-level scheme. To initialize the procedure suitable approximations are required for $\mathbf{u}_h^1$, $\mathbf{u}_h^2$ and η_h^2. Here we state our assumptions on these initial approximates. (An initialization pro-

cedure is presented in Appendix A.2)

$$\|\mathbf{u}^n - \mathbf{u}_h^n\|_{\mathbf{X}}^2 + \|\eta^n - \eta_h^n\|_{L^2(\Omega_1)}^2 \leq C(\Delta t)^4 + C\left(h^{2k+2} + h^{2m+2}\right) \text{ for } n = 0, 1, 2. \quad (6.7)$$

Approximation scheme

The approximation scheme we investigate is: *Given* $\eta_0 \in R_h$, *for* $n = 3, \ldots, N$, *determine* $(\mathbf{u}_h^n, p_h^n, \lambda_h^n, \eta_h^n) \in \mathbf{X}_h \times Q_h \times \Lambda_h \times R_h$ *satisfying*:

Scheme 6.1.

$$a_1(\mathbf{u}_h^n, \mathbf{v}) + b(p_h^n, \mathbf{v}) + \langle \lambda_h^n, \mathbf{v} \cdot \mathbf{n}_1 \rangle = (\mathbf{f}^n, \mathbf{v}) \ \forall \mathbf{v} \in \mathbf{X}_{1,h}, \quad (6.8)$$

$$b(q, \mathbf{u}_h^n) = 0 \ \forall q \in Q_{1,h}, \quad (6.9)$$

$$a_2(\mathbf{u}_h^n, \mathbf{v}) + b(p_h^n, \mathbf{v}) + d(\mathbf{u}_h^n, \mathbf{v}) + \langle \lambda_h^n, \mathbf{v} \cdot \mathbf{n}_2 \rangle = (\mathbf{f}^n, \mathbf{v}) \ \forall \mathbf{v} \in \mathbf{X}_{2,h}, \quad (6.10)$$

$$b(q, \mathbf{u}_h^n) = 0 \ \forall q \in Q_{2,h}, \quad (6.11)$$

$$\langle \mathbf{u}_1 \cdot \mathbf{n}_1 + \mathbf{u}_2 \cdot \mathbf{n}_2, \nu \rangle = 0 \ \forall \nu \in \Lambda_h, \quad (6.12)$$

$$(d_t \eta_h^n, r) + (g(|\widetilde{\mathbf{u}}_h^n|)\overline{\eta}_h^n, r) = 0 \ \forall r \in R_h. \quad (6.13)$$

Now we focus on the computability of Scheme 6.1. First, we prove that the porosity η_h^n can be computed from previous approximations and derive some of its properties.

Lemma 6.1. *Given* η_h^{n-1}, $\mathbf{u}_h^{n-3}$, $\mathbf{u}_h^{n-2}$ *and* $\mathbf{u}_h^{n-1}$, *there exists a unique solution* $\eta_h^n \in R_h$ *satisfying* (6.13).

Proof. See Lemma 4.1. □

Nonnegativity of the discrete porosity

In this section we give conditions under which the discrete porosity η_h is nonnegative and bounded above by the initial porosity η_0. These results are consistent with the properties derived for the continuous porosity and mirror the propositions derived in Section 5.6. The key assumption in this section is that the finite element space for the discrete porosity is composed of piecewise constant functions, i.e., we replace R_h in (6.13) with $R_{h,0}$.

Lemma 6.2. *Let $\eta_h^{n-1} \in R_{h,0}$ be given with $\eta_h^{n-1} \geq 0$ in Ω_1. Furthermore, assume $\Delta t < 2/g_{\max}$. Then, the solution $\eta_h^n \in R_{h,0}$ given in Lemma 6.1 is nonnegative in Ω_1.*

Proof. Assume there exists $K \in \mathcal{T}_h^1$ such that $\eta_h^n|_K < 0$. Substituting $r = \eta_h^n \chi_K \in R_{h,0}$ into (6.13), where χ_K is the indicator function of the cell K yields

$$(\eta_h^n, \eta_h^n)_K + \frac{\Delta t}{2} \left(g(|\widetilde{\mathbf{u}}_h^n|)\, \eta_h^n, \eta_h^n\right)_K = \left(\eta_h^{n-1} \left(1 - \frac{\Delta t}{2} g(|\widetilde{\mathbf{u}}_h^n|)\right), \eta_h^n\right)_K . \tag{6.14}$$

Owing to the nonnegativity of $g(\,\cdot\,)$, and the nonnegativity of η_h^{n-1}, it follows that the left hand side of (6.14) is positive while the right hand side is nonpositive. This is a contradiction. Hence, no such cell K can exist. We conclude that η_h^n is nonnegative in Ω_1. □

Lemma 6.3. *Let $\eta_h^{n-1} \in R_{h,0}$ be given, with $\eta_h^{n-1} \geq 0$ in Ω_1. Furthermore, assume $\Delta t < 2/g_{\max}$. Then, the solution $\eta_h^n \in R_{h,0}$ given in Lemma 6.1 satisfies $\eta_h^n \leq \eta_h^{n-1}$ in Ω_1. Equivalently, η_h^n is monotonically decreasing on n.*

Proof. Assume there exists $K \in \mathcal{T}_h^1$ such that $\left(\eta_h^n - \eta_h^{n-1}\right)\big|_K > 0$. Substituting $r = \left(\eta_h^n - \eta_h^{n-1}\right) \chi_K \in R_{h,0}$ into (6.13), we obtain

$$\left(\eta_h^n - \eta_h^{n-1}, \eta_h^n - \eta_h^{n-1}\right)_K = -\frac{\Delta t}{2} \left(g(|\widetilde{\mathbf{u}}_h^n|)\, \left(\eta_h^n + \eta_h^{n-1}\right), \eta_h^n - \eta_h^{n-1}\right)_K . \tag{6.15}$$

Observe that the left hand side of (6.15) is positive. Moreover, owing to Lemma 6.2, the nonnegativity of $g(\,\cdot\,)$ and the nonnegativity of η_h^{n-1}, the right hand side of (6.15) is nonpositive. This is a contradiction. Thus, no such cell K exists, and $\eta_h^n \leq \eta_h^{n-1}$ in Ω_1. □

Corollary 6.1. *Let $\eta_h^{n-1} \in R_{h,0}$ be given with $0 \leq \eta_h^{n-1} \leq \eta_0$ in Ω_1. Furthermore, assume $\Delta t < 2/g_{\max}$. Then, the solution $\eta_h^n \in R_{h,0}$ given in Lemma 6.1 is also bounded above by η_0 in Ω_1.*

Proof. This is a direct consequence of Lemma 6.3. □

Although the porosity η_h^n can be computed, the Stokes and Darcy problems are still coupled through the interfacial pressure λ_h^n. Our objective is to decouple them in order to take advantage of the robust and efficient solvers that are available for each individual problem. We achieve the decoupling by treating λ as a given quantity, and proceed to demonstrate that we can systematically compute a sequence of approximations $\{\lambda_k\}_{k=1}^{\infty} \subset \Lambda_h$, such that (6.12) is satisfied. In the discussion that follows we assume that η_h is known.

6.2 Decoupling the problems

Let $\lambda \in \Lambda_h$ and $\mathbf{f} \in \mathbf{X}'$ be given. We introduce the following problems.

Problem 6.1. *Find $(\mathbf{u}_{1,h}, p_{1,h}) \in \mathbf{X}_{1,h} \times Q_{1,h}^0$ satisfying*

$$a_1(\eta_h; \mathbf{u}_{1,h}, \mathbf{v}_1) + b(p_{1,h}, \mathbf{v}_1) = (\mathbf{f}, \mathbf{v}_1) - \langle \lambda, \mathbf{v}_1 \cdot \mathbf{n}_1 \rangle \;\; \forall \mathbf{v}_1 \in \mathbf{X}_{1,h},$$

$$b(q_1, \mathbf{u}_{1,h}) = 0 \;\forall q_1 \in Q_{1,h},$$

Problem 6.2. *Find* $(\mathbf{u}_{2,h}, p_{2,h}) \in \mathbf{X}_{2,h} \times Q^0_{2,h}$ *satisfying*

$$a_2(\mathbf{u}_{2,h}, \mathbf{v}_2) + b(p_{2,h}, \mathbf{v}_2) + d(\eta_h; \mathbf{u}_{2,h}, \mathbf{v}_2) = (\mathbf{f}, \mathbf{v}_2) - \langle \lambda, \mathbf{v}_2 \cdot \mathbf{n}_2 \rangle \ \ \forall \mathbf{v}_2 \in \mathbf{X}_{2,h},$$
$$b(q_2, \mathbf{u}_{2,h}) = 0 \ \forall q_2 \in Q_{2,h}.$$

Remark 6.4. *Let* $\mathbf{v}_1 \in \mathbf{H}_{\rm div}(\Omega_1)$. *Then,* $\mathbf{v}_1 \cdot \mathbf{n}_1 \in H^{-1/2}(\partial\Omega_1)$ *and*

$$\|\mathbf{v}_1 \cdot \mathbf{n}_1\|_{H^{-1/2}(\partial\Omega_1)} \leq C \, \|\mathbf{v}_1\|_{\mathbf{H}_{\rm div}(\Omega_1)},$$

for some positive constant $C = C(\Omega_1)$. *However, for* $\lambda \in \Lambda = H^{1/2}(\Gamma)$, *the duality pairing* $\langle \lambda, \mathbf{v}_1 \cdot \mathbf{n}_1 \rangle_\Gamma$ *is in general not well-defined. Let* $E^{1/2}_\Gamma : H^{1/2}(\Gamma) \to H^{1/2}(\partial\Omega_1)$ *be the operator defined in Lemma 2.1 of [45]. Then,*

$$\langle \lambda, \mathbf{v}_1 \cdot \mathbf{n}_1 \rangle_\Gamma = \left\langle E^{1/2}_\Gamma \lambda, \mathbf{v}_1 \cdot \mathbf{n}_1 \right\rangle_{\partial\Omega_1} \leq \|E^{1/2}_\Gamma \lambda\|_{H^{1/2}(\partial\Omega_1)} \, \|\mathbf{v}_1 \cdot \mathbf{n}_1\|_{H^{-1/2}(\partial\Omega_1)}$$
$$\leq C \, \|\lambda\|_{H^{1/2}(\Gamma)} \, \|\mathbf{v}_1\|_{\mathbf{H}_{\rm div}(\Omega_1)} = C \, \|\lambda\|_\Lambda \, \|\mathbf{v}_1\|_{\mathbf{H}_{\rm div}(\Omega_1)}.$$

The next lemma provides a norm estimate for $\mathbf{u}_{i,h}$, $i = 1, 2$.

Lemma 6.4. *There exists a positive constant* C *such that*

$$\|\mathbf{u}_{i,h}\|_{\mathbf{X}_i} \leq C \left(\|\mathbf{f}\| + \|\lambda\|_\Lambda \right), \ \text{and} \ \ \langle \lambda, \mathbf{u}_{i,h} \cdot \mathbf{n}_i \rangle \leq C \, \|\lambda\|_\Lambda \, \|\mathbf{u}_{i,h}\|_{\mathbf{X}_i} \quad i = 1, 2.$$

Proof. Owing to the coercivity of $a_1(\eta_h, \cdot, \cdot)$ and $a_2(\cdot, \cdot)$, we set $(\mathbf{v}_i, q_i)$ in Problem 6.1 and Problem 6.2 to $(\mathbf{u}_{i,h}, p_{i,h})$ and obtain

$$C_i \, \|\mathbf{u}_{i,h}\|^2_{\mathbf{X}_i} \leq \|\mathbf{f}\| \, \|\mathbf{u}_{i,h}\|_{\mathbf{X}_i} + |\langle \lambda, \mathbf{u}_{i,h} \cdot \mathbf{n}_i \rangle| \quad i = 1, 2, \tag{6.16}$$

where $C_1 = \beta_{\min}$ and $C_2 = 1$. From Remark 6.4, we obtain

$$\langle \lambda, \mathbf{u}_{1,h} \cdot \mathbf{n}_1 \rangle_\Gamma \leq C \, \|\lambda\|_\Lambda \, \|\mathbf{u}_{1,h}\|_{\mathbf{X}_1}, \tag{6.17}$$

and owing to the trace theorem (see Theorem 5.1)

$$\langle \lambda, \mathbf{u}_{2,h} \cdot \mathbf{n}_2 \rangle_\Gamma \leq \|\lambda\|_{L^2(\Gamma)} \, \|\mathbf{u}_{2,h} \cdot \mathbf{n}_2\|_{L^2(\Gamma)} \leq C_{\mathcal{T}} \, \|\lambda\|_\Lambda \, \|\mathbf{u}_{2,h}\|_{\mathbf{X}_2}. \tag{6.18}$$

Hence, from (6.16), (6.17) and (6.18), the result follows. □

Corollary 6.2. *Let $\lambda \in \Lambda$ be given. The operators $\ell_i : \mathbf{X}_i \to \mathbb{R}$, given by*

$$\ell_i(\mathbf{v}) = (\mathbf{f}, \mathbf{v}) - \langle \lambda, \mathbf{v} \cdot \mathbf{n}_i \rangle \quad i = 1, 2,$$

are bounded linear functionals.

Proof. The result is a direct consequence of Cauchy-Schwarz, (6.17) and (6.18). □

Remark 6.5. *The well-posedness of Problem 6.1 and Problem 6.2 follows from the theory of saddle-point problems, Corollary 6.2 and the inf-sup condition for the spaces $\mathbf{X}_{i,h} \times Q_{i,h}$, $i = 1, 2$ (see [18]).*

In view of Remark 6.5, let $(\mathbf{u}_{1,h}, p_{1,h}) \in \mathbf{X}_{1,h} \times Q^0_{1,h}$ be the unique solution to Problem 6.1 and $(\mathbf{u}_{2,h}, p_{2,h}) \in \mathbf{X}_{2,h} \times Q^0_{2,h}$ be the unique solution to Problem 6.2.

Notation 6.3. *We denote by $\mathbf{u}_{i,h}(\lambda, \mathbf{f})$ and $p_{i,h}(\lambda, \mathbf{f})$, $i = 1, 2$, the dependency of the velocity and pressure on the interfacial pressure λ and the forcing term $\mathbf{f}$.*

Remark 6.6. *Note that, in general, $p_h = (p_{1,h}, p_{2,h}) \in Q_{1,h} \times Q_{2,h}$ is not in Q_h. However, one can enforce this by adding the appropriate constant.*

Now we define a few operators that allow us to translate condition (6.12) into a minimization problem.

Definition 6.1. *The operator* $G : \Lambda_h \to \Lambda_{h,0} \times \Lambda_h$ *is given by*

$$G(\lambda) = \begin{pmatrix} \rho\left(\mathbf{u}_{1,h}(\lambda, \mathbf{f}); \mathbf{u}_{2,h}(\lambda, \mathbf{f})\right) \\ \delta\lambda \end{pmatrix},$$

where $\rho : \mathbf{X}_1 \times \mathbf{X}_2 \times \Gamma \to \mathbb{R}$*, is a piecewise constant function given by*

$$\rho\left(\mathbf{u}_{1,h}(\lambda, \mathbf{f}); \mathbf{u}_{2,h}(\lambda, \mathbf{f})\right)(x) = \sum_{S \in W_h} \frac{\chi_S(x)}{|S|} \int_S \mathbf{u}_{1,h}(\lambda, \mathbf{f}) \cdot \mathbf{n}_1 + \mathbf{u}_{2,h}(\lambda, \mathbf{f}) \cdot \mathbf{n}_2 \, d\Gamma, \quad (6.19)$$

χ_S *is the indicator function of the set* S*, and* $\delta \in (0, 1)$ *is a penalization parameter.*

The concept of differentiation for operators is also required in the subsequent analysis.

Definition 6.2. *Let* X*,* Y *be convex topological vector spaces. Let* $A : X \to Y$ *and* $U \subset X$ *be open. Then, the Gâteaux differential* $dA(x; \nu)$ *of* A *at* $x \in U$ *in the direction of* $\nu \in X$ *is defined as:*

$$dA(x; \nu) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(A(x + \varepsilon\nu) - A(x)\right).$$

Note that $dA(x; \cdot) : X \to Y$*.*

Equipped with Definition 6.2, we compute the Gâteaux differential of G.

Lemma 6.5. *Let* $\lambda, \nu \in \Lambda_h$*. Then, the Gâteaux differential of* G *at* λ*,* $dG(\lambda; \cdot)$:

$\Lambda_h \to \Lambda_{h,0} \times \Lambda_h$, *in the direction of* ν *is given by*

$$dG(\lambda;\nu) = \begin{pmatrix} \rho\left(\mathbf{u}_{1,h}(\nu,\mathbf{0});\mathbf{u}_{2,h}(\nu,\mathbf{0})\right) \\ \delta\nu \end{pmatrix}.$$

Proof. Owing to the bilinearity of $a_1(\eta_h,\,\cdot\,,\,\cdot\,)$, $a_2(\,\cdot\,,\,\cdot\,)$, $b(\,\cdot\,,\,\cdot\,)$ and $d(\eta_h,\,\cdot\,,\,\cdot\,)$, and the definition of Problem 6.1 and Problem 6.2, it follows that

$$\left(\mathbf{u}_{i,h}(\lambda+\varepsilon\nu,\mathbf{f}) - \mathbf{u}_{i,h}(\lambda,\mathbf{f})\right)/\varepsilon = \mathbf{u}_{i,h}(\nu,\mathbf{0}) \text{ for } i=1,2. \tag{6.20}$$

Hence, in view of (6.19) and (6.20), we obtain

$$\begin{aligned}\Big(\rho\left(\mathbf{u}_{1,h}(\lambda+\varepsilon\nu,\mathbf{f});\mathbf{u}_{2,h}(\lambda+\varepsilon\nu,\mathbf{f})\right) - \rho\left(\mathbf{u}_{1,h}(\lambda,\mathbf{f});\mathbf{u}_{2,h}(\lambda,\mathbf{f})\right)\Big)/\varepsilon \\ = \rho\left(\mathbf{u}_{1,h}(\nu,\mathbf{0});\mathbf{u}_{2,h}(\nu,\mathbf{0})\right).\end{aligned} \tag{6.21}$$

Finally, owing to (6.21) and Definition 6.2, the proposition follows. □

We proceed to introduce the adjoint of G.

Lemma 6.6. *Let* $\gamma \in \Lambda_{h,0}$. *Define* $\mathfrak{u}_{1,h}(\gamma) \in \mathbf{X}_{1,h}$, $\mathfrak{p}_{1,h}(\gamma) \in Q^0_{1,h}$ *to be the velocity and pressure obtained by solving the problem*

$$\begin{aligned} a_1(\eta_h;\mathfrak{u}_{1,h},\mathbf{v}_1) + b(\mathfrak{p}_{1,h},\mathbf{v}_1) &= \langle\gamma,\rho\left(\mathbf{v}_1,\mathbf{0}\right)\rangle \;\; \forall \mathbf{v}_1 \in \mathbf{X}_{1,h}, \\ b(q_1,\mathfrak{u}_{1,h}) &= 0 \;\forall q_1 \in Q_{1,h}. \end{aligned} \tag{6.22}$$

Similarly, define $\mathfrak{u}_{2,h}(\gamma) \in \mathbf{X}_{2,h}$, $\mathfrak{p}_{2,h}(\gamma) \in Q^0_{2,h}$ *to be the velocity and pressure obtained*

by solving the problem

$$a_2(\mathfrak{u}_{2,h}, \mathbf{v}_2) + b(\mathfrak{p}_{2,h}, \mathbf{v}_2) + d(\eta_h; \mathfrak{u}_{2,h}, \mathbf{v}_2) = \langle \gamma, \rho\,(\mathbf{0}, \mathbf{v}_2)\rangle \;\; \forall \mathbf{v}_2 \in \mathbf{X}_{2,h}, \qquad (6.23)$$

$$b(q_2, \mathfrak{u}_{2,h}) = 0 \;\forall q_2 \in Q_{2,h}.$$

Then, the adjoint of $dG(\lambda;\,\cdot\,)$, $dG^*(\lambda;\,\cdot\,,\,\cdot\,) : \Lambda_{h,0} \times \Lambda_h \to \Lambda_h$, *is given by*

$$dG^*(\lambda; \gamma, \nu) = -\,(\mathfrak{u}_{1,h}(\gamma) \cdot \mathbf{n}_1 + \mathfrak{u}_{2,h}(\gamma) \cdot \mathbf{n}_2)|_\Gamma + \delta\,\nu.$$

Proof. Let $i = 1, 2$. First recall that the pair $(\mathbf{u}_{i,h}(\nu, \mathbf{0}), p_{i,h}(\nu, \mathbf{0}))$ solves Problem 6.1 and Problem 6.2 with $\mathbf{f} = \mathbf{0}$ and λ replaced by ν. Set the corresponding test functions $(\mathbf{v}_i, q_i)$ in Problem 6.1 and Problem 6.2 to $(\mathfrak{u}_{i,h}(\gamma), \mathfrak{p}_{i,h}(\gamma))$. Add the resulting equations from Problem 6.1 and Problem 6.2. Similarly, set the test functions $(\mathbf{v}_i, q_i)$ in systems (6.22) and (6.23) to $(\mathbf{u}_{i,h}(\nu, \mathbf{0}), p_{i,h}(\nu, \mathbf{0}))$, and add the resulting equations. Hence, the left hand sides from both expressions are equal and we obtain

$$-\,\langle \nu, \mathfrak{u}_{1,h}(\gamma) \cdot \mathbf{n}_1 + \mathfrak{u}_{2,h}(\gamma) \cdot \mathbf{n}_2\rangle = \Big\langle \gamma, \rho\Big(\mathbf{u}_{1,h}(\nu, \mathbf{0}); \mathbf{u}_{2,h}(\nu, \mathbf{0})\Big)\Big\rangle\,. \qquad (6.24)$$

Let $\xi \in \Lambda_h$. Now observe that owing to (6.24)

$$\begin{aligned}\left(dG(\lambda;\nu), \begin{pmatrix}\gamma\\ \xi\end{pmatrix}\right) &= \Big\langle \gamma, \rho\Big(\mathbf{u}_{1,h}(\nu, \mathbf{0}); \mathbf{u}_{2,h}(\nu, \mathbf{0})\Big)\Big\rangle + \delta\,\langle \nu, \xi\rangle \\ &= -\,\langle \nu, \mathfrak{u}_{1,h}(\gamma) \cdot \mathbf{n}_1 + \mathfrak{u}_{2,h}(\gamma) \cdot \mathbf{n}_2\rangle + \delta\,\langle \nu, \xi\rangle \\ &= (\nu, dG^*(\lambda; \gamma, \xi))\,.\end{aligned}$$

This concludes the proof. □

Next we show that $dG(\lambda;\,\cdot\,)$ has closed range. The importance of this property lies in that fact that it allows us to define a least squares solution.

Theorem 6.1. *$dG(\lambda;\,\cdot\,)$ has closed range.*

Proof. Let $\left\{f^k\right\}_{k=1}^{\infty} \subset \text{Range}\,(dG(\lambda;\,\cdot\,))$ with $f^k \to f$ in $\Lambda_{h,0} \times \Lambda_h$. Thus, for every f^k, there exists $\nu^k \in \Lambda_h$ such that

$$f^k = \begin{pmatrix} \rho\left(\mathbf{u}_{1,h}(\nu^k, \mathbf{0}); \mathbf{u}_{2,h}(\nu^k, \mathbf{0})\right) \\ \delta\nu^k \end{pmatrix}.$$

Owing to the fact that $f^k \to f$, it follows that there exists $\nu \in \Lambda_h$ such that $\nu^k \to \nu$ in Λ_h. Now note that due to the linearity of $\mathbf{u}_{i,h}(\,\cdot\,, \mathbf{0})$,

$$\left(\mathbf{u}_{i,h}(\nu, \mathbf{0}) - \mathbf{u}_{i,h}(\nu^k, \mathbf{0})\right) = \mathbf{u}_{i,h}(\nu - \nu^k, \mathbf{0}) \text{ for } i = 1, 2. \tag{6.25}$$

Hence, in view of (6.19) and (6.25),

$$\begin{aligned} &\rho\left(\mathbf{u}_{1,h}(\nu, \mathbf{0}); \mathbf{u}_{2,h}(\nu, \mathbf{0})\right) - \rho\left(\mathbf{u}_{1,h}(\nu^k, \mathbf{0}); \mathbf{u}_{2,h}(\nu^k, \mathbf{0})\right) \\ &\qquad = \rho\left(\mathbf{u}_{1,h}(\nu - \nu^k, \mathbf{0}); \mathbf{u}_{2,h}(\nu - \nu^k, \mathbf{0})\right). \end{aligned} \tag{6.26}$$

Now observe that for $i = 1, 2$, Lemma 6.4 yields

$$\begin{aligned} \left(\int_\Gamma \mathbf{u}_{i,h}(\nu - \nu^k, \mathbf{0}) \cdot \mathbf{n}_i \, d\Gamma\right)^2 &= \left\langle 1, \mathbf{u}_{i,h}(\nu - \nu^k, \mathbf{0}) \cdot \mathbf{n}_i \right\rangle^2 \\ &\leq C\ |\Gamma|\ \|\mathbf{u}_{i,h}(\nu - \nu^k, \mathbf{0})\|_{\mathbf{X}_i}^2 \leq C\ |\Gamma|\ \|\nu - \nu^k\|_\Lambda^2 \xrightarrow{k\to\infty} 0. \end{aligned} \tag{6.27}$$

Thus, owing to (6.19), (6.26), and (6.27)

$$\rho\left(\mathbf{u}_{1,h}(\nu^k,\mathbf{0});\mathbf{u}_{2,h}(\nu^k,\mathbf{0})\right) \to \rho\left(\mathbf{u}_{1,h}(\nu,\mathbf{0});\mathbf{u}_{2,h}(\nu,\mathbf{0})\right) \text{ in } \Lambda_{h,0}.$$

Consequently,

$$f = \begin{pmatrix} \rho\left(\mathbf{u}_{1,h}(\nu,\mathbf{0});\mathbf{u}_{2,h}(\nu,\mathbf{0})\right) \\ \delta\nu \end{pmatrix},$$

and $f \in \text{Range}\left(dG(\lambda;\,\cdot\,)\right)$. We conclude that $dG(\lambda;\,\cdot\,)$ has closed range. □

Now we introduce the concept of a least squares solution.

Definition 6.3. *Let $A : X \to Y$ be a linear bounded operator between Hilbert spaces with closed range. Let A^* denote the adjoint of A and let P be the orthogonal projection operator onto the range of A. We say that x is a least squares solution to the equation $Ax = f$ if $Ax = Pf$.*

The next lemma justifies the name least squares and makes the connection to the normal equation.

Lemma 6.7 (See [48] pg. 39)**.** *The following statements are equivalent.*

1. *x is a least squares solution to $Ax = f$*
2. *$\|Ax - f\|_Y \leq \|Aw - f\|_Y \ \forall w \in X$*
3. *$A^*Ax = A^*f$*

Proof.

$(1 \to 2)$ Note that since $Ax = Pf$, it follows that $(Ax - f) \perp \text{Range}(A)$. Hence, writing $Aw - f = (Aw - Ax) + (Ax - f)$, and using the Pythagorean theorem, we obtain

$$\begin{aligned} \|Aw - f\|_Y^2 &= \|Aw - Ax\|_Y^2 + \|Ax - f\|_Y^2 \\ &= \|Aw - Ax\|_Y^2 + \|Pf - f\|_Y^2. \end{aligned}$$

Thus, setting $w = x$ minimizes $\|Aw - f\|_Y^2$.

$(2 \to 3)$ Let $w \in X$ be such that $Aw = Pf$. Then, writing $Ax - f = (Ax - Pf) + (Pf - f)$, and using the Pythagorean theorem, yields

$$\begin{aligned} \|Ax - f\|_Y^2 &= \|Ax - Pf\|_Y^2 + \|Pf - f\|_Y^2 \\ &= \|Ax - Pf\|_Y^2 + \|Aw - f\|_Y^2 \\ &\geq \|Ax - Pf\|_Y^2 + \|Ax - f\|_Y^2, \end{aligned}$$

which implies that $Ax = Pf$. In particular $Ax - f \in \text{Range}(A)^\perp = \ker(A^*)$. Hence, $A^*(Ax - f) = 0$, which leads to $A^*Ax = A^*f$.

$(3 \to 1)$ From $A^*Ax = A^*f$, it follows that $Ax - f \in \ker(A^*) = \text{Range}(A)^\perp$. Thus, $P(Ax - f) = 0$, implying $Ax = Pf$.

□

Now we consider the following minimization problem

$$\min_{\lambda \in \Lambda_h} \|G(\lambda)\|_{L^2(\Gamma) \times \Lambda}^2, \tag{6.28}$$

which intends to enforce the flux condition (6.12). In order to solve (6.28), we let λ^0 be an initial guess for the optimal interfacial pressure and find a correction function

ν^*, such that

$$\begin{aligned}\|G(\lambda^0) + dG(\lambda^0;\nu^*)\|^2_{L^2(\Gamma)\times\Lambda} &= \min_{\nu\in\Lambda_h} \|G(\lambda^0) + dG(\lambda^0;\nu)\|^2_{L^2(\Gamma)\times\Lambda} \\ &= \min_{\nu\in\Lambda_h} \|dG(\lambda^0;\nu) - \left(-G(\lambda^0)\right)\|^2_{L^2(\Gamma)\times\Lambda}. \end{aligned} \tag{6.29}$$

We solve (6.29) by means of the normal equation: Find $\nu \in \Lambda_h$ satisfying

$$dG^*\left(\lambda^0; dG(\lambda^0;\nu)\right) = -dG^*\left(\lambda^0; G(\lambda_0)\right). \tag{6.30}$$

Remark 6.7. *In view of Theorem 6.1, Definition 6.3, and Lemma 6.7, it follows that* (6.30) *is well-posed.*

The standard algorithm to approximate (6.30) is the conjugate gradient for the least squares problem.

Algorithm 6.1 Conjugate gradient for the least squares problem

1: $r \leftarrow b - Ax^0$

2: $s \leftarrow A^* r$

3: $p \leftarrow s$

4: $c \leftarrow \|s\|^2$

5: **while** $\|b - Ax\| > \varepsilon$ and $\sqrt{c} > \varepsilon$ **do**

6: $\quad q \leftarrow Ap$

7: $\quad \sigma \leftarrow c / |q|^2$

8: $\quad x \leftarrow x + \sigma p$

9: $\quad r \leftarrow r - \sigma q$

10: $\quad s \leftarrow A^* r$

11: $\quad c' \leftarrow \|s\|^2$

12: $\quad \tau \leftarrow c'/c$

13: $\quad c \leftarrow c'$

14: $\quad p \leftarrow s + \tau p$

15: **end while**

Now we state the algorithm we use to solve (6.30).

Algorithm 6.2 Solve the normal equation (6.30)

1: Choose $\lambda^0, \nu^0 \in \Lambda_h$

2: $A := dG(\lambda^0; \cdot\,)$

3: $b \leftarrow -G(\lambda^0)$

4: $x^0 \leftarrow \nu^0$

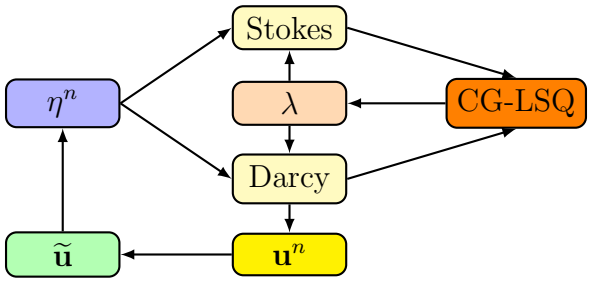
5: Use Algorithm 6.1 to solve $A^* A x = A^* b$

6: $\lambda \leftarrow \lambda^0 + x$ (interfacial pressure)

We summarize the results of this section describing a typical iteration of

Scheme 6.1.

Algorithm 6.3 One iteration of Scheme 6.1



1. Use the previously computed approximations $\mathbf{u}_h^{n-3}$, $\mathbf{u}_h^{n-2}$, and $\mathbf{u}_h^{n-1}$ to obtain $\widetilde{\mathbf{u}}_h^n$.
2. Use $\widetilde{\mathbf{u}}_h^n$ and η_h^{n-1} to compute the porosity η_h^n.
3. Smooth η_h^n to obtain $\eta_h^{n,s}$.
4. Use $\eta_h^{n,s}$ and Algorithm 6.2 to obtain λ_h^n.
5. Compute $\mathbf{u}_{i,h}^n(\lambda_h^n)$ and $p_{i,h}^n(\lambda_h^n)$.

We conclude this section stating an *a priori* error estimate for Scheme 6.1.

Conjecture 6.1. *Let* $(\mathbf{u}, p, \eta)$ *satisfy* (6.1)-(6.6), *and* $(\mathbf{u}_h^n, p_h^n, \eta_h^n)$ *satisfy Scheme 6.1. Furthermore, assume* $C_{\mathcal{S}(\eta_h^n)}$ *given in* (4.12) *is bounded by* $C_{\mathcal{S}}\|\eta^n\|_{m+1}$. *Then, for* Δt *sufficiently small there exists* $C > 0$ *independent of* h *and* Δt, *such that for* $n = 1, 2, \ldots, N$,

$$\|\mathbf{u}^n - \mathbf{u}_h^n\|_{\mathbf{X}} + \|p^n - p_h^n\| + \|\eta^n - \eta_h^n\| \leq C\left((\Delta t)^2 + h^{k+1} + h^{m+1}\right). \tag{6.31}$$

In particular, for $k = m = 1$, *we obtain second order convergence in space and time.*

Remark 6.8. *The proposition given in Conjecture 6.1 is not stated as a theorem in view that we can not guarantee the convergence of Algorithm 6.3 for every time-step.*

Furthermore, the error analysis for the Darcy problem with deposition included in Appendix A.1 applies only to the bilinear form $a_1(\eta, \cdot, \cdot)$ and not the general form $a(\eta, \cdot, \cdot)$ defined in (5.19).

In the next section we explore the convergence properties of Scheme 6.1.

6.3 Numerical experiments

In this section we verify that the convergence properties of Scheme 6.1 are consistent with Conjecture 6.1.

Experiment 6.1 Consider the physical parameters

$$\beta(\eta) = \eta^2 + 0.1, \quad \Psi(\eta) = 1, \quad \mu = 1/2, \quad g(|\mathbf{u}|) = |\mathbf{u}|^2 + 1, \tag{6.32}$$

and let $\Omega_1 = (0,1) \times (0,1)$, $\Omega_2 = (0,1) \times (1,2)$ and $\Gamma = (0,1) \times \{1\}$. We partition Ω_i into square cells. On Ω_2 we use the Taylor-Hood element pair, i.e., Q_2 elements for the velocity and Q_1 elements for the pressure. On Ω_1 we use the Raviart-Thomas element of degree 1, RT_1 for the velocity and discQ_1 elements for the pressure. The interfacial pressure λ is approximated on Γ using Q_2 elements and the porosity η is approximated using discQ_1 elements.

Remark 6.9. *To compute $\rho(\cdot, \cdot)$ (see Definition 6.1 and Notation 6.1), every face F_i on the interface Γ is uniformly refined twice so that $|W_{h,i}| = 16$.*

The boundary conditions for the Darcy problem are imposed weakly, i.e., through the weak form, while the boundary conditions for the Stokes problem are imposed strongly, i.e., by setting the degrees of freedom corresponding to the equa-

tion

$$\mathbf{u} = \mathbf{g} \text{ on } \partial\Omega_2 \setminus \Gamma,$$

where $\mathbf{g}$ is a given function. We consider successively finer meshes $\mathcal{T}_h$ and smaller time-steps Δt, $\Delta t \propto h$, and compute the error between the exact solution

$$\begin{aligned}
\eta &= 0.8 - \delta\,\pi^2\,t^2 \sin(\pi x)\sin(\pi y) - \frac{1}{2}\,t^2 \sin(\pi x)\sin(\pi y)\,, \\
\mathbf{u}_1 &= \left(-x(\sin(y)\exp(1) + 2(y-1)), -\cos(y)\exp(1) + (y-1)^2\right)^T \cos(t), \\
p_1 &= \left(-\sin(y)\exp(1) + \cos(x)\exp(y) + y^2 - 2y + 1\right)\,\cos(t), \\
\mathbf{u}_2 &= \left((y-1)^2 x^3, -\exp(1)\cos(y)\right)^T \cos(t), \\
p_2 &= \left(\cos(x)\exp(y) + y^2 2y + 1\right)\,\cos(t),
\end{aligned}$$

and the numerical approximations under different discrete norms for a time horizon $T = 0.5$. The smoothed porosity is approximated using Q_1 elements by solving the boundary value problem: Find $\eta_h^s \in R_h^s$ satisfying

$$\begin{aligned}
(\delta\,\nabla\eta_h^s, \nabla\varphi) + (\eta_h^s, \varphi) = (\eta_h, \varphi) \quad \forall\varphi \in R_h^s, \\
\eta_h^s = \eta_h \text{ on } \partial\Omega_1,
\end{aligned} \tag{6.33}$$

where $\delta = 0.05$. The corresponding exact smoothed porosity is

$$\eta^s(\mathbf{x}, t) = 0.8 - \frac{1}{2}\,t^2 \sin(\pi x)\sin(\pi y)\,.$$

For $\|\cdot\|_U$ a norm on some generic space U and f_h^n a generic numerical approximation

to $f(t_n)$, $n = 0, \ldots, N$, the discrete norm under consideration is defined as follows,

$$|||f - f_h|||_{L^2(0,T;U)} = \left(\sum_{k=0}^{N} \|f(t_k) - f_h^k\|_U^2 \, \Delta t \right)^{1/2}.$$

Moreover, for h_1 and h_2 two different mesh parameters, the numerical convergence rates are computed using the formula

$$\log \left(\frac{|||f - f_{h_1}|||_{L^2(0,T;U)}}{|||f - f_{h_2}|||_{L^2(0,T;U)}} \right) / \log \left(\frac{h_1}{h_2} \right).$$

The results are summarized in Table 6.1. A graphical depiction of Table 6.1 can be found in Figure 6.1. Note that the numerical convergence rates are in agreement with Conjecture 6.1.

Experiment 6.2 For this experiment we validate our numerical implementation in 3D. In view that Experiment 6.1 supports the correctness of the time-stepping scheme and the spatial discretization in 2D, it remains to verify the implementation of the spatial discretization in 3D. Let $\Omega_1 = (0,1) \times (0,1) \times (0,1)$, $\Omega_2 = (0,1) \times (0,1) \times (1,2)$ and $\Gamma = (0,1) \times (0,1) \times \{1\}$. We consider the same finite element spaces of Experiment 6.1 and the spatial parameters given in (6.32). The domains Ω_i are

Table 6.1: Numerical results for Experiment 6.1.

Stokes domain

h	Δt	$\lvert\lvert\lvert \mathbf{u} - \mathbf{u}_h \rvert\rvert\rvert_{L^2(0,T;H^1(\Omega_2))}$	Rate	$\lvert\lvert\lvert p - p_h \rvert\rvert\rvert_{L^2(0,T;L^2(\Omega_2))}$	Rate
0.354	$5.00 \cdot 10^{-2}$	$6.38 \cdot 10^{-3}$	–	$1.23 \cdot 10^{-2}$	–
0.177	$2.50 \cdot 10^{-2}$	$1.54 \cdot 10^{-3}$	2.05	$3.00 \cdot 10^{-3}$	2.03
0.088	$1.25 \cdot 10^{-2}$	$3.78 \cdot 10^{-4}$	2.02	$7.42 \cdot 10^{-4}$	2.01
0.044	$6.25 \cdot 10^{-3}$	$9.37 \cdot 10^{-5}$	2.01	$1.85 \cdot 10^{-4}$	2.00
0.022	$3.13 \cdot 10^{-3}$	$2.33 \cdot 10^{-5}$	2.01	$4.62 \cdot 10^{-5}$	2.00
	Expected rate		2		2

Darcy domain

h	Δt	$\lvert\lvert\lvert \mathbf{u} - \mathbf{u}_h \rvert\rvert\rvert_{L^2(0,T;\mathbf{H}_{\mathrm{div}}(\Omega_1))}$	Rate	$\lvert\lvert\lvert p - p_h \rvert\rvert\rvert_{L^2(0,T;L^2(\Omega_1))}$	Rate
0.354	$5.00 \cdot 10^{-2}$	$8.67 \cdot 10^{-2}$	–	$8.55 \cdot 10^{-3}$	–
0.177	$2.50 \cdot 10^{-2}$	$2.60 \cdot 10^{-2}$	1.74	$2.11 \cdot 10^{-3}$	2.02
0.088	$1.25 \cdot 10^{-2}$	$7.27 \cdot 10^{-3}$	1.84	$5.28 \cdot 10^{-4}$	2.00
0.044	$6.25 \cdot 10^{-3}$	$1.97 \cdot 10^{-3}$	1.88	$1.33 \cdot 10^{-4}$	1.99
0.022	$3.13 \cdot 10^{-3}$	$5.19 \cdot 10^{-4}$	1.93	$3.34 \cdot 10^{-5}$	1.99
	Expected rate		2		2

		Porosity		Divergence in Ω_1
h	Δt	$\lvert\lvert\lvert \eta - \eta_h \rvert\rvert\rvert_{L^2(0,T;L^2(\Omega_1))}$	Rate	$\lvert\lvert\lvert \nabla \cdot \mathbf{u}_h \rvert\rvert\rvert_{L^2(0,T;L^2(\Omega_1))}$
0.354	$5.00 \cdot 10^{-2}$	$1.68 \cdot 10^{-2}$	–	$2.01 \cdot 10^{-11}$
0.177	$2.50 \cdot 10^{-2}$	$5.64 \cdot 10^{-3}$	1.58	$4.57 \cdot 10^{-11}$
0.088	$1.25 \cdot 10^{-2}$	$1.66 \cdot 10^{-3}$	1.77	$1.40 \cdot 10^{-10}$
0.044	$6.25 \cdot 10^{-3}$	$4.56 \cdot 10^{-4}$	1.86	$3.83 \cdot 10^{-10}$
0.022	$3.13 \cdot 10^{-3}$	$1.20 \cdot 10^{-4}$	1.92	$1.09 \cdot 10^{-9}$
	Expected rate		2	

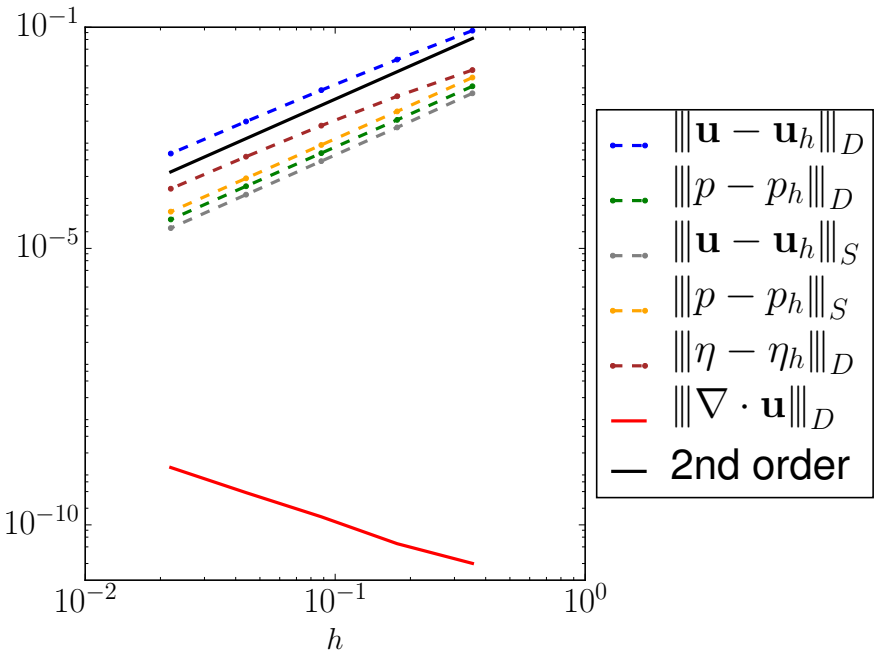


Figure 6.1: Convergence rates for Experiment 6.1. For vector arguments, the norm $|||\cdot|||_D$ corresponds to the discrete $L^2(0,T;\mathbf{H}_{\text{div}}(\Omega_1))$ norm. For scalar inputs the norm $|||\cdot|||_D$ is the discrete $L^2(0,T;L^2(\Omega_1))$ norm. Similarly, the norm $|||\cdot|||_S$ is the $L^2(0,T;H^1(\Omega_2))$ for vector arguments and the $L^2(0,T;L^2(\Omega_2))$ for scalar inputs.

partitioned into hexahedra. To compute the convergence rates we use

$$
\begin{aligned}
\eta^s &= 0.8 - \cos(y)\exp(x)\sin(z), \\
\mathbf{u}_1 &= \begin{pmatrix} -x(\sin(y)\exp(z) + 2(y-1)) \\ y^2 - 2yz - \cos(y)\exp(z) + \exp(x) \\ -\exp(y)x + (z-1)^2 \end{pmatrix}, \\
p_1 &= (y-1)^2 + \exp(y)\cos(x) - \exp(z)\sin(y), \\
\mathbf{u}_2 &= \begin{pmatrix} \exp(2z)\cos(y)\sin(x) - 0.5\exp(y)z - 0.25\exp(y) \\ -0.5\exp(y)xz - 0.25\exp(y)x + \cos(x)\exp(y+2z) \\ \exp(y)xz \end{pmatrix}, \\
p_2 &= \exp(y)z\cos(x) - \exp(y)x + (y-1)^2 - \exp(1)\sin(y).
\end{aligned}
$$

Analogous to Experiment 6.1, for $\|\cdot\|_U$ a norm on the generic space U, and h_1, h_2 two different mesh parameters, the numerical convergence rate is given by the formula

$$\log\left(\frac{\|f - f_{h_1}\|_U}{\|f - f_{h_2}\|_U}\right) / \log\left(\frac{h_1}{h_2}\right).$$

The results are summarized in Table 6.2. Observe that the numerical convergence rates are in agreement with the theory.

In the next chapter, we explore an optimization application related to the Stokes-Darcy problem. For the remainder of this document we refer to the numerical approximations obtained from Scheme 6.1 as solutions to the Stokes-Darcy problem with deposition.

Table 6.2: Numerical results for Experiment 6.2.

Stokes domain

h	$\|\mathbf{u}-\mathbf{u}_h\|_{H^1(\Omega_2)}$	Rate	$\|\mathbf{u}-\mathbf{u}_h\|_{L^2(\Omega_2)}$	Rate	$\|p-p_h\|_{L^2(\Omega_2)}$	Rate
0.433	$8.00 \cdot 10^{-1}$	–	$3.09 \cdot 10^{-2}$	–	$3.94 \cdot 10^{-2}$	–
0.217	$2.02 \cdot 10^{-1}$	1.99	$3.89 \cdot 10^{-3}$	2.99	$4.07 \cdot 10^{-3}$	3.27
0.108	$5.06 \cdot 10^{-2}$	2.00	$4.88 \cdot 10^{-4}$	3.00	$7.97 \cdot 10^{-4}$	2.35
	Expected rate	2				2

Darcy domain

h	$\|\mathbf{u}-\mathbf{u}_h\|_{\mathbf{H}_{\mathrm{div}}(\Omega_1)}$	Rate	$\|p-p_h\|_{L^2(\Omega_1)}$	Rate	$\|\nabla \cdot \mathbf{u}\|_{L^2(\Omega_1)}$
0.433	$1.65 \cdot 10^{-2}$	–	$1.51 \cdot 10^{-2}$	–	$1.63 \cdot 10^{-10}$
0.217	$3.94 \cdot 10^{-3}$	2.07	$2.82 \cdot 10^{-3}$	2.42	$3.67 \cdot 10^{-10}$
0.108	$9.68 \cdot 10^{-4}$	2.03	$6.85 \cdot 10^{-4}$	2.04	$1.39 \cdot 10^{-9}$
	Expected rate	2		2	

Chapter 7

An optimization application for the Stokes-Darcy problem

7.1 Introduction

When designing a filter one typically is interested in maximizing the lifetime of the filtration unit, maximizing the amount of material that is captured throughout its operation, and minimizing the cost incurred in its construction. To approach this problem, one option is to develop a mathematical model for the filtration process and optimize over design parameters such as the geometry of the filter and the physical properties of the filtration material. One example in this direction is [42], where filtration in a polymer fiber melt-spinning process is considered. Therein, the authors relate the filter lifetime to the maximum allowable pressure drop across the filter and optimize two competing criteria: (i) the filter lifetime, and (ii) the amount of debris that escapes the filter. The optimization involves sorting the outputs of an extrusion filter simulator by means of a genetic algorithm. Another example that considers the same filtration setting is [43], where the authors replace the multicriteria objective

with a single functional which incorporates the additional constraint that the amount of escaped debris can not exceed a threshold value. In this chapter we follow some of the ideas discussed in [42].

Building upon the model and the numerical approximation for the filtration problem, we investigate an optimization application of the Stokes-Darcy problem with deposition (SDD). We consider five numerical experiments. Experiment 7.3 motivates the optimization problem that is defined in Section 7.2. Experiment 7.4 optimizes over a family of linear porosity profiles, while Experiment 7.5 optimizes over a family of parabolic porosity profiles. The remaining two experiments are the 3D analogs of Experiment 7.4, and Experiment 7.5.

Experiment 7.3 Let $\Omega_1 = (0,1) \times (0,1)$ denote the Darcy domain, $\Omega_2 = (0,1) \times (1,2)$ denote the Stokes domain, and $\Gamma = (0,1) \times \{1\}$ their common interface. Furthermore, define the inflow boundary $\Gamma_{\text{in}} = (0,1) \times \{2\}$, the outflow boundary $\Gamma_{\text{out}} = (0,1) \times \{0\}$, the Darcy boundary $\Gamma_1 = \partial\Omega_1 \setminus (\Gamma \cup \Gamma_{\text{out}})$, and the Stokes boundary $\Gamma_2 = \partial\Omega_2 \setminus (\Gamma \cup \Gamma_{\text{in}})$. A graphical depiction of the computational domain is given in Figure 7.1. We partition Ω_i into 256 square cells and use the $RT_1 \times \text{disc}Q_1$ pair in Ω_1, and the Taylor-Hood element pair $(Q_2)^2 \times Q_1$ in Ω_2. The porosity is approximated using $\text{disc}Q_0$ elements, and the smooth porosity is computed with the differential filter (6.33) using Q_1 elements and $\delta = 10^{-8}$. The interfacial pressure is approximated with Q_2 elements and the function $\rho(\,\cdot\,,\,\cdot\,)$ (see (6.19)) is computed using piecewise constant elements. The physical parameters are given by:

$$\beta(\eta) = C_\beta \frac{(1-\eta)^2}{\eta^3}, \quad \Psi(\eta) = 1, \quad \mu = 1/2, \quad g(|\mathbf{u}|) = |\mathbf{u}|\,. \tag{7.1}$$

In (7.1), the expression for $\beta(\,\cdot\,)$ arises from the well-known Kozeny-Carman equation

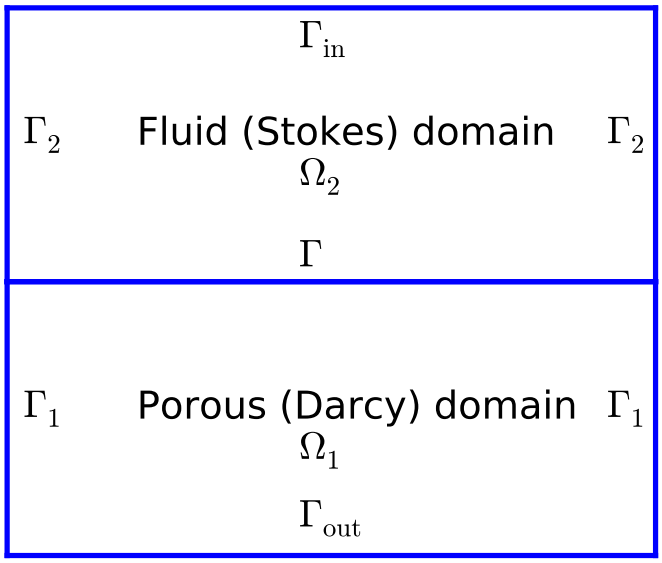


Figure 7.1: Computational domain for Experiment 7.3.

[75], applicable for laminar flow through a bed packed with spherical particles. The constant C_β typically depends upon the viscosity of the fluid and the mean diameter of the particles in the porous media. In Section 3.2 we discussed that a typical profile for $g(\,\cdot\,)$ is a unimodal function. However, since we are interested in slow to moderate flows where the values of $g(|\mathbf{u}|)$ lie to the left of the maximum of $g(\,\cdot\,)$, the function $g(\,\cdot\,)$ is suitably approximated by $g(|\mathbf{u}|) = |\mathbf{u}|$. In this experiment we assume that $C_\beta = 1$ and the viscosity in the fluid (Stokes) domain is $\mu = 0.5$. For the initial porosity profile, we consider a vertical channel in the porous medium given by

$$\text{channel}(y) = 0.5 + 0.35 \sin(2\pi\, y)\,,$$

and assume the porosity attains a maximum $\eta_{\max} = 0.9$ along the channel and decreases at an exponential rate $\tau = 7.5$, proportional to the horizontal distance to the channel. We further assume the minimum porosity is given by $\eta_{\min} = 0.5$. The

expression for η_0 satisfying the aforementioned characteristics is

$$\eta_0(x, y) = \eta_{\min} + (\eta_{\max} - \eta_{\min}) \exp\left(-\tau \left|\text{channel}(y) - x\right|\right).$$

We set zero flux boundary conditions on Γ_1 and homogeneous Dirichlet boundary conditions on Γ_2. On the inflow boundary Γ_{in} we impose the parabolic profile $\mathbf{u}_{\text{in}} = (0,\, 4x(x-1))^T$, and enforce weakly the condition $p = 0$ on the outflow boundary Γ_{out}. With $T = 20$ and $\Delta t = 0.045$, we study the evolution of the porosity $\eta(\mathbf{x}, t)$, the maximum pressure at the inflow, $p_{\text{in}}(t) = \|p(t)\|_{L^\infty(\Gamma_{\text{in}})}$, and the amount of particulate that was retained by the filter for an operation time of t units, $\omega(t)$. The function $\omega(\,\cdot\,) : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}$ is given by

$$\omega(t) = \int_{\Omega_1} (\eta_0(\mathbf{x}, t) - \eta(\mathbf{x}))\rho_{\text{particulate}}\, d\Omega_1,$$

where we have chosen the density of the particulate in the porous (Darcy) domain to be $\rho_{\text{particulate}} = 1$. Physically, we expect to observe two different phenomena. First, as time progresses, the regions of high porosity, e.g., the channel, should fill with the particulate that is dispersed in the fluid, thereby decreasing the overall porosity of the medium. Second, in view that the volumetric flow rate and density are constant, the pressure required to push the fluid through the filter should increase as the porosity decreases. The first phenomenon is simply the process of filtration. In particular, it is often important to quantify the mass of captured particulate in order to compute the filtration efficiency. The second phenomenon is relevant in industry for the following reasons. High pressure demands on the pump typically imply high requirements of energy and consequently an increase in the cost of operation. Furthermore, under high pressures, the filtration medium, e.g., a porous membrane, can be subjected to

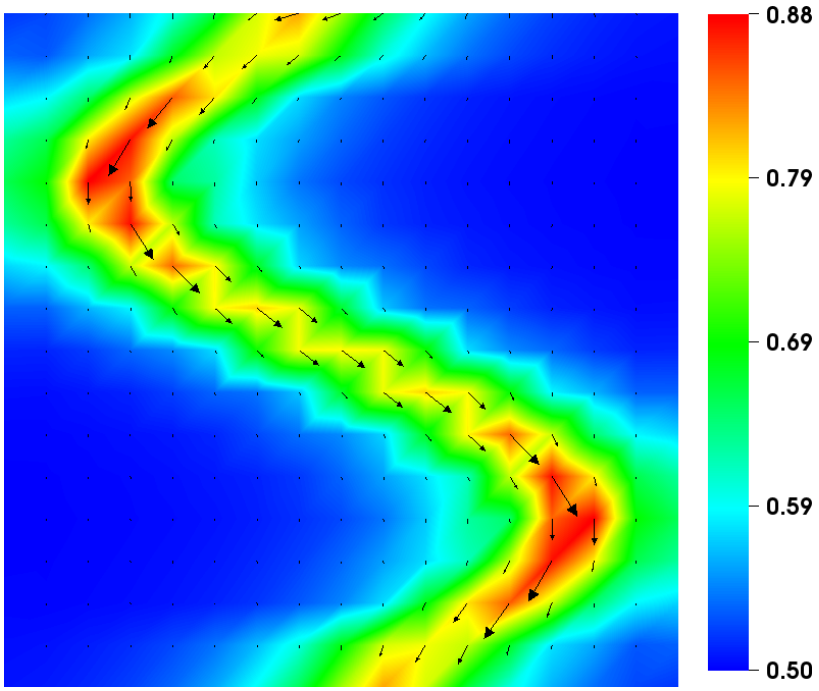


Figure 7.2: Initial profile of the smoothed porosity in Experiment 7.3. The arrows indicate the direction of the flow in the porous domain Ω_1.

$p_{\text{in}}(0)$	$p_{\text{in}}(20)$	$\omega(20)$	$\max_{(\mathbf{x},t)\in\Omega_1\times[0,20]} \lvert\mathbf{u}_1(\mathbf{x},t)\rvert$
4.98	8.8×10^4	0.564	3.924

Table 7.1: Results for Experiment 7.3.

considerable levels of stress. These mechanical forces can lead to deformation, tearing, or irreversible damage of the equipment.

The results are summarized in Table 7.1, Figure 7.2, Figure 7.3, and Figure 7.4. Note that the pressure at the inflow Γ_{in} increased by four orders of magnitude with respect to the original value in a time period of $T = 20$ units. Furthermore, Figure 7.4 depicts how the porosity in Ω_1 diminished due to the deposition of material. In particular, 0.564 units of mass were captured during the filtration process. In summary, Experiment 7.3 shows that for a given porosity profile η_0 and time horizon T, we can compute the maximum pressure at the inflow of the filtration unit Ω and

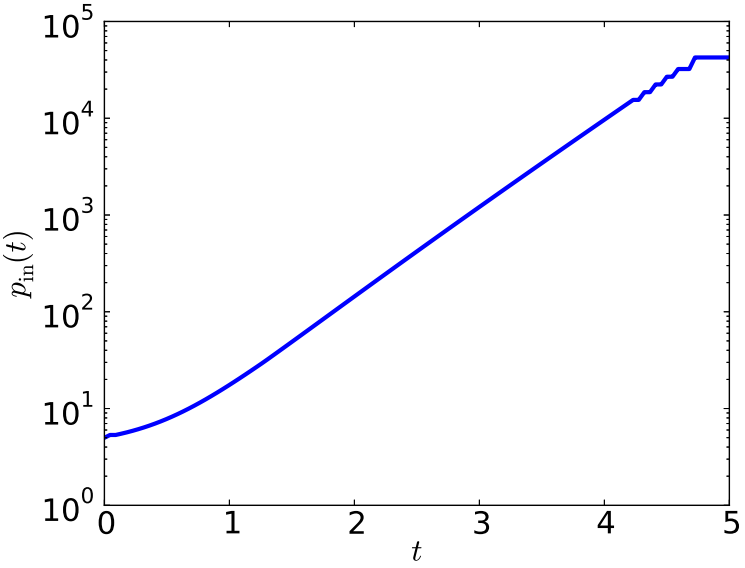


Figure 7.3: Temporal evolution of the pressure along the inflow boundary in Experiment 7.3.

the total amount of material captured by the filter. In the next section we investigate the effect of varying the initial porosity profile η_0 on the values of p_{in} and $\omega(T)$.

7.2 Optimization problem

Having in mind the setting discussed in Section 7.1, we introduce the functions we aim to optimize. One relates the lifetime of the filter to the pressure at the inflow, and the other describes the amount of material captured during the lifetime of the filter.

Definition 7.1. *Let η_0 be a given initial porosity profile for the SDD problem. Then, $T(\eta_0)$ is the first time when the pressure at the inflow attains the maximum allowed value $p_{\max} = 2000$, i.e., $T(\eta_0) = \min\{t \in (0, \infty) \mid p_{\text{in}}(t) \geq p_{\max}\}$. Furthermore, let $\omega(\eta_0)$ represent the mass of captured material during the lifetime of the filter, i.e., $\omega(\eta_0) = \omega(T(\eta_0))$.*

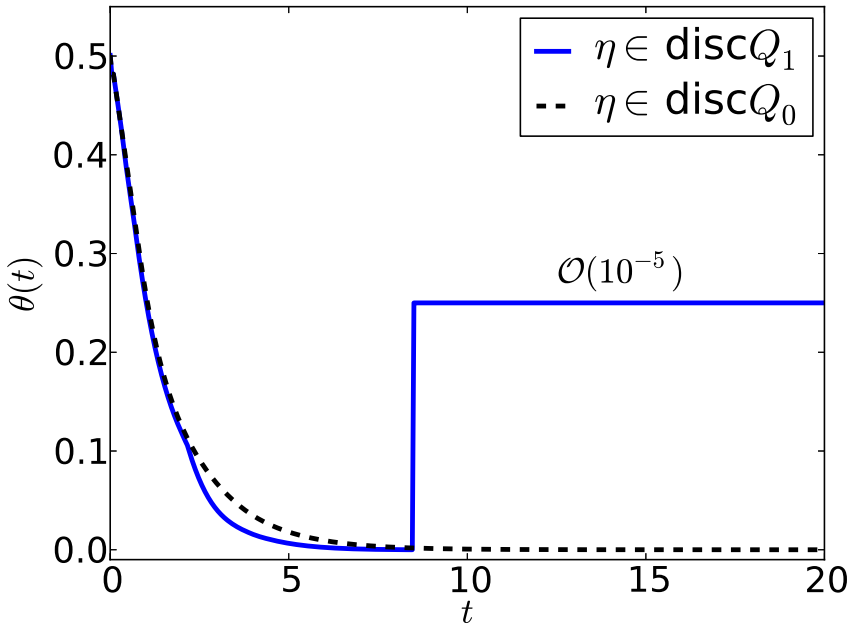


Figure 7.4: Comparison of the temporal evolution of the minimum of the porosity in Experiment 7.3 using discQ_1 and discQ_0 elements. The function θ is given by $\theta(t) = \max\{z(t), -0.25\ \text{sign}(z(t))\}$, where $z(t) = \min_{\mathbf{x}\in\Omega_1}\{\eta_h(\mathbf{x}, t)\}$. Note that, in agreement with Lemma 6.2, the porosity remains nonnegative for $\eta_h \in R_{h,0}$. In the case where $\eta_h \in \text{disc}Q_1$, the porosity becomes negative at $t = 8.505$. The negative values are in the order of 10^{-5}.

Now we state the optimization problem and define the concept of optimality.

Problem 7.1. *Define* $\mathfrak{X} = \{\eta_0 \in L^\infty(\Omega_1) \mid 0 < \eta_0(\mathbf{x}) < 1 \text{ for a.e. } \mathbf{x} \in \Omega_1\}$. *Let* $\mathcal{X} \subseteq \mathfrak{X}$ *, denote the space of all feasible porosity profiles. The optimization problem we consider is:* $\max_{\eta_0 \in \mathcal{X}} \mathbf{f}(\eta_0)$, *where* $\mathbf{f} = (f_1, f_2) : \mathcal{X} \to \mathbb{R}^2$, *and* $f_1(\eta) := T(\eta)$, $f_2(\eta) := \omega(\eta)$.

Definition 7.2 (Pareto optimality)**.** *We say that* $\eta^* \in \mathcal{X}$ *is efficient or Pareto optimal for Problem 7.1 if there is no* $\eta' \in \mathcal{X}$ *such that* $f_i(\eta') \geq f_i(\eta^*)$ *for* $i = 1, 2$, *and* $f_i(\eta') > f_i(\eta^*)$ *for some* i. *In the case where such an* η' *exists, we say that* η^* *is dominated by* η'.

Note that Definition 7.2 allows the possibility of multiple optima. Also observe that the dimension of the space $\mathfrak{X}$ is infinite. Thus, in order to simplify the problem, we make $\mathcal{X}$ finite dimensional by restricting the possible initial porosity profiles to a one parameter family. This idea is explored in the following numerical experiments. We consider linear and parabolic profiles. We start exploring the 2D setting and subsequently follow with the 3D case.

2D setting

For the numerical experiments in this section we use the same spatial discretization as for Experiment 7.3 and set $\Delta t = 0.001$.

Experiment 7.4 Let $\eta_{\max} = 0.9$ and $\eta_{\min} = 0.1$. Consider the set of porosity profiles given by

$$\mathcal{X}_1^{2\mathrm{D}} = \{\eta_\tau \in \mathfrak{X} \,|\, \tau \in [0,1],\, \eta_\tau(x,y) = (b(\tau) - a(\tau))\, y + a(\tau)\}\,,$$
$$a(\tau) = (\eta_{\max} - \eta_{\min})\,\tau + \eta_{\min}, \quad b(\tau) = (\eta_{\min} - \eta_{\max})\,\tau + \eta_{\max}. \tag{7.2}$$

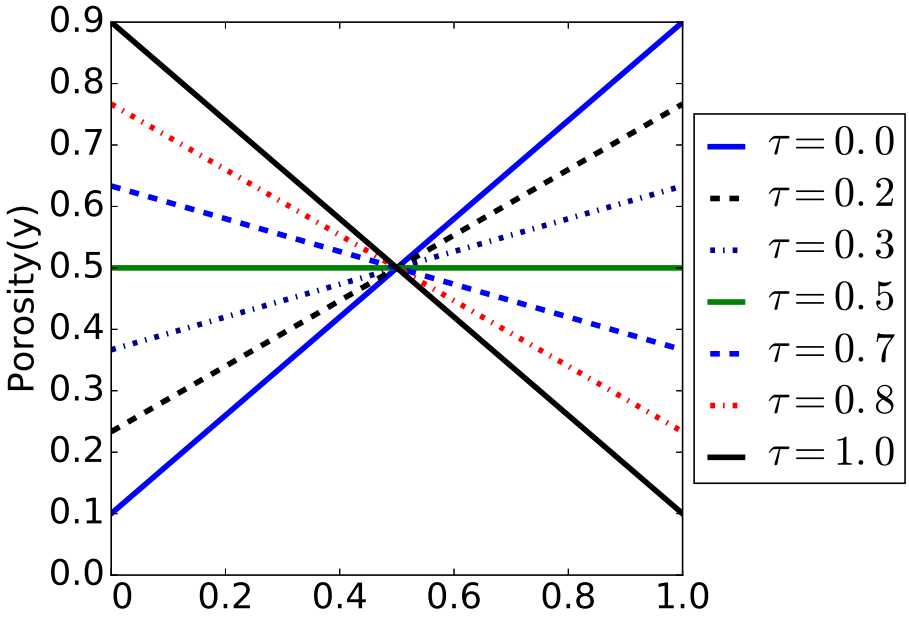


Figure 7.5: Porosity profiles in Experiment 7.4 (2D).

τ	T	$\omega(T)$	Type
0.00	2.062	0.37567	Affine
0.24	2.840	0.42564	Affine
0.50	3.026	0.43389	Constant
0.74	2.863	0.42599	Affine
1.00	2.063	0.37363	Affine

Table 7.2: Results for Experiment 7.4.

Observe that an arbitrary element of $\mathcal{X}_1^{2D}$ is a linear porosity profile in the y-direction. For example, for $\tau = 0$, we have a uniform gradient where the maximum porosity $\eta_{\max}$ occurs at the top $\{y = 1\}$ and the minimum porosity $\eta_{\min}$ occurs at the bottom $\{y = 0\}$. The opposite situation arises for $\tau = 1$. Finally, $\tau = 1/2$ results in a constant porosity profile $\eta \equiv (\eta_{\max} + \eta_{\min})/2$. These observations are depicted in Figure 7.5.

Remark 7.1. *The initial volume of void space in* Ω_1 *is* $(\eta_{\max} + \eta_{\min})/2 = 1/2$ *for all* $\eta_\tau \in \mathcal{X}_1^{2D}$, $0 \leq \tau \leq 1$.

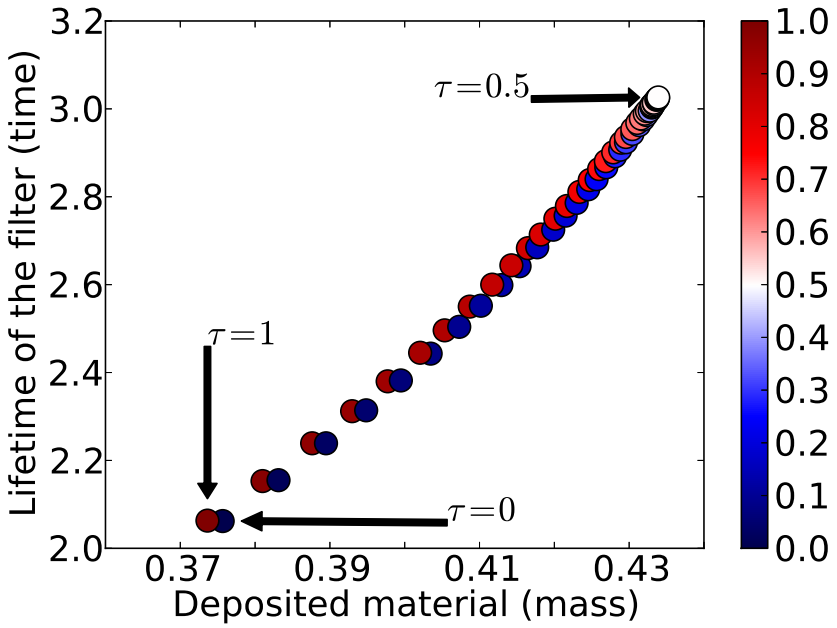


Figure 7.6: Results for Experiment 7.4 in 2D. The color bar indicates the magnitude of τ.

To solve Problem 7.1 we consider a uniform partition of the interval $[0, 1]$ consisting of 63 points τ_k and solve the SDD using $\eta_{\tau_k} \in \mathcal{X}_1^{2\mathrm{D}}$ as an initial porosity profile. We record the time at which the pressure attains $p_{\max}$ and the amount of captured material during that time period. The results are summarized in Figure 7.6 and Table 7.2. The results indicate that for a filter with a constant gradient along the y-direction, the configuration which yields the longest lifetime and largest retention of material is one close to $\tau = 0.5$, i.e., the porosity is essentially uniform (constant) throughout Ω_1. The filter lifetime and mass of captured material for $\tau = 0.50$ is roughly 47% and 16%, respectively, larger than that corresponding to either $\tau = 0$ or $\tau = 1$. Another important observation is that for both extreme values $\tau = 0$ and $\tau = 1$, the performance of the filter is practically the same. We conclude that the optimal η^* to Problem 7.1 over $\mathcal{X}_1^{2\mathrm{D}}$ is contained in a neighborhood of the constant

function $\eta = 0.5$.

Experiment 7.5 When designing a filter, one has to take into account the distribution of the filtration medium. The material that fills the filter, often called the packing material, is composed of small particles that when compacted together give rise to channels (void space) of variable length and size. The more densely packed the material is, the less void space and, consequently, the smaller the porosity. As the packing material is pushed against the boundary of the container, the walls prevent the constituent particles from attaining the compact configuration that they would normally exhibit in an obstructed region. Hence, under uniform packing conditions, the porosity is expected to be higher close to the boundary and lower at the center. Evidence of this phenomenon can be found in [77]. To facilitate the design process, we investigate the performance of a family of filters that transitions from a parabolic profile with maximum occurring at the boundary to a constant porosity value. This family is given by

$$\begin{aligned}
&\mathcal{X}_2^{2\mathrm{D}} = \{\eta_\tau \in \mathfrak{X} \,|\, \tau \in [0,1],\, \eta_\tau(x,y) = U(\tau)\,(b(\tau) + q(x)\,(a(\tau) - b(\tau)))\}\,,\\
&a(\tau) = (1/2 - \eta_{\min})\tau + \eta_{\min}, \quad b(\tau) = (1/2 - \eta_{\max})\tau + \eta_{\max},\\
&U(\tau) = 6/(\tau + 5), \quad q(x) = 4\,x\,(1-x),
\end{aligned}$$

where $\eta_{\max} = 0.75$ and $\eta_{\min} = 0.25$.

Remark 7.2. *The initial volume of void space in* Ω_1 *is* $(\eta_{\max} + \eta_{\min})/2 = 1/2$ *for all* $\eta_\tau \in \mathcal{X}_2^{2\mathrm{D}}$, $0 \le \tau \le 1$.

The elements in $\mathcal{X}_2^{2\mathrm{D}}$ transition from a parabolic profile ($\tau = 0$) with the maximum porosity at $x = 0, 1$ and minimum porosity at $x = 1/2$, to a constant profile ($\tau = 1$) with $\eta = (\eta_{\max} + \eta_{\min})/2 = 1/2$. A graphical representation of this transition

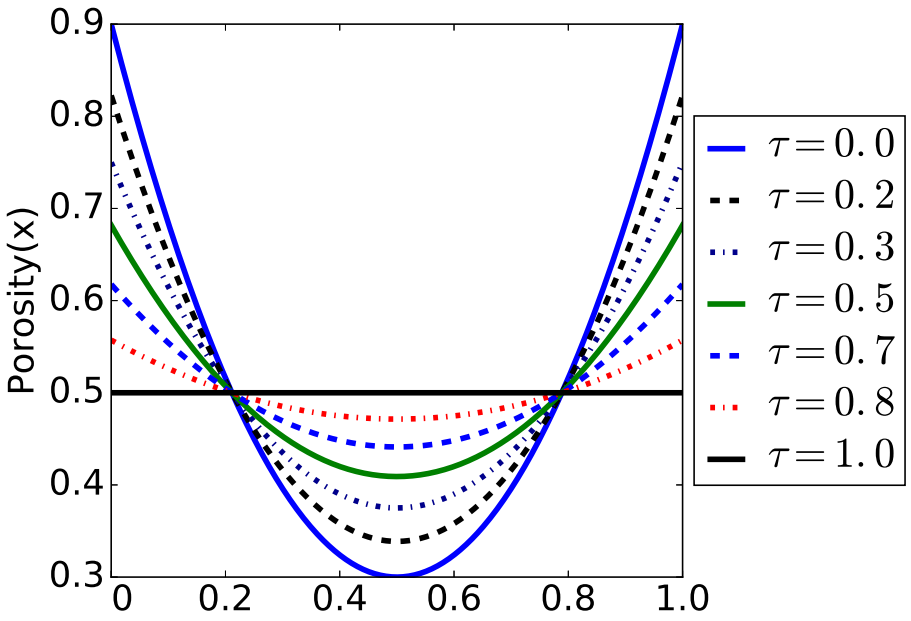


Figure 7.7: Parabolic profiles described by $\mathcal{X}_2^{2D}$ in Experiment 7.5.

τ	T	$\omega(T)$	Type
0.00	2.903	0.43338	Parabolic
0.26	2.958	0.43352	Parabolic
0.50	2.994	0.43371	Parabolic
0.71	3.013	0.43380	Parabolic
1.00	3.026	0.43389	Constant

Table 7.3: Results for Experiment 7.5.

is given in Figure 7.7. Analogous to Experiment 7.4, we partition the interval $[0, 1]$ into 63 points τ_k and solve the SDD for each $\eta_{\tau_k} \in \mathcal{X}_2^{2D}$. The results are depicted in Figure 7.8 and Table 7.3. The results indicate that as we deviate from the constant porosity profile ($\tau = 1$), the filter lifetime diminishes gradually. However, the amount of deposited material remains fairly constant. When compared with the absolute maximum ($\tau = 1$), at the absolute minimum $\tau = 0$ the values for T and $\omega(T)$ are 4.065% and 0.117% lower, respectively. We conclude that the optimal η^* to Problem 7.1 over $\mathcal{X}_2^{2D}$ belongs to a neighborhood of the constant porosity profile $\eta = 1/2$.

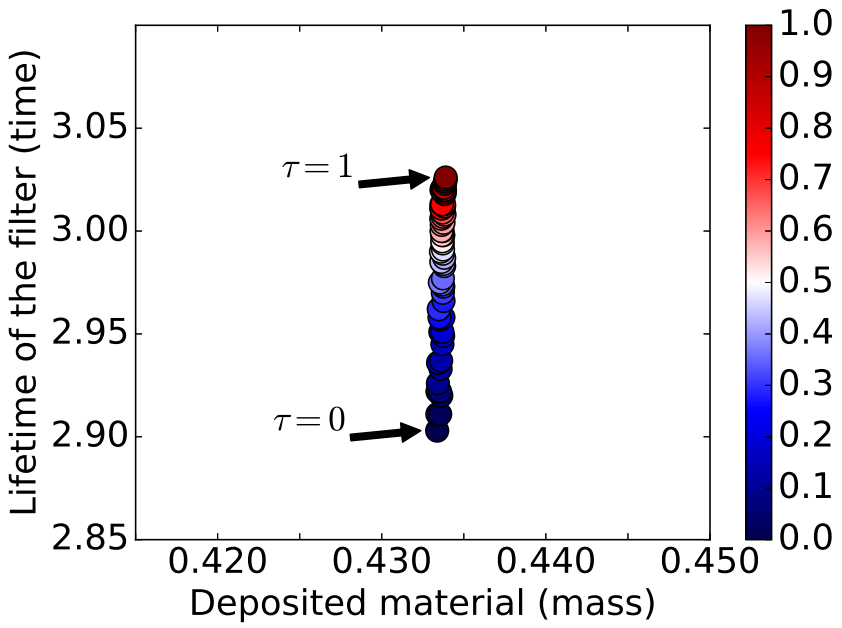


Figure 7.8: Results for Experiment 7.5. The color bar indicates the magnitude of τ.

The next experiment explores initial porosity profiles that are, in a certain sense, opposite to what was considered in Experiment 7.5.

Experiment 7.6 In contrast to Experiment 7.5, we design a family of parabolic profiles that attain minimum porosity at the wall and maximum porosity at the center. The collection of parabolic profiles is given by

$$\mathcal{X}_3^{2\mathrm{D}} = \{\eta_\tau \in \mathfrak{X} \,|\, \tau \in [0,1],\, \eta_\tau(x,y) = U(\tau)\,(a(\tau) + q(x)\,(b(\tau) - a(\tau)))\}\,,$$
$$a(\tau) = (1/2 - \eta_{\min})\tau + \eta_{\min}, \quad b(\tau) = (1/2 - \eta_{\max})\tau + \eta_{\max},$$
$$U(\tau) = 150/(187 - 37\tau), \quad q(x) = 4\,x\,(1-x),$$

where $\eta_{\max} = 0.87$ and $\eta_{\min} = 0.13$, discretize uniformly the interval $[0,1]$ into 63 points τ_k, and solve the SDD for each $\eta_{\tau_k} \in \mathcal{X}_3^{2\mathrm{D}}$.

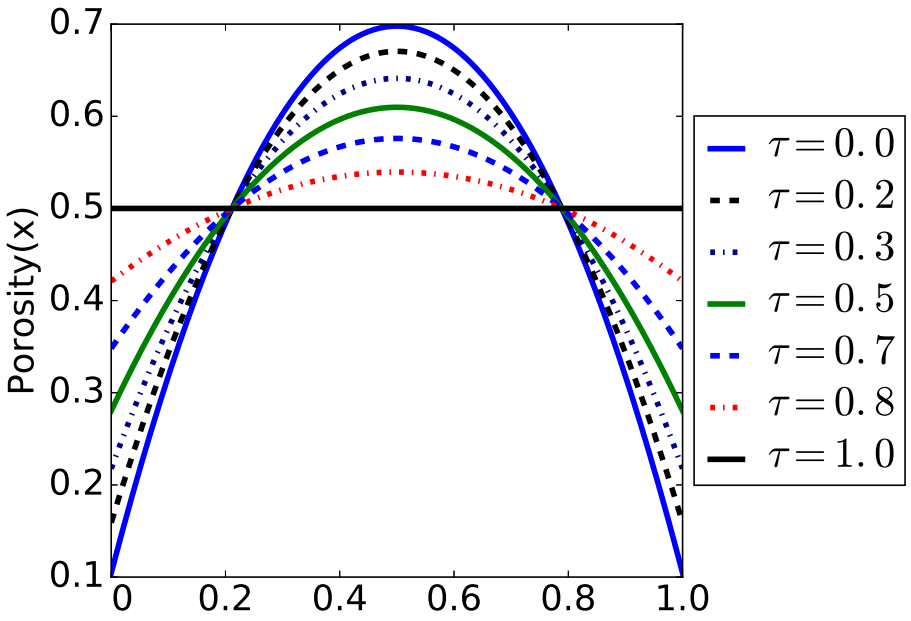


Figure 7.9: Parabolic profiles described by $\mathcal{X}_3^{2D}$ in Experiment 7.6.

τ	T	$\omega(T)$	Type
0.00	2.905	0.43384	Parabolic
0.26	2.960	0.43373	Parabolic
0.50	2.999	0.43385	Parabolic
0.76	3.018	0.43371	Parabolic
1.00	3.026	0.43389	Constant

Table 7.4: Results for Experiment 7.6.

Remark 7.3. *The initial volume of void space in* Ω_1 *is* $(\eta_{\max} + \eta_{\min})/2 = 1/2$ *for all* $\eta_\tau \in \mathcal{X}_3^{2D}$, $0 \leq \tau \leq 1$.

Examples of elements of $\mathcal{X}_3^{2D}$ are shown in Figure 7.9. The results are summarized in Figure 7.10 and Table 7.4. Similar to Experiment 7.5, the constant initial porosity profile $\tau = 1$ outperforms the considered filters for $\tau < 1$. In this case the filter corresponding to $\tau = 1$ exhibited a lifetime 4.165% longer than the filter corresponding to $\tau = 0$. Hence, for the physical parameters under consideration, using an initial parabolic profile that qualitatively matches the inflow velocity profile appears

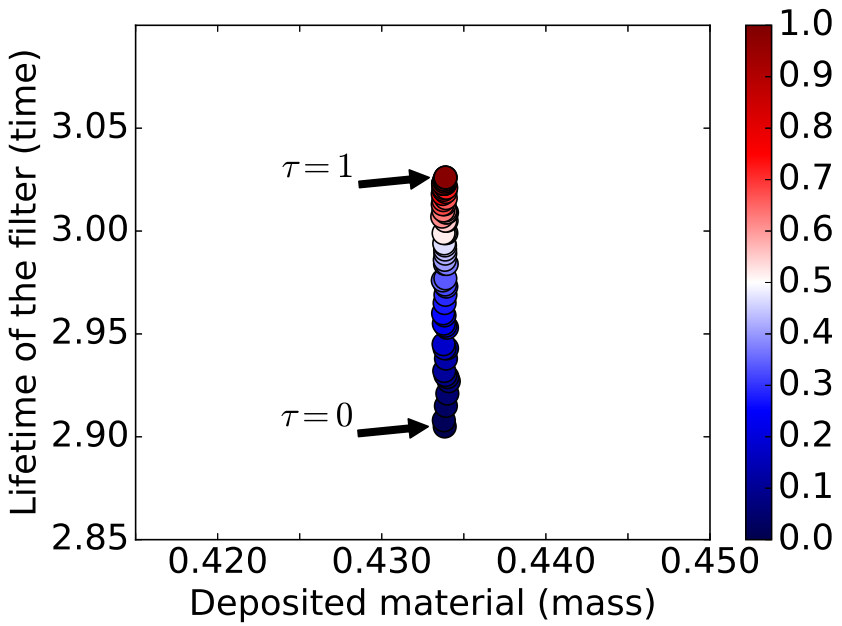


Figure 7.10: Results for Experiment 7.6. The color bar indicates the magnitude of τ.

to confer no advantage over a constant initial porosity profile.

Though a constant initial porosity profile resulted in the best performance for Experiment 7.4, Experiment 7.5, and Experiment 7.6, the question of optimizing Problem 7.1 in 2D over $\mathfrak{X}$ remains open.

Now we present 3D analogs to Experiment 7.4, Experiment 7.5 and Experiment 7.6.

3D setting

Let $\Omega_1 = (0,1) \times (0,1) \times (0,1)$ denote the Darcy domain, $\Omega_2 = (0,1) \times (0,1) \times (1,2)$ denote the Stokes domain, and $\Gamma = (0,1) \times (0,1) \times \{1\}$ their common interface. Moreover, let $\Gamma_{\text{in}} = (0,1) \times (0,1) \times \{2\}$ represent the inflow boundary and $\Gamma_{\text{out}} = (0,1) \times (0,1) \times \{0\}$ the outflow boundary. Define $\Gamma_1 = \partial\Omega_1 \setminus (\Gamma \cup \Gamma_{\text{out}})$ and

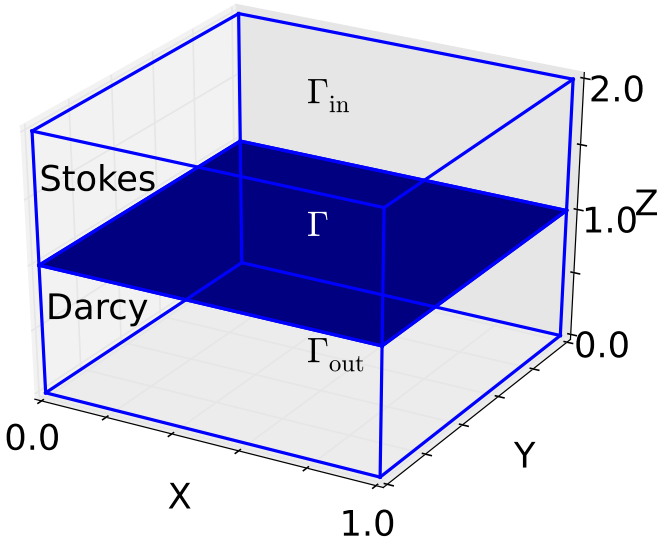


Figure 7.11: Computational domain in the 3D setting. The colored plane denotes the interface Γ.

$\Gamma_2 = \partial\Omega_2 \setminus (\Gamma \cup \Gamma_{\text{in}})$. We partition Ω_i into 64 cubic cells, set $\Delta t = 0.005$, and use the same physical parameters as in Experiment 7.3. Furthermore, no flux boundary conditions are imposed on Γ_1 and homogeneous Dirichlet boundary conditions are imposed on Γ_2. Finally, we set the inflow velocity $\mathbf{u}_{\text{in}} = \mathbf{u}|_{\Gamma_{\text{in}}}$ to the parabolic profile $\mathbf{u}_{\text{in}} = (0,\, 0,\, -16xy(x-1)(y-1))^T$ and enforce weakly $p = 0$ along Γ_{out}. The computational domain is depicted in Figure 7.11. In Experiment 7.7 the interval $[0, 1]$ is uniformly partitioned into 64 points τ_k and for Experiment 7.8 and Experiment 7.9 into 20 points. The SDD is solved for each η_{τ_k} in a given family of porosity profiles.

Experiment 7.7 Consider the family of linear porosity profiles on z

$$\mathcal{X}_1^{3\text{D}} = \{\eta_\tau \in \mathfrak{X} \,|\, \tau \in [0,1],\, \eta_\tau(x,y,z) = (b(\tau) - a(\tau))\, z + a(\tau)\}\,, \tag{7.3}$$

τ	T	$\omega(T)$	Type
0.00	3.975	0.41685	Affine
0.24	4.615	0.43693	Affine
0.50	4.825	0.44207	Constant
0.76	4.610	0.43586	Affine
1.00	3.945	0.41352	Affine

Table 7.5: Results for Experiment 7.7.

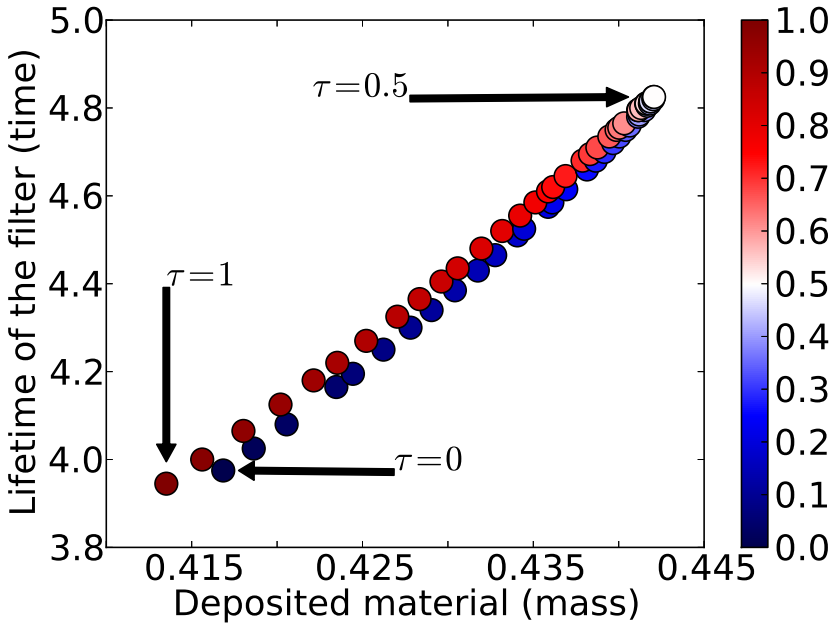


Figure 7.12: Results for Experiment 7.7. The color bar indicates the magnitude of τ.

with $a(\,\cdot\,)$, $b(\,\cdot\,)$, $\eta_{\max}$, and $\eta_{\min}$ as given by (7.2).

Remark 7.4. *The initial volume of void space in* Ω_1 *is* $(\eta_{\max} + \eta_{\min})/2 = 1/2$ *for all* $\eta_\tau \in \mathcal{X}_1^{3\mathrm{D}}$, $0 \leq \tau \leq 1$.

Recall from Figure 7.5 that $\tau = 0$ corresponds to a linearly decreasing porosity profile, $\tau = 1$ corresponds to a linearly increasing porosity profile, and $\tau = 0.5$ to a constant profile. The results are summarized in Table 7.5 and Figure 7.12. We observe that, qualitatively, the filters in 2D from Experiment 7.4, and the filters in 3D from Experiment 7.7 behave similarly. In particular, a near constant porosity profile is

τ	T	$\omega(T)$	Type
0.00	4.570	0.44138	Parabolic
0.25	4.660	0.44132	Parabolic
0.50	4.735	0.44153	Parabolic
0.75	4.775	0.44137	Parabolic
1.00	4.810	0.44168	Constant

Table 7.6: Results for Experiment 7.8.

optimal in both settings. Also note that the initial porosity profile corresponding to $\tau = 0$ yields better results than those generated by $\tau = 1$. That is, under the given conditions, a filter where the initial porosity is maximal at the top and decreases to its minimum at the bottom performs better than one where the initial porosity is minimal at the top and increases towards the bottom.

Experiment 7.8 Consider the family of parabolic profiles given by

$$\begin{aligned}
&\mathcal{X}_2^{3\mathrm{D}} = \{\eta_\tau \in \mathfrak{X} \,|\, \tau \in [0,1],\, \eta_\tau(x,y,z) = U(\tau)\,(b(\tau) + q(x,y)\,(a(\tau) - b(\tau)))\}\,,\\
&a(\tau) = (1/2 - \eta_{\min})\tau + \eta_{\min}, \quad b(\tau) = (1/2 - \eta_{\max})\tau + \eta_{\max},\\
&U(\tau) = 45/(49 - 4\tau), \quad q(x,y) = 16\,x\,y\,(x-1)\,(y-1),\\
&\eta_{\max} = 0.9, \quad \eta_{\min} = 0.1.
\end{aligned} \tag{7.4}$$

Remark 7.5. *The initial volume of void space in* Ω_1 *is* $(\eta_{\max} + \eta_{\min})/2 = 1/2$ *for all* $\eta_\tau \in \mathcal{X}_2^{3\mathrm{D}}$, $0 \le \tau \le 1$.

Note that for $\tau = 0$ we have an initial porosity profile where the porosity is maximal on the walls of the filter and minimal at the center. As τ increases, the initial porosity becomes more uniform until it reaches the constant value $\eta = 1/2$ when $\tau = 1$. A graphical representation for the case $\tau = 0$ is shown in Figure 7.13. The results are summarized in Figure 7.14 and Table 7.6. The results suggest that

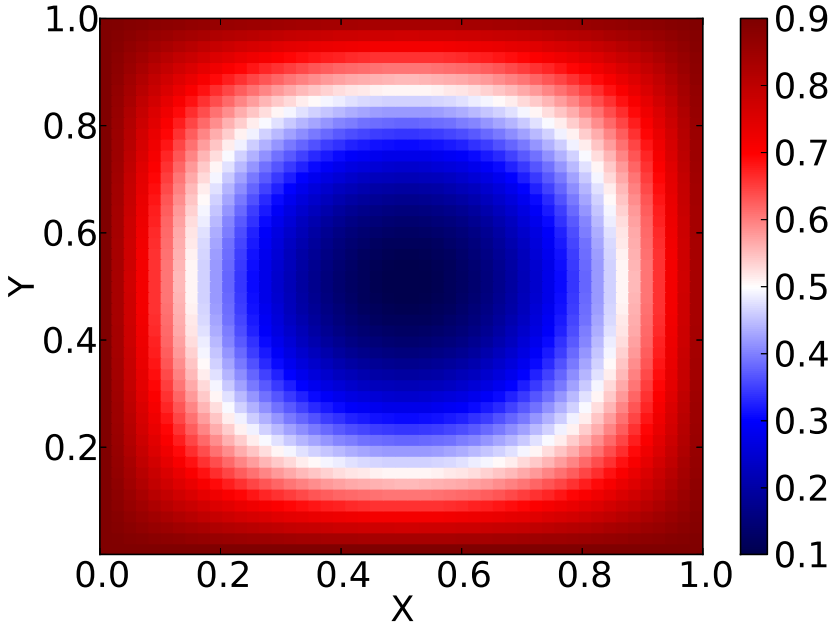


Figure 7.13: Porosity profile in Experiment 7.8 when $\tau = 0$. The colorbar indicates the magnitude of the porosity.

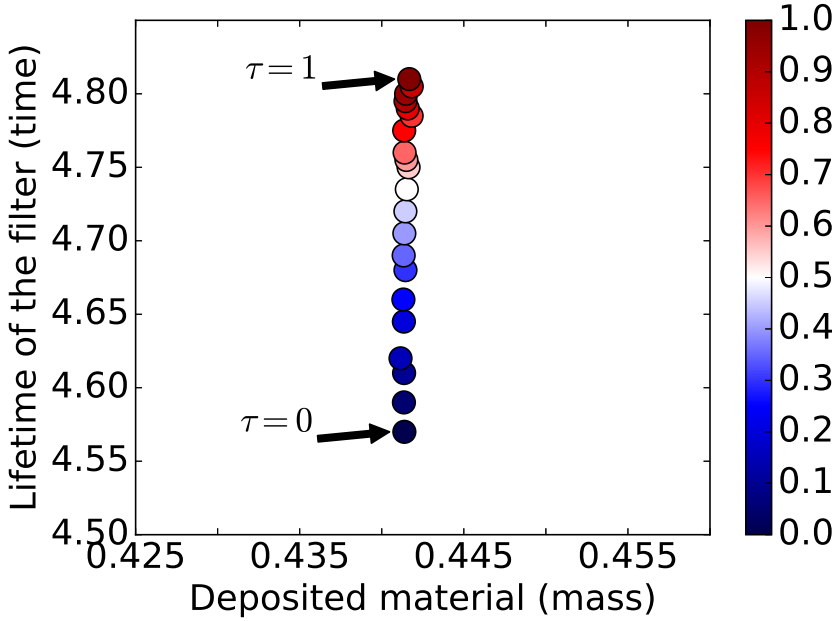


Figure 7.14: Results for Experiment 7.8. The color bar indicates the magnitude of τ.

τ	T	$\omega(T)$	Type
0.00	4.555	0.44062	Parabolic
0.25	4.685	0.44115	Parabolic
0.55	4.780	0.44160	Parabolic
0.75	4.815	0.44190	Parabolic
1.00	4.820	0.44193	Constant

Table 7.7: Results for Experiment 7.9.

a constant initial porosity profile is optimal for this family of filters.

Experiment 7.9 Opposite to what was done in Experiment 7.8, we consider parabolic profiles whose porosity is maximal at the center and minimal at the walls of the filter. Define

$$\mathcal{X}_3^{3\mathrm{D}} = \{\eta_\tau \in \mathfrak{X} \,|\, \tau \in [0,1],\, \eta_\tau(x,y,z) = U(\tau)\,(a(\tau) + q(x,y)\,(b(\tau) - a(\tau)))\}\,,$$
$$a(\tau) = (1/2 - \eta_{\min})\tau + \eta_{\min}, \quad b(\tau) = (1/2 - \eta_{\max})\tau + \eta_{\max},$$
$$U(\tau) = 450/(37\tau + 413), \quad q(x,y) = 16\,x\,y\,(x-1)\,(y-1),$$
$$\eta_{\max} = 0.87, \quad \eta_{\min} = 0.13.$$

Remark 7.6. *The initial volume of void space in* Ω_1 *is* $(\eta_{\max} + \eta_{\min})/2 = 1/2$ *for all* $\eta_\tau \in \mathcal{X}_3^{3\mathrm{D}}$, $0 \le \tau \le 1$.

The case $\tau = 0$ is depicted in Figure 7.15. The case $\tau = 1$ is the constant porosity profile $\eta = 1/2$. The results are summarized in Figure 7.16 and Table 7.7. We note that these results are consistent with the 2D setting explored in Experiment 7.6. In summary, in 2D and 3D a constant porosity profile appears to yield a higher mass retention and filter lifetime than using a parabolic profile with either maximal porosity at the center and minimimal at the walls, or with minimal porosity at the center and maximal at the boundary. Just like in the 2D case, the question of optimizing over the entire space $\mathfrak{X}$ remains open.

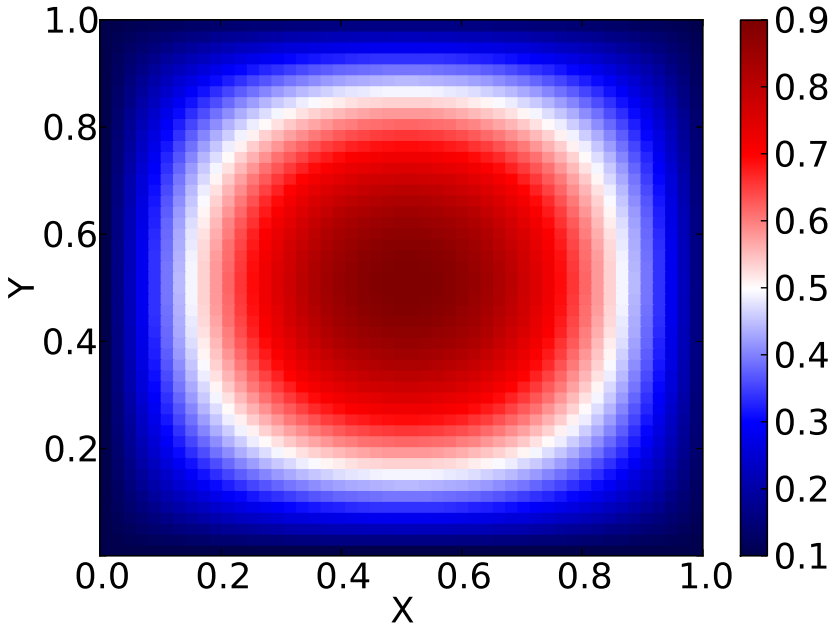


Figure 7.15: Parabolic profile for the porosity corresponding to $\tau = 0$ in Experiment 7.9. The colorbar denotes the magnitude of the porosity.

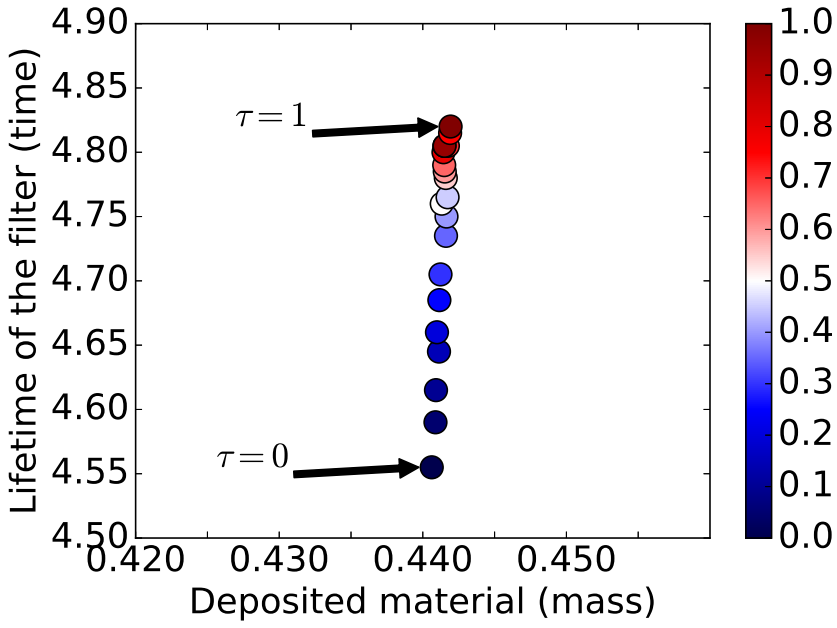


Figure 7.16: Results for Experiment 7.9. The color bar indicates the magnitude of τ.

Chapter 8

Conclusion

Summary

In the first part of this thesis a filtration model was presented. The functions under consideration were the velocity, pressure, and porosity. The model combined a nonlinear Darcy problem to describe the fluid flow and a time-dependent deposition equation to characterize the evolution of the porosity. Existence and uniqueness of a solution to the governing equations was established. Subsequently, a numerical scheme capable of approximating the solution of the continuous problem was investigated. It was proven that the scheme is uniquely solvable and converges to the exact solution as the discretization parameters Δt and h tend to zero. Numerical experiments confirmed the analytical results. Furthermore, the performance of different filters was compared by varying the initial porosity profile.

In the second part of this thesis, we considered a generalization of the initial filtration model. The new system of equations coupled the previously analyzed augmented nonlinear Darcy problem to the Stokes equations. Well-posedness of the coupled sys-

tem with appropriate boundary and interfacial conditions was established. Moreover, consistent with the physics, the porosity was shown to be nonnegative and bounded above by the initial porosity. Thereafter, we introduced a numerical scheme for the coupled problem capable of preserving the nonnegativity and boundedness of the porosity. Numerical experiments followed, showing that the numerical approximations for the porosity are indeed bounded between zero and one. Furthermore, the computations showed that the numerical convergence rates in space and time for the coupled problem are in agreement with the conjectured estimates. Lastly, we performed computational experiments in 2D and 3D with a focus on filter optimization. Specifically, we defined a multicriteria objective function and optimized over uniparametric families of linear and parabolic initial porosity profiles.

Future research

An interesting extension of this work is to remove the assumption of an affine dependence of the velocity on the pressure gradient in the porous domain and allow for nonlinear effects. Concretely, we aim to replace the Darcy law with a Darcy-Forchheimer model of the form

$$\beta(\eta)\,\mathbf{u} + \gamma\,\mathbf{u}\,|\mathbf{u}| + \nabla p = \mathbf{f},$$

where γ is a nonnegative constant. In this setting, after deriving the weak form, the Darcy velocity is to be found in the space $(L^3(\Omega_1))^d$, while the pressure is to be found in the intersection of the fractional Sobolev space $W^{1,3/2}(\Omega_1)$ and the Hilbert space $L^2_0(\Omega_1)$. Incorporating the nonlinearity allows for more realistic simulations by taking into account the inertial effects. Analysis of the well-posedness of the Darcy-

Forchheimer model using techniques of monotonicity, coercivity and hemi-continuity of a nonlinear operator can be found in [47]. After completing the analytical part of this problem, a thorough numerical comparison between the linear and nonlinear regimes would follow in order to better understand the inertial effects.

We observe that the model introduced and analyzed in Chapter 5 assumes a quasi-static flow in the Stokes domain. A more realistic model would be to consider the time-dependent Stokes equations in this subdomain.

A further extension is to also include inertial effect in the fluid domain and replace the Stokes equations with the more general time-dependent Navier-Stokes. Also desirable, would be to incorporate heat exchange effects and consider a nonisothermal flow. An axisymmetric model in 2D that couples the Darcy-Forchheimer and the compressible Navier-Stokes equations to a vertical wellbore model is discussed in [5]. Therein, the authors not only consider a nonlinear dependency of the flow on the velocity, but also allow for variations in density and temperature, which are modeled through the use of the Peng-Robinson equation of state. The analytical techniques discussed in this work could be modified to incorporate variations in porosity.

Appendices

Appendix A Error analysis for Darcy with deposition

A.1 Approximation Scheme

In this section we use the discrete operators

$$d_t f^n = \frac{f^n - f^{n-1}}{\Delta t}, \quad \overline{f}^n = \frac{f^n + f^{n-1}}{2}, \quad \widetilde{f}^n = f^{n-1} + \frac{1}{2} f^{n-2} - \frac{1}{2} f^{n-3},$$
$$d_{2t} f^n = \frac{f^n - f^{n-2}}{2\Delta t}, \quad \breve{f}^n = \frac{f^n + f^{n-2}}{2}, \tag{1}$$

and the next two lemmas

Lemma A.1. *For $f^k \in L^2((0,T); L^2(\Omega))$*

$$\begin{aligned}
\|\overline{f}^n\|^2 &\leq \frac{1}{2}(\|f^n\|^2 + \|f^{n-1}\|^2), \\
\|\widetilde{f}^n\|^2 &\leq 2\|f^{n-1}\|^2 + \|f^{n-2}\|^2 + \|f^{n-3}\|^2, \\
\widetilde{f}^n &= 2\overline{f}^{n-1} - \overline{f}^{n-2}, \\
\|\widetilde{f}^n\|^2 &\leq 8\|\overline{f}^{n-1}\|^2 + 2\|\overline{f}^{n-2}\|^2.
\end{aligned}$$

Proof. Follows directly from the definition of the discrete operators and the triangle inequality. □

Lemma A.2. *Let $f(\,\cdot\,, t) \in L^2(\Omega))$ with $f(\mathbf{x}, \cdot)$ sufficiently regular so that the follow-*

ing estimates are well-defined. Then, the following inequalities hold.

$$\begin{aligned}
\|\overline{f}^n - f^{n-1/2}\|^2 &\leq \frac{1}{48}(\Delta t)^3 \int_{t_{n-1}}^{t_n} \|f_{tt}\|^2 \, dt, \\
\|\breve{f}^2 - f^1\|^2 &\leq \frac{1}{3}(\Delta t)^4 \sup_{t_0 \leq t \leq t_2} \|f_{tt}(\cdot, t)\|^2, \\
\|d_t f^n - f_t^{n-1/2}\|^2 &\leq \frac{1}{1280}(\Delta t)^3 \int_{t_{n-1}}^{t_n} \|f_{ttt}\|^2 \, dt, \\
\|d_{2t} f^2 - f_t^1\|^2 &\leq \frac{1}{80}(\Delta t)^4 \sup_{t_0 \leq t \leq t_2} \|f_{ttt}(\cdot, t)\|^2, \\
\|\widetilde{f}^n - f^{n-1/2}\|^2 &\leq \frac{39}{8}(\Delta t)^3 \int_{t_{n-3}}^{t_{n-1/2}} \|f_{tt}\|^2 \, dt, \\
\|d_t f^n\|^2 &\leq \frac{1}{\Delta t} \int_{t_{n-1}}^{t_n} \|f_t\|^2 \, dt.
\end{aligned}$$

Proof. See [34]. □

Recall that the approximation scheme we investigate in Section 4.1 is: *Given* $\eta_0 \in R_h$*, for* $n = 3, \ldots, N$*, determine* $(\mathbf{u}_h^n, p_h^n, \eta_h^n) \in X_h \times Q_h \times R_h$ *satisfying*

$$\begin{aligned}
(\beta(\eta_h^{n,s})\mathbf{u}_h^n + \beta(\eta_h^{n,s})\mathbf{b}^n, \mathbf{v}) - (p_h^n, \nabla \cdot \mathbf{v}) &= (\mathbf{f}^n, \mathbf{v}) \ \forall \mathbf{v} \in X_h, \\
(q, \nabla \cdot \mathbf{u}_h^n) &= 0 \ \forall q \in Q_h, \\
(d_t \eta_h^n, r) + (g(|\widetilde{\mathbf{u}}_h^n + \mathbf{b}^{n-1/2}|)\overline{\eta}_h^n, r) &= 0 \ \forall r \in R_h.
\end{aligned}$$

Also recall that we defined the discrete quantities

$$\begin{aligned}
\boldsymbol{\Lambda}^n &= \mathbf{u}^n - \mathcal{U}^n, & \mathbf{E}^n &= \mathcal{U}^n - \mathbf{u}_h^n, \\
\psi^n &= \eta^n - \tau^n, & F^n &= \tau^n - \eta_h^n, \\
\varepsilon_{\mathbf{u}} &= \mathbf{u}^n - \mathbf{u}_h^n, & \varepsilon_\eta &= \eta^n - \eta_h^n.
\end{aligned}$$

in (4.14).

First note that (4.1)-(4.3) satisfies

$$(\beta(\eta^{n,s})(\mathbf{u}^n + \mathbf{b}^n), \mathbf{v}) = (\mathbf{f}^n, \mathbf{v}) \, \forall \mathbf{v} \in Z_h, \tag{2}$$

$$\begin{aligned}(d_t\eta^n, r) + (g(|\widetilde{\mathbf{u}}^n + \mathbf{b}^{n-1/2}|)\overline{\eta}^n, r) &= (d_t\eta^n - \frac{\partial \eta^{n-1/2}}{\partial t}, r) \\ &+ \ (g(|\widetilde{\mathbf{u}}^n + \mathbf{b}^{n-1/2}|)\overline{\eta}^n - g(|\mathbf{u}^{n-1/2} + \mathbf{b}^{n-1/2}|)\eta^{n-1/2}, r) \, \forall r \in R_h.\end{aligned} \tag{3}$$

Subtracting (4.8) and (4.10) from (2) and (3), respectively, we obtain

$$(\beta(\eta^{n,s})(\mathbf{u}^n + \mathbf{b}^n) - \beta(\eta_h^{n,s})(\mathbf{u}_h^n + \mathbf{b}^n), \mathbf{v}) = 0 \ \forall \mathbf{v} \in Z_h, \tag{4}$$

$$(d_t\varepsilon_\eta^n, r) + (g(|\widetilde{\mathbf{u}}^n + \mathbf{b}^{n-1/2}|)\overline{\eta}^n - g(|\widetilde{\mathbf{u}}_h^n + \mathbf{b}^{n-1/2}|)\overline{\eta}_h^n, r) = I^n(r) \ \forall r \in R_h \tag{5}$$

where

$$\begin{aligned}I^n(r) &= (d_t\eta^n - \frac{\partial \eta^{n-1/2}}{\partial t}, r) + (g(|\widetilde{\mathbf{u}}^n + \mathbf{b}^{n-1/2}|)\overline{\eta}^n \\ &\qquad - g(|\mathbf{u}^{n-1/2} + \mathbf{b}^{n-1/2}|)\eta^{n-1/2}, r) \\ &= (d_t\eta^n - \frac{\partial \eta^{n-1/2}}{\partial t}, r) + ((g(|\widetilde{\mathbf{u}}^n + \mathbf{b}^{n-1/2}|) - g(|\mathbf{u}^{n-1/2} + \mathbf{b}^{n-1/2}|))\overline{\eta}^n \\ &+ g(|\mathbf{u}^{n-1/2} + \mathbf{b}^{n-1/2}|)(\overline{\eta}^n - \eta^{n-1/2}), r).\end{aligned}$$

Rewriting (4) yields

$$((\beta(\eta^{n,s}) - \beta(\eta_h^{n,s}))(\mathbf{u}^n + \mathbf{b}^n) + \beta(\eta_h^{n,s})(\mathbf{u}^n - \mathbf{u}_h^n), \mathbf{v}) = 0, \forall \mathbf{v} \in Z_h,$$

or, equivalently,

$$(\beta(\eta_h^{n,s})\varepsilon_{\mathbf{u}}, \mathbf{v}) = ((\beta(\eta_h^{n,s}) - \beta(\eta^{n,s}))(\mathbf{u}^n + \mathbf{b}^n), \mathbf{v}), \forall \mathbf{v} \in Z_h. \tag{6}$$

Setting $\varepsilon_{\mathbf{u}} = \mathbf{E}^n + \mathbf{\Lambda}^n$ and $\mathbf{v} = \mathbf{E}^n$ in (6), we obtain

$$(\beta(\eta_h^{n,s})\mathbf{E}^n, \mathbf{E}^n) = ((\beta(\eta_h^{n,s}) - \beta(\eta^{n,s}))(\mathbf{u}^n + \mathbf{b}^n), \mathbf{E}^n) - (\beta(\eta_h^{n,s})\mathbf{\Lambda}^n, \mathbf{E}^n). \tag{7}$$

Using assumptions $\boldsymbol{A\beta 2}$ and $\boldsymbol{A\beta 3}$, (7) yields

$$\beta_{\min}\|\mathbf{E}^n\|^2 \leq \|\mathbf{u}^n + \mathbf{b}^n\|_\infty \beta_{\text{Lip}}\|\eta_h^{n,s} - \eta^{n,s}\|\|\mathbf{E}^n\| + \beta_{\max}\|\mathbf{\Lambda}^n\|\|\mathbf{E}^n\|.$$

Thus,

$$\beta_{\min}\|\mathbf{E}^n\| \leq \beta_{\text{Lip}}\|\mathbf{u}^n + \mathbf{b}^n\|_\infty\|\eta_h^{n,s} - \eta^{n,s}\| + \beta_{\max}\|\mathbf{\Lambda}^n\|. \tag{8}$$

Now note that from $\boldsymbol{A\eta^s 1}$

$$\begin{aligned}\|\mathcal{S}(\eta_h^n) - \eta^{n,s}\| &\leq |\Omega|^{1/2}\|\mathcal{S}(\eta_h^n) - \eta^{n,s}\|_{L^\infty(\Omega)} \leq |\Omega|^{1/2}C_s\|\eta_h^n - \eta^n\| \\ &\leq |\Omega|^{1/2}C_s(\|\psi^n\| + \|F^n\|).\end{aligned}$$

Thus,

$$\begin{aligned}\|\eta_h^{n,s} - \eta^{n,s}\| &\leq \|\eta_h^{n,s} - \mathcal{S}(\eta_h^n)\| + \|\mathcal{S}(\eta_h^n) - \eta^{n,s}\| \\ &\leq \|\eta_h^{n,s} - \mathcal{S}(\eta_h^n)\| + |\Omega|^{1/2}C_s(\|\psi^n\| + \|F^n\|).\end{aligned} \tag{9}$$

Using (9) in (8) yields

$$\begin{aligned}\beta_{\min}\|\mathbf{E}^n\| \leq{}& \beta_{\text{Lip}}\|\mathbf{u}^n + \mathbf{b}^n\|_\infty\|\eta_h^{n,s} - \mathcal{S}(\eta_h^n)\| + \beta_{\max}\|\mathbf{\Lambda}^n\| \\ &+ \beta_{\text{Lip}}|\Omega|^{1/2}C_s\|\mathbf{u}^n + \mathbf{b}^n\|_\infty(\|\psi^n\| + \|F^n\|).\end{aligned} \tag{10}$$

From the fact that for $\{a_i\}_{i=1}^q \subset \mathbb{R}$, $\left(\sum_{i=1}^q a_i\right)^2 \leq q \sum_{i=1}^q a_i^2$, we obtain

$$\begin{aligned}\|\mathbf{E}^n\|^2 \leq 3\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}\|\mathbf{u}^n + \mathbf{b}^n\|_\infty^2 \|\eta_h^{n,s} - \mathcal{S}(\eta_h^n)\|^2 + 3\frac{\beta_{\max}^2}{\beta_{\min}^2}\|\mathbf{\Lambda}^n\|^2 \\ + 6\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}\,|\Omega|\,C_s^2\|\mathbf{u}^n + \mathbf{b}^n\|_\infty^2(\|\psi^n\|^2 + \|F^n\|^2). \qquad (11)\end{aligned}$$

We now focus on (5) and rewrite it as

$$\begin{aligned}(d_t\varepsilon_\eta, r) = ((g(|\widetilde{\mathbf{u}}_h^n + \mathbf{b}^{n-1/2}|) - g(|\widetilde{\mathbf{u}}^n + \mathbf{b}^{n-1/2}|))\overline{\eta}^n, r) \\ + (g(|\widetilde{\mathbf{u}}_h^n + \mathbf{b}^{n-1/2}|)(\overline{\eta}_h^n - \overline{\eta}^n), r) + I^n(r), \forall r \in R_h. \qquad (12)\end{aligned}$$

Substituting $\varepsilon_\eta = F^n + \psi^n$ and $r = \overline{F}^n$ in (12) yields

$$\begin{aligned}(d_tF^n, \overline{F}^n) = -(d_t\psi^n, \overline{F}^n) + ((g(|\widetilde{\mathbf{u}}_h^n + \mathbf{b}^{n-1/2}|) - g(|\widetilde{\mathbf{u}}^n + \mathbf{b}^{n-1/2}|))\overline{\eta}^n, \overline{F}^n) \\ + (g(|\widetilde{\mathbf{u}}_h^n + \mathbf{b}^{n-1/2}|)(\overline{\eta}_h^n - \overline{\eta}^n), \overline{F}^n) + I^n(\overline{F}^n). \qquad (13)\end{aligned}$$

Rewriting and bounding the terms in (13), we obtain

$$(d_tF^n, \overline{F}^n) = \frac{\|F^n\|^2 - \|F^{n-1}\|^2}{2\Delta t}, \qquad (14)$$

$$(d_t\psi^n, \overline{F}^n) \leq \frac{1}{2}(\|d_t\psi^n\|^2 + \|\overline{F}^n\|), \qquad (15)$$

$$
\begin{aligned}
(g(|\widetilde{\mathbf{u}}_h^n + \mathbf{b}^{n-1/2}|) - g(|\widetilde{\mathbf{u}}^n + \mathbf{b}^{n-1/2}|))\overline{\eta}^n, \overline{F}^n) & \\
&\leq \|\overline{\eta}^n\|_\infty \, \|g(|\widetilde{\mathbf{u}}_h^n + \mathbf{b}^{n-1/2}|) - g(|\widetilde{\mathbf{u}}^n + \mathbf{b}^{n-1/2}|)\| \, \|\overline{F}^n\| \\
&\leq \|\overline{\eta}^n\|_\infty \, g_{\text{Lip}} \, \| \, |\widetilde{\mathbf{u}}_h^n + \mathbf{b}^{n-1/2}| - |\widetilde{\mathbf{u}}^n + \mathbf{b}^{n-1/2}| \, \| \, \|\overline{F}^n\| \\
&\leq g_{\text{Lip}} \|\overline{\eta}^n\|_\infty \|\widetilde{\mathbf{u}}_h^n - \widetilde{\mathbf{u}}^n\| \, \|\overline{F}^n\| \\
&\leq \frac{1}{2} g_{\text{Lip}}^2 \|\overline{\eta}^n\|_\infty^2 \|\widetilde{\mathbf{u}}_h^n - \widetilde{\mathbf{u}}^n\|^2 + \frac{1}{2} \|\overline{F}^n\|^2 \\
&\leq g_{\text{Lip}}^2 \|\overline{\eta}^n\|_\infty^2 (\|\widetilde{\mathbf{E}}^n\|^2 + \|\widetilde{\mathbf{\Lambda}}^n\|^2) + \frac{1}{2} \|\overline{F}^n\|^2, \qquad (16)
\end{aligned}
$$

and

$$
\begin{aligned}
(g(|\widetilde{\mathbf{u}}_h^n + \mathbf{b}^{n-1/2}|)(\overline{\eta}_h^n - \overline{\eta}^n), \overline{F}^n) &\leq g_{\max} \|\overline{\eta}_h^n - \overline{\eta}^n\| \|\overline{F}^n\| \\
&\leq \frac{1}{2} g_{\max}^2 \|\overline{\eta}_h^n - \overline{\eta}^n\|^2 + \frac{1}{2} \|\overline{F}^n\| \\
&\leq g_{\max}^2 (\|\overline{F}^n\|^2 + \|\overline{\psi}^n\|^2) + \frac{1}{2} \|\overline{F}^n\|, \qquad (17)
\end{aligned}
$$

which follow from Cauchy-Schwarz and assumptions ***Ag2***-***Ag3***. For each term in $I^n(\overline{F}^n)$, we compute the bounds

$$
(d_t \eta^n - \frac{\partial \eta^{n-1/2}}{\partial t}, \overline{F}^n) \leq \frac{1}{2} \|d_t \eta^n - \frac{\partial \eta^{n-1/2}}{\partial t}\|^2 + \frac{1}{2} \|\overline{F}^n\|^2, \qquad (18)
$$

$$
\begin{aligned}
((g(|\widetilde{\mathbf{u}}^n + \mathbf{b}^{n-1/2}|) - g(|\mathbf{u}^{n-1/2} + \mathbf{b}^{n-1/2}|))\overline{\eta}^n, \overline{F}^n) & \\
&\leq \|\overline{\eta}^n\|_\infty g_{\text{Lip}} \|\widetilde{\mathbf{u}}^n - \mathbf{u}^{n-1/2}\| \|\overline{F}^n\| \\
&\leq \frac{1}{2} g_{\text{Lip}}^2 \|\overline{\eta}^n\|_\infty^2 \|\widetilde{\mathbf{u}}^n - \mathbf{u}^{n-1/2}\|^2 + \frac{1}{2} \|\overline{F}^n\|^2, \qquad (19)
\end{aligned}
$$

and

$$(g(|\mathbf{u}^{n-1/2}+\mathbf{b}^{n-1/2}|)(\overline{\eta}^n-\eta^{n-1/2}),\overline{F}^n) \\ \leq \frac{g_{\max}^2}{2}\|\overline{\eta}^n-\eta^{n-1/2}\|^2+\frac{1}{2}\|\overline{F}^n\|^2. \tag{20}$$

Consequently, combining (18)-(20) yields

$$|I^n(\overline{F}^n)| \leq \frac{1}{2}\|d_t\eta^n-\frac{\partial\eta^{n-1/2}}{\partial t}\|^2+\frac{1}{2}g_{\mathrm{Lip}}^2\|\overline{\eta}^n\|_\infty^2\|\widetilde{\mathbf{u}}^n-\mathbf{u}^{n-1/2}\|^2 \\ +\frac{g_{\max}^2}{2}\|\overline{\eta}^n-\eta^{n-1/2}\|^2+\frac{3}{2}\|\overline{F}^n\|^2. \tag{21}$$

Thus, using (14)-(18) and (21) in (13) and multiplying by $2\Delta t$, we obtain the bound

$$\|F^n\|^2-\|F^{n-1}\|^2 \leq \Delta t\|d_t\psi^n\|^2+2\Delta t g_{\mathrm{Lip}}^2\|\overline{\eta}^n\|_\infty^2(\|\widetilde{\mathbf{E}}^n\|^2+\|\widetilde{\mathbf{\Lambda}}^n\|^2) \\ +2\Delta t g_{\max}^2\|\overline{\psi}^n\|^2+\Delta t(6+2g_{\max}^2)\|\overline{F}^n\|^2+\Delta t R^n(\mathbf{u},\eta). \tag{22}$$

where

$$R^n(\mathbf{u},\eta)=\|d_t\eta^n-\frac{\partial\eta^{n-1/2}}{\partial t}\|^2+g_{\mathrm{Lip}}^2\|\overline{\eta}^n\|_\infty^2\|\widetilde{\mathbf{u}}^n-\mathbf{u}^{n-1/2}\|^2 \\ +g_{\max}^2\|\overline{\eta}^n-\eta^{n-1/2}\|^2. \tag{23}$$

Note that owing to Lemma A.2, we can further bound (23) to obtain

$$R^n(\mathbf{u},\eta) \leq \frac{1}{1280}(\Delta t)^3\int_{t_{n-1}}^{t_n}\|\eta_{ttt}\|^2\,dt+g_{\mathrm{Lip}}^2\|\overline{\eta}^n\|_\infty^2\frac{39}{8}(\Delta t)^3\int_{t_{n-3}}^{t_{n-1/2}}\|\mathbf{u}_{tt}\|^2\,dt \\ +g_{\max}^2\frac{1}{48}(\Delta t)^3\int_{t_{n-1}}^{t_n}\|\eta_{tt}\|^2\,dt. \tag{24}$$

Using Lemma A.1 to bound the terms $\|\widetilde{\mathbf{E}}^n\|^2$, $\|\widetilde{\mathbf{\Lambda}}^n\|^2$, $\|\overline{\psi}^n\|^2$, $\|\overline{F}^n\|^2$, and Lemma A.2

to bound

$$\Delta t\|d_t\psi^n\|^2 \le \int_{t_{n-1}}^{t_n} \|\psi_t\|^2\,dt,$$

in (22), yields

$$\begin{aligned}\|F^n\|^2 - \|F^{n-1}\|^2 \le \int_{t_{n-1}}^{t_n} \|\psi_t\|^2\,dt + 2\Delta t g_{\text{Lip}}^2\|\overline{\eta}^n\|_\infty^2(2\|\mathbf{E}^{n-1}\|^2 + \|\mathbf{E}^{n-2}\|^2 \\ + \|\mathbf{E}^{n-3}\|^2 + 2\|\mathbf{\Lambda}^{n-1}\|^2 + \|\mathbf{\Lambda}^{n-2}\|^2 + \|\mathbf{\Lambda}^{n-3}\|^2) + 2\Delta t\, g_{\max}^2\frac{1}{2}(\|\psi^n\|^2 \\ + \|\psi^{n-1}\|^2) + \Delta t(6 + 2g_{\max}^2)\frac{1}{2}(\|F^n\|^2 + \|F^{n-1}\|^2) + \Delta t R^n(\mathbf{u},\eta). \qquad (25)\end{aligned}$$

Summing (25) from $n = 3$ to $n = \ell$, and bounding

$$\|\overline{\eta}^n\|_\infty^2 = \frac{1}{4}\|\eta^n + \eta^{n-1}\|_\infty^2 \le \frac{2}{4}(\|\eta^n\|_\infty^2 + \|\eta^{n-1}\|_\infty^2),$$

we obtain

$$\begin{aligned}\|F^\ell\|^2 - \|F^2\|^2 \le 8g_{\text{Lip}}^2\Delta t\sum_{n=0}^{\ell-1}|||\eta|||_\infty^2(\|\mathbf{E}^n\|^2 + \|\mathbf{\Lambda}^n\|^2) + 2g_{\max}^2\Delta t\sum_{n=2}^{\ell}\|\psi^n\|^2 \\ + \Delta t(6 + 2g_{\max}^2)\sum_{n=2}^{\ell}\|F^n\|^2 + \Delta t\sum_{n=3}^{\ell}R^n(\mathbf{u},\eta) + \int_{t_2}^{t_\ell}\|\psi_t\|^2\,dt. \qquad (26)\end{aligned}$$

In view of the bound for $\|\mathbf{E}^n\|^2$ given in (11), (26) becomes

$$\begin{aligned}
\|F^\ell\|^2 - \|F^2\|^2 \le\ & 8(1+3\frac{\beta_{\max}^2}{\beta_{\min}^2})g_{\text{Lip}}^2\Delta t\sum_{n=0}^{\ell-1}|||\eta|||_\infty^2\|\mathbf{\Lambda}^n\|^2 \\
& + \Delta t\sum_{n=1}^{\ell}(2g_{\max}^2 + 48g_{\text{Lip}}^2|||\eta|||_\infty^2\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}|\Omega|C_s^2\|\mathbf{u}^n+\mathbf{b}^n\|_\infty^2)\|\psi^n\|^2 \\
& + \Delta t\sum_{n=1}^{\ell}(6+2g_{\max}^2 + 48g_{\text{Lip}}^2|||\eta|||_\infty^2\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}|\Omega|C_s^2\|\mathbf{u}^n+\mathbf{b}^n\|_\infty^2)\|F^n\|^2 \\
& + \Delta t\sum_{n=1}^{\ell-1}24g_{\text{Lip}}^2|||\eta|||_\infty^2\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}\|\mathbf{u}^n+\mathbf{b}^n\|_\infty^2\|\eta_h^{n,s}-\mathcal{S}(\eta_h^n)\|^2 \\
& + \Delta t\sum_{n=1}^{\ell-1}R^n(\mathbf{u},\eta) + \int_{t_2}^{t_\ell}\|\psi_t\|^2\,dt. \qquad (27)
\end{aligned}$$

Define

$$\begin{aligned}
\widetilde{w}_1^n &= 8(1+3\frac{\beta_{\max}^2}{\beta_{\min}^2})g_{\text{Lip}}^2|||\eta|||_\infty^2, \\
\widetilde{w}_2^n &= 2g_{\max}^2 + 48g_{\text{Lip}}^2\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}|||\eta|||_\infty^2\,|\Omega|\,C_s^2\|\mathbf{u}^n+\mathbf{b}^n\|_\infty^2, \\
\widetilde{w}_3^n &= 6 + w_2^n, \\
\widetilde{w}_4^n &= 24g_{\text{Lip}}^2|||\eta|||_\infty^2\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}\|\mathbf{u}^n+\mathbf{b}^n\|_\infty^2, \\
\widetilde{w}_5^n &= \frac{39}{8}g_{\text{Lip}}^2|||\eta|||_\infty^2,
\end{aligned}$$

and let $w_i = |||\widetilde{w}_i|||_\infty$. Thus, using (24) and the definition of the w_i in (27), yields

$$\begin{aligned}
\|F^\ell\|^2 - \|F^2\|^2 \le\ & \Delta t\sum_{n=0}^{\ell}\Big(w_1\|\mathbf{\Lambda}^n\|^2 + w_2\|\psi^n\|^2 + w_3\|F^n\|^2 \\
& + w_4\|\eta_h^{n,s}-\mathcal{S}(\eta_h^n)\|_2^2\Big) + \int_{t_2}^{t_\ell}\|\psi_t\|^2\,dt + (\Delta t)^4\,w_5\int_0^{t_{\ell-1/2}}\|\mathbf{u}_{tt}\|^2\,dt \\
& + (\Delta t)^4\,\frac{g_{\max}^2}{48}\int_{t_2}^{t_\ell}\|\eta_{tt}\|^2\,dt + \frac{(\Delta t)^4}{1280}\int_{t_2}^{t_\ell}\|\eta_{ttt}\|^2\,dt. \qquad (28)
\end{aligned}$$

Choosing Δt so that $\Delta t\, w_3 < 1/2$, using the discrete Gronwall's lemma [60] in (28) and the fact that $\Delta t\, \ell \leq \Delta t\, N = T$, results in the bound

$$\begin{aligned}\|F^{\ell}\|^2 \leq K\Delta t \sum_{n=0}^{\ell} \Big(w_1 \|\mathbf{\Lambda}^n\|^2 + w_2 \|\psi^n\|^2 + w_4 \|\eta_h^{n,s} - \mathcal{S}(\eta_h^n)\|_2^2 \Big) \\ + K\Big(\int_{t_2}^{t_\ell} \|\psi_t\|^2\, dt + (\Delta t)^4\, w_5 \int_0^{t_{\ell-1/2}} \|\mathbf{u}_{tt}\|^2\, dt + \|F^2\|^2 \\ + (\Delta t)^4\, \frac{g_{\max}^2}{48} \int_{t_2}^{t_\ell} \|\eta_{tt}\|^2\, dt + \frac{(\Delta t)^4}{1280} \int_{t_2}^{t_\ell} \|\eta_{ttt}\|^2\, dt \Big), \end{aligned} \tag{29}$$

where

$$K = \exp\left(\frac{T\, w_3}{1 - \Delta t\, w_3} \right) \leq \exp\left(2T\, w_3\right).$$

Using elements of order k for the velocity and elements of order l for the porosity in the corresponding approximating spaces, we obtain the bounds

$$\begin{aligned} &\Delta t \sum_{n=0}^{\ell} \Big(w_1 \|\mathbf{\Lambda}^n\|^2 + w_2 \|\psi^n\|^2 + w_4 \|\eta_h^{n,s} - \mathcal{S}(\eta_h^n)\|_2^2 \Big) \\ &\leq \Delta t \sum_{n=0}^{\ell} \Big(w_1 C h^{2k+2} \|\mathbf{u}^n\|_{k+1}^2 + w_2 C h^{2m+2} \|\eta^n\|_{m+1}^2 + w_4 \|\eta_h^{n,s} - \mathcal{S}(\eta_h^n)\|_2^2 \Big) \\ &\leq w_1 C h^{2k+2} |||\mathbf{u}|||_{k+1}^2 + w_2 C h^{2m+2} |||\eta|||_{m+1}^2 + w_4 |||\eta_h^s - \mathcal{S}(\eta_h)|||_2^2, \text{ and} \end{aligned} \tag{30}$$

$$\int_{t_2}^{t_\ell} \|\psi_t\|^2\, dt \leq C h^{2m+2} \int_{t_2}^{t_\ell} \|\eta_t\|_{m+1}^2\, dt. \tag{31}$$

Substituting (30) and (31) in (29), yields

$$\begin{aligned}\|F^{\ell}\|^2 \leq K\Big(& w_1 C h^{2k+2} |||\mathbf{u}|||_{k+1}^2 + w_2 C h^{2m+2} |||\eta|||_{m+1}^2 + w_4 |||\eta_h^s - \mathcal{S}(\eta_h)|||_2^2 \\ & + C h^{2m+2} \int_{t_2}^{t_\ell} \|\eta_t\|_{m+1}^2 \, dt + (\Delta t)^4 \, w_5 \int_0^{t_{\ell-2}} \|\mathbf{u}_{tt}\|^2 \, dt + \|F^2\|^2 \\ & + (\Delta t)^4 \, \frac{g_{\max}^2}{48} \int_{t_2}^{t_\ell} \|\eta_{tt}\|^2 \, dt + \frac{(\Delta t)^4}{1280} \int_{t_2}^{t_\ell} \|\eta_{ttt}\|^2 \, dt\Big). \end{aligned} \tag{32}$$

We are now in position to bound the term $\|\mathbf{E}^{\ell}\|^2$. Replacing (32) in (11), we obtain

$$\begin{aligned}\|\mathbf{E}^{\ell}\|^2 \leq\ & 3\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2} \|\mathbf{u}^{\ell} + \mathbf{b}^{\ell}\|_{\infty}^2 \|\eta_h^{\ell,s} - \mathcal{S}(\eta_h^{\ell})\|^2 + 3C\frac{\beta_{\max}^2}{\beta_{\min}^2} h^{2k+2} \|\mathbf{u}^{\ell}\|_{k+1}^2 \\ & + 6\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2} |\Omega| C_s^2 \|\mathbf{u}^{\ell} + \mathbf{b}^{\ell}\|_{\infty}^2 \Big(C h^{2m+2} \|\eta^{\ell}\|_{m+1}^2 + K\big(w_1 C h^{2k+2} |||\mathbf{u}|||_{k+1}^2 \\ & + w_2 C h^{2m+2} |||\eta|||_{m+1}^2 + w_4 |||\eta_h^s - \mathcal{S}(\eta_h)|||_2^2 + C h^{2m+2} \int_{t_2}^{t_\ell} \|\eta_t\|_{m+1}^2 \, dt \\ & + (\Delta t)^4 \, w_5 \int_0^{t_{\ell-2}} \|\mathbf{u}_{tt}\|^2 \, dt + \|F^2\|^2 + (\Delta t)^4 \, \frac{g_{\max}^2}{48} \int_{t_2}^{t_\ell} \|\eta_{tt}\|^2 \, dt \\ & + \frac{(\Delta t)^4}{1280} \int_{t_2}^{t_\ell} \|\eta_{ttt}\|^2 \, dt\big)\Big). \end{aligned}$$

Using the fact that

$$\|\mathbf{u}^{\ell} - \mathbf{u}_h^{\ell}\|^2 \leq 2(\|\mathbf{E}^{\ell}\|^2 + \|\mathbf{\Lambda}^{\ell}\|^2) \text{ and } \|\eta^{\ell} - \eta_h^{\ell}\|^2 \leq 2(\|F^{\ell}\|^2 + \|\psi^{\ell}\|^2),$$

we obtain the estimates

$$\begin{aligned}\|\mathbf{u}^\ell - \mathbf{u}_h^\ell\|^2 \leq 6\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}\|\mathbf{u}^\ell + \mathbf{b}^\ell\|_\infty^2\|\eta_h^{\ell,s} - \mathcal{S}(\eta_h^\ell)\|^2 + (2 + 6\frac{\beta_{\max}^2}{\beta_{\min}^2})Ch^{2k+2}\|\mathbf{u}^\ell\|_{k+1}^2 \\ + 12\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}|\Omega|C_s^2\|\mathbf{u}^\ell + \mathbf{b}^\ell\|_\infty^2\Big(Ch^{2m+2}\|\eta^\ell\|_{m+1}^2 + K\big(w_1Ch^{2k+2}|||\mathbf{u}|||_{k+1}^2 \\ + w_2Ch^{2m+2}|||\eta|||_{m+1}^2 + w_4|||\eta_h^s - \mathcal{S}(\eta_h)|||_2^2 + Ch^{2m+2}\int_{t_2}^{t_\ell}\|\eta_t\|_{m+1}^2\,dt \\ + (\Delta t)^4\,w_5\int_0^{t_{\ell-2}}\|\mathbf{u}_{tt}\|^2\,dt + \|F^2\|^2 + (\Delta t)^4\,\frac{g_{\max}^2}{48}\int_{t_2}^{t_\ell}\|\eta_{tt}\|^2\,dt \\ + \frac{(\Delta t)^4}{1280}\int_{t_2}^{t_\ell}\|\eta_{ttt}\|^2\,dt\big)\Big),\end{aligned} \tag{33}$$

and

$$\begin{aligned}\|\eta^\ell - \eta_h^\ell\|^2 \leq 2K\Big(w_1Ch^{2k+2}|||\mathbf{u}|||_{k+1}^2 + w_2Ch^{2m+2}|||\eta|||_{m+1}^2 + w_4|||\eta_h^s - \mathcal{S}(\eta_h)|||_2^2 \\ + Ch^{2m+2}\int_{t_2}^{t_\ell}\|\eta_t\|_{m+1}^2\,dt + (\Delta t)^4\,w_5\int_0^{t_{\ell-2}}\|\mathbf{u}_{tt}\|^2\,dt + \|F^2\|^2 \\ + (\Delta t)^4\,\frac{g_{\max}^2}{48}\int_{t_2}^{t_\ell}\|\eta_{tt}\|^2\,dt + \frac{(\Delta t)^4}{1280}\int_{t_2}^{t_\ell}\|\eta_{ttt}\|^2\,dt\Big) + 2Ch^{2m+2}\|\eta^\ell\|_{m+1}^2.\end{aligned} \tag{34}$$

Remark .1. *Owing to* (4.12)*, the error estimates* (33) *and* (34) *are optimal in view that there exists a constant* $C = C(\Omega,\mathbf{u},\eta,\mathcal{S},\beta,g)$ *such that*

$$\begin{aligned}\|\mathbf{u}^\ell - \mathbf{u}_h^\ell\| &\leq C\left(h^{k+1} + h^{m+1} + (\Delta t)^2\right),\\ \|\eta^\ell - \eta_h^\ell\| &\leq C\left(h^{k+1} + h^{m+1} + (\Delta t)^2\right).\end{aligned}$$

We now derive an error estimate corresponding to the pressure. For that purpose, recall that the the finite element pair $RT_k \times P_k$ is estable [18], i.e., it satisfies

the the discrete inf-sup condition

$$0 < \gamma \leq \inf_{q_h \in Q_h} \sup_{\mathbf{v}_h \in X_h} \frac{(q_h, \nabla \cdot \mathbf{v}_h)}{\|q_h\| \, \|\mathbf{v}_h\|_X}. \tag{35}$$

Thus, owing to (3.35), (4.1) and (35), we obtain for $\mathcal{Q}^n \in Q_h$

$$\begin{aligned}
\gamma \, \|p_h^n - \mathcal{Q}^n\| &\leq \sup_{\mathbf{v}_h \in X_h} \frac{(p_h^n - \mathcal{Q}^n, \nabla \cdot \mathbf{v}_h)}{\|\mathbf{v}_h\|_X} = \sup_{\mathbf{v}_h \in X_h} \frac{(p_h^n, \nabla \cdot \mathbf{v}_h) - (\mathcal{Q}^n, \nabla \cdot \mathbf{v}_h)}{\|\mathbf{v}_h\|_X} \\
&= \sup_{\mathbf{v}_h \in X_h} \frac{(\beta(\eta_h^{n,s})(\mathbf{u}_h^n + \mathbf{b}^n), \mathbf{v}_h) - (\mathbf{f}^n, \mathbf{v}_h) - (\mathcal{Q}^n, \nabla \cdot \mathbf{v}_h)}{\|\mathbf{v}_h\|_X} \\
&\leq \sup_{\mathbf{v}_h \in X_h} \frac{(\beta(\eta_h^{n,s})(\mathbf{u}_h^n + \mathbf{b}^n), \mathbf{v}_h) - (\beta(\eta^{n,s})(\mathbf{u}^n + \mathbf{b}^n), \mathbf{v}_h)}{\|\mathbf{v}_h\|_X} \\
&+ \sup_{\mathbf{v}_h \in X_h} \frac{(p^n, \nabla \cdot \mathbf{v}_h) - (\mathcal{Q}^n, \nabla \cdot \mathbf{v}_h)}{\|\mathbf{v}_h\|_X} \\
&= \sup_{\mathbf{v}_h \in X_h} \frac{(\beta(\eta_h^{n,s})(\mathbf{u}_h^n - \mathbf{u}^n), \mathbf{v}_h) + ((\mathbf{u}^n + \mathbf{b}^n)(\beta(\eta_h^{n,s}) - \beta(\eta^{n,s})), \mathbf{v}_h)}{\|\mathbf{v}_h\|_X} \\
&+ \sup_{\mathbf{v}_h \in X_h} \frac{(p^n - \mathcal{Q}^n, \nabla \cdot \mathbf{v}_h)}{\|\mathbf{v}_h\|_X} \\
&\leq \beta_{\max} \|\mathbf{u}_h^n - \mathbf{u}^n\| + \beta_{\mathrm{Lip}} \|\mathbf{u}^n + \mathbf{b}^n\|_\infty \|\eta_h^{n,s} - \eta^{n,s}\| + \|p^n - \mathcal{Q}^n\|.
\end{aligned} \tag{36}$$

Finally, combining the bound

$$\|p^n - p_h^n\| \leq \|p^n - \mathcal{Q}^n\| + \|p_h^n - \mathcal{Q}^n\|,$$

with (36), yields

$$\begin{aligned}
\|p^n - p_h^n\| \leq \gamma^{-1} \, &(\beta_{\max} \|\mathbf{u}_h^n - \mathbf{u}^n\| + \beta_{\mathrm{Lip}} \|\mathbf{u}^n + \mathbf{b}^n\|_\infty \|\eta_h^{n,s} - \eta^{n,s}\|) \\
&+ (1 + \gamma^{-1}) \, \|p_h^n - \mathcal{Q}^n\|.
\end{aligned} \tag{37}$$

Hence, in view of (37) and Remark .1 , we obtain the estimate

$$\|p^\ell - p_h^\ell\| \leq C\left(h^{k+1} + h^{m+1} + (\Delta t)^2\right).$$

A.2 Initialization of the Approximation Scheme

In this section we provide the initialization steps that guarantee optimal convergence (see (4.7)) for the numerical scheme introduced in Section 4.1. We split the process into three steps. In Step 1 we construct a suboptimal approximation $(\mathbf{z}_h^2, \pi_h^2)$ to $(\mathbf{u}^2, \eta^2)$ using $(\mathbf{u}_h^0, \eta_h^0)$. Then, in Step 2, we introduce the approximation $(\mathbf{z}_h^1, \pi_h^1)$ $= \frac{1}{2}\left((\mathbf{z}_h^2, \pi_h^2) + (\mathbf{u}_h^0, \eta_h^0)\right)$ and use it to obtain an optimal approximation $(\mathbf{u}_h^2, \eta_h^2)$ to $(\mathbf{u}^2, \eta^2)$. Finally, in Step 3, the average between $(\mathbf{u}_h^2, \eta_h^2)$ and $(\mathbf{u}_h^0, \eta_h^0)$ yields an optimal approximation $(\mathbf{u}_h^1, \eta_h^1)$ to $(\mathbf{u}^1, \eta^1)$.

We use a notation similar to that discussed in Section A.1. For $\mathcal{U}^n, \mathbf{z}_h^n \in Z_h$ and τ^n, $\pi_h^n \in R_h$, define

$$\begin{aligned} \mathbf{\Lambda}_{\mathbf{z}}^n &= \mathbf{u}^n - \mathcal{U}^n, & \mathbf{E}_{\mathbf{z}}^n &= \mathcal{U}^n - \mathbf{z}_h^n, \\ \psi_\pi^n &= \eta^n - \tau^n, & F_\pi^n &= \tau^n - \pi_h^n, \\ \varepsilon_{\mathbf{z}} &= \mathbf{u}^n - \mathbf{z}_h^n, & \varepsilon_\pi &= \eta^n - \pi_h^n. \end{aligned}$$

Step 1

Given $\eta_h^0 \in R_h$, and $\mathbf{z}_h^0 = \mathbf{u}_h^0 \in X_h$, determine $(\mathbf{z}_h^2, \pi_h^2) \in Z_h \times R_h$ satisfying

$$(\beta(\pi_h^{2,s})\mathbf{z}_h^2 + \beta(\pi_h^{2,s})\mathbf{b}^2, \mathbf{v}) = (\mathbf{f}^2, \mathbf{v}) \ \forall \mathbf{v} \in Z_h, \tag{38}$$

$$(d_{2t}\pi_h^2, r) + (g(|\mathbf{z}_h^0 + \mathbf{b}^0|)\pi_h^2, r) = 0 \ \forall r \in R_h. \tag{39}$$

First note that from (4.1)-(4.3), $\mathbf{u}$ and η satisfy

$$(\beta(\eta^{2,s})(\mathbf{u}^2+\mathbf{b}^2),\mathbf{v}) = (\mathbf{f}^2,\mathbf{v})\ \forall \mathbf{v}\in Z_h, \tag{40}$$

$$\begin{aligned}(d_{2t}\eta^2,r) + (g(|\mathbf{u}^0+\mathbf{b}^0|)\eta^2,r) &= (d_{2t}\eta^2 - \frac{\partial\eta^2}{\partial t},r) + (g(|\mathbf{u}^0+\mathbf{b}^0|)\eta^2,r)\\ &- (g(|\mathbf{u}^2+\mathbf{b}^2|)\eta^2,r)\ \forall r\in R_h.\end{aligned} \tag{41}$$

Subtracting (38) and (39) from (40) and (41), respectively, we obtain

$$(\beta(\eta^{2,s})(\mathbf{u}^2+\mathbf{b}^2) - \beta(\pi_h^{2,s})(\mathbf{z}_h^2+\mathbf{b}^2),\mathbf{v}) = 0\ \forall \mathbf{v}\in Z_h, \tag{42}$$

$$(d_{2t}\varepsilon_\pi^2,r) + (g(|\mathbf{u}^0+\mathbf{b}^0|)\eta^2 - g(|\mathbf{z}_h^0+\mathbf{b}^0|)\pi_h^2,r) = I_\pi^2(r)\ \forall r\in R_h, \tag{43}$$

where

$$I_\pi^2(r) = (d_{2t}\eta^2 - \frac{\partial\eta^2}{\partial t},r) + (g(|\mathbf{u}^0+\mathbf{b}^0|)\eta^2 - g(|\mathbf{u}^2+\mathbf{b}^2|)\eta^2,r). \tag{44}$$

Rewriting (42) yields

$$((\beta(\eta^{2,s}) - \beta(\pi_h^{2,s}))(\mathbf{u}^2+\mathbf{b}^2) + \beta(\pi_h^{2,s})(\mathbf{u}^2-\mathbf{z}_h^2),\mathbf{v}) = 0\ \forall \mathbf{v}\in Z_h,$$

or, equivalently

$$(\beta(\pi_h^{2,s})\varepsilon_{\mathbf{z}},\mathbf{v}) = ((\beta(\pi_h^{2,s}) - \beta(\eta^{2,s}))(\mathbf{u}^2+\mathbf{b}^2),\mathbf{v}),\forall \mathbf{v}\in Z_h. \tag{45}$$

Setting $\varepsilon_{\mathbf{z}} = \mathbf{E}_{\mathbf{z}}^2 + \mathbf{\Lambda}_{\mathbf{z}}^2$ and $\mathbf{v} = \mathbf{E}_{\mathbf{z}}^2$ in (45), we obtain

$$(\beta(\pi_h^{2,s})\mathbf{E}_{\mathbf{z}}^2,\mathbf{E}_{\mathbf{z}}^2) = ((\beta(\pi_h^{2,s}) - \beta(\eta^{2,s}))(\mathbf{u}^2+\mathbf{b}^2),\mathbf{E}_{\mathbf{z}}^2) - (\beta(\pi_h^{2,s})\mathbf{\Lambda}_{\mathbf{z}}^2,\mathbf{E}_{\mathbf{z}}^2). \tag{46}$$

Using assumptions $\boldsymbol{A\beta 2}$ and $\boldsymbol{A\beta 3}$ in (46) yields

$$\beta_{\min}\|\mathbf{E}_{\mathbf{z}}^2\|^2 \leq \|\mathbf{u}^2+\mathbf{b}^2\|_\infty \beta_{\mathrm{Lip}}\|\pi_h^{2,s}-\eta^{2,s}\|\|\mathbf{E}_{\mathbf{z}}^2\| + \beta_{\max}\|\mathbf{\Lambda}_{\mathbf{z}}^2\|\|\mathbf{E}_{\mathbf{z}}^2\|.$$

Proceeding as in the general case (see (11)) we obtain the bound

$$\begin{aligned}\|\mathbf{E}_{\mathbf{z}}^2\|^2 \leq 3\frac{\beta_{\mathrm{Lip}}^2}{\beta_{\min}^2}\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2\|\pi_h^{2,s}-\mathcal{S}(\pi_h^2)\|^2 + 3\frac{\beta_{\max}^2}{\beta_{\min}^2}\|\mathbf{\Lambda}_{\mathbf{z}}^2\|^2 \\ + 6\frac{\beta_{\mathrm{Lip}}^2}{\beta_{\min}^2}\,|\Omega|\,C_s^2\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2(\|\psi_\pi^2\|^2+\|F_\pi^2\|^2). \qquad (47)\end{aligned}$$

Now rewrite (43) as

$$\begin{aligned}(d_{2t}\varepsilon_\pi^2, r) = ((g(|\mathbf{z}_h^0+\mathbf{b}^0|)-g(|\mathbf{u}^0+\mathbf{b}^0|))\eta^2, r) \\ + (g(|\mathbf{z}_h^0+\mathbf{b}^0|)(\pi_h^2-\eta^2), r) + I_\pi^2(r)\ \forall r\in R_h. \qquad (48)\end{aligned}$$

and substitute $\varepsilon_\pi = F_\pi^2+\psi_\pi^2$, and $r=\breve{F}_\pi^2$ in (48) to obtain

$$\begin{aligned}(d_{2t}F_\pi^2, \breve{F}_\pi^2) = -(d_t\psi_\pi^2, \breve{F}_\pi^2) + ((g(|\mathbf{z}_h^0+\mathbf{b}^0|)-g(|\mathbf{u}^0+\mathbf{b}^0|))\eta^2, \breve{F}_\pi^2) \\ + (g(|\mathbf{z}_h^0+\mathbf{b}^0|)(\pi_h^2-\eta^2), \breve{F}_\pi^2) + I_\pi^2(\breve{F}_\pi^2). \qquad (49)\end{aligned}$$

Bounding the terms in (49) and observing that $\|\mathbf{E}_{\mathbf{z}}^0\| = \|F_\pi^0\| = 0$, yields

$$(d_{2t}F_\pi^2, \breve{F}_\pi^2) = \frac{\|F_\pi^2\|^2}{4\Delta t}, \qquad (50)$$

$$(d_{2t}\psi_\pi^2, \breve{F}_\pi^2) \leq \frac{1}{2}(\|d_{2t}\psi_\pi^2\|^2 + \|\breve{F}_\pi^2\|^2), \qquad (51)$$

$$
\begin{aligned}
&(g(|\mathbf{z}_h^0+\mathbf{b}^0|)-g(|\mathbf{u}^0+\mathbf{b}^0|))\eta^2,\breve{F}_\pi^2) \\
&\qquad \le \|\eta^2\|_\infty \, \|g(|\mathbf{z}_h^0+\mathbf{b}^0|)-g(|\mathbf{u}^0+\mathbf{b}^0|)\| \, \|\breve{F}_\pi^2\| \\
&\qquad \le \|\eta^2\|_\infty \, g_{\mathrm{Lip}} \, \| \, |\mathbf{z}_h^0+\mathbf{b}^0|-|\mathbf{u}^0+\mathbf{b}^0| \, \| \, \|\breve{F}_\pi^2\| \\
&\qquad \le g_{\mathrm{Lip}}\|\eta^2\|_\infty \|\mathbf{z}_h^0-\mathbf{u}^0\| \, \|\breve{F}_\pi^2\| \\
&\qquad \le \frac{1}{2} g_{\mathrm{Lip}}^2 \|\eta^2\|_\infty^2 \|\mathbf{z}_h^0-\mathbf{u}^0\|^2 + \frac{1}{2}\|\breve{F}_\pi^2\|^2 \\
&\qquad \le g_{\mathrm{Lip}}^2 \|\eta^2\|_\infty^2 (\|\mathbf{E}_{\mathbf{z}}^0\|^2 + \|\mathbf{\Lambda}_{\mathbf{z}}^0\|^2) + \frac{1}{2}\|\breve{F}_\pi^2\|^2, \qquad (52)
\end{aligned}
$$

and

$$
\begin{aligned}
(g(|\mathbf{z}_h^0+\mathbf{b}^0|)(\pi_h^2-\eta^2),\breve{F}_\pi^2) &\le g_{\max}\|\pi_h^2-\eta^2\|\|\breve{F}_\pi^2\| \\
&\le \frac{1}{2} g_{\max}^2 \|\pi_h^2-\eta^2\|^2 + \frac{1}{2}\|\breve{F}_\pi^2\|^2 \\
&\le g_{\max}^2(\|\breve{F}_\pi^2\|^2 + \|\breve{\psi}_\pi^2\|^2) + \frac{1}{2}\|\breve{F}_\pi^2\|^2. \qquad (53)
\end{aligned}
$$

Bounding each term in $I_\pi^2(\breve{F}_\pi^2)$ (see (44)), we obtain

$$
(d_{2t}\eta^2 - \frac{\partial \eta^2}{\partial t}, \breve{F}_\pi^2) \le \frac{1}{2}\|d_{2t}\eta^2 - \frac{\partial \eta^2}{\partial t}\|^2 + \frac{1}{2}\|\breve{F}_\pi^2\|^2, \qquad (54)
$$

and

$$
\begin{aligned}
&((g(|\mathbf{u}^0+\mathbf{b}^0|)-g(|\mathbf{u}^2+\mathbf{b}^2|))\eta^2,\breve{F}_\pi^2) \\
&\qquad \le \|\eta^2\|_\infty g_{\mathrm{Lip}}\|(\mathbf{u}^2+\mathbf{b}^2)-(\mathbf{u}^0+\mathbf{b}^0)\|\|\breve{F}_\pi^2\| \\
&\qquad \le \frac{1}{2} g_{\mathrm{Lip}}^2 \|\eta^2\|_\infty^2 \|(\mathbf{u}^2+\mathbf{b}^2)-(\mathbf{u}^0+\mathbf{b}^0)\|^2 + \frac{1}{2}\|\breve{F}_\pi^2\|^2, \qquad (55)
\end{aligned}
$$

Consequently, in view of (44), (54) and (55),

$$|I_\pi^2(\breve{F}_\pi^2)| \le \frac{1}{2}\|d_{2t}\eta^2 - \frac{\partial \eta^2}{\partial t}\|^2 + \frac{1}{2}g_{\text{Lip}}^2\|\eta^2\|_\infty^2\|(\mathbf{u}^2+\mathbf{b}^2)-(\mathbf{u}^0+\mathbf{b}^0)\|^2 + \|\breve{F}_\pi^2\|^2. \tag{56}$$

Thus, combining (49)-(53), (56) and multiplying by $4\Delta t$, yields

$$\|F_\pi^2\|^2 \le 2\Delta t\|d_{2t}\psi_\pi^2\|^2 + 4\Delta t g_{\text{Lip}}^2\|\eta^2\|_\infty^2\|\mathbf{\Lambda}_\mathbf{z}^0\|^2 + 4\Delta t g_{\max}^2\|\breve{\psi}_\pi^2\|^2 + 2\Delta t(5+2g_{\max}^2)\|\breve{F}^2\|^2 + 2\Delta t R_\pi^2(\mathbf{u},\mathbf{b},\eta), \tag{57}$$

where

$$R_\pi^2(\mathbf{u},\mathbf{b},\eta) = \|d_{2t}\eta^2 - \frac{\partial \eta^2}{\partial t}\|^2 + g_{\text{Lip}}^2\|\eta^2\|_\infty^2\|(\mathbf{u}^2+\mathbf{b}^2)-(\mathbf{u}^0+\mathbf{b}^0)\|^2. \tag{58}$$

Now note that

$$\begin{aligned}\int_\Omega((\mathbf{u}^2+\mathbf{b}^2)-(\mathbf{u}^0+\mathbf{b}^0))^2\,d\Omega &= \int_\Omega\left(\int_{t_0}^{t_2}(\mathbf{u}+\mathbf{b})_t(\cdot,t)\,dt\right)^2 d\Omega \\ &\le 2\int_\Omega \Delta t\int_{t_0}^{t_2}((\mathbf{u}+\mathbf{b})_t(\cdot,t))^2\,dt\,d\Omega = 2\Delta t\int_{t_0}^{t_2}\|(\mathbf{u}+\mathbf{b})_t(\cdot,t)\|^2\,dt \\ &\le 4(\Delta t)^2\sup_{t_0\le t\le t_2}\|(\mathbf{u}+\mathbf{b})_t(\cdot,t)\|^2.\end{aligned} \tag{59}$$

Similarly, using the fact that

$$\eta^2 = \eta^0 + \int_{t_0}^{t_2}\eta_t\,dt = \eta^0 + 2\Delta t\,\eta_t^2 + \int_{t_0}^{t_2}\eta_{tt}(t_2-t)\,dt,$$

yields

$$\begin{aligned}\int_\Omega (d_{2t}\eta^2 - \frac{\partial \eta^2}{\partial t})^2 \, d\Omega &= \int_\Omega \left(\frac{1}{2\Delta t} \int_{t_0}^{t_2} \eta_{tt}(t_2 - t)\, dt \right)^2 d\Omega \\ &\leq \frac{1}{4(\Delta t)^2} \int_\Omega \left(\int_{t_0}^{t_2} \eta_{tt}^2 \, dt \int_{t_0}^{t_2} (t_2 - t)^2 \, dt \right) d\Omega \\ &= \frac{\Delta t}{12} \int_\Omega \int_{t_0}^{t_2} \eta_{tt}^2 \, dt \, d\Omega = \frac{\Delta t}{12} \int_{t_0}^{t_2} \|\eta_{tt}(\cdot, t)\|^2 \, dt \\ &\leq \frac{(\Delta t)^2}{12} \sup_{t_0 \leq t \leq t_2} \|\eta_{tt}(\cdot, t)\|^2. \end{aligned} \tag{60}$$

Owing to (59) and (60), we can further bound (58) to obtain

$$\begin{aligned} R_\pi^2(\mathbf{u}, \mathbf{b}, \eta) \leq \frac{(\Delta t)^2}{12} &\sup_{t_0 \leq t \leq t_2} \|\eta_{tt}(\cdot, t)\|^2 \\ &+ 4(\Delta t)^2 g_{\text{Lip}}^2 \|\eta^2\|_\infty^2 \sup_{t_0 \leq t \leq t_2} \|(\mathbf{u} + \mathbf{b})_t(\cdot, t)\|^2. \end{aligned} \tag{61}$$

Making use of the triangle inequality to bound the term $\|\breve{\psi}_\pi^2\|^2$ and Lemma A.2 to bound

$$\|d_{2t}\psi_\pi^2\|^2 \leq \frac{1}{2\Delta t} \int_{t_0}^{t_2} \|(\psi_\pi)_t\|^2 \, dt$$

in (57), implies

$$\begin{aligned} \|F_\pi^2\|^2 \leq &\int_{t_0}^{t_2} \|(\psi_\pi)_t\|^2 \, dt + 4\Delta t g_{\text{Lip}}^2 \|\eta^2\|_\infty^2 \|\mathbf{\Lambda}_{\mathbf{z}}^0\|^2 + 4\Delta t \, g_{\max}^2 \frac{1}{2} (\|\psi_\pi^2\|^2 + \|\psi_\pi^0\|^2) \\ &+ 2\Delta t (5 + 2g_{\max}^2) \frac{1}{4} \|F_\pi^2\|^2 + 2\Delta t R_\pi^2(\mathbf{u}, \mathbf{b}, \eta). \end{aligned} \tag{62}$$

Choosing Δt in (62) so that $\Delta t \, (5/2 + g_{\max}^2) < 1$, and defining

$$K_1 = (1 - \Delta t \, (5/2 + g_{\max}^2))^{-1},$$

we obtain

$$\begin{aligned}\|F_\pi^2\|^2 &\le K_1\Big(\int_{t_0}^{t_2}\|(\psi_\pi)_t\|^2\,dt + 4\Delta t\, g_{\text{Lip}}^2\|\eta^2\|_\infty^2\|\mathbf{\Lambda}_{\mathbf{z}}^0\|^2\\ &\quad + 2\Delta t\, g_{\max}^2(\|\psi_\pi^2\|^2 + \|\psi_\pi^0\|^2) + 2\Delta t\, R_\pi^2(\mathbf{u},\mathbf{b},\eta)\Big)\\ &\le K_1\Big(Ch^{2m+2}\int_{t_0}^{t_2}\|\eta_t\|_{m+1}^2 + 4C\Delta t\, g_{\text{Lip}}^2\|\eta^2\|_\infty^2 h^{2k+2}\|\mathbf{u}^0\|_{k+1}^2\\ &\quad + 2C\Delta t\, g_{\max}^2 h^{2m+2}(\|\eta^2\|_{m+1}^2 + \|\eta^0\|_{m+1}^2) + \frac{(\Delta t)^3}{6}\sup_{t_0\le t\le t_2}\|\eta_{tt}(\cdot,t)\|^2\\ &\quad + 8(\Delta t)^3 g_{\text{Lip}}^2\|\eta^2\|_\infty^2\sup_{t_0\le t\le t_2}\|(\mathbf{u}+\mathbf{b})_t(\cdot,t)\|^2\Big). \end{aligned}\tag{63}$$

Replacing (63) in (47), yields

$$\begin{aligned}\|\mathbf{E}_{\mathbf{z}}^2\|^2 &\le 3\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2\|\pi_h^{2,s}-\mathcal{S}(\pi_h^2)\|^2 + 3\frac{\beta_{\max}^2}{\beta_{\min}^2}\|\mathbf{\Lambda}_{\mathbf{z}}^2\|^2\\ &\quad + 6\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}|\Omega|\,C_s^2\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2\Big(Ch^{2m+2}\|\eta^2\|_{m+1}^2 + K_1\big(Ch^{2m+2}\int_{t_0}^{t_2}\|\eta_t\|_{m+1}^2\\ &\quad + 4C\Delta t\, g_{\text{Lip}}^2\|\eta^2\|_\infty^2 h^{2k+2}\|\mathbf{u}^0\|_{k+1}^2 + 2C\Delta t\, g_{\max}^2 h^{2m+2}(\|\eta^2\|_{m+1}^2+\|\eta^0\|_{m+1}^2)\\ &\quad + \frac{(\Delta t)^3}{6}\sup_{t_0\le t\le t_2}\|\eta_{tt}(\cdot,t)\|^2 + 8(\Delta t)^3 g_{\text{Lip}}^2\|\eta^2\|_\infty^2\sup_{t_0\le t\le t_2}\|(\mathbf{u}+\mathbf{b})_t(\cdot,t)\|^2\big)\Big). \end{aligned}\tag{64}$$

Finally, combining the bounds

$$\|\mathbf{u}^2-\mathbf{z}_h^2\|^2 \le 2(\|\mathbf{E}_{\mathbf{z}}^2\|^2+\|\mathbf{\Lambda}_{\mathbf{z}}^2\|^2),\ \ \|\eta^2-\pi_h^2\|^2 \le 2(\|F_\pi^2\|^2+\|\psi_\pi^2\|^2),$$

with (64), yields

$$\|\mathbf{u}^2-\mathbf{z}_h^2\|^2 \le 6\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2\|\pi_h^{2,s}-\mathcal{S}(\pi_h^2)\|^2+(2+6\frac{\beta_{\max}^2}{\beta_{\min}^2})Ch^{2k+2}\|\mathbf{u}^2\|_{k+1}^2$$
$$+12\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}\,|\Omega|\,C_s^2\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2\Big(Ch^{2m+2}\|\eta^2\|_{m+1}^2+K_1\big(Ch^{2m+2}\int_{t_0}^{t_2}\|\eta_t\|_{m+1}^2$$
$$+4C\Delta t g_{\text{Lip}}^2\|\eta^2\|_\infty^2h^{2k+2}\|\mathbf{u}^0\|_{k+1}^2+2C\Delta t\,g_{\max}^2h^{2m+2}(\|\eta^2\|_{m+1}^2+\|\eta^0\|_{m+1}^2)$$
$$+\frac{(\Delta t)^3}{6}\sup_{t_0\le t\le t_2}\|\eta_{tt}(\cdot,t)\|^2+8(\Delta t)^3g_{\text{Lip}}^2\|\eta^2\|_\infty^2\sup_{t_0\le t\le t_2}\|(\mathbf{u}+\mathbf{b})_t(\cdot,t)\|^2\big)\Big), \quad (65)$$

and

$$\|\eta^2-\pi_h^2\|^2 \le 2K_1\Big(Ch^{2m+2}\int_{t_0}^{t_2}\|\eta_t\|_{m+1}^2+4C\Delta tg_{\text{Lip}}^2\|\eta^2\|_\infty^2\,h^{2k+2}\|\mathbf{u}^0\|_{k+1}^2$$
$$+2C\Delta t\,g_{\max}^2h^{2m+2}(\|\eta^2\|_{m+1}^2+\|\eta^0\|_{m+1}^2)+\frac{(\Delta t)^3}{6}\sup_{t_0\le t\le t_2}\|\eta_{tt}(\cdot,t)\|^2$$
$$+8(\Delta t)^3g_{\text{Lip}}^2\|\eta^2\|_\infty^2\sup_{t_0\le t\le t_2}\|(\mathbf{u}+\mathbf{b})_t(\cdot,t)\|^2\Big)+2Ch^{2m+2}\|\eta^2\|_{m+1}. \quad (66)$$

Step 2

In this section we consider the following problem: *For* $\pi_h^1=(\pi_h^2+\pi_h^0)/2\in R_h$, *and* $\mathbf{z}_h^1=(\mathbf{z}_h^2+\mathbf{z}_h^0)/2\in X_h$, *determine* $(\mathbf{u}_h^2,\eta_h^2)\in Z_h\times R_h$ *satisfying*

$$(\beta(\eta_h^{2,s})\mathbf{u}_h^2+\beta(\eta_h^{2,s})\mathbf{b}^2,\mathbf{v})=(\mathbf{f}^2,\mathbf{v})\ \forall\mathbf{v}\in Z_h, \quad (67)$$

$$(d_{2t}\eta_h^2,r)+(g(|\mathbf{z}_h^1+\mathbf{b}^1|)\breve{\eta}_h^2,r)=0\ \forall r\in R_h. \quad (68)$$

First note that, from (4.1)-(4.3), $(\mathbf{u},\eta)$ satisfies

$$(\beta(\eta^{2,s})(\mathbf{u}^2+\mathbf{b}^2),\mathbf{v})=(\mathbf{f}^2,\mathbf{v})\ \forall\mathbf{v}\in Z_h, \quad (69)$$

$$(d_{2t}\eta^2,r)+(g(|\mathbf{u}^1+\mathbf{b}^1|)\eta^1,r)=(d_{2t}\eta^2-\frac{\partial\eta^1}{\partial t},r)\ \forall r\in R_h. \quad (70)$$

Subtracting (67) and (68) from (69) and (70), respectively, we obtain

$$(\beta(\eta^{2,s})(\mathbf{u}^2+\mathbf{b}^2)-\beta(\eta_h^{2,s})(\mathbf{u}_h^2+\mathbf{b}^2),\mathbf{v})=0\ \forall\mathbf{v}\in Z_h, \tag{71}$$

$$(d_{2t}\varepsilon_\eta^2,r)+(g(|\mathbf{u}^1+\mathbf{b}^1|)\eta^1-g(|\mathbf{z}_h^1+\mathbf{b}^1|)\breve{\eta}_h^2,r)=I^2(r)\ \forall r\in R_h, \tag{72}$$

where

$$I^2(r)=(d_{2t}\eta^2-\frac{\partial\eta^1}{\partial t},r). \tag{73}$$

Rewriting (71), yields

$$((\beta(\eta^{2,s})-\beta(\eta_h^{2,s}))(\mathbf{u}^2+\mathbf{b}^2)+\beta(\eta_h^{2,s})(\mathbf{u}^2-\mathbf{u}_h^2),\mathbf{v})=0\ \forall\mathbf{v}\in Z_h,$$

or, equivalently

$$(\beta(\eta_h^{2,s})\varepsilon_{\mathbf{u}},\mathbf{v})=((\beta(\eta_h^{2,s})-\beta(\eta^{2,s}))(\mathbf{u}^2+\mathbf{b}^2),\mathbf{v})\ \forall\mathbf{v}\in Z_h, \tag{74}$$

where $\varepsilon_{\mathbf{u}}=\mathbf{E}^2+\mathbf{\Lambda}^2$ with $\mathbf{E}^2$, $\mathbf{\Lambda}^2$ as defined in (4.14). Setting $\mathbf{v}=\mathbf{E}^2$ in (74) to obtain

$$(\beta(\eta_h^{2,s})\mathbf{E}^2,\mathbf{E}^2)=((\beta(\eta_h^{2,s})-\beta(\eta^{2,s}))(\mathbf{u}^2+\mathbf{b}^2),\mathbf{E}^2)-(\beta(\eta_h^{2,s})\mathbf{\Lambda}^2,\mathbf{E}^2), \tag{75}$$

and using assumptions ***Aβ2*** and ***Aβ3*** in (75), yields

$$\beta_{\min}\|\mathbf{E}^2\|^2\leq\|\mathbf{u}^2+\mathbf{b}^2\|_\infty\,\beta_{\mathrm{Lip}}\|\eta_h^{2,s}-\eta^{2,s}\|\|\mathbf{E}^2\|+\beta_{\max}\|\mathbf{\Lambda}^2\|\|\mathbf{E}^2\|. \tag{76}$$

Proceeding as in the general case (see (11)) to bound (76), we obtain

$$\|\mathbf{E}^2\|^2 \leq 3\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2\|\eta_h^{2,s}-\mathcal{S}(\eta_h^2)\|^2+3\frac{\beta_{\max}^2}{\beta_{\min}^2}\|\mathbf{\Lambda}^2\|^2 \\ +6\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}\,|\Omega|\,C_s^2\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2(\|\psi^2\|^2+\|F^2\|^2). \tag{77}$$

We now rewrite (72) as

$$(d_{2t}\varepsilon_\eta^2, r) = ((g(|\mathbf{z}_h^1+\mathbf{b}^1|)-g(|\mathbf{u}^1+\mathbf{b}^1|))\eta^1, r) \\ +(g(|\mathbf{z}_h^1+\mathbf{b}^1|)(\breve{\eta}_h^2-\eta^1), r)+I^2(r), \forall r\in R_h, \tag{78}$$

and substitute $\varepsilon_\eta = F^2+\psi^2$ and $r=\breve{F}^2$ in (78), which yields

$$(d_{2t}F^2,\breve{F}^2) = -(d_t\psi^2,\breve{F}^2)+((g(|\mathbf{z}_h^1+\mathbf{b}^1|)-g(|\mathbf{u}^1+\mathbf{b}^1|))\eta^1,\breve{F}^2) \\ +(g(|\mathbf{z}_h^1+\mathbf{b}^1|)(\breve{\eta}_h^2-\eta^1),\breve{F}^2)+I^2(\breve{F}^2). \tag{79}$$

Bounding all the terms in (79) and using $\|F^0\|=0$, we obtain

$$(d_{2t}F^2,\breve{F}^2) = \frac{\|F^2\|^2}{4\Delta t}, \tag{80}$$

$$(d_{2t}\psi^2,\breve{F}^2) \leq \frac{1}{2}(\|d_{2t}\psi^2\|^2+\|\breve{F}^2\|), \tag{81}$$

$$
\begin{aligned}
&(g(|\mathbf{z}_h^1+\mathbf{b}^1|)-g(|\mathbf{u}^1+\mathbf{b}^1|))\eta^1,\breve{F}^2)\\
&\qquad\le \|\eta^1\|_\infty \|g(|\mathbf{z}_h^1+\mathbf{b}^1|)-g(|\mathbf{u}^1+\mathbf{b}^1|)\|\,\|\breve{F}^2\|\\
&\qquad\le \|\eta^1\|_\infty\, g_{\mathrm{Lip}}\,\|\,|\mathbf{z}_h^1+\mathbf{b}^1|-|\mathbf{u}^1+\mathbf{b}^1|\,\|\,\|\breve{F}^2\|\\
&\qquad\le g_{\mathrm{Lip}}\|\eta^1\|_\infty\|\mathbf{z}_h^1-\mathbf{u}^1\|\,\|\breve{F}^2\|\\
&\qquad\le \frac{1}{2}g_{\mathrm{Lip}}^2\|\eta^1\|_\infty^2\|\mathbf{z}_h^1-\mathbf{u}^1\|^2+\frac{1}{2}\|\breve{F}^2\|^2 \qquad (82)
\end{aligned}
$$

and

$$
\begin{aligned}
(g(|\mathbf{z}_h^1+\mathbf{b}^1|)(\breve{\eta}_h^2-\eta^1),\breve{F}^2) &\le g_{\max}\|\breve{\eta}_h^2-\eta^1\|\|\breve{F}^2\|\\
&\le \frac{1}{2}g_{\max}^2\|\breve{\eta}_h^2-\eta^1\|^2+\frac{1}{2}\|\breve{F}^2\|^2\\
&\le g_{\max}^2(\|\breve{\eta}_h^2-\breve{\eta}^2\|^2+\|\breve{\eta}^2-\eta^1\|^2)+\frac{1}{2}\|\breve{F}^2\|^2\\
&\le g_{\max}^2(2\|\breve{F}_h^2\|+2\|\breve{\psi}^2\|^2+\|\breve{\eta}^2-\eta^1\|^2)+\frac{1}{2}\|\breve{F}^2\|^2. \qquad (83)
\end{aligned}
$$

Observe that the terms $\|\mathbf{z}_h^1-\mathbf{u}^1\|^2$ in (82) and $I^2(\breve{F}^n)$ in (73) can be bounded as follows:

$$
\begin{aligned}
\|\mathbf{z}_h^1-\mathbf{u}^1\|^2 &= \|\frac{1}{2}(\mathbf{z}_h^0+\mathbf{z}_h^2)-\mathbf{u}^1\|^2\\
&\le 2\|\frac{1}{2}(\mathbf{z}_h^0+\mathbf{z}_h^2)-\frac{1}{2}(\mathbf{u}^0+\mathbf{u}^2)\|^2+2\|\frac{1}{2}(\mathbf{u}^0+\mathbf{u}^2)-\mathbf{u}^1\|^2\\
&\le \|\mathbf{z}_h^0-\mathbf{u}^0\|^2+\|\mathbf{z}_h^2-\mathbf{u}^2\|^2+2\|\frac{1}{2}(\mathbf{u}^0+\mathbf{u}^2)-\mathbf{u}^1\|^2\\
&= \|\mathbf{\Lambda}^0\|^2+\|\mathbf{z}_h^2-\mathbf{u}^2\|^2+2\|\breve{\mathbf{u}}^2-\mathbf{u}^1\|^2,\\
I^2(\breve{F}^n) = (d_{2t}\eta^2-\frac{\partial\eta^1}{\partial t},\breve{F}^2) &\le \frac{1}{2}\|d_{2t}\eta^2-\frac{\partial\eta^1}{\partial t}\|^2+\frac{1}{2}\|\breve{F}^n\|^2. \qquad (84)
\end{aligned}
$$

Thus, combining (79)-(84) and multiplying by $4\Delta t$, yields

$$\begin{aligned}\|F^2\|^2 \leq\ & 2\Delta t\|d_{2t}\psi^2\|^2 + 2\Delta t g_{\text{Lip}}^2\|\eta^1\|_\infty^2\big(\|\mathbf{\Lambda}^0\|^2 + \|\mathbf{z}_h^2 - \mathbf{u}^2\|^2 + 2\|\breve{\mathbf{u}}^2 - \mathbf{u}^1\|^2\big)\\ &+ 8\Delta t g_{\max}^2\|\breve{\psi}^2\|^2 + 4\Delta t g_{\max}^2\|\breve{\eta}^2 - \eta^1\|^2 + 2\Delta t(4 + 4g_{\max}^2)\|\breve{F}^2\|^2\\ &+ 2\Delta t\|d_{2t}\eta^2 - \frac{\partial\eta^1}{\partial t}\|^2. \end{aligned} \tag{85}$$

Recalling the definition of $\|\breve{F}^2\|$ (see (1)) and choosing Δt in (85) so that $2\Delta t\,(1 + g_{\max}^2) < 1$, yields

$$\begin{aligned}\|F^2\|^2 \leq K_2\Big(& 2\Delta t\|d_{2t}\psi^2\|^2 + 2\Delta t g_{\text{Lip}}^2\|\eta^1\|_\infty^2(\|\mathbf{\Lambda}^0\|^2 + \|\mathbf{z}_h^2 - \mathbf{u}^2\|^2\\ &+ 2\|\breve{\mathbf{u}}^2 - \mathbf{u}^1\|^2) + 8\Delta t g_{\max}^2\|\breve{\psi}^2\|^2 + 4\Delta t g_{\max}^2\|\breve{\eta}^2 - \eta^1\|^2\\ &+ 2\Delta t\|d_{2t}\eta^2 - \frac{\partial\eta^1}{\partial t}\|^2\Big), \end{aligned} \tag{86}$$

where

$$K_2 = (1 - 2\Delta t(1 + g_{\max}^2))^{-1}.$$

Using the bounds given Lemma A.2 and (65) in (86), we obtain

$$\begin{aligned}\|F^2\|^2 \leq K_2\Big(& Ch^{2m+2}\int_{t_0}^{t_2}\|\eta_t\|_{m+1}^2\,dt + 2\Delta t g_{\text{Lip}}^2\|\eta^1\|_\infty^2\Big(Ch^{2k+2}\|\mathbf{u}^0\|_{k+1}^2\\ &+ 6\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}\|\mathbf{u}^2 + \mathbf{b}^2\|_\infty^2\|\pi_h^{2,s} - \mathcal{S}(\pi_h^2)\|^2 + (2 + 6\frac{\beta_{\max}^2}{\beta_{\min}^2})Ch^{2k+2}\|\mathbf{u}^2\|_{k+1}^2\\ &+ 12\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}\,|\Omega|\,C_s^2\|\mathbf{u}^2 + \mathbf{b}^2\|_\infty^2\big(Ch^{2m+2}\|\eta^2\|_{m+1}^2 + K_1\big(Ch^{2m+2}\int_{t_0}^{t_2}\|\eta_t\|_{m+1}^2\\ &+ 4C\Delta t g_{\text{Lip}}^2\|\eta^2\|_\infty^2 h^{2k+2}\|\mathbf{u}^0\|_{k+1}^2 + 2C\Delta t\,g_{\max}^2 h^{2m+2}(\|\eta^2\|_{m+1}^2 + \|\eta^0\|_{m+1}^2)\\ &+ \frac{(\Delta t)^3}{6}\sup_{t_0\leq t\leq t_2}\|\eta_{tt}(\cdot,t)\|^2 + 8(\Delta t)^3 g_{\text{Lip}}^2\|\eta^2\|_\infty^2\sup_{t_0\leq t\leq t_2}\|(\mathbf{u}+\mathbf{b})_t(\cdot,t)\|^2\big)\big)\end{aligned}$$

$$+\frac{2}{3}(\Delta t)^4 \sup_{t_0\le t\le t_2}\|\mathbf{u}_{tt}(\cdot,t)\|^2\Big) + 4C\Delta t g_{\max}^2 h^{2m+2}(\|\eta^2\|_{m+1}^2 + \|\eta^0\|_{m+1}^2)$$
$$+\frac{4}{3}g_{\max}^2(\Delta t)^5 \sup_{t_0\le t\le t_2}\|\eta_{tt}(\cdot,t)\|^2 + \frac{1}{40}(\Delta t)^5 \sup_{t_0\le t\le t_2}\|\eta_{ttt}(\cdot,t)\|^2\Big). \tag{87}$$

Inserting (87) in (77), yields

$$\begin{aligned}
\|\mathbf{E}^2\|^2 \le\ & 3\frac{\beta_{\mathrm{Lip}}^2}{\beta_{\min}^2}\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2\|\eta_h^{2,s}-\mathcal{S}(\eta_h^2)\|^2 + 3C\frac{\beta_{\max}^2}{\beta_{\min}^2}h^{2k+2}\|\mathbf{u}^2\|_{k+1}^2\\
&+6\frac{\beta_{\mathrm{Lip}}^2}{\beta_{\min}^2}|\Omega|\,C_s^2\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2\Big(Ch^{2m+2}\|\eta^2\|_{m+1}^2 + K_2\Big(Ch^{2m+2}\int_{t_0}^{t_2}\|\eta_t\|_{m+1}^2\,dt\\
&+2\Delta t g_{\mathrm{Lip}}^2\|\eta^1\|_\infty^2\Big(Ch^{2k+2}\|\mathbf{u}^0\|_{k+1}^2 + 6\frac{\beta_{\mathrm{Lip}}^2}{\beta_{\min}^2}\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2\|\pi_h^{2,s}-\mathcal{S}(\pi_h^2)\|^2\\
&+(2+6\frac{\beta_{\max}^2}{\beta_{\min}^2})Ch^{2k+2}\|\mathbf{u}^2\|_{k+1}^2 + 12\frac{\beta_{\mathrm{Lip}}^2}{\beta_{\min}^2}|\Omega|\,C_s^2\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2\big(Ch^{2m+2}\|\eta^2\|_{m+1}^2\\
&+K_1\big(Ch^{2m+2}\int_{t_0}^{t_2}\|\eta_t\|_{m+1}^2 + 4C\Delta t g_{\mathrm{Lip}}^2\|\eta^2\|_\infty^2 h^{2k+2}\|\mathbf{u}^0\|_{k+1}^2\\
&+2C\Delta t\, g_{\max}^2 h^{2m+2}(\|\eta^2\|_{m+1}^2+\|\eta^0\|_{m+1}^2) + \frac{(\Delta t)^3}{6}\sup_{t_0\le t\le t_2}\|\eta_{tt}(\cdot,t)\|^2\\
&+8(\Delta t)^3 g_{\mathrm{Lip}}^2\|\eta^2\|_\infty^2\sup_{t_0\le t\le t_2}\|(\mathbf{u}+\mathbf{b})_t(\cdot,t)\|^2\big)\big) + \frac{2}{3}(\Delta t)^4\sup_{t_0\le t\le t_2}\|\mathbf{u}_{tt}(\cdot,t)\|^2\Big)\\
&+4C\Delta t g_{\max}^2 h^{2m+2}(\|\eta^2\|_{m+1}^2+\|\eta^0\|_{m+1}^2) + \frac{4}{3}g_{\max}^2(\Delta t)^5\sup_{t_0\le t\le t_2}\|\eta_{tt}(\cdot,t)\|^2\\
&+\frac{1}{40}(\Delta t)^5\sup_{t_0\le t\le t_2}\|\eta_{ttt}(\cdot,t)\|^2\Big)\Big).
\end{aligned} \tag{88}$$

Combining the bounds

$$\|\mathbf{u}^2-\mathbf{u}_h^2\|^2 \le 2(\|\mathbf{E}^2\|^2+\|\mathbf{\Lambda}^2\|^2),\ \|\eta^2-\eta_h^2\|^2 \le 2(\|F^2\|^2+\|\psi^2\|^2),$$

with (88), results in the estimates

$$\|\mathbf{u}^2-\mathbf{u}_h^2\|^2 \le (6\frac{\beta_{\max}^2}{\beta_{\min}^2}+2)Ch^{2k+2}\|\mathbf{u}^2\|_{k+1}^2 + 6\frac{\beta_{\mathrm{Lip}}^2}{\beta_{\min}^2}\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2\|\eta_h^{2,s}-\mathcal{S}(\eta_h^2)\|^2$$

$$
\begin{aligned}
&+ 12\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}\,|\Omega|\,C_s^2\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2\Big(Ch^{2m+2}\|\eta^2\|_{m+1}^2+K_2\Big(Ch^{2m+2}\int_{t_0}^{t_2}\|\eta_t\|_{m+1}^2\,dt\\
&+ 2\Delta t g_{\text{Lip}}^2\|\eta^1\|_\infty^2\Big(Ch^{2k+2}\|\mathbf{u}^0\|_{k+1}^2+6\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2\|\pi_h^{2,s}-\mathcal{S}(\pi_h^2)\|^2\\
&+ (2+6\frac{\beta_{\max}^2}{\beta_{\min}^2})Ch^{2k+2}\|\mathbf{u}^2\|_{k+1}^2+12\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}\,|\Omega|\,C_s^2\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2\big(Ch^{2m+2}\|\eta^2\|_{m+1}^2\\
&+ K_1\big(Ch^{2m+2}\int_{t_0}^{t_2}\|\eta_t\|_{m+1}^2+4C\Delta t g_{\text{Lip}}^2\|\eta^2\|_\infty^2h^{2k+2}\|\mathbf{u}^0\|_{k+1}^2\\
&+ 2C\Delta t\,g_{\max}^2h^{2m+2}(\|\eta^2\|_{m+1}^2+\|\eta^0\|_{m+1}^2)+\frac{(\Delta t)^3}{6}\sup_{t_0\le t\le t_2}\|\eta_{tt}(\cdot,t)\|^2\\
&+ 8(\Delta t)^3g_{\text{Lip}}^2\|\eta^2\|_\infty^2\sup_{t_0\le t\le t_2}\|(\mathbf{u}+\mathbf{b})_t(\cdot,t)\|^2\big)\big)+\frac{2}{3}(\Delta t)^4\sup_{t_0\le t\le t_2}\|\mathbf{u}_{tt}(\cdot,t)\|^2\Big)\\
&+ 4C\Delta t g_{\max}^2h^{2m+2}(\|\eta^2\|_{m+1}^2+\|\eta^0\|_{m+1}^2)+\frac{4}{3}g_{\max}^2(\Delta t)^5\sup_{t_0\le t\le t_2}\|\eta_{tt}(\cdot,t)\|^2\\
&+ \frac{1}{40}(\Delta t)^5\sup_{t_0\le t\le t_2}\|\eta_{ttt}(\cdot,t)\|^2\Big)\Big). \qquad (89)
\end{aligned}
$$

and

$$
\begin{aligned}
\|\eta^2-\eta_h^2\|^2 \le\ & 2K_2\Big(Ch^{2m+2}\int_{t_0}^{t_2}\|\eta_t\|_{m+1}^2\,dt+2\Delta t g_{\text{Lip}}^2\|\eta^1\|_\infty^2\Big(Ch^{2k+2}\|\mathbf{u}^0\|_{k+1}^2\\
&+ 6\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2\|\pi_h^{2,s}-\mathcal{S}(\pi_h^2)\|^2+(2+6\frac{\beta_{\max}^2}{\beta_{\min}^2})Ch^{2k+2}\|\mathbf{u}^2\|_{k+1}^2\\
&+ 12\frac{\beta_{\text{Lip}}^2}{\beta_{\min}^2}\,|\Omega|\,C_s^2\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2\big(Ch^{2m+2}\|\eta^2\|_{m+1}^2+K_1\big(Ch^{2m+2}\int_{t_0}^{t_2}\|\eta_t\|_{m+1}^2\\
&+ 4C\Delta t g_{\text{Lip}}^2\|\eta^2\|_\infty^2h^{2k+2}\|\mathbf{u}^0\|_{k+1}^2+2C\Delta t\,g_{\max}^2h^{2m+2}(\|\eta^2\|_{m+1}^2+\|\eta^0\|_{m+1}^2)\\
&+ \frac{(\Delta t)^3}{6}\sup_{t_0\le t\le t_2}\|\eta_{tt}(\cdot,t)\|^2+8(\Delta t)^3g_{\text{Lip}}^2\|\eta^2\|_\infty^2\sup_{t_0\le t\le t_2}\|(\mathbf{u}+\mathbf{b})_t(\cdot,t)\|^2\big)\big)\\
&+ \frac{2}{3}(\Delta t)^4\sup_{t_0\le t\le t_2}\|\mathbf{u}_{tt}(\cdot,t)\|^2\Big)+4C\Delta t g_{\max}^2h^{2m+2}(\|\eta^2\|_{m+1}^2+\|\eta^0\|_{m+1}^2)\\
&+ \frac{4}{3}g_{\max}^2(\Delta t)^5\sup_{t_0\le t\le t_2}\|\eta_{tt}(\cdot,t)\|^2+\frac{1}{40}(\Delta t)^5\sup_{t_0\le t\le t_2}\|\eta_{ttt}(\cdot,t)\|^2\Big)\\
&+ 2Ch^{2m+2}\|\eta^2\|_{m+1}^2. \qquad (90)
\end{aligned}
$$

Step 3

Define $\mathbf{u}_h^1 = \breve{\mathbf{u}}_h^2$, $\eta_h^1 = \breve{\eta}_h^2$ and note that

$$\begin{aligned}\|\mathbf{u}^1 - \mathbf{u}_h^1\|^2 &\le 2\|\mathbf{u}^1 - \breve{\mathbf{u}}^2\|^2 + 2\|\breve{\mathbf{u}}^2 - \breve{\mathbf{u}}_h^2\|^2 \\ &\le 2\|\mathbf{u}^1 - \breve{\mathbf{u}}^2\|^2 + \|\mathbf{u}^2 - \mathbf{u}_h^2\|^2 + \|\mathbf{u}^0 - \mathbf{u}_h^0\|^2, \\ \|\eta^1 - \eta_h^1\|^2 &\le 2\|\eta^1 - \breve{\eta}^2\|^2 + \|\eta^2 - \eta_h^2\|^2 + \|\eta^0 - \eta_h^0\|^2. \end{aligned} \tag{91}$$

In view of Lemma A.2 and the previously computed bounds (89), (90) and (91), we obtain

$$\begin{aligned}\|\mathbf{u}^1 - \mathbf{u}_h^1\|^2 &\le (6\frac{\beta_{\max}^2}{\beta_{\min}^2} + 2)Ch^{2k+2}\|\mathbf{u}^2\|_{k+1}^2 + 6\frac{\beta_{\mathrm{Lip}}^2}{\beta_{\min}^2}\|\mathbf{u}^2 + \mathbf{b}^2\|_\infty^2\|\eta_h^{2,s} - \mathcal{S}(\eta_h^2)\|^2 \\ &+ 12\frac{\beta_{\mathrm{Lip}}^2}{\beta_{\min}^2}|\Omega|\, C_s^2\|\mathbf{u}^2 + \mathbf{b}^2\|_\infty^2\Big(Ch^{2m+2}\|\eta^2\|_{m+1}^2 + K_2\Big(Ch^{2m+2}\int_{t_0}^{t_2}\|\eta_t\|_{m+1}^2\,dt \\ &+ 2\Delta t g_{\mathrm{Lip}}^2\|\eta^1\|_\infty^2\Big(Ch^{2k+2}\|\mathbf{u}^0\|_{k+1}^2 + 6\frac{\beta_{\mathrm{Lip}}^2}{\beta_{\min}^2}\|\mathbf{u}^2 + \mathbf{b}^2\|_\infty^2\|\pi_h^{2,s} - \mathcal{S}(\pi_h^2)\|^2 \\ &+ (2 + 6\frac{\beta_{\max}^2}{\beta_{\min}^2})Ch^{2k+2}\|\mathbf{u}^2\|_{k+1}^2 + 12\frac{\beta_{\mathrm{Lip}}^2}{\beta_{\min}^2}|\Omega|\, C_s^2\|\mathbf{u}^2 + \mathbf{b}^2\|_\infty^2\big(Ch^{2m+2}\|\eta^2\|_{m+1}^2 \\ &+ K_1\big(Ch^{2m+2}\int_{t_0}^{t_2}\|\eta_t\|_{m+1}^2 + 4C\Delta t g_{\mathrm{Lip}}^2\|\eta^2\|_\infty^2 h^{2k+2}\|\mathbf{u}^0\|_{k+1}^2 \\ &+ 2C\Delta t\, g_{\max}^2 h^{2m+2}(\|\eta^2\|_{m+1}^2 + \|\eta^0\|_{m+1}^2) + \frac{(\Delta t)^3}{6}\sup_{t_0\le t\le t_2}\|\eta_{tt}(\cdot,t)\|^2 \\ &+ 8(\Delta t)^3 g_{\mathrm{Lip}}^2\|\eta^2\|_\infty^2\sup_{t_0\le t\le t_2}\|(\mathbf{u}+\mathbf{b})_t(\cdot,t)\|^2\big)\big) + \frac{2}{3}(\Delta t)^4\sup_{t_0\le t\le t_2}\|\mathbf{u}_{tt}(\cdot,t)\|^2\Big) \\ &+ 4C\Delta t g_{\max}^2 h^{2m+2}(\|\eta^2\|_{m+1}^2 + \|\eta^0\|_{m+1}^2) + \frac{4}{3}g_{\max}^2(\Delta t)^5\sup_{t_0\le t\le t_2}\|\eta_{tt}(\cdot,t)\|^2 \\ &+ \frac{1}{40}(\Delta t)^5\sup_{t_0\le t\le t_2}\|\eta_{ttt}(\cdot,t)\|^2\Big)\Big) + \frac{2}{3}(\Delta t)^4\sup_{t_0\le t\le t_2}\|\mathbf{u}_{tt}(\cdot,t)\|^2 + Ch^{2k+2}\|\mathbf{u}^0\|_{k+1}^2. \end{aligned} \tag{92}$$

and

$$\begin{aligned}\|\eta^1-\eta_h^1\|^2 \le 2K_2\Big(Ch^{2m+2}\int_{t_0}^{t_2}\|\eta_t\|_{m+1}^2\,dt + 2\Delta t g_{\mathrm{Lip}}^2\|\eta^1\|_\infty^2\Big(Ch^{2k+2}\|\mathbf{u}^0\|_{k+1}^2 \\ + 6\frac{\beta_{\mathrm{Lip}}^2}{\beta_{\min}^2}\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2\|\pi_h^{2,s}-\mathcal{S}(\pi_h^2)\|^2 + (2+6\frac{\beta_{\max}^2}{\beta_{\min}^2})Ch^{2k+2}\|\mathbf{u}^2\|_{k+1}^2 \\ + 12\frac{\beta_{\mathrm{Lip}}^2}{\beta_{\min}^2}\,|\Omega|\,C_s^2\|\mathbf{u}^2+\mathbf{b}^2\|_\infty^2\big(Ch^{2m+2}\|\eta^2\|_{m+1}^2 + K_1\big(Ch^{2m+2}\int_{t_0}^{t_2}\|\eta_t\|_{m+1}^2 \\ + 4C\Delta t g_{\mathrm{Lip}}^2\|\eta^2\|_\infty^2 h^{2k+2}\|\mathbf{u}^0\|_{k+1}^2 + 2C\Delta t\, g_{\max}^2 h^{2m+2}(\|\eta^2\|_{m+1}^2+\|\eta^0\|_{m+1}^2) \\ + \frac{(\Delta t)^3}{6}\sup_{t_0\le t\le t_2}\|\eta_{tt}(\cdot,t)\|^2 + 8(\Delta t)^3 g_{\mathrm{Lip}}^2\|\eta^2\|_\infty^2\sup_{t_0\le t\le t_2}\|(\mathbf{u}+\mathbf{b})_t(\cdot,t)\|^2\big)\big) \\ + \frac{2}{3}(\Delta t)^4\sup_{t_0\le t\le t_2}\|\mathbf{u}_{tt}(\cdot,t)\|^2\Big) + 4C\Delta t g_{\max}^2 h^{2m+2}(\|\eta^2\|_{m+1}^2+\|\eta^0\|_{m+1}^2) \\ + \frac{4}{3}g_{\max}^2(\Delta t)^5\sup_{t_0\le t\le t_2}\|\eta_{tt}(\cdot,t)\|^2 + \frac{1}{40}(\Delta t)^5\sup_{t_0\le t\le t_2}\|\eta_{ttt}(\cdot,t)\|^2\Big) \\ + 2Ch^{2m+2}\|\eta^2\|_{m+1}^2 + \frac{2}{3}(\Delta t)^4\sup_{t_0\le t\le t_2}\|\eta_{tt}(\cdot,t)\|^2 + Ch^{2m+2}\|\eta^0\|_{m+1}^2. \qquad (93)\end{aligned}$$

Consequently, estimates (89), (90), (92) and (93), yield the optimal bound

$$\|\mathbf{u}^n-\mathbf{u}_h^n\|_X^2 + \|\eta^n-\eta_h^n\|^2 \le C\,(\Delta t)^4 + C\left(h^{2k+2}+h^{2m+2}\right)\text{ for } n=0,1,2.$$

□

Bibliography

[1] R.A. Adams and J.J.F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2003.

[2] P.M. Adler, J. Thovert, and V.V. Mourzenko. *Fractured porous media*. Oxford University Press, Oxford, 2013.

[3] G. Allaire. Homogenization of the Navier-Stokes equations with a slip boundary condition. *Comm. Pure Appl. Math.*, 44(6):605–641, 1991.

[4] G. Allaire. Homogenization and two-scale convergence. *SIAM J. Math. Anal.*, 23(6):1482–1518, 1992.

[5] M. Amara, D. Capatina, and L. Lizaik. Coupling of Darcy-Forchheimer and compressible Navier-Stokes equations with heat transfer. *SIAM J. Sci. Comput.*, 31(2):1470–1499, 2008/09.

[6] T. Arbogast and D.S. Brunson. A computational method for approximating a Darcy-Stokes system governing a vuggy porous medium. *Comput. Geosci.*, 11(3):207–218, 2007.

[7] C. Ayan, N. Colley, G. Cowan, E. Ezekwe, M. Wannell, P. Goode, F. Halford, J. Joseph, A. Mongini, and J. Pop. Measuring permeability anisotropy: The latest approach. *Oilfield Review*, 6(4):24–35, 1994.

[8] M. Azaïez, F. Ben Belgacem, C. Bernardi, and N. Chorfi. Spectral discretization of Darcy's equations with pressure dependent porosity. *Appl. Math. Comput.*, 217(5):1838–1856, 2010.

[9] L. Badea, M. Discacciati, and A. Quarteroni. Numerical analysis of the Navier-Stokes/Darcy coupling. *Numer. Math.*, 115(2):195–227, 2010.

[10] W. Bangerth, T. Heister, L. Heltai, G. Kanschat, M. Kronbichler, M. Maier, B. Turcksin, and T.D. Young. The `deal.II` library, version 8.2. *Archive of Numerical Software*, 3, 2015.

[11] V. Barbu, G. Da Prato, and M. Röckner. *Stochastic porous media equations*, volume 2163 of *Lecture Notes in Mathematics*. Springer, [Cham], 2016.

[12] S. Bartels, M. Jensen, and R. Müller. Discontinuous Galerkin finite element convergence for incompressible miscible displacement problems of low regularity. *SIAM J. Numer. Anal.*, 47(5):3720–3743, 2009.

[13] R.G. Bartle. *The elements of integration*. John Wiley & Sons, Inc., New York-London-Sydney, 1966.

[14] G.K. Batchelor. *An introduction to fluid dynamics*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, paperback edition, 1999.

[15] J. Bear. *Dynamics of fluids in porous media*. Environmental Science series (New York). Elsevier, 1972.

[16] G.S. Beavers and D.D. Joseph. Boundary conditions at a naturally permeable wall. *Journal of Fluid Mechanics*, 30(1):197–207, 1967.

[17] R.B. Bird, W.E. Stewart, and E.N. Lightfoot. *Transport phenomena*. John Wiley & Sons, 2007.

[18] D. Boffi, F. Brezzi, and M. Fortin. *Mixed finite element methods and applications*, volume 44 of *Springer Series in Computational Mathematics*. Springer, Heidelberg, 2013.

[19] D. Braess. *Finite elements*. Cambridge University Press, Cambridge, third edition, 2007. Theory, fast solvers, and applications in elasticity theory, Translated from the German by L.L. Schumaker.

[20] S.C. Brenner. Korn's inequalities for piecewise H^1 vector fields. *Math. Comp.*, 73(247):1067–1087, 2004.

[21] Y. Cao, M. Gunzburger, X. He, and X. Wang. Parallel, non-iterative, multi-physics domain decomposition methods for time-dependent Stokes-Darcy systems. *Math. Comp.*, 83(288):1617–1644, 2014.

[22] Y. Cao, M. Gunzburger, X. Hu, F. Hua, X. Wang, and W. Zhao. Finite element approximations for Stokes-Darcy flow with Beavers-Joseph interface conditions. *SIAM J. Numer. Anal.*, 47(6):4239–4256, 2010.

[23] Y. Cao, M. Gunzburger, F. Hua, and X. Wang. Coupled Stokes-Darcy model with Beavers-Joseph interface boundary condition. *Commun. Math. Sci.*, 8(1):1–25, 2010.

[24] A. Çeşmelioğlu and B. Rivière. Analysis of time-dependent Navier-Stokes flow coupled with Darcy flow. *J. Numer. Math.*, 16(4):249–280, 2008.

[25] W. Chen, P. Chen, M. Gunzburger, and N. Yan. Superconvergence analysis of FEMs for the Stokes-Darcy system. *Math. Methods Appl. Sci.*, 33(13):1605–1617, 2010.

[26] W. Chen, M. Gunzburger, D. Sun, and X. Wang. Efficient and long-time accurate second-order methods for the Stokes-Darcy system. *SIAM J. Numer. Anal.*, 51(5):2563–2584, 2013.

[27] Z. Chen and R. Ewing. Mathematical analysis for reservoir models. *SIAM J. Math. Anal.*, 30(2):431–453, 1999.

[28] M. Cui and N. Yan. A posteriori error estimate for the Stokes-Darcy system. *Math. Methods Appl. Sci.*, 34(9):1050–1064, 2011.

[29] G. de Marsily. *Quantitative Hydrogeology: Groundwater Hydrology for Engineers*. Academic Press, 1986.

[30] Z. Ding. A proof of the trace theorem of Sobolev spaces on Lipschitz domains. *Proc. Amer. Math. Soc.*, 124(2):591–600, 1996.

[31] M. Discacciati, E. Miglio, and A. Quarteroni. Mathematical and numerical models for coupling surface and groundwater flows. *Appl. Numer. Math.*, 43(1-2):57–74, 2002.

[32] M.G. Edwards, R.D. Lazarov, and I. Yotov, editors. *Locally conservative numerical methods for flow in porous media*. Springer, Dordrecht, 2002. Comput. Geosci. **6** (2002), no. 3-4.

[33] W. Ehlers and J. Bluhm, editors. *Porous media*. Springer-Verlag, Berlin, 2002. Theory, experiments and numerical applications.

[34] V.J. Ervin and N. Heuer. Approximation of time-dependent, viscoelastic fluid flow: Crank-Nicolson, finite element approximation. *Numer. Methods Partial Differential Equations*, 20(2):248–283, 2004.

[35] V.J. Ervin, E.W. Jenkins, and H. Lee. Approximation of the Stokes-Darcy system by optimization. *J. Sci. Comput.*, 59(3):775–794, 2014.

[36] V.J. Ervin, E.W. Jenkins, and S. Sun. Coupled generalized nonlinear Stokes flow with flow through a porous medium. *SIAM J. Numer. Anal.*, 47(2):929–952, 2009.

[37] V.J. Ervin, E.W. Jenkins, and S. Sun. Coupling nonlinear Stokes and Darcy flow using mortar finite elements. *Appl. Numer. Math.*, 61(11):1198–1222, 2011.

[38] V.J. Ervin, H. Lee, and A.J. Salgado. Generalized Newtonian fluid flow through a porous medium. *Journal of Mathematical Analysis and Applications*, 433(1):603 – 621, 2016.

[39] A. Fasano, M. Primicerio, and G. Baldini. Mathematical models for espresso coffee preparation. In F. Hodnett, editor, *Proceedings of the Sixth European Conference on Mathematics in Industry August 27–31, 1991 Limerick*, pages 137–140. Vieweg+Teubner Verlag, Wiesbaden, 1992.

[40] M. Feng, R. Qi, R. Zhu, and B. Ju. Stabilized Crouzeix-Raviart element for the coupled Stokes and Darcy problem. *Appl. Math. Mech. (English Ed.)*, 31(3):393–404, 2010.

[41] X. Feng. On existence and uniqueness results for a coupled system modeling miscible displacement in porous media. *J. Math. Anal. Appl.*, 194(3):883–910, 1995.

[42] K.R. Fowler, E.W. Jenkins, C.L. Cox, B. McClune, and B. Seyfzadeh. Design analysis of polymer filtration using a multiobjective genetic algorithm. *Separation Science and Technology*, 43(4):710–726, 2008.

[43] K.R. Fowler, E.W. Jenkins, and S.M. LaLonde. Understanding the effects of polymer extrusion filter layering configurations using simulation-based optimization. *Optimization and Engineering*, 11(2):339–354, Jun 2010.

[44] G.P. Galdi. *An introduction to the mathematical theory of the Navier-Stokes equations. Vol. I*, volume 38 of *Springer Tracts in Natural Philosophy*. Springer-Verlag, New York, 1994. Linearized steady problems.

[45] J. Galvis and M. Sarkis. Non-matching mortar discretization analysis for the coupling Stokes-Darcy equations. *Electron. Trans. Numer. Anal.*, 26:350–384, 2007.

[46] V. Girault, F. Murat, and A. Salgado. Finite element discretization of Darcy's equations with pressure dependent porosity. *M2AN Math. Model. Numer. Anal.*, 44(6):1155–1191, 2010.

[47] V. Girault and M. F. Wheeler. Numerical discretization of a Darcy-Forchheimer model. *Numer. Math.*, 110(2):161–198, 2008.

[48] C.W. Groetsch. *Generalized inverses of linear operators: representation and approximation*. Marcel Dekker, Inc., New York-Basel, 1977. Monographs and Textbooks in Pure and Applied Mathematics, No. 37.

[49] J.K. Hale. *Ordinary differential equations*. Robert E. Krieger Publishing Co., Inc., Huntington, N.Y., second edition, 1980.

[50] J.G. Heywood and R. Rannacher. Finite-element approximation of the nonstationary Navier-Stokes problem. IV. Error analysis for second-order time discretization. *SIAM J. Numer. Anal.*, 27(2):353–384, 1990.

[51] J. Huang, N. Gretz, and S. Weinfurter. Filtration markers and determination methods for the assessment of kidney function. *European Journal of Pharmacology*, 790:92 – 98, 2016. The pharmacology of kidney regeneration.

[52] A. Hunt, R. Ewing, and B. Ghanbarian. *Percolation theory for flow in porous media*, volume 880 of *Lecture Notes in Physics*. Springer, Cham, third edition, 2014.

[53] D.B. Ingham and I. Pop, editors. *Transport phenomena in porous media. II*. Pergamon Press, Oxford, 2002.

[54] W. Jäger and A. Mikelić. On the interface boundary condition of Beavers, Joseph, and Saffman. *SIAM J. Appl. Math.*, 60(4):1111–1127, 2000.

[55] J.D. Jansen. *A systems description of flow through porous media*. SpringerBriefs in Earth Sciences. Springer, Cham, 2013.

[56] G. Kanschat and B. Rivière. A strongly conservative finite element method for the coupling of Stokes and Darcy flow. *J. Comput. Phys.*, 229(17):5933–5943, 2010.

[57] T. Karper, K. Mardal, and R. Winther. Unified finite element discretizations of coupled Darcy-Stokes flow. *Numer. Methods Partial Differential Equations*, 25(2):311–326, 2009.

[58] P.R. Kvietys and D.N. Granger. Role of intestinal lymphatics in interstitial volume regulation and transmucosal water transport. *Annals of the New York Academy of Sciences*, 1207:E29–E43, 2010.

[59] S. Kwon, S. Lew, and R.S. Chamberlain. Leukocyte filtration and postoperative infections. *Journal of Surgical Research*, 205(2):499 – 509, 2016.

[60] W. Layton. *Introduction to the numerical analysis of incompressible viscous flows*, volume 6 of *Computational Science & Engineering*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008.

[61] W. Layton, H. Tran, and C. Trenchea. Analysis of long time stability and errors of two partitioned methods for uncoupling evolutionary groundwater-surface water flows. *SIAM J. Numer. Anal.*, 51(1):248–272, 2013.

[62] W. Layton, H. Tran, and X. Xiong. Long time stability of four methods for splitting the evolutionary Stokes-Darcy problem into Stokes and Darcy subproblems. *J. Comput. Appl. Math.*, 236(13):3198–3217, 2012.

[63] W.J. Layton, F. Schieweck, and I. Yotov. Coupling fluid flow with porous media flow. *SIAM J. Numer. Anal.*, 40(6):2195–2218 (2003), 2002.

[64] D.A. Nield. The Beavers-Joseph boundary condition and related matters: A historical and critical note. *Transport in Porous Media*, 78(3):537, 2009.

[65] K.R. Rajagopal. On a hierarchy of approximate models for flows of incompressible fluids through porous solids. *Math. Models Methods Appl. Sci.*, 17(2):215–252, 2007.

[66] B.H. Rivera and M.G. Rodriguez. Characterization of Airborne Particles Collected from Car Engine Air Filters Using SEM and EDX Techniques. *International Journal of Environmental Research and Public Health*, 13(10):1–16, 10 2016.

[67] B.M. Rivière and N.J. Walkington. Convergence of a discontinuous Galerkin method for the miscible displacement equation under low regularity. *SIAM J. Numer. Anal.*, 49(3):1085–1110, 2011.

[68] H. Rui and R. Zhang. A unified stabilized mixed finite element method for coupling Stokes and Darcy flows. *Comput. Methods Appl. Mech. Engrg.*, 198(33-36):2692–2699, 2009.

[69] P.G. Saffman. On the boundary condition at the surface of a porous medium. *Studies in Applied Mathematics*, 50(2):93–101, 1971.

[70] W.T. Sha. *Novel porous media formulation for multiphase flow conservation equations*. Cambridge University Press, Cambridge, 2011.

[71] L.J. Thibodeaux. *Environmental chemodynamics: Movement of chemicals in air, water, and soil*, volume 110. John Wiley & Sons, 1996.

[72] L. Traverso, T.N. Phillips, and Y. Yang. Mixed finite element methods for groundwater flow in heterogeneous aquifers. *Comput. & Fluids*, 88:60–80, 2013.

[73] J. M. Urquiza, D. N'Dri, A. Garon, and M.C. Delfour. Coupling Stokes and Darcy equations. *Appl. Numer. Math.*, 58(5):525–538, 2008.

[74] S. Whitaker. Flow in porous media I. A theoretical derivation of Darcy's law. *Transp. Porous Media*, 1:3–25, 1986.

[75] P. Xu and B. Yu. Developing a new form of permeability and Kozeny-Carman constant for homogeneous porous media by means of fractal geometry. *Advances in Water Resources*, 31(1):74 – 81, 2008.

[76] V.A. Yangali-Quintanilla, A.H. Holm, M. Birkner, S. D'Antonio, H.W. Stoltze, M. Ulbricht, and X. Zheng. A fast and reliable approach to benchmark low pressure hollow fibre filtration membranes for water purification. *Journal of Membrane Science*, 499:515 – 525, 2016.

[77] W. Zhang, K.E. Thompson, A.H. Reed, and L. Beenken. Relationship between packing structure and porosity in fixed beds of equilateral cylindrical particles. *Chemical Engineering Science*, 61(24):8060 – 8074, 2006.