Stabilized Finite Elements for Resource Modeling and Optimization

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Simon Taylor

Kiel, 2025

Erster Gutachter: Prof. Dr. Malte Braack

Zweiter Gutachter: Prof. Dr. Thomas Slawig

Tag der mündlichen Prüfung: 3.6.2025

Preface

This thesis is concerned with the application of systems of partial differential equations (PDEs) to investigate various real-world problems encountered in the fields of geophysics, fish biomass sustainability, and suspension-sedimentation processes. The common factor to these disparate fields are the components of the mathematical models used to describe them: they each employ a specific composite of numerical approximation by finite elements, fluid dynamics, and the convection of some kind of material to describe the progress of a physical process through time.

- In geodynamics one is interested in the flow and interaction between the molten magma and solid rock in the Earth's subsurface layers. Of interest to us was improving on recent advances in models of the long-term behaviour of the lithosphere coupled to the two-phase melt/matrix interaction.

- The viability of ever increasing fish harvesting relies heavily on population modeling that attempts to predict future availability based on common assumptions and simplifications. In this project we developed a fisheries optimization model that for the first time investigated a spatially distributed domain with real parameters and a competing species interaction.

- In the sedimentation mechanism one has to deal with a highly sensitive nonlinear coupling between systems of PDEs that depict how particles in a suspension are convected in a vector field, and settle due to gravity. Our contribution to this field

was the first application of a recent stabilization technique to investigate sediment accumulation in convective dominant flows.

In this context, we consider the two fundamental principles available in computational fluid dynamics for the simulation of a physical quantity's movement: convection under directional forces, such as gravity or ocean currents, and diffusion. Additionally, in the sustainability model the concept of population growth through the application of the Logistic model is examined in conjunction with biomass movement.
The traditional tools available for the simulation of a directional flow involve solving the general Navier-Stokes equations, a complete system of PDEs that characterize the motion of viscous fluids which must be solved simultaneously for both the fluid velocity and pressure. In certain flow regimes, such as slow moving fluids, or highly viscous flows, the Stokes equations are more applicable since they disregard the negligible inertial forces that describe the fluid's acceleration. This specialized simplification is valid when the physical characteristics of the problem exclude the possibility of high velocities and turbulent flow. This is the physical environment under which all the problems considered in this thesis are solved. The second form of movement commonly encountered in natural systems is that of diffusive motion, generally described by a single scalar PDE via the second-order Laplacian operator. Combined together with the flow equations, one arrives at a system that describes how a quantity of particles disperses locally as it is carried in a fluid's stream. Such models can rarely, if ever, be solved analytically and to that end we employ the finite-element method for efficient numerical approximations together with multiple stabilization techniques to arrive at robust and reproducible results.

vi

Einleitung

Diese Arbeit befasst sich mit der Anwendung von Systemen partieller Differentialgleichungen (PDEs), um verschiedene reale Probleme zu untersuchen, die in den Bereichen Geophysik, Nachhaltigkeit der Fischbiomasse und Suspensionssedimentationsprozesse auftreten. Der gemeinsame Faktor dieser unterschiedlichen Bereiche sind die Komponenten der mathematischen Modelle, die zu ihrer Beschreibung verwendet werden: Sie verwenden jeweils eine spezifische Kombination aus numerischer Annäherung durch finite Elemente, Fluiddynamik und Konvektion einer Art von Material, um den Verlauf eines physikalischen Prozesses über die Zeit zu beschreiben.

- In der Geodynamik interessiert man sich für den Fluss und die Wechselwirkung zwischen geschmolzenem Magma und festem Gestein in den unterirdischen Schichten der Erde. Unser Ziel war es, aktuelle Fortschritte in Modellen des langfristigen Verhaltens der Lithosphäre, gekoppelt an die Zwei-Phasen-Schmelze/Matrix-Wechselwirkung, zu verbessern.

- Die Durchführbarkeit einer stetig wachsenden Fischernte hängt stark von der Populationsmodellierung ab, die versucht, die zukünftige Verfügbarkeit auf der Grundlage gängiger Annahmen und Vereinfachungen vorherzusagen. In diesem Projekt haben wir ein Fischereioptimierungsmodell entwickelt, das erstmals einen räumlich verteilten Bereich mit realen Parametern und einer Interaktion konkurrierender Arten untersucht.

- Im Sedimentationsmechanismus hat man es mit einer hochsensiblen, nichtlinearen

Kopplung zwischen Systemen von partiellen Differentialgleichungen zu tun, die beschreiben, wie Partikel in einer Suspension in einem Vektorfeld konvektiert werden und aufgrund der Schwerkraft sedimentieren. Unser Beitrag zu diesem Feld war die erstmalige Anwendung einer neuen Stabilisierungstechnik zur Untersuchung der Sedimentansammlung in konvektionsdominierten Strömungen.

In diesem Zusammenhang betrachten wir die beiden grundlegenden Prinzipien, die in der numerischen Strömungsmechanik für die Simulation der Bewegung einer physikalischen Größe zur Verfügung stehen: Konvektion unter gerichteten Kräften, wie Schwerkraft oder Meeresströmungen, und Diffusion. Zusätzlich wird im Nachhaltigkeitsmodell das Konzept des Populationswachstums durch die Anwendung des logistischen Modells in Verbindung mit der Biomassebewegung untersucht.
Die traditionellen Werkzeuge zur Simulation einer gerichteten Strömung umfassen die Lösung der allgemeinen Navier-Stokes-Gleichungen, eines vollständigen Systems von partiellen Differentialgleichungen, das die Bewegung viskoser Flüssigkeiten charakterisiert und gleichzeitig für sowohl die Flüssigkeitsgeschwindigkeit als auch den Druck gelöst werden muss.
In bestimmten Strömungsregimen, wie langsam fließenden Flüssigkeiten oder hochviskosen Strömungen, sind die Stokes-Gleichungen anwendbarer, da sie die vernachlässigbaren Trägheitskräfte, die die Beschleunigung der Flüssigkeit beschreiben, außer Acht lassen. Diese spezielle Vereinfachung ist gültig, wenn die physikalischen Eigenschaften des Problems die Möglichkeit hoher Geschwindigkeiten und turbulenter Strömungen ausschließen.

Dies ist das physikalische Umfeld, in dem alle in dieser Arbeit betrachteten Probleme gelöst werden. Die zweite Form der Bewegung, die in natürlichen Systemen häufig vorkommt, ist die Diffusionsbewegung, die im Allgemeinen durch eine einzelne skalare PDE über den Laplacian-Operator zweiter Ordnung beschrieben wird. Kombiniert mit den Strömungsgleichungen erhält man ein System, das beschreibt, wie sich eine Menge von Teilchen lokal ausbreitet, während sie in einer Flüssigkeitsströmung mitgeführt wird. Solche Modelle können, wenn überhaupt, nur selten analytisch gelöst werden. Zu diesem Zweck verwenden wir die Finite-Elemente-Methode für effiziente numerische Näherungen zusammen mit mehreren Stabilisierungstechniken, um robuste und reproduzierbare Ergebnisse zu erzielen.

Contents

Introduction 1

The objective of these accumulated works consist of the design, solution and analysis of models based on various forms of incompressible flow problems, variously combined with other transport problems. The two fundamental principles which are used throughout the works presented here are: the fluid flow equations of the Stokes system, and the transport of material due the combined effects of convection and diffusion.

Stokes flow

The motion of an incompressible Newtonian fluid is described by the continuous time-dependent incompressible Stokes equations. Consider a bounded domain $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$ and a final time $T > 0$. In their simplest form and with constant density the incompressible Stokes equations read,

$$\partial_t \mathbf{u} - \nu \Delta \mathbf{u} + \nabla p = \mathbf{f} \qquad \text{in } \Omega \times (0, T], \tag{1.1}$$

$$\operatorname{div} \mathbf{u} = 0 \qquad \text{in } \Omega \times (0, T], \tag{1.2}$$

where $\mathbf{u}$ and p denote the fluid velocity and pressure, respectively. The body forces are denoted by $\mathbf{f}$ and ν is a kinematic viscosity. For a unique solution the equations (1.1)-(1.2) have to be equipped with appropriate boundary conditions and an initial condition for the velocity. In incompressible flow problems the pressure field is not obtained from an equation of state, but it establishes itself instantaneously and acts as a Lagrange multiplier. Therefore the pressure must be integrated in time implicitly (Löhner, 2008). While approximating the solution to equations (1.1)-(1.2) with a finite difference or finite element discretization one has to cope with the saddle point nature of the equation system, which

arises from the velocity-pressure coupling. When implicit time integration is considered, the spatial discretization of the problem at a certain time gives rise to a saddle point system of size $(\mathcal{N}_u+\mathcal{N}_p)\times(\mathcal{N}_u+\mathcal{N}_p)$, where $\mathcal{N}_u$ and $\mathcal{N}_p$ are the degrees of freedom for the discrete velocities and pressure, respectively. For an overview of direct and iterative methods for solving a saddle point system we refer to (Benzi et al., 2005) and the references therein.

Darcy Flow

An interesting derivative of low Reynolds number (viscous) flow equations deals with time changing phase transitions of the flowing medium. One such application is the two-phase flow that models a liquid flowing through a porous medium. Darcy flow defines this flow regime where the velocity $\mathbf{u}(t)$ of the fluid moving through the porous medium is proportional to the pressure gradient, and inversely proportional to the viscosity of the fluid. The time-dependent governing equations by (McKenzie, 1984) which are used to model magma dynamics by the flow of melted mantle fluid through a porous solid *matrix* are,

$$\frac{\partial \rho_f \phi}{\partial t} + \nabla \cdot (\rho_f \phi \mathbf{u}_f) = \Gamma \tag{1.3}$$

$$\frac{\partial \rho_s (1-\phi)}{\partial t} + \nabla \cdot [\rho_s (1-\phi)\mathbf{u}_s] = -\Gamma \tag{1.4}$$

$$\phi(\mathbf{u}_f - \mathbf{u}_s) = -\frac{K}{\mu}(\nabla p - \rho_f \mathbf{g}) \tag{1.5}$$

$$\nabla p = \nabla \cdot \left[\eta(\nabla \mathbf{u}_s + \nabla \mathbf{u}_s^T)\right] + \nabla \left[\left(\zeta - \frac{2}{3}\eta\right)\nabla \cdot \mathbf{u}_s\right] + \bar{\rho}\mathbf{g} \tag{1.6}$$

where ϕ is the matrix porosity, ρ_f and ρ_s are the fluid and solid densities, p is the fluid pressure, $\mathbf{u}_f$ and $\mathbf{u}_s$ are the fluid and solid velocity fields, Γ is the melting rate, i.e. the rate of mass transfer from solid to fluid, K is the permeability, μ is the melt viscosity, and $\mathbf{g}$ is the acceleration due to gravity. In this formulation mass is conserved for the fluid and solid across equations (1.3) and (1.4) which allows for mass transfer between the phases via Γ. An extended version of Darcy's law governing the separation of melt from solid is described by equation (1.5). Momentum conservation of the solid phase which is modeled as a compressible, inertia-free viscous fluid, is shown by (1.6), where η and ζ are the shear and bulk viscosities of the solid and $\bar{\rho} = \rho_f\phi + \rho_s(1-\phi)$ is the two-phase mean density of the system. These equations taken together consistently couple the viscous mantle deformation via Stokes flow, with Darcy's law for the fluid. Since the solid stresses and

fluid pressure are coupled, the fluid pressure reacts to solid deformation and gravitational forces which drives fluid flow and varies the porosity. These variations can then feed back through the connections of the permeability and viscosity, which can lead to phenomena such as localized flow and the emergence of nonlinear magma waves.

Convection-diffusion-reaction

The convection-diffusion equation describes the movement of particles or energy $u(t)$ via the processes of convection due to an applied velocity field, such as a Stokes velocity $\mathbf{b}(t)$, and diffusion, the general stochastic movement of any particles by a gradient in Gibbs free energy. The time-dependent version is,

$$\partial_t u - \varepsilon \Delta u + (\mathbf{b} \cdot \nabla) u + \sigma u = f \quad \text{in } \Omega \tag{1.7}$$

where u is the convected and diffused quantity and σ is a reaction field. Here, ε is a positive diffusion coefficient which can also be non-constant in some applications, such as those presented in chapters 3 and 5. Furthermore, to obtain a well-posed problem (1.7) has to be equipped with appropriate boundary conditions on the boundary $\partial\Omega$ of Ω. The right hand side f is known as a source or sink term. Similar to the Stokes system, instabilities can arise when the diffusion constant is much smaller than the spatial mesh width, i.e. $\varepsilon \ll h$, the Galerkin discretization of (1.7) suffers from spurious oscillations (Johnson et al., 1984; Hughes and Brooks, 1979). In chapter 2, we introduce a theoretical framework for stabilised Galerkin methods.

Boundary conditions

To complete the definitions of any of the above models one needs to explicitly state the applicable boundary conditions $\partial\Omega$ of Ω that are relevant to the problem under consideration. The three types that describe what most physical properties that may be encountered on the perimeter of a simulated domain are,

- The Dirichlet condition: a fixed type where the value of the sought quantity is set to a constant or function value on a part of the boundary $\partial\Omega_D \subset \Omega$,

$$u = g_1 \quad \text{on } \partial\Omega_D$$

  In the case of a flow problem with a *no-slip* condition, one commonly sets $g_1 = 0$ on

$\partial\Omega_D$.

- Neumann condition: a flux type of condition where one does not want to explicitly set an exact value on the boundary, but rather the rate at which the quantity may pass through. The flux is prescribed to a function value g_2 at the desired section of the boundary $\partial\Omega_N \subset \Omega$ by,

$$\frac{\partial u}{\partial \mathbf{n}} = g_2 \quad \text{on } \partial\Omega_N$$

  where $\mathbf{n}$ is the outward facing unit normal vector. In the case of a convection-diffusion problem with a *no-flux* condition, one does not want the material u to escape the domain but knowing the solution along the boundary is desired. One therefore sets the Neumann condition to zero via $g_2 = 0$ on $\partial\Omega_N$ and leaves the Dirichlet condition unprescribed.

- Robin condition: a weighted combination of the Dirichlet and Neumann types

$$au + b\frac{\partial u}{\partial \mathbf{n}} = g_3 \quad \text{on } \partial\Omega_R$$

  with non-zero weights a and b which may also be non-constant functions. Common applications are in heat transfer problems as convective boundary conditions, and electromagnetic problems, as impedance boundary conditions.

Discretization of flow and transport problems

2

In this section we present spatio-temporal discretizations for flow and transport problems which form a starting point for the techniques utilized in Chapters 3, 4 and 5. The continuous flow and convection equations just introduced are in general not solvable analytically since the solutions are sought in infinite dimensional spaces. By applying the Galerkin method one can approximate the continuous function spaces by finite dimensional subspaces, of which one such choice is given by the finite element method (FEM), a method well suited to discretizing PDEs (Ern and Guermond, 2013). These subspaces are realized by decomposing the bounded domain Ω into a finite set of cells K of size h_K. Similarly, the complete time interval I under consideration is also split into a finite number of (usually equal sized) subintervals, further explained in section 2.1.2. Then the continuous solution can be approximated by piecewise polynomial trial functions defined on each partition. We will start by introducing stable space discretizations for the time-dependent Stokes equations. Following that, we will turn our attention to the time-dependent convection-diffusion-reaction equation in order to discuss the instabilities caused by convection dominance, and the remedies that may be applied.

2.1 Preliminaries

If there exists a positive constant C which is independent of viscosity, diffusivity constant and mesh size such that $\alpha \leq C\beta$, then we will write $\alpha \lesssim \beta$.

Remark 2.1 (Notation)**.** *We assume that $\Omega \in \mathbb{R}^d$, $d = 2, 3$, is a polygonal or polyhedral domain with boundary Γ. For a subdomain ω of Ω, we consider the usual Sobolev spaces*

$W^{m,p}(\omega)$ with norm $\|\cdot\|_{m,p}$ and semi-norm $|\cdot|_{m,p}$. In the case $p = 2$, we set $H^m(\omega) = W^{m,2}(\omega)$. The L^2 inner product on ω is denoted by $(\cdot,\cdot)_\omega$. For $\omega = \Omega$ we omit the index ω. We further define

$$\begin{aligned} H_0^1(\Omega) &:= \{v \in L^2(\Omega) \ : \ v|_\Gamma = 0\}, \\ L_0^2(\Omega) &:= \{q \in L^2(\Omega) \ : \ (q,1) = 0\} \end{aligned}$$

and consider the function spaces

$$\mathbf{V} := [H_0^1(\Omega)]^d, \ Q := L_0^2(\Omega) \text{ and } S := H^1(\Omega).$$

Remark 2.2. *The ansatz spaces $\mathbf{V}$ and Q satisfy the following inf-sup condition:*

$$\exists \beta > 0 : \inf_{q \in Q\setminus\{0\}} \sup_{\mathbf{v} \in \mathbf{V}\setminus\{\mathbf{0}\}} \frac{(q, div\,\mathbf{v})}{\|q\|_0 |\mathbf{v}|_1} \geq \beta. \tag{2.1}$$

The compatibility condition (2.1) between the velocity space $\mathbf{V}$ and the pressure space Q is an important tool for the well-posedness of the solutions to the Stokes and Navier-Stokes equations. For details refer to (John, 2016).

2.1.1 Finite element spaces

Let $\mathcal{T}_h(\Omega)$ be a shape-regular, admissible decomposition of Ω into d-simplices or quadrilaterals/hexahedra.

Simplices: Let $\hat{K}$ be the unit $d-$simplex and $T_K : \hat{K} \to K$ be the affine mapping from reference $d-$simplex onto an arbitrary cell $K \in \mathcal{T}_h(\Omega)$. The space of all polynomials of degree k on $\hat{K}$ is denoted with $\mathbb{P}_k(\hat{K})$. The finite element space on simplices is defined as

$$\mathcal{P}_{k,h} := \{\phi \in C(\bar{\Omega}) : \phi|_K \circ T_K \in \mathbb{P}_k(\hat{K}), \ K \in \mathcal{T}_h(\Omega)\}.$$

Quadrilaterals and hexahedra: Let $\hat{K}$ be the unit hypercube and $T_K : \hat{K} \to K$ be the multilinear mapping from reference hypercube onto an arbitrary cell $K \in \mathcal{T}_h(\Omega)$. The space of polynomials on $\hat{K}$ of maximal degree k in each coordinate direction is denoted

with $\mathbb{Q}_k(\hat{K})$. The finite element space on quadrilaterals/hexahedra is defined as

$$\mathcal{Q}_{k,h} := \{\phi \in C(\bar{\Omega}) : \phi|_K \circ T_K \in \mathbb{Q}_k(\hat{K}),\ K \in \mathcal{T}_h(\Omega)\}.$$

For the well-posedness of the discrete system, it is required that the discrete counterpart of the compatibility condition (2.1) be valid, i.e.

$$\exists \beta_h > 0 : \inf_{q \in Q_h \setminus \{0\}} \sup_{\mathbf{v} \in \mathbf{V}_h \setminus \{\mathbf{0}\}} \frac{(q, \operatorname{div} \mathbf{v})}{\|q\|_0 |\mathbf{v}|_1} \geq \beta_h. \tag{2.2}$$

The equal-order pairs $Q_k^d \times Q_k$ used in this work do not fulfill the discrete inf-sup condition (2.2). An example of one possible choice of finite element space fulfilling (2.2) is the so called Taylor-Hood pair:

$$\begin{aligned} \mathbf{V}_h \times Q_h :=& ([\mathcal{P}_{k+1,h}]^d \cap \mathbf{V}) \times (\mathcal{P}_{k,h} \cap \mathbf{Q}), \\ \text{or } \mathbf{V}_h \times Q_h :=& ([\mathcal{Q}_{k+1,h}]^d \cap \mathbf{V}) \times (\mathcal{Q}_{k,h} \cap \mathbf{Q}) \end{aligned}$$

with $k \in \mathbb{N}$.

2.1.2 Implicit and explicit temporal discretization

In this section we present the two time stepping methods applied for discretizing our initial value problems in the time domain. We denote the time at the n^{th} time-step by t_n and the computed solution at the n^{th} step by $u_n := u(t_n)$. For ease of implementation we always use a constant step size $k := t_n - t_{n-1} > 0$. The two schemes of interest to us are:

(a) The first-order backwards Euler (BE) method which reads,

$$u_{n+1} = u_n + k f(u_{n+1}, t_{n+1})$$

Note that the right hand side term $f(u_{n+1}, t_{n+1})$ is not known and the method is implicit in u_{n+1}. In the case that f is nonlinear, one has to solve this by finding the roots of the implicit equation using, for example, the Newton method to minimize the residual at every iteration n. This is computationally expensive in relation to the simpler forward Euler scheme but the scheme offers unconditional stability, also known as *A-stability*.

(b) Heun's second-order explicit method, which is a special case from the class of two-stage, or predictor-corrector, Runge-Kutta methods. The scheme is as follows,

$$\tilde{u}_{n+1} = u_n + kf(u_n, t_n)$$
$$u_{n+1} = u_n + \frac{k}{2}\left[f(u_n, t_n) + f(\tilde{u}_{n+1}, t_{n+1})\right]$$

This is a so-called predictor-corrector method as the first step makes a *prediction* $\tilde{u}_{n+1}$ of the solution at the next time point using an explicit scheme and, with this information, the second stage *corrects* and updates the computation giving the updated solution at u_{n+1}. This is a conditionally stable method due to the explicit calculation at each stage, i.e. the right hand sides contain previously calculated and known quantities.

Remark 2.3. *For stability reasons explicit time-marching methods are subject to the necessary Courant–Friedrichs–Lewy (CFL) condition when solving PDEs numerically. The consequence of this is that timestep k has a fixed upper bound dependent on the spatial discretization h of the form $k \leq h$.*

A generalized form of implicit methods such as the BE method above is the θ-scheme, which allows one to change a parameter $\theta \in (0, 1]$ to select the type of discretization scheme according to,

$$\frac{1}{\theta k}u_n + a(u_n) = f(t_n) + \frac{1-\theta}{\theta}f(t_{n-1}) + \frac{1}{\theta k}u_{n-1} - \frac{1-\theta}{\theta}a(u_{n-1})$$

where k is the time step size as before and $a(u_n)$ is the bilinear left hand side. Selecting $\theta = 1/2$ configures the time discretizer as the (not strongly A-stable) Crank–Nicolson method, and choosing $\theta = 1$ gives the BE-scheme above.

2.2 The time-dependent Stokes equations

The time-dependent Stokes equations read: find the velocity $\mathbf{u} : \Omega \to \mathbb{R}^d$ and the pressure $p : \Omega \to \mathbb{R}$ such that:

$$\partial_t \mathbf{u} - \nu \Delta \mathbf{u} + \nabla p = \mathbf{f} \qquad \text{in } \Omega_T := \Omega \times (0, T], \tag{2.3}$$

$$\operatorname{div} \mathbf{u} = 0 \qquad \text{in } \Omega_T, \tag{2.4}$$

$$\mathbf{u} = \mathbf{0} \qquad \text{on } \Gamma \times (0, T], \tag{2.5}$$

$$\mathbf{u}(0) = \mathbf{u}_0 \qquad \text{in } \Omega, \tag{2.6}$$

where $\mathbf{f} : \Omega \to \mathbb{R}^d$ is a forcing term, $\mathbf{u}_0$ is an initial velocity field and $\nu > 0$ is a positive constant viscosity. For simplicity we consider homogeneous Dirichlet conditions on Γ. We denote the outer normal on the boundary Γ by $\mathbf{n}$.

The θ-scheme discretization of the equations (2.3)-(2.6) in time reads with $\theta = 1$: given $\mathbf{u}_n$ solve for $\mathbf{u} = \mathbf{u}_{n+1}$ and $p = p_{n+1}$

$$\frac{\mathbf{u} - \mathbf{u}_n}{k} - \nu \Delta \mathbf{u} + \nabla p = \mathbf{f}_{n+1} \quad \text{in } \Omega, \quad \mathbf{u}|_\Gamma = \mathbf{0}, \tag{2.7}$$

$$\operatorname{div} \mathbf{u} = 0 \qquad \text{in } \Omega. \tag{2.8}$$

As above, the velocities $\mathbf{u}_n$ and pressure p_n are calculated implicitly at each time-step for a given source term $\mathbf{f}_{n+1} := \mathbf{f}(t_{n+1})$ and old velocities $\mathbf{u}_n$. One important characteristic property of the semi-discrete formulation (2.7)-(2.8) is the coupling between velocity and pressure is preserved. Therefore, at each time-step, problems of a saddle-point type have to be solved. The corresponding weak formulation for the time-discrete problem, and multiplying by k for readability, reads: find $(\mathbf{u}, p) \in \mathbf{V} \times Q$ such that

$$(\mathbf{u} - \mathbf{u}_n, \mathbf{v}) + k\, A(\mathbf{u}, p; q, \mathbf{v}) = k(\mathbf{f}_{n+1}, \mathbf{v}) \quad \forall \mathbf{v} \in V, \tag{2.9}$$

$$(\operatorname{div} \mathbf{u}, q) = 0 \qquad \forall q \in Q. \tag{2.10}$$

where $A(\mathbf{u}, p) = \nu(\nabla \mathbf{u}, \nabla \mathbf{v}) - (p, \operatorname{div} \mathbf{v})$ is now the bilinear form of the left hand side.

A straightforward spatial discretization of (2.9)-(2.10) provides the fully discrete formula-

tion. Find $(\mathbf{u}^h, p^h) \in \mathbf{V}_h \times Q_h$ such that

$$(\mathbf{u}^h - \mathbf{u}_n^h, \mathbf{v}) + k\, A(\mathbf{u}^h, p^h; q, \mathbf{v}) = k(\mathbf{f}_{n+1}^h, \mathbf{v}) \quad \forall \mathbf{v} \in V_h, \tag{2.11}$$

$$(\operatorname{div} \mathbf{u}^h, q) = 0 \qquad \forall q \in Q_h \tag{2.12}$$

with space discretized bilinear form $A(\mathbf{u}^h, p^h) = \nu(\nabla \mathbf{u}^h, \nabla \mathbf{v}) - (p^h, \operatorname{div} \mathbf{v})$.

Remark 2.4. *On very fine meshes or coarser three-dimensional meshes, solving instantaneously for* $(\mathbf{u}^h, p^h)$ *gives rise to very large matrix systems. Hence, it becomes a challenging task to find solutions in a reasonable computational time.*

Remark 2.5. *It is possible to perform an explicit time integration for the Stokes system* (2.11) *and* (2.12)*. It is well known that their stability is conditional on time step restrictions (McDonough, 2007). In several applications, such as climate research, very small time steps are physically relevant (Wan et al., 2015). In such situations the equations for the velocities are solved explicitly, so that the numerically most expensive step is an elliptic pressure problem.*

2.3 Darcy Flow

Discretizing the mass and momentum conservation system (1.3) - (1.6) for Darcy flow requires the melt-fraction to be simplified somewhat. Utilizing the identity,

$$\frac{1}{\rho}\nabla\rho = \nabla(\ln \rho) \approx \frac{\partial \rho}{\partial p}\mathbf{g}$$

the gradients arising from the application of Darcy's law can be closely related to the gravitational forces. After further manipulation described in detail in chapter 3 and setting the appropriate function spaces $\mathbf{X}^h := \mathbf{V}^h \times Q^h$, we arrive at the weak formulation

$$A(\mathbf{u}_n^h, p_n^h; \mathbf{v}, q) = F[\bar{\rho}_n](\mathbf{v}, q) \quad \forall (\mathbf{v}, q) \in \mathbf{X}^h \tag{2.13}$$

with bilinear form $A(\mathbf{u}_n^h, p_n^h; \mathbf{v}, q)$ and right hand side $F[\bar{\rho}_n](\mathbf{v}, q)$. As a time discretization, the explicit Heun method from section 2.1.2 was selected due to its second order convergence rate. First, for clarity, define the term

$$R_m(\phi, \mathbf{u}; \psi) := (\rho_s^{-1}\Gamma_m, \psi) - ((1-\phi)\mathbf{u}, \nabla\psi) + \int_{\delta\Omega} (\phi - 1)\mathbf{u} \cdot \mathbf{n}\psi \, \mathrm{ds}$$

The explicit predictor step can then be calculated by,

$$(\tilde{\phi}_n^h, \psi) = (\phi_{n-1}^h, \psi) + kR_{n-1}(\phi_{n-1}^h, \mathbf{u}_{n-1}^h; \psi) \qquad \forall \psi \in Q_h$$

providing the predictor mean density $\tilde{\bar{\rho}} = \bar{\rho}(\tilde{\phi}_n^h)$, predictor pressure $\tilde{p}_n^h$, and predictor velocity $\tilde{\mathbf{u}}_n^h$ by calculating,

$$A(\tilde{\mathbf{u}}_n^h, \tilde{p}_n^h; \mathbf{v}, q) = F[\tilde{\bar{\rho}}_n](\mathbf{v}, q) \quad \forall (\mathbf{v}, q) \in \mathbf{X}^h$$

With the predictor step values fed in to the corrector,

$$(\phi_n^h, \psi) = \frac{1}{2}\left(\phi_{n-1}^h + \tilde{\phi}_n^h, \psi\right) + \frac{k}{2} R_n(\tilde{\phi}_n^h, \tilde{\mathbf{u}}_n^h; \psi)$$

giving the new melt fraction, one obtains the mean density, velocity and pressure at the next timestep by solving (2.13).

2.4 The convection-diffusion-reaction equation

We now turn our attention to the scalar convection-diffusion-reaction problem which describes the behavior of scalar quantities like temperature and concentrations, amongst other things. While the scalar field of interest is being transported in a flow field with velocity $\mathbf{b}$ (convection), the molecular transport (diffusion) takes place with a rate ε, and reactions with rate σ. The convection-diffusion problem with homogeneous Dirichlet values reads,

$$\partial_t u - \varepsilon \Delta u + (\mathbf{b} \cdot \nabla) u + \sigma u \ = f \ \text{ in } \Omega, \qquad u|_\Gamma \ = \ 0 \tag{2.14}$$

where $f \in L^\infty(\Omega)$ is a given source term and $\varepsilon > 0$. In addition to the regularity assumptions $\sigma \in L^\infty(\Omega)$ and $\mathbf{b} \in W^{1,\infty}(\Omega)^d$ we further require that $\sigma(x) - \frac{1}{2}\mathrm{div}\, \mathbf{b}(x) \geq \sigma_0 > 0 \quad \forall x \in \Omega$. The ratio of the convective transport rate to the diffusive transport rate is described by the Pèclet number $Pe := \varepsilon^{-1} \|\mathbf{b}\|_\infty L$, where L denotes the characteristic length scale. When convection dominates this leads to high Pèclet numbers which in turn can lead to numerical approximations that oscillate or are unstable. To mitigate this, the numerical solver can increase the number of subdivisions of Ω, i.e. refine the mesh, as a way to reduce L to the point where the local Pèclet number is small enough.

Using the function space $V := H_0^1(\Omega)$ a variational formulation of (2.14) reads: find $u \in V$ such that

$$A(u,v) \;:=\; \varepsilon(\nabla u, \nabla v) + ((\mathbf{b}\cdot\nabla)u, v) + (\sigma u, v) \;=\; (f,v) \qquad \forall v \in V. \tag{2.15}$$

2.4.1 Galerkin discretization

The continuous Galerkin method consists of replacing the test and solution spaces in (2.15) with a finite dimensional subspace $V_h \subset V$: find $u_h \in V_h$ such that

$$A(u_h, v_h) \;=\; F(v_h) \qquad \forall v_h \in V_h. \tag{2.16}$$

The following Lemma shows the coercivity of the bilinear form A with respect to the energy norm $|\!|\!| v |\!|\!|_{1,\varepsilon} := \varepsilon |v|_1^2 + \sigma_0 \|v\|^2$.

Lemma 2.6. *It holds*

$$A(v,v) \;\geq |\!|\!| v |\!|\!|_{1,\varepsilon}^2 \qquad \forall v \in V. \tag{2.17}$$

Proof. Integration by parts and considering $\sigma(x) - \frac{1}{2}\mathrm{div}\,\mathbf{b}(x) \geq \sigma_0 \quad \forall x \in \Omega$ proves the claim. □

If $\varepsilon \lesssim h\|\mathbf{b}\|_\infty$, then diffusion is dominated by convection and the Galerkin method leads to unstable solutions. The existence of these instabilities in the continuous Galerkin method is a known issue and can be remedied by stabilization, and in some cases higher mesh refinement. For an extensive treatment of the subject we refer to (Lube, 1994). In the diffusive limit ($\varepsilon \to 0$) the norm $|\!|\!| v |\!|\!|_{1,\varepsilon}$ loosens the control over gradients, since the diffusivity constant ε scales the $|v|_1^2$ part of the norm $|\!|\!| v |\!|\!|_{1,\varepsilon}$.

Theorem 2.7. *Assume that u and u_h are the solutions of* (2.15) *and* (2.16)*, respectively. It holds*

$$|\!|\!| u - u_h |\!|\!|_{1,\varepsilon} \leq c \inf_{v_h \in V_h} \|u - v_h\|_{1,2}. \tag{2.18}$$

where $c = \varepsilon^{\frac{1}{2}} + \sigma_0^{-\frac{1}{2}}\big(\|\mathbf{b}\|_\infty + \|\sigma - div\,\mathbf{b}\|_\infty\big)$.

Proof. We observe that

$$\begin{aligned}|A(u,v)| &\leq |\varepsilon(\nabla u, \nabla v) + ((\mathbf{b}\cdot\nabla)u, v) + (\sigma u, v)| \\ &\leq \varepsilon|u|_1|v|_1 + |-((\operatorname{div}\mathbf{b})u, v) - (u, (\mathbf{b}\cdot\nabla)v + (\sigma u, v))| \\ &\leq \varepsilon|u|_1|v|_1 + \|\sigma - \operatorname{div}\mathbf{b}\|_\infty\|u\|_0\|v\|_0 + \|\mathbf{b}\|_\infty\|u\|_0|v|_1 \\ &\leq c|\!|\!|u|\!|\!|_{1,\varepsilon}\|v\|_{1,2}\end{aligned}$$

with $c = \varepsilon^{\frac{1}{2}} + \sigma_0^{-\frac{1}{2}}\big(\|\mathbf{b}\|_\infty + \|\sigma - \operatorname{div}\mathbf{b}\|_\infty\big)$. Hence, together with the coercivity result (2.6) the Galerkin orthogonality leads to

$$|\!|\!|u - u_h|\!|\!|_{1,\varepsilon}^2 \leq A(u - u_h, u - u_h) = A(u - u_h, u - v_h) \leq c|\!|\!|u - u_h|\!|\!|_{1,\varepsilon}\|u - v_h\|_{1,2}$$

which concludes the proof. □

Remark 2.8. *For $\varepsilon \gtrsim h$, the error estimate in Theorem 2.7 is optimal. In the case $\varepsilon \lesssim h$, one possible way of obtaining error estimates for the gradients is to employ the inverse estimate*

$$\begin{aligned}\|\nabla(u - u_h)\|_0 &\leq \|\nabla(u - v_h)\|_0 + \|\nabla(v_h - u_h)\|_0 \\ &\leq \|\nabla(u - v_h)\|_0 + \frac{\mu_{inv}}{h}\|v_h - u_h\|_0 \\ &\leq \|\nabla(u - v_h)\|_0 + \frac{\mu_{inv}}{h}\|u - v_h\|_0 + \frac{\mu_{inv}}{\sigma_0^{\frac{1}{2}}h}|\!|\!|u - u_h|\!|\!|_{1,\varepsilon}.\end{aligned}$$

Hence, together with Theorem 2.7 one arrives at the error estimate:

$$\|\nabla(u - u_h)\|_0 \lesssim \frac{1}{h}\inf_{v_h\in V_h}\|u - v_h\|_{1,2} \qquad \forall v \in V$$

which is worse by one order with respect to the mesh size h compared to the optimal order.

Stabilized FEMs serve as a remedy for this phenomena. In Section 2.4.3 we present an overview of some of the most common stabilization techniques. We restrict ourselves to limited instances of linear stabilization. We refer to (Donea and Huerta, 2003; Morton, 2019; Roos et al., 2008) for early Petrov-Galerkin Methods and other stabilization techniques.

2.4.2 Fully discretized convection-diffusion-reaction

The time discretization following the θ-scheme in the implicit Euler configuration now becomes, given u_n^h solve for $u^h = u_{n+1}$,

$$(u^h - u_n^h, v) + k\, A(u^h, v) = k(f_{n+1}^h, v) \qquad \forall v \in V_h. \tag{2.19}$$

2.4.3 Summary of stabilization methods

The idea of stabilized methods is to equip the solution space V with a norm

$$\|\!| \cdot |\!\|_S \;:=\; \Big(|\!\| v |\!\|_{1,\varepsilon}^2 + \sum_{K \in \mathcal{T}_h} |\!\| v |\!\|_K^2 \Big)^{\frac{1}{2}},$$

which is clearly a stronger norm than the energy norm $|\!\| v |\!\|_{1,\varepsilon}$. This is achieved by introducing additional terms for each cell $K \in \mathcal{T}_h$ in the variational formulation. In each cell, the increased stability of the solution is due to the semi-norm $|\!\| v |\!\|_K$. Let $A_s : V \times V \to \mathbb{R}$ be a continuous bilinear form and $F_s : V \to \mathbb{R}$ be a continuous linear form over V. A common characteristic property of the stabilized methods is to find $u_h \in V_h$ such that,

$$A_s(u_h, v_h) \;=\; F_s(v_h) \qquad \forall v_h \in V_h \tag{2.20}$$

instead of (2.16). We will present several assumptions, which suffice for the solution of (2.20) to exhibit a quasi-optimal convergence order. Afterwards, we will list some stabilized methods, which satisfy those assumptions and fit in the error analysis presented below.

Well-posedness and a priori error estimates

We start with an assumption which is necessary and sufficient for the well-posedness of (2.20).

Assumption 2.9. *It holds*

$$(i) \quad \exists\, \beta > 0, \qquad \inf_{u \in V_h \setminus \{0\}} \sup_{v \in V_h \setminus \{0\}} \frac{A_s(u, v)}{|\!\| u |\!\|_S |\!\| v |\!\|_S} \geq \beta, \tag{2.21}$$

$$(ii) \quad \textit{for each } v \in V_h, \quad (A_s(u_h, v) = 0 \quad \forall u_h \in V_h) \implies v = 0. \tag{2.22}$$

Theorem 2.10. *If Assumption 2.9 is satisfied, then* (2.20) *has a unique solution.*

Proof. We refer to (Ern and Guermond, 2013). □

Assumption 2.11. *There exist a semi-norm* $\|\!| \cdot |\!\|_{cont}$ *over* V_h *and* $C_1 > 0$ *such that*

$$A_s(u, v_h) \leq C_1 |\!| u |\!|_{cont} |\!| v_h |\!|_S \qquad \forall u \in V,\ v_h \in V_h.$$

As an upper bound for the interpolation error of $u \in H^{k+1}(\Omega)$ we define

$$E_K^2(u) := (\varepsilon + h_K) h_K^{2k} |u|_{k+1,K}^2 \quad \forall K \in \mathcal{T}_h \text{ and } u \in H^{k+1}(\Omega).$$

Assumption 2.12. *For every* $u \in H^{k+1}(\Omega)$ *there exist* $i_h u \in V_h$ *and* $C_2 > 0$ *such that*

$$|\!| u - i_h u |\!|_S + |\!| u - i_h u |\!|_{cont} \leq C_2 \Big(\sum_{K \in \mathcal{T}_h} E_K^2(u) \Big)^{\frac{1}{2}}.$$

Moreover, it is necessary for the strong solution of (2.14) to solve the stabilized variational formulation (2.20) *at least* up to an accuracy, so that the consistency error is compensated by the interpolation error.

Assumption 2.13. *For the solution* $u \in H^{k+1}(\Omega)$ *of* (2.14)*, there exists* $C_3 > 0$:

$$\sup_{v \in V_h \setminus \{0\}} \frac{|F_s(u, v) - A_s(u, v)|}{|\!| v |\!|_S} \leq C_3 \Big(\sum_{K \in \mathcal{T}_h} E_K^2(u) \Big)^{\frac{1}{2}}.$$

Theorem 2.14. *Assume that the Assumptions 2.9-2.13 are fulfilled. Then it holds*

$$|\!| u - u_h |\!|_S \lesssim \Big(C_2 + \frac{C_1 C_2 + C_3}{\beta} \Big) \Big(\sum_{K \in \mathcal{T}_h} (\varepsilon + h_K) h_K^{2k} |u|_{k+1,K}^2 \Big)^{\frac{1}{2}}. \tag{2.23}$$

Proof. From the approximated Galerkin orthogonality

$$A_s(u - u_h, v_h) = A_s(u, v_h) - F_s(u, v_h)$$

we observe

$$A_s(e_h, v_h) = -A_s(\eta, v_h) - A_s(u, v_h) + F_s(u, v_h)$$

with $\eta := u - i_h u$ and $e_h := i_h u - u_h$. Therefore, together with the triangle inequality, (2.21) provides

$$\begin{aligned}
\|\!|u - u_h|\!\|_S &\leq \|\!|\eta|\!\|_S + \|\!|e_h|\!\|_S \\
&\leq \|\!|\eta|\!\|_S + \frac{1}{\beta} \sup_{v \in V_h \setminus \{0\}} \frac{A_s(e_h, v)}{\|\!|v|\!\|_S} \\
&\leq \|\!|\eta|\!\|_S + \frac{1}{\beta} \sup_{v \in V_h \setminus \{0\}} \frac{|A_s(\eta, v)|}{\|\!|v|\!\|_S} + \frac{1}{\beta} \sup_{v \in V_h \setminus \{0\}} \frac{\left|A_s(u, v) - F_s(v)\right|}{\|\!|v|\!\|_S}. \qquad (2.24)
\end{aligned}$$

The claim follows from Assumptions 2.11- 2.13. □

Remark 2.15. *The a priori error estimate in Theorem 2.14 is quasi-optimal in the $H^1(\Omega)$ norm. However, it is not optimal in the $L^2(\Omega)$ norm, since a $h^{\frac{1}{2}}$ factor is missing.*

2.4.4 Examples of stabilized methods

Many stabilization techniques fulfil the assumptions above and fit into the a priori error estimates in Theorem 2.14. They perturb the bilinear form A and the linear form F as

$$A_s(u, v) := A(u, v) + \sum_{K \in \mathcal{T}_H(\Omega)} S_K(u, v) \qquad (2.25)$$

$$F_s(v) := F(v) + \sum_{K \in \mathcal{T}_H(\Omega)} F_K(v) \qquad (2.26)$$

where $S_K(u, v) : V(K) \times V(K) \to \mathbb{R}$ and $F_K(v) : V(K) \to \mathbb{R}$. In Table 2.1 we list several stabilized methods together with their stabilization terms $S_K(u, v)$, right-hand side contributions $F_K(v)$ and the local semi-norms $\|\!|u|\!\|_K$ in each cell K. Some of their properties are:

- The Streamline-Upwinding-Petrov-Galerkin (SUPG) method (Brooks and Hughes, 1982) is a residual based method and is one of the most popular stabilization techniques. It introduces artificial diffusion in the streamline direction in a consistent way.

- The Galerkin-Least-Squares (GLS) technique (Hughes, Franca, et al., 1989) is an-

	$S_K(u,v)$	$F_K(v)$	$\|\|\|u\|\|\|_K$
SUPG	$\delta_K(\mathcal{L}u, \mathbf{b}\cdot\nabla v)_K$	$\delta_K(f, \mathbf{b}\cdot\nabla v)_K$	$\delta_K\|\mathbf{b}\cdot\nabla u\|_K^2$
GLS	$\delta_K(\mathcal{L}u, \mathcal{L}v)_K$	$\delta_K(f, \mathcal{L}v)_K$	$\delta_K\big(\|\mathbf{b}\cdot\nabla u\|_K^2 + \|\sigma u - \varepsilon\Delta v\|_K^2\big)$
SV	$\delta_K(\nabla(u-\pi u), \nabla(v-\pi v))_K$ $\delta_K(\mathbf{b}\cdot\nabla(u-\pi u), \mathbf{b}\cdot\nabla(v-\pi v))_K$	- -	$S_K(u,u) + H_K\|\mathbf{b}\cdot\nabla u\|_K^2$
LPS	$\delta_K(\kappa(\nabla u), \kappa(\nabla v))_K$	-	$S_K(u,u)$
CIP	$\gamma_K([\|\nabla u\|], [\|\nabla v\|])_{\partial K}$	-	$S_K(u,u) + H_K\|\mathbf{b}\cdot\nabla u\|_K^2$

Table 2.1: Stabilization terms (left), right-hand side contributions (middle) and the additional control in the stabilization norm (right) for each cell K.

other example of residual based stabilization methods. It differs from SUPG depending on the choice of differential operator acting on the test functions.

- The subgrid viscosity (SV) (Guermond, 1999; Guermond, 2001) method introduces artificial diffusion only in the fluctuation space. The main ingredient of the scheme is the L^2 stable projection $\pi : V_h \to V_H$ where $V_H \subset V_h$ is a coarse space, such that

$$\exists\, c_\beta > 0,\ c_0 \geq 0 \qquad \sup_{v\in V_h\setminus\{0\}} \frac{(\mathbf{b}\cdot\nabla u, v)}{\|v\|_0} \geq c_\beta\|\mathbf{b}\cdot\nabla u\|_0 - c_0\|u\|_0 \quad \forall u \in V_H. \tag{2.27}$$

  The artificial diffusion acts only on the fluctuation space $V_h - V_H$, thus the gradients of the fluctuations are stabilized.

- The local projection stabilization (LPS) method (R. Becker and Braack, 2004) stabilizes the fluctuations of the gradients. A fluctuation operator $\kappa : L^2(\Omega) \to L^2(\Omega)$ and a discontinuous discrete space $D_h \subset L^2(\Omega)$ constitute the main mechanism. To this end, a macro-partition $\mathcal{M}_h(\Omega)$ of Ω (e.g. $\mathcal{M}_h(\Omega) = \mathcal{T}_{2h}(\Omega)$) and local projections $\pi_M : L^2(M) \to D_h(M)\ \forall M \in \mathcal{M}_h(\Omega)$ are introduced. Through local projections a global projection $\pi : L^2(\Omega) \to D_h$ is defined by $(\pi v)|_M = \pi_M(v|_M)$ $\forall M \in \mathcal{M}_h(\Omega),\ \forall v \in L^2(\Omega)$. The fluctuation operator is then defined as $\kappa := id - \pi$ and assumed to fulfil

$$\|\kappa q\|_M \lesssim h_M^l |q|_{l,M} \quad \forall q \in H^l(M),\ \forall M \in \mathcal{M}_h(\Omega), 0 \leq l \leq k.$$

Moreover, it is required that

$$\exists\, c_\beta > 0, \forall M \in \mathcal{M}_h(\Omega) \inf_{q_h \in D_h(M)} \sup_{v_h \in Y_h(M)} \frac{(v_h, q_h)}{\|v_h\|_M \|q_h\|_M} \geq c_\beta \tag{2.28}$$

where $Y_h \subset V$ is a continuous finite element space and $Y_h(M) := \{w|_M \ : \ w \in Y_h, \ w = 0 \text{ on } \Omega \backslash M\}$. Another requirement is the existence of an interpolation operator $i_h : H^1(\Omega) \to Y_h \cap H_0^1(\Omega)$ such that

$$\|v - i_h v\|_K + h_K |v - i_h v|_K \lesssim h_K^l \|v\|_{l,\tilde{\omega}(K)} \ \forall v \in H^l(\tilde{\omega}(K)), \ \forall K \in \mathcal{T}_h(\Omega), 0 \leq l \leq k+1$$

where $\tilde{\omega}(K) = \cup_{\overline{K'} \cap \overline{K} \neq \emptyset} K'$. Different choices of the solution and projection spaces are given in (Matthies et al., 2007; Knobloch, 2010). We note that, the LPS method is weakly consistent. A variant of the LPS method which is stable in the SUPG norm is introduced in (Knobloch, 2009).

- The continuous interior penalty (CIP) method (Burman and Hansbo, 2004) controls the jumps of gradients over element faces. For $v \in V_h$, the jump over a face E in a fixed direction $\mathbf{n}_E$ is defined by

$$[|v|]_E(\mathbf{x}) := \begin{cases} \lim_{t \to 0^+} (v(\mathbf{x} - t\mathbf{n}_E) - v(\mathbf{x} + t\mathbf{n}_E)) & \text{if } E \not\subset \partial\Omega, \\ 0 & \text{if } E \subset \partial\Omega. \end{cases}$$

.

Remark 2.16. *In the convection dominated regime ($K \in \mathcal{T}_h(\Omega) : \ \varepsilon \lesssim h_K \|\mathbf{b}\|_{\infty,K}$), the user chosen parameters are suggested as $\delta_K \sim h_K$, $\gamma_K \sim h_K^2$. We refer to (Ern and Guermond, 2013) and (Roos et al., 2008) for justifications of the Assumptions 2.9-2.13 regarding the stabilization techniques listed in the Table 2.1.*

2.5 Nonlinear Optimization

We now provide an brief overview of the optimal control theory and optimality conditions as applied to the problem of chapter 4. The minimization of an objective, or *cost* under the constraint of partial differential equation (PDE) models is an important goal in physical and economic processes. This allows, for example a macroeconomic PDE model (Achdou et al., 2014) of income and wealth distribution to be steered to some desirable outcome.

For further real-world examples see the textbooks by (Kirk, 1971) and (Jahn, 2014).
We start by denoting the variables conventionally used in this field. The control variable embedded in the PDE that we wish to vary is denoted by q, the state variable by u and the adjoint state by z. Define the bounded time interval of the problem as $I := (0, T)$ with $0 < T < \infty$. Now let V and H be Hilbert spaces and let the dual space of V be denoted by V^*. For the generalized definition of the state equation we define the spatial differential operator $L : R \times V \to V^*$. We also define specific inner products for the state equation on the time interval with the subscript I,

$$(v, w)_I := \int_I (v(t), w(t))\,\mathrm{d}t \qquad \text{and} \qquad A(q, u)(\varphi) := \int_I \bar{a}(q(t), u(t))(\varphi(t))\,\mathrm{d}t$$

where $\bar{a} : R \times V \times V \to R$ is the semi-linear form of operator L. Let R be a spatial Hilbert space for the control $q(t)$. Next, we define the Hilbert space for the weak formulation of the state as $X := W(I)$, where

$$W(I) = \left\{ v \,|\, v \in L^2(I, V) \text{ and } \partial_t v \in L^2(I, V^*) \right\}$$

For the controls we choose the Hilbert space as a subspace of $L^2(I, R)$ which we define as $Q \subseteq L^2(I, R)$.

The primal equation

We consider the state equations as a general parabolic PDE initial value problem on time interval I,

$$\partial_t u(t) + L(q(t), u(t)) = f(t) \qquad \forall t \in I \tag{2.29}$$

$$u(0) = u_0(q(0)) \tag{2.30}$$

with $q(t) \in R$ and $u(t)$ in V the solution of state equation. The right hand side is given by $f(t) \in L^2(I, V^*)$ and the initial condition is $u_0 : R \to H$. The state equation can then be expressed compactly in the weak form: Find for a given control $q \in Q$ a state $u \in X$, such that

$$(\partial_t u, \varphi)_I + A(q, u)(\varphi) + (u(0), q(0)) = (f, \varphi)_I + (u_0(q), \varphi(0)) \quad \forall \varphi \in X \tag{2.31}$$

where the initial condition is coupled to the state equation by $\varphi(0)$.

The cost functional

A general objective functional J can be defined with three components, each individually dependent on u, $u(T)$ and q respectively. With the mappings $J_1 : V \to \mathbb{R}$ and $J_2 : H \to \mathbb{R}$ as $J : Q \times X \to \mathbb{R}$ the objective is defined as,

$$J(q,u) := \int_I J_1(u(t))\,\mathrm{d}t + J_2(u(T)) + \frac{\alpha}{2}||q||_Q^2 \tag{2.32}$$

where the $\frac{\alpha}{2}||q||_Q^2$ term is an explicit regularization factor to ensure uniqueness of the optimal solution. The optimization problem can now be stated as follows,

$$\left.\begin{array}{l} \text{minimize } J(q,u) \text{ subject to} \\ \text{PDE constraints (2.31)} \\ (q,u) \in Q \times X \end{array}\right\} \tag{2.33}$$

For the necessary or sufficient optimality conditions that are required for a well-posed optimization problem, one needs to take the first and second order derivatives of $J(q,u)$ to determine its minimum and curvature. The traditional sensitivity approach is too costly if the all gradients are required. An adjoint approach has been developed by (R. Becker and Vexler, 2007) that relies on an implicit *reduced cost functional*. For its definition we now assume the existence of a solution operator $S : Q \to X$ which maps the control q to the unique solution $u = S(q)$ of the weak state (2.31). This assumption permits the definition of the reduced cost functional $j : Q \to \mathbb{R}$ as $j(q) := J(q, S(q))$ and hence the (implicit in u) optimization problem can be stated,

$$\left.\begin{array}{c} \text{minimize } j(q) \\ q \in Q \end{array}\right\} \tag{2.34}$$

On the question of existence, (Dacorogna, 1989) proved that if j is weakly lower semi-continuous and coercive over Q, then the minimization problem (2.34) has at least one solution $q \in Q$.

The question of uniqueness can be answered using a convexity argument: If the reduced functional j fulfils the existence criteria of Dacorogna and is in addition strongly convex on Q, with $\lambda \in (0,1)$ and $q_1, q_2 \in Q$ and $q_1 \neq q_2$ such that,

$$j(\lambda q_1 + (1-\lambda)q_2) < \lambda j(q_1) + (1-\lambda) j(q_2)$$

then minimization problem (2.34) has a unique solution. For the proof by contradiction, see (Meidner, 2008).
For the necessary and sufficient optimality conditions we reproduce the theorems from (Meidner, 2008) as definitions without proofs and refer the reader to (Tröltzsch, 2005) for an explanation of the technicalities. The fundamental conditions required for a well-posed optimization problem are the first and second necessary optimality conditions, as well as a second order sufficient condition:

Definition 2.17 (First order necessary optimality condition)**.** *Let the reduced functional j be Gateaux differentiable on an open subset $Q_0 \subseteq Q$. If $q \in Q_0$ is a local optimal solution of the optimization problem* (2.34)*, then there holds the first order necessary optimality condition*

$$j'(q)(\delta q) = 0 \qquad \forall \delta q \in Q$$

Definition 2.18 (Second order necessary optimality condition)**.** *Let the reduced functional j be two times continuously Fréchet differentiable on an open subset $Q_0 \subseteq Q$. If $q \in Q_0$ is a local optimal solution of the optimization problem* (2.34)*, then there holds the second order necessary optimality condition*

$$j''(q)(\delta q, \delta q) \geq 0 \qquad \forall \delta q \in Q$$

Definition 2.19 (Second order sufficient optimality condition)**.** *Let the reduced functional j be two times continuously Fréchet differentiable on a neighborhood $Q_0 \subseteq Q$ of q, Moreover, let the control q fulfill the first order necessary optimality condition*

$$j'(q)(\delta q) = 0$$

and assume the existence of $\gamma > 0$ such that the second order sufficient optimality condition

$$j''(q)(\delta q, \delta q) \geq \gamma \|\delta q\|_Q^2 \qquad \forall \delta q \in Q$$

is valid, Then a constant $\rho > 0$ exists such that the quadratic growth condition

$$j(q + \delta q) \geq j(q) + \frac{\gamma}{4}\|\delta q\|_Q^2$$

holds for all $\delta q \in Q$ with $\|\delta q\|_Q \leq \rho$. As a consequence q is a local solution of the optimization problem (2.34)*.*

These definitions depend on the concepts of differentiability in normed vector spaces with Gateaux and Fréchet derivatives. For an complete explanation of these concepts refer to (Jahn, 2014). For the computational application, (Roland Becker et al., 2007) explains the theory and implementation of methods for evaluating the required derivatives numerically by finite element methods.

The dual equation

When numerically solving a standalone PDE, the system is naturally discretized for an approximation to be computed, as demonstrated by the spatio-temporal examples above. In optimization one has the choice to first establish the necessary optimality conditions on the continuous level of the primal, dual and optimality equations, and then discretize them suitably for a numerical solver, as in the *optimize-then-discretize* approach (Collis et al., 2002). An alternative approach, *discretize-then-optimize*, is to derive the optimality system from an already discretized primal system. It is known that in the presence of some kinds of stabilization terms, as described in section 2.4.4, the two approaches do not commute. For an in-depth explanation of how the control is affected by the choice of stabilization, and the justification for using LPS, particularly for flow problems, we refer to (Braack, 2009).

For an optimize-then-discretize approach we can define the Lagrangian $\mathcal{L} : V \times Q \to \mathbb{R}$ based on the explicit objective (2.32) to formulate the adjoint equation as,

$$\mathcal{L}(q, u, Z) := J(q, u) + A(q, u)$$

Here the space of Lagrange multipliers is $Z := H^1(T, 0; V)$. For clarity we have not discretized the system and have dropped the h superscript. For the derivative of $A(q, u)$ with respect to u at point (q, u) we use the notation $DA(q, u)$, similarly $DJ_1(U)$ is the derivative of J_1 w.r.t. U at point U and DJ_2, that of J_2 w.r.t. $U(T)$. The adjoint equation now reads,

$$\begin{aligned} -\frac{\partial Z}{\partial t} + DA(q, u) &= DJ_1(u) \\ Z|_{t=T} &= DJ_2 \end{aligned} \tag{2.35}$$

The third and final requirement for a well posed optimality system is the variational in-

equality,

$$\int_I \partial_q J(q,u)(p-q) \geq 0 \qquad \forall p \in Q \tag{2.36}$$

where p is the quantity that varies in relation to q. The complete optimality system is then fully defined by the primal PDE state equations (2.31), the dual equation (2.35) and the variational inequality (2.36).

2.6 Overview of publications

In the following we give an overview of the contents corresponding to each publication.

Paper 1

The first paper deals with a problem in geophysics and the dynamics of mantle convection and melt migration. In this application the melt migration is modeled by coupling Stokes flow containing a non-constant viscosity with the Darcy flow described in chapter 1. In this arrangement the Stokes component captures the slow long-term movements of the earths mantle and lithosphere, whilst the Darcy flow models the two-phase regime. The work is an extension of a previous model that required the use of the compaction pressure making the underlying operator non-coercive. This would require the use of either mixed inf-sup stable elements, or applicable stabilization were the system to be solved with the finite element method. The original contribution in this work was the application of a coercive non-symmetric linear operator, allowing the use of simple equal-order elements. The existence and uniqueness of the solutions was proven, as well as the coercivity of the bilinear form. As a numerical validation, a stationary example with a known solution was presented to prove the expected convergence rate for Q_2 elements, as well as a super-convergence in the case of Q_1. For a further example, a time-dependent system was discretized according to the Heun scheme from section 2.1.2 and second order convergence was observed.

Paper 2

The second paper proposes an optimal marine biomass harvesting model. A modified non-convex cost functional that incorporates a penalty term for vessels removing biomass from marine protected areas is minimized by the optimal control principles presented in

section 2.5 by application of the Stokes and convection-diffusion equations described above. The setting is the south-western quadrant of the Baltic Sea where LPS stabilized Stokes flow as documented in 2.4.4 is used to simulate ocean currents. A convective term is also stabilized with the SUPG method. As a coupling term in a predator-prey model we utilized the relatively recent predator-dependent sigmoidal functional response $f(u, v)$ of Crowley and Martin (Crowley and Martin, 1989),

$$f(u, v) = \frac{\omega u}{(1 + \omega a u)(1 + bv)}$$

A predator-prey dynamical system, with the functional response $f(u, v)$ constitute the state, or constraints, of the minimization problem. These state equations were implemented as close as possible to physical reality by selecting biological parameters from research published in peer-reviewed journals. A penalty was shown to exist that balances the harvesting and predation pressure such that the harvested species biomass is not driven below unsustainable levels.

Paper 3

The third paper is devoted to the problem of stabilizing a sediment transport system in three dimensions. A composition of Stokes flow and the convection-diffusion-reaction equation from section 2.4, with a special flux term to describe the movement and settling of a suspension in a flow field. What sets this class of problem apart is the coupling method: the flow equations have a variable and nonlinear viscosity dependent on the concentration, and the convection equation has a nonlinear variable diffusion coefficient. This gives rise to a highly nonlinear and sensitive problem to solve. Most recent work in this area has involved finding numerical solutions with the finite difference method with some form of streamline diffusion stabilization, i.e. SUPG/GLS as described in section 2.4.4, or various other penalty methods. In this work, the problem was stabilized using the LPS method and numerically solved for the non-stationary case in three dimensions.

	Journal	Status
Paper 1	*Journal of Applied Mathematics and Computing*, Vol. 65, 2021, pp. 273-293 DOI: 10.1007/s12190-020-01391-y © Springer The original publication is available at `https://doi.org/10.1007/s12190-020-01391-y`	peer-reviewed published
Paper 2	*Natural Resource Modeling* Vol. 38, Issue 1, Feb. 2025 DOI: 10.1111/nrm.70000 © The Authors (CC BY 4.0) The original publication is available at `https://doi.org/10.1111/nrm.70000`	peer-reviewed published
Paper 3	*Computational and Applied Mathematics*	submitted 26.02.25

Table 2.2: Publications and corresponding journals.

Paper 1: Equal-order finite element approximation for mantle-melt transport

3

The original publication is available at
https://doi.org/10.1007/s12190-020-01391-y

Journal of Applied Mathematics and Computing (2021) 65:273–293
https://doi.org/10.1007/s12190-020-01391-y

ORIGINAL RESEARCH

Equal-order finite element approximation for mantle-melt transport

Malte Braack[1] · **Kamel Nafa**[2] · **Simon Taylor**[1]

Received: 25 February 2020 / Revised: 23 June 2020 / Accepted: 26 June 2020 / Published online: 10 July 2020

Abstract

Mantle convection and melt migration are important processes for understanding Earth's dynamics and how they relate to observations at the surface. Recently it has been established that melt migration can be modelled by coupling variable-viscosity Stokes flow and Darcy flow, where Stokes flow generally captures the long-term behaviour of the mantle and lithosphere, and Darcy flow models the two-phase regime. It is known that approximating the solution by finite element methods requires the use of mixed inf-sup stable elements or additional stabilization terms. Here, we propose a formulation with a coercive non-symmetric linear operator which allows the use of simple equal-order elements.

Keywords Elliptic differential equations · Porous media · Finite elements · Error estimates

Mathematics Subject Classifications 35J25 · 76Sxx · 65N30

1 Introduction

The need to solve systems of coupled Navier-Stokes and Darcy flow arises is various fields, such as modelling the interaction of surface water and groundwater aquifers [1,

This work was supported by the SQU Internal grant IG/SCI/DOMS/17/02 and Oman Research council (TRC) grant RC/SCI/DOMS/16/01.

Malte Braack
braack@math.uni-kiel.de

Kamel Nafa
nkamel@squ.edu.om

Simon Taylor
taylor@math.uni-kiel.de

[1] Mathematical Seminar, University of Kiel, Ludewig-Meyn-Str.4, 24098 Kiel, Germany

[2] Department of Mathematics, College of Science, P.O.Box 36, Al Khoudh 123, Muscat, Oman

2], blood flow problems [3,4], and fuel cell dynamics [5,6]. Approximating the solution of coupled Stokes-Darcy systems in most methods requires solving Stokes or Darcy on adjacent sub-domains coupled with appropriate interface boundary conditions [7–10], and [11]. In melt migration modelling the partially molten rock cannot be handled using the approach above because we cannot identify the boundary between the solid and fluid phases. The model derived by Mc McKenzie [12] for melt migration assumes a dual continuum mixture of solid matrix and fluid melt. The mixing parameter is the porosity φ, i.e. the volume fraction of fluid melt, which is assumed to be much smaller than 1 and may be zero in parts of the domain where there is no fluid melt. The motion of the solid is governed by Stokes flow, and the melt is transported according to Darcy's law. In addition, he included a compacting relation that relates the solid and fluid pressures.

Most of the previous studies of geodynamics do not consider that melt migration can be modelled by coupling variable-viscosity Stokes flow and Darcy flow, where Stokes flow generally captures the long-term behaviour of the mantle and lithosphere, and Darcy flow models the two-phase regime. Their studies either do not consider melting or treat it in a simplified way [13–16]. Others have simplified the model by approximating the two-phase flow and the effects of compaction [17–19], and [20]. The correct setting was later proposed by [21]. However, the required computational effort was limited to 2D problems, and although these models take into account the compaction of the solid matrix to allow melt to be expelled or to flow in, they treat both individual phases as incompressible and often assume their densities to be constant.

The discretization of a similar (but not identical) complex melting problem is suitable described by Schimenz et al. [33] by using a mixed discontinuous Galerkin method in vertical direction and a Fourier method in vertical direction for the elliptic-hyperbolic system of equations.

In [23], the authors derive such a model for 2D and 3D simulations. However, the model they used require the so-called compaction pressure as an additional unknown in the system which not only increases the dimension of the system, but also makes the underlying operator non-coercive (see, [21–23]). Thus, approximating the solution using the finite element method requires the use of mixed inf-sup stable elements or additional stabilization terms. Here, we propose a formulation with a coercive non-symmetric linear operator which allow the use of simple equal-order elements.

2 Governing equations

Let $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, be a Lipschitz-domain. The model we consider here is a two-phase flow consisting of melt and a deformable matrix. The melt fraction φ defines averaged quantities $\bar{X}$ out of solid (matrix) X_s and fluid X_f quantities (the subscripts f and s stand for fluid and solid, respectively):

$$\bar{X} = (1 - \varphi)X_s + \varphi X_f. \tag{1}$$

The governing equations for a two-phase flow were given by McKenzie in [12]. They include the equations for conservation of mass, momentum, and energy. The mass

Table 1 Physical quantities and associated symbols

Variable	Symbol	Unit
Solid velocity	$\mathbf{u}_s$	$m/year$
Fluid velocity	$\mathbf{u}_f$	$m/year$
Fluid pressure	p_f	Pa
Compaction pressure	p_c	Pa
Shear viscosity of the solid	η	$Pa \cdot year$
Bulk viscosity of the solid	ξ	$Pa \cdot year$
Darcy coefficient	K_D	Pa
Gravitational force	$\mathbf{g}$	$m^2/year$
Fluid density	ϱ_f	$kg \cdot m^{-3}$
Solid density	ϱ_s	$kg \cdot m^{-3}$
Mean density	$\bar{\varrho}$	$kg \cdot m^{-3}$
Mean compressibility	α	$kg \cdot m^{-3} Pa^{-1}$
Mass force	f	$Pa \cdot kg \cdot m^2/year$
Melt production rate	Γ	$kg/(m^3 year)$

conservation for fluid and solid are described by the two equations

$$\frac{\partial(\varrho_f \varphi)}{\partial t} + \nabla \cdot (\varrho_f \varphi \mathbf{u}_f) = \Gamma, \tag{2}$$

$$\frac{\partial(\varrho_s (1-\varphi))}{\partial t} + \nabla \cdot (\varrho_s (1-\varphi) \mathbf{u}_s) = -\Gamma. \tag{3}$$

Where, ϱ_f and ϱ_s are the fluid- and solid densities, $\mathbf{u}_f$ and $\mathbf{u}_s$ the fluid- and solid velocities, respectively, t the time, and Γ the melt production rate following a volume element of matrix. The conservation of momentum of the melt is given by

$$\varphi(\mathbf{u}_f - \mathbf{u}_s) = -K_D (\nabla p_f - \varrho_f \mathbf{g}), \tag{4}$$

with the constant $K_D = k_\varphi/\eta_f$, where k_φ is the permeability and η_f is the melt viscosity, p_f the pressure within the melt, $\mathbf{g}$ the gravity acceleration vector. For the case of vanishing solid velocity, $\mathbf{u}_s = 0$, Eq. (4) reduces to the well known Darcy's law. Further quantities and associated symbols used in this paper are given in Table 1.

The conservation of momentum of the matrix is given by

$$-\nabla \cdot T\mathbf{u}_s + \nabla p_f = \bar{\varrho}\mathbf{g}, \tag{5}$$

where T is the strain tensor given by

$$T\mathbf{u} := \eta(\nabla \mathbf{u} + \nabla \mathbf{u}^T) + (\xi - \tfrac{2}{3}\eta)(\nabla \cdot \mathbf{u})I, \tag{6}$$

with the shear viscosity of the solid η, the bulk viscosity of the solid ξ, and the identity matrix I .

2.1 Simplified equations at steady densities

In the case that the densities do not vary with time i.e. $\partial\varrho_{f,s}/\partial t = 0$, Eqs. (2) and (3) reduce to:

$$\frac{\partial\varphi}{\partial t} + \nabla\cdot(\varphi\mathbf{u}_f) = \frac{\Gamma}{\varrho_f} - \frac{\varphi}{\varrho_f}\mathbf{u}_f\cdot\nabla\varrho_f \tag{7}$$

$$-\frac{\partial\varphi}{\partial t} + \nabla\cdot((1-\varphi)\mathbf{u}_s) = -\frac{\Gamma}{\varrho_s} - \frac{1-\varphi}{\varrho_s}\mathbf{u}_s\cdot\nabla\varrho_s. \tag{8}$$

Adding Eqs. (7) and (8) we obtain

$$\nabla\cdot\left[\varphi\mathbf{u}_f + (1-\varphi)\mathbf{u}_s\right] = \Gamma\left(\frac{1}{\varrho_f} - \frac{1}{\varrho_s}\right) - \frac{\varphi}{\varrho_f}\mathbf{u}_f\cdot\nabla\varrho_f - \frac{1-\varphi}{\varrho s}\mathbf{u}_s\cdot\nabla\varrho_s.$$

Now, we can eliminate the fluid velocity $\mathbf{u}_f$, using Darcy's law (4):

$$\begin{aligned}\nabla\cdot\mathbf{u}_s - \nabla\cdot(K_D\nabla p_f) + K_D\mathbf{g}\cdot\nabla\varrho_f \\ = \Gamma\left(\frac{1}{\varrho_f} - \frac{1}{\varrho_s}\right) + \left(-\frac{\varphi}{\varrho_f}\mathbf{u}_s + K_D\left(\frac{1}{\varrho_f}\nabla p_f - \mathbf{g}\right)\right)\cdot\nabla\varrho_f, \\ -\frac{1-\varphi}{\varrho_s}\mathbf{u}_s\cdot\nabla\varrho_s - \varrho_f\nabla\cdot(K_D\mathbf{g}).\end{aligned} \tag{9}$$

Several terms in Eq. (9) contain gradients of the solid and fluid density, respectively. As proposed in [23], the logaritmic derivatives $\nabla(\ln\varrho_{f,s}) = \varrho_{f,s}^{-1}\nabla\varrho_{f,s}$ can be assumed to be closely related to the gravitational forces:

$$\frac{1}{\varrho_s}\nabla\varrho_s = \nabla(\ln\varrho_s) \approx \frac{\partial\varrho_s}{\partial p_s}\mathbf{g} \quad\text{and}\quad \frac{1}{\varrho_f}\nabla\varrho_f = \nabla(\ln\varrho_f) \approx \frac{\partial\varrho_f}{\partial p_f}\mathbf{g}.$$

Assuming that the isothermal compressibilities, $\kappa_s := \varrho_s^{-1}\frac{\partial\varrho_s}{\partial p_s}$ and $\kappa_f := \varrho_f^{-1}\frac{\partial\varrho_f}{\partial p_f}$, are constant and

rearranging terms leads to the following reformulation (approximation) of (9):

$$\nabla\cdot\mathbf{u}_s - \nabla\cdot(K_D\nabla p_f) - K_D\kappa_f\varrho_f\mathbf{g}\cdot\nabla p_f + \left(\varphi\varrho_f\kappa_f + (1-\varphi)\varrho_s\kappa_s\right)(\mathbf{u}_s\cdot\mathbf{g}) = f,$$

with given forcing

$$f := \Gamma\left(\frac{1}{\varrho_f} - \frac{1}{\varrho_s}\right) - 2K_D\kappa_f\varrho_f^2|\mathbf{g}|^2 - \varrho_f\nabla\cdot(K_D\mathbf{g}).$$

32

Thus, the two-phase PDE system reduces to the original two-phase PDE system given by McKenzie [12]:

$$-\nabla \cdot T\mathbf{u}_s + \nabla p_f = \bar{\varrho}\mathbf{g} \tag{10}$$
$$\nabla \cdot \mathbf{u}_s - \nabla \cdot (K_D \nabla p_f) + \beta \mathbf{g} \cdot \nabla p_f + \alpha \mathbf{u}_s \cdot \mathbf{g} = \mathbf{f} \tag{11}$$

with α, the weighted average of the flux of compressibility of solid and fluid densities with respect to the corresponding pressures, and transport-scaling parameter β given by

$$\alpha := \overline{\varrho\kappa} = \varphi\varrho_f\kappa_f + (1-\varphi)\varrho_s\kappa_s, \tag{12}$$
$$\beta := -K_D\kappa_f\varrho_f. \tag{13}$$

2.2 Time dependent melt fraction

Although we previously assumed that the flow field is in temporal equilibrium, this is not necessarily the case for the melt fraction φ. The corresponding equation is given by

$$\varrho_s \frac{\partial \varphi}{\partial t} + \nabla \cdot (\varrho_s(\varphi - 1)\mathbf{u}_s) = \Gamma. \tag{14}$$

Assuming $\varrho_s > 0$ a.e. in Ω, this equation can also be written in the form

$$\frac{\partial \varphi}{\partial t} + \nabla \cdot ((\varphi - 1)\mathbf{u}_s) + (\varphi - 1)\kappa_s\varrho_s^2\mathbf{u}_s \cdot \mathbf{g} = \frac{\Gamma}{\varrho_s}. \tag{15}$$

2.3 Boundary conditions

The system of Eqs. (10), (11) and (15) has to be supplemented with suitable boundary conditions. We impose Dirichlet conditions for $\mathbf{u}_s$ and the lithostatic pressure gradient for p_f:

$$\mathbf{u}_s = \mathbf{u}_0 \quad \text{on} \quad \partial\Omega, \tag{16}$$
$$\nabla p_f \cdot \mathbf{n} = \varrho_s \mathbf{g} \cdot \mathbf{n} \quad \text{on} \quad \partial\Omega. \tag{17}$$

The boundary condition for φ are of Dirichlet type

$$\varphi = \varphi_0 \quad \text{on } \partial\Omega_-,$$

on the inflow part of the boundary $\partial\Omega_- := \{x \in \partial\Omega : \mathbf{u}_s \cdot \mathbf{n} < 0\}$.

2.4 Compaction pressure

In order to solve the PDE system (10), (11) numerically several authors treat it as a system of three-field saddle point problems [21,23], and [22] by introducing a secondary pressure p_c, called compaction pressure, defined as

$$p_c = -\xi \nabla \cdot \mathbf{u}_s. \tag{18}$$

Then, Eq. (10) reduces to the form

$$-\nabla \cdot (\tilde{T}\mathbf{u}_s) + \nabla(p_f + p_c) = \bar{\varrho}\mathbf{g}. \tag{19}$$

Where, the reduced strain tensor $\tilde{T}\mathbf{u}_s$ is given by

$$\tilde{T}\mathbf{u}_s := \eta(\nabla \mathbf{u}_s + \nabla \mathbf{u}_s^T) - \frac{2}{3}\eta(\nabla \cdot \mathbf{u}_s)\mathbf{I}. \tag{20}$$

The resulting system (11), (18) and (19) now has three unknowns $\mathbf{u}_s$, p_f, and p_c. The corresponding system is of saddle point structure, so that its finite element discretization is not possible without considering inf-sup stable elements or special stabilization techniques, for instance the pressure stabilized Petrov-Galerkin method (PSPG) or the local projection pressure stabilization (LPS). We refer to [24,26,29], and [25] for details. In this work we will follow a simpler approach by dealing with the original equations. So, the compaction pressure p_c will not be a primary variable. In case this quantity is needed for special purposes, it can be recovered from $\mathbf{u}_s$ by solving the discrete system corresponding to (18), as we shall explain below.

3 Weak formulation and discretization for constant melt fraction

Let us first consider the case that the melt fraction φ is given and constant in time. Then the density ϱ becomes a given constant and we only have two unknown variables p_f and $\mathbf{u}_s$, determined by a linear stationary system of equations, i.e. without any time dependency. Moreover, let us assume for ease of presentation that we have homogeneous Dirichlet data for the velocities, $\mathbf{u}_0 = 0$. The extension to non-homogeneous Dirichlet data is straight forward.

3.1 Weak formulation

The associated function spaces are

$$\mathbf{X} := \mathbf{V} \times Q, \quad \mathbf{V} := H_0^1(\Omega)^d, \quad Q := H^1(\Omega) \cap L_0^2(\Omega).$$

As bilinear form $A : \mathbf{X} \times \mathbf{X} \to \mathbb{R}$ we introduce

34

$$A(\mathbf{u}, p; \mathbf{v}, q) := (T\mathbf{u}, \nabla \mathbf{v}) - (p, \nabla \cdot \mathbf{v})$$

$$+(\nabla \cdot \mathbf{u}, q) + (K_D \nabla p, \nabla q) + (\beta \mathbf{g} \cdot \nabla p, q) + (\alpha \mathbf{u} \cdot \mathbf{g}, q).$$

The right hand side is given by

$$F[\bar{\varrho}](\mathbf{v}, q) := (\bar{\varrho}\mathbf{g}, \mathbf{v}) + (f, q) + \int_{\partial\Omega} K_D \varrho_s \mathbf{g} \cdot \mathbf{n} q \, ds. \tag{21}$$

In order to have this bilinear form A and the right hand side $F[\bar{\varrho}]$ to be well-defined, it is sufficient to make certain regularity assumptions on the involved coefficients. This will be discussed below.

We seek $\mathbf{u}_s \in \mathbf{V}$ and $p_f \in Q$ s.t.

$$A(\mathbf{u}_s, p_f; \mathbf{v}, q) = F[\bar{\varrho}](\mathbf{v}, q) \qquad \forall (\mathbf{v}, q) \in \mathbf{X}. \tag{22}$$

Assumption 1 We assume the following regularities for the coefficients

$$\begin{aligned} \varrho_s, K_D, \mathbf{g} &\in C(\bar{\Omega}), \\ \varrho_f &\in W^{1,3}(\Omega) \cap L^\infty(\Omega), \\ \kappa_f, \kappa_s, &\in L^\infty(\Omega), \\ \varphi &\in L^{3/2}(\Omega), \\ \Gamma &\in H^{-1}(\Omega), \end{aligned}$$

and the partial densities are bounded from below as

$$\varrho_s, \varrho_f \geq \epsilon > 0 \quad \text{a.e. in } \Omega.$$

The regularity assumptions for $\varrho_s, \varrho_f, \kappa_s$ and κ_f are sufficient to ensure that α determined by (12) has the regularity $\alpha \in L^{3/2}(\Omega)$. This can be used to show the well-posedness of the term $(\alpha \mathbf{u} \cdot \mathbf{g}, q)$ in the bilinear form A:

$$\begin{aligned} |(\alpha \mathbf{u} \cdot \mathbf{g}, q)| &= \|\mathbf{g}\|_{L^\infty} \|\alpha \mathbf{u} q\|_{L^1(\Omega)} \\ &\leq \|\mathbf{g}\|_{L^\infty(\Omega)} \|\mathbf{u}\|_{L^6(\Omega)^d} \|\alpha\|_{L^{3/2}(\Omega)^d} \|\mathbf{g}\|_{L^6(\Omega)^d} \\ &\leq \|\mathbf{g}\|_{L^\infty(\Omega)} \|\mathbf{u}\|_{H^1(\Omega)^d} \|\alpha\|_{L^{3/2}(\Omega)^d} \|q\|_{H^1(\Omega)}. \end{aligned} \tag{23}$$

Here we used the Sobolev embedding $H^1(\Omega) \hookrightarrow L^6(\Omega)$, for $d = 2, 3$.

Lemma 1 *Under the regularity assumption (Assumption 1) the linear form $F[\bar{\varrho}]$ is a dual form of $\boldsymbol{X}$, i.e. $F[\bar{\varrho}] \in \boldsymbol{X}'$.*

Proof Let us now check the individual term of $F[\bar{\varrho}]$ in view of appropriate definition. The boundary integral is well-defined due to the assumed continuity of K_D, $\mathbf{g}$, ρ_s and the trace inequality:

$$\left| \int_{\partial\Omega} K_D \varrho_s \mathbf{g} \cdot \mathbf{n} q \, ds \right| \leq C \|q\|_{H^1(\Omega)},$$

with a constant $C = C(\Omega, K_D, \varrho_s, \mathbf{g})$. The term $(\bar{\varrho}\mathbf{g}, \mathbf{v})$ is bounded as

$$|(\bar{\varrho}\mathbf{g}, \mathbf{v})| \leq \|\mathbf{g}\|_{L^\infty(\Omega)} \|\bar{\varrho}\|_{H^{-1}(\Omega)} \|\mathbf{v}\|_{H^1(\Omega)}.$$

The regularity $\bar{\varrho} \in H^{-1}(\Omega)$ is obtained by Eq. (1) and $\varphi \in L^{3/2} \subset H^{-1}(\Omega)$ and $\varrho_s, \varrho_f \in L^\infty(\Omega)$. It remains to show an upper bound for the dual pairing of forcing $f \in H^{-1}(\Omega)$ and $q \in H^1(\Omega)$:

$$\langle f, q \rangle \leq \left(c_2 \|\Gamma\|_{H^{-1}(\Omega)} + c_3\right) \|q\|_{H^1(\Omega)} + c_4 |\varrho_f q|_{H^1(\Omega)} + c_5 \|q\|_{L^2(\partial\Omega)},$$

with $c_2 = c_2(\epsilon)$, $c_3 = c_3(K_D, \kappa_f, \mathbf{g}, \varrho_f)$, $c_4 = c_4(K_D, \mathbf{g})$ and $c_5 = c_5(K_D, \mathbf{g}, \varrho_f)$. By the trace inequality we can bound $\|q\|_{L^2(\partial\Omega)}$ by the H^1-norm of q. The Hölder inequality and Sobolev embedding $H^1(\Omega) \hookrightarrow L^6(\Omega)$ yields

$$\begin{aligned} |\varrho_f q|_{H^1(\Omega)} &\leq \|\nabla \varrho_f q\|_{L^2(\Omega)} + \|\varrho_f \nabla q\|_{L^2(\Omega)} \\ &\leq \|\nabla \varrho_f\|_{L^3(\Omega)} \|q\|_{L^6(\Omega)} + \|\varrho_f\|_{L^\infty(\Omega)} \|\nabla q\|_{L^2(\Omega)} \\ &\leq c_6 \|q\|_{H^1(\Omega)} \end{aligned}$$

with $c_6 = c_6(\varrho_f)$ due to the assumption $\varrho_f \in W^{1,3}(\Omega)$. In summary we obtain $\|f\|_{H^{-1}(\Omega)} \leq C$, and hence $F[\bar{\varrho}] \in \mathbf{X}'$. □

Remark 1 Even a stronger assumption $\varphi \in L^\infty(\Omega)$ is not a severe restriction, because, from the physical point of view, $0 \leq \varphi \leq 1$ is reasonable. The lower bound for ϱ_f and ϱ_s are necessary for ensuring that the right hand side f is properly defined.

3.2 Existence and uniqueness of solutions

Throughout this section we presuppose that Assumption 1 is valid. In order to prove the continuity and the coercivity of the bilinear form $A(\mathbf{u}, p; \mathbf{v}, q)$, and the continuity of the right hand side linear form, we equip the product space $\mathbf{X}$ with the norm:

$$|||(\mathbf{u}, p)||| := \left(\|\eta^{1/2} \nabla \mathbf{u}\|^2 + \|(2(\xi + \tfrac{1}{3}\eta))^{1/2} \nabla \cdot \mathbf{u}\|^2 + \|K_D^{1/2} \nabla p\|^2 \right)^{1/2}. \quad (24)$$

Here and in what follows we use the notation $\|\cdot\|$ for the L^2-norm over Ω.

Lemma 2 *[Continuity] The bilinear form $A(\mathbf{u}, p; \mathbf{v}, q)$ is continuous with respect to the triple norm $|||\cdot|||$.*

Proof The assumed regularity of ϱ_s, ϱ_f and φ ensures $\alpha \in L^{3/2}(\Omega)$. Hence, we can use (23) to bound the reaction term. Further, we have $\beta \in L^\infty(\Omega)$ so that a term-by-term estimate, use of the Poincaré inequality leads to

$$A(\mathbf{u}, p; \mathbf{v}, q) \leq C_1 |||(\mathbf{u}, p)||| |||(\mathbf{v}, q)|||$$

36

with the constant $C_1 = C_1(\Omega, K_D, \beta, \eta, \mathbf{g}, \varrho_s, \varrho_f, \varphi, \kappa_f, \kappa_s)$. □

In order to show the coercivity of this bilinear form, we first show the non-negativity of the strain tensor when it is tested diagonally:

Lemma 3 *For constant η it holds*

$$(T\boldsymbol{u}, \nabla\boldsymbol{u}) = \|\eta^{1/2}\nabla\boldsymbol{u}\|^2 + \|(\xi + \tfrac{1}{3}\eta)^{1/2}\nabla\cdot\boldsymbol{u}\|^2 \qquad \forall\boldsymbol{u}\in\boldsymbol{V}.$$

Proof Because of $I : \nabla\mathbf{u} = \nabla\cdot\mathbf{u}$ we obtain

$$\begin{aligned}(T\mathbf{u}, \nabla\mathbf{u}) &= \|\eta^{1/2}\nabla\mathbf{u}\|^2 + (\eta\nabla\mathbf{u}^T, \nabla\mathbf{u}) + ((\xi - \tfrac{2}{3}\eta)(\nabla\cdot\mathbf{u})I, \nabla\mathbf{u})\\ &= \|\eta^{1/2}\nabla\mathbf{u}\|^2 + \sum_{i,j}(\eta\partial_j\mathbf{u}_i, \partial_i\mathbf{u}_j) + \|(\xi - \tfrac{2}{3}\eta)^{1/2}\nabla\cdot\mathbf{u}\|^2.\end{aligned}$$

where $\partial_j\mathbf{u}_i = \dfrac{\partial\mathbf{u}_i}{\partial x_j}$ for $i, j = 1, \ldots, d$.

a. Let us first assume, that $\mathbf{u}\in(H_0^1(\Omega)\cap C^2(\Omega))^d$. Due to this regularity of $\mathbf{u}$ the terms in the sum above can be integrated by parts:

$$\begin{aligned}\sum_{i,j}(\partial_j\mathbf{u}_i, \partial_i\mathbf{u}_j) &= -\sum_{i,j}(\partial_i\partial_j\mathbf{u}_i, \mathbf{u}_j) \;=\; -\sum_{i,j}(\partial_j\partial_i\mathbf{u}_i, \mathbf{u}_j)\\ &= \sum_{i,j}(\partial_i\mathbf{u}_i, \partial_j\mathbf{u}_j) \;=\; \left(\sum_i\partial_i\mathbf{u}_i, \sum_j\partial_j\mathbf{u}_j\right)\\ &= \|\nabla\cdot\mathbf{u}\|^2.\end{aligned}$$

Using this in the equation derived above, yields the assertion.

b. For less regular $\mathbf{u}$, the same equation is obtained by a density argument: We take the limit of a smooth limiting sequence $(\mathbf{u}_n)_{n\in\mathbb{N}}\subset(H_0^1(\Omega)\cap C^2(\Omega))^d$, $\mathbf{u}_n\to\mathbf{u}$ in $\mathbf{V}_0$:

$$\begin{aligned}\sum_{i,j}(\partial_j\mathbf{u}_i, \partial_i\mathbf{u}_j) &= \lim_{n\to\infty}\sum_{i,j}(\partial_j(\mathbf{u}_n)_i, \partial_i(\mathbf{u}_n)_j)\\ &= \lim_{n\to\infty}\|\nabla\cdot\mathbf{u}_n\|^2 \;=\; \|\nabla\cdot\mathbf{u}\|^2.\end{aligned}$$

□

Lemma 4 (Coercivity) *We assume that*

$$\|\boldsymbol{g}\|_{L^\infty(\Omega)^d}\|\overline{\varrho\kappa}\|_{L^{3/2}(\Omega)} \le c \tag{25}$$

with a certain constant $c = c(\eta, K_D)$. Then the bilinear form A is continuous and $\boldsymbol{X}$-coercive; in particular

$$A(\boldsymbol{u}, p; \boldsymbol{u}, p) \ge \frac{1}{2}|||(\boldsymbol{u}, p)||| \quad \forall(\boldsymbol{u}, p)\in\boldsymbol{X}. \tag{37}$$

Proof The assertion follows by diagonal testing, use of the previous Lemma, and the fact that the convective term vanishes when tested diagonally i.e. $(\mathbf{g}\cdot\nabla p,\, p) = 0$. The only critical (not necessarily positive) term is the zero order reaction term $(\alpha\mathbf{u}\cdot\mathbf{g},\, p)$. However, due to the assumption on α, this term can be bounded by (23) and Young's inequality as follows

$$\begin{aligned}|(\alpha\mathbf{u}\cdot\mathbf{g},\, p)| &\le \|\alpha\|_{L^{3/2}(\Omega)}\|\mathbf{g}\|_{L^\infty(\Omega)}\|\mathbf{u}\|_{H^1(\Omega)}\|p\|_{H^1(\Omega)}\\ &\le \frac{1}{2}C(\eta,\, K_D)\|\alpha\|_{L^{3/2}(\Omega)}\|\mathbf{g}\|_{L^\infty(\Omega)}\left(\|\eta^{1/2}\mathbf{u}\|^2_{H^1(\Omega)^d} + \|K_D^{1/2}p\|^2_{H^1(\Omega)}\right)\\ &\le \frac{1}{2}(\|\eta^{1/2}\nabla\mathbf{u}\|^2 + \|K_D^{1/2}\nabla p\|^2).\end{aligned}$$

This leads to the desired lower bound for $A(\mathbf{u},\, p;\, \mathbf{u},\, p)$. □

Remark 2 The constant c in (25) depends on η and K_D. In the special case of constant η and K_D this constant is just $c = \sqrt{\eta K_D}/c_\Omega$, where c_Ω is the Poincaré constant.

Theorem 1 *With the same assumptions as in the previous Lemma, Eq. (22) has an unique solution* $(\boldsymbol{u}_s,\, p_f) \in \boldsymbol{X}$ *and*

$$\begin{aligned}&\|\eta^{1/2}\nabla\boldsymbol{u}_s\|^2 + \|(\xi + \tfrac{1}{3}\eta)^{1/2}\nabla\cdot\boldsymbol{u}_s\| + \|K_D^{1/2}\nabla p_f\|^2\\ &\quad\le \|\bar{\varrho}\boldsymbol{g}\|^2_{H^{-1}(\Omega)^d} + \|f\|^2_{H^{-1}(\Omega)} + \|K_D\varrho_s\boldsymbol{g}\cdot\boldsymbol{n}\|^2_{L^2(\partial\Omega)}.\end{aligned}$$

Proof Taking into account that the quantities $\bar{\varrho}$ and f are functions of φ and Γ, respectively, we see that the H^{-1}-norms of $\bar{\varrho}$ and f are well-defined. Now, the assertion is an immediate consequence of the previous Lemma and the theorem of Lax-Milgram. □

3.3 Equal-order finite elements

The coercivity of the problem allows us to use several standard equal order elements for the discrete velocity $\mathbf{u}^h$ and the discrete pressure p^h. Let T_h be a shape regular partition of Ω into d-dimensional simplices, quadrilaterals or hexahedra [30]. The diameter of a cell $K \in T_h$ will be denoted by h_K and the mesh parameter $h = \max\{h_K | K \in T_h\}$ represents the maximum diameter of all cells. Let $S^h \subset H^1(\Omega)$ be a finite element space of continuous, piecewise polynomial functions defined over T_h, with degree of interpolation order $r \ge 1$. We consider triangulations T_h of shape regular elements, so that an interpolation operator $i_h : H^1(\Omega) \to S^h$ exists (Scott-Zhang interpolation [31]) with the following properties for all $1 \le l \le r+1$, all $K \in T_h$, and all $u \in H^l(\omega_K)$:

$$\|u - i_h u\|_K \lesssim h_K^l\, |u|_{H^l(\omega_K)}\,, \tag{26}$$

$$\|\nabla(u - i_h u)\|_K \lesssim h_K^{l-1}\, |u|_{H^l(\omega_K)}\,. \tag{27}$$

Here, ω_K denotes a union of cells in the neighbourhood of K, and the expression $a \lesssim b$ means $a \le cb$ with the constant c independent of the mesh parameter h.

Possible choices on shape regular meshes are P_r or Q_r elements, or finite element spaces containing these spaces, e.g. locally enriched spaces. Here, we consider equal-order finite element spaces for velocity and pressure:

$$\mathbf{V}^h := (S^h)^d \cap \mathbf{V}, \quad Q^h := S^h \cap Q, \text{ and } \mathbf{X}^h := \mathbf{V}^h \times Q^h.$$

Now, the linear system to be solved reads as follows: Find $(\mathbf{u}^h, p^{h)} \in \mathbf{X}^h$ such that

$$A(\mathbf{u}^h, p^h; \mathbf{v}, q) = F[\bar{\varrho}](\mathbf{v}, q), \quad \forall (\mathbf{v}, q) \in \mathbf{X}^h. \tag{28}$$

The main advantage of the equal-order discretization of the problem is that the degrees of freedom of all variables can be assigned to the same geometrical identities, e.g. to vertices. Therefore, the discrete systems can be designed in a block-wise manner, so that an efficient block preconditioner or block smoother for multigrid solvers can be utilized.

3.4 A priori error estimate

Theorem 2 *We make the same assumptions as in Theorem 1. The discrete solution of Eq. (28)* $(\mathbf{u}^h, p^h) \in \mathbf{X}^h$ *is unique. If* $(\mathbf{u}_s, p_f) \in H^l(\Omega)^{d+1}$ *with* $1 \leq l \leq r+1$ *we have the following discretization error estimate*

$$|||(\boldsymbol{u}_s - \boldsymbol{u}^h, p_f - p^h)||| \leq C h^{l-1} \left(|\mathbf{u}_s|_{H^l(\Omega)} + |p_f|_{H^l(\Omega)} \right).$$

Proof Due to coercivity (Lemma 4) we can apply Cea's lemma, see e.g. [30]:

$$|||(\mathbf{u}_s - \mathbf{u}^h, p_f - p^h)||| \leq C |||(\mathbf{u}_s - \boldsymbol{i}_h \mathbf{u}, p_f - i_h p_f)|||,$$

with arbitrary interpolation $\boldsymbol{i}_h : H^l(\Omega)^d \to \mathbf{V}_h$ and $i_h : H^l(\Omega) \to Q_h$. For $l = 1$ we choose the Scott-Zhang interpolant, and for $l \geq 2$ we take the nodal interpolant. This leads e.g. to

$$\| K_D^{1/2} \nabla (p - i_h p) \| \leq c h^{l-1} \| p \|_{H^l(\Omega)}$$

with a constant c depending only on K_D and Ω. The other parts in the norm $|||\cdot|||$ are obtained analogously. □

3.5 Post-processing of the compaction pressure

Working with equal-order elements on simplices, i.e. with P_r elements, the divergence of the discrete velocity, $\nabla \cdot \mathbf{u}_h$, is piece-wise polynomial of order $r-1$ but discontinuous across element faces/edges. Hence, for constant (or cell-wise constant) bulk viscosity, the discrete compaction pressure $p_c^h := -\xi \nabla \cdot \mathbf{u}^h$ is of the same type. Therefore, the evaluation of this quantity inside the cells is straight forward. The situation is different for its evaluation on vertices or edges/faces due to the discontinuity. If the discrete

compaction pressure is required on the vertices, denoted by p_c^h, a common strategy is to define it by the L^2-projection onto Q_h,

$$p_c^h \in Q^h : \quad (p_c^h, q) = (-\xi \nabla \cdot \mathbf{u}^h, q) \quad \forall q \in Q^h. \tag{29}$$

Lemma 5 *Under the same assumptions as in Theorem 1,* $(\mathbf{u}_s, p_f) \in H^l(\Omega)^{d+1}$ *and* $p_c \in H^{l-1}(\Omega)$ *with* $1 \le l \le r+1$ *we have*

$$\|p_c - p_c^h\| \le Ch^{l-1} \left(|\boldsymbol{u}_s|_{H^l(\Omega)} + |p_f|_{H^l(\Omega)} + |p_c|_{H^{l-1}(\Omega)} \right).$$

Proof Let $\tilde{p}_c^h \in Q^h$ be the solution of the problem

$$(\tilde{p}_c^h, q) = (-\xi \nabla \cdot \mathbf{u}_s, q) \quad \forall q \in Q^h.$$

Then we have by Cea's Lemma and standard interpolation results

$$\|p_c - \tilde{p}_c^h\| \le C \inf_{q^h \in Q^h} \|p_c - q^h\| \ \le \ Ch^{l-1} |p_c|_{H^{l-1}(\Omega)}.$$

By stability of the discrete equations and Theorem 2, we have

$$\|p_c^h - \tilde{p}_c^h\| \le \|\xi \nabla \cdot (\mathbf{u}_s - \mathbf{u}^h)\| \ \le \ Ch^{l-1} (|\mathbf{u}_s|_{H^l(\Omega)} + |p_f|_{H^l(\Omega)}),$$

with a constant C depending on ξ and η. With the triangle inequality we arrive at the desired estimate. □

For more regular compaction pressure, $p_c \in H^l(\Omega)$, we see that the obtained accuracy for the compaction pressure is of one order less (with respect to the mesh size h) than the optimal interpolation error. However, there are well known methods to increase the accuracy by special gradient recovery techniques as an alternative to solving Eq. (29). We refer to the classical work of Zienciewicz and Zhu [32].

4 Variational formulation and discretization for variable melt fraction rate

In this section we propose a second order explicit Runge-Kutta scheme to solve the hyperbolic PDE for the melt fraction φ. However, we start with the first order forward Euler method, because it will be an intermediate step in the higher-order time stepping scheme.

4.1 First-order forward Euler method

We first discretize the hyperbolic Eq. (15) for φ in time by using a forward Euler with time step $k := t_n - t_{n-1} > 0$. The semi-discrete equation for $\varphi_n \approx \varphi(t_n)$ reads

$$\varphi_n = \varphi_{n-1} + k \left(\varrho_s^{-1} \Gamma_{n-1} + \nabla \cdot ((1 - \varphi_{n-1}) \mathbf{u}_{n-1}) \right).$$

Taking this equation in variational form and applying Q_r elements results in the discrete system for $\varphi_n^h \in Q^h$:

$$(\varphi_n^h, \psi) = (\varphi_{n-1}^h, \psi) + kR_{n-1}(\varphi_{n-1}^h, \mathbf{u}_{n-1}^h; \psi) \quad \forall \psi \in Q^h, \tag{30}$$

with

$$R_m(\varphi, \mathbf{u}; \psi) := (\varrho_s^{-1}\Gamma_m, \psi) - ((1-\varphi)\mathbf{u}, \nabla\psi) + \int_{\partial\Omega} (\varphi - 1)\mathbf{u}\cdot\mathbf{n}\psi \, ds.$$

Solving Eq. (30) requires basically inversion of the mass matrix. Afterwards, the mean density can be updated:

$$\bar{\varrho}_n := \bar{\varrho}(\varphi_n^h) = (1-\varphi_n^h)\varrho_s + \varphi_n^h \varrho_f. \tag{31}$$

Finally, the equation for $(\mathbf{u}_n^h, p_n^h) \in \mathbf{X}^h$ is solved:

$$A(\mathbf{u}_n^h, p_n^h; \mathbf{v}, q) = F[\bar{\varrho}_n](\mathbf{v}, q), \quad \forall (\mathbf{v}, q) \in \mathbf{X}^h, \tag{32}$$

with the right hand side $F_n[\bar{\varrho}_n]$ as given in Eq. (21). The entire algorithm for variable melt fraction rate looks now as follows:

`Algorithm (forward Euler)`

1. Initialize $\mathbf{u}_0^h$, p_0^h and φ_0^h, set $n := 0$ and $t_n := 0$.
2. Increase $n \to n+1$ and set $t_n := t_{n-1} + k$.
3. Make one forward Euler step (30) to determine φ_n^h.
4. Update mean density $\bar{\varrho}_n$ by (31).
5. Solve the linear problem (32) to determine p_n^h and $\mathbf{u}_n^h$.
6. If $t_n \leq T$ goto 2.

Note that we have a time step restriction (CFL condition) of the form $k \leq h$ for stability issues. An alternative without such a time step restriction would be the usage of an implicit time stepping scheme for φ_n^h. However, in this case, the corresponding equation includes $\mathbf{u}_n^h$. On the other hand, the quantities $\alpha(t_n)$ in (12) and $\bar{\varrho}(t_n)$ include φ_n^h, so that the equations for φ_n^h and (28) become mutually coupled. Solving this system is much more numerically expensive than the semi-explicit algorithm presented above.

4.2 Second-order forward Heun method

In order to obtain a second order method, me apply a second-order explicit Runge-Kutta method for φ, for instance the Heun method. This time stepping scheme consists of the forward Euler as a predictor step:

$$(\varphi_{*,n}^h, \psi) = (\varphi_{n-1}^h, \psi) + kR_{n-1}(\varphi_{n-1}^h, \mathbf{u}_{n-1}^h; \psi) \quad \forall \psi \in Q^h, \tag{33}$$

leading to a predictor mean density $\bar{\varrho}_{*,n} = \bar{\varrho}(\varphi^h_{*,n})$, and a predictor velocity and pressure, $\mathbf{u}^h_{*,n}$ and $p^h_{*,n}$ respectively, given as solution of

$$A(\mathbf{u}^h_{*,n}, p^h_{*,n}; \mathbf{v}, q) = F[\bar{\varrho}_{*,n}](\mathbf{v}, q), \quad \forall (\mathbf{v}, q) \in \mathbf{X}^h. \tag{34}$$

Afterwards, the new melt fraction is obtained by solving

$$(\varphi^h_n, \psi) = \frac{1}{2}\left(\varphi^h_{n-1} + \varphi^h_{*,n}, \psi\right) + \frac{k}{2} R_n(\varphi^h_{*,n}, \mathbf{u}^h_{*,n}; \psi), \tag{35}$$

followed by the new mean density $\bar{\varrho}_n = \bar{\varrho}(\varphi^h_n)$, and new velocity and pressure by solving (32).

The numerical cost of the Heun method is just a factor of two compared to the forward Euler, but leads to a substantial increase in accuracy. This will be demonstrated in the numerical examples below. The corresponding algorithm looks as follows:

Algorithm (Heun method)

1. Initialize $\mathbf{u}^h_0$, p^h_0 and φ^h_0, set $n := 0$ and $t_n := 0$.
2. Increase $n \to n+1$ and set $t_n := t_{n-1} + k$.
3. Make one forward Euler step (33) to determine $\varphi^h_{*,n}$.
4. Update mean density $\bar{\varrho}_{*,n}$ (similar to (31)).
5. Solve the linear problem (34) to determine $p^h_{*,n}$ and $\mathbf{u}^h_{*,n}$.
6. Update melt fraction φ^h_n by inversion of mass matrix, (35).
7, Get new mean density $\bar{\varrho}_n$ according to (31).
8. Update solid velocity and fluid pressure by solving (32).
9. If $t_n \leq T$ goto 2.

5 Numerical example

In this section we test the methodology for two 2D examples, a stationary problem and a time-dependent problem. Both cases are designed in such a way that the exact solution is known.

5.1 Stationary example

To validate the analysis and the error estimates of the proposed scheme, we choose as an example the two-dimensional problem given in [23] with a known exact solution:

42

$$\mathbf{u}_s(z) = \left(\frac{1}{10}e^z, -\frac{3}{40}e^z\right)^T \quad \text{and} \quad p_f(z) = p_0 - z - \frac{27}{200}e^z.$$

The problem is solved on the unit square $\Omega := (0, 1)^2$ in the $xz-$ plane. The Darcy coefficient K_D and the melt fraction φ depend only on the variable z :

$$K_D(z) = \frac{149}{45} + \frac{1}{30}e^z, \ \varphi(z) = 1 - 0.3e^z.$$

The shear and bulk viscosities η and ξ, respectively, are given by

$$\eta(x) = \frac{1}{2}e^{2x},$$
$$\xi(x, z) = e^{-z} + \frac{2}{3}e^{2x} + 1.$$

The resulting compaction pressure according to (18) is

$$p_c(x, z) = \frac{3}{40}\left(1 + \frac{2}{3}e^{2x+z} + e^z\right).$$

We further set $\varrho_f(z) := e^{-z}$, $\varrho_s(z) := 1.2\varrho_f(z)$, gravitation $\mathbf{g}(z) := (0, -e^z)^T$, and $\Gamma(z) := \frac{27}{1000}e^z$. For the isothermal compressibilities we choose constant values $\kappa_f = 1$ and $\kappa_s = 5/6$. Hence, $\alpha(z) = e^{-z}$ and $\beta(z) = -K_D\kappa_f\varrho_f = -\frac{149}{45}e^{-z} - \frac{1}{30}$. The boundary data is given by the exact solutions, i.e. for the velocity $\mathbf{u}_0 = \mathbf{u}_s\ |_{\partial\Omega}$, and for the pressure $f_2 = -\partial_n e^{-z}\ |_{\partial\Omega}$. Now we can obtain the forcing term by basic calculations

$$f = \Gamma\left(\frac{1}{\varrho_f} - \frac{1}{\varrho_s}\right) - 2K_D\kappa_f\varrho_f^2|\mathbf{g}|^2 - \varrho_f\nabla\cdot(K_D\mathbf{g}) = \frac{9}{2000}e^{2z} - \frac{149}{45},$$
$$\bar{\varrho}\mathbf{g} = \left(0, -\frac{4}{5} - \frac{3}{50}e^z\right)^T.$$

In Fig. 1 we display the discretization errors obtained for Q_1 (left figure) and Q_2 (right figure) approximations. In the Q_1 result the L^2 errors of p and $\mathbf{u}$ are of order $O(h^2)$. The error in the gradient (i.e. in the H^1-semi norm of p and $\mathbf{u}$ is of order $O(h)$. For the Q_2 approximation the convergence is enhanced, where the L^2 errors in p and $\mathbf{u}$ are $O(h^3)$. The errors $|\mathbf{u} - \mathbf{u}_h|_{H^1(\Omega)}$ and $|p - p_h|_{H^1(\Omega)}$ have order $O(h^2)$. This corresponds exactly to the a priori estimate in Theorem 2.

Moreover, we are interested in the error behavior of the post-processed compaction pressure p_c. In Fig. 2 we plot the corresponding discretization errors for Q_1 and Q_2 elements. We obtain for Q_1-elements an error $\|p_c - p_c^h\| = \mathcal{O}(h^{3/2})$, and for Q_2-elements an error $\|p_c - p_c^h\| = \mathcal{O}(h^2)$. In comparison with the theoretical expectation, these results corresponds exactly with the theory for Q_2, and show a super-convergence behaviour for Q_1.

In order to get an idea about the error distribution, in Fig. 3 we depict the discretization errors on a sequence of meshes for Q_1-elements. The errors are reduced uniformly under mesh refinement.

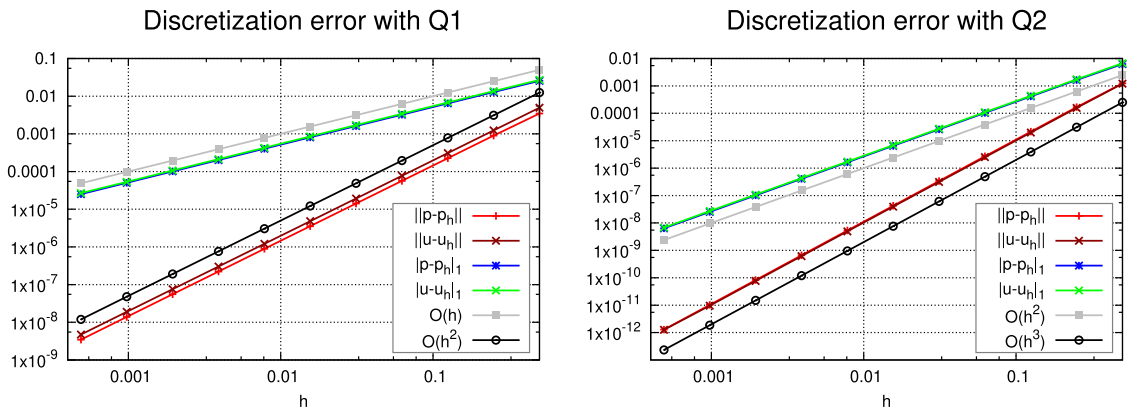
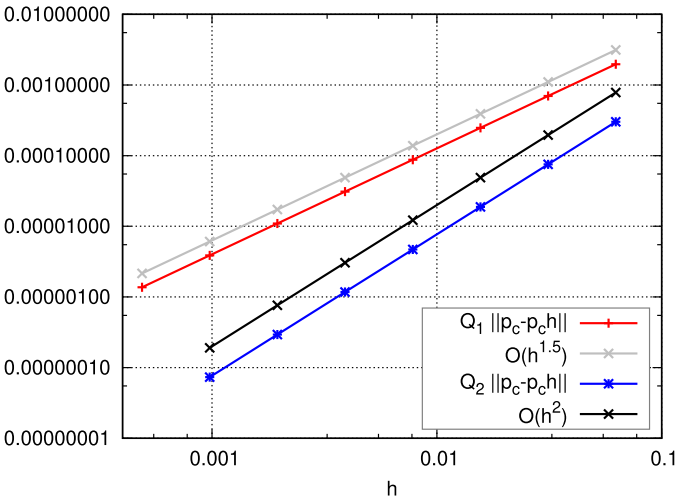


Fig. 1 Discretization error obtained with Q_1 (left) and Q_2 approximation (right)

Fig. 2 Discretization error of the compaction pressure, $\|p_c - p_c^h\|$ obtained with Q_1 and Q_2 approximation



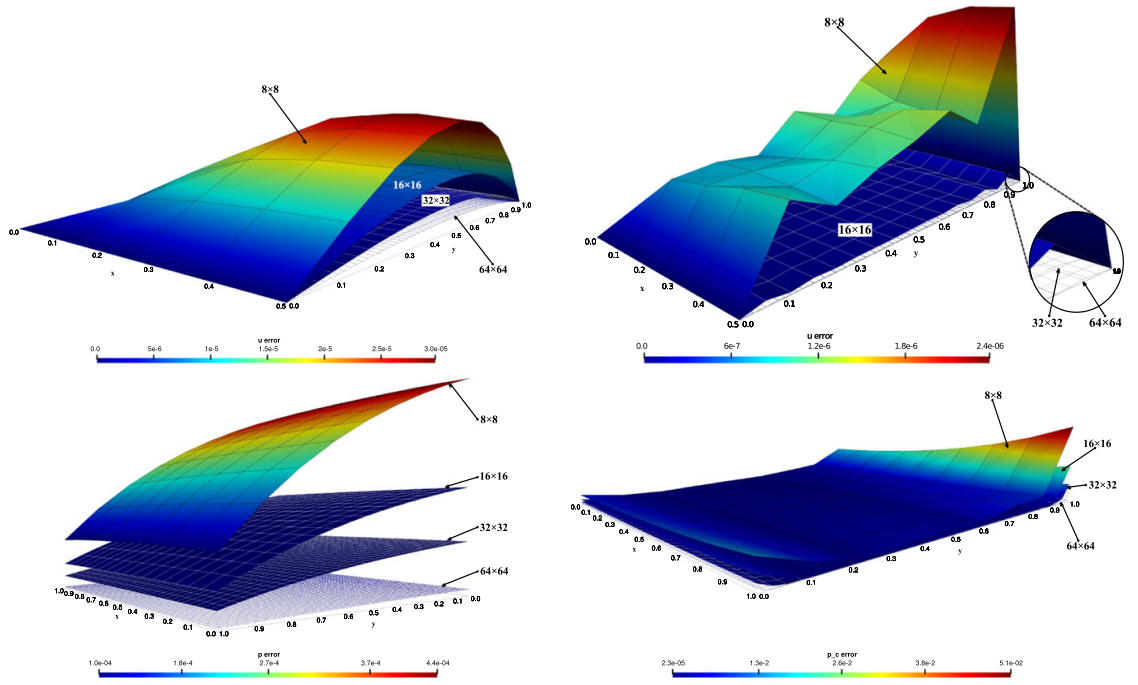


Fig. 3 Discretization errors in example 5.1: $|\mathbf{u} - \mathbf{u}_h|$ with Q_1 approximation (top left), with Q_2 approximation (top right), $p - p_h$ (lower left) and $p_c - p_c^h$ (lower right) with Q_1 on a sequence of meshes. The colors and elevations are scaled individually for each picture

5.2 Time-dependent example

We extend the example in the previous section to the time-dependent case. The exact solution now reads:

$$\begin{aligned}\mathbf{u}_s(z,t) &= e^{z+t}\left(\frac{1}{10}, -\frac{3}{40}\right)^T,\\ p_f(z,t) &= p_0 - z - \frac{27}{200}e^z + (1-e^t)(-0.2z + 0.135e^z),\\ \varphi(z,t) &= e^t(1-0.3e^z).\end{aligned}$$

The coefficients η, ξ and K_D remain independent of time and, hence, identical to their values in the stationary case (previous subsection). The quantities Γ, f and f_2 are adapted in such a way that the solution given above solves the set of equations. Obviously, the solution is designed in such a way that we recover for $t = 0$ the same solution as in the stationary example: $\mathbf{u}_s(z,t) = \mathbf{u}_s(z,0)e^t$, $p_f(z,t) = p_f(z,0) + q(z,t)$ with $q(z,t) = (1-e^t)(-0.2z+0.135e^z)$ and $\varphi(z,t) = \varphi(z,0)e^t$. The averaged density becomes

$$\begin{aligned}\bar{\varrho}(z,t) &= \varphi(z,t)\varrho_f(z) + (1-\varphi(z,t))\varrho_s(z)\\ &= e^t\bar{\varrho}(z,0) + (1-e^t)\varrho_s(z).\end{aligned}$$

It is straightforward to verify that (10) and (11) are satisfied. The compaction pressure then becomes

$$p_c(x,z,t) = -\xi(x,z)\nabla\cdot\mathbf{u}(z,t) = \frac{3}{40}e^t\left(1 + \frac{2}{3}e^{2x+z} + e^z\right),$$

and finally the pressure forcing term results to

$$f(z,t) = e^t\left(\frac{149}{225} + \frac{9}{2000}e^{2z}\right) - \frac{298}{75}.$$

The initial forcing at $t = 0$, i.e. $f(z,0)$, is identical to $f(z)$ of the stationary example.

5.2.1 Forward Euler

In order to validate the temporal error we first have a look at the error in melt fraction in different norms, see Fig. 4. In the L^2- and the L^∞-norm we observe first order convergence with respect to the time step, $\|\varphi - \varphi_k^h\|_{L^2(\Omega)} \sim \|\varphi - \varphi_k^h\|_{L^\infty(\Omega)} \sim k$. The H^1-seminorm also begins to reduce with first order, but stagnates for smaller time steps. This is not unexpected, since Eq. (15) does not enforce H^1-regularity nor H^1-stability.

For the pressure and velocity variables the spatial error is much larger than the temporal error. The reason for this is probably that the time step does not enter directly

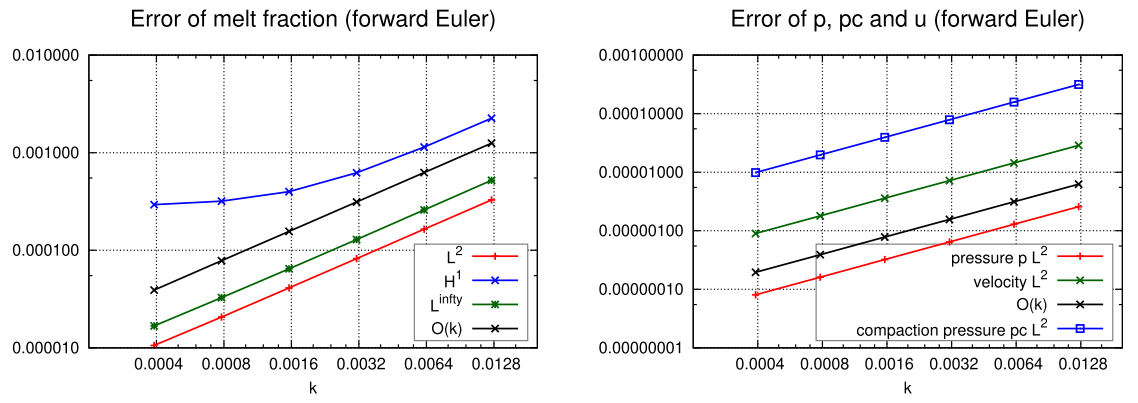


Fig. 4 Discretization error of melt fraction (left) and pressure and velocity (right) obtained with forward Euler (Sect. 5.2.1)

into Eq. (22), but the temporal discretization enters only implicitly by the mean density $\bar{\varrho}$ which itself depends on the time dependent melt fraction φ. However, in order to visualize the temporal impact we plot the temporal error only by considering the quantities $p^h - p_n^h$ and $\mathbf{u}^h - \mathbf{u}_n^h$ at $t = 0.1$, where p^h and $\mathbf{u}^h$ are pressure and velocity, respectively, discretized in space but with the correct mean density $\bar{\varrho}(t_n)$, i.e.

$$A(\mathbf{u}^h, p^h; \mathbf{v}, q) = F[\bar{\varrho}(t_n)](\mathbf{v}, q) \quad \forall (\mathbf{v}, q) \in \mathbf{X}^h.$$

In Fig. 4 (right figure) we observe first oder convergence of the error in the L^2-norms, $\|p^h - p_n^h\| \sim k$ and $\|\mathbf{u}^h - \mathbf{u}_n^h\| \sim k$ for fixed spatial mesh size h. The compaction pressure, obtained by post-processing, also converges with first order: $\|(p_c)^h - (p_c)_n^h\| \sim k$.

5.2.2 Heun method

For the Heun method on a fixed spatial mesh we obtain second order convergence for the melt fraction $\|\varphi_k^h - \varphi\|_{L^2(\Omega)} \sim k^2$ in L^2-, H^1- and L^∞-norm (Fig. 5 left) until the (fixed) spatial error dominates and leads to stagnation of the total error. The reason that this stagnation is not observed with the forward Euler method is that with Heun, the error is a factor 100 smaller than with Euler (with $k = 4 \cdot 10^{-3}$). This stagnation appears earlier on in the H^1-norm. For solid velocity, fluid pressure and compaction pressure we obtain second order convergence as well (Fig. 5 right). Here, no stagnation appears which is due to the fact that we (once more) only depict the temporal error, e.g. $\|\mathbf{u}^h - \mathbf{u}_n^h\|_{L^2(\Omega)}$, so that spatial effects are excluded.

5.3 Example in 3D

As third example we choose a configuration without known analytical solution in three spatial dimensions. The computational domain is $\Omega := (-1, 1) \times (-1, 1) \times (0, 1)$. The setting is stationary with the following choice of melt fraction and Darcy coefficient, respectively:

46

$$\gamma(x, y, z) = e^{-20x^2y^2z},$$

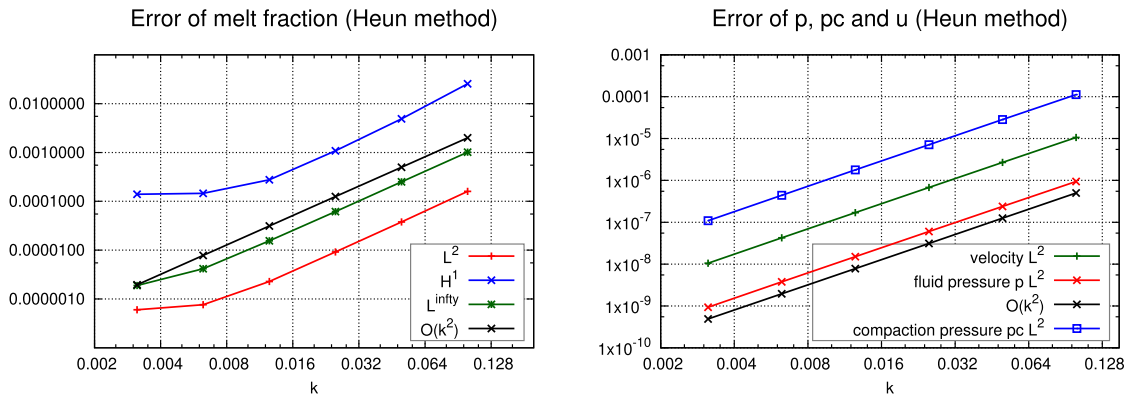


Fig. 5 Discretization error of melt fraction (left) and of the pressure and velocity (right) for the time-dependent case with the Heun method in the L^2-norm (Sect. 5.2.2)

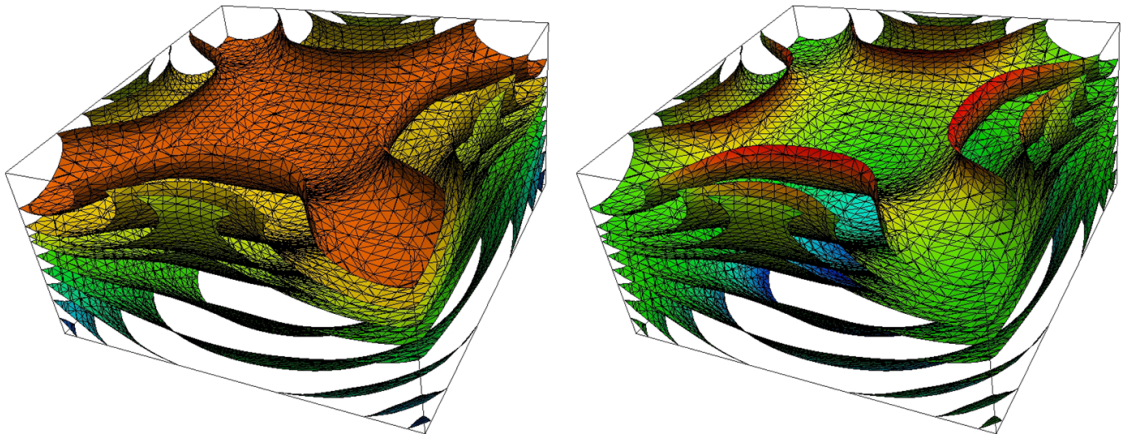


Fig. 6 Pressure p (left) and compaction pressure p_c in the 3D-configuration (Sect. 5.3)

$$\varphi(x, y, z) = 1 - x^2 y^2 z,$$
$$K_D(x, y, z) = (1.01 - x^2 y^2 z)^{-1}.$$

Shear and bulk viscosities are simply $\eta = \xi = 1$. Gravitation has only a vertical component, $\mathbf{g} = -10e_3$. The remaining coefficients are chosen as $\varrho_s = 20$, $\varrho_f = 10$, $\kappa_f = 0.05$, $\kappa_s = \kappa_f/6$. The velocity data for the boundary conditions is given by $\mathbf{u}_0 = 0$. The two pressures are shown in Fig. 6.

6 Summary

We propose a variational formulation for modeling mantle-melt transport with a coercive bilinear form for solid velocity $\mathbf{u}_s$ and fluid pressure p_f. The compaction pressure p_c is determined by a post-processing step if needed. We obtain existence and uniqueness of solutions, and we derived an a priori error estimate for equal-order finite elements. For the time-dependent case, we propose a splitting method which consists of an hyperbolic equation for the melt fraction φ and an elliptic problem for solid velocity and fluid pressure. The time-discretization for the melt fraction is carried out by explicit schemes (forward Euler or Heun method), so that the equation for velocity

and pressure decouples from the equation for the melt fraction, i.e. only information of previous time steps of $\mathbf{u}_s$ and p_f enters into the equation for φ. In numerical examples with known exact solutions we demonstrate that the expected convergence rates with respect to the spatial mesh size h and with respect to the time step k are obtained.

Acknowledgements Open Access funding provided by Projekt DEAL.

References

1. Discacciati, M., Miglio, E., Quarteroni, A.: Mathematical and numerical models for coupling surface. Appl. Numer. Math. **43**, 57–74 (2002)
2. Rivière, B.: Analysis of a discontinuous finite element method for the coupled Stokes and Darcy problems. J. Sci. Comput. **22**(1), 479–500 (2005)
3. Badea, L., Discacciati, M., Quarteroni, A.: Numerical analysis of the Navier–Stokes/Darcy coupling. Numer. Math. **115**(2), 195–227 (2010)
4. D' Angelo, C., Zunino, P.: Robust numerical approximation of coupled Stokes' and Darcy's flows applied to vascular hemodynamics and biochemical transport. ESAIM Math. Model. Numer. Anal. **45**(3), 447–476 (2011)
5. Wanga, Z.H., Wanga, C.Y., Chen, K.S.: Two-phase flow and transport in the air cathode of proton exchange membrane fuel cells. J. Power Sources **94**, 40–50 (2001)
6. Xie, X., Xu, J., Xue, G.: Uniformly-stable finite element methods for Brinkman–Stokes–Darcy Models. J. Comp. Math. **26**, 437–455 (2008)
7. Urquiz, J.M., Ndri, D., Garon, A., Delfour, M.C.: Coupling Stokes and Darcy equations. Appl. Numer. Math. **58**(1), 525–538 (2008)
8. Karper, T., Mardal, K.A., Winther, R.: Simplified finite element discretizations of coupled Darcy–Stokes flow. Numer. Methods Partial Differ. Equ. **25**(2), 311–326 (2009)
9. Pacquaut, G., Bruchon, J., Moulin, N., Drapier, S.: Combining a level-set method and a mixed stabilized P1/P1 formulation for coupling Stokes–Darcy flows. Internat. J. Numer. Methods Fluids **69**, 459–480 (2012)
10. Braack, M., Nafa, K.: A monolithic finite element discretization for coupled Darcy–Stokes flow. PAMM Proc. Appl. Math. Mech. **13**, 243–244 (2013)
11. Nafa, K.: Equal order approximations enriched with bubbles for coupled Stokes–Darcy problem. J. Comput. Appl. Math. **270**, 275–282 (2014)
12. McKenzie, D.: The generation and compaction of partially molten rock. J. Petrol. **25**(3), 713–765 (1984)
13. Schmeling, H.: Partial melting and melt segregation in a convecting mantle. In: Bagdassarov, N., Laporte, D., Thompson, A.B. (eds.) Physics and Chemistry of Partially Molten Rocks, pp. 141–178. Springer, Berlin (2000)
14. Gerya, T.: Future directions in subduction modeling. J. Geodyn. **52**(5), 344–378 (2011)
15. Sobolev, S.V., Sobolev, A.V., Kuzmin, D.V., Krivolutskaya, N.A., Petrunin, A.G., Arndt, N.T., Radko, V.A., Vasiliev, Y.R.: Linking mantle plumes, large igneous provinces and environmental catastrophes. Nature **477**(7364), 312–316 (2011)
16. Ballmer, M.D., Ito, G., Wolfe, C.J., Solomon, S.C.: Double layering of a thermochemical plume in the upper mantle beneath Hawaii. Earth Planet. Sci. Lett. **376**, 155–164 (2013)

48

17. Ruedas, T., Schmeling, H., Marquart, G., Kreutzmann, A., Junge, A.: Temperature and melting of a ridge-centred plume with application to Iceland. Part I: dynamics and crust production. Geophys. J. Int. **158**(2), 729–743 (2004)
18. Cagnioncle, A.-M., Parmentier, E.M., Elkins-Tanton, L.T.: Effect of solid flow above a subducting slab on water distribution and melting at convergent plate boundaries. J. geophys. Res. **112**(B9), (2007)
19. Schmeling, H., Marquart, G.: Crustal accretion and dynamic feedback on mantle melting of a ridge centred plume: the Iceland case. Tectonophysics **447**(1–4), 31–52 (2008)
20. Mittelstaedt, E., Ito, G., van Hunen, J.: Repeat ridge jumps associated with plume-ridge interaction, melt transport, and ridge migration. J. geophys. Res. **116**(B1), (2011)
21. Keller, T., May, D.A., Kaus, B.J.P.: Numerical modelling of magma dynamics coupled to tectonic deformation of lithosphere and crust. Geophys. J. Int. **195**(3), 1406–1442 (2013)
22. Rhebergen, S., Wells, G.N., Wathen, A.J., Katz, R.F.: Three-field block-preconditioners for models of coupled magma/mantle dynamics. SIAM J. Sci. Comput. **37**(5), 2270–2294 (2015)
23. Dannberg, J., Heister, T.: Compressible magma/mantle dynamics: 3-D adaptive simulations in ASPECT. Geophys. J. Int. **207**(3), 1343–1366 (2016)
24. Becker, R., Braack, M.: A finite element pressure gradient stabilization for the Stokes equations based on local projections. Calcolo **38**(4), 173–199 (2001)
25. Nafa, K., Wathen, A.J.: Local projection stabilized Galerkin approximations for the generalized Stokes problem. Comput. Methods. Appl. Mech. Engrg. **198**(5–8), 877–883 (2009)
26. Brooks, A.N., Hughes, T.J.R.: Streamline upwind/Petrov–Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier–Stokes equations. Comput. Methods Appl. Mech. Eng. **32**(1–3), 199–259 (1982)
27. Srámek, O., Ricard, Y., Bercovici, D.: Simultaneous melting and compaction in deformable two-phase media. Geophys. J. Int. **168**(3), 964–982 (2007)
28. Thomas, J.R., Hughes, T.J.R., Leopoldo, P., Franca, L.F., Balestra, M.: A new finite element formulation for computational fluid dynamics: V. Circumventing the babuška-brezzi condition: a stable Petrov–Galerkin formulation of the stokes problem accommodating equal-order interpolations. Comput. Methods. Appl. Mech. Eng. **59**(1), 85–99 (1986)
29. Matthies, G., Skrzypacz, P., Tobiska, L.: A unified convergence analysis for local projection stabilisations applied to the Oseen problem. M2AN Math. Model Numer. Anal. **41**(4), 713–742 (2007)
30. Ern, A., Guermond, J.-L.: Theory and Practice of Finite Elements. Springer, New York (2004)
31. Scott, L., Zhang, S.: Finite element interpolation of nonsmooth functions satisfying boundary conditions. Math. Comput. **54**(190), 483–493 (1990)
32. Zienkiewicz, O.C., Zhu, J.Z.: The superconvergent patch recovery and a posteriori error estimates. Part I: the recovery technique. Int. J. Numer. Meth. Eng. **33**, 1331–1364 (1992)
33. Schiemenz, A.R., Hesse, M.A., Hesthaven, J.S.: Modeling magma dynamics with a mixed Fourier collocation-discontinuous Galerkin method. Comm. Comput. Phys. **10**(2), 433–452 (2011)

Paper 2: Optimized Predator-Prey and MPA Based Fishing Strategies for the Baltic Sea

4

The original publication is available at
https://doi.org/10.1111/nrm.70000

Natural Resource Modeling

WILEY

Natural Resource Modeling

RESEARCH ARTICLE OPEN ACCESS

Optimized Predator-Prey and MPA Based Fishing Strategies for the Baltic Sea

Simon Taylor | Malte Braack

Department of Applied Mathematics, Christian-Albrechts-UniversitÃđt zu Kiel, Kiel, Germany

Correspondence: Simon Taylor (taylor@math.uni-kiel.de)

Received: 12 July 2024 | **Revised:** 23 January 2025 | **Accepted:** 28 January 2025

Keywords: Baltic Sea | Crowley-Martin function | marine protected areas | numerical simulation | optimal control | predator-prey model | stage structure

ABSTRACT

Marine protected areas (MPAs) have become the de-facto mechanism for marine species conservation and management and the ecological benefits of no take zones (NTZs) are supported by an expanding scientific interest in their effectiveness. However, in the south-western Baltic region there is a paucity of studies on the validity of MPAs, possibly due to the lack of enforcement. In this work we investigate the effects of nonconstant harvesting of a resource simultaneously under predation pressure. We use a functional response of the predator dependent Crowley-Martin type in age-structured state equations of a fixed time-horizon optimal control problem, wherein a pre-existing MPA may be encountered. Where available, only peer-reviewed data and biomass stock levels are used for biological parameters in the state equations. Our results for the prey component of the model show that a moderate penalty exists that yields the same benefits as a fully closed reserve. However, taking into account the whole system we observe that an NTZ benefits the predator species too.

1 | Introduction

Marine protected areas (MPAs) are increasingly recognized as essential tools for sustaining renewable fish biomass. Following the adoption of the EU Common Fisheries Policy in 1983, the EU committed to environmental goals like *Aichi Target 11*—aiming to protect 10% of biodiversity-rich coastal and marine regions (UN Convention on Biological Diversity 2012). While MPA designations have grown, effective implementation remains challenging. Issues such as permitting or ignoring trawling within reserves (Dureuil et al. 2018) and severe underreporting of total removals when relying solely on landings data (Zeller et al. 2011) undermine conservation efforts. In the Baltic Sea, for instance, a report by the Helsinki Commission (HELCOM 2013) found that only 53 of 163 MPAs address harvesting in their management plans, with just two fully prohibiting fishing.

With this in mind we incorporate an MPA into an optimization problem to investigate the spatial-temporal harvesting of a biomass based on the model proposed by Meng et al. (Meng et al. 2014), by extending it to two dimensions. We selected this model as a starting point for a number of reasons. It supports an age stage structure so we can consider unharvested immature and harvested mature components separately by the inclusion of one control term. Its stability properties have already been proven by (Meng et al. 2014; Hoang 2020). Finally, its response function is of the Crowley & Martin (Crowley and Martin 1989) type which accounts for predator interference, a desirable property when considering a piscivorous predator, such as Atlantic Cod that exhibits peer aggression (Cohen et al. 1990). Feeding rates decline with increasing predator density regardless of prey abundance. This interference affects feeding rates continuously, whether predators are handling or searching for prey.

Our extension employs time-dependent partial differential equations (PDEs) to describe the dynamics of mature predators and prey in the Baltic Sea. These equations incorporate spatial biomass movement via diffusion and advection, driven by an

Summary

- Our modeling work shows that the degree of penalization in MPAs may trigger different fishing strategies and lead to very different final fish stocks for predator and prey. When considering population models for fisheries management, the inclusion of the spatial domain to the model has been shown to be of crucial importance.
- Since most harvested marine ecosystems exist somewhere within a predator-prey hierarchy, these results lead us to recommend including a harvester-predator competition element in fisheries models whenever possible.
- Both the temporal and spacial trajectories of fishing should be subject to standardized monitoring for the fair imposition of penalties for intra-MPA harvesting. Therefore the territories surrounding the MPAs should pool reserve management resources to have an effective overview to provide feedback from monitoring.

external vector field. Additionally, a control term is introduced into the prey biomass equation to represent the harvesting rate, transforming the predator-prey system into a dynamic optimal control problem over a finite time horizon. Unlike many studies where reserves dynamically emerge based on problem parameters, this project examines a pre-existing reserve with fixed geometry and location, enabling simulations to focus on the duration and intensity of protection.

To provide meaningful gradients for the diffusive and advective terms, we simulate the presence of a marine sanctuary where biomass removal is discouraged within its boundaries through a penalty regime (e.g., increased access costs). The optimization problem is solved using real biological data where available, or otherwise with biologically plausible parameter estimates. We analyze the effects of varying the penalty for harvesting within the MPA on the predator-prey interplay. Our evaluation includes the spatial distribution of fish populations and the temporal dynamics of overall fish biomass, providing insights into the ecosystem-wide impact of MPA management strategies.

The paper is organized as follows: Section 2 introduces the controlled state equations, which include a dependency on a Stokes vector field, along with the objective function to be minimized. Additionally, the finite-element implementation and the relevant stabilization terms are thoroughly explained. Section 3 details the finite-element mesh, initial conditions, a justification for the chosen biological parameter values supported by peer-reviewed sources, and summarizes them in Table 2. In Section 4 we present numerical results for both non-penalized (open MPA) and penalized harvesting conditions, exploring a wide range of penalty parameters. This approach highlights qualitatively distinct outcomes under varying levels of harvesting restriction.

2 | Definition of the Optimal Control Problem

The design of our optimization problem is underpinned by a large body of existing work on the topics of population models and harvesting. The approach we take is informed most notably by the following studies.

2.1 | Literature Overview

Neubert (Neubert 2003) introduced a dimensionless length parameter for a one-dimensional logistic-growth reaction-diffusion model with Dirichlet boundary conditions. Without predefined reserves, he identified an optimal harvesting strategy maximizing steady-state biomass, revealing emergent no-take reserves. For small domains, reserves centered in the habitat maximized biomass; for large domains, reserves formed a network with maximum harvesting in between. Kurata & Shi (Kurata and Shi 2008) extended this model by minimizing a biological energy functional, assuming a given total global fishing rate and determined its local distribution. They demonstrated that a no-effort subdomain always exists, constrained by maximal and average efforts, and derived the minimal patch size for species survival under habitat loss. Neubert & Herrera (Neubert and Herrera 2008) analyzed the Gordon-Schaefer model to explore conditions where reserves enhance rent, stock, and employment simultaneously, contrary to traditional open-access outcomes. Ding & Lenhart (Ding and Lenhart 2009) generalized Neubert's work to stationary multi-dimensional systems, deriving an optimality system and exploring domain scaling on fish density, harvesting, and costs via numerical methods. Joshi et al. (Joshi et al. 2009) presented a dynamic extension with diffusion and advection, studying how parameters like initial stock, time horizon, and advection influence reserve formation and placement. Kelly et al. (Kelly et al. 2016) examined boundary condition choices in nonlinear parabolic PDE models with a spatially distributed control. Braack et al. (Braack et al. 2018) incorporated sustainability into fisheries models, transforming the optimal control problem into a non-convex form, and proved solution existence and optimality conditions. Upmann & Gromov (Upmann and Gromov 2023) analyzed finite and infinite time horizon models, detailing scenarios where initial stock influences effort trajectories, with implications for long-term sustainability. Predator-prey models emphasize coupling populations through predation rates, with classic Lotka-Volterra equations introducing nonlinear dynamics. Holling (Holling 1959) expanded on this model with dynamic functional responses (types I, II, III), later enhanced to include predator dependence by DeAngelis et al. (DeAngelis et al. 1975), Beddington (Beddington 1975), and others. Empirical approaches by Arditi & Ginzburg (Arditi and Ginzburg 1989) and Essington & Hansson (Essington and Hansson 2004) refined these responses, incorporating predator-prey size interactions (e.g., Aljetlawi et al. (Aljetlawi et al. 2004)). Crowley & Martin (Crowley and Martin 1989) introduced a pre-emption model integrating interference and prey handling, suited for species with peer aggression. Skalski & Gilliam (Skalski and Gilliam 2001) validated interference effects on consumption, rejecting Holling's type II model in favor of Crowley-Martin in most tests. This function has since gained traction in multi-species models, focusing on system dynamics, equilibria, and stability (Ali and Jazar 2013; Sambath et al. 2013; Meng et al. 2014; Chen et al. 2018). Among these, Meng et al. (Meng et al. 2014) incorporated age structure, later shown by Hoang (Hoang 2020) to have an asymptotically stable positive equilibrium under certain conditions.

2.2 | State Equations

Let $u_i = u_i(x, t)$ denote the population biomass of juvenile prey and $u = u(x, t), v = v(x, t)$ the mature prey and predator biomasses respectively in the smooth bounded domain $\Omega \subset \mathbb{R}^2$ for time $t \in [0, T]$. Further, let the subdomain $\Omega_{\text{MPA}} \in \Omega$ designate the marine reserve. The age-structured trophic interaction model is then described by the coupled initial value advection-diffusion problems,

$$\partial_t u_i = \gamma u_i \left(1 - \frac{u_i}{K}\right) - d u_i. \tag{1}$$

$$\partial_t u = \epsilon_u \Delta u + d u_i - f(u, v)v - d_1 u - \mathbf{b} \cdot \nabla u - qu. \tag{2}$$

$$\partial_t v = \epsilon_v \Delta v + \delta f(u, v)v - d_2 v. \tag{3}$$

$$\nabla u \cdot \mathbf{n} = \nabla v \cdot \mathbf{n} = 0 \quad \text{on}(0, T) \times \partial\Omega. \tag{4}$$

$$\text{where} \quad f(u, v) = \frac{\omega u}{(1 + \omega a u)(1 + bv)}. \tag{5}$$

The right hand side of (1) shows the rate of change of the juvenile population is determined by a standard logistic term with carrying capacity K, and a negative term due to maturity recruitment signified by rate d. The mature prey component (2) is formulated as a PDE to account for movement due to diffusion via the Laplacian operator Δ, and advection due to ocean currents, signified by vector $\mathbf{b}$. The prey has in addition the combined negative effects of the predator feeding rate function $f(u, v)$, a natural mortality rate d_1, and the harvesting term qu we wish to optimize via the spatial-temporal control variable $q(x, t), \forall\, x \in \Omega_T$ and finite time $t \in [0, T]$. The predator density equation (3) is simplified in comparison to (1) and (2) based on two assumptions: the influence of low velocity currents on strong swimming predators produces a negligible advective affect, and mortality due to bycatches is minimal due to highly focused fishing practices with the aid of fish-finding sonar devices, and selective fishing equipment. The natural (no-flux) Neumann boundary conditions (4) mean that individuals cannot escape the domain through $\partial\Omega$, the coastline boundary. The MPA does not explicitly feature in the equation system (1)–(5), but it is incorporated into the cost function to be optimized, as described below.

The advective term $\mathbf{b} \cdot \nabla u$ in Eq. (2) represents the influence of ocean currents on the mature prey. In contrast, stochastic movement is modeled by the diffusive term. The parameter ϵ_u governs the balance between these two modes of movement. The variable $\mathbf{b}$, representing ocean currents, which introduces advection into the state equations is a divergence free stationary Stokes flow field given by

$$-\mu \Delta \mathbf{b} + \nabla p = 0. \tag{6}$$

$$\nabla \cdot \mathbf{b} = 0. \tag{7}$$

$$\mathbf{b} = \mathbf{b}_{\text{in}}(x) \quad \text{on } \Gamma_{\text{in}}. \tag{8}$$

$$\mu \frac{\partial \mathbf{b}}{\partial n} - p\mathbf{n} = \mathbf{0} \quad \text{on } \Gamma_{\text{out}}. \tag{9}$$

$$\mathbf{b} = \mathbf{0} \quad \text{on } \partial\Omega \backslash (\Gamma_{\text{in}} \cup \Gamma_{\text{out}}). \tag{10}$$

At the inflow boundary Γ_{in} (8), a parabolic velocity function $\mathbf{b}_{\text{in}}(x)$ is specified to prevent discontinuities with the adjacent Dirichlet boundary elements in condition (10). The outflow Γ_{out} (9) is then left as a "do-nothing" artificial boundary in the normal direction $\mathbf{n}$. The choice of elementary Stokes flow, in contrast to a full Navier-Stokes treatment, is justified by a lack of any macro-scale or localized turbulence due to very slow flow rates. Averaged flow velocities as monitored by hydrographic agencies (BSH 2023; IOW 2023) show a generally mild variance of low velocities between 0.1 and 0.3 m.s^{-1} ensuring low Reynolds number laminar flows that are stable throughout the entire Baltic. All parameters for the population and flow equations are summarized in Table 1.

2.3 | Cost Function

The objective function $J(q, u)$ we seek to minimize for the state u and control variable q in equation (2) is a modified version of that introduced by Braack et al. (Braack et al. 2018),

$$J(q, u) := \int_0^T e^{-\rho t} \int_\Omega q(r - u)\mathrm{d}x\mathrm{d}t - \lambda \int_\Omega u(T)\mathrm{d}x + \frac{\alpha}{2} \int_0^T \int_\Omega q^2 \mathrm{d}x\mathrm{d}t. \tag{11}$$

TABLE 1 | Parameters and units for the state and Stokes flow equations. The most suitable units applicable are: mass in kilotons, length in kilometers, and time in years. All parameters are strictly positive with the exception of a and b which have only a non-negative constraint.

Parameter	Description	Units
ϵ_u, ϵ_v	Biomass diffusivity	$\text{km}^2\ \text{year}^{-1}$
K	Environment carrying capacity	ktons
d	Transition rate to mature prey	year^{-1}
γ	Intrinsic growth of immature prey*	year^{-1}
d_1	Natural mortality of mature prey	year^{-1}
d_2	Maximum mortality of mature predators**	year^{-1}
δ	Predator conversion coefficient	Dimensionless
ω	Predator attack rate	year^{-1}
a	Handling time	hours
b	Interference amongst predators	Dimensionless
μ	Salt water viscosity	N.s.m^{-2}

*Bounded by $\gamma > d$ to ensure asymptotic stability in the model Jacobian (Meng et al. 2014).
**By *maximum* we mean the maximum of natural mortality M and fishing mortality F as defined by ICES.

We explain the arising terms as follows:

- The product qr represents the costs associated with harvesting a remote resource (see, e.g., [Sirén and Parvinen 2015; Upmann and Behringer 2020]), where $r = r(x)$ is a location-dependent cost factor. This factor includes the travel costs from the home harbor at x_0 to the specific harvesting location x, calculated as the Euclidean distance $\|x - x_0\|_2$. Additionally, penalty terms for entering the MPA are incorporated into r. Denoting the characteristic function of the MPA region by χ, we incorporate a penalization factor σ by multiplying it with χ. The penalization factor σ is a specified value determined locally by resource managers for fishing vessels entering Ω_{MPA}. Consequently, the cost factor r is defined as:

$$r = r(x) := \|x - x_0\|_2 + \sigma\chi(x), \qquad \chi(x) = \begin{cases} 1 & x \in \Omega_{\text{MPA}} \\ 0 & \text{otherwise} \end{cases}. \tag{12}$$

- The fishery revenue is represented by the term $-qu$, the product of fishing activity and the present biomass of mature prey. The negative sign reflects the goal of minimizing J.
- The exponential term serves as a discount factor with non-negative discount rate ρ, determining the impact of future rewards on J. All terms are integrated over location x and time t.
- The integral over $u(T)$ represents the remaining non-harvested biomass at terminal time T. The coefficient λ accounts for a surplus for future harvesting and emphasizes sustainable practices.
- Finally, the integral over the quadratic term q^2 is a standard regularization term in the cost functional, a so-called L^2-Tikhonov regularization (see e.g. Gerth [Gerth 2021]). The factor α preceding the integral is a small positive parameter, depending on the minimum mesh size h, typically given by $\alpha \approx h^2$.

The complete optimization problem is then: minimize objective functional (11) subject to the PDE constraints (1)–(4) and the pointwise control constraint from the set of admissible controls

$$Q_{ad} := \{q \in L^\infty(\Omega_T) : 0 \le q \le q_{\max}\}. \tag{13}$$

In relation to state equation (2) the condition $q = 0$ signifies no fishing allowed, and $q = q_{\max}$ is the maximum permitted harvesting rate as determined by fishing regulations. The control q is strictly non-negative due to the assumption that there is no artificial restocking of fish.

Our optimal control problem is to determine the best strategy q of harvesting under the restriction that the biomass of mature prey u solves a PDE system for the spatial-temporal control variable q:

$$q^* = \arg\min_{q \in Q_{ad}} \{J(q, u) : u \text{ is solution of (1)–(4)}\}.$$

2.4 | Discretization and Stabilization

To minimize run times of the optimal control solver, the Stokes velocity field $\mathbf{b}$ is generated independently as a stationary LPS-stabilized (Braack and Becker 2001) equal-order finite element solution. This constant vector is then re-used in each iteration of the optimal control solver.

Following an optimize-then-discretize approach, see for example Heinkenschloss & Leykekhman (Heinkenschloss and Leykekhman 2010), the state system (1)–(3) is discretized according to a standard Galerkin recipe, given test functions $\phi \in L^2(0, T; V)$, where $V := H^1(\Omega)$ is the usual notation denoting the Sobolev space with first (weak) derivatives in the $L^2(\Omega)$ space. Hence, u is a L^2-function in time and $u(t) \in V$ for all $0 < t < T$. The bilinear forms are then,

$$\begin{aligned} a_1(u_i, \phi) &= d\int_\Omega u_i\phi \mathrm{d}x - \gamma\int_\Omega u_i\left(1 - \tfrac{u_i}{K}\right)\phi \mathrm{d}x \\ a_2(u_i, u, v, q, \phi) &= \epsilon_u\int_\Omega \nabla u \cdot \nabla\phi \mathrm{d}x - d\int_\Omega u_i\phi \mathrm{d}x \\ &+ \int_\Omega f(u, v)v\phi \mathrm{d}x + d_1\int_\Omega u\phi \mathrm{d}x + \int_\Omega(\mathbf{b} \cdot \nabla u)\phi \mathrm{d}x + \int_\Omega qu\phi \mathrm{d}x \\ a_3(u, v, \phi) &= \epsilon_v\int_\Omega \nabla v \cdot \nabla\phi \mathrm{d}x - \delta\int_\Omega f(u, v)v\phi \mathrm{d}x + d_2\int_\Omega v\phi \mathrm{d}x. \end{aligned} \tag{14}$$

With the notation $U := (u_i, u, v)$ and $\Phi := (\phi_1, \phi_2, \phi_3)$ we define

$$A(U, q; \Phi) := a_1(u_i; \phi_1) + a_2(u_i, u, v, q; \phi_2) + a_3(u, v; \phi_3).$$

The control q is discretized as cell-wise constants, that is, by P_0 finite elements on the same mesh. Further, let $V_h \subset V$ be the space of bilinear finite elements on quadrilateral meshes and the discretized control space $Q_h \subset Q_{ad} \cap P_0$. It is well-known that the convective term $(\mathbf{b} \cdot \nabla u, \phi)$ in the semilinear form a_2 gives rise to instabilities if this term becomes dominant in comparison to the diffusive term. To be more specific, instabilities begin to appear if $h|\mathbf{b}| > 1/\epsilon_u$, where h is the mesh size. To prevent such instabilities, numerical dissipation should be included. Among the best methods, with low numerical dissipation, is the Streamline-upwind-Petrov-Galerkin (SUPG) method, see Johnson (Johnson 1985). This method adds the following terms to the Galerkin formulation:

$$S_h(U, q; \phi) := \sum_T \delta_T(\mathbf{b} \cdot \nabla u - \epsilon\Delta u - du_i + f(u, v)v + d_1u + qu, \mathbf{b} \cdot \nabla\phi)_T.$$

It is a sum over the elements T, and on each element T it consists of the product of the residual of the equation, multiplied by the streamline derivative of the test function, $\mathbf{b} \cdot \nabla \cdot \phi$, weighted by a stabilization parameter $\delta_T = \delta_0 h_T$ and a constant $\delta_0 > 0$. The stabilized semi-linear form now reads

$$A_h(U, q; \Phi) := A(U, q; \Phi) + S_h(U, q; \phi_2).$$

The semi-discrete state equation (continuous in time, discrete in space) reads:

$$U \in H^1\left(0, T; V_h^3\right): \quad \frac{\partial U}{\partial t} + A_h(U, q; \Phi) = 0 \quad \forall\, \Phi \in V_h^3, \tag{15}$$

and initial condition $U(0) = U_0$. Note that this form of stabilization is not consistent in the sense that the strong solution U and control q do not satisfy the discrete equations. This is due to the missing temporal derivative in S_h. The incorporation of the temporal derivative would lead to a consistent method and with a better a priori error estimate. However, the discretization error is in our case several magnitudes smaller than the modeling error. Therefore, this inconsistency is irrelevant in terms of accuracy. However, without the temporal derivative in S_h, the scheme is much easier to solve. Our Lagrange functional $\mathcal{L} : V_h^3 \times Q_h \times V_h^3 \to \mathbb{R}$ is defined as

$$\mathcal{L}(U, q, Z) \coloneqq J(U, q) + A_h(U, q; Z).$$

To simplify differentiability of the first term of the Lagrangian (as well as the numerical implementation), we reformulate the objective functional (11) into a three-component reduced cost functional as described by Becker et al. (Becker et al. 2007): $J(u_i, u, v, q) = J_1(u_i, u, v) + J_2(u(T)) + J_3(q)$. To eliminate the

$$-\int_0^T e^{-\rho t} \int_\Omega qu \, \mathrm{d}x\mathrm{d}t,$$

term in (11) we multiply the state equations (1)–(3) with $e^{-\rho t}$ and integrate over Ω_T using partial integration in t, giving

$$\begin{aligned} J(u_i, u, v, q) &= J_1(u_i, u, v) + J_2(u(T)) + J_3(q) \\ &= \int_0^T \int_\Omega e^{-\rho t}\left\{\rho u + d_1 u - d u_i + f(u, v) + \mathbf{b} \cdot \nabla u\right\} \mathrm{d}x\mathrm{d}t \\ &\quad + \int_\Omega \left\{e^{-\rho T} u(T) - \lambda u(T) - u_0\right\} \mathrm{d}x \\ &\quad + \int_0^T \int_\Omega \left\{e^{-\rho t} q r + \tfrac{\alpha}{2} q^2\right\} \mathrm{d}x\mathrm{d}t. \end{aligned}$$

For the formulation of the adjoint equation, we introduce $DA_h(U, q)$ that is the derivative of A_h w.r.t. U at point (U, q). With this notation, the adjoint equation reads

$$\begin{aligned} Z = (z_i, z_u, z_v) \in H^1\left(T, 0; V_h^3\right): -\tfrac{\partial Z}{\partial t} + DA_h(U, q)(\Phi, Z) \\ = -e^{-\rho t} q \phi_2 \quad \forall\, \Phi = (\phi_1, \phi_2, \phi_3) \in V_h^3, \\ Z|_{t=T} = -\lambda, \end{aligned} \tag{16}$$

where the right hand side of (16) are the derivatives of J_1 w.r.t. U and J_2 w.r.t. $U(T)$, respectively. The third optimality condition is then the variational inequality,

$$\int_0^T \int_\Omega (e^{-\rho t}(r - u) + \alpha q - u z_u))(p - q) \mathrm{d}x\mathrm{d}t \geq 0 \quad \forall\, p \in Q_h. \tag{17}$$

The complete optimality system is then fully defined by the primal system (15), the dual equation (16), and the variational inequality (17). For further details on the derivation of this optimality system and reduced cost functional, see for example Braack et al. (Braack et al. 2018) and Becker et al. (Becker et al. 2007). The implicit Euler method is employed for time discretization and the complete optimization system is solved by a semi-smooth Newton method and Trust Region iteration.

3 | The Domain, Model Parameters and Initial Conditions

3.1 | Computational Domain

The southeastern quadrant of the Baltic serves as the domain and is discretized as a shape regular mesh of 5 × 5 km Q_1 elements. See Figure 1 showing the domain boundary and other key features.

3.2 | Determining Model Parameters

As a species choice for an application of the model we select the prey u_i and u as *Clupea harengus* (herring). As one of the most commercially important species within the northern hemisphere, it is heavily fished. It is however robust, recovers rapidly from population crashes and is currently listed as Least Concern on the IUCN Red List (Herdson and Priede 2010). Although Baltic herring are a distinct subspecies of Atlantic herring, they are biologically similar enough to be considered equally. For a predator, we select *Gadus morhua* (cod) as it is considered an apex predator in its local environment and is also of interest as another harvested species. Where possible we draw mainly from ICES data from western Baltic subdivisions 20–24 (ICES 2019a, b) for the selection of biological parameters.

3.2.1 | *Clupea harengus*

As a highly shoaling and schooling fish, their rate of diffusion out of the shoal can be very low. Due to a lack of peer reviewed data or other applicable experiments, likely due to the difficulty of tagging and monitoring such small fish (unlike the larger gadids), we have to assume a biologically sensible diffusion coefficient.

The growth rate γ for immature herring was taken as the average from samples in Fishbase (Froese and Pauly 2021) which are in turn the growth parameter extracted from a fit to the von Bertalanffy growth function.

The recruitment coefficient d is well documented in the ICES herring recruitment data (ICES 2019b), as well as Fishbase.

3.2.2 | *Gadus morhua*

The observations of cod swimming speed on which a diffusion constant could be based vary by some orders of magnitude. Løkkeborg (Løkkeborg 1998) individually tagged cod with acoustic transmitters and recorded their roaming activity in a North Sea fjord. He measured an area they covered at 99–226 $m^2\ hr^{-1}$. Taking the low point of that range gives a

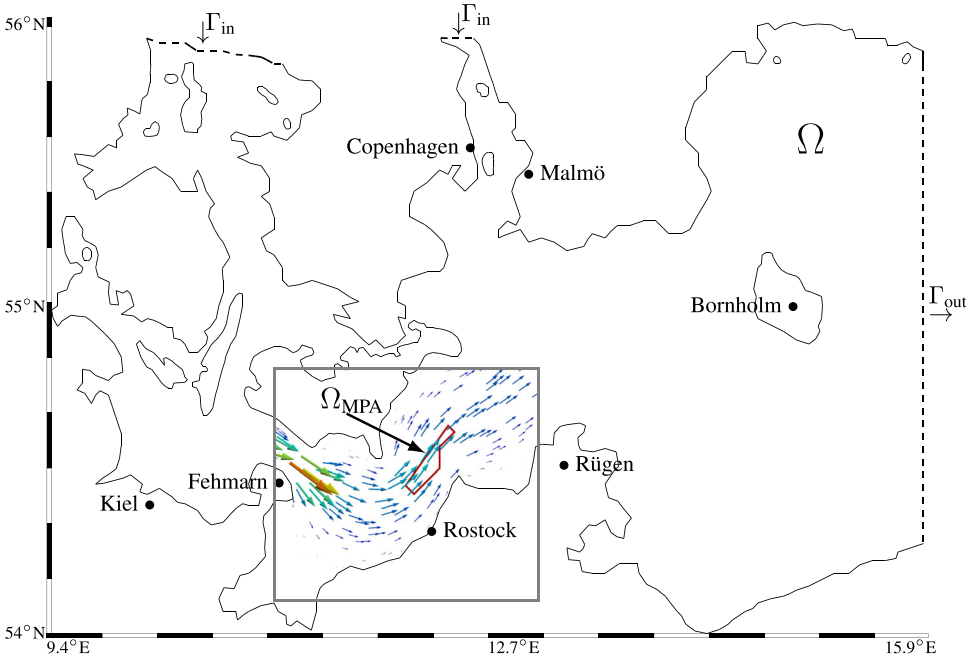


FIGURE 1 | The computational domain encompassing ICES subdivisions 20–24, discretized by 5×5 km Q_1 finite elements, with the outflow along the entire eastern dashed edge, two inlets and an otherwise closed boundary. The inset shows a subsection of the flow field **b** through MPA232.

diffusion rate of $\epsilon_v \approx 0.9$ km^2 year^{-1}. However, he only managed to obtain reliable data from eight fish and the value also varies considerably due to the time of day and other foraging activity. The FAO Species Catalog (Cohen et al. 1990) also contains a wide range of swimming data that could provide bounds on a diffusion coefficient.

We chose the mortality rate d_2 as the maximum of the natural (M) as observed by studies such as (Froese and Pauly 2021) (M = 0.26), (Sinclair 2001) (M = 0.4), and fishing effort (F) as documented by ICES (ICES 2019a) to reflect the current depressed state of the Gadid biomass in the Baltic today which is currently listed on HELCOM's Red List as endangered due to overfishing. In addition, serious environmental pressures, such as eutrophication, contribute irregularly to their mortality.

The conversion coefficient δ, which represents the efficiency with which the consumed biomass is converted into somatic growth, is also well researched, albeit predominantly in controlled aquaculture studies. These studies in which environmental parameters are controlled and/or monitored, see for example Andersen & Riis-Vestergaard (Andersen and Riis-Vestergaard 2003), Bjornsson & Steinarsson (Bjornsson and Steinarsson 2002) and Bethke et al. (Bethke et al. 2013), all point out that the feed conversion ratio is sensitively temperature dependent. Field studies, such as those conducted by Neuenfeldt (Neuenfeldt et al. 2019), using stomach content data, found a reasonable fit for the net conversion efficiency κ. Fishbase (Froese and Pauly 2021), and another database compiled by Brey et al. (Brey et al. 2010), provide further resources for determining appropriate conversion values. For our application we take the midpoint between the lowest and highest published values we could find.

For the handling time a, we interpret it as defined originally by Crowley & Martin (Crowley and Martin 1989) to mean the time spent processing the prey after capture, that is, the digestion time until renewed hunger instigates a renewed attack. Digestion studies, see for example Knutsen & Salvanes (Knutsen and Salvanes 1999), discuss the stages of prey processing and also the time $H_d = 0.2t_{\max}$ (hours), which we incorporate here, to be the time taken to consume the prey and resume feeding before the entire prey bolus is digested. They also point out the independence of predator size found by dos Santos & Jobling (dos Santos and Jobling 1991), who conducted a digestion study involving a wide size range of herring fed gadids.

Attack rate ω and predator interference b values for gadids preying on clupeids have been established by Essington & Hansson (Essington and Hansson 2004), who broke the available herring data down into five age classes and achieved a model fit based on the attack rate/predator dependence model of Hassell & Varley (Hassell and Varley 1969). We also note the prior contribution of Arditi & Akçakaya (Arditi and Akçakaya 1990), also based on the Hassell & Varley model, who noted the tendency to underestimate the level of predator dependence when fitting to a linear functional response. From the work of Essington & Hansson, we make use of their attack rate a and interference θ for herring and cod age classes 3.

Table 2 summarizes the values selected for all variables in the numerical model. As a home port, we set the x_0 location to Rostock, due to it having the highest capacity fishing fleet in the region (Edebohls et al. 2022).

3.3 | Initial Data

Initial conditions for the three populations were taken from the 2010 ICES Baltic Sea stock assessments (ICES 2019b). We used $u_0 = 100$ ktons from the spawning stock biomass (SSB) estimate. Based on the same stock data and the age ratio $u_i \approx 0.1u$

shown by Grygiel & Wyszyński (Grygiel and Wyszyński 2002) we set $u_{i0} = 10$ ktons uniformly distributed throughout the domain. For the initial cod biomass, we took $v_0 = 15$ ktons, the SSB from the corresponding cod assessment (ICES 2019a).

The control constraint, which represents the highest permissible rate of extraction to prevent population collapse, was set to $q_{\max} = 0.35\ \text{year}^{-1}$ as taken from the ICES herring stock assessment (ICES 2019b) $\mathrm{F}_{\mathrm{MSY}}$ value, defined as the fishing mortality consistent with achieving a maximum sustainable yield (MSY). For an interesting review summarizing the MSY metric, see Quinn (Quinn 2003). As a reserve, we select MPA232 (Kadetrinne), a 100 km^2 patch located on the western edge of the Arkona Basin, and designate it as a subdomain by Ω_{MPA}. This particular MPA is chosen out of practicality for a numerical simulation; the traveling costs term $r(x)$ in the cost functional is the straight-line distance from the origin port, and all of MPA232 is reachable from Rostock. Formulating the r-term on a curved path through out the domain would require a computationally expensive pathfinding algorithm. MPA232 is classified as *managed* (HELCOM 2013), in the sense that landings have to be reported and private fishing is forbidden. In this MPA, regulatory measures are currently undergoing a complex coordination process at both the national and European levels (der Justiz 2022). For the regularization parameter on this mesh scale, we set $\alpha = 10^{-5}$ and the salt water viscosity $\mu = 1.25 \times 10^{-3}\ \text{N.s.m}^{-2}$ at 5°C.

TABLE 2 | Parameter values applied to the implementation of the state equations.

Parameter	Description	Value	Source
ϵ_u	Prey diffusivity	$0.5\ \text{km}^2\ \text{year}^{-1}$	No data
ϵ_v	Predator diffusivity	$0.9\ \text{km}^2\ \text{year}^{-1}$	Løkkeborg (1998)
d	Transition rate to mature prey	$0.25\ \text{year}^{-1}$	Froese and Pauly (2021), ICES (2019b)
γ	Intrinsic growth of immature prey	$0.3\ \text{year}^{-1}$	Froese and Pauly (2021), Pauly (1978), Cohen et al. (1990)
d_1	Natural mortality of mature prey	$0.22\ \text{year}^{-1}$	Froese and Pauly (2021)
d_2	Maximum mortality of mature predators	$0.5\ \text{year}^{-1}$	ICES (2019a), Sinclair (2001)
δ	Predator conversion coefficient	0.3	Neuenfeldt et al. (2019), Bethke et al. (2013)
ω	Predator attack rate	$0.2\ \text{year}^{-1}$	Essington and Hansson (2004)
a	Handling time	2.05×10^{-3} years	Knutsen and Salvanes (1999)
b	Interference amongst predators	0.6	Essington and Hansson (2004)
K	Environment carrying capacity	200 ktons	No data

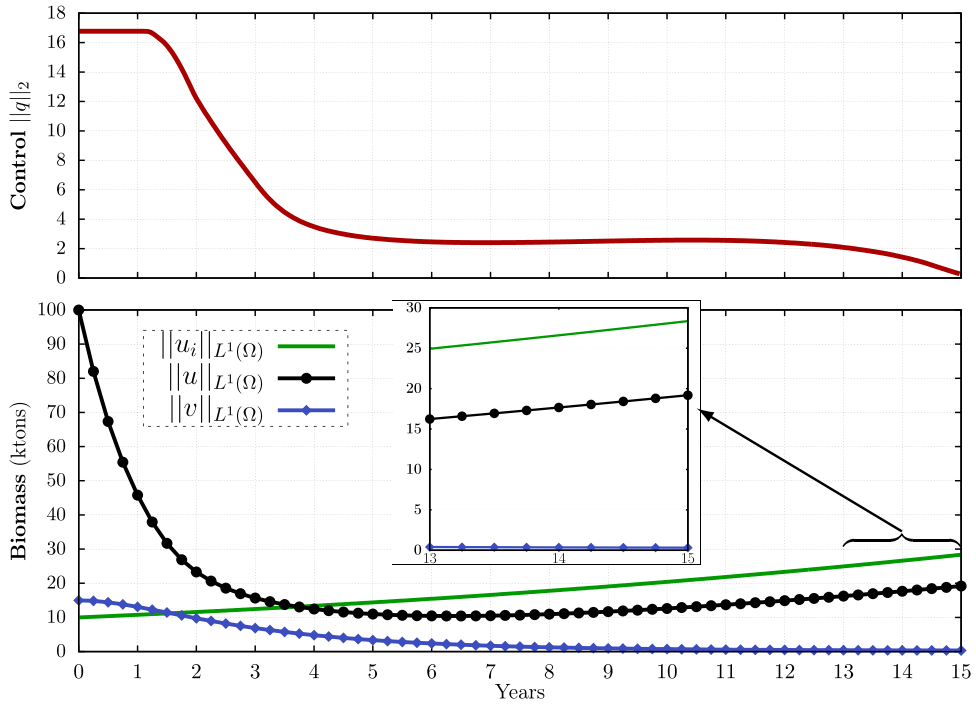


FIGURE 2 | The optimal solutions to the state equations (1)–(3) subject to cost functional (11) with an open MPA condition $\chi(x) = 0$.

4 | Results

4.1 | Nonpenalized Optimal Control Problem

Penalty σ in equation (12) is set to zero, forcing $r(x)$ in the cost functional to consist only of transport costs, that is, $r(x) = \|x\|_2$, the direct distance to the fishing port. The population evolution was simulated for 15 years; long enough to allow the population densities to reach a stable equilibrium, shown in Figure 2. Evident is an initial ($0 \leq t \lesssim 1.25$ years) maximum fishing rate ($q = q_{\max}$) due to initially high levels of u and a steady-state for u in the range $t \in [\sim 6, \sim 7]$ years. Moreover, a continuous reduction of the cod population can be observed, supposably due to the high fishing activity and subsequent shortage of prey. We

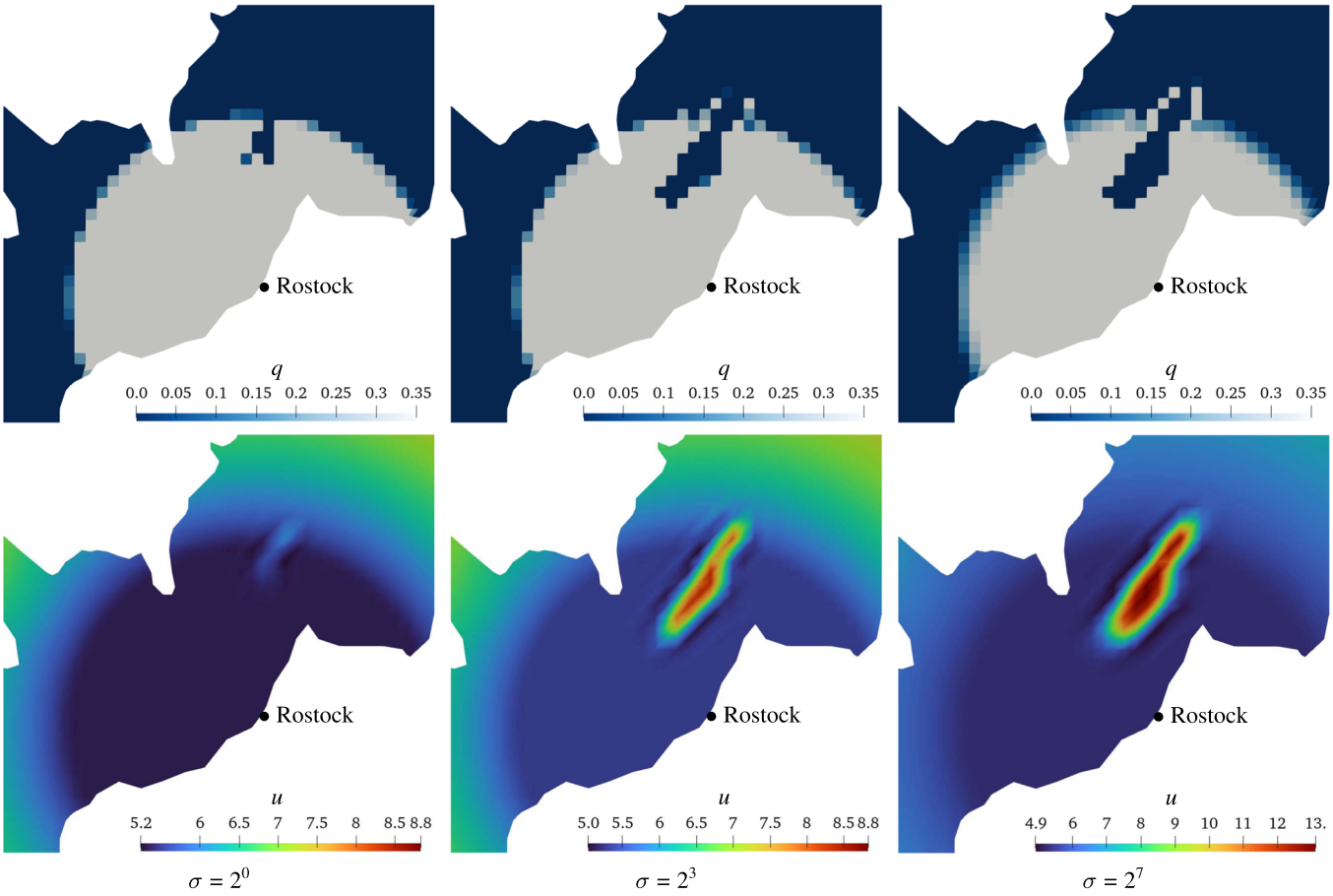


FIGURE 3 | The control and state solutions at $t = 2.75$ years showing the emergence of MPA232 (Kadetrinne) as penalty σ increases. A maximum fishing rate (high q) also emerges around the perimeter of Ω_{MPA}. The radial gradient centered at Rostock with higher q values, and correspondingly reduced biomass u, is a result of the vessel transport costs in the cost functional.

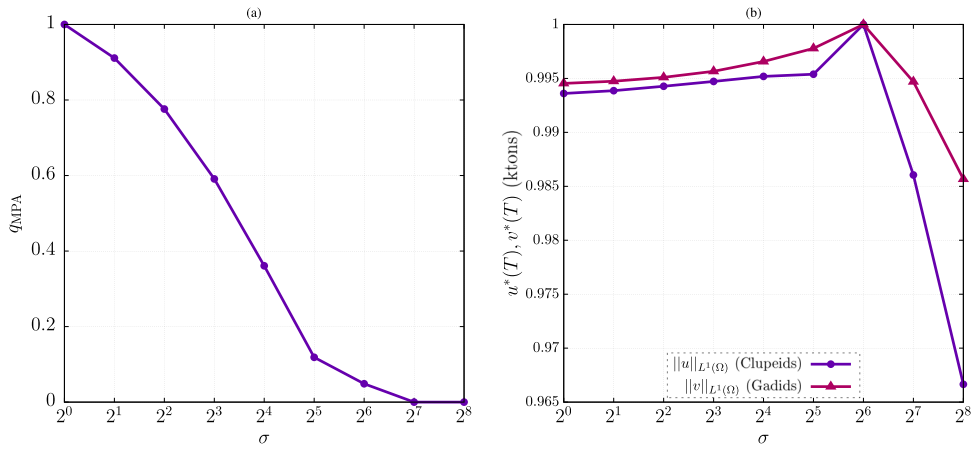


FIGURE 4 | The changing control q_{MPA} and optimal solutions. The plots show the impact of the penalty parameter σ on the fishing rate inside the MPA, q_{MPA} (a), and the resulting biomass of herring u^* and cod v^* (b) at $T = 6$ years and $\rho = 10^{-5}$. For $\sigma \geq 2^7$, the onset of a no take zone can be observed.

also like to mention the analysis by Hutchings (Hutchings 2000), who showed there is little evidence of rapid gadid recovery and they may still not experience recovery as much as 15 years after 45%–99% declines.

4.2 | Penalized Fishing Strategy

To simulate the enforcement of a penalty regime, parameter $r(x)$ was reenabled to contain both the transport and penalty costs by setting $\sigma > 0$ in equation (12). The penalty was then increased exponent-wise within the range $\sigma \in [2^0, 2^8]$. The remaining parameters remain consistent with the nonpenalized version above.

Figure 3 shows the control and state solutions on the mesh in the vicinity of Ω_{MPA} at an early time point, showing how the MPA appears due to addition of the σ term in (12). Observe that the MPA first emerges from the furthest northern point from Rostock where the penalty applies, and expands southwards. This is due to the furthest point having higher transport costs, leading to lower fishing and higher prey stock u. To further investigate the effect of penalization on rate q, we define a local variable q_{MPA} as the integral over all the finite-element cells contained within Ω_{MPA} over the total run time,

$$q_{\text{MPA}} := \int_0^T \int_{\Omega_{\text{MPA}}} q(x,t)\,\mathrm{d}x\mathrm{d}t. \tag{18}$$

To account for the case where a cell overlaps the boundary of the MPA, we only considered cells that had at least three out of four Q_1 element vertices contained within the MPA.

Figure 4A shows the effect of the penalty parameter σ on the fishing rate q_{MPA} within Ω_{MPA} (left plot) and the corresponding normalized optimal solutions for the biomass (u^* and v^*) (right plot) at the time horizon T. The biomass shows a peak at $\sigma = 2^6$. At higher penalties, q_{MPA} is decreasing to zero (i.e. a closed reserve), but it also leads to reduction of prey due to predator

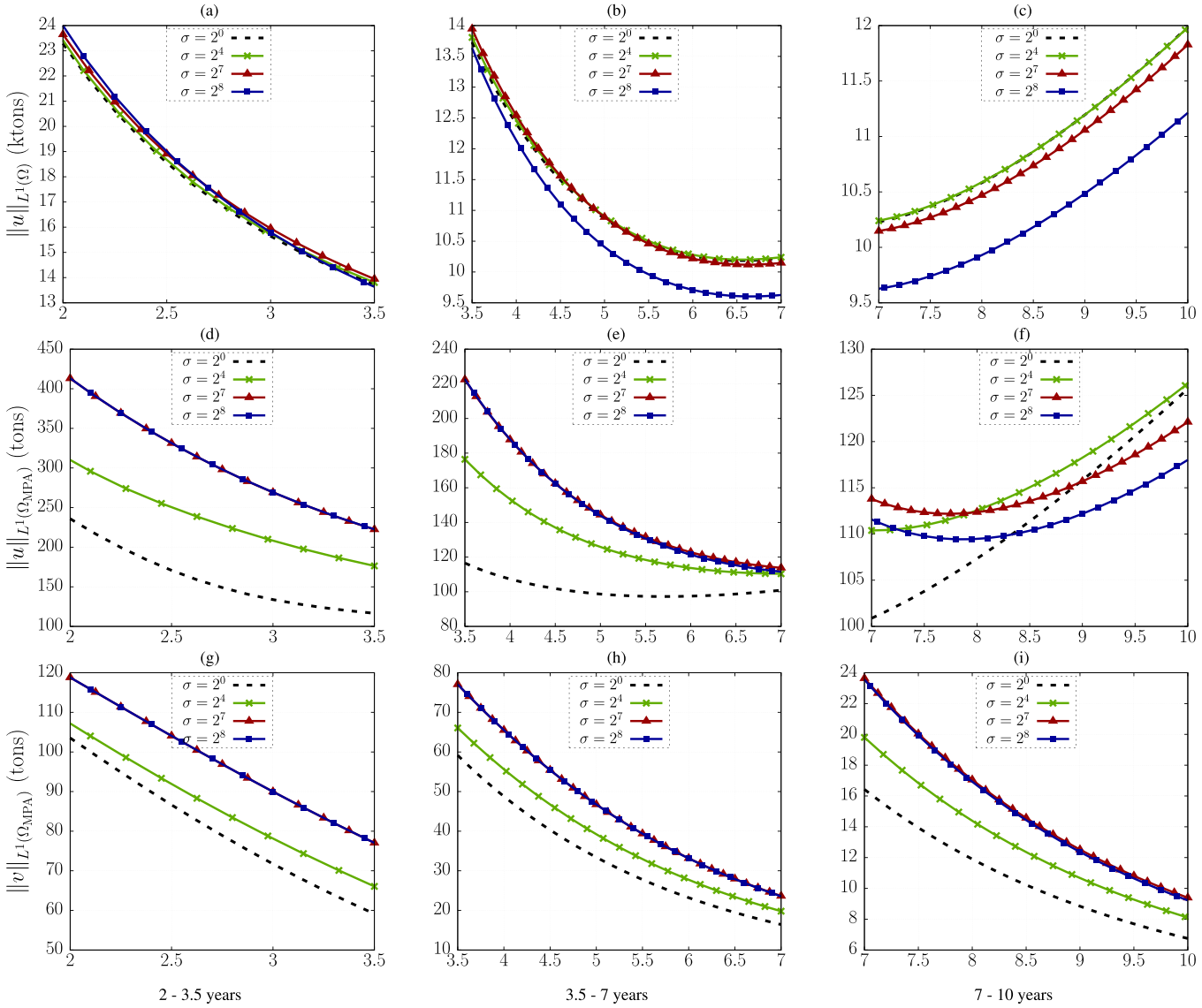


FIGURE 5 | (a–i) Evolution of the penalized fishing problem for three time intervals of interest under five penalty regimes. The top row shows the total prey biomass for the entire domain. The bottom two rows show the prey u and predator v biomasses respectively, contained within the MPA.

dominance. The small difference between the maximum and minimum can be explained by the relatively small area of Ω_{MPA} in relation to the whole domain.

The temporal dynamics of the biomass for different penalty values σ are shown in Figure 5. The top row (A–C) depicts the overall biomass of prey inside the entire domain Ω. For all values of σ, the biomass decreases in the first 6–7 years due to massive harvesting in Ω. Afterwards, in year 7–10, the MPA regulation leads to a recovery of the fish stock, see plot (C). Surprisingly, higher penalization leads to lower values of prey. To understand this, the dynamics inside the MPA is important. This is shown in the middle row plots (D–F) and the lower row plots (G–I) of Figure 5. In the initial phase, the higher penalties induce an expected higher level of u inside the MPA, see plot (D), which in turn drives the predator levels higher for all subsequent times, as in plots (G–I). However, in the intermediate phase, plot (E), there is a convergence in the prey biomass at the latter end of the interval as predation catches up. In the late phase, plot (F), the expectation that a fishing ban ($\sigma \geq 2^7$; $q_{\text{MPA}} = 0$) necessarily generates a higher population is reversed due to the dominant predation. Overall, a complete reserve closure initially provides only a short-term benefit for the domain wide amount of biomass, which then becomes detrimental in the medium to long-term. The reserve sees a longer lasting benefit from no harvesting but only up to an intermediate time, where the later phase then sees an equal benefit when transitioning to a medium ($\sigma = 2^4$) penalty. There are then three scenarios:

1. The maximum penalty ($\sigma = 2^8$) provides a whole-domain benefit for the prey only in the initial period and favors the predators for all times.
2. A high penalty ($\sigma = 2^7$) benefits both species in the short and intermediate term and is only detrimental to the prey in the long term.
3. The lower penalties only start to benefit the prey in and out the reserve at the end of the 10 year period but disadvantage the predators for all times.

Figure 6A,B show the prey levels in the reserve and everywhere outside it ($\Omega \setminus \Omega_{\text{MPA}}$) as a function of the penalty at time points corresponding to the interval endpoints in Figure 5. It can be seen in Figure 6A that for the shorter time spans (2 and 3.5 years), an increasing penalty leads to a monotonic increase, then plateau in the biomass. For the 7 year time span, there is at first a growth in the biomass level until a maximum at $\sigma = 2^6$, after which there is a decay. In the long term (10 years), the levels remain constant for almost all σ values and then also starts falling off after $\sigma \geq 2^6$. In Figure 6B (outside the reserve), the only time span to show the slightest growth is the also shortest. Note that the 10 year level in Figure 6A,B is always higher than that for 7 years. This is due to the convex shape of the growth in u as the juveniles u_i approach their carrying capacity K, and the 7 year sample point being in the region of the minimum in Figure 5B,E. Graphs 6C,D are the corresponding predator population, showing their population levels driven higher within Ω_{MPA} as a direct dependency on σ, whereas their levels remain constant on the outside of the MPA.

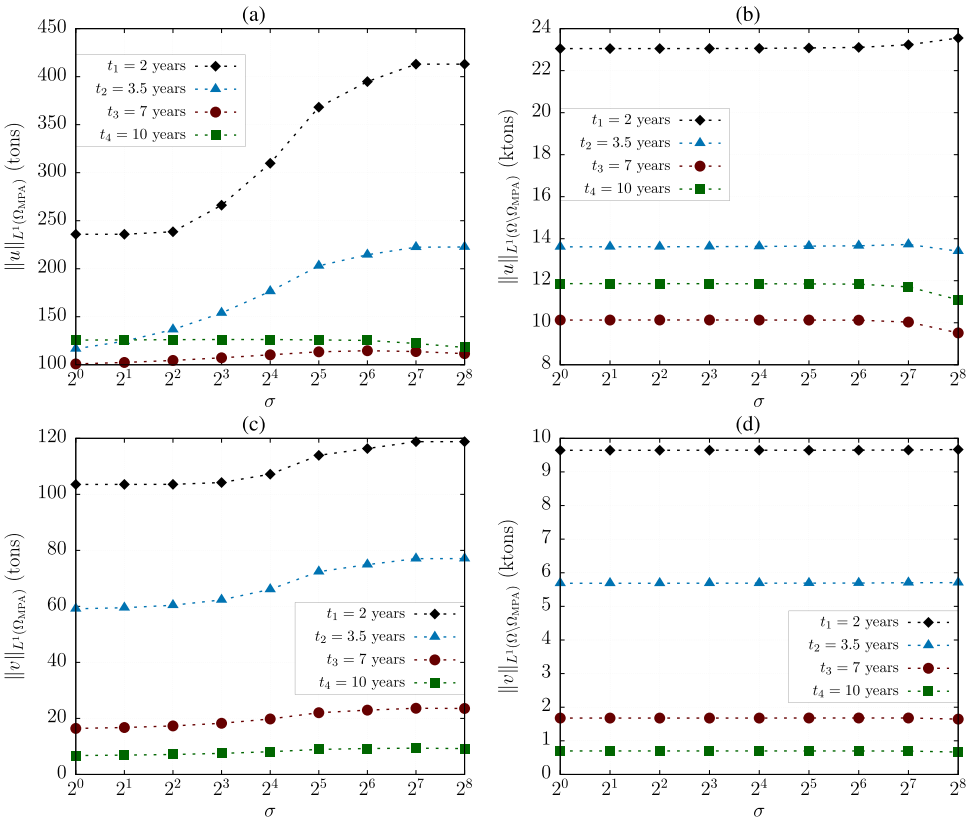


FIGURE 6 | (a–d) Snapshots of the mature prey u (top row) and predator v (bottom) populations at 2, 3.5, 7, and 10 year time points for all levels of penalty σ within the MPA (left column) and its effect everywhere outside (right column).

Hence, we observe that a high penalty effectively discourages human fishing, offering short-term protection for the prey. However, this also results in an increase in the predator population, which ultimately harms the prey in the long term. Moreover, a low penalty allows both humans and predators to exploit the prey, creating competition between the two. This consistently puts the predator at a disadvantage.

As a further enhancement, one could consider dynamic MPAs with a time-dependent penalty, $\sigma = \sigma(t)$. This approach would enable the use of a high penalty to initially help the prey population recover, followed by a lower penalty later to allow for human fishing. Such a strategy could potentially prevent the predator population from becoming excessively large, which could harm the prey in the later stages.

5 | Conclusion

We have presented the results of a fishing optimization problem that considers how the prey species in an age-structured predator-prey system is affected by predation and harvesting in the region of an MPA in the Baltic Sea. Somewhat surprisingly, the penalization of vessels entering the reserve did not incur a simple response of increased prey biomass, but rather displayed a more nuanced relationship between the species in the system. We identified three scenarios: (a) For maximal penalty in the MPA, the prey initially benefits across the entire domain but favors predators throughout the entire time horizon. (b) A high penalty provides short- and intermediate-term benefits to both prey and predators, but negatively impacts prey in the long term. (c) At lower penalties, the prey benefits both inside and outside the reserve only toward the end of the 10-year period, while consistently disadvantaging predators.

It is reasonably well established that MPAs exhibit a biological response (Lester et al. 2009; Lester and Halpern 2008) and how this is manifested using known biological parameters is seen in our results. Central to our problem design is the compliance of fishing vessels to not enter a fixed protected area, or when they do, the imposition of a penalty for doing so. The implementation of such a compliance regime is not unfeasible due to the political makeup of the Baltic region. First, since all the neighboring exclusive economic zones are directly adjacent to each other there can be no unregulated, unreported, and unauthorized (IUU) *high seas* fishing within the basin. Second, the geofencing of ocean patches and tracking vessels traversing them has become feasible with the adoption of the automatic identification system (AIS), a GPS position broadcast system. The designation of the MPAs is already communally managed by the surrounding member states, all that remains seems to be the willingness to enforce compliance.

We are aware of a modest bias in our parameter sources due to the dominance of aquaculture studies in the available literature, which is driven by a commercial interest in growth efficiency in farmed fish. The publication of more field study data would therefore be helpful for the further refinement of our analysis. Another direction of future research would be the addition of a temperature dimension to the model. Temperature variations are known to directly affect the feeding rate, handling time (Knutsen and Salvanes 1999), swim speed, and hence diffusion (Claireaux et al. 1995), and attack rate (Lang et al. 2012) of *Gadus morhua*. Given the difficulty in obtaining such data in field trials, it would nevertheless make an interesting comparison to our aquaculture influenced results.

Author Contributions

Simon Taylor: investigation, conceptualization, visualization, writing – review and editing, software, formal analysis, writing – original draft, methodology, validation, data curation. **Malte Braack:** conceptualization, writing – original draft; supervision, methodology.

Acknowledgments

Open Access funding enabled and organized by Projekt DEAL.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Ali, N., and M. Jazar. 2013. "Global Dynamics of a Modified Leslie-Gower Predator-Prey Model With Crowley-Martin Functional Responses." *Journal of Applied Mathematics and Computing* 43: 271–293.

Aljetlawi, A. A., E. Sparrevik, and K. Leonardsson. 2004. "Prey-Predator Size-Dependent Functional Response: Derivation and Rescaling to the Real World." *Journal of Animal Ecology* 73, no. 2: 239–252. https://doi.org/10.1111/jae.2004.73.issue-2.

Andersen, N. G., and J. Riis-Vestergaard. 2003. "The Effects of Food Consumption Rate, Body Size and Temperature on Net Food Conversion Efficiency in Saithe and Whiting." *Journal of Fish Biology* 62, no. 2: 395–412. https://doi.org/10.1046/j.1095-8649.2003.00030.x.

Arditi, R., and H. R. Akçakaya. 1990. "Underestimation of Mutual Interference of Predators." *Oecologia* 83, no. 3: 358–361. https://doi.org/10.1007/BF00317560.

Arditi, R., and L. R. Ginzburg. 1989. "Coupling in Predator-Prey Dynamics: Ratio-Dependence." *Journal of Theoretical Biology* 139, no. 3: 311–326. https://doi.org/10.1016/S0022-5193(89)80211-5.

Becker, R., D. Meidner, and B. Vexler. 2007. "Efficient Numerical Solution of Parabolic Optimization Problems by Finite Element Methods." *Optimization Methods and Software* 22, no. 5: 813–833. https://doi.org/10.1080/10556780701228532.

Beddington, J. R. 1975. "Mutual Interference Between Parasites or Predators and its Effect on Searching Efficiency." *The Journal of Animal Ecology* 44, no. 1: 331–340. https://doi.org/10.2307/3866.

Bethke, E., M. Bernreuther, and R. Tallman. 2013. "Feed Efficiency Versus Feed Conversion Ratio–Demonstrated on Feeding Experiments With Juvenile Cod (*Gadus morhua*)." *SSRN* 55, no. 1: 1–200. https://doi.org/10.2139/ssrn.2313137.

Bjornsson, B., and A. Steinarsson. 2002. "The Food-Unlimited Growth Rate of Atlantic Cod (*Gadus morhua*)." *Canadian Journal of Fisheries and Aquatic Sciences* 59, no. 3: 494–502. https://doi.org/10.1139/f02-028.

Braack, M., and R. Becker. 2001. "A Finite Element Pressure Gradient Stabilization for the Stokes Equations Based on Local Projections." *CALCOLO* 38: 173–199.

Braack, M., M. F. Quaas, B. Tews, and B. Vexler. 2018. "Optimization of Fishing Strategies in Space and Time as a Non-Convex Optimal Control

Problem." *Journal of Optimization Theory and Applications* 178, no. 3: 950–972. https://doi.org/10.1007/s10957-018-1304-7.

Brey, T., C. Müller-Wiegmann, Z. M. Zittier, and W. Hagen. 2010. "Body Composition in Aquatic Organisms – A Global Data Bank of Relationships Between Mass, Elemental Composition and Energy Content." *Journal of Sea Research* 64, no. 3: 334–340. https://doi.org/10.1016/j.seares.2010.05.002. http://www.thomas-brey.de/science/virtualhandbook/.

BSH. 2023. Das bundesamt für seeschifffahrt und hydrographie (the german federal maritime and hydrographic agency). https://www.bsh.de/EN/DATA/Predictions/Currents/currents_node.html.

Chen, S., J. Wei, and J. Yu. 2018. "Stationary Patterns of a Diffusive Predator–Prey Model With Crowley–Martin Functional Response." *Nonlinear Analysis: Real World Applications* 39: 33–57.

Claireaux, G., D. M. Webber, S. R. Kerr, and R. G. Boutilier. 1995. "Physiology and Behaviour of Free-Swimming Atlantic Cod (*Gadus morhua*) Facing Fluctuating Temperature Conditions." *Journal of Experimental Biology* 198, no. 1: 49–60. https://doi.org/10.1242/jeb.198.1.49.

Cohen, D., T. Inada, T. Iwamoto, and N. Scialabba. 1990. "Fao Species Catalogue. Vol. 10. Gadiform Fishes of the World (Order Gadiformes). An Annotated and Illustrated Catalogue of Cods, Hakes, Grenadiers and Other Gadiform Fishes Known to Date." *Food and Agriculture Organization (FAO) of the United Nations Catalogue* 10: 44–47.

Crowley, P. H., and E. K. Martin. 1989. "Functional Responses and Interference Within and Between Year Classes of a Dragonfly Population." *Journal of the North American Benthological Society* 8, no. 3: 211–221. https://doi.org/10.2307/1467324.

DeAngelis, D. L., R. A. Goldstein, and R. V. O'Neill. 1975. "A Model for Tropic Interaction." *Ecology* 56, no. 4: 881–892. https://doi.org/10.2307/1936298.

der Justiz, B. 2022. Bekanntmachung bundesanzeiger banz at 08.02.2022 b7. https://www.bundesanzeiger.de/pub/publication/G5qzLAhuWIG48leQ6n4/content/211211000705M001/BAnzAT08022022B700.pdf.

Ding, W., and S. Lenhart. 2009. "Optimal Harvesting of a Spatially Explicit Fishery Model." *Natural Resource Modeling* 22, no. 2: 173–211. https://doi.org/10.1111/nrm.2009.22.issue-2.

dosSantos, J., and M. Jobling. 1991. "Factors Affecting Gastric Evacuation in Cod, *Gadus morhua* L., Fed Single-Meals of Natural Prey." *Journal of Fish Biology* 38, no. 5: 697–713. https://doi.org/10.1111/jfb.1991.38.issue-5.

Dureuil, M., K. Boerder, K. A. Burnett, R. Froese, and B. Worm. 2018. "Elevated Trawling Inside Protected Areas Undermines Conservation Outcomes in a Global Fishing Hot Spot." *Science* 362, no. 6421: 1403–1407. https://doi.org/10.1126/science.aau0561.

Edebohls, I., M. Niemann, J. Berkenhagen, R. Döring, and A. Schröder. 2022. "Steckbrief Zur Meeresfischerei in Deutschland." *Johann Heinrich von Thünen-Institut*: 1–25. https://literatur.thuenen.de/digbib_extern/dn065720.pdf.

Essington, T. E., and S. Hansson. 2004. "Predator-Dependent Functional Responses and Interaction Strengths in a Natural Food Web." *Canadian Journal of Fisheries and Aquatic Sciences* 61, no. 11: 2215–2226. https://doi.org/10.1139/f04-146.

Froese, R., and D. Pauly. 2021. Fishbase. https://www.fishbase.de.

Gerth, D. 2021. "A New Interpretation of (Tikhonov) Regularization." *Inverse Problems* 37, no. 6: 064002. https://doi.org/10.1088/1361-6420/abfb4d.

Grygiel, W., and M. Wyszyński. 2002. "Sexual Maturation of the Southern Baltic Herring and Sprat (1980–2001)." *ICES Scientific Reports* 21. https://www.ices.dk/sites/pub/CM%20Doccuments/2002/O/O2102.PDF.

Hassell, M., and G. Varley. 1969. "New Inductive Population Model for Insect Parasites and Its Bearing on Biological Control." *Nature* 223, no. 5211: 1133–1137. https://doi.org/10.1038/2231133a0.

Heinkenschloss, M., and D. Leykekhman. 2010. "Local Error Estimates for Supg Solutions of Advection-Dominated Elliptic Linear-Quadratic Optimal Control Problems." *SIAM Journal on Numerical Analysis* 47, no. 6: 4607–4638. https://doi.org/10.1137/090759902.

Janica, B, E Jan, and B Penina. 2013. "Helcom Protect–Overview of the Status of the Network of Baltic Sea Marine Protected Areas." *HELCOM* 16: 19.

Herdson, D., and I. Priede. 2010. Clupea Harengus. The IUCN Red List of Threatened Species 2010.

Hoang, M. T. 2020. "On the Global Asymptotic Stability of a Predator–Prey Model With Crowley–Martin Function and Stage Structure for Prey." *Journal of Applied Mathematics and Computing* 64: 765.

Holling, C. S. 1959. "The Components of Predation as Revealed by a Study of Small-Mammal Predation of the European Pine Sawfly." *Canadian Entomologist* 91, no. 5: 293–320. https://doi.org/10.4039/Ent91293-5.

Hutchings, J. A. 2000. "Collapse and Recovery of Marine Fishes." *Nature* 406, no. 6798: 882–885. https://doi.org/10.1038/35022565.

ICES. 2019a. Cod (Gadus morhua) in Subdivisions 22-24, Western Baltic Stock (Western Baltic Sea).

ICES. 2019b. Herring (clupea harengus) in Subdivisions 20-24, Spring Spawners (Skagerrak, Kattegat, and Western Baltic).

IOW. 2023. Leibniz Institute For Baltic Sea Research Warnemünde. https://www.io-warnemuende.de.

Johnson, C. 1985. "Streamline Diffusion Methods for Problems in Fluid Mechanics." *Finite Elements in Fluids* 6: 251–261.

Joshi, H. R., G. E. Herrera, S. Lenhart, and M. G. Neubert. 2009. "Optimal Dynamic Harvest of a Mobile Renewable Resource." *Natural Resource Modeling* 22, no. 2: 322–343. https://doi.org/10.1111/nrm.2009.22.issue-2.

Kelly, M. R. J., Y. Xing, and S. Lenhart. 2016. "Optimal Fish Harvesting for a Population Modeled by a Nonlinear Parabolic Partial Differential Equation." *Natural Resource Modeling* 29, no. 1: 36–70. https://doi.org/10.1111/nrm.2016.29.issue-1.

Knutsen, I., and A. G. Salvanes. 1999. "Temperature-Dependent Digestion Handling Time in Juvenile Cod and Possible Consequences of Prey Choice." *Marine Ecology-Progress Series* 181: 61–79. https://doi.org/10.3354/meps181061.

Kurata, K., and J. Shi. 2008. "Optimal Spatial Harvesting Strategy and Symmetry-Breaking." *Applied Mathematics and Optimization* 58, no. 1: 89–110. https://doi.org/10.1007/s00245-007-9032-7.

Lang, B., B. C. Rall, and U. Brose. 2012. "Warming Effects on Consumption and Intraspecific Interference Competition Depend on Predator Metabolism." *Journal of Animal Ecology* 81, no. 3: 516–523. https://doi.org/10.1111/jane.2012.81.issue-3.

Lester, S. E., and B. S. Halpern. 2008. "Biological Responses in Marine No-Take Reserves Versus Partially Protected Areas." *Marine Ecology Progress Series* 367: 49–56. https://doi.org/10.3354/meps07599.

Lester, S. E., B. S. Halpern, K. Grorud-Colvert, et al. 2009. "Biological Effects Within No-Take Marine Reserves: A Global Synthesis." *Marine Ecology Progress Series* 384: 33–46. https://doi.org/10.3354/meps08029.

Løkkeborg, S. 1998. "Feeding Behaviour of Cod, *Gadus morhua*: Activity Rhythm and Chemically Mediated Food Search." *Association for the Study of Animal Behaviour* 56: 371–378.

Meng, X.-Y., H.-F. Huo, H. Xiang, and Q. yu Yin. 2014. "Stability in a Predator–Prey Model With Crowley–Martin Function and Stage Structure for Prey." *Applied Mathematics and Computation* 232, no. 3: 810–819. https://doi.org/10.1016/j.amc.2014.01.139.

Neubert, M. G. 2003. "Marine Reserves and Optimal Harvesting." *Ecology Letters* 6, no. 9: 843–849. https://doi.org/10.1046/j.1461-0248.2003.00493.x.

Neubert, M. G., and G. E. Herrera. 2008. "Triple Benefits From Spatial Resource Management." *Theoretical Ecology* 1, no. 1: 5–12. https://doi.org/10.1007/s12080-007-0009-6.

Neuenfeldt, S., V. Bartolino, A. Orio, et al. 2019. "Feeding and Growth of Atlantic Cod (*Gadus morhua* L.) in the Eastern Baltic Sea Under Environmental Change." *ICES Journal of Marine Science* 77, no. 2: 624–632.

Pauly, D. 1978. "A Preliminary Compilation of Fish Length Growth Parameters." *Ber. Inst. Meereskd. Christian-Albrechts-Univ. Kiel* 55, no. 1: 1–200.

Quinn, T. J. 2003. "Ruminations on the Development and Future of Population Dynamics Models in Fisheries." *Natural Resource Modeling* 16, no. 4: 341–392. https://doi.org/10.1111/nrm.2003.16.issue-4.

Sambath, M., S. Gnanavel, and K. Balachandran. 2013. "Stability and Hopf Bifurcation of a Diffusive Predator–Prey Model With Predator Saturation and Competition." *Applicable Analysis* 92, no. 12: 2439–2456. https://doi.org/10.1080/00036811.2012.742185.

Sinclair, A. F. 2001. "Natural Mortality of Cod (*Gadus morhua*) in the Southern Gulf of St Lawrence." *ICES Journal of Marine Science* 58, no. 1: 1–10. https://doi.org/10.1006/jmsc.1999.0490.

Sirén, A., and K. Parvinen. 2015. "A Spatial Bioeconomic Model of the Harvest of Wild Plants and Animals." *Ecological Economics* 116, no. 1: 201–210. https://doi.org/10.1016/j.ecolecon.2015.04.015.

Skalski, G. T., and J. F. Gilliam. 2001. "Functional Responses With Predator Interference: Viable Alternatives to the Holling Type II Model." *Ecology* 82, no. 11: 3083–3092. https://doi.org/10.1890/0012-9658(2001)082[3083:FRWPIV]2.0.CO;2.

UN Convention on Biological Diversity. 2012. Target 11 Strategic Plan 2011–2020–Technical Rationale. https://www.cbd.int/sp/targets/rationale/target-11.

Upmann, T., and S. Behringer. 2020. "Harvesting a Remote Renewable Resource." *Theoretical Ecology* 13, no. 4: 459–480. https://doi.org/10.1007/s12080-020-00453-y.

Upmann, T., and D. Gromov. 2023. "The Structure of Optimal Solutions for Harvesting a Renewable Resource." *Natural Resource Modeling* 36, no. 1: e12355. https://doi.org/10.1111/nrm.v36.1.

Zeller, D., P. Rossing, S. Harper, L. Persson, S. Booth, and D. Pauly. 2011. "The Baltic Sea: Estimates of Total Fisheries Removals 1950–2007." *Fisheries Research* 108, no. 2: 356–363. https://doi.org/10.1016/j.fishres.2010.10.024.

Paper 3: A local-projection stabilized sedimentation model with convection-dominant flow

5

Submitted to *Computational and Applied Mathematics*
26.2.2025

A local-projection stabilized sedimentation model with convection-dominant flow

Simon Taylor[1*] and Malte Braack[1]

[1*]Department of Applied Mathematics, Christian-Albrechts-Universität zu Kiel, 24118, Kiel, Germany.

*Corresponding author(s). E-mail(s): taylor@math.uni-kiel.de;
Contributing authors: braack@math.uni-kiel.de;

Abstract

We analyze the existence of solutions for the finite-element discretized stationary and time-dependent sedimentation-consolidation partial differential equations (PDEs) using recent advances in the local-projection stabilization (LPS) method. The sedimentation model studied here couples gravitationally forced Stokes flow and a convection-diffusion equation for the solids concentration, with boundary conditions selected to mimic physical applications. This system is highly nonlinear and sensitive due to the non-constant velocity and concentration coupling terms, as well as the nonlinear flux term in the convection equation. We provide an overview of pre-existing methods already developed for solving this class of problem, establish the coercivity of the underlying operator, as well as conditions for the existence of discrete solutions. For convection-dominated regimes using non-*inf-sup* stable finite elements with LPS, we demonstrate that stabilization effectively eliminates interior and boundary layers in the velocity and pressure solutions. Numerical examples in two and three dimensions for the non-stationary case illustrate the theoretical results.

Keywords: LPS Stabilization, Finite Elements, Sedimentation Models, Numerical Methods

MSC Classification: 35A01 , 35A25 , 35G61 , 35Q35 , 65N30 , 65N12

1 Introduction

The computational simulation of gravitational sedimentation and consolidation is crucial for various industrial and natural processes. Despite the highly nonlinear nature of these problems, most stabilization methods have focused on residual-based approaches, or their variations. This paper explores the application of recent advances in LPS for flow and convection problems applied to sedimentation-consolidation under gravity.

Such models have been used in optimizing gravity thickeners for solid-liquid separation in mineral processing (Benn et al., 2018) and wastewater treatment (Ekama et al., 1997). A gravitational thickener is a large container where an untreated suspension enters, sediment concentrates at the bottom, and clear liquid overflows. Flocculants are often added to accelerate separation. For foundational studies on gravity thickeners and flocculated settling, see (Fitch, 1966), and for the theory of settling convection and boundaries, refer to (Hill et al., 1977; Acrivos and Herbolzheimer, 1979).

A general mathematical model for sedimentation-consolidation of flocculated suspensions was developed by (Bürger et al., 2000), where a Navier-Stokes-type system was coupled with a convection-diffusion equation via concentration-dependent diffusion and viscosity functions. They analyzed solvability in two and three dimensions, depending on phase and mixture viscosities, and presented numerical results for a one-dimensional case using a finite difference operator splitting method. Subsequent numerical solutions in two and three dimensions were explored in (Bürger et al., 2015; Ruiz-Baier and Torres, 2015) using finite volume-element methods, while (Kleine and Reddy, 2005) applied stabilized finite elements. (Bürger et al., 2015) modeled a settling tank with inlet and outlet boundaries, whereas (Ruiz-Baier and Torres, 2015) focused on gravity settling in closed channels, introducing a variational multiscale stabilization for the Navier-Stokes component. Previously, (Bürger et al., 2012) proposed a stabilized finite volume element method (FVM) using transfer maps to align test and trial spaces, producing a scheme structurally similar to an edge-based stabilization method.

These models, including ours, are based on Kynch's hindered settling theory (Kynch, 1952), which describes how increasing local particle concentration slows sedimentation. The settling velocity $V(\phi)$ is a monotonically decreasing function of concentration, playing a crucial role in the sediment transport flux function. For these problems, Stokes viscosity is modeled as a nonlinear function, where an exponent controls the onset and slope. The diffusivity function is also nonlinear and material-dependent, often fitted to experimental data. For example, (Bürger et al., 2000) calibrated their model using flocculated kaolin deposition experiments. The combination of strong nonlinearities and coupling between the incompressibility constraint and field equations results in a highly sensitive system, making accurate numerical solutions challenging. The finite-volume method (FVM) is widely used due to its conservation properties and implementation simplicity, similar to finite differences. It achieves optimal error estimates in the H^1-norm, comparable to the linear finite-element method (FEM) (Ewing et al., 2002; Ronghua, 1987), and optimal order estimates in the L^2-norm are also attainable (Ewing et al., 2002). Studies on FVM for generalized Stokes problems (Chou and Kwak, 1998; Rui, 2005) leveraged its connection to FEM, obtaining error estimates through finite-element techniques satisfying the discrete inf-sup condition.

In this work, we apply the LPS method (Becker and Braack, 2001) to these models. LPS stabilizes gradient fluctuations by constructing a global projection from local ones, avoiding computationally expensive second-order derivatives. Recent extensions of LPS to non-Dirichlet boundary conditions, such as Neumann-type stabilization for inflow and outflow conditions (Braack, 2018), motivate our approach. Here, we apply LPS to stabilize both the flow and transport components simultaneously. The key advantage of LPS for coupled PDE systems is its avoidance of additional couplings between unknowns due to the absence of second-order derivatives. To our knowledge, the full system in Section 2 has not yet been stabilized using LPS.

These models belong to a broader class of variable-coefficient coupled flow-transport problems relevant to other fields, such as thermal convection flows with temperature-dependent coefficients (Tabata and Tagami, 2005; Lorca and Boldrini, 1999), natural convection (Jha et al., 2023), mantle-magma dynamics with Darcy flow (Dannberg and Heister, 2016), and soil science (Lee et al., 2000). Our results may thus be of interest beyond sedimentation modeling.

The paper is arranged as follows: Section 2 introduces the continuous equations and parameter definitions. Section 3 presents the discretization and stabilization terms. Finally, Section 4 demonstrates two numerical examples: a 2D in-outflow setup and a 3D case with flow obstacles.

2 Problem description

The consolidation of a sediment suspension in a fluid is governed by the coupling of Stokes flow with gravitational forcing, alongside a convection-diffusion-reaction equation. The choice of the Stokes model is well-suited for describing suspension motion, as sedimentation occurs slowly under a predominantly gravitational body force. This justifies focusing on the laminar macroscopic flow over large temporal scales, where turbulence would merely delay particle deposition. The suspended solid phase is thus treated as a continuous phase within a two-phase system.

We consider the incompressible flow of a fluid in a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$, where $d \in \{2, 3\}$, with boundary $\Gamma = \partial\Omega$. The time interval of interest is $(0, T)$, and we define the space-time domain as $\Omega_T := \Omega \times (0, T)$. The Stokes problem, governing the velocity field $\boldsymbol{u} : \Omega_T \to \mathbb{R}^d$ and pressure $p : \Omega_T \to \mathbb{R}$, is subject to a no-slip (homogeneous Dirichlet) condition on part of the boundary. A zero-stress (natural outflow) condition is imposed on Γ_{out}. For the transport equation, the solid concentration $\phi : \Omega_T \to \mathbb{R}$ satisfies a Neumann boundary condition on Γ, except at the inflow boundary Γ_{in}, where a smooth Dirichlet condition is prescribed. This configuration permits material entry into Ω via Stokes convection while ensuring containment due to the no-flux barrier.

The governing system of equations is given by:

$$\partial_t \phi - \operatorname{div}(\kappa(\phi)\nabla\phi) + \boldsymbol{u} \cdot \nabla\phi = \nabla \cdot \boldsymbol{f}(\phi) \qquad \text{in } \Omega_T, \tag{1}$$

$$\partial_t \boldsymbol{u} - \operatorname{div}(\mu(\phi)\boldsymbol{\varepsilon}(\boldsymbol{u}) - p\mathbf{I}) = \phi\, \boldsymbol{g} \qquad \text{in } \Omega_T, \tag{2}$$

$$\operatorname{div} \boldsymbol{u} = 0 \qquad \text{in } \Omega_T, \tag{3}$$

with boundary conditions

$$\begin{aligned} \boldsymbol{u} &= 0 && \text{on } \Gamma \times (0,T) \setminus (\Gamma_{\text{in}} \cup \Gamma_{\text{out}}) && (4)\\ \nabla\phi \cdot \boldsymbol{n} &= 0 && \text{on } \Gamma \times (0,T) \setminus \Gamma_{\text{in}} && (5)\\ \boldsymbol{u} &= \boldsymbol{u}_{\text{in}} && \text{on } \Gamma_{\text{in}} \times (0,T) && (6)\\ \phi &= \phi_{\text{in}} && \text{on } \Gamma_{\text{in}} \times (0,T) && (7)\\ \mu(\phi)\nabla\boldsymbol{u} \cdot \boldsymbol{n} - p\boldsymbol{n} &= -p_0\boldsymbol{n} && \text{on } \Gamma_{\text{out}} \times (0,T) && (8) \end{aligned}$$

and initial condition

$$\boldsymbol{u}(0) = \boldsymbol{u}_0 \qquad \text{in } \Omega. \tag{9}$$

As usual, $\boldsymbol{n}$ denotes the boundary normal vector, and t represents time. The gravitational force is given by $\boldsymbol{g} = \mathrm{g}\,\boldsymbol{k}$, where $\boldsymbol{k}$ is the downward-facing unit normal. The right-hand side of (1) represents the divergence of the Kynch flux function, a vector oriented in the downward direction. This flux function was first introduced in (Kynch, 1952):

$$\boldsymbol{f}(\phi) := v_\infty \phi\, V(\phi)\boldsymbol{k}, \quad V(\phi) := \begin{cases} (1-\phi/\phi_{max})^{n_{\text{RZ}}}, & 0 \le \phi \le \phi_{\max},\\ 0, & \phi < 0 \text{ or } \phi > \phi_{max}, \end{cases} \tag{10}$$

with Stokes settling velocity $v_\infty \ge 0$, and the so called hindered settling factor $V(\phi)$, which represents the slip velocity of particles in a suspension, and $n_{\text{RZ}} \in \mathbb{N}$. A nominal maximal concentration ϕ_{max} is used to prevent singularities.

The body force term $\phi\,\boldsymbol{g} : \Omega \to \mathbb{R}^d$ on the right-hand side of (2) accounts for local fluctuations in ϕ that influence the mixture flow. The strain tensor, defined as $\boldsymbol{\varepsilon}(\boldsymbol{u}) = \frac{1}{2}(\nabla\boldsymbol{u} + \nabla\boldsymbol{u}^{\mathrm{T}})$, appears inside the divergence term in (2) due to the presence of a variable viscosity $\mu(\phi)$. The viscosity $\mu(\phi)$ is a nonlinear, concentration-dependent Newtonian viscosity, as described by (McLachlan, 1986; Ruiz-Baier and Torres, 2015):

$$\mu(\phi) := \mu_0(1-\phi/\tilde{\phi}_{max})^{-\beta}, \quad \beta \in \mathbb{N},$$

with minimal viscosity $\mu_0 > 0$, and nominal maximum concentration $\tilde{\phi}_{max} \ge \phi_{max}$.

If flocculation is neglected, self-diffusion can be modeled solely by the constant $\kappa(\phi) \equiv D_0$, which remains a viable approach. However, to include flocculation effects, the function $\kappa(\phi)$ represents a local, nonlinear, concentration-dependent diffusion term, as documented in (Guazzelli and Hinch, 2011; Guazzelli et al., 2011):

$$\kappa(\phi) := D_0 - \frac{v_\infty \phi\, V(\phi)\sigma_{\text{e}}'(\phi)}{(\rho_s - \rho_f)\mathrm{g}\phi}, \quad \text{where } \sigma_{\text{e}}(\phi) := \frac{\sigma_0\alpha}{\phi_c^\alpha}\phi^{\alpha-1}, \quad \alpha \in \mathbb{N}. \tag{11}$$

The *effective solid stress function* $\sigma_{\text{e}}(\phi)$ models sediment compressibility, particularly when particle flocculation is considered. It satisfies $\sigma_e \in C^2(\mathbb{R})$ with $\sigma_e' \ge 0$, ensuring

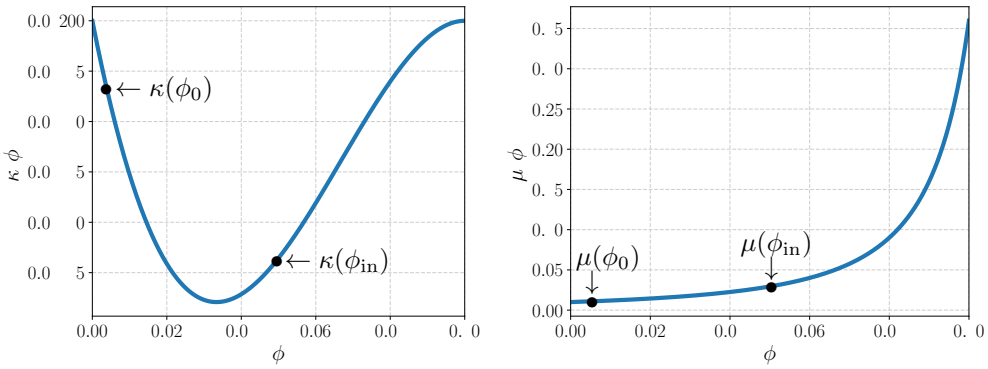


Fig. 1: The nonlinear characteristics of the continuous concentration-dependent local diffusion (left) and viscosity functions (right) for given initial and inflow conditions.

non-negative compressive stress. Figure 1 illustrates example profiles of the diffusion and viscosity functions, highlighting potential variations in initial and boundary conditions. By κ_0 we denote the lower bound

$$\kappa_0 := \inf \{\kappa(\phi) \ : \ 0 \leq \phi \leq \phi_{max}\} .$$

We assume that this lower bound is positive, $\kappa_0 > 0$.

The coupling of the forced Stokes system with the transport equation in (1) introduces several issues that affect the numerical stability of a solution:

- For the convection-dominated makeup, i.e. $\kappa(\phi)$ small, the problem leads to a failure in the pure Galerkin approximation to find a solution at the discrete level without spurious over- and undershoots.
- Additionally, the method for solving the system demands the use of *inf-sup* stable finite element pairs for pressure and velocity, unless some form of stabilization is utilized.
- Further still, the solution method also has to ensure the local conservation of mass.

3 Variational formulation and discretization

In the case of a stationary problem, the equation system (1)-(3) can be reformulated into

$$\begin{aligned} -\text{div}(\kappa(\phi)\nabla\phi) + \boldsymbol{u} \cdot \nabla\phi - \nabla \cdot \boldsymbol{f}(\phi) &= d_\phi \qquad \text{in } \Omega, &(12)\\ -\text{div}(\mu(\phi)\boldsymbol{\varepsilon}(\boldsymbol{u}) - p\mathbf{I}) - \phi\,\boldsymbol{g} &= \boldsymbol{d}_u \qquad \text{in } \Omega, &(13)\\ \text{div}\,\boldsymbol{u} &= 0 \qquad \text{in } \Omega, &(14) \end{aligned}$$

with right-hand sides d_ϕ and $\boldsymbol{d}_u$, accounting for the inhomogeneous Dirichlet data (6)-(8). The sought solution is then given by $\phi + \phi_{\text{in}}$, $\boldsymbol{u} + \boldsymbol{u}_{\text{in}}$, and p, where ϕ_{in} and $\boldsymbol{u}_{\text{in}}$

are extensions of the boundary data to Ω. We now present the variational formulation and the finite-element discretization for the stationary problem.

3.1 Variational formulation for the stationary equations

The corresponding Hilbert spaces are defined as follows:

$$\begin{aligned} V_\phi &:= \{\phi \in H^1(\Omega), : \phi = 0 \text{ on } \Gamma_{\text{in}}\}, \\ \boldsymbol{V}_u &:= \{\boldsymbol{u} \in H^1(\Omega)^3 : \boldsymbol{u} = 0 \text{ on } \partial\Omega \setminus \Gamma_{\text{out}}\}, \\ V_p &:= L^2_0(\Omega). \end{aligned}$$

Thus, we have $\phi \in V_\phi$, $\boldsymbol{u} \in \boldsymbol{V}_u$, and $p \in V_p$. The inner products in $L^2(\Omega)$ and $L^2(\Gamma)$ are denoted by $(\cdot,\cdot)$ and $(\cdot,\cdot)_\Gamma$, respectively. The $L^2(\Omega)$-norm will be denoted by $\|\cdot\|$. We now introduce the following three forms:

$$\begin{aligned} a_1(\phi, \boldsymbol{u}; \varphi) &:= (\kappa(\phi)\nabla\phi, \nabla\varphi) + (\boldsymbol{u}\cdot\nabla\phi, \varphi) + (\boldsymbol{f}(\phi), \nabla\varphi) - (\boldsymbol{f}(\phi)\cdot\boldsymbol{n}, \varphi)_\Gamma \\ &\quad + \frac{1}{2}((\operatorname{div}\boldsymbol{u})\phi, \varphi), \qquad (15) \\ a_2(\phi, \boldsymbol{u}; \varphi) &:= (\mu(\phi)\nabla\boldsymbol{u}, \nabla\varphi) - (\phi\boldsymbol{g}, \varphi), \qquad (16) \\ b(\boldsymbol{u}; \varphi) &:= (\operatorname{div}\boldsymbol{u}, \varphi). \qquad (17) \end{aligned}$$

We assemble the three variables $\phi, \boldsymbol{u}$ and p in the tuple $\boldsymbol{x} := [\phi, \boldsymbol{u}, p] \in \boldsymbol{X} := V_\phi \times \boldsymbol{V}_u \times V_p$. Denoting the test functions by $\boldsymbol{y} := [y_1, \boldsymbol{y}_2, y_3]$, we define the form $A : \boldsymbol{X} \times \boldsymbol{X} \to \mathbb{R}$ by

$$A(\boldsymbol{x}, \boldsymbol{y}) := a_1(\phi, \boldsymbol{u}; y_1) + a_2(\phi, \boldsymbol{u}; \boldsymbol{y}_2) - b(\boldsymbol{y}_2, p) + b(\boldsymbol{u}, y_3). \qquad (18)$$

Furthermore, we define the functional $\boldsymbol{d} \in \boldsymbol{X}'$ defined by $\langle \boldsymbol{d}, \boldsymbol{y}\rangle := (d_\phi, y_1) + (\boldsymbol{d}_u, \boldsymbol{y}_2)$. The equation can now be written in the compact form

$$\boldsymbol{x} \in \boldsymbol{X} : \quad A(\boldsymbol{x}, \boldsymbol{y}) = \langle \boldsymbol{d}, \boldsymbol{y}\rangle \quad \forall \boldsymbol{y} \in \boldsymbol{X}. \qquad (19)$$

3.2 Weak continuity of the underlying nonlinear operator

In this section, we establish the weak continuity of the operator defined above. This property is crucial for proving the existence of weak solutions.

Lemma 3.1 *The operator introduced in (18) is sequentially weakly continuous. Specifically, if $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ is a weakly convergent sequence in $\boldsymbol{X}$ such that $\boldsymbol{x}_n \rightharpoonup \boldsymbol{x}$, then for every $\boldsymbol{y} \in \boldsymbol{X}$, it holds that:*

$$\lim_{n\to\infty} A(\boldsymbol{x}_n, \boldsymbol{y}) = A(\boldsymbol{x}, \boldsymbol{y}).$$

Proof All linear components are obviously continuous, so we focus on the nonlinear terms of the operator.

Assume $\phi_n \rightharpoonup \phi$ in V_ϕ and $\boldsymbol{u}_n \rightharpoonup \boldsymbol{u}$ in $\boldsymbol{V}_u$. By a compactness argument, the weak convergence of $\boldsymbol{x}_n = [\phi_n, \boldsymbol{u}_n, p_n]$ implies convergence in L^2-norm for the H^1-components:

$$\lim_{n\to\infty} (\|\phi_n - \phi\| + \|\boldsymbol{u}_n - \boldsymbol{u}\|) = 0.$$

Using the trace theorem (Adams, 1975), which states that the trace operator $\gamma : H^1(\Omega) \to L^q(\partial\Omega)$ is continuous for $2 \le q \le 4$ and compact for $2 \le q < 4$, we obtain:

$$\lim_{n\to\infty} \|\phi_n - \phi\|_{L^2(\partial\Omega)} = 0.$$

Since $\|V'\|_{L^\infty} \le n_{RZ}$, we get:

$$\lim_{n\to\infty} \left(\|\phi_n - \phi\|_{L^2(\partial\Omega)} + \|V(\phi_n) - V(\phi)\|_{L^2(\partial\Omega)} \right) \to 0.$$

For a test function $\varphi \in C^\infty(\Omega)$ with $\varphi|_{\Gamma_{\text{in}}} = 0$, we analyze the convective term:

$$T_{1,n} := (\operatorname{div}(\boldsymbol{u}_n\phi_n - \boldsymbol{u}\phi), \varphi).$$

Splitting and estimating yields:

$$T_{1,n} = (\operatorname{div}((\boldsymbol{u}_n - \boldsymbol{u})\phi), \varphi) + (\operatorname{div}(\boldsymbol{u}(\phi_n - \phi)), \varphi) + (\operatorname{div}((\boldsymbol{u}_n - \boldsymbol{u})(\phi_n - \phi)), \varphi).$$

The first term vanishes for $n \to \infty$ by weak convergence $\boldsymbol{u}_n \rightharpoonup \boldsymbol{u}$, and the second term vanishes by weak convergence $\phi_n \rightharpoonup \phi$. The third is controlled using boundedness and convergence in the L^2-norm:

$$\begin{aligned}
&(\operatorname{div}((\boldsymbol{u}_n - \boldsymbol{u})(\phi_n - \phi)), \varphi) \\
&= ((\phi_n - \phi)\operatorname{div}(\boldsymbol{u}_n - \boldsymbol{u}), \varphi) + ((\boldsymbol{u}_n - \boldsymbol{u})\nabla(\phi_n - \phi), \varphi) \\
&\le \left(\underbrace{\|\phi_n - \phi\|}_{\to 0} \underbrace{\|\operatorname{div}(\boldsymbol{u}_n - \boldsymbol{u})\|}_{<\infty} + \underbrace{\|\boldsymbol{u}_n - \boldsymbol{u}\|}_{\to 0} \underbrace{\|\nabla(\phi_n - \phi)\|}_{<\infty} \right) \|\varphi\|_{L^\infty}.
\end{aligned}$$

This results in $\lim_{n\to\infty} T_{1,n} = 0$. For the bulk term with nonlinear flux $\boldsymbol{f}$:

$$T_{2,n} := |(\boldsymbol{f}(\phi_n) - \boldsymbol{f}(\phi), \nabla\varphi)| + \left| \int_{\partial\Omega} (\boldsymbol{f}(\phi_n) - \boldsymbol{f}(\phi)) \cdot \boldsymbol{n}\varphi \, ds \right|,$$

we estimate the differences in the L^1-norm:

$$\begin{aligned}
T_{2,n} &\le \|\boldsymbol{f}(\phi_n) - \boldsymbol{f}(\phi)\|_{L^1} \|\nabla\varphi\|_{L^\infty} + \|\boldsymbol{f}(\phi_n) - \boldsymbol{f}(\phi)\|_{L^1(\partial\Omega)} \|\nabla\varphi\|_{L^\infty} \\
&\le |v_\infty||\boldsymbol{k}| \left(\|\phi V(\phi) - \phi_n V(\phi_n)\|_{L^1} + \|\phi V(\phi) - \phi_n V(\phi_n)\|_{L^1(\partial\Omega)} \right) \|\nabla\varphi\|_\infty.
\end{aligned}$$

The two L^1-integrals converge, because:

$$\begin{aligned}
\|\phi V(\phi) - \phi_n V(\phi_n)\|_{L^1} &\le \|(\phi - \phi_n)V(\phi_n)\|_{L^1} + \|\phi(V(\phi) - V(\phi_n))\|_{L^1} \\
&\le (|\Omega|^{1/2} + \|\phi\|\|V'\|_{L^\infty}) \underbrace{\|\phi - \phi_n\|}_{\to 0}, \\
\|\phi V(\phi) - \phi_n V(\phi_n)\|_{L^1(\partial\Omega)} &\le (|\partial\Omega|^{1/2} + \|\phi\|_{L^2(\partial\Omega)}\|V'\|_{L^\infty(\partial\Omega)}) \underbrace{\|\phi - \phi_n\|_{L^2(\partial\Omega)}}_{\to 0}.
\end{aligned}$$

This implies $\lim_{n\to\infty} T_{2,n} = 0$. For the diffusive term with nonlinear coefficient

$$T_{3,n} := |(\kappa(\phi_n)\nabla\phi_n - \kappa(\phi)\nabla\phi, \nabla\varphi)|,$$

we have $\nabla\kappa(\phi) = \kappa'(\phi)\nabla\phi \in L^2(\Omega)$, hence, $\kappa(\phi) \in H^1(\Omega)$. This is used to derive the following upper bound:

$$T_{3,n} \le |((\kappa(\phi_n) - \kappa(\phi))\nabla\phi_n, \nabla\varphi)| + |(\phi_n - \phi, \operatorname{div}(\kappa(\phi)\nabla\varphi))|$$

$$
\begin{aligned}
&+ \left| \int_{\partial\Omega} (\phi_n - \phi)\kappa(\phi)\nabla\varphi \cdot \boldsymbol{n}\, ds \right| \\
&\leq \|\kappa(\phi_n) - \kappa(\phi)\| \|\nabla\phi_n\| \|\nabla\varphi\|_{L^\infty} + \underbrace{\|\phi_n - \phi\|}_{\to 0} \|\mathrm{div}\,(\kappa(\phi)\nabla\varphi)\| \\
&\quad + \underbrace{\|\phi_n - \phi\|_{L^2(\partial\Omega)}}_{\to 0} \|\kappa(\phi)\nabla\varphi \cdot \boldsymbol{n}\|_{L^2(\partial\Omega)}.
\end{aligned}
$$

Using the convergence $\|\kappa(\phi_n) - \kappa(\phi)\| \leq \|\kappa'\|_{L^\infty} \|\phi - \phi_n\| \to 0$ and $\|\phi_n - \phi\| \to 0$ we arrive at $\lim_{n\to\infty} T_{3,n} = 0$. Combining all these results yields:

$$
\lim_{n\to\infty} A(\boldsymbol{x}_n, \boldsymbol{y}) \;=\; A(\boldsymbol{x}, \boldsymbol{y}) \quad \forall \boldsymbol{y} \in C^\infty(\Omega)^{d+2}.
$$

Using density and continuity with respect to the second argument, we conclude weak-* convergence:

$$
A(\boldsymbol{x}_n, \cdot) \overset{*}{\rightharpoonup} A(\boldsymbol{x}, \cdot).
$$

□

3.3 Coercivity of the underlying nonlinear operator

In this section, we establish the nonlinear coercivity of the operator restricted to divergence-free velocities. We define the subspace of divergence-free functions:

$$
\boldsymbol{X}_0 := V_\phi \times (\boldsymbol{V}_u \cap \{\boldsymbol{u} \in H^1(\Omega)^d : \mathrm{div}\, \boldsymbol{u} = 0\}),
$$

and introduce the operator $A_0 : \boldsymbol{X}_0 \times \boldsymbol{X}_0 \to \mathbb{R}$, defined by restricting A:

$$
A_0(\phi, \boldsymbol{u}; y_1, \boldsymbol{y}_2) := A(\phi, \boldsymbol{u}, 0; y_1, \boldsymbol{y}_2, 0) = a_1(\phi, \boldsymbol{u}; y_1) + a_2(\phi, \boldsymbol{u}; \boldsymbol{y}_2).
$$

This reformulation highlights the behavior of the nonlinear operator on the divergence-free subspace, setting the stage for demonstrating its coercivity.

Lemma 3.2 *If $\kappa_0 > 0$ is sufficiently large, the coercivity condition is established as*

$$
A_0(\phi, \boldsymbol{u}; \phi, \boldsymbol{u}) \;\geq\; \frac{\mu_0}{2}\|\nabla\phi\|^2 + \frac{\kappa_0}{4}\|\nabla\boldsymbol{u}\|^2 \quad \forall [\phi, \boldsymbol{u}] \in \boldsymbol{X}_0. \tag{20}
$$

The lower bound for κ_0 depends on v_∞, $\boldsymbol{k}$, μ_0, $\boldsymbol{g}$, Ω.

Proof 1. Testing (15) with $\varphi = \phi$:

$$
a_1(\phi, \boldsymbol{u}; \phi) \;=\; \|\kappa(\phi)^{1/2}\nabla\phi\|^2 + (\boldsymbol{f}(\phi), \nabla\phi) - (\boldsymbol{f}(\phi)\cdot\boldsymbol{n}, \phi)_\Gamma.
$$

With $\boldsymbol{f} \circ \phi \in L^2(\Omega)$ and $\|\boldsymbol{f}(\phi)\| \leq |v_\infty| \|\boldsymbol{k}\| \|\phi\|$, we apply the Poincaré and trace inequalities and obtain the bound:

$$
|(\boldsymbol{f}(\phi), \nabla\phi)| + |(\boldsymbol{f}(\phi)\cdot\boldsymbol{n}, \phi)_\Gamma| \;\leq\; \|\phi\| |v_\infty| |\boldsymbol{k}| \left(\int_\Omega |\nabla\phi| + \int_\Gamma |\phi| \right) \leq c\|\nabla\phi\|^2, \tag{21}
$$

with a constant $c > 0$ depending on v_∞, $\boldsymbol{k}$ and Ω. If $\kappa_0 \geq 2c$, then

$$
a_1(\phi, \boldsymbol{u}; \phi) \;\geq\; \frac{\kappa_0}{2}\|\nabla\phi\|^2.
$$

2. Testing of (16) with $\varphi = \boldsymbol{u}$, using Poincaré. and Young's inequality, leads to

$$\begin{aligned} a_2(\phi, \boldsymbol{u}; \boldsymbol{u}) &= \|\mu(\phi)^{1/2}\nabla \boldsymbol{u}\|^2 + (\phi \boldsymbol{g}, \boldsymbol{u}) \\ &\geq \mu_0 \|\nabla u\|^2 - |\boldsymbol{g}| \|\phi\| \|\boldsymbol{u}\| \\ &\geq \frac{\mu_0}{2}\|\nabla u\|^2 - c_\Omega \mu_0^{-1} |\boldsymbol{g}|^2 \|\nabla \phi\|^2. \end{aligned}$$

3. If $\kappa_0 \geq 4c_\Omega |\mu_0^{-1}\boldsymbol{g}|^2$, then

$$A_0(\phi, \boldsymbol{u}; \phi, \boldsymbol{u}) = a_1(\phi, \boldsymbol{u}; \phi) + a_2(\phi, \boldsymbol{u}; \boldsymbol{u}) \geq \frac{\mu_0}{2}\|\nabla \boldsymbol{u}\|^2 + \frac{\kappa_0}{4}\|\nabla \phi\|^2.$$

This confirms the coercivity condition (20). □

3.4 Existence of weak solutions

The following Proposition states the existence of solutions:

Proposition 3.3 *For sufficiently large $\kappa_0 > 0$, the system (12)-(14), combined with boundary conditions (4)-(8), admits a weak solution $\boldsymbol{x} = [\phi, \boldsymbol{u}, p] \in \boldsymbol{X}$.*

Proof Since the bilinear form a_3 is independent of p, equation (19) forms a saddle-point problem. We decompose it into two subproblems:
1. Velocity-solid concentration pair: Find $[\phi, \boldsymbol{u}] \in \boldsymbol{X}_0$ such that:

$$A_0(\phi, \boldsymbol{u}; y_1, \boldsymbol{y}_2) = (b_\phi, y_1) + (\boldsymbol{b}_u, \boldsymbol{y}_2) \quad \forall [y_1, \boldsymbol{y}_2] \in \boldsymbol{X}_0. \tag{22}$$

The form A_0 is coercive (Lemma 3.2) and sequentially weak continuous (Lemma 3.1). Thus, according to Theorem IV.1.2 from (Girault and Raviart, 1986), the nonlinear equation (22) has a solution.

2. Pressure recovery: Determine $p \in V_p$ by

$$b(\boldsymbol{y}_2, p) = (\mu(\phi)\nabla \boldsymbol{u}, \nabla \boldsymbol{y}_2) - (\phi g, \boldsymbol{y}_2) \quad \forall \boldsymbol{y}_2 \in \boldsymbol{V}_u. \tag{23}$$

The bilinear form $b(\cdot, \cdot)$ satisfies an inf-sup condition. Theorem IV.1.4 from (Girault and Raviart, 1986) then guarantees a solution for subproblems (23), completing the proof. □

3.5 Finite element formulation for the stationary equations

To discretize equation (19), we employ equal-order finite elements. We denote the underlying quadrilateral mesh (for $d = 2$) or hexahedral mesh (for $d = 3$) by $\mathcal{T}_h$, where h denotes the mesh size. We define the finite-dimensional subspace $\boldsymbol{X}_h \subset \boldsymbol{X}$, consisting of continuous, piecewise bi-linear Q_1 functions on each cell $K \in \mathcal{T}_h$ for the discrete variable $\boldsymbol{x}_h \in \boldsymbol{X}_h$. Since equal-order elements are not inf-sup stable for the bilinear form $b(\cdot, \cdot)$, stabilization is required. The stabilized discrete form $A_h : \boldsymbol{X}_h \times \boldsymbol{X}_h \to \mathbb{R}$ is defined as:

$$A_h(\boldsymbol{x}, \boldsymbol{y}) := A(\boldsymbol{x}, \boldsymbol{y}) + S_\beta(\phi, y_1) + S_\delta(p, y_3).$$

where $\boldsymbol{x} = [\phi, \boldsymbol{u}, p]$ and $\boldsymbol{y} = [y_1, \boldsymbol{y}_2, y_3]$.

Both S_β and S_δ use a local projection operator $\pi_h := id - \kappa_h$, where $\kappa_h : L^2(\Omega)^d \to D_h^d$ is a fluctuation operator onto the space D_h of patch-wise discontinuous, constant functions. For further details, we refer to (Becker and Braack, 2001).

- **Pressure stabilization:** To address the lack of discrete inf-sup stability, we use:

$$S_\delta(p,y) := \sum_{K\in\mathcal{T}_{2h}} \delta_K \int_K \pi_h\nabla p\cdot\pi_h\nabla y$$

  where $\delta_K = h_K^2$.
- **Convective stabilization:** If κ is small compared to $|\boldsymbol{u}|$, the convective term in a_1 requires stabilization via:

$$S_\beta(\phi,y) := \sum_K \beta_K \int_K \pi_h\nabla\phi\cdot\pi_h\nabla y,$$

  where $\beta_K = \beta_0 h_K$.

Finally, we solve the nonlinear system:

$$\boldsymbol{x}_h\in\boldsymbol{X}_h: \quad A_h(\boldsymbol{x}_h,\boldsymbol{y}_h) = \langle \boldsymbol{b},\boldsymbol{y}_h\rangle \quad \forall \boldsymbol{y}_h\in\boldsymbol{X}_h. \tag{24}$$

3.6 A general solvability result for finite-dimensional nonlinear equations

To establish the existence of discrete solutions for the nonlinear system (22), the approach used in the proof of Proposition 3.3 is not directly applicable. The main challenge arises from the additional coupling of pressure in the divergence equation, which disturbs the saddle-point structure. As a result, we resort to the existence theory developed in (Becker and Braack, 2024a), along with its extension to specific systems in (Becker and Braack, 2024b).

The starting point is a Banach space of the form $X := Y\times Q$ and an operator $\mathcal{A}: X\to X'$ defined for $x=[v,p]^T\in X$ as follows:

$$\mathcal{A}(x) := \begin{pmatrix} C(v)\\ 0\end{pmatrix} + \begin{bmatrix} L & -B^*\\ B & T\end{bmatrix}\begin{bmatrix} v\\ p\end{bmatrix}, \tag{25}$$

where $L\in L(Y,Y')$ (i.e. a linear continus operator from Y into its dual Y'), $T\in L(Q,Q')$, and $B\in L(V,Q')$ are linear operators. The operator $C: Y\to Y'$ is not necessarily linear, but sequentially weak continuity is required: $x_n \rightharpoonup x$ implies $C(x_n)\overset{*}{\rightharpoonup} C(x)$. For a given right-hand side $b\in Y'\times\{0\}$, a sufficient condition for the existence of a solution to the equation

$$x\in X: \quad \mathcal{A}(x) = b \tag{26}$$

is provided by the following Proposition, see Prop. 4.3 in (Becker and Braack, 2024b).

Proposition 3.4 *The operator $\mathcal{A}: X\to X'$ is assumed to be of form (25) with the properties:*

- *(A1) L is coercive.*
- *(A2a) It holds $\langle Tp, p\rangle_{Q',Q} \geq 0$, for all $p \in Q$.*
- *(A2b) There exists $\Theta \in L(Q, Y)$ which is onto, and a constant $c_T \in \mathbb{R}$ s.t.*

$$\langle B^* p, \Theta p\rangle_{Y',Y} \geq \frac{1}{2}\|p\|_Q^2 - c_T \langle Tp, p\rangle_{Q',Q} \qquad \forall p \in Q.$$

- *(A3) It holds $\langle C(u), u\rangle_{Y',Y} \geq 0$ for all $u \in Y$.*
- *(A4) $\|C(u)\|_{Y'} \leq c_N \|u\|_Y^2$ for all $u \in Y$.*

Then, equation (26) has a solution.

Note that assumption (A2b) acts as a discrete inf-sup condition.

3.7 Solvability of the finite-dimensional sedimentation equations

Now, we apply Proposition 3.4 to establish the solvability of the sediment equations discretized using the finite element method described above.

Proposition 3.5 *For sufficiently large $\kappa_0 > 0$, equation (24) has a solution.*

Proof We begin by reformulating equation (22) to match the structure required in Proposition 3.4. The operator in question takes the form in (25), where the variable v corresponds to the tuple $[\phi, \boldsymbol{u}]$. Specifically, we incorporate the diffusive components of a_1 and a_2 into the linear operator L, which is defined as:

$$\langle L[\phi, \boldsymbol{u}], [y_1, \boldsymbol{y}_2]\rangle := \frac{\kappa_0}{2}(\nabla\phi, \nabla y_1) + \frac{\mu_0}{2}(\nabla \boldsymbol{u}, \nabla \boldsymbol{y}_2) - (\phi \boldsymbol{g}, \boldsymbol{y}_2) + S_\beta(\phi, y_1).$$

The remaining (nonlinear) part is given by:

$$\begin{aligned}\langle C[\phi, \boldsymbol{u}], [y_1, \boldsymbol{y}_2]\rangle \;=\; & (\boldsymbol{u}\cdot\nabla\phi, y_1) + \tfrac{1}{2}((\operatorname{div}\boldsymbol{u})\phi, y_1) + (\boldsymbol{f}(\phi), \nabla y_1) - (\boldsymbol{f}(\phi)\cdot\boldsymbol{n}, y_1)_\Gamma \\ & + ((\kappa(\phi) - \tfrac{1}{2}\kappa_0)\nabla\phi, \nabla y_1) + ((\mu(\phi) - \tfrac{1}{2}\mu_0))\nabla\boldsymbol{u}, \nabla\boldsymbol{y}_2).\end{aligned}$$

As already shown in Lemma 3.1, the nonlinear terms are sequentially weakly continuous. The operator T consists of the pressure-pressure coupling $\langle Tp, q\rangle = S_\delta(p, q)$. We now verify properties (A1)-(A4) step by step.
(A1): We establish a lower bound for L. We have the following estimate:

$$\begin{aligned}\langle L[\phi, \boldsymbol{u}], [\phi, \boldsymbol{u}]\rangle \;&=\; \frac{\kappa_0}{2}\|\nabla\phi\|^2 + \frac{\mu_0}{2}\|\nabla\boldsymbol{u}\|^2 + S_\beta(\phi, \phi) - (\phi\boldsymbol{g}, \boldsymbol{u}) \\ &\geq\; (\tfrac{1}{2}\kappa_0 - c|\boldsymbol{g}|\mu_0^{-1})\|\nabla\phi\|^2 + \frac{\mu_0}{2}\|\nabla\boldsymbol{u}\|^2.\end{aligned}$$

This shows that L is coercive, provided that $\kappa_0 > 0$ is sufficiently large, thus verifying property (A1).
(A2a): This property is clearly satisfied because $\langle Tp, p\rangle = S_\delta(p, p) \geq 0$.
(A2b): This property is shown in (Becker and Braack, 2024b) for LPS stabilization S_δ.
(A3): From the identity $(\boldsymbol{u}\cdot\nabla\phi, \phi) + \frac{1}{2}((\operatorname{div}\boldsymbol{u})\phi, \phi) = 0$ we define $\kappa_\delta := \kappa(\phi) - \frac{1}{2}\kappa_0 \geq \frac{1}{2}\kappa_0 > 0$

and $\mu_\delta := \mu(\phi) - \frac{1}{2}\mu_0 \geq \frac{1}{2}\mu_0 > 0$. With this notation, we obtain the following estimate for the nonlinear term:

$$\langle C[\phi, \boldsymbol{u}], [\phi, \boldsymbol{u}] \rangle \;=\; (\boldsymbol{f}(\phi), \nabla\phi) - (\boldsymbol{f}(\phi) \cdot \boldsymbol{n}, \phi)_\Gamma + \|\kappa_\delta^{1/2} \nabla\phi\|^2_{L^2(\Omega)} + \|\mu_\delta^{1/2} \nabla\boldsymbol{u}\|^2_{L^2(\Omega)}.$$

Using the upper bound from (21), we obtain:

$$\langle C[\phi, \boldsymbol{u}], [\phi, \boldsymbol{u}] \rangle \;\geq\; (\tfrac{1}{2}\kappa_0 - c)\|\nabla\phi\|^2_{L^2(\Omega)} + \tfrac{1}{2}\mu_0 \|\nabla\boldsymbol{u}\|^2_{L^2(\Omega)}$$

For sufficiently large κ_0, this verifies property (A3).
(A4) To show that the norm of $C[\phi, \boldsymbol{u}]$ grows at most quadratically, we analyze the scaling behavior of its individual terms. The convective terms $(\boldsymbol{u} \cdot \nabla\phi, y_1) + \frac{1}{2}((\operatorname{div}\boldsymbol{u})\phi, y_1)$ scale quadratically. The diffusive terms $((\kappa(\phi) - \frac{1}{2}\kappa_0)\nabla\phi, \nabla y_1)$ and $((\mu(\phi) - \mu_0))\nabla\boldsymbol{u}, \nabla\boldsymbol{y}_2)$, scale linearly, as the coefficients $\kappa(\phi)$ and $\mu(\phi)$ are bounded in $L^\infty(\Omega)$. Finally, the nonlinearity of $\boldsymbol{f}(\phi)$ contributes at most linear growth, given that $\|\boldsymbol{f}(\phi)\| \leq c\|\phi\|$. Therefore, the overall norm of $C[\phi, \boldsymbol{u}]$ satisfies the required quadratic bound. □

3.8 Finite element formulation for the time-dependent equations

Discretizing the time dependence with the implicit (backward) Euler method results for $\boldsymbol{x}_n := [\phi_n, \boldsymbol{u}_n, p_n]$ in the system

$$\tau\mathcal{I}\boldsymbol{x}_n + \mathcal{A}(\boldsymbol{x}_n) \;=\; \tau\mathcal{I}\boldsymbol{x}_{n-1} \quad n \geq 1, \tag{27}$$

with initial conditions ϕ_0 and $\boldsymbol{u}_0$, the time step $\tau = t_n - t_{n-1}$, and the operator

$$\langle \mathcal{I}\boldsymbol{x}, \boldsymbol{y} \rangle \;:=\; (\phi, y_1) + (\boldsymbol{u}, \boldsymbol{y}_2).$$

The solvability of the system (27) follows immediately with Proposition (3.5) of the stationary case and the L^2-coercivity of the operator $\mathcal{I}$ w.r.t. the variables ϕ and $\boldsymbol{u}$: $\langle \mathcal{I}\boldsymbol{x}, \boldsymbol{x} \rangle = \|\phi\|^2 + \|\boldsymbol{u}\|^2$.

4 Numerical examples

We now present two numerical examples to demonstrate the results of section 3. The triangulation employs equal-order Q_1 LPS-stable finite element quadrilaterals in 2D and Q_1 cuboids in 3D. In the 2D example, the elements are perfectly shape-regular in the interior. In the 3D case, the elements remain regular in the interior but become more loosely quasi-uniform towards the inflow and outflow regions, where they are constricted into a tighter space. The inflow $\boldsymbol{u}_{\text{in}}$ is prescribed using a smooth parabolic profile to ensure that the Dirichlet conditions on the edges of the neighboring faces are well-defined. To coerce asymmetry in the laminar stream in each case, an obstacle is placed in the flow path. The simulation is conducted over the time interval $T := [0 : 300]$, which is discretized into fixed timesteps of $\Delta t = 0.01$ using the implicit Euler method as the time integrator. In both examples, the stabilization parameters are chosen as $\delta_K^T \propto h_K$ and $\delta_K^F \propto h_K$. The nonlinear system is solved using the Newton method with a tolerance of 10^{-9}. The resulting linear systems are solved using the multigrid method with LU preconditioning.

Parameter	Description	Value
D_0	Hydrodynamic self-diffusion constant	$0.012\ \mathrm{m}^2\ \mathrm{sec}^{-1}$
ρ_s	Sediment density	$1750\ \mathrm{kg\ m}^{-3}$
ρ_f	Salt water density	$998\ \mathrm{kg\ m}^{-3}$
v_∞	Particle Stokes velocity	$0.01\ \mathrm{m\ sec}^{-1}$
$\phi_{\max}$	Maximum sediment concentration	$0.1\ \mathrm{kg\ m}^{-3}$
$\tilde{\phi}_{\max}$	Nominal maximum sediment concentration	$0.12\ \mathrm{kg\ m}^{-3}$
β	Newtonian viscosity exponent	2
n_{RZ}	Sediment dependent exponent	2
g	Gravitational acceleration	$9.8\ \mathrm{m\ s}^{-2}$
σ_0	Flocculation scale	0.1 Pa
α	Flocculation compressibility exponent	3
ϕ_c	Critical sediment concentration	$0.08\ \mathrm{kg\ m}^{-3}$
Initial and Boundary Conditions		
$\boldsymbol{u}_0$	Initial velocity	$(0,0,0)^{\mathrm{T}}\ \mathrm{m\ sec}^{-1}$
p_0	Stationary pressure solution	$\phi g z$
ϕ_0	Initial concentration gradient	$5\times10^{-3}(H-z^2)/H^2\ \mathrm{kg\ m}^{-3}$
μ_0	Viscosity scale	0.01
$\boldsymbol{u}_{\mathrm{in}}$	Inlet flow velocity	$(0.1,0)^{\mathrm{T}},\ (0,0.1,0)^{\mathrm{T}}\ \mathrm{m\ sec}^{-1}$
ϕ_{in}	Inflowing sediment concentration	$0.05\ \mathrm{kg\ m}^{-3}$

Table 1: Parameters and units for the transport and flow equations designed for a convection-dominant setting, i.e. $\|\boldsymbol{u}\| \gg \kappa(\phi)$. H is the maximum domain height, i.e. the fluid depth.

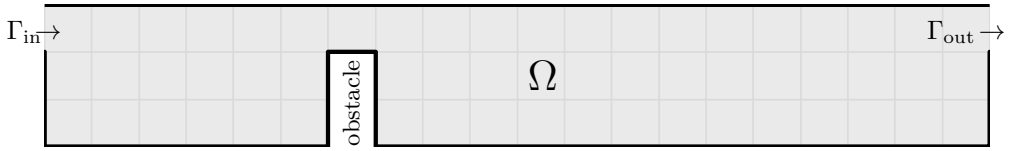


Fig. 2: The 2D domain Ω showing the in- and outflow boundaries and the flow hindrance. The gray shaded region represents the Q_1-elements on the coarsest mesh.

4.1 Parameters and initial conditions

The model parameters used in the examples are listed in Table 1. They were selected so as to induce a strong convection-dominant flow, allowing us to assess the effectiveness of the LPS stabilization on the system. To impose an initial concentration gradient, we set $\phi(z) = \phi_0(H - z^2)/H^2$, resulting in a smooth vertical profile of $\phi(z) \in [0, \phi_0]$, where H is the maximum domain height.

4.2 2D Example

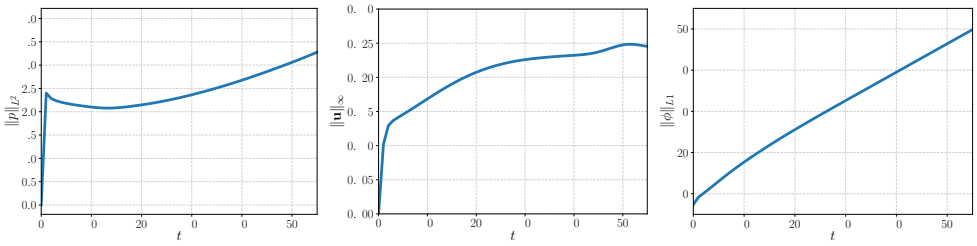
The two dimensional example domain was constructed as a long $1 \times 7\,\mathrm{m}$ channel, with a 0.3 m width inlet positioned at the upper left and a 0.3 m width outflow at the upper right. Figure 2 provides a detailed illustration of the domain with the mesh

shown at its coarsest refinement. To perturb the fluid flow and break symmetry in the stream direction, a rectangular obstacle was placed on the bottom, one-third of the way from the inlet. Such obstacles (or *baffles*) are commonly found in water clarifiers at treatment facilities and have been modeled by researchers, including (Rao et al., 2007).

Figures 3(a) and 3(b) depict the solution norms $\|p\|_{L^2}$ and $\|\boldsymbol{u}\|_{\infty}$ for the initial settling phase of the pressure and velocity. Figure 3(c) depicts the expected growth in the L_1-norm due to sediment accretion. This growth occurs as material is convected into the domain from the Dirichlet condition on the inflow face and accumulates within the domain, owing to the no-flux boundary condition imposed on the convection equation.

The numerical solutions to equations (1)-(8) are presented as pressure-velocity and concentration pairs in Figure 4 for three interesting time points, illustrating the progression of the convected material through the channel. In the left chamber, the pressure is higher than in the right and develops a vertical gradient as sediment accumulates at a higher concentration due to containment by the barrier. This behavior is expected, as the Stokes equation (2) directly depends on ϕ.

The strongly convective nature of the flow is evident in Figures 4(b), (d), and (f). After the initial establishment of pressure and velocity solutions, achieved within 10 iterations at $t = 1$, Figure 4(b) shows only a minimal diffusive layer at the leading edge of the concentration front as it begins to convect rightwards. The downward flow driven by gravitational forces becomes visible in Figures 4(d) and (f). Over time, the diffusive layer gradually expands, and by Figure 4(f), it stretches further along the stream direction, feeding into the right-hand chamber. This behavior contrasts with what may be observed using more diffusive methods such as artificial diffusion or upwinding (Caussignac and Touzani, 1988), often causing a more pronounced or smeared diffusive layer. Additionally, Figure 4(c) captures a transient circular motion in the velocity field within the left chamber, which quickly dissipates. Interestingly, no such recirculation was observed when the material flowed over the obstacle into the right-hand chamber.



(a) (b) (c)

Fig. 3: The relevant norms of the two-dimensional pressure and velocity solution over time, and the increasing L_1-norm of the concentration.

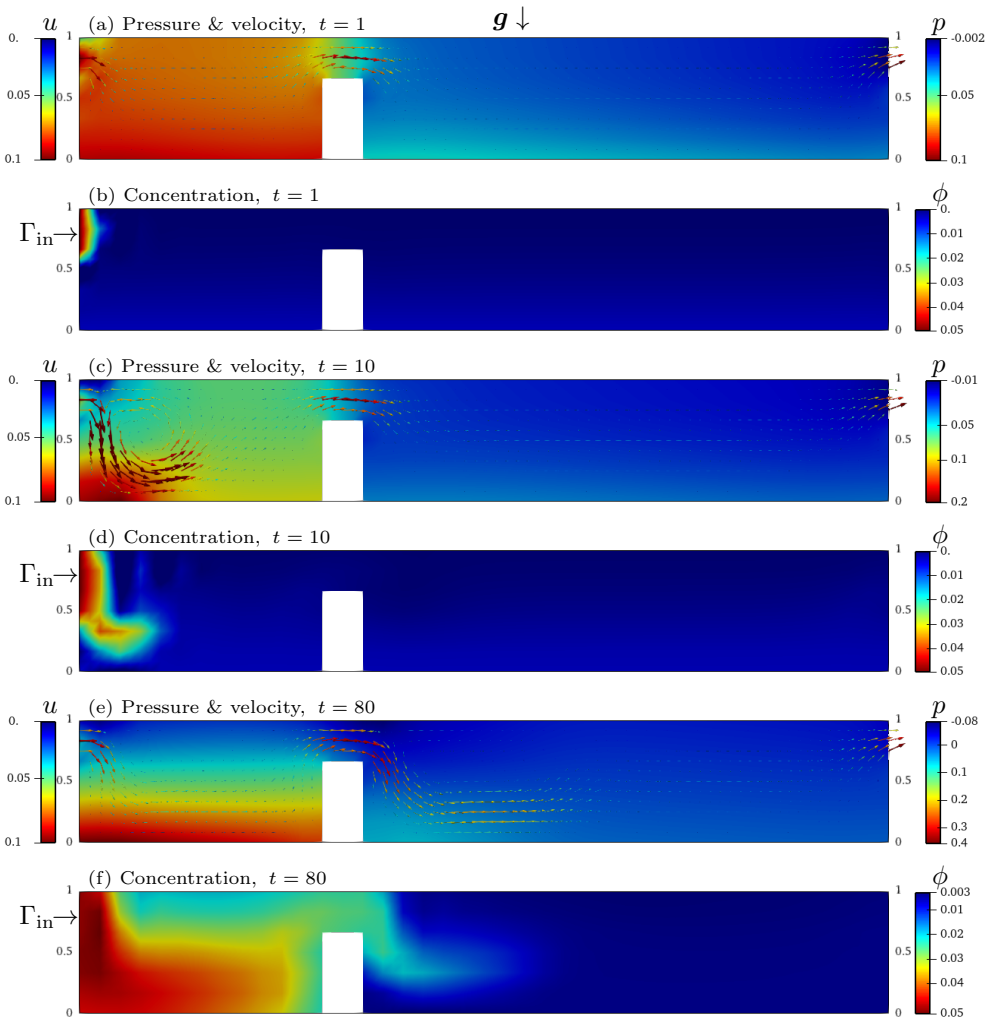


Fig. 4: LPS stabilized convection-dominated sedimentation flow in 2D at three time points, $t = 1$, $t = 10$, and $t = 80$. The white rectangular shape is the obstacle. For Figures (a), (c) and (e) showing both u and p, the colored arrows represent the direction and magnitude of the velocity $\boldsymbol{u}$, and p as the background color gradient.

4.3 3D Example

The 3D domain is also constructed as an elongated channel with elevated inflow and outflow faces, sloping down into a sunken basin in the center. The layout is shown in Figure 5. In this case, the obstacle is a three-dimensional chicane that forces the flow through an S-shaped constriction at the base of the ramp. As a result, the stream must change direction in all three dimensions as it progresses from the inlet to the outlet.

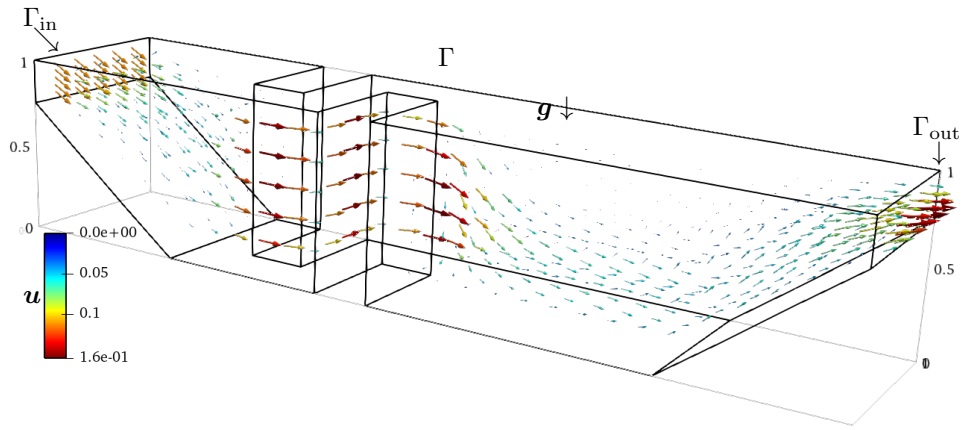
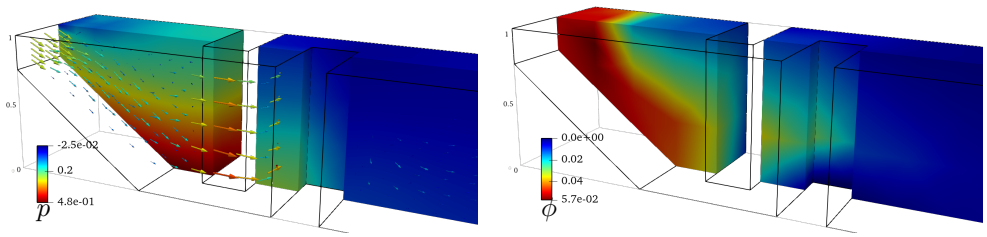


Fig. 5: The 3D domain Ω detailing the chicane, boundaries Γ, Γ_{in}, Γ_{out}, and the velocity solution $\boldsymbol{u}(t)$ after a settling time at $t = 120$.



(a) Pressure (b) Concentration ϕ

Fig. 6: Vertical cutaway through the center of the 3D domain at $t = 50$ to illustrate the internal pressure and concentration near the inlet, which has a bulge in the interior due to the no-slip velocity boundary condition. The dark black lines are the domain edges.

All parameters remain consistent with those used in the 2D example. As seen in Figure 6(a), a higher pressure region establishes itself in the first chamber before the obstacle, similar to the behavior observed in the 2D case. Figure 7 shows two concentration contours that illustrate convection through the chicane. The higher concentration contour, shown in red, accumulates near the inflow and begins to flow around the obstacle. A noticeable drop-off in concentration height occurs as the material moves further from the inlet. The blue-shaded contour highlights the leading edge of the convected quantity ϕ.

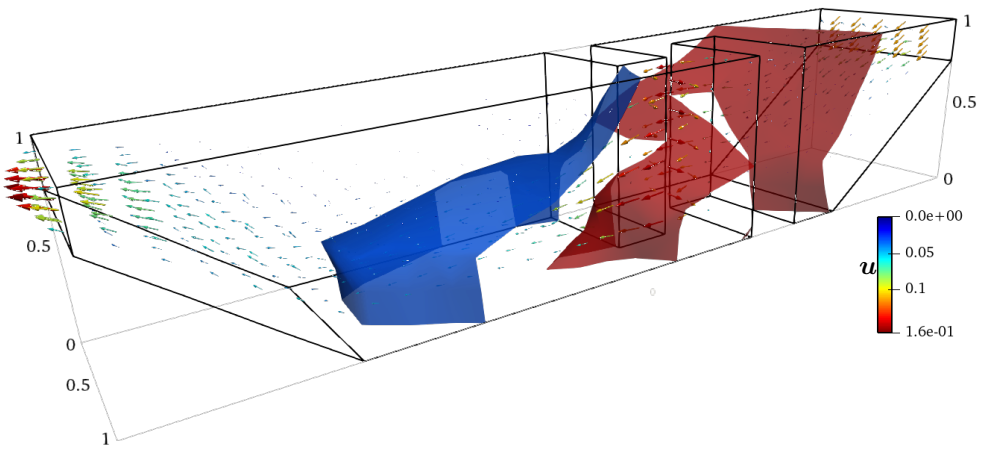


Fig. 7: Contours of material under convection through the chicane shown from the rear for improved visibility, with contours for $\phi = 0.045$ (semi-transparent red), and $\phi = 0.025$ (blue).

5 Conclusions

In this study, we have developed the solution theory for a nonlinear system of partial differential equations that describe the sedimentation-consolidation process. We employed the finite element method (FEM) combined with local-projection stabilization and demonstrated, via numerical applications of the theory, the stability of our approach. Our results show that this method effectively captures the complex dynamics of sediment transport, including the interplay between fluid flow and particle settling. The incorporation of LPS has proven to be essential in addressing numerical instabilities that often arise in simulations of sediment-laden flows, particularly in convective-dominated flows and the associated nonlinear behavior.

Our findings indicate that the combination of FEM and LPS provides a high fidelity in simulation results for such problems. This pairing offers a powerful tool for predicting sedimentation behavior across various environmental and industrial applications. Furthermore, the computational efficiency of this methodology makes it suitable for larger, more complex models, enabling further exploration of sediment dynamics under different conditions.

Future research could extend this framework to include multi-phase flows and examine the effects of varying particle sizes and shapes, particularly in cases where flocculation plays a critical role. This would further enhance our understanding of sedimentation processes. Moreover, an a priori error estimate for these kinds of nonlinear equations, would be desirable. Overall, this work contributes to advancing computational techniques in sedimentation dynamics and opens a new branch for future investigations into sediment transport phenomena, with significant implications for environmental management and engineering practices.

Statements and Declarations

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

References

Adams, R.A.: Sobolev Spaces. Pure and Applied Mathematics, vol. 65, p. 268. Academic Press, Inc., New York-London (1975)

Acrivos, A., Herbolzheimer, E.: Enhanced sedimentation in settling tanks with inclined walls. Journal of Fluid Mechanics **92**(3), 435–457 (1979) https://doi.org/10.1017/S0022112079000720

Becker, R., Braack, M.: A finite element pressure gradient stabilization for the stokes equations based on local projections. CALCOLO **38**, 173–199 (2001) https://doi.org/10.1007/s10092-001-8180-4

Becker, R., Braack, M.: The concept of mapped coercivity for nonlinear operators in Banach spaces. to appear 2025 in J. Functional Anal. (2024). https://doi.org/10.48550/arXiv.2305.16783

Becker, R., Braack, M.: Mapped coercivity for the stationary Navier-Stokes equations and their finite element approximations (2024). https://doi.org/10.48550/arXiv.2412.00494

Benn, F.A., Fawell, P.D., Halewood, J., Austin, P.J., Costine, A.D., Jones, W.G., Francis, N.S., Druett, D.C., Lester, D.: Sedimentation and consolidation of different density aggregates formed by polymer-bridging flocculation. Chemical Engineering Science **184**, 111–125 (2018) https://doi.org/10.1016/j.ces.2018.03.037

Bürger, R., Kumar, S., Ruiz-Baier, R.: Discontinuous finite volume element discretization for coupled flow–transport problems arising in models of sedimentation. Journal of Computational Physics **299**, 446–471 (2015) https://doi.org/10.1016/j.jcp.2015.07.020

Braack, M.: Local projection stabilization for the stokes equation with Neumann condition. Computer Methods in Applied Mechanics and Engineering **334**, 507–522 (2018) https://doi.org/10.1016/j.cma.2018.02.008

Bürger, R., Ruiz-Baier, R., Torres, H.: A stabilized finite volume element formulation for sedimentation-consolidation processes. SIAM Journal on Scientific Computing **34**(3), 265–289 (2012) https://doi.org/10.1137/110836559

Bürger, R., Wendland, W.L., Concha, F.: Model equations for gravitational sedimentation-consolidation processes. ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik **80**(2), 79–92 (2000) https://doi.org/10.1002/(SICI)1521-4001(200002)80:2⟨79::AID-ZAMM79⟩3.0.CO;2-Y

Chou, S.H., Kwak, D.Y.: A covolume method based on rotated bilinears for the generalized stokes problem. SIAM Journal on Numerical Analysis **35**(2), 494–507 (1998) https://doi.org/10.1137/S0036142996299964

Caussignac, P., Touzani, R.: Linear conforming and nonconforming upwind finite elements for the convection-diffusion equation. IMA J. Numer. Anal. **8**(1), 85–103

(1988) https://doi.org/10.1093/imanum/8.1.85
Dannberg, J., Heister, T.: Compressible magma/mantle dynamics: 3-d, adaptive simulations in aspect. Geophysical Journal International **207**(3), 1343–1366 (2016) https://doi.org/10.1093/gji/ggw329
Ekama, G.A., Barnard, J.L., Günthert, F.W.: Secondary settling tanks : theory, modelling, design and operation. (1997). https://api.semanticscholar.org/CorpusID:108099753
Ewing, R.E., Lin, T., Lin, Y.: On the accuracy of the finite volume element method based on piecewise linear polynomials. SIAM Journal on Numerical Analysis **39**(6), 1865–1888 (2002). Accessed 2024-08-13
Fitch, B.: Current theory and thickener design. Industrial & Engineering Chemistry **58**(10), 18–28 (1966) https://doi.org/10.1021/ie50682a006
Guazzelli, E., Hinch, J.: Fluctuations and instability in sedimentation. Annual Review of Fluid Mechanics **43**(Volume 43, 2011), 97–116 (2011) https://doi.org/10.1146/annurev-fluid-122109-160736
Guazzelli, E., Morris, J.F., Pic, S.: A Physical Introduction to Suspension Dynamics. Cambridge Texts in Applied Mathematics, vol. 1. Cambridge University Press, Cambridge (2011). https://doi.org/10.1017/CBO9780511894671
Girault, V., Raviart, P.-A.: Finite Element Methods for Navier-Stokes Equations. Springer Series in Computational Mathematics, vol. 5, p. 374. Berlin (1986). https://doi.org/10.1007/978-3-642-61623-5
Hill, W.D., Rothfus, R.R., Li, K.: Boundary-enhanced sedimentation due to settling convection. International Journal of Multiphase Flow **3**(6), 561–583 (1977) https://doi.org/10.1016/0301-9322(77)90030-1
Jha, B.K., Musa, M.K., Ajibade, A.O.: Time-dependent natural convection fluid flow of stably stratified fluid in an asymmetrically heated vertical channel filled with an anisotropic porous material. Heat Transfer **52**(1), 665–686 (2023) https://doi.org/10.1002/htj.22711
Kleine, D., Reddy, B.D.: Finite element analysis of flows in secondary settling tanks. International Journal for Numerical Methods in Engineering **64**(7), 849–876 (2005) https://doi.org/10.1002/nme.1373
Kynch, G.J.: A theory of sedimentation. Trans. Faraday Soc. **48**, 166–176 (1952) https://doi.org/10.1039/TF9524800166
Lorca, S.A., Boldrini, J.L.: The initial value problem for a generalized boussinesq model. Nonlinear Analysis: Theory, Methods & Applications **36**(4), 457–480 (1999) https://doi.org/10.1016/S0362-546X(97)00635-4
Lee, S.-R., Kim, Y.-S., Kim, Y.-S.: Analysis of sedimentation/consolidation by finite element method. Computers and Geotechnics **27**(2), 141–160 (2000) https://doi.org/10.1016/S0266-352X(00)00014-8
McLachlan, D.S.: Equations for the conductivity of macroscopic mixtures. Journal of Physics C: Solid State Physics **19**(9), 1339 (1986) https://doi.org/10.1088/0022-3719/19/9/007
Ruiz-Baier, R., Torres, H.: Numerical solution of a multidimensional sedimentation problem using finite volume-element methods. Applied Numerical Mathematics **95**, 280–291 (2015) https://doi.org/10.1016/j.apnum.2013.12.006 . Fourth Chilean

Workshop on Numerical Analysis of Partial Differential Equations (WONAPDE 2013)

Rao, R.R., Mondy, L.A., Altobelli, S.A.: Instabilities during batch sedimentation in geometries containing obstacles: A numerical and experimental study. International Journal for Numerical Methods in Fluids **55**(8), 723–735 (2007) https://doi.org/10.1002/fld.1483

Ronghua, L.: Generalized difference methods for a nonlinear dirichlet problem. SIAM Journal on Numerical Analysis **24**(1), 77–88 (1987). Accessed 2024-08-13

Rui, H.-x.: Analysis on a finite volume element method for stokes problems. Acta Mathematicae Applicatae Sinica **21**(3), 359–372 (2005) https://doi.org/10.1007/s10255-005-0243-x

Tabata, M., Tagami, D.: Error estimates of finite element methods for nonstationary thermal convection problems with temperature-dependent coefficients. Numerische Mathematik **100**(2), 351–372 (2005) https://doi.org/10.1007/s00211-005-0589-2

Conclusion and outlook 6

In this thesis we focused on solving three distinct applied problems found in natural processes with the aim of extending current research results.

Regarding our application to the geosciences, we improved the efficiency of the recently proposed methods involving coupled Stokes-Darcy flow by proposing a coercive linear operator that allowed for the use of pure equal-order finite element pairs without the need for additional stabilization terms in the bilinear form. Thus, while still complying with the well known CFL condition (2.3) time-step restriction, this allowed for an explicit time discretization in which only an inversion of the mass matrix was required.

Our investigation into sustainable fishing was focused on the enhancement and optimization of a spatio-temporal fishing model using the concepts of optimal control discussed in section 2.5. To achieve this we proposed broadening current models to take in to consideration the environmental effects of species predation, ocean currents, and trawler monitoring in relation to MPAs. To bring an additional level of realism to our numerical results, we simulated the system only using previously documented biological parameters where they were available. For undocumented parameters, such as a diffusion coefficient for the biomass, we calculated biologically plausible values based on known limits, for example swimming speed. Some of these parameters could only be sourced from aquaculture studies which are, admittedly, only loosely related to more realistic values one would find in free-field studies. Since our model serves as a form of prototype, future work in this area could be the incorporation of more complete trawler numbers, their harvesting capabilities, as well as all the MPAs they have within their range. Model extensions could also

take into account the species specific cyclical reproduction and maturation rates. A future model expansion should also take into consideration the effects of water temperature on the biomass life-cycle. This would involve an extension in to three dimensions so that the movement of fish between stratified layers of varying temperatures could better reflect their temperature sensitivity and accessibility to harvesting.

Finally, regarding our interest in the sedimentation process, we focused on stabilizing the nonlinear and highly sensitive system through the application of the LPS method as detailed in section 2.4.4. Prior to this research, finding sedimentation-consolidation solutions was almost exclusively focused on the finite difference method with variations of residual-based stabilization methods. The work proved the coercivity of the underlying nonlinear operator and the existence of weak solutions in our application. Our findings showed that the incorporation of LPS as a stabilization mechanism in a finite element solver for this class of problem is effective in providing smooth solutions that effectively address numerical instabilities.

Bibliography

Achdou, Yves et al. (2014). "Partial differential equation models in macroeconomics". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 372.2028. DOI: 10.1098/rsta.2013.0397.

Becker, R. and M. Braack (2004). "A two-level stabilization scheme for the Navier-Stokes equations". In: *Numerical mathematics and advanced applications*. Ed. by M. Feistauer et al. Springer, pp. 123–130.

Becker, R. and B. Vexler (2007). "Optimal control of the convection-diffusion equation using stabilized finite element methods". In: *Numer. Math.* 106, pp. 349–367. DOI: 10.1007/s00211-007-0067-0.

Becker, Roland, Dominik Meidner, and Boris Vexler (2007). "Efficient numerical solution of parabolic optimization problems by finite element methods". In: *Optimization Methods and Software* 22.5, pp. 813–833. DOI: 10.1080/10556780701228532.

Benzi, M., G.H. Golub, and J. Liesen (2005). "Numerical solution of saddle point problems". In: *Acta numerica* 14, pp. 1–137.

Braack, M. (2009). "Optimal Control in Fluid Mechanics by Finite Elements with Symmetric Stabilization". In: *SIAM Journal on Control and Optimization* 48.2, pp. 672–687. DOI: 10.1137/060653494. URL: https://doi.org/10.1137/060653494.

Brooks, A.N. and T.J. Hughes (1982). "Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations". In: *Computer Methods in Applied Mechanics and Engineering* 32.1, pp. 199–259. DOI: 10.1016/0045-7825(82)90071-8.

Burman, E. and P. Hansbo (2004). "Edge stabilization for Galerkin approximations of convection-diffusion-reaction problems". In: *Computer methods in applied mechanics and engineering* 193.15-16, pp. 1437–1453.

Collis, Samuel et al. (Apr. 2002). "Analysis of the Streamline Upwind/Petrov Galerkin Method Applied to the Solution of Optimal Control Problems". In.

Crowley, Philip H. and Elizabeth K. Martin (1989). "Functional Responses and Interference within and between Year Classes of a Dragonfly Population". In: *Journal of the North American Benthological Society* 8.3, pp. 211–221. DOI: 10.2307/1467324.

Dacorogna, B. (1989). "Direct Methods in the Calculus of Variations". In: *Appl. Math. Sci.* 78.

Donea, J. and A. Huerta (2003). *Finite element methods for flow problems*. John Wiley & Sons.

Ern, A. and J.L. Guermond (2013). *Theory and practice of finite elements*. Vol. 159. Springer Science & Business Media.

Guermond, J.L. (1999). "Stabilization of Galerkin approximations of transport equations by subgrid modeling". In: *ESAIM: Mathematical Modelling and Numerical Analysis* 33.6, pp. 1293–1316.

— (2001). "Subgrid stabilization of Galerkin approximations of linear monotone operators". In: *IMA Journal of Numerical Analysis* 21.1, pp. 165–197.

Hughes, T.J. and A.N. Brooks (1979). "A multi-dimensional upwind scheme with no crosswind diffusion". In: *Finite Element Methods for Convection Dominated Flows*. Ed. by T.J. Hughes. Vol. 34. New York: American Society of Mechanical Engineers, pp. 19–35.

Hughes, T.J., L.P. Franca, and G.M. Hulbert (1989). "A new finite element formulation for computational fluid dynamics: VIII. The Galerkin/least-squares method for advective-diffusive equations". In: *Computer methods in applied mechanics and engineering* 73.2, pp. 173–189.

Jahn, Johannes (2014). *Introduction to the Theory of Nonlinear Optimization*. Springer Berlin, Heidelberg.

John, V. (2016). *Finite Element Methods for Incompressible Flow Problems*. Springer Series in Computational Mathematics. Springer International Publishing. ISBN: 9783319457505.

Johnson, C., U. Navert, and J. Pitkaranta (1984). "Finite element methods for linear hyperbolic problems". In: *Computer methods in applied mechanics and engineering* 45, pp. 285–312.

Kirk, Donald E. (1971). *Optimal control theory—an introduction*. Prentice Hall Inc.

Knobloch, P. (2009). "On a variant of the local projection method stable in the SUPG norm". In: *Kybernetika* 45.4, pp. 634–645.

— (2010). "A generalization of the local projection stabilization for convection-diffusion-reaction equations". In: *SIAM journal on numerical analysis* 48.2, pp. 659–680.

Löhner, R. (2008). *Applied computational fluid dynamics techniques: an introduction based on finite element methods*. John Wiley & Sons.

Lube, Gert (Jan. 1994). "Stabilized Galerkin Finite Element Methods for Convection Dominated and Incompressible Flow Problems". In: *Inst. of Mathematics, Banach Cent. Publ* 29, pp. 85–104. DOI: 10.4064/-29-1-85-104.

Matthies, G., P. Skrzypacz, and L. Tobiska (2007). "A unified convergence analysis for local projection stabilisations applied to the Oseen problem". In: *ESAIM: Mathematical Modelling and Numerical Analysis* 41.4, pp. 713–742.

McDonough, J. M. (2007). "Lectures in computational fluid dynamics of incompressible flow: Mathematics, algorithms and implementations". In.

McKenzie, D. (1984). "The generation and compaction of partially molten rock". In: *J. Petrol.* 25.3, pp. 713–765.

Meidner, Dominik (Mar. 2008). "Adaptive Space-Time Finite Element Methods for Optimization Problems Governed by Nonlinear Parabolic Systems". Available at https://doi.org/10.11588/heidok.00008272. PhD thesis. Heidelberg, Germany: Ruprecht-Karls-Universität.

Morton, K.W. (2019). *Numerical solution of convection-diffusion problems*. CRC Press.

Roos, H.G., M. Stynes, and L. Tobiska (2008). *Robust numerical methods for singularly perturbed differential equations: convection-diffusion-reaction and flow problems*. Vol. 24. Springer Science & Business Media.

Tröltzsch, F. (2005). *Optimale Steuerung partieller Differentialgleichungen*. Vieweg und Sohn.

Wan, H. et al. (2015). "Short-term time step convergence in a climate model". In: *Journal of Advances in Modeling Earth Systems* 7.1, pp. 215–225. DOI: 10.1002/2014MS000368.

Erklärung

Hiermit versichere ich,

I. dass diese Abhandlung - abgesehen von der Beratung durch meinen Betreuer Malte Braack - nach Inhalt und Form meine eigene ist,

II. dass diese Arbeit zum Teil an einer anderen Stelle bereits veröffentlicht, bzw. zur Veröffentlichung eingereicht wurde, im Detail:

   i. *"Equal-order finite element approximation for mantle-melt transport"* zusammen mit Malte Braack, Journal of Applied Mathematics and Computing, Vol. 65, 2020, pp. 273-293, veröffentlicht.

   ii. *"Optimized Predator-Prey and MPA Based Fishing Strategies for the Baltic Sea"* zusammen mit Malte Braack, Natural Resource Modeling, Vol. 65, 2025, veröffentlicht.

   iii. *"A local-projection stabilized sedimentation model with convection-dominant flow"* zusammen mit Malte Braack, Computational and Applied Mathematics, 2025, eingereicht.

III. dass diese Arbeit unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft entstanden ist,

IV. dass die eingereichten Veröffentlichungen einen Eigenanteil gemäß des Schreibens meines Betreuers vorweisen

V. und dass mir kein akademischer Grad entzogen wurde.

Simon Taylor