

Stokes Equations in Geophysics: Forward and Inverse Problems

DISSERTATION

zur Erlangung des Grades
Doktor der Naturwissenschaften

am Fachbereich Physik, Mathematik und Informatik
der Johannes Gutenberg-Universität
in Mainz

vorgelegt von

Lukas Holbach

geboren in Adenau, Deutschland

Mainz, den 16. Januar 2024

1. Berichtstatter: Prof. Dr. Martin Hanke-Bourgeois
2. Berichtstatter: Prof. Dr. Georg Stadler

Datum der mündlichen Prüfung: 26. April 2024

Acknowledgments

removed from online version

Abstract

Motion of rock inside the Earth is commonly modeled by incompressible Stokes equations using various rheological models. Most models for the effective viscosity are characterized by severe nonlinearities and large viscosity jumps over small length scales as well as unknown or uncertain material properties, which lead to many theoretical and practical questions about the governing systems. This thesis addresses two aspects of geodynamic Stokes equations: In the first part, well-posedness of Stokes systems with two highly nonlinear rheologies, which are important for realistic simulations of flow in the Earth's mantle, is discussed from a theoretical point of view. After formulating the instantaneous incompressible Stokes equations as employed throughout this thesis, well-posedness of the linear viscous model is reviewed while introducing concepts of convex optimization. Subsequently, we investigate Stokes systems using a non-Newtonian power-law viscosity that is also used for one of the computational inverse problems we address later. Using methods from convex analysis, well-posedness of the system is shown in a Sobolev space that depends on the degree of nonlinearity. As a next step, we investigate the corresponding Stokes system for a specific type of visco-plastic rheology. If a positive lower bound is added to the effective viscosity, well-posedness can be shown in a Sobolev space by means of a convex energy minimization problem. However, without the lower bound for the effective viscosity, a solution must be sought in the larger space of functions of bounded deformation. Existence of a corresponding energy minimizer is shown under a smallness assumption on the body forces acting on the fluid.

In the second part of this thesis, we address the issue of inferring unknown subsurface material parameters and geometric structures from surface or near-surface measurements of the solution of the Stokes equations. First, we introduce principal stress directions as a data type for geodynamical inversions. Formulating the inversion as a regularized infinite-dimensional optimization problem constrained by the Stokes equations, the adjoint equations are derived employing a formal Lagrangian approach. In simplified numerical examples, we demonstrate the potential use of this data type in real-world applications. Finally, we address the inverse problem of simultaneously reconstructing subsurface structures and material parameters from surface observations. Employing the Bayesian inversion framework enables quantification of uncertainties in the inferred parameters and shapes. We formulate an infinite-dimensional inverse problem using a level-set approach to describe boundaries between different materials, which thus allows for jumps in the material parameters. By proposing a heuristic for constructing tailored level-set priors representing seismic knowledge, we build a bridge between seismic and mechanical models. In two numerical examples, we demonstrate the potential of our approach by studying the uncertainties of and trade-offs between subsurface geometric structures, densities, and viscosities using a dimension-robust Markov chain Monte Carlo sampling method to approximate the posterior distribution.

Contents

Acknowledgments	iii
Abstract	v
1 Introduction	1
1.1 Stokes Equations in Geodynamics	1
1.2 Well-Posedness of the Linear Stokes Equations with Variable Viscosity	7
1.2.1 A Saddle-Point Problem	8
1.2.2 Convex Optimization Approach	13
1.3 Structure of the Thesis	16
2 Non-Newtonian Power-law Rheology: Well-Posedness of the Forward Problem	19
2.1 Rheological Model	20
2.2 Weak Formulation of the Forward Problem	21
2.3 Energy Minimization	25
2.4 Existence and Uniqueness of the Pressure	30
2.5 Conclusion	31
3 Ideal Visco-plastic Rheology: Existence and (Non-)Uniqueness of Solutions to the Stokes Equations	33
3.1 Rheological Model	34
3.2 The Regularized Case: $\underline{\eta} > 0$	35
3.3 Ideal Visco-plastic Rheology: $\underline{\eta} = 0$	42
3.3.1 Functions of Bounded Deformation and a Solution Space	42
3.3.2 Generalizing the Objective Function	46
3.3.3 Existence of Solutions	51
3.4 Conclusion and Outlook	54
4 Adjoint-based Inversion of Principal Stress Directions	57
4.1 Motivation	57
4.2 Governing Equations	58
4.2.1 Rheological Model	58
4.2.2 Well-Posedness of the Forward Problem	58
4.3 Inverse Problem Using Principal Stress Directions	59
4.3.1 Adjoint Equations and Adjoint-based Gradients	62
4.4 Discretization	64
4.4.1 Forward Problem	64
4.4.2 Discrete Adjoint-based Gradient	65
4.4.3 Discrete Adjoint Equations	66

Contents

4.5	Numerical Results	68
4.5.1	Model Setup	68
4.5.2	Forward Simulations	70
4.5.3	Inversion Results	70
4.6	Conclusion and Outlook	71
5	Identifying Geometric Structures and Rheological Properties Inside the Earth: A Bayesian Level Set Method	77
5.1	Motivation	77
5.2	Well-Posedness of the Forward Problem	79
5.3	Formulating the Inverse Problem	80
5.3.1	Level Set Parameterization for Uncertain Geometric Structures	80
5.4	Well-Posedness of the Bayesian Inverse Problem	84
5.4.1	Linear Stokes Equations with Surface Measurements	86
5.5	Prior Distribution for the Level Set Function	89
5.5.1	Incorporating Seismic Images as Prior Knowledge	91
5.6	Numerical Methods	92
5.6.1	Discretization and Prior Distribution	92
5.6.2	Approximating the Posterior Distribution	93
5.7	Numerical Experiments	95
5.7.1	Model 1: Sinkers	96
5.7.2	Model 2: Subduction Zone	104
5.8	Conclusion and Outlook	109
	List of Symbols	111
	List of Figures	115
	List of Tables	117
	Bibliography	119

1 CHAPTER

Introduction

Gaining knowledge of geometric structures and rheological parameters inside the Earth is a major challenge in geodynamics. Most information about the interior of the Earth has been obtained through seismic tomography, where velocity anomalies of seismic waves caused by earthquakes are inferred from their measured travel times. These velocity anomalies, which arise due to the reflection and refraction of seismic waves by different compositional structures, provide images of the Earth's interior. However, they tend to be blurry and there is no direct translation from recovered velocity anomalies to material parameters like rock densities or viscosities.

Nonetheless, such information is of great interest. The exploration of salt domes, for instance, is important for the oil industry because oil reservoirs are often found close to them. The investigation of subduction zones is an example that requires rheological information in far deeper regions. For a better understanding of earthquakes and plate motion, improving knowledge about subsurface structures and material characteristics is crucial. Therefore, combining seismic images with models for the mechanical long-term behavior of the Earth, which explicitly depend on the rheological parameters, is a promising approach that has been the subject of various research projects (see e.g. Gurnis and Hager, 1988; Stadler et al., 2010; Kaus, 2016).

This thesis addresses the problem of inferring geometric structures and material parameters deep inside the Earth from sparse surface measurements using geodynamic models. Furthermore, we discuss well-posedness of the underlying equations for different rheological models that are important for realistic simulations of flow in the Earth's mantle.

1.1 Stokes Equations in Geodynamics

In this section, we introduce the governing equations for motion inside the Earth with a focus on the rheological models employed in this thesis. We mainly follow Schubert, Turcotte, and Olson (2001, Chapter 6) and Batchelor (2000, Chapters 2 and 3), which contain detailed explanations of the models and approximations used here. There is clear evidence both in nature and from laboratory experiments that the slow deformation of solid rocks can be modeled as a viscous fluid. Since phenomena in seismology occur on the order of seconds whereas mantle convection and plate motion happen significantly slower over millions of years, we only consider the instantaneous model and neglect advective terms. The applications we are interested in either affect the Earth's mantle or a two-dimensional vertical slice of it. Hence, the model domains will be subsets of \mathbb{R}^d , $d \in \{2, 3\}$.

The governing equations are based on two well-known physical principles, the conservation of mass and the conservation of momentum. There are two types of forces that can act on a fluid and are relevant for our model: (i) volumetric forces, which we collect in a vector-valued function \mathbf{f} , and (ii) surface forces, which can be represented through $\boldsymbol{\sigma}\boldsymbol{\nu}$, where $\boldsymbol{\sigma}$ denotes the (symmetric) stress tensor and $\boldsymbol{\nu}$ the unit-length outer normal vector. Newton's second law of motion requires balance of forces for any control volume $A \subset \mathbb{R}^d$, $d \in \{2, 3\}$, with boundary ∂A :

$$\int_A \mathbf{f}(\mathbf{x}) \, d\mathbf{x} = - \int_{\partial A} \boldsymbol{\sigma}(\mathbf{x})\boldsymbol{\nu}(\mathbf{x}) \, ds = - \int_A \operatorname{div} \boldsymbol{\sigma}(\mathbf{x}) \, d\mathbf{x}, \quad (1.1)$$

where the divergence theorem is used to obtain the last equality. Since this is valid for any volume A of the fluid, we obtain the *momentum* or *force balance equation*:

$$- \operatorname{div} \boldsymbol{\sigma} = \mathbf{f}. \quad (1.2)$$

While this equation holds for arbitrary volumetric forces \mathbf{f} , gravity is the largest force in this context and we will thus often assume that it is the only body force applied to the fluid, i.e., $\mathbf{f}(\mathbf{x}) = \rho(\mathbf{x})\mathbf{g}$, where ρ is the density of the fluid and $\mathbf{g} := g\mathbf{e}_d$ with the gravitational acceleration g and the d -th unit vector \mathbf{e}_d .

The mass of a fluid in some control volume A is given by $\int_A \rho(\mathbf{x}) \, d\mathbf{x}$, whereas the mass transported out of this volume due to the velocity field \mathbf{u} is described by $\int_{\partial A} \rho(\mathbf{x})\mathbf{u}(\mathbf{x}) \cdot \boldsymbol{\nu}(\mathbf{x}) \, ds$. Here, \cdot denotes the Euclidean inner product. The change of mass with time is thus

$$\frac{d}{dt} \int_A \rho(\mathbf{x}) \, d\mathbf{x} = - \int_{\partial A} \rho(\mathbf{x})\mathbf{u}(\mathbf{x}) \cdot \boldsymbol{\nu}(\mathbf{x}) \, ds = - \int_A \operatorname{div}(\rho(\mathbf{x})\mathbf{u}(\mathbf{x})) \, d\mathbf{x}, \quad (1.3)$$

where we again used the divergence theorem. Since this is valid for any control volume A and the density in the instantaneous model does not depend on time, we obtain the *continuity equation (conservation of mass)*:

$$- \operatorname{div}(\rho\mathbf{u}) = 0. \quad (1.4)$$

In the context of this thesis, where the equations describe flow in the Earth's mantle, we assume the effects of compressibility to be negligible. Employing the Boussinesq approximation (see e.g. Schubert, Turcotte, and Olson (2001, Section 6.10) or Spiegel and Veronis (1960)), we replace (1.4) with its incompressible version

$$- \operatorname{div} \mathbf{u} = 0. \quad (1.5)$$

To close the system of conservation equations (1.2) and (1.5), we employ a constitutive relation of the form

$$\boldsymbol{\sigma}(\mathbf{u}, p) = 2\eta\dot{\boldsymbol{\epsilon}}(\mathbf{u}) - p\mathbf{I}, \quad (1.6)$$

where $\dot{\boldsymbol{\epsilon}}(\mathbf{u}) = \frac{1}{2}(\nabla\mathbf{u} + \nabla\mathbf{u}^\top)$ is the strain rate tensor, p the pressure, and \mathbf{I} the second-order identity tensor. Depending on the specific application, the effective isotropic viscosity of the fluid, $\eta = \eta(\mathbf{x}, \mathbf{u}, p)$, may depend explicitly on space as well as on the velocity and pressure fields. In the latter case, the system of partial

differential equations becomes nonlinear and both the mathematical and numerical treatment can be a delicate task.

Before we discuss some rheological models with their potential applications in geodynamics, we review the governing equations as employed throughout this thesis. Let $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, be a bounded domain with boundary $\partial\Omega$. Summarizing (1.2), (1.5) and (1.6) yields the *instantaneous incompressible Stokes equations*:

$$-\operatorname{div} \boldsymbol{\sigma} = \mathbf{f} \quad \text{in } \Omega, \quad (1.7a)$$

$$-\operatorname{div} \mathbf{u} = 0 \quad \text{in } \Omega, \quad (1.7b)$$

$$\boldsymbol{\sigma} = 2\eta\dot{\boldsymbol{\epsilon}}(\mathbf{u}) - p\mathbf{I} \quad \text{in } \Omega. \quad (1.7c)$$

Furthermore, we employ homogeneous normal flow and free slip boundary conditions everywhere on the boundary:

$$\mathbf{u} \cdot \boldsymbol{\nu} = 0 \quad \text{on } \partial\Omega, \quad (1.8a)$$

$$\mathbf{T}(\boldsymbol{\sigma}\boldsymbol{\nu}) = \mathbf{0} \quad \text{on } \partial\Omega, \quad (1.8b)$$

where $\mathbf{T} := \mathbf{I} - \boldsymbol{\nu}\boldsymbol{\nu}^\top$ denotes the tangential projection on $\partial\Omega$. The understanding behind (1.8) is that no material can leave the domain and the shear stresses vanish on the boundary. For the remainder of this thesis, we refer to (1.7) and (1.8) as the *forward problem* of the Stokes equations.

The choice of rheological model has significant influence on the flow behavior and thus on both the mathematical and computational treatment of the Stokes equations. A detailed discussion of the Earth's rheology can be found in Ranalli (1995). The most simple model assumes a constant viscosity in the entire domain, $\eta \equiv \eta_0 > 0$. However, the viscosity in the Earth's mantle is highly heterogeneous with variations of several orders of magnitude and it is therefore necessary to consider at least spatial dependence. A linear constitutive relation with spatially varying viscosity could be employed in salt domes, e.g., which occur only a few kilometers underneath the Earth's surface, because nonlinear effects are negligible in this local model (Baumann, Kaus, and Popov, 2014). While a linear, spatially varying viscosity is not the most realistic rheological model on a global scale, it drastically facilitates numerical solution of the governing equations and has therefore been applied to mantle convection as well (see e.g. Bunge, Richards, and Baumgardner, 1996; Horbach, Bunge, and Oeser, 2014; Heister et al., 2017). On the other hand, nonlinear rheologies have been shown to be crucial for Earth's plate motion (Alisic et al., 2012).

Viscous and plastic deformation account for the most relevant nonlinear effects in mantle flow. In particular, the viscous rheology can be described by a *power-law viscosity*

$$\eta_{\text{visc}}(\mathbf{x}, \mathbf{u}) = \eta_0(\mathbf{x}, n, T) \dot{\epsilon}_{\text{II}}(\mathbf{u})^{\frac{1}{n}-1}. \quad (1.9)$$

Here, $n \geq 1$ is the power-law exponent, $\eta_0(\mathbf{x}, n, T)$ the Arrhenius law dependent on the temperature, T , and $\dot{\epsilon}_{\text{II}}(\mathbf{u}) = \sqrt{\frac{1}{2}\dot{\boldsymbol{\epsilon}}(\mathbf{u}) : \dot{\boldsymbol{\epsilon}}(\mathbf{u})}$ is the square root of the second invariant of the strain rate tensor, where $:$ denotes the Frobenius inner product. For $n = 1$, (1.9) becomes linear and describes a Newtonian fluid, coinciding with the spatially dependent model above. If the power-law exponent is larger than 1, (1.9) models a nonlinear non-Newtonian strain-rate-weakening viscosity. This

rheology is realistic in the upper mantle and the lithosphere and has been employed in several studies (see e.g. Worthen et al., 2014; Räss et al., 2017). We discuss selected theoretical and practical aspects of this model in Chapters 2 and 4.

Plastic deformation can be modeled by

$$\eta_{\text{plast}}(\mathbf{x}, \mathbf{u}, p) = \frac{\tau_y(\mathbf{x}, p)}{2 \dot{\epsilon}_{\text{II}}(\mathbf{u})}, \quad (1.10)$$

where τ_y is the yield stress, at which — intuitively — the material fails and plastic deformation occurs. The most commonly used models for plastic yielding in geodynamics include the von Mises yield criterion, which assumes the yield stress to be constant everywhere (von Mises, 1913). A depth-dependent version of this considers τ_y to be spatially dependent, but independent of the pressure and velocity fields. Alternatively, one can employ a pressure-dependent Drucker-Prager yield criterion (Drucker and Prager, 1952). To model deformation in the Earth’s mantle, the viscous and plastic models (1.9) and (1.10) are often combined through

$$\eta(\mathbf{x}, \mathbf{u}, p) = \min(\eta_{\text{visc}}(\mathbf{x}, \mathbf{u}), \eta_{\text{plast}}(\mathbf{x}, \mathbf{u}, p)), \quad (1.11)$$

resulting in a *visco-plastic* rheology (see e.g. Moresi et al., 2007; Kaus, 2010; Stadler et al., 2010; Rudi, 2019). Extreme nonlinearities in these models give rise to a number of numerical issues. In Kaus (2010), e.g., different angles of shear bands, regions in which plastic yielding can occur, are encountered for different grid resolutions. This is one motivation for us to study the *well-posedness* of a visco-plastic model in Chapter 3, i.e., we discuss existence, uniqueness, and stability of solutions to the Stokes equations with a visco-plastic rheology.

While the numerical solution of the Stokes equations is not a central aspect of this work, it is a delicate task in itself. Hence, we give a short overview of the issues arising and refer the reader to research handling these topics. Many of the above mentioned rheological models for the Earth’s mantle contain severe nonlinearities. From a numerical perspective, this raises issues to construct stable iterative methods to solve the forward problem. Additionally, local nonlinearities, e.g. at plate boundaries, influence plate motion on a global scale and thus need to be resolved with the underlying mesh. Since a mesh that globally resolves such small length scales leads to prohibitively many degrees of freedom, adaptive mesh refinement and multigrid methods are often employed. Furthermore, viscosity jumps of several orders of magnitude occurring at plate boundaries result in ill-conditioned linear equation systems that need to be solved in each iteration of the nonlinear solver. Approaches to solving the Stokes equations efficiently in parallel and to dealing with the above mentioned issues can be found, e.g., in Burstedde et al. (2008), Tackley (2008), Stadler et al. (2010), Rudi et al. (2015), Kaus et al. (2016), and Rudi, Stadler, and Ghattas (2017).

The main focus of this thesis is the identification of rheological parameters and geometric structures in regions deep inside the Earth from sparse surface or near-surface measurements. In general, the process of inferring unknown physical quantities of interest from noisy measurements using a mathematical model is often referred to as an *inverse problem*. Most inverse problems of practical interest do not have a unique solution or the solution is unstable under data perturbations, i.e., they are

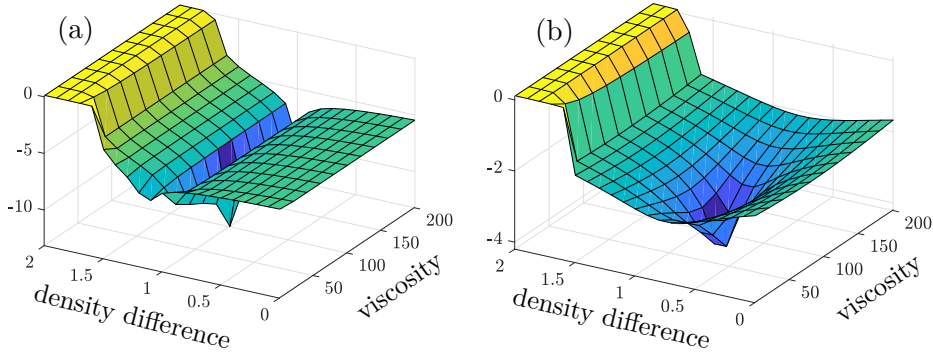


Figure 1.1 Examples of objective functions evaluated on a regular grid in the parameter space: (a) data misfit without regularization, and (b) using Tikhonov regularization. The x - and y -axes show an effective viscosity and the density difference, respectively, whereas both the colors and the z -axis indicate the values of the logarithm of the objective function. In both plots, the data misfit corresponds to measurements of principal stress directions, which are introduced in Chapter 4. All units are nondimensional.

ill-posed (see e.g. Hanke, 2017). In the context of inverse Stokes problems, a unique solution (density field) does not exist even if the velocity field is known in the entire domain since the flow is mainly driven by density differences, not their specific values. In practical examples, one can merely measure a small number of plate velocities and stresses close to the surface, making the inverse problem extremely ill-posed.

Regularization methods are one way to cope with this issue and have been studied comprehensively (see e.g. Engl, Hanke, and Neubauer, 1996). In general, due to noise and model errors, one cannot expect to be able to reproduce measurements exactly with the mathematical model and numerical approximations chosen. Therefore, one approach to solving inverse problems is to find parameters minimizing an *objective function*, which in the simplest case is a least-squares misfit between observed data and model output. The ill-posedness of the problem can lead to multiple local or even global minima of the data misfit, or a misfit that can be insensitive to changes in some parameters (see Figure 1.1a). The latter results in a flat objective function, which can be difficult to minimize numerically. The idea of Tikhonov regularization is to include an additional penalty term in the objective function to promote certain parameter values or structures, incorporating knowledge one already has about the inference parameters. In the Stokes context, one might know a plausible range of density values or the approximate location of an expected jump in viscosities. Ideally, the regularized objective function will have only one minimum and will be sensitive to changes in every inference parameter (see Figure 1.1b). However, the penalty term adds a bias to the solution of the inverse problem and thus needs to be chosen carefully to balance regularization and bias.

Another approach to dealing with ill-posedness is to employ the Bayesian inversion framework. Here, prior beliefs are stated in terms of probabilities and serve as regularization. The solution to a Bayesian inverse problem, the so-called *posterior distribution*, is a conditional probability distribution for the inference parameters

given the measured data, which additionally enables quantification of uncertainties. While this approach is, under certain assumptions, reasonably well understood and computationally feasible for linear forward operators and Gaussian random fields (Stuart, 2010; Bui-Thanh et al., 2013), the problem becomes significantly more challenging when the forward operator governing the inverse problem is nonlinear, in which case one usually has to rely on sampling techniques to approximate the distribution. Note that choosing a prior distribution introduces a bias to the solution of the inverse problem, similar to choosing a regularization term in the deterministic inversion outlined above. In fact, under the assumption of additive Gaussian noise and a Gaussian prior, maximizers of the posterior probability distribution coincide with minimizers of a prior-regularized objective function and the two approaches can thus be seen as two sides of the same coin.

Regardless of the specific inversion approach chosen, parameters of interest in geodynamic inverse problems include the density field ρ as well as rheological parameters depending on the effective viscosity model used, e.g., the reference viscosity η_0 and/or the power-law exponent n in (1.9), or the yield stress τ_y in (1.10). These quantities depend on the material and can be seen as piecewise constants in the context of mantle flow since they exhibit large variation over small length scales. While seismic tomography provides images of subsurface structures, they typically suffer from blurry edges and it is not clear how exactly recovered seismic velocity anomalies relate to the rheological parameters that we aim to identify. Inverting for the material parameters as spatially dependent functions without additional information about the geometric structures is almost hopeless due to the severe ill-posedness of the inverse problem (cf. Worthen et al., 2014).

Several works have assumed interfaces between different materials to be known and inferred constant material parameters within these structures from surface velocities (see e.g. Baumann and Kaus, 2015; Ratnaswamy, Stadler, and Gurnis, 2015). In Chapter 4, we extend these approaches by including principal stress directions as measurements and derive adjoint-based expressions for the gradient of the objective function that can be computed efficiently. Although the mathematical framework presented there allows for inversion of parameter fields, we restrict the numerical examples to inversion of piecewise constants in predefined regions due to the above reasons. These models are thus limited in their ability to decipher the trade-offs between rheology and density and their ability to assess the role of uncertain spatial structures.

Alternatively, one can rely on parameterizations of geometric structures and additionally invert for them. These include explicit parameterizations of geometric interfaces (see e.g. Carpio, Iakunin, and Stadler, 2020), for which a moderate number of parameters may be sufficient. However, this approach limits geometric flexibility. Employing an implicit parameterization such as describing geometric structures as level sets of an auxiliary function has become a particularly popular tool for geometric inverse problems due to the shape flexibility it provides (Santosa, 1996). In Chapter 5, we build on Iglesias, Lu, and Stuart (2016) to formulate a level-set based Bayesian inverse problem to simultaneously infer subsurface geometric structures and rheological parameters from surface measurements of plate velocities and normal tractions. This allows us to additionally quantify uncertainties in the inferred

parameters and shapes. Information about subsurface structures from seismic images is included by constructing tailored level-set priors.

However, before any inverse problem is addressed, well-posedness of the governing forward model is a necessary prerequisite.

1.2 Well-Posedness of the Linear Stokes Equations with Variable Viscosity

In the following, we review well-posedness of the Stokes equations (1.7) and (1.8) with linear constitutive relation and spatially varying viscosity, i.e., we show existence of a unique solution along with stability estimates and continuous dependence of the solution on the data. While the results themselves are well-known and follow with standard methods, this exemplary discussion highlights the tools typically needed and the strategies pursued for proving well-posedness with more complicated rheologies like the ones we approach in Chapters 2 and 3. Basic knowledge of functional analysis and partial differential equations is expected. An overview of common notations used throughout this thesis can be found in the List of Symbols in the appendix.

Unless stated otherwise, we assume in the entire thesis that the bounded domain $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, is a polytope, i.e., $\Omega \subset \mathbb{R}^2$ is a polygon and $\Omega \subset \mathbb{R}^3$ is a polyhedron. Consider a linear constitutive relation, i.e., $\eta \in L_+^\infty(\Omega)$ is independent of \mathbf{u} and p and essentially bounded from above and below by positive constants. We start by defining appropriate solution spaces for the velocity and pressure fields. Let

$$V := \left\{ \mathbf{u} \in H^1(\Omega)^d \mid \mathbf{u} \cdot \boldsymbol{\nu} = 0 \text{ on } \partial\Omega \right\}, \quad (1.12)$$

$$Q := L_0^2(\Omega) := \left\{ p \in L^2(\Omega) \mid \int_{\Omega} p \, d\mathbf{x} = 0 \right\}, \quad (1.13)$$

where the boundary condition on \mathbf{u} is to be understood in the sense of traces. Since the trace operator from $H^1(\Omega)^d$ to $L^2(\partial\Omega)^d$ is linear and continuous (see e.g. Nečas, 2012, Chapter 3, Theorem 1.1), this is also true for the trace of the normal component. Therefore, $(V, \langle \cdot, \cdot \rangle_{H^1})$ is a Hilbert space as a closed subspace of the Hilbert space $H^1(\Omega)^d$. Similarly, the function $p \mapsto \int_{\Omega} p \, d\mathbf{x}$, $p \in L^2(\Omega)$, is continuous and linear as Ω is bounded and thus its nullspace is a closed subspace of $L^2(\Omega)$, implying that $(Q, \langle \cdot, \cdot \rangle_{L^2})$ is a Hilbert space as well.

This allows us to derive the weak form of the Stokes equations: Multiplying equation (1.7a) with an arbitrary test function $\mathbf{v} \in V$, integrating over Ω , and using Green's formula, we obtain

$$\begin{aligned} \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} &= - \int_{\Omega} \mathbf{v} \cdot \operatorname{div} \boldsymbol{\sigma} \, d\mathbf{x} \\ &= \int_{\Omega} \nabla \mathbf{v} : \boldsymbol{\sigma} \, d\mathbf{x} - \int_{\partial\Omega} \underbrace{\mathbf{v} \cdot \boldsymbol{\sigma} \boldsymbol{\nu}}_{=0} \, ds \\ &= \int_{\Omega} \nabla \mathbf{v} : 2\eta \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) \, d\mathbf{x} - \int_{\Omega} \nabla \mathbf{v} : p \mathbf{I} \, d\mathbf{x} \\ &= \int_{\Omega} 2\eta \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) : \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \, d\mathbf{x} - \int_{\Omega} p \operatorname{div} \mathbf{v} \, d\mathbf{x}, \end{aligned} \quad (1.14)$$

where we used the (pointwise) identity

$$\begin{aligned}
 \dot{\boldsymbol{\epsilon}}(\mathbf{u}) : \dot{\boldsymbol{\epsilon}}(\mathbf{v}) &= \dot{\boldsymbol{\epsilon}}(\mathbf{v}) : \dot{\boldsymbol{\epsilon}}(\mathbf{u}) \\
 &= \frac{1}{2} \nabla \mathbf{v} : \dot{\boldsymbol{\epsilon}}(\mathbf{u}) + \frac{1}{2} \nabla \mathbf{v}^\top : \dot{\boldsymbol{\epsilon}}(\mathbf{u}) \\
 &= \frac{1}{2} \nabla \mathbf{v} : \dot{\boldsymbol{\epsilon}}(\mathbf{u}) + \frac{1}{2} (\nabla \mathbf{v}^\top)^\top : (\dot{\boldsymbol{\epsilon}}(\mathbf{u}))^\top \\
 &= \nabla \mathbf{v} : \dot{\boldsymbol{\epsilon}}(\mathbf{u})
 \end{aligned} \tag{1.15}$$

in the last line. Furthermore, the boundary integral vanishes due to the boundary conditions: According to (1.8b),

$$\boldsymbol{\sigma} \boldsymbol{\nu} = \boldsymbol{\nu} \boldsymbol{\nu}^\top \boldsymbol{\sigma} \boldsymbol{\nu}$$

and thus (1.12) implies

$$\mathbf{v} \cdot \boldsymbol{\sigma} \boldsymbol{\nu} = (\mathbf{v} \cdot \boldsymbol{\nu}) \boldsymbol{\nu}^\top \boldsymbol{\sigma} \boldsymbol{\nu} = 0$$

on $\partial\Omega$. Finally, multiplying (1.7b) with arbitrary $q \in Q$, integrating over Ω and adding the resulting equation to (1.14) yields the *weak form of the Stokes equations* (1.7): Find $(\mathbf{u}, p) \in V \times Q$ such that

$$\int_{\Omega} 2\eta \dot{\boldsymbol{\epsilon}}(\mathbf{u}) : \dot{\boldsymbol{\epsilon}}(\mathbf{v}) \, d\mathbf{x} - \int_{\Omega} p \operatorname{div} \mathbf{v} \, d\mathbf{x} - \int_{\Omega} q \operatorname{div} \mathbf{u} \, d\mathbf{x} = \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} \tag{1.16}$$

for all $(\mathbf{v}, q) \in V \times Q$. Note that a natural space for the right-hand side force \mathbf{f} is V' , the dual space of V . However, in certain situations we restrict the choices to subspaces $L^r(\Omega)^d \subset V'$, $r \geq 2$.

1.2.1 A Saddle-Point Problem

We follow Girault and Raviart (1986, Chapter I, § 4) to prove well-posedness of (1.16). Let

$$\begin{aligned}
 a: V \times V &\rightarrow \mathbb{R}, \\
 (\mathbf{u}, \mathbf{v}) &\mapsto \int_{\Omega} 2\eta \dot{\boldsymbol{\epsilon}}(\mathbf{u}) : \dot{\boldsymbol{\epsilon}}(\mathbf{v}) \, d\mathbf{x},
 \end{aligned} \tag{1.17}$$

$$\begin{aligned}
 b: V \times Q &\rightarrow \mathbb{R}, \\
 (\mathbf{u}, p) &\mapsto - \int_{\Omega} p \operatorname{div} \mathbf{u} \, d\mathbf{x}.
 \end{aligned} \tag{1.18}$$

Using this notation, the weak form (1.16) consists of finding $(\mathbf{u}, p) \in V \times Q$ such that

$$a(\mathbf{u}, \mathbf{v}) + b(\mathbf{v}, p) = \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} \quad \text{for all } \mathbf{v} \in V, \tag{1.19a}$$

$$b(\mathbf{u}, q) = 0 \quad \text{for all } q \in Q. \tag{1.19b}$$

This *saddle-point problem* has a unique solution if the following criteria are met:

1.2 Well-Posedness of the Linear Stokes Equations with Variable Viscosity

1. a and b are *continuous* bilinear forms, i.e., there are constants $c_1, c_2 > 0$ such that

$$a(\mathbf{u}, \mathbf{v}) \leq c_1 \|\mathbf{u}\|_{H^1} \|\mathbf{v}\|_{H^1} \quad \text{for all } \mathbf{u}, \mathbf{v} \in V, \quad (1.20a)$$

$$b(\mathbf{u}, q) \leq c_2 \|\mathbf{u}\|_{H^1} \|q\|_{L^2} \quad \text{for all } \mathbf{u} \in V, q \in Q. \quad (1.20b)$$

2. The bilinear form a is *coercive*, i.e., there is a constant $c_3 > 0$ such that

$$a(\mathbf{u}, \mathbf{u}) \geq c_3 \|\mathbf{u}\|_{H^1}^2 \quad \text{for all } \mathbf{u} \in V. \quad (1.21)$$

3. The bilinear form b satisfies the *inf-sup condition*, i.e., there is a constant $c_4 > 0$ such that

$$\inf_{0 \neq q \in Q} \sup_{0 \neq \mathbf{v} \in V} \frac{b(\mathbf{v}, q)}{\|\mathbf{v}\|_{H^1} \|q\|_{L^2}} \geq c_4. \quad (1.22)$$

Due to the linearity of the integral and the occurring differential operators, a and b define bilinear forms. Note that this approach is not suitable if the viscosity η depends on the velocity field \mathbf{u} as a is nonlinear then. Since

$$|\dot{\boldsymbol{\varepsilon}}(\mathbf{u})| \leq \frac{1}{2} \left(|\nabla \mathbf{u}| + |\nabla \mathbf{u}^\top| \right) = |\nabla \mathbf{u}|, \quad (1.23)$$

where $|\cdot| = \sqrt{\cdot \cdot}$ is the Frobenius norm, and

$$(\operatorname{div} \mathbf{u})^2 = \left(\sum_{i=1}^d \frac{\partial u_i}{\partial x_i} \right)^2 \leq d \sum_{i=1}^d \left(\frac{\partial u_i}{\partial x_i} \right)^2 \leq d \sum_{i,j=1}^d \left(\frac{\partial u_i}{\partial x_j} \right)^2 = d |\nabla \mathbf{u}|^2, \quad (1.24)$$

straightforward estimates yield¹

$$a(\mathbf{u}, \mathbf{v}) \leq 2 \|\eta\|_\infty \|\mathbf{u}\|_{H^1} \|\mathbf{v}\|_{H^1} \quad \text{for all } \mathbf{u}, \mathbf{v} \in V, \quad (1.25)$$

$$b(\mathbf{u}, q) \leq \sqrt{d} \|\mathbf{u}\|_{H^1} \|q\|_{L^2} \quad \text{for all } \mathbf{u} \in V, q \in Q, \quad (1.26)$$

and thus a and b are continuous.

Proving coercivity of a relies on *Korn's inequality*, which is a crucial tool for the solution theory of Stokes equations. Under appropriate conditions, Korn's inequality states that there is a constant $C_K > 0$ such that

$$\|\dot{\boldsymbol{\varepsilon}}(\mathbf{u})\|_{L^2} \geq C_K \|\mathbf{u}\|_{H^1} \quad \text{for all } \mathbf{u} \in V \quad (1.27)$$

and can thus be seen as an analog of Poincaré's inequality suitable to the deformation tensor $\dot{\boldsymbol{\varepsilon}}(\mathbf{u})$. Intuitively, the inequality is valid if we can guarantee that $\mathbf{0}$ is the only element of the deformation tensor's nullspace that lies in V . The nullspace of $\dot{\boldsymbol{\varepsilon}}$ consists of so-called *rigid displacements*, i.e., functions $\mathbf{r}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$ with a vector $\mathbf{b} \in \mathbb{R}^d$ and a skew-symmetric matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$, i.e., $\mathbf{A} = -\mathbf{A}^\top$. In the present case, the boundary condition $\mathbf{r} \cdot \boldsymbol{\nu} = 0$ everywhere on the polytopic boundary $\partial\Omega$ implies that $\mathbf{0}$ is indeed the only rigid displacement in V and thus Korn's inequality

¹Inequalities (1.23) and (1.24) imply $\|\dot{\boldsymbol{\varepsilon}}(\mathbf{u})\|_{L^r} \leq \|\nabla \mathbf{u}\|_{L^r}$ and $\|\operatorname{div} \mathbf{u}\|_{L^r} \leq \sqrt{d} \|\nabla \mathbf{u}\|_{L^r}$ for arbitrary $1 \leq r \leq \infty$ if $\mathbf{u} \in W^{1,r}(\Omega)^d$.

holds on V . A more general and mathematically rigorous discussion can be found in Proposition 2.4, which includes V as a special case.

Coercivity of the bilinear form a now follows immediately from Korn's inequality: For arbitrary $\mathbf{u} \in V$, we obtain

$$a(\mathbf{u}, \mathbf{u}) = \int_{\Omega} 2\eta \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) : \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) \, d\mathbf{x} \geq 2\eta_{\min} \|\dot{\boldsymbol{\varepsilon}}(\mathbf{u})\|_{L^2}^2 \geq 2\eta_{\min} C_K^2 \|\mathbf{u}\|_{H^1}^2, \quad (1.28)$$

where $\eta_{\min} > 0$ denotes the essential infimum of $\eta \in L_+^\infty(\Omega)$.

It remains to establish the inf-sup condition (1.22). Let $q \in Q \setminus \{0\}$ be arbitrary. Due to Girault and Raviart (1986, Chapter I, Lemma 3.2), there exist $\mathbf{v} = \mathbf{v}(q) \in V$ and a constant $C > 0$ such that

$$-q = \operatorname{div} \mathbf{v}, \quad \|\mathbf{v}\|_{H^1} \leq C \|q\|_{L^2}. \quad (1.29)$$

We therefore obtain

$$\frac{b(\mathbf{v}, q)}{\|\mathbf{v}\|_{H^1}} = \frac{-\int_{\Omega} q \operatorname{div} \mathbf{v} \, d\mathbf{x}}{\|\mathbf{v}\|_{H^1}} = \frac{\|q\|_{L^2}^2}{\|\mathbf{v}\|_{H^1}} \geq \frac{\|q\|_{L^2}}{C}, \quad (1.30)$$

which implies

$$\sup_{0 \neq \mathbf{v} \in V} \frac{b(\mathbf{v}, q)}{\|\mathbf{v}\|_{H^1} \|q\|_{L^2}} \geq \frac{1}{C}. \quad (1.31)$$

Since q was arbitrary,

$$\inf_{0 \neq q \in Q} \sup_{0 \neq \mathbf{v} \in V} \frac{b(\mathbf{v}, q)}{\|\mathbf{v}\|_{H^1} \|q\|_{L^2}} \geq \frac{1}{C} \quad (1.32)$$

and we have verified (1.22). From this, the existence of a unique solution $(\mathbf{u}, p) \in V \times Q$ to the Stokes equations (1.16) follows.

The next step is to show stability of the solution, which is done by proving that the norms of \mathbf{u} and p are bounded by the right-hand side forcing term. Due to the coercivity estimate (1.21) and the weak form (1.19a) and (1.19b), we obtain

$$c_3 \|\mathbf{u}\|_{H^1}^2 \leq a(\mathbf{u}, \mathbf{u}) = \int_{\Omega} \mathbf{f} \cdot \mathbf{u} \, d\mathbf{x} \leq \|\mathbf{f}\|_{V'} \|\mathbf{u}\|_{H^1}, \quad (1.33)$$

or equivalently,

$$\|\mathbf{u}\|_{H^1} \leq c_3^{-1} \|\mathbf{f}\|_{V'}. \quad (1.34)$$

An analogous estimate for the pressure follows from the inf-sup condition (1.22), which implies, using the weak form (1.19a) and estimates (1.20a) and (1.34),

$$\begin{aligned} \|p\|_{L^2} &\leq c_4^{-1} \sup_{0 \neq \mathbf{v} \in V} \frac{b(\mathbf{v}, p)}{\|\mathbf{v}\|_{H^1}} \\ &= c_4^{-1} \sup_{0 \neq \mathbf{v} \in V} \frac{\int_{\Omega} \mathbf{f} \cdot \mathbf{v} + a(\mathbf{u}, -\mathbf{v})}{\|\mathbf{v}\|_{H^1}} \\ &\leq c_4^{-1} \sup_{0 \neq \mathbf{v} \in V} \frac{\|\mathbf{f}\|_{V'} \|\mathbf{v}\|_{H^1} + c_1 \|\mathbf{u}\|_{H^1} \|\mathbf{v}\|_{H^1}}{\|\mathbf{v}\|_{H^1}} \\ &= c_4^{-1} (\|\mathbf{f}\|_{V'} + c_1 \|\mathbf{u}\|_{H^1}) \\ &\leq c_4^{-1} (1 + c_1 c_3^{-1}) \|\mathbf{f}\|_{V'}. \end{aligned} \quad (1.35)$$

1.2 Well-Posedness of the Linear Stokes Equations with Variable Viscosity

With the help of these a-priori estimates (1.34) and (1.35), we are able to prove continuous dependence of the solution (\mathbf{u}, p) on the right-hand side \mathbf{f} and the viscosity η . Denote $(\hat{\mathbf{u}}, \hat{p})$ as the unique solution of (1.19a) and (1.19b) corresponding to the perturbed data $\hat{\mathbf{f}} \in V'$, $\hat{\eta} \in L_+^\infty(\Omega)$. By adding zero to equation (1.19a) in a suitable way and rearranging terms, we obtain

$$\int_{\Omega} 2\eta \dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}}) : \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \, d\mathbf{x} - \int_{\Omega} \hat{p} \operatorname{div} \mathbf{v} \, d\mathbf{x} = \int_{\Omega} \hat{\mathbf{f}} \cdot \mathbf{v} \, d\mathbf{x} + \int_{\Omega} 2(\eta - \hat{\eta}) \dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}}) : \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \, d\mathbf{x} \quad (1.36)$$

for all $\mathbf{v} \in V$. Subtracting the weak form (1.19a) corresponding to \mathbf{f} and η from this equation yields

$$\begin{aligned} \int_{\Omega} 2\eta \dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}} - \mathbf{u}) : \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \, d\mathbf{x} - \int_{\Omega} (\hat{p} - p) \operatorname{div} \mathbf{v} \, d\mathbf{x} &= \int_{\Omega} (\hat{\mathbf{f}} - \mathbf{f}) \cdot \mathbf{v} \, d\mathbf{x} \\ &+ \int_{\Omega} 2(\eta - \hat{\eta}) \dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}}) : \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \, d\mathbf{x} \end{aligned} \quad (1.37)$$

for all $\mathbf{v} \in V$. Choosing $\mathbf{v} := \hat{\mathbf{u}} - \mathbf{u}$, the pressure integral vanishes since $(\hat{p} - p) \in Q$ and (1.19b) holds for both \mathbf{u} and $\hat{\mathbf{u}}$. Using the coercivity (1.21), the left-hand side of (1.37) is thus bounded from below by

$$\int_{\Omega} 2\eta \dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}} - \mathbf{u}) : \dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}} - \mathbf{u}) \, d\mathbf{x} = a(\hat{\mathbf{u}} - \mathbf{u}, \hat{\mathbf{u}} - \mathbf{u}) \geq c_3 \|\hat{\mathbf{u}} - \mathbf{u}\|_{H^1}^2. \quad (1.38)$$

On the other hand, the right-hand side of (1.37) is bounded from above by

$$\begin{aligned} &\int_{\Omega} (\hat{\mathbf{f}} - \mathbf{f}) \cdot (\hat{\mathbf{u}} - \mathbf{u}) \, d\mathbf{x} + \int_{\Omega} 2(\eta - \hat{\eta}) \dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}}) : \dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}} - \mathbf{u}) \, d\mathbf{x} \\ &\leq \|\hat{\mathbf{f}} - \mathbf{f}\|_{V'} \|\hat{\mathbf{u}} - \mathbf{u}\|_{H^1} + 2 \|\hat{\eta} - \eta\|_{\infty} \|\dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}})\|_{L^2} \|\dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}} - \mathbf{u})\|_{L^2} \\ &\leq \|\hat{\mathbf{f}} - \mathbf{f}\|_{V'} \|\hat{\mathbf{u}} - \mathbf{u}\|_{H^1} + 2c_3^{-1} \|\hat{\eta} - \eta\|_{\infty} \|\hat{\mathbf{f}}\|_{V'} \|\hat{\mathbf{u}} - \mathbf{u}\|_{H^1}, \end{aligned} \quad (1.39)$$

where we used the a-priori estimate (1.34) in the last line. Combining (1.38) and (1.39) yields

$$\|\hat{\mathbf{u}} - \mathbf{u}\|_{H^1} \leq C \left(\|\hat{\mathbf{f}} - \mathbf{f}\|_{V'} + \|\hat{\mathbf{f}}\|_{V'} \|\hat{\eta} - \eta\|_{\infty} \right) \quad (1.40)$$

for some constant $C > 0$ and thus the velocity field \mathbf{u} depends continuously on the right-hand side \mathbf{f} and viscosity η . Furthermore, an analogous statement for the pressure follows with estimates (1.34) and (1.40) using the inf-sup condition (1.22)

and equation (1.37):

$$\begin{aligned}
\|\hat{p} - p\|_{L^2} &\leq c_4^{-1} \sup_{0 \neq \mathbf{v} \in V} \frac{b(\mathbf{v}, \hat{p} - p)}{\|\mathbf{v}\|_{H^1}} \\
&= c_4^{-1} \sup_{0 \neq \mathbf{v} \in V} \left[\left(- \int_{\Omega} 2\eta \dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}} - \mathbf{u}) : \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \, d\mathbf{x} + \int_{\Omega} (\hat{\mathbf{f}} - \mathbf{f}) \cdot \mathbf{v} \, d\mathbf{x} \right. \right. \\
&\quad \left. \left. + \int_{\Omega} 2(\eta - \hat{\eta}) \dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}}) : \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \, d\mathbf{x} \right) / \|\mathbf{v}\|_{H^1} \right] \\
&\leq c_4^{-1} \sup_{0 \neq \mathbf{v} \in V} \left[\left(2 \|\eta\|_{\infty} \|\hat{\mathbf{u}} - \mathbf{u}\|_{H^1} \|\mathbf{v}\|_{H^1} + \|\hat{\mathbf{f}} - \mathbf{f}\|_{V'} \|\mathbf{v}\|_{H^1} \right. \right. \\
&\quad \left. \left. + 2 \|\hat{\eta} - \eta\|_{\infty} \|\hat{\mathbf{u}}\|_{H^1} \|\mathbf{v}\|_{H^1} \right) / \|\mathbf{v}\|_{H^1} \right] \tag{1.41} \\
&\leq c_4^{-1} \left(2 \|\eta\|_{\infty} C \left(\|\hat{\mathbf{f}} - \mathbf{f}\|_{V'} + \|\hat{\mathbf{f}}\|_{V'} \|\hat{\eta} - \eta\|_{\infty} \right) + \|\hat{\mathbf{f}} - \mathbf{f}\|_{V'} \right. \\
&\quad \left. + 2c_3^{-1} \|\hat{\eta} - \eta\|_{\infty} \|\hat{\mathbf{f}}\|_{V'} \right) \\
&\leq \tilde{C} \left(\|\hat{\mathbf{f}} - \mathbf{f}\|_{V'} + \|\hat{\mathbf{f}}\|_{V'} \|\hat{\eta} - \eta\|_{\infty} \right)
\end{aligned}$$

for some constant $\tilde{C} > 0$.

In Chapter 5, we assume more regularity on the right-hand side, namely $\mathbf{f} \in L^r(\Omega)^d \subseteq L^2(\Omega)^d$ for some $r \geq 2$. In this case, we also obtain continuous dependence of the stress tensor $\boldsymbol{\sigma} \in H(\text{div}; \Omega)$ on the data. First note that the above estimates hold as well if the V' norm is replaced by the L^2 norm. Furthermore, the weak form (1.14) states that

$$\int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} = \int_{\Omega} \nabla \mathbf{v} : \boldsymbol{\sigma} \, d\mathbf{x} \tag{1.42}$$

for all $\mathbf{v} \in C_c^{\infty}(\Omega; \mathbb{R}^d) \subset V$, implying that the distributional divergence of $\boldsymbol{\sigma}$ is $-\mathbf{f} \in L^2(\Omega)^d$ and thus $\boldsymbol{\sigma} \in H(\text{div}; \Omega)$. As before, denoting $\hat{\boldsymbol{\sigma}} = 2\hat{\eta}\dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}}) - \hat{p}\mathbf{I}$ as the stress tensor corresponding to data $\hat{\mathbf{f}} \in L^2(\Omega)^d$ and $\hat{\eta} \in L_+^{\infty}(\Omega)$,

$$\begin{aligned}
\|\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\|_{L^2} &= \|2(\hat{\eta}\dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}}) - \eta\dot{\boldsymbol{\varepsilon}}(\mathbf{u})) - (\hat{p} - p)\mathbf{I}\|_{L^2} \\
&\leq 2 \|\hat{\eta} - \eta\|_{\infty} \|\dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}})\|_{L^2} + 2 \|\eta\|_{\infty} \|\dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}} - \mathbf{u})\|_{L^2} + \sqrt{d} \|\hat{p} - p\|_{L^2}, \tag{1.43}
\end{aligned}$$

and using estimates (1.34), (1.40) and (1.41) yields

$$\|\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\|_{L^2} \leq \hat{C} \left(\|\hat{\mathbf{f}} - \mathbf{f}\|_{L^2} + \|\hat{\mathbf{f}}\|_{L^2} \|\hat{\eta} - \eta\|_{\infty} \right) \tag{1.44}$$

for some constant $\hat{C} > 0$. We therefore obtain

$$\begin{aligned}
\|\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\|_{H(\text{div})} &\leq \|\text{div}(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})\|_{L^2} + \|\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\|_{L^2} \\
&\leq \|\hat{\mathbf{f}} - \mathbf{f}\|_{L^2} + \hat{C} \left(\|\hat{\mathbf{f}} - \mathbf{f}\|_{L^2} + \|\hat{\mathbf{f}}\|_{L^2} \|\hat{\eta} - \eta\|_{\infty} \right) \\
&\leq (\hat{C} + 1) \left(\|\hat{\mathbf{f}} - \mathbf{f}\|_{L^2} + \|\hat{\mathbf{f}}\|_{L^2} \|\hat{\eta} - \eta\|_{\infty} \right), \tag{1.45}
\end{aligned}$$

1.2 Well-Posedness of the Linear Stokes Equations with Variable Viscosity

which is the desired continuous dependence of the stress tensor on the data. Note that combining the a-priori estimates for velocity and pressure, (1.34) and (1.35), immediately yields an a-priori estimate for the stress tensor:

$$\begin{aligned}
\|\boldsymbol{\sigma}\|_{H(\operatorname{div})} &\leq \|\operatorname{div} \boldsymbol{\sigma}\|_{L^2} + \|\boldsymbol{\sigma}\|_{L^2} \\
&\leq \|\mathbf{f}\|_{L^2} + 2\|\eta\|_{\infty} \|\dot{\boldsymbol{\varepsilon}}(\mathbf{u})\|_{L^2} + \sqrt{d}\|p\|_{L^2} \\
&\leq \|\mathbf{f}\|_{L^2} + 2\|\eta\|_{\infty} c_3^{-1} \|\mathbf{f}\|_{L^2} + c_4^{-1} \sqrt{d} (1 + c_1 c_3^{-1}) \|\mathbf{f}\|_{L^2} \\
&= \bar{C} \|\mathbf{f}\|_{L^2}
\end{aligned} \tag{1.46}$$

for some constant $\bar{C} > 0$.

Remark 1.1. For the well-posedness of the Bayesian inverse problem treated in Chapter 5, we need continuous dependence of the solution on the viscosity η in a Lebesgue space $L^s(\Omega)$ for some finite $1 \leq s < \infty$. By applying Hölder's inequality,

$$\begin{aligned}
\int_{\Omega} 2(\eta - \hat{\eta}) \dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}}) : \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \, d\mathbf{x} &\leq 2\|(\hat{\eta} - \eta) \dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}})\|_{L^2} \|\dot{\boldsymbol{\varepsilon}}(\mathbf{v})\|_{L^2} \\
&\leq 2\|\hat{\eta} - \eta\|_{L^{r'}} \|\dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}})\|_{L^r} \|\mathbf{v}\|_{H^1}
\end{aligned} \tag{1.47}$$

for arbitrary $\mathbf{v} \in V$, where $2 \leq r$, $r' \leq \infty$, $\frac{1}{r} + \frac{1}{r'} = \frac{1}{2}$. Hence, with the same arguments as before, we obtain an estimate of the form

$$\|\hat{\mathbf{u}} - \mathbf{u}\|_{H^1} \leq C \left(\|\hat{\mathbf{f}} - \mathbf{f}\|_{L^2} + \|\dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}})\|_{L^r} \|\hat{\eta} - \eta\|_{L^{r'}} \right), \tag{1.48}$$

analogous to (1.39) and (1.40) if $r = 2$. However, for $r > 2$ the exponent $s := r' = 2r/(r-2)$ is finite and yields the desired estimate in $L^s(\Omega)$ if we can control the norm $\|\dot{\boldsymbol{\varepsilon}}(\hat{\mathbf{u}})\|_{L^r}$. The latter follows from regularity theory about the Stokes equations, which yields stability estimates of the form (1.34) in $W^{1,r}(\Omega)^d$ if $\mathbf{f} \in L^r(\Omega)^d$ (see e.g. Amrouche and Rejaiba (2014, Theorem 3.7) or Bonito, DeVore, and Nочetto (2013, Sections 1 and 2) for analogous estimates and the corresponding ideas of the proof). Due to the Sobolev embeddings for $r > 2$ in the bounded domain Ω , combining these results with the above estimates yields

$$\|\hat{\mathbf{u}} - \mathbf{u}\|_{H^1} \leq C \left(\|\hat{\mathbf{f}} - \mathbf{f}\|_{L^2} + \|\hat{\mathbf{f}}\|_{L^r} \|\hat{\eta} - \eta\|_{L^s} \right) \tag{1.49}$$

for some constant $C > 0$ and thus the continuous dependence of the velocity field on the viscosity $\eta \in L_+^{\infty}(\Omega) \subset L^s(\Omega)$ under the additional assumption $\mathbf{f} \in L^r(\Omega)^d$. Following the same arguments, we also obtain continuous dependence of the pressure and stress tensor on $\mathbf{f} \in L^r(\Omega)^d \subset L^2(\Omega)^d$ and $\eta \in L_+^{\infty}(\Omega) \subset L^s(\Omega)$.

1.2.2 Convex Optimization Approach

An alternative way of proving well-posedness is to formulate the Stokes equations as a constrained optimization problem. In the first step, the pressure is eliminated from the equations by only considering divergence-free velocity fields. The corresponding weak form is then reformulated as a minimization problem and well-posedness can be shown using methods of convex analysis. Existence and uniqueness of the pressure

field can be argued by establishing an inf-sup condition, analogous to Section 1.2.1. As this approach will be employed for the nonlinear constitutive relations discussed in Chapters 2 and 3, we explain the key ideas for the linear case $\eta \in L_+^\infty(\Omega)$ here. Furthermore, we focus on the methods for proving existence of a unique solution and the stability estimates since continuous dependence on the data requires different tools for nonlinear models (see e.g. Zeidler, 1990).

Let

$$V_0 := \left\{ \mathbf{u} \in H^1(\Omega)^d \mid \operatorname{div} \mathbf{u} = 0 \text{ on } \Omega, \mathbf{u} \cdot \boldsymbol{\nu} = 0 \text{ on } \partial\Omega \right\} \subset V, \quad (1.50)$$

equipped with the H^1 topology. Since $\operatorname{div}: H^1(\Omega)^d \rightarrow L^2(\Omega)$ is continuous, V_0 is a Hilbert space as a closed subspace of the Hilbert space V . Using V_0 as the ansatz and test space in (1.19), the weak formulation reduces to finding $\mathbf{u} \in V_0$ such that

$$a(\mathbf{u}, \mathbf{v}) = \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} \quad (1.51)$$

for all $\mathbf{v} \in V_0$, since all velocity fields in V_0 are divergence-free. Introducing the *energy functional*

$$\Phi: V_0 \rightarrow \mathbb{R}, \quad \mathbf{u} \mapsto \frac{1}{2}a(\mathbf{u}, \mathbf{u}) - \int_{\Omega} \mathbf{f} \cdot \mathbf{u} \, d\mathbf{x}, \quad (1.52)$$

we obtain

$$\begin{aligned} \Phi(\mathbf{u} + \mathbf{v}) &= \frac{1}{2}a(\mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v}) - \int_{\Omega} \mathbf{f} \cdot (\mathbf{u} + \mathbf{v}) \, d\mathbf{x} \\ &= \Phi(\mathbf{u}) + a(\mathbf{u}, \mathbf{v}) - \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} + \frac{1}{2}a(\mathbf{v}, \mathbf{v}) \end{aligned} \quad (1.53)$$

for arbitrary $\mathbf{u}, \mathbf{v} \in V_0$, where

$$0 \leq \frac{a(\mathbf{v}, \mathbf{v})}{2 \|\mathbf{v}\|_{H^1}} \leq \frac{c_1 \|\mathbf{v}\|_{H^1}^2}{2 \|\mathbf{v}\|_{H^1}} = \frac{c_1}{2} \|\mathbf{v}\|_{H^1} \rightarrow 0 \quad \text{for } \|\mathbf{v}\|_{H^1} \rightarrow 0. \quad (1.54)$$

Therefore,

$$\Phi'(\mathbf{u}) = a(\mathbf{u}, \cdot) - \langle \mathbf{f}, \cdot \rangle \quad (1.55)$$

is the Fréchet derivative of Φ at \mathbf{u} and the first-order necessary conditions for an extremal point $\mathbf{u} \in V_0$ of Φ are exactly the weak form (1.51):

$$0 \stackrel{!}{=} \Phi'(\mathbf{u})\mathbf{v} = a(\mathbf{u}, \mathbf{v}) - \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} \quad (1.56)$$

for all $\mathbf{v} \in V_0$. Since the quadratic functional $\mathbf{u} \mapsto a(\mathbf{u}, \mathbf{u})$ is strictly convex on V_0 due to the coercivity of a and therefore Φ is also strictly convex, minimizing Φ is equivalent to the first-order optimality conditions being satisfied. We thus seek a solution of the minimization problem

$$\inf_{\mathbf{v} \in V_0} \Phi(\mathbf{v}) \quad (1.57)$$

using methods of convex analysis. Note that such a minimizer is unique if it exists due to the strict convexity of Φ .

Remark 1.2. To enforce the divergence-free condition in (1.57), one can also introduce a Lagrange multiplier. Defining the Lagrangian functional

$$\mathcal{L}: V \times Q \rightarrow \mathbb{R}, \quad (\mathbf{u}, p) \mapsto \Phi(\mathbf{u}) - \int_{\Omega} p \operatorname{div} \mathbf{u} \, d\mathbf{x}, \quad (1.58)$$

the first-order optimality conditions read

$$\begin{aligned} 0 &\stackrel{!}{=} \mathcal{L}'(\mathbf{u}, p)(\mathbf{v}, q) \\ &= \Phi'(\mathbf{u})\mathbf{v} - \int_{\Omega} p \operatorname{div} \mathbf{v} \, d\mathbf{x} - \int_{\Omega} q \operatorname{div} \mathbf{u} \, d\mathbf{x} \\ &= \int_{\Omega} 2\eta \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) : \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \, d\mathbf{x} - \int_{\Omega} \mathbf{f} \cdot \mathbf{v} - \int_{\Omega} p \operatorname{div} \mathbf{v} \, d\mathbf{x} - \int_{\Omega} q \operatorname{div} \mathbf{u} \, d\mathbf{x} \end{aligned} \quad (1.59)$$

for all $(\mathbf{v}, q) \in V \times Q$, which is exactly the weak formulation (1.16). Hence, the pressure field p can be seen as a Lagrange multiplier for the divergence-free condition.

In order to prove existence of a solution to the optimization problem (1.57), we employ the *direct method in the calculus of variations* (see e.g. Dacorogna, 2007). The general procedure is as follows:

1. Let $(\mathbf{u}_n)_n \subset V_0$ be an *extremal sequence* for (1.57), i.e.,

$$\lim_{n \rightarrow \infty} \Phi(\mathbf{u}_n) = \inf_{\mathbf{v} \in V_0} \Phi(\mathbf{v}). \quad (1.60)$$

2. Show existence of a subsequence $(\mathbf{u}_{n_k})_k \subseteq (\mathbf{u}_n)_n$ converging to some $\mathbf{u} \in V_0$ with respect to a suitable topology.
3. Prove that Φ is *lower semi-continuous* at \mathbf{u} with respect to the topology of Item 2. Then,

$$\inf_{\mathbf{v} \in V_0} \Phi(\mathbf{v}) \leq \Phi(\mathbf{u}) \leq \liminf_{k \rightarrow \infty} \Phi(\mathbf{u}_{n_k}) = \inf_{\mathbf{v} \in V_0} \Phi(\mathbf{v}), \quad (1.61)$$

and thus $\mathbf{u} = \arg \min_{\mathbf{v} \in V_0} \Phi(\mathbf{v})$ is a solution to the minimization problem.

Item 2 is usually accomplished by showing that Φ is *coercive*², i.e.,

$$\Phi(\mathbf{v}) \rightarrow \infty \text{ for } \|\mathbf{v}\|_{H^1} \rightarrow \infty, \quad (1.62)$$

since this implies boundedness³ of the extremal sequence $(\mathbf{u}_n)_n$. In the present case, since V_0 is reflexive as a Hilbert space, any bounded sequence has a weakly convergent subsequence (see e.g. Werner, 2018, Theorem III.3.7) and thus we need to show in Item 3 that Φ is weakly lower semi-continuous. If the minimization problem is stated in a non-reflexive space as in Chapter 3, it is often still possible to take advantage of the boundedness of the extremal sequence to obtain a subsequence that is convergent with respect to a different topology. In any case, the proof follows the general ideas stated above.

²Note that despite being the same word, this term describes a different property than in Section 1.2.1.

However, the connection for linear problems can be seen in (1.63).

³As long as the objective function Φ does not take the value $+\infty$ everywhere.

In the linear case, where Φ of (1.52) is a quadratic functional, the coercivity of the bilinear form a immediately implies that the objective function is coercive:

$$\begin{aligned}\Phi(\mathbf{v}) &= \frac{1}{2}a(\mathbf{v}, \mathbf{v}) - \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} \\ &\geq \frac{1}{2}c_3 \|\mathbf{v}\|_{H^1}^2 - \|\mathbf{f}\|_{V_0'} \|\mathbf{v}\|_{H^1} \quad \longrightarrow \infty \quad \text{for } \|\mathbf{v}\|_{H^1} \rightarrow \infty\end{aligned}\tag{1.63}$$

since the constants are positive and the quadratic term dominates. We therefore obtain a subsequence $(\mathbf{u}_{n_k})_k \subseteq (\mathbf{u}_n)_n$ converging weakly to a limit point $\mathbf{u} \in V_0$:

$$\mathbf{u}_{n_k} \xrightarrow{k \rightarrow \infty} \mathbf{u} \quad \text{in } V_0.\tag{1.64}$$

A convenient property of convex functions is that lower semi-continuity implies weak lower semi-continuity (Ekeland and Temam, 1999, Chapter I, Corollary 2.2). It is therefore sufficient to prove continuity of Φ , which we already know since we proved that Φ is even Fréchet differentiable. Hence, \mathbf{u} is the unique minimizer of Φ and the unique solution to the weak form (1.51).

Stability of the solution follows from estimate (1.63) and the minimizing property of \mathbf{u} :

$$0 = \Phi(\mathbf{0}) \geq \Phi(\mathbf{u}) \geq \frac{1}{2}c_3 \|\mathbf{u}\|_{H^1}^2 - \|\mathbf{f}\|_{V_0'} \|\mathbf{u}\|_{H^1},\tag{1.65}$$

which immediately implies

$$\|\mathbf{u}\|_{H^1} \leq 2c_3^{-1} \|\mathbf{f}\|_{V_0'}.\tag{1.66}$$

We have already seen in Section 1.2.1 that the inf-sup condition holds on $V \times Q$, and thus existence of a unique pressure field $p \in Q$ such that (\mathbf{u}, p) is the unique weak solution of the mixed formulation (1.16) is guaranteed (Girault and Raviart, 1986, Chapter I, Corollary 4.1). Since the a-priori estimate for p only relied on the inf-sup condition and the stability estimate for \mathbf{u} , stability as well as continuous dependence on the data follow in complete analogy to Section 1.2.1.

1.3 Structure of the Thesis

This thesis is roughly separated into two parts: Chapters 2 and 3 deal with well-posedness of the forward Stokes equations for different rheological models, whereas Chapters 4 and 5 investigate possibilities to recover subsurface material parameters and geometric structures in computational inverse problems governed by the Stokes equations.

Chapter 2 — Non-Newtonian Power-law Rheology: Well-Posedness of the Forward Problem

Convection in the Earth's mantle is often modeled by an Arrhenius-type power-law viscosity. We discuss mathematical properties of this highly nonlinear non-Newtonian rheology, which we use as the forward model for the inverse problem in Chapter 4. The main contributions are:

1. We state the forward problem as a convex optimization problem for the velocity field in a subspace of the Sobolev space $W^{1,r}(\Omega)^d$, where r depends on the power-law parameter.
2. Using methods of convex analysis, existence of a unique solution to this optimization problem is proved. By means of an appropriate inf-sup condition on a subspace of $W^{1,r}(\Omega)^d \times L^{r'}(\Omega)$ due to Amrouche and Girault (1994), existence of a unique pressure field corresponding to the velocity field is established. Additionally, we provide stability estimates for the solution.

Chapter 3 — Ideal Visco-plastic Rheology: Existence and (Non-)Uniqueness of Solutions to the Stokes Equations

We consider a highly nonlinear visco-plastic rheology model. Adding a positive lower bound to the effective viscosity, which can be seen as a type of regularization, yields existence of a unique regularized solution in a subspace of the Hilbert space $H^1(\Omega)^d$, while the problem can become singular without regularization and the solution must be sought in the space of functions of bounded deformations, $BD(\Omega)$, a non-reflexive Banach space. The main contributions are:

1. We derive a convex energy functional and show that minimizing this functional is equivalent to finding a weak solution of the Stokes equations with regularized visco-plastic rheology.
2. For strictly positive regularization parameters, well-posedness of the problem in a subspace of $H^1(\Omega)^d$ is proved.
3. We generalize the optimization problem for the case without regularization by extending the energy functional to $BD(\Omega)$.
4. Existence of a corresponding energy minimizer is shown in a subspace of $BD(\Omega)$ under a smallness assumption on the body forces acting on the fluid. However, uniqueness cannot be guaranteed in general.

Chapter 4 — Adjoint-based Inversion of Principal Stress Directions

The work in this chapter has been published in a peer-reviewed journal (Reuber et al., 2020). We address the inverse problem of inferring rheological parameters inside the Earth from sparse near-surface measurements. Taking advantage of efficient adjoint-based gradients, the feasibility of a deterministic inversion using principal stress directions as data is explored. The main contributions are:

1. We introduce principal stress directions as a data type in geodynamical inversions and formulate the problem as a regularized infinite-dimensional constrained optimization problem.
2. Employing a formal Lagrangian approach, the adjoint equations are derived both in their weak and strong form. This allows us to compute derivatives of the objective function efficiently at the cost of solving one additional linear partial differential equation.

3. The adjoint-based derivatives are implemented and incorporated into the C++ software package LaMEM (Kaus et al., 2016), an efficient 3D finite-difference Stokes solver built on PETSc (Balay et al., 2023).⁴
4. In simplified numerical examples, we are able to recover piecewise constant subsurface material parameters from near-surface measurements, emphasizing the potential of including principal stress directions as data in real-world applications.

Chapter 5 — Identifying Geometric Structures and Rheological Properties Inside the Earth: A Bayesian Level Set Method

Most of this chapter has been published in a peer-reviewed journal (Holbach, Gurnis, and Stadler, 2023). We consider the inverse problem of reconstructing subsurface structures and material parameters like viscosity or density from surface observations of plate velocities and normal tractions. Treating this problem in the Bayesian framework allows us to quantify uncertainties in the inferred parameters and shapes. The main contributions are:

1. We formulate the infinite-dimensional Bayesian inverse problem of identifying geometric structures and rheological parameters inside the Earth using a level-set approach to describe boundaries between different materials.
2. Well-posedness of the Bayesian inverse problem: We apply the theory of Iglesias, Lu, and Stuart (2016) to show that the posterior probability distribution is well-defined and Lipschitz continuous with respect to the observations in the Hellinger distance along with stability estimates.
3. We propose a method to construct tailored level-set prior distributions with adjusted mean and covariance representing seismic knowledge, building a bridge between seismic and mechanical models.
4. The framework is implemented in a Python code based on FEniCS (Logg, Mardal, and Wells, 2012), an open-source software library for solving partial differential equations using the finite element method.
5. In two numerical examples, we study the trade-offs between geometry, density, and viscosity using dimension-robust Markov chain Monte Carlo sampling methods to approximate the posterior distribution and quantify uncertainties in the inferred shapes and material parameters. Additionally, we investigate how informative different data types are and demonstrate the benefits of combining them.

⁴The implementation was mainly executed by Georg Reuber and Anton Popov.

2 CHAPTER

Non-Newtonian Power-law Rheology: Well-Posedness of the Forward Problem

In this chapter, we discuss well-posedness of the Stokes equations with a non-Newtonian power-law rheology as mentioned in Chapter 1. This model describes the governing equations of the computational inverse problem we address in Chapter 4.

Power-law viscosities have been used to model the flow in a wide range of quite different applications. Besides the numerical analysis of general non-Newtonian fluids (Glowinski and Xu, 2011), such models have been employed in the geophysical context in magma dynamics (McKenzie, 1984), mantle convection (Schubert, Turcotte, and Olson, 2001), and other geodynamical studies (Baumann and Kaus, 2015; Räss et al., 2017). Furthermore, power-law rheologies have been used in glaciology to describe ice sheet dynamics (Hutter, 1983), where the model for the effective viscosity is known as Glen’s law. Isaac, Stadler, and Ghattas (2015) developed scalable solvers for Stokes equations with shear-thinning power-law viscosities based on high-order finite-element discretizations with applications to the Antarctic ice sheet; compared to the model we describe in Section 2.1, the effective viscosity was regularized to avoid infinite viscosity for vanishing deformations. In Jouvét and Rappaz (2011), existence of a unique weak solution of this regularized model was shown with mixed boundary conditions that are important for realistic modeling of ice sheet dynamics. In addition to geophysical applications, shear-thickening power-law rheologies have been considered in hemodynamics (Hundertmark-Zaušková, Lukáčová-Medvidová, and Nečasová, 2016). Apart from the different physical behavior of the flow — the effective viscosity increases when the shear stresses increase — the authors investigated an unsteady fluid-structure interaction problem, in which the domain also evolved with time.

Well-posedness of the Stokes system using various rheological models has been studied extensively (see e.g. Ladyzhenskaya, 2003; Giga and Novotný, 2018, and references therein), but to the best of the author’s knowledge not the particular model described in Section 2.1 with its specific boundary conditions, which governs the inverse problem in Chapter 4. Thus, for the reader’s convenience we include a self-contained presentation of the well-posedness of this model, following the usual strategy from the pertinent literature.

This chapter is structured as follows: After describing the rheological model, we define appropriate function spaces for weak solutions of the forward problem. Then, we eliminate the pressure from the equations by considering the subspace of divergence-free velocity fields and derive a strictly convex energy functional whose

symbol	meaning	function space
$\tilde{\eta}_0$	viscosity prefactor	$L_+^\infty(\Omega)$
$\dot{\epsilon}_0$	reference strain rate	\mathbb{R}_+
n	power-law exponent	\mathbb{R}_+
E_a	activation energy	$L_+^\infty(\Omega)$
p_l	lithostatic pressure	$L_+^\infty(\Omega)$
V_a	activation volume	$L_+^\infty(\Omega)$
R	universal gas constant	\mathbb{R}_+
T	temperature	$\mathcal{C}(\bar{\Omega}) \cap L_+^\infty(\Omega)$

Table 2.1 Rheological parameters and corresponding function spaces.

first-order optimality conditions are the weak form of the Stokes equations in this subspace. Existence of a unique minimizer of the energy functional is proved using methods of convex analysis. By means of an appropriate inf-sup condition due to Amrouche and Girault (1994), we establish existence of a unique pressure field corresponding to the velocity field. Additionally, we give stability estimates for the solution.

2.1 Rheological Model

We consider an *Arrhenius-type power-law viscosity* (see e.g. Ranalli, 1995, Chapters 10 and 12)

$$\eta = \eta(\mathbf{x}, \mathbf{u}) = \tilde{\eta}_0 \left(\frac{\dot{\epsilon}_{\text{II}}(\mathbf{u})}{\dot{\epsilon}_0} \right)^{\frac{1}{n}-1} \exp \left(\frac{E_a + p_l V_a}{nRT} \right). \quad (2.1)$$

The viscosity prefactor $\tilde{\eta}_0$ can be seen as the viscosity at some reference strain rate $\dot{\epsilon}_0$. The power-law exponent $n \geq 1$ controls the nonlinear behavior of the fluid through the square root of the second invariant of the strain rate tensor, $\dot{\epsilon}_{\text{II}}(\mathbf{u}) = \sqrt{\frac{1}{2} \dot{\boldsymbol{\epsilon}}(\mathbf{u}) : \dot{\boldsymbol{\epsilon}}(\mathbf{u})}$. If $n = 1$, the model is linear and mathematically equivalent to the rheology discussed in Section 1.2. For $n > 1$, however, the exponent becomes negative and therefore (2.1) describes a strain-rate-weakening (or shear-thinning) nonlinear non-Newtonian viscosity. Typical values for the power-law exponent in geodynamical models range from 1 to 4 (cf. Ranalli, 1995, Chapter 10).

The exponential term in the definition of the power-law viscosity is an Arrhenius-type law that includes the viscosity's dependence on temperature, T . Furthermore, it is influenced by the power-law exponent n as well as the activation energy E_a , activation volume V_a , the universal gas constant R , and the lithostatic pressure p_l . Note that the latter is not the solution variable p of the Stokes equations but instead a prescribed depth-dependent approximation.

Most of the parameters in (2.1) are material-dependent or constant. Table 2.1 lists mathematical function spaces of the rheological parameters that are physically plausible. While the power-law exponent n may vary throughout the Earth's mantle, there is uncertainty in its specific values. Furthermore, it is often assumed to be constant throughout the upper mantle and lithosphere (e.g. Zhong, Gurnis, and

2.2 Weak Formulation of the Forward Problem

Moresi, 1998; Baumann and Kaus, 2015; Rudi, 2019). Since the model domains we consider in the simulations of Chapter 4 are located in these regions of the Earth, we assume n to be constant here as well.

Given the assumptions from Table 2.1, it follows that the mathematical properties of the viscosity model (2.1) can be equivalently described via

$$\eta = \eta(\mathbf{x}, \mathbf{u}) = \eta_0(\mathbf{x}) |\dot{\boldsymbol{\epsilon}}(\mathbf{u})|^{\frac{1}{n}-1}, \quad (2.2)$$

where $|\cdot|$ denotes the Frobenius norm and

$$\eta_0 := \tilde{\eta}_0 \left(\sqrt{2} \dot{\boldsymbol{\epsilon}}_0 \right)^{1-\frac{1}{n}} \exp \left(\frac{E_a + p_l V_a}{nRT} \right) \in L_+^\infty(\Omega) \quad (2.3)$$

is a new reference viscosity that includes all the spatial dependencies and constants of (2.1) that do not depend on the velocity field \mathbf{u} . Hence, we prove well-posedness using the notation of (2.2).

Note that the expression for the effective viscosity (2.2) has a singularity for $\dot{\boldsymbol{\epsilon}}(\mathbf{u}) = \mathbf{0}$. However, the physically relevant quantity in the Stokes equations is the viscous stress tensor

$$\boldsymbol{\tau}(\mathbf{u}) = 2\eta(\mathbf{u})\dot{\boldsymbol{\epsilon}}(\mathbf{u}) = 2\eta_0(\mathbf{x}) |\dot{\boldsymbol{\epsilon}}(\mathbf{u})|^{\frac{1}{n}-1} \dot{\boldsymbol{\epsilon}}(\mathbf{u}), \quad (2.4)$$

which should vanish if there are no deformations. Due to $n \geq 1$, this is consistent with the above definition (2.2) since, for arbitrary $\mathbf{E} \in \mathbb{R}^{d \times d} \setminus \{\mathbf{0}\}$,

$$|\mathbf{E}|^{\frac{1}{n}-1} \mathbf{E} \rightarrow \mathbf{0} \quad \text{for} \quad \mathbf{E} \rightarrow \mathbf{0}. \quad (2.5)$$

Therefore, we read $|\mathbf{0}|^{\frac{1}{n}-1} \mathbf{0}$ as $\mathbf{0}$ and $\boldsymbol{\tau}(\mathbf{u})$ is well-defined for arbitrary velocity fields \mathbf{u} .

2.2 Weak Formulation of the Forward Problem

Let $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, be a bounded domain with polytopic boundary $\partial\Omega$, i.e., Ω is a polygon in $2D$ and a polyhedron in $3D$. To find an appropriate function space for weak solutions of the Stokes equations with power-law viscosity, recall the weak form of the linear viscous case treated in Section 1.2.2: Find a divergence-free velocity field \mathbf{u} satisfying

$$\int_{\Omega} 2\eta\dot{\boldsymbol{\epsilon}}(\mathbf{u}) : \dot{\boldsymbol{\epsilon}}(\mathbf{v}) \, d\mathbf{x} = \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} \quad (2.6)$$

for all divergence-free test functions \mathbf{v} . Choosing $\mathbf{v} := \mathbf{u}$ and using the effective viscosity (2.2), the left-hand side takes the form

$$\int_{\Omega} 2\eta_0 |\dot{\boldsymbol{\epsilon}}(\mathbf{u})|^{\frac{1}{n}-1} \dot{\boldsymbol{\epsilon}}(\mathbf{u}) : \dot{\boldsymbol{\epsilon}}(\mathbf{u}) \, d\mathbf{x} = \int_{\Omega} 2\eta_0 |\dot{\boldsymbol{\epsilon}}(\mathbf{u})|^{\frac{1}{n}+1} \, d\mathbf{x}. \quad (2.7)$$

This quantity, also known as viscous dissipation, should be finite, and thus it seems natural to choose $W^{1, \frac{1}{n}+1}(\Omega)^d$ as the ansatz space. Note that this is also consistent with the linear viscous case, $n = 1$, for which we proved existence of a unique solution in a subspace of $H^1(\Omega)^d = W^{1,2}(\Omega)^d$.

Let $n \geq 1$ be given. For $r(n) := \left(\frac{1}{n} + 1\right) \in (1, 2]$, define the spaces

$$V := \left\{ \mathbf{u} \in W^{1,r(n)}(\Omega)^d \mid \mathbf{u} \cdot \boldsymbol{\nu} = 0 \text{ on } \partial\Omega \right\}, \quad (2.8)$$

$$V_0 := \left\{ \mathbf{u} \in W^{1,r(n)}(\Omega)^d \mid \operatorname{div} \mathbf{u} = 0 \text{ on } \Omega, \mathbf{u} \cdot \boldsymbol{\nu} = 0 \text{ on } \partial\Omega \right\} \subset V, \quad (2.9)$$

both equipped with the topology of $W^{1,r(n)}(\Omega)^d$. As usual, the boundary condition is to be understood in the sense of traces; for the sake of completeness, we state the corresponding trace theorem, which can be found in Nečas, 2012, Chapter 3, Theorem 1.1:

Theorem 2.1. *Let $\Omega \subset \mathbb{R}^d$, $d \geq 2$, be a bounded Lipschitz domain with boundary $\partial\Omega$, and define $r' := r/(r-1)$ for $1 < r < \infty$. There is a linear, bounded, and surjective trace operator*

$$\gamma: W^{1,r'}(\Omega) \rightarrow W^{1-\frac{1}{r'},r'}(\partial\Omega) \quad (2.10)$$

satisfying $\gamma(v) = v|_{\partial\Omega}$ for all $v \in C^\infty(\bar{\Omega})$. Furthermore, for all $v \in W^{1,r'}(\Omega)$ and $\mathbf{u} \in W^{1,r}(\Omega)^d$, we obtain a Green's formula:

$$\int_{\partial\Omega} v \mathbf{u} \cdot \boldsymbol{\nu} \, ds = \int_{\Omega} \mathbf{u} \cdot \nabla v \, d\mathbf{x} + \int_{\Omega} v \operatorname{div} \mathbf{u} \, d\mathbf{x}, \quad (2.11)$$

where

$$v|_{\partial\Omega} \in W^{1-\frac{1}{r'},r'}(\partial\Omega), \quad \mathbf{u} \cdot \boldsymbol{\nu}|_{\partial\Omega} \in W^{1-\frac{1}{r},r}(\partial\Omega)$$

holds in the trace sense.

For the remainder of this chapter, we use the short notation $r := r(n) = \frac{1}{n} + 1$ since $n \geq 1$ is fixed.

Proposition 2.2. *$(V, \|\cdot\|_{W^{1,r}})$ and $(V_0, \|\cdot\|_{W^{1,r}})$ are reflexive Banach spaces.*

PROOF. Since $r > 1$, $W^{1,r}(\Omega)^d$ is a reflexive Banach space (see e.g. Brezis, 2011, Proposition 9.1) and it suffices to show that V and V_0 are closed subspaces of $W^{1,r}(\Omega)^d$.

Since the trace operator from $W^{1,r}(\Omega)^d$ to $W^{1-\frac{1}{r},r}(\partial\Omega)^d$ is continuous, the boundary condition $\mathbf{u} \cdot \boldsymbol{\nu} = 0$ holds for limit points $\mathbf{u} \in W^{1,r}(\Omega)^d$ of arbitrary convergent sequences in V or V_0 , and thus V is closed.

Due to the inequality

$$\|\operatorname{div} \mathbf{u}\|_{L^r} \leq \sqrt{d} \|\mathbf{u}\|_{W^{1,r}}, \quad (2.12)$$

the linear operator $\operatorname{div}: W^{1,r}(\Omega)^d \rightarrow L^r(\Omega)$ is continuous. Therefore, $\operatorname{div} \mathbf{u} = 0$ on Ω for limit points $\mathbf{u} \in W^{1,r}(\Omega)^d$ of arbitrary convergent sequences in V_0 , which in turn yields $\mathbf{u} \in V_0$ and V_0 is closed. \square

Recall that the weak form of the Stokes equations from Section 1.2.1 involves the dual pairing

$$\langle p, \operatorname{div} \mathbf{v} \rangle = \int_{\Omega} p \operatorname{div} \mathbf{v} \, d\mathbf{x}. \quad (2.13)$$

2.2 Weak Formulation of the Forward Problem

Since $\operatorname{div} \mathbf{v} \in L^r(\Omega)$ for $\mathbf{v} \in V$, we choose the space

$$Q := L_0^{r'}(\Omega) := \left\{ p \in L^{r'}(\Omega) \mid \int_{\Omega} p \, d\mathbf{x} = 0 \right\}, \quad (2.14)$$

where $r' := r(n)' = (1+n) \in [2, \infty)$ is the Hölder conjugate of $r(n) = 1 + 1/n$, for the pressure. Since Ω is bounded, the linear functional $p \mapsto \int_{\Omega} p \, d\mathbf{x}$, $p \in L^{r'}(\Omega)$, is continuous and thus $(Q, \|\cdot\|_{L^{r'}})$ is a reflexive Banach space as a closed subspace of $L^{r'}(\Omega)$. Note that $L^{r'}(\Omega)^d \subseteq L^r(\Omega)^d$ and thus $W^{1,r'}(\Omega)^d \subseteq W^{1,r}(\Omega)^d$ due to $r' \geq r$ and Ω being bounded.

Following the same arguments as in Sections 1.2.1 and 1.2.2, we obtain the weak form of the Stokes equations (1.7) and (1.8) with power-law rheology (2.2): Given $\mathbf{f} \in V'$, find $(\mathbf{u}, p) \in V \times Q$ satisfying

$$\int_{\Omega} 2\eta_0 |\dot{\boldsymbol{\varepsilon}}(\mathbf{u})|^{\frac{1}{n}-1} \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) : \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \, d\mathbf{x} - \int_{\Omega} p \operatorname{div} \mathbf{v} \, d\mathbf{x} - \int_{\Omega} q \operatorname{div} \mathbf{u} \, d\mathbf{x} = \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} \quad (2.15)$$

for all $(\mathbf{v}, q) \in V \times Q$. For arbitrary $\mathbf{u} \in V$, we observe that the viscous stress tensor $\boldsymbol{\tau}(\mathbf{u}) = 2\eta_0 |\dot{\boldsymbol{\varepsilon}}(\mathbf{u})|^{\frac{1}{n}-1} \dot{\boldsymbol{\varepsilon}}(\mathbf{u})$ lies in $L^{r'}(\Omega)^{d \times d}$:

$$\begin{aligned} \|\boldsymbol{\tau}(\mathbf{u})\|_{L^{r'}} &\leq 2 \|\eta_0\|_{\infty} \left(\int_{\Omega} \left| |\dot{\boldsymbol{\varepsilon}}(\mathbf{u})|^{\frac{1}{n}-1} \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) \right|^{1+n} \, d\mathbf{x} \right)^{1/(1+n)} \\ &= 2 \|\eta_0\|_{\infty} \left(\int_{\Omega} |\dot{\boldsymbol{\varepsilon}}(\mathbf{u})|^{\frac{1}{n}(1+n)} \, d\mathbf{x} \right)^{1/(1+n)} \\ &= 2 \|\eta_0\|_{\infty} \|\dot{\boldsymbol{\varepsilon}}(\mathbf{u})\|_{L^r}^{r/r'} < \infty. \end{aligned} \quad (2.16)$$

Hence, the left-hand side of (2.15) is indeed well-defined due to Hölder's inequality.

As a first step towards proving well-posedness of (2.15), we show that Korn's inequality holds on V and thus also on $V_0 \subset V$. We start by showing that the only element in V that is in the strain rate tensor's nullspace is $\mathbf{0}$.

Lemma 2.3. *Let $\dot{\boldsymbol{\varepsilon}}: V \rightarrow L^r(\Omega)^{d \times d}$. Then its nullspace only contains the zero velocity field, i.e., $\mathcal{N}(\dot{\boldsymbol{\varepsilon}}) = \{\mathbf{0}\}$.*

PROOF. The nullspace of $\dot{\boldsymbol{\varepsilon}}$ consists of rigid displacements (see e.g. Temam, 1985, Chapter I, Lemma 1.1), i.e., functions of the form

$$\mathbf{r}: \Omega \rightarrow \mathbb{R}^d, \quad \mathbf{x} \mapsto \mathbf{A}\mathbf{x} + \mathbf{b}, \quad \mathbf{b} \in \mathbb{R}^d, \quad \mathbf{A} \in \mathbb{R}^{d \times d} \text{ with } \mathbf{A} = -\mathbf{A}^{\top}. \quad (2.17)$$

Let $\mathbf{r} \in V$ be a rigid displacement. According to the trace theorem (Theorem 2.1),

$$0 = \mathbf{r}(\mathbf{x}) \cdot \boldsymbol{\nu}(\mathbf{x}) = (\mathbf{A}\mathbf{x} + \mathbf{b}) \cdot \boldsymbol{\nu}(\mathbf{x}) \quad (2.18)$$

holds for almost every $\mathbf{x} \in \partial\Omega$.

Since $\Omega \subset \mathbb{R}^d$ is a polytope, there are subsets $\Gamma_1, \dots, \Gamma_d \subset \partial\Omega$, each with positive $(d-1)$ -dimensional Lebesgue measure, which can be parameterized by hyperplanes whose normal vectors are linearly independent (see Figure 2.1), i.e.,

$$\Gamma_i = \{\mathbf{x} \in \mathbb{R}^d \mid \mathbf{x} \cdot \boldsymbol{\nu}_i - \alpha_i = 0\} \cap \partial\Omega, \quad \alpha_i \in \mathbb{R}, \quad \boldsymbol{\nu}_i \in \mathbb{R}^d \setminus \{\mathbf{0}\}, \quad i \in \{1, \dots, d\}, \quad (2.19)$$

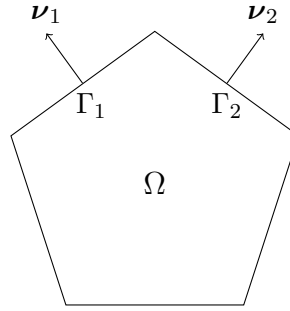


Figure 2.1 Sketch of a potential polygonal domain $\Omega \subset \mathbb{R}^2$ as in the proof of Lemma 2.3. The outer normals ν_1 and ν_2 of two neighboring sides, Γ_1 and Γ_2 , are necessarily linearly independent. In three dimensions, one can instead choose three (neighboring) planes whose normals are linearly independent.

and $\{\nu_1, \dots, \nu_d\}$ forms a basis of \mathbb{R}^d . Due to (2.18) and $\mathbf{A} = -\mathbf{A}^\top$, for arbitrary $i \in \{1, \dots, d\}$, we obtain

$$\begin{aligned}
 0 &= -(\mathbf{A}\mathbf{x} + \mathbf{b}) \cdot \nu_i \\
 &= -(\mathbf{A}\mathbf{x}) \cdot \nu_i - \mathbf{b} \cdot \nu_i \\
 &= -\mathbf{x} \cdot \mathbf{A}^\top \nu_i - \mathbf{b} \cdot \nu_i \\
 &= \mathbf{x} \cdot \mathbf{A}\nu_i - \mathbf{b} \cdot \nu_i
 \end{aligned} \tag{2.20}$$

for all $\mathbf{x} \in \Gamma_i$. As (2.20) defines a hyperplane¹ and is valid for all $\mathbf{x} \in \Gamma_i$, the two hyperplanes must coincide. Therefore, the corresponding normal vectors are linearly dependent, i.e., there are $\lambda_i \in \mathbb{R}$ such that

$$\mathbf{A}\nu_i = \lambda_i \nu_i, \quad i \in \{1, \dots, d\}. \tag{2.21}$$

However, eigenvalues of skew-symmetric matrices are imaginary, which immediately implies $\lambda_i = 0$ for all $i \in \{1, \dots, d\}$. Since $\{\nu_1, \dots, \nu_d\}$ forms a basis of \mathbb{R}^d , (2.21) yields $\mathbf{A} = \mathbf{0}$. Plugging this into (2.20), we obtain

$$\mathbf{b} \cdot \nu_i = 0, \quad i \in \{1, \dots, d\}, \tag{2.22}$$

which implies $\mathbf{b} = \mathbf{0}$, again because the normal vectors form a basis of \mathbb{R}^d . Hence, $\mathbf{r} \equiv \mathbf{0}$ and thus $\mathcal{N}(\dot{\boldsymbol{\varepsilon}}) = \{\mathbf{0}\}$. \square

The above result allows us to prove Korn's inequality on V . There are many situations in which Korn's inequality holds and the proofs often follow similar ideas. Since the inequality is of such central importance for solution theory about Stokes equations, we carry the proof out following the strategy of Oleinik (1992, Chapter I, Theorem 2.5), where a similar theorem is proved in $H^1(\Omega)^d$.

Proposition 2.4. *There is a constant $C_K > 0$ such that*

$$\|\dot{\boldsymbol{\varepsilon}}(\mathbf{u})\|_{L^r} \geq C_K \|\mathbf{u}\|_{W^{1,r}} \tag{2.23}$$

for all $\mathbf{u} \in V$.

¹If $\nu_i \in \mathcal{N}(\mathbf{A})$, this is technically not true. However, it is what we actually aim to show and no further arguments are needed in this case.

PROOF. Assume (2.23) is not valid. Then there is a sequence $(\mathbf{u}_m)_m \subset V$ such that

$$\|\dot{\boldsymbol{\varepsilon}}(\mathbf{u}_m)\|_{L^r} < \frac{1}{m} \|\mathbf{u}_m\|_{W^{1,r}} \quad \text{for all } m \in \mathbb{N}, \quad (2.24)$$

or equivalently after normalizing,

$$\|\dot{\boldsymbol{\varepsilon}}(\mathbf{u}_m)\|_{L^r} < \frac{1}{m}, \quad \|\mathbf{u}_m\|_{W^{1,r}} = 1 \quad \text{for all } m \in \mathbb{N}, \quad (2.25)$$

which immediately implies

$$\dot{\boldsymbol{\varepsilon}}(\mathbf{u}_m) \xrightarrow{m \rightarrow \infty} \mathbf{0} \quad \text{in } L^r(\Omega)^{d \times d}. \quad (2.26)$$

Since $(\mathbf{u}_m)_m$ is bounded in $W^{1,r}(\Omega)^d$ and V is weakly sequentially closed, there is a subsequence, without loss of generality using the same indices, and $\mathbf{u} \in V$ such that

$$\mathbf{u}_m \xrightarrow{m \rightarrow \infty} \mathbf{u} \quad \text{in } W^{1,r}(\Omega)^d. \quad (2.27)$$

The compact embedding of $W^{1,r}(\Omega)^d$ into $L^r(\Omega)^d$ (see Adams and Fournier, 2003, Theorem 6.3 with $1 < r < \infty$) implies

$$\mathbf{u}_m \xrightarrow{m \rightarrow \infty} \mathbf{u} \quad \text{in } L^r(\Omega)^d. \quad (2.28)$$

Furthermore, according to Geymonat, Suquet, and Nedelec (1986, Theorem 1) and since $1 < r < \infty$, there is a constant $c > 0$ such that

$$\|\mathbf{v}\|_{W^{1,r}} \leq c (\|\mathbf{v}\|_{L^r} + \|\dot{\boldsymbol{\varepsilon}}(\mathbf{v})\|_{L^r}) \quad (2.29)$$

for all $\mathbf{v} \in W^{1,r}(\Omega)^d$, which yields

$$\|\mathbf{u}_m - \mathbf{u}_l\|_{W^{1,r}} \leq c (\|\mathbf{u}_m - \mathbf{u}_l\|_{L^r} + \|\dot{\boldsymbol{\varepsilon}}(\mathbf{u}_m - \mathbf{u}_l)\|_{L^r}), \quad m, l \in \mathbb{N}, \quad (2.30)$$

and due to (2.26) and (2.28), $(\mathbf{u}_m)_m$ is a Cauchy sequence in $W^{1,r}(\Omega)^d$. Since $W^{1,r}(\Omega)^d$ is complete, $(\mathbf{u}_m)_m$ converges strongly in $W^{1,r}(\Omega)^d$, necessarily to its weak limit $\mathbf{u} \in V$. Additionally, $\|\mathbf{u}\|_{W^{1,r}} = 1$ since $\|\mathbf{u}_m\|_{W^{1,r}} = 1$ for all $m \in \mathbb{N}$.

The continuity of $\dot{\boldsymbol{\varepsilon}}: V \rightarrow L^r(\Omega)^{d \times d}$ and (2.26) now yield $\dot{\boldsymbol{\varepsilon}}(\mathbf{u}) = \mathbf{0}$, implying $\mathbf{u} = \mathbf{0}$ according to Lemma 2.3. However, this is a contradiction to $\|\mathbf{u}\|_{W^{1,r}} = 1$. \square

Remark 2.5. Although $r(n) \in (1, 2]$ due to the physical requirements on n , the statement remains true for $1 < s < \infty$ on the analogously defined space $V \subset W^{1,s}(\Omega)^d$. In fact, the proof does not change since the tools and citations used are valid for arbitrary $1 < s < \infty$. However, Korn's inequality does not hold for $s = 1$ or $s = \infty$ (see Ornstein, 1962, and references therein), which will raise issues in Chapter 3.

2.3 Energy Minimization

Similar to Section 1.2.2, we first eliminate the pressure from the equations and show existence of a unique divergence-free velocity field $\mathbf{u} \in V_0$ satisfying

$$\int_{\Omega} 2\eta_0 |\dot{\boldsymbol{\varepsilon}}(\mathbf{u})|^{\frac{1}{n}-1} \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) : \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \, d\mathbf{x} = \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} \quad (2.31)$$

for all (divergence-free) test functions $\mathbf{v} \in V_0$.

To reformulate the weak form as a constrained minimization problem, we introduce the function

$$g: \Omega \times \mathbb{R}^{d \times d} \rightarrow \mathbb{R}, \quad (\mathbf{x}, \mathbf{E}) \mapsto \frac{2n}{n+1} \eta_0(\mathbf{x}) |\mathbf{E}|^{\frac{1}{n}+1}, \quad (2.32)$$

with the corresponding functional

$$\mathcal{G}: V_0 \rightarrow \mathbb{R}, \quad \mathbf{u} \mapsto \int_{\Omega} g(\mathbf{x}, \dot{\mathbf{e}}(\mathbf{u})(\mathbf{x})) \, d\mathbf{x}, \quad (2.33)$$

and define the energy functional

$$\Phi: V_0 \rightarrow \mathbb{R}, \quad \mathbf{u} \mapsto \mathcal{G}(\mathbf{u}) - \int_{\Omega} \mathbf{f} \cdot \mathbf{u} \, d\mathbf{x}. \quad (2.34)$$

Proposition 2.6. *The energy functional Φ defined in (2.34) is strictly convex. Furthermore, the first-order necessary conditions for minimizers of Φ coincide with the weak form of the Stokes equations (2.31).*

PROOF. We start by showing strict convexity of $g(\mathbf{x}, \cdot)$ for almost every $\mathbf{x} \in \Omega$ and deduce strict convexity of \mathcal{G} from this. Let $t \in (0, 1)$ and $\mathbf{E} \neq \mathbf{E}^* \in \mathbb{R}^{d \times d}$ be arbitrary. Since the real function $y \mapsto y^{\frac{1}{n}+1}$ is strictly monotonically increasing for $y \geq 0$, we obtain

$$\begin{aligned} g(\mathbf{x}, t\mathbf{E} + (1-t)\mathbf{E}^*) &= \frac{2n}{n+1} \eta_0(\mathbf{x}) |t\mathbf{E} + (1-t)\mathbf{E}^*|^{\frac{1}{n}+1} \\ &\leq \frac{2n}{n+1} \eta_0(\mathbf{x}) (t|\mathbf{E}| + (1-t)|\mathbf{E}^*|)^{\frac{1}{n}+1}, \end{aligned} \quad (2.35)$$

and equality holds if and only if one of the following cases occurs:

- (a) $\mathbf{E} = \mathbf{0}$,
- (b) $\mathbf{E}^* = \mathbf{0}$,
- (c) $\mathbf{E} = \lambda \mathbf{E}^*$ for some $\lambda > 0$.

If none of the above conditions holds, the (strict) convexity of the real function $y \mapsto y^{\frac{1}{n}+1}$ yields

$$\begin{aligned} g(\mathbf{x}, t\mathbf{E} + (1-t)\mathbf{E}^*) &< \frac{2n}{n+1} \eta_0(\mathbf{x}) (t|\mathbf{E}| + (1-t)|\mathbf{E}^*|)^{\frac{1}{n}+1} \\ &\leq \frac{2n}{n+1} \eta_0(\mathbf{x}) \left(t|\mathbf{E}|^{\frac{1}{n}+1} + (1-t)|\mathbf{E}^*|^{\frac{1}{n}+1} \right) \\ &= tg(\mathbf{x}, \mathbf{E}) + (1-t)g(\mathbf{x}, \mathbf{E}^*). \end{aligned} \quad (2.36)$$

In case (a), we obtain strict inequality via

$$\begin{aligned} |t\mathbf{E} + (1-t)\mathbf{E}^*|^{\frac{1}{n}+1} &= |(1-t)\mathbf{E}^*|^{\frac{1}{n}+1} = (1-t)^{\frac{1}{n}+1} |\mathbf{E}^*|^{\frac{1}{n}+1} \\ &< (1-t)|\mathbf{E}^*|^{\frac{1}{n}+1} = t|\mathbf{E}|^{\frac{1}{n}+1} + (1-t)|\mathbf{E}^*|^{\frac{1}{n}+1} \end{aligned} \quad (2.37)$$

since $0 < 1 - t < 1$, $\mathbf{E}^* \neq \mathbf{0} = \mathbf{E}$, and $\frac{1}{n} + 1 > 1$. Case (b) can be treated analogously. If condition (c) holds, note that $\lambda \neq 1$ since $\mathbf{E} \neq \mathbf{E}^*$. Therefore, $\lambda |\mathbf{E}^*| \neq |\mathbf{E}^*|$ and the strict convexity and monotonicity of the real function $y \mapsto y^{\frac{1}{n}+1}$ yield

$$\begin{aligned} |t\mathbf{E} + (1-t)\mathbf{E}^*|^{\frac{1}{n}+1} &= (t\lambda |\mathbf{E}^*| + (1-t)|\mathbf{E}^*|)^{\frac{1}{n}+1} \\ &< t\lambda^{\frac{1}{n}+1} |\mathbf{E}^*|^{\frac{1}{n}+1} + (1-t)|\mathbf{E}^*|^{\frac{1}{n}+1} \\ &= t|\mathbf{E}|^{\frac{1}{n}+1} + (1-t)|\mathbf{E}^*|^{\frac{1}{n}+1}. \end{aligned} \quad (2.38)$$

Combining the cases, we have shown that $g(\mathbf{x}, \cdot)$ is strictly convex for almost every $\mathbf{x} \in \Omega$.

Now, let $\mathbf{u} \neq \mathbf{v} \in V_0$ be arbitrary. To take advantage of the strict convexity of $g(\mathbf{x}, \cdot)$, we need to ensure that $\dot{\mathbf{e}}(\mathbf{u}) \neq \dot{\mathbf{e}}(\mathbf{v})$ on a subset of Ω that has positive Lebesgue measure. Assume the contrary is true, i.e., $\dot{\mathbf{e}}(\mathbf{u}) = \dot{\mathbf{e}}(\mathbf{v})$ almost everywhere in Ω . Korn's inequality (Proposition 2.4) then yields

$$0 = \|\dot{\mathbf{e}}(\mathbf{u}) - \dot{\mathbf{e}}(\mathbf{v})\|_{L^r} = \|\dot{\mathbf{e}}(\mathbf{u} - \mathbf{v})\|_{L^r} \geq C_K \|\mathbf{u} - \mathbf{v}\|_{W^{1,r}} \geq 0, \quad (2.39)$$

implying $\mathbf{u} = \mathbf{v}$ and thus a contradiction. For arbitrary $t \in (0, 1)$, we therefore obtain

$$\begin{aligned} \mathcal{G}(t\mathbf{u} + (1-t)\mathbf{v}) &= \int_{\Omega} g(\mathbf{x}, \dot{\mathbf{e}}(t\mathbf{u} + (1-t)\mathbf{v})(\mathbf{x})) \, d\mathbf{x} \\ &= \int_{\Omega} g(\mathbf{x}, t\dot{\mathbf{e}}(\mathbf{u})(\mathbf{x}) + (1-t)\dot{\mathbf{e}}(\mathbf{v})(\mathbf{x})) \, d\mathbf{x} \\ &< \int_{\Omega} \left(tg(\mathbf{x}, \dot{\mathbf{e}}(\mathbf{u})(\mathbf{x})) + (1-t)g(\mathbf{x}, \dot{\mathbf{e}}(\mathbf{v})(\mathbf{x})) \right) \, d\mathbf{x} \\ &= t\mathcal{G}(\mathbf{u}) + (1-t)\mathcal{G}(\mathbf{v}). \end{aligned} \quad (2.40)$$

Since the function $\mathbf{u} \mapsto -\int_{\Omega} \mathbf{f} \cdot \mathbf{u} \, d\mathbf{x}$, $\mathbf{u} \in V_0$, is linear, Φ is strictly convex as the sum of a linear and a strictly convex function.

In the next step, we investigate the first-order necessary conditions for minimizers of Φ . First note that, since $\frac{1}{n} + 1 > 1$, the directional derivative of $g_{\mathbf{x}} := g(\mathbf{x}, \cdot)$ at $\mathbf{E} \in \mathbb{R}^{d \times d}$ in direction $\mathbf{E}^* \in \mathbb{R}^{d \times d}$ is given by

$$g'_{\mathbf{x}}(\mathbf{E})\mathbf{E}^* = 2\eta_0(\mathbf{x}) |\mathbf{E}|^{\frac{1}{n}-1} \mathbf{E} : \mathbf{E}^* \quad (2.41)$$

for almost every $\mathbf{x} \in \Omega$, where we again read $|\mathbf{0}|^{\frac{1}{n}-1} \mathbf{0}$ as $\mathbf{0}$. Let $\mathbf{u}, \mathbf{v} \in V_0$ be arbitrary and define the auxiliary function $\varphi(t) := \mathcal{G}(\mathbf{u} + t\mathbf{v})$. Our aim is to determine $\varphi'(0) = \mathcal{G}'(\mathbf{u})\mathbf{v}$. If we can show that we are allowed to switch integral and differential, we obtain

$$\begin{aligned} \varphi'(t) &= \frac{d}{dt} \int_{\Omega} g(\mathbf{x}, \dot{\mathbf{e}}(\mathbf{u} + t\mathbf{v})(\mathbf{x})) \, d\mathbf{x} \\ &= \int_{\Omega} \frac{\partial}{\partial t} g(\mathbf{x}, \dot{\mathbf{e}}(\mathbf{u})(\mathbf{x}) + t\dot{\mathbf{e}}(\mathbf{v})(\mathbf{x})) \, d\mathbf{x} \\ &= \int_{\Omega} g'_{\mathbf{x}}(\dot{\mathbf{e}}(\mathbf{u}) + t\dot{\mathbf{e}}(\mathbf{v}))\dot{\mathbf{e}}(\mathbf{v}) \, d\mathbf{x}. \end{aligned} \quad (2.42)$$

Since we are only interested in $\varphi'(0)$, we assume $|t| \leq 1$ and show that the last integrand is dominated by an integrable function independent of t :

$$\begin{aligned} |g'_x(\dot{\mathbf{e}}(\mathbf{u}) + t\dot{\mathbf{e}}(\mathbf{v}))\dot{\mathbf{e}}(\mathbf{v})| &= 2|\eta_0(\mathbf{x})| |\dot{\mathbf{e}}(\mathbf{u}) + t\dot{\mathbf{e}}(\mathbf{v})|^{\frac{1}{n}-1} (\dot{\mathbf{e}}(\mathbf{u}) + t\dot{\mathbf{e}}(\mathbf{v})) : \dot{\mathbf{e}}(\mathbf{v}) \\ &\leq 2\|\eta_0\|_\infty |\dot{\mathbf{e}}(\mathbf{u}) + t\dot{\mathbf{e}}(\mathbf{v})|^{\frac{1}{n}} |\dot{\mathbf{e}}(\mathbf{v})| \\ &\leq 2\|\eta_0\|_\infty (|\dot{\mathbf{e}}(\mathbf{u})| + |\dot{\mathbf{e}}(\mathbf{v})|)^{\frac{1}{n}} |\dot{\mathbf{e}}(\mathbf{v})|. \end{aligned} \quad (2.43)$$

Since $\frac{1}{n}r' = (r-1)r' = r$ and $\dot{\mathbf{e}}(\mathbf{u}), \dot{\mathbf{e}}(\mathbf{v}) \in L^r(\Omega)^{d \times d}$, Hölder's inequality yields that $2\|\eta_0\|_\infty (|\dot{\mathbf{e}}(\mathbf{u})| + |\dot{\mathbf{e}}(\mathbf{v})|)^{\frac{1}{n}} |\dot{\mathbf{e}}(\mathbf{v})| \in L^1(\Omega)$. Using Lebesgue's dominated convergence theorem, (2.42) is thus valid for $t = 0$:

$$\mathcal{G}'(\mathbf{u})\mathbf{v} = \varphi'(0) = \int_\Omega 2\eta_0 |\dot{\mathbf{e}}(\mathbf{u})|^{\frac{1}{n}-1} \dot{\mathbf{e}}(\mathbf{u}) : \dot{\mathbf{e}}(\mathbf{v}) \, d\mathbf{x}. \quad (2.44)$$

As the functional $\mathbf{u} \mapsto -\int_\Omega \mathbf{f} \cdot \mathbf{u} \, d\mathbf{x}$, $\mathbf{u} \in V_0$, is linear and bounded, the first-order necessary conditions for minimizers of Φ read

$$0 = \Phi'(\mathbf{u})\mathbf{v} = \int_\Omega 2\eta_0 |\dot{\mathbf{e}}(\mathbf{u})|^{\frac{1}{n}-1} \dot{\mathbf{e}}(\mathbf{u}) : \dot{\mathbf{e}}(\mathbf{v}) \, d\mathbf{x} - \int_\Omega \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} \quad \text{for all } \mathbf{v} \in V_0, \quad (2.45)$$

which is exactly the weak form (2.31). \square

This result allows us to prove well-posedness of the Stokes equations with power-law rheology using the minimization problem

$$\inf_{\mathbf{v} \in V_0} \Phi(\mathbf{v}). \quad (2.46)$$

Theorem 2.7. *Let $\mathbf{f} \in V' \subset V'_0$, $\eta_0 \in L^\infty_+(\Omega)$, and $n \geq 1$ be given. There is a unique minimizer $\mathbf{u} \in V_0$ of the energy functional Φ defined in (2.34). Furthermore, the solution satisfies the stability estimate*

$$\|\mathbf{u}\|_{W^{1,r}} \leq \left(\frac{n+1}{2n} \eta_{0,\min}^{-1} C_K^{-\frac{1}{n}-1} \|\mathbf{f}\|_{V'_0} \right)^n, \quad (2.47)$$

where $\eta_{0,\min} > 0$ is the essential infimum of η_0 and $C_K > 0$ the constant from Korn's inequality (2.23).

PROOF. We employ the direct method in the calculus of variations. Note that a minimizer of Φ , in case it exists, is unique since Φ is strictly convex (see Proposition 2.6). Let $(\mathbf{u}_m)_m \subset V_0$ be an extremal sequence for (2.46). Using Korn's inequality, we obtain for arbitrary $\mathbf{v} \in V_0$:

$$\begin{aligned} \Phi(\mathbf{v}) &= \int_\Omega \frac{2n}{n+1} \eta_0(\mathbf{x}) |\dot{\mathbf{e}}(\mathbf{v})|^r \, d\mathbf{x} - \int_\Omega \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} \\ &\geq \frac{2n}{n+1} \eta_{0,\min} \|\dot{\mathbf{e}}(\mathbf{v})\|_{L^r}^r - \|\mathbf{f}\|_{V'_0} \|\mathbf{v}\|_{W^{1,r}} \\ &\geq \frac{2n}{n+1} \eta_{0,\min} C_K^r \|\mathbf{v}\|_{W^{1,r}}^r - \|\mathbf{f}\|_{V'_0} \|\mathbf{v}\|_{W^{1,r}} \\ &\rightarrow \infty \quad \text{for } \|\mathbf{v}\|_{W^{1,r}} \rightarrow \infty, \end{aligned} \quad (2.48)$$

since the constants are positive and $r = \frac{1}{n} + 1 > 1$. Hence, Φ is coercive and the extremal sequence $(\mathbf{u}_m)_m$ bounded. According to Proposition 2.2, V_0 is reflexive and we therefore obtain a subsequence $(\mathbf{u}_{m_k})_k \subseteq (\mathbf{u}_m)_m$ and $\mathbf{u} \in V_0$ such that

$$\mathbf{u}_{m_k} \xrightarrow{k \rightarrow \infty} \mathbf{u} \quad \text{in } V_0. \quad (2.49)$$

Since Φ is (strictly) convex, it suffices to show that Φ is continuous, which we achieve by applying Proposition 1.1 of Chapter IV in Ekeland and Temam (1999). First note that g is a *Carathéodory function*, i.e.,

1. for all $\mathbf{E} \in \mathbb{R}^{d \times d}$, the function $\mathbf{x} \mapsto g(\mathbf{x}, \mathbf{E})$ is measurable,
2. for almost all $\mathbf{x} \in \Omega$, the function $\mathbf{E} \mapsto g(\mathbf{x}, \mathbf{E})$ is continuous.

Furthermore, for arbitrary $\mathbf{E} \in L^r(\Omega)^{d \times d}$, the function $g(\cdot, \mathbf{E}(\cdot))$ lies in $L^1(\Omega)$:

$$\int_{\Omega} |g(\mathbf{x}, \mathbf{E}(\mathbf{x}))| \, d\mathbf{x} \leq \frac{2n}{n+1} \|\eta_0\|_{\infty} \|\mathbf{E}\|_{L^r}^r < \infty. \quad (2.50)$$

The proposition cited above now yields continuity of $\mathbf{E} \mapsto g(\cdot, \mathbf{E}(\cdot))$ as a function from $L^r(\Omega)^{d \times d}$ to $L^1(\Omega)$. Since

$$\begin{aligned} \|\dot{\varepsilon}(\mathbf{u})\|_{L^r} &\leq \|\nabla \mathbf{u}\|_{L^r} \leq \|\mathbf{u}\|_{W^{1,r}} \quad \text{and} \\ \left| - \int_{\Omega} \mathbf{f} \cdot \mathbf{u} \, d\mathbf{x} \right| &\leq \|\mathbf{f}\|_{V'_0} \|\mathbf{u}\|_{W^{1,r}} \end{aligned} \quad (2.51)$$

for arbitrary $\mathbf{u} \in W^{1,r}(\Omega)^d$ and both expressions are linear in \mathbf{u} , combining the results shows that Φ is continuous as a composition of continuous functions. Therefore, Φ is also weakly lower semi-continuous (Ekeland and Temam, 1999, Chapter I, Corollary 2.2) and we obtain

$$\inf_{\mathbf{v} \in V_0} \Phi(\mathbf{v}) \leq \Phi(\mathbf{u}) \leq \liminf_{k \rightarrow \infty} \Phi(\mathbf{u}_{m_k}) = \inf_{\mathbf{v} \in V_0} \Phi(\mathbf{v}), \quad (2.52)$$

which implies $\mathbf{u} = \arg \min_{\mathbf{v} \in V_0} \Phi(\mathbf{v})$.

To prove stability of the solution, we take advantage of (2.48) and the minimizing property of \mathbf{u} :

$$0 = \Phi(\mathbf{0}) \geq \Phi(\mathbf{u}) \geq \frac{2n}{n+1} \eta_{0,\min} C_K^r \|\mathbf{u}\|_{W^{1,r}}^{\frac{1}{n}+1} - \|\mathbf{f}\|_{V'_0} \|\mathbf{u}\|_{W^{1,r}} \quad (2.53)$$

yields

$$\|\mathbf{u}\|_{W^{1,r}}^{\frac{1}{n}} \leq \frac{n+1}{2n} \eta_{0,\min}^{-1} C_K^{-r} \|\mathbf{f}\|_{V'_0}, \quad (2.54)$$

from which estimate (2.47) follows. \square

Corollary 2.8. *For given $\mathbf{f} \in V' \subset V'_0$, $\eta_0 \in L^{\infty}_+(\Omega)$, and $n \geq 1$, there is a unique weak solution $\mathbf{u} \in V_0$ to the Stokes equations with power-law rheology (2.31).*

2.4 Existence and Uniqueness of the Pressure

Existence of a unique pressure field relies on establishing an inf-sup condition for the spaces V and Q . The ideas from Section 1.2.1, where the ansatz spaces for the velocity and pressure fields were Hilbert spaces, carry over to the present case of reflexive Banach spaces (cf. Girault and Raviart, 1986, Chapter I, §4).

Lemma 2.9. *There is a constant $C > 0$ such that*

$$\inf_{0 \neq q \in Q} \sup_{0 \neq \mathbf{v} \in V} \frac{-\int_{\Omega} q \operatorname{div} \mathbf{v} \, d\mathbf{x}}{\|\mathbf{v}\|_{W^{1,r}} \|q\|_{L^{r'}}} \geq C. \quad (2.55)$$

PROOF. Due to Amrouche and Girault (1994, Corollary 3.2), there exists a constant $C > 0$ satisfying

$$\sup_{0 \neq \mathbf{v} \in W_0^{1,r}(\Omega)^d} \frac{-\int_{\Omega} q \operatorname{div} \mathbf{v} \, d\mathbf{x}}{\|\mathbf{v}\|_{W^{1,r}}} \geq C \|q\|_{L^{r'}} \quad (2.56)$$

for every $q \in Q$. Since $W_0^{1,r}(\Omega)^d \subset V$, the supremum can only take a larger value if we replace $W_0^{1,r}(\Omega)^d$ by V . Therefore,

$$\sup_{0 \neq \mathbf{v} \in V} \frac{-\int_{\Omega} q \operatorname{div} \mathbf{v} \, d\mathbf{x}}{\|\mathbf{v}\|_{W^{1,r}}} \geq C \|q\|_{L^{r'}} \quad (2.57)$$

and the inf-sup condition (2.55) holds. \square

Validity of the inf-sup condition guarantees existence of a unique pressure field solving the Stokes equations and allows us to derive a stability estimate. For the sake of completeness, we state this result together with the results from Section 2.3:

Theorem 2.10. *For given $\mathbf{f} \in V'$, $\eta_0 \in L_+^\infty(\Omega)$, and $n \geq 1$, there is a unique solution $(\mathbf{u}, p) \in V \times Q$ to the mixed weak form of the Stokes equations with power-law rheology (2.15). The solution satisfies the stability estimates*

$$\|\mathbf{u}\|_{W^{1,r}} \leq \left(\frac{n+1}{2n} \eta_{0,\min}^{-1} C_K^{-\frac{1}{n}-1} \|\mathbf{f}\|_{V'} \right)^n \quad (2.58)$$

and

$$\|p\|_{L^{r'}} \leq C^{-1} \left(\|\eta_0\|_\infty \frac{n+1}{n} \eta_{0,\min}^{-1} C_K^{-\frac{1}{n}-1} + 1 \right) \|\mathbf{f}\|_{V'}, \quad (2.59)$$

where $\eta_{0,\min} > 0$ is the essential infimum of η_0 and $C_K, C > 0$ are the constants from Korn's inequality (2.23) and the inf-sup condition (2.55), respectively.

PROOF. Existence of a unique solution $(\mathbf{u}, p) \in V \times Q$ follows from Corollary 2.8 and Lemma 2.9 using Theorem 2.2 of Howell and Walkington (2011).

It only remains to show the stability estimate for the pressure field p . Using the inf-sup condition (2.55), the weak form (2.15) with $q \equiv 0$ as well as Hölder's

inequality, we obtain

$$\begin{aligned}
\|p\|_{L^{r'}} &\leq C^{-1} \sup_{0 \neq \mathbf{v} \in V} \frac{-\int_{\Omega} p \operatorname{div} \mathbf{v} \, d\mathbf{x}}{\|\mathbf{v}\|_{W^{1,r}}} \\
&= C^{-1} \sup_{0 \neq \mathbf{v} \in V} \frac{-\langle \boldsymbol{\tau}(\mathbf{u}), \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \rangle + \langle \mathbf{f}, \mathbf{v} \rangle}{\|\mathbf{v}\|_{W^{1,r}}} \\
&\leq C^{-1} \sup_{0 \neq \mathbf{v} \in V} \frac{\|\boldsymbol{\tau}(\mathbf{u})\|_{L^{r'}} \|\dot{\boldsymbol{\varepsilon}}(\mathbf{v})\|_{L^r} + \|\mathbf{f}\|_{V'} \|\mathbf{v}\|_{W^{1,r}}}{\|\mathbf{v}\|_{W^{1,r}}} \\
&\leq C^{-1} (\|\boldsymbol{\tau}(\mathbf{u})\|_{L^{r'}} + \|\mathbf{f}\|_{V'}).
\end{aligned} \tag{2.60}$$

Due to (2.16), the $L^{r'}$ norm of the viscous stress tensor $\boldsymbol{\tau}(\mathbf{u})$ can be bounded from above by the $W^{1,r}$ norm of \mathbf{u} . Combining this with the stability estimate (2.47) yields

$$\begin{aligned}
\|p\|_{L^{r'}} &\leq C^{-1} \left(2 \|\eta_0\|_{\infty} \|\dot{\boldsymbol{\varepsilon}}(\mathbf{u})\|_{L^r}^{r/r'} + \|\mathbf{f}\|_{V'} \right) \\
&\leq C^{-1} \left(2 \|\eta_0\|_{\infty} \|\mathbf{u}\|_{W^{1,r}}^{r/r'} + \|\mathbf{f}\|_{V'} \right) \\
&\leq C^{-1} \left(2 \|\eta_0\|_{\infty} \left(\frac{n+1}{2n} \eta_{0,\min}^{-1} C_K^{-\frac{1}{n}-1} \|\mathbf{f}\|_{V'} \right)^{nr/r'} + \|\mathbf{f}\|_{V'} \right), \\
&= C^{-1} \left(\|\eta_0\|_{\infty} \frac{n+1}{n} \eta_{0,\min}^{-1} C_K^{-\frac{1}{n}-1} + 1 \right) \|\mathbf{f}\|_{V'},
\end{aligned} \tag{2.61}$$

where we used the relation $nr = r'$. \square

2.5 Conclusion

We have shown the existence of a unique velocity-pressure pair (\mathbf{u}, p) solving the weak form of the Stokes equations with an Arrhenius-type nonlinear non-Newtonian power-law rheology, which is a practically relevant model for a wide range of physical applications. The solution spaces depend on the power-law exponent n and thus on the degree of nonlinearity in the effective viscosity. Additionally, we provided stability estimates for the solution and therefore ensured well-posedness of this forward problem.

A natural extension of the present model is to allow the power-law exponent to vary spatially, e.g., $n \in L_+^{\infty}(\Omega)$. The key difference to a model with constant power-law exponent is the requirement for L^r spaces with spatially varying exponents r , which fit into the framework of so-called *Orlicz spaces* (see e.g. Adams and Fournier, 2003, Chapter 8). To ensure existence of a solution in this setting, some type of Korn's inequality on a suitable function space is needed. For certain classes of Orlicz spaces using different boundary conditions than the ones employed here, such inequalities have been proved (Breit and Diening, 2011).

A further discussion of space-dependent power laws, however, is beyond the scope of this work because variations of n can be neglected in our applications of Chapter 4, as we elaborated in the beginning of Section 2.1.

3 CHAPTER

Ideal Visco-plastic Rheology: Existence and (Non-)Uniqueness of Solutions to the Stokes Equations

When modeling plate motion on a global scale, it is more realistic to account for both viscous and plastic deformation inside the Earth. As outlined in Chapter 1, this is often accomplished by defining the effective viscosity, η , as the minimum of a model for viscous deformation, η_{visc} , and a model for plastic deformation, η_{plast} , i.e.,

$$\eta = \min(\eta_{\text{visc}}, \eta_{\text{plast}}). \quad (3.1)$$

In this chapter, we discuss well-posedness of Stokes equations for a specific choice of such a visco-plastic rheology.

There are numerous geodynamic studies using various visco-plastic models (e.g. Moresi et al., 2007; Stadler et al., 2010; Rudi, 2019). One approach is to mimic plasticity through power-law viscosities as in Chapter 2 with large power-law exponents (Zhong, Gurnis, and Moresi, 1998). The resulting rheology, which the authors refer to as pseudo-plastic, is characterized by a very slow increase of stresses with higher strain rates. The idea behind this approach is to avoid unrealistically high stresses in practical examples, and thus imitate the behavior of plastic deformation. Spiegelman, May, and Wilson (2016) have performed numerical studies and formally discussed solvability of Stokes equations for selected examples with another visco-plastic rheology. Instead of taking the minimum of a viscous and plastic model as in (3.1), they use a so-called inverse viscosity mixing, i.e., the effective viscosity is defined as the inverse of the sum of the reciprocals of the viscous and plastic models. Due to the extreme nonlinearities characteristic of visco-plastic rheologies, the numerical studies highlight the difficulties of finding a solution to the corresponding discrete Stokes systems — in several examples, convergence of the solutions cannot be observed when using a Newton solver, even after hundreds of iterations. Similar issues occur in Kaus (2010), where a visco-plastic rheology similar to (3.1) is employed. In their numerical study, the angles of shear bands, regions of high strain rates and thus plastic deformation, depend on the discretization of the simulation domain.

In Rudi, Shih, and Stadler (2020), an efficient Newton solver for certain visco-plastic rheologies including the minimum model (3.1) was developed. Furthermore, an energy functional was defined and it was formally shown that the first-order necessary conditions are the weak form of the corresponding Stokes system. In their

numerical simulations, a lower bound of the viscosity was employed in the form of

$$\eta = \underline{\eta} + \min(\eta_{\text{visc}}, \eta_{\text{plast}}), \quad (3.2)$$

where $\underline{\eta} \geq 0$ serves as a regularization parameter. Subsequently, the authors hinted at existence and uniqueness of solutions depending on whether the regularized model ($\underline{\eta} > 0$) or the unregularized model ($\underline{\eta} = 0$) is considered. In this chapter, we take an analogous energy minimization perspective and investigate both regularized and unregularized problems rigorously.

Complementary to the references for mathematical studies of Stokes systems given in Chapter 2, Málek et al. (1996) have discussed measure-valued solutions for a class of non-Newtonian fluids with strain-rate-dependent viscosities. However, lower bounds on the viscous stress tensor were assumed that are violated by the rheological model of this chapter. While Lanzendörfer (2011) additionally allowed for a pressure-dependent viscosity, differentiability assumptions on the constitutive relationship had to be made that are also not satisfied by the visco-plastic model considered in our work. Besides systems describing fluid dynamics, plastic deformation is a fundamental research topic in solid mechanics (e.g. Hill, 1998; Han and Reddy, 2013; Temam, 1985). While the problems exhibit similarities, in solid mechanics one solves for displacement fields rather than velocity and pressure fields. Nevertheless, similar mathematical approaches can be employed, and in particular the solution space and concepts presented in this chapter follow the theory of Temam (1985).

This chapter is structured as follows: After specifying the rheological model in Section 3.1, we prove well-posedness of the regularized model in a subspace of H^1 in Section 3.2 using arguments from convex analysis and a well-known inf-sup condition. In Section 3.3, we generalize the optimization problem for $\underline{\eta} = 0$ and define a suitable solution space, which turns out to be a subspace of the space of functions of bounded deformation. Existence of minimizers of the generalized objective function is proved under a smallness assumption on the right-hand side forcing term, whereas uniqueness of a solution cannot be guaranteed in general. Additionally, stability estimates are provided.

3.1 Rheological Model

As mentioned above, we consider a *visco-plastic rheology* (see e.g. Moresi et al., 2007; Kaus, 2010; Stadler et al., 2010; Rudi, 2019) of the form

$$\eta = \eta(\mathbf{x}, \mathbf{u}) = \underline{\eta} + \min\left(\eta_r(\mathbf{x}), \frac{\tau_y(\mathbf{x})}{2 \dot{\epsilon}_{\text{II}}(\mathbf{u})}\right), \quad (3.3)$$

where $\eta_r \in L_+^\infty(\Omega)$ denotes the spatially dependent reference viscosity in the viscous regime. The second argument of the minimum models plastic deformation using a von Mises yield criterion (e.g. Fulsack, 1995; von Mises, 1913). Here, $\dot{\epsilon}_{\text{II}}(\mathbf{u}) = \sqrt{\frac{1}{2} \dot{\epsilon}(\mathbf{u}) : \dot{\epsilon}(\mathbf{u})}$ is the square root of the second invariant of the strain rate tensor and $\tau_y \in L_+^\infty(\Omega)$ is the yield stress, at which — intuitively — the flow changes its behavior from viscous (and thus linear for the above model) to plastic (and thus nonlinear). Furthermore, $\underline{\eta} \geq 0$ is a lower bound of the viscosity and serves as a

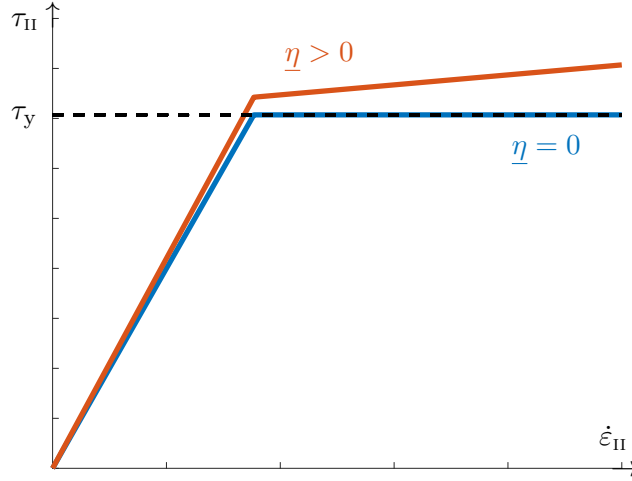


Figure 3.1 Viscous stresses τ_{II} as a function of the strain rates $\dot{\epsilon}_{II}$ at some fixed point. The red graph corresponds to $\underline{\eta} > 0$, whereas the blue graph corresponds to $\underline{\eta} = 0$. In the latter case, the stresses are bounded from above by the yield stress τ_y , which is indicated by a black dashed line.

regularization or “elliptification” as explained below. Formally, the above model (3.3) can be seen as the minimum of two power-law viscosities with exponents $n = 1$ (linear viscous) and $n = \infty$ (plastic).

Defining the viscous stress tensor as $\boldsymbol{\tau}(\mathbf{u}) := 2\eta(\mathbf{u})\dot{\boldsymbol{\epsilon}}(\mathbf{u})$, we obtain the pointwise inequality

$$\tau_{II}(\mathbf{u}) = 2\eta(\mathbf{u})\dot{\epsilon}_{II}(\mathbf{u}) = 2\left(\underline{\eta} + \min\left(\eta_r, \frac{\tau_y}{2\dot{\epsilon}_{II}(\mathbf{u})}\right)\right)\dot{\epsilon}_{II}(\mathbf{u}) \leq 2\underline{\eta}\dot{\epsilon}_{II}(\mathbf{u}) + \tau_y, \quad (3.4)$$

i.e., the viscous stresses are bounded from above by the yield stress τ_y for $\underline{\eta} = 0$. We refer to this case as the *ideal visco-plastic rheology*. However, the effective viscosity (3.3) can get arbitrarily small in this case and it is not possible to obtain ellipticity in the usual solution spaces. While this issue is resolved by choosing $\underline{\eta} > 0$, the stresses can get arbitrarily large in this case (see also Figure 3.1), which is problematic from a physical point of view. In Sections 3.2 and 3.3, we will see that a significantly larger solution space is needed for the unregularized case.

Since it will be more convenient for the mathematical theory carried out in this chapter, we use the equality $\sqrt{2}\dot{\epsilon}_{II}(\mathbf{u}) = |\dot{\boldsymbol{\epsilon}}(\mathbf{u})|$, where $|\cdot|$ denotes the Frobenius norm, to rewrite the effective viscosity as

$$\eta = \eta(\mathbf{x}, \mathbf{u}) = \underline{\eta} + \min\left(\eta_r(\mathbf{x}), \frac{\tau_y(\mathbf{x})}{\sqrt{2}|\dot{\boldsymbol{\epsilon}}(\mathbf{u})|}\right). \quad (3.5)$$

Furthermore, in the entire chapter we assume the bounded domain $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, to be a two-dimensional polygon or a three-dimensional polyhedron.

3.2 The Regularized Case: $\underline{\eta} > 0$

If the effective viscosity is bounded below by a strictly positive constant, i.e., $\underline{\eta} > 0$, existence of a solution in a Hilbert space can be shown with standard arguments. In

particular, we will see that a unique solution can be found in the same spaces as in the linear viscous case treated in Section 1.2. Note that we employ homogeneous normal flow and free slip boundary conditions throughout the entire thesis. For the remainder of this section, let $\underline{\eta} > 0$ be fixed.

Analogous to Section 1.2, we define the Hilbert spaces

$$V := \left\{ \mathbf{u} \in H^1(\Omega)^d \mid \mathbf{u} \cdot \boldsymbol{\nu} = 0 \text{ on } \partial\Omega \right\} \subset H^1(\Omega)^d, \quad (3.6)$$

$$V_0 := \left\{ \mathbf{u} \in H^1(\Omega)^d \mid \operatorname{div} \mathbf{u} = 0 \text{ on } \Omega, \mathbf{u} \cdot \boldsymbol{\nu} = 0 \text{ on } \partial\Omega \right\} \subset V, \quad (3.7)$$

$$Q := L_0^2(\Omega) := \left\{ p \in L^2(\Omega) \mid \int_{\Omega} p \, d\mathbf{x} = 0 \right\} \subset L^2(\Omega), \quad (3.8)$$

equipped with the standard H^1 and L^2 topology, respectively. The weak form of the Stokes equations (1.7) and (1.8) with visco-plastic rheology (3.5) is then obtained in the same way as in Section 1.2.1 and reads: Given $\mathbf{f} \in V'$, find $(\mathbf{u}, p) \in V \times Q$ such that

$$\begin{aligned} \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} &= \int_{\Omega} 2\underline{\eta} \dot{\boldsymbol{\epsilon}}(\mathbf{u}) : \dot{\boldsymbol{\epsilon}}(\mathbf{v}) \, d\mathbf{x} \\ &+ \int_{\Omega} \min \left(2\underline{\eta}_r, \frac{\sqrt{2}\tau_y}{|\dot{\boldsymbol{\epsilon}}(\mathbf{u})|} \right) \dot{\boldsymbol{\epsilon}}(\mathbf{u}) : \dot{\boldsymbol{\epsilon}}(\mathbf{v}) \, d\mathbf{x} \\ &- \int_{\Omega} p \operatorname{div} \mathbf{v} \, d\mathbf{x} \\ &- \int_{\Omega} q \operatorname{div} \mathbf{u} \, d\mathbf{x} \end{aligned} \quad (3.9)$$

holds for all $(\mathbf{v}, q) \in V \times Q$. Note that

$$\int_{\Omega} \left| \min \left(2\underline{\eta}_r, \frac{\sqrt{2}\tau_y}{|\dot{\boldsymbol{\epsilon}}(\mathbf{u})|} \right) \dot{\boldsymbol{\epsilon}}(\mathbf{u}) : \dot{\boldsymbol{\epsilon}}(\mathbf{v}) \right| \, d\mathbf{x} \leq 2 \|\eta_r\|_{\infty} \|\dot{\boldsymbol{\epsilon}}(\mathbf{u})\|_{L^2} \|\dot{\boldsymbol{\epsilon}}(\mathbf{v})\|_{L^2} < \infty \quad (3.10)$$

and thus every term in the weak form is indeed well-defined.

To examine existence and uniqueness of solutions, we take a convex optimization approach analogous to Chapter 2 and Section 1.2.2. As a first step, we eliminate the pressure from the equations by considering the weak form on the subspace V_0 : Given $\mathbf{f} \in V' \subset V_0'$, find $\mathbf{u} \in V_0$ such that

$$\begin{aligned} \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} &= \int_{\Omega} 2\underline{\eta} \dot{\boldsymbol{\epsilon}}(\mathbf{u}) : \dot{\boldsymbol{\epsilon}}(\mathbf{v}) \, d\mathbf{x} \\ &+ \int_{\Omega} \min \left(2\underline{\eta}_r, \frac{\sqrt{2}\tau_y}{|\dot{\boldsymbol{\epsilon}}(\mathbf{u})|} \right) \dot{\boldsymbol{\epsilon}}(\mathbf{u}) : \dot{\boldsymbol{\epsilon}}(\mathbf{v}) \, d\mathbf{x} \end{aligned} \quad (3.11)$$

holds for all $\mathbf{v} \in V_0$.

The next step is to find a convex energy functional whose first-order necessary conditions are the weak form (3.11). Looking at (3.11), it is straightforward to find antiderivatives for the two terms in the first line. On the other hand, the directional derivative of the Frobenius norm at $\mathbf{E} \neq \mathbf{0}$ in direction \mathbf{E}^* is given by $(\mathbf{E} : \mathbf{E}^*)/|\mathbf{E}|$, and rewriting the integrand of the last term as

$$\min \left(2\underline{\eta}_r, \frac{\sqrt{2}\tau_y}{|\dot{\boldsymbol{\epsilon}}(\mathbf{u})|} \right) \dot{\boldsymbol{\epsilon}}(\mathbf{u}) : \dot{\boldsymbol{\epsilon}}(\mathbf{v}) = \min \left(2\underline{\eta}_r |\dot{\boldsymbol{\epsilon}}(\mathbf{u})|, \sqrt{2}\tau_y \right) \frac{\dot{\boldsymbol{\epsilon}}(\mathbf{u}) : \dot{\boldsymbol{\epsilon}}(\mathbf{v})}{|\dot{\boldsymbol{\epsilon}}(\mathbf{u})|} \quad (3.12)$$

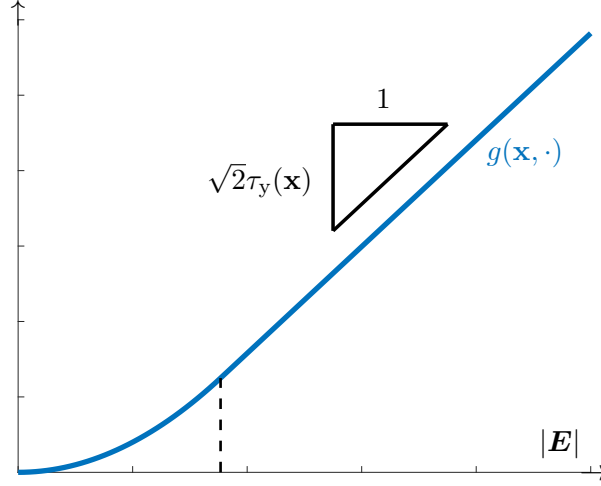


Figure 3.2 Graph of $g(\mathbf{x}, \cdot)$ for fixed $\mathbf{x} \in \Omega$ as a function of $|\mathbf{E}|$. The dashed line indicates the transition between the quadratic and linear parts of g , which occurs at $|\mathbf{E}| = \tau_y(\mathbf{x})/(\sqrt{2}\eta_r(\mathbf{x}))$. The triangle visualizes the slope of the linear part, $\sqrt{2}\tau_y(\mathbf{x})$.

motivates the definition of the function

$$g: \Omega \times \mathbb{R}^{d \times d} \rightarrow \mathbb{R},$$

$$(\mathbf{x}, \mathbf{E}) \mapsto \begin{cases} \eta_r(\mathbf{x}) |\mathbf{E}|^2 & \text{if } 2\eta_r(\mathbf{x}) |\mathbf{E}| < \sqrt{2}\tau_y(\mathbf{x}), \\ \sqrt{2}\tau_y(\mathbf{x}) |\mathbf{E}| - \frac{\tau_y^2(\mathbf{x})}{2\eta_r(\mathbf{x})} & \text{else.} \end{cases} \quad (3.13)$$

Note that the (only spatially dependent) shift in the linear part of g is chosen such that the function is continuous (see also the proof of Proposition 3.1). Corresponding to g , we define the functional

$$\mathcal{G}: L^1(\Omega)^{d \times d} \rightarrow \mathbb{R}, \quad \mathbf{E} \mapsto \int_{\Omega} g(\mathbf{x}, \mathbf{E}(\mathbf{x})) \, d\mathbf{x}. \quad (3.14)$$

For fixed $\mathbf{x} \in \Omega$, the function $g(\mathbf{x}, \cdot)$ is sometimes referred to as the Huber loss (Huber, 1964; Hintermüller and Stadler, 2006), and \mathcal{G} can be seen as a (weighted) regularization of the L^1 norm (see Figure 3.2); it is easy to verify that \mathcal{G} is indeed well-defined on $L^1(\Omega)^{d \times d}$. From a physical perspective, the quadratic part of g belongs to the viscous regime of the fluid, whereas the linear part corresponds to plastic deformation. Furthermore, the slope is bounded from above by the slope of the linear part, which causes issues in the unregularized case, i.e., $\underline{\eta} = 0$, as we will see in Section 3.3. Next, we prove that solutions of (3.11) coincide with minimizers of the energy functional

$$\Phi_{\underline{\eta}}: V_0 \rightarrow \mathbb{R}, \quad \mathbf{u} \mapsto \mathcal{G}(\dot{\boldsymbol{\varepsilon}}(\mathbf{u})) + \underline{\eta} \int_{\Omega} \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) : \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) \, d\mathbf{x} - \int_{\Omega} \mathbf{f} \cdot \mathbf{u} \, d\mathbf{x}. \quad (3.15)$$

Proposition 3.1. *The energy functional $\Phi_{\underline{\eta}}: V_0 \rightarrow \mathbb{R}$ is strictly convex. In particular, the functionals $\mathcal{G}: L^1(\Omega)^{d \times d} \rightarrow \mathbb{R}$ and $\mathcal{G} \circ \dot{\boldsymbol{\varepsilon}}: V_0 \rightarrow \mathbb{R}$ are convex. Furthermore, the first-order necessary conditions for minimizers of $\Phi_{\underline{\eta}}$ coincide with the weak form of the visco-plastic Stokes equations (3.11).*

PROOF. To prove convexity of \mathcal{G} , we introduce the auxiliary function

$$\hat{g}_{\mathbf{x}}: \mathbb{R} \rightarrow \mathbb{R}, \quad c \mapsto \begin{cases} \eta_r(\mathbf{x})c^2 & \text{if } 2\eta_r(\mathbf{x})c < \sqrt{2}\tau_y(\mathbf{x}), \\ \sqrt{2}\tau_y(\mathbf{x})c - \frac{\tau_y^2(\mathbf{x})}{2\eta_r(\mathbf{x})} & \text{else,} \end{cases} \quad (3.16)$$

for fixed $\mathbf{x} \in \Omega$. First note that $\hat{g}_{\mathbf{x}}$ is continuous:

$$\begin{aligned} \hat{g}_{\mathbf{x}} \left(\frac{\sqrt{2}\tau_y(\mathbf{x})}{2\eta_r(\mathbf{x})} - \right) &= \eta_r(\mathbf{x}) \frac{\tau_y^2(\mathbf{x})}{2\eta_r^2(\mathbf{x})} \\ &= \sqrt{2}\tau_y(\mathbf{x}) \frac{\sqrt{2}\tau_y(\mathbf{x})}{2\eta_r(\mathbf{x})} - \frac{\tau_y^2(\mathbf{x})}{2\eta_r(\mathbf{x})} \\ &= \hat{g}_{\mathbf{x}} \left(\frac{\sqrt{2}\tau_y(\mathbf{x})}{2\eta_r(\mathbf{x})} + \right), \end{aligned} \quad (3.17)$$

where “ $-$ ” and “ $+$ ” in the argument of $\hat{g}_{\mathbf{x}}$ indicate the limit of the function value when the argument is approached from below and above, respectively. Note further that the function is convex and strictly monotonically increasing on the non-negative real numbers. For arbitrary $\mathbf{E}, \mathbf{E}^* \in L^1(\Omega)^{d \times d}$ and $t \in (0, 1)$, we therefore obtain

$$\begin{aligned} \mathcal{G}(t\mathbf{E} + (1-t)\mathbf{E}^*) &= \int_{\Omega} \hat{g}_{\mathbf{x}}(|t\mathbf{E} + (1-t)\mathbf{E}^*|) \, d\mathbf{x} \\ &\leq \int_{\Omega} \hat{g}_{\mathbf{x}}(t|\mathbf{E}| + (1-t)|\mathbf{E}^*|) \, d\mathbf{x} \\ &\leq \int_{\Omega} (t\hat{g}_{\mathbf{x}}(|\mathbf{E}|) + (1-t)\hat{g}_{\mathbf{x}}(|\mathbf{E}^*|)) \, d\mathbf{x} \\ &= t\mathcal{G}(\mathbf{E}) + (1-t)\mathcal{G}(\mathbf{E}^*), \end{aligned} \quad (3.18)$$

which implies convexity of \mathcal{G} as well as $\mathcal{G} \circ \dot{\epsilon}$. Due to Korn’s inequality (Proposition 2.4), the quadratic functional

$$\mathbf{u} \mapsto \underline{\eta} \int_{\Omega} \dot{\epsilon}(\mathbf{u}) : \dot{\epsilon}(\mathbf{u}) \, d\mathbf{x} \quad (3.19)$$

is strictly convex on V_0 . Since the functional $\mathbf{u} \mapsto \langle \mathbf{f}, \mathbf{u} \rangle$, $\mathbf{u} \in V_0$, is linear, the objective function $\Phi_{\underline{\eta}} = \mathcal{G}(\dot{\epsilon}(\cdot)) + \underline{\eta} \|\dot{\epsilon}(\cdot)\|_{L^2}^2 - \langle \mathbf{f}, \cdot \rangle$ is strictly convex on V_0 .

It remains to determine the first-order necessary conditions for minimizers of $\Phi_{\underline{\eta}}$. First, observe that $\hat{g}_{\mathbf{x}}$ is continuously differentiable with derivative

$$\hat{g}'_{\mathbf{x}}(c) = \begin{cases} 2\eta_r(\mathbf{x})c & \text{if } 2\eta_r(\mathbf{x})c < \sqrt{2}\tau_y(\mathbf{x}), \\ \sqrt{2}\tau_y(\mathbf{x}) & \text{else,} \end{cases} \quad (3.20)$$

since

$$\hat{g}'_{\mathbf{x}} \left(\frac{\sqrt{2}\tau_y(\mathbf{x})}{2\eta_r(\mathbf{x})} - \right) = 2\eta_r(\mathbf{x}) \frac{\sqrt{2}\tau_y(\mathbf{x})}{2\eta_r(\mathbf{x})} = \sqrt{2}\tau_y(\mathbf{x}) = \hat{g}'_{\mathbf{x}} \left(\frac{\sqrt{2}\tau_y(\mathbf{x})}{2\eta_r(\mathbf{x})} + \right). \quad (3.21)$$

Hence, the directional derivative of the function $g_{\mathbf{x}} := g(\mathbf{x}, \cdot) = \hat{g}_{\mathbf{x}} \circ |\cdot|$ at $\mathbf{E} \in \mathbb{R}^{d \times d}$ in a direction $\mathbf{E}^* \in \mathbb{R}^{d \times d}$ is given by

$$g'_{\mathbf{x}}(\mathbf{E})\mathbf{E}^* = \begin{cases} 2\eta_r(\mathbf{x})\mathbf{E} : \mathbf{E}^* & \text{if } 2\eta_r(\mathbf{x})|\mathbf{E}| < \sqrt{2}\tau_y(\mathbf{x}), \\ \sqrt{2}\tau_y(\mathbf{x}) \frac{\mathbf{E} : \mathbf{E}^*}{|\mathbf{E}|} & \text{else.} \end{cases} \quad (3.22)$$

Note that

$$\begin{aligned} g'_{\mathbf{x}}(\mathbf{E})\mathbf{E}^* &= \mathbf{E} : \mathbf{E}^* \begin{cases} 2\eta_r(\mathbf{x}) & \text{if } 2\eta_r(\mathbf{x})|\mathbf{E}| < \sqrt{2}\tau_y(\mathbf{x}), \\ \frac{\sqrt{2}\tau_y(\mathbf{x})}{|\mathbf{E}|} & \text{else,} \end{cases} \\ &= \min\left(2\eta_r(\mathbf{x}), \frac{\sqrt{2}\tau_y(\mathbf{x})}{|\mathbf{E}|}\right) \mathbf{E} : \mathbf{E}^* \end{aligned} \quad (3.23)$$

for arbitrary $\mathbf{E}, \mathbf{E}^* \in \mathbb{R}^{d \times d}$, which was our motivation for the definition of g . Furthermore, as already suggested through Figure 3.2,

$$|g'_{\mathbf{x}}(\mathbf{E})\mathbf{E}^*| \leq \min\left(2\eta_r(\mathbf{x}), \frac{\sqrt{2}\tau_y(\mathbf{x})}{|\mathbf{E}|}\right) |\mathbf{E}| |\mathbf{E}^*| \leq \sqrt{2}\tau_y(\mathbf{x}) |\mathbf{E}^*|. \quad (3.24)$$

With the above preparations, we are ready to state the first-order necessary conditions. Let $\mathbf{u}, \mathbf{v} \in V_0$ be arbitrary and define the auxiliary function $\varphi(t) := \mathcal{G}(\dot{\varepsilon}(\mathbf{u} + t\mathbf{v}))$. For arbitrary $t \in \mathbb{R}$,

$$\begin{aligned} \varphi'(t) &= \frac{d}{dt} \int_{\Omega} g(\mathbf{x}, \dot{\varepsilon}(\mathbf{u} + t\mathbf{v})(\mathbf{x})) \, d\mathbf{x} \\ &= \int_{\Omega} \frac{\partial}{\partial t} g(\mathbf{x}, \dot{\varepsilon}(\mathbf{u})(\mathbf{x}) + t\dot{\varepsilon}(\mathbf{v})(\mathbf{x})) \, d\mathbf{x} \\ &= \int_{\Omega} g'_{\mathbf{x}}(\dot{\varepsilon}(\mathbf{u}) + t\dot{\varepsilon}(\mathbf{v}))\dot{\varepsilon}(\mathbf{v}) \, d\mathbf{x}, \end{aligned} \quad (3.25)$$

where switching integral and differential is justified by Lebesgue's dominated convergence theorem and (3.24):

$$|g'_{\mathbf{x}}(\dot{\varepsilon}(\mathbf{u}) + t\dot{\varepsilon}(\mathbf{v}))\dot{\varepsilon}(\mathbf{v})| \leq \sqrt{2} \|\tau_y\|_{\infty} |\dot{\varepsilon}(\mathbf{v})| \in L^2(\Omega) \subset L^1(\Omega). \quad (3.26)$$

Consequently, combining the above arguments yields

$$(\mathcal{G} \circ \dot{\varepsilon})'(\mathbf{u})\mathbf{v} = \varphi'(0) = \int_{\Omega} g'_{\mathbf{x}}(\dot{\varepsilon}(\mathbf{u}))\dot{\varepsilon}(\mathbf{v}) \, d\mathbf{x} = \int_{\Omega} \min\left(2\eta_r, \frac{\sqrt{2}\tau_y}{|\dot{\varepsilon}(\mathbf{u})|}\right) \dot{\varepsilon}(\mathbf{u}) : \dot{\varepsilon}(\mathbf{v}) \, d\mathbf{x}. \quad (3.27)$$

With a simple calculation, the directional derivatives of the remaining quadratic and linear terms in $\Phi_{\underline{\eta}}$ can be determined and we obtain the following first-order necessary conditions for minimizers $\mathbf{u} \in V_0$ of $\Phi_{\underline{\eta}}$:

$$\begin{aligned} 0 \stackrel{!}{=} \Phi'_{\underline{\eta}}(\mathbf{u})\mathbf{v} &= \int_{\Omega} \min\left(2\eta_r, \frac{\sqrt{2}\tau_y}{|\dot{\varepsilon}(\mathbf{u})|}\right) \dot{\varepsilon}(\mathbf{u}) : \dot{\varepsilon}(\mathbf{v}) \, d\mathbf{x} \\ &\quad + \int_{\Omega} 2\underline{\eta} \dot{\varepsilon}(\mathbf{u}) : \dot{\varepsilon}(\mathbf{v}) \, d\mathbf{x} - \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} \end{aligned} \quad (3.28)$$

for all $\mathbf{v} \in V_0$, which is indeed the weak form of the Stokes equations (3.11). \square

In view of Proposition 3.1, minimizing (3.15) is thus equivalent to finding weak solutions of (3.11). Before proving existence of a unique minimizer, we state the following:

Lemma 3.2. *The functionals $\mathcal{G}: L^1(\Omega)^{d \times d} \rightarrow \mathbb{R}$ and $\mathcal{G} \circ \dot{\boldsymbol{\varepsilon}}: V_0 \rightarrow \mathbb{R}$ are Lipschitz continuous.*

PROOF. First note that for almost all $\mathbf{x} \in \Omega$, the function $g(\mathbf{x}, \cdot)$ is Lipschitz continuous due to (3.24) and the mean value theorem (e.g. Werner, 2018, Theorem III.5.4):

$$|g(\mathbf{x}, \mathbf{E}_1) - g(\mathbf{x}, \mathbf{E}_2)| \leq \sqrt{2}\tau_y(\mathbf{x}) |\mathbf{E}_1 - \mathbf{E}_2| \quad (3.29)$$

for all $\mathbf{E}_1, \mathbf{E}_2 \in \mathbb{R}^{d \times d}$. This immediately implies

$$\begin{aligned} |\mathcal{G}(\mathbf{E}_1) - \mathcal{G}(\mathbf{E}_2)| &\leq \int_{\Omega} |g(\mathbf{x}, \mathbf{E}_1(\mathbf{x})) - g(\mathbf{x}, \mathbf{E}_2(\mathbf{x}))| \, d\mathbf{x} \\ &\leq \int_{\Omega} \sqrt{2}\tau_y(\mathbf{x}) |\mathbf{E}_1(\mathbf{x}) - \mathbf{E}_2(\mathbf{x})| \, d\mathbf{x} \\ &\leq \sqrt{2} \|\tau_y\|_{\infty} \|\mathbf{E}_1 - \mathbf{E}_2\|_{L^1}, \end{aligned} \quad (3.30)$$

and thus \mathcal{G} is Lipschitz continuous. Since Ω is bounded and $L^2(\Omega)^{d \times d}$ is therefore continuously embedded in $L^1(\Omega)^{d \times d}$, Lipschitz continuity of $\mathcal{G} \circ \dot{\boldsymbol{\varepsilon}}: V_0 \rightarrow \mathbb{R}$ follows from the linearity and continuity of $\dot{\boldsymbol{\varepsilon}}: V_0 \rightarrow L^2(\Omega)^{d \times d}$ (cf. (1.23)). \square

Theorem 3.3. *Let $\mathbf{f} \in V' \subset V'_0$. There is a unique minimizer $\mathbf{u} \in V_0$ of the functional $\Phi_{\underline{\eta}}$ defined in (3.15). Furthermore, the solution satisfies the stability estimate*

$$\|\mathbf{u}\|_{H^1} \leq \underline{\eta}^{-1} C_K^{-2} \|\mathbf{f}\|_{V'_0}, \quad (3.31)$$

where $C_K > 0$ is the constant from Korn's inequality (2.23).

PROOF. We employ the direct method in the calculus of variations. As seen in Proposition 3.1, $\Phi_{\underline{\eta}}$ is strictly convex and a minimizer, if it exists, is thus unique.

Let $(\mathbf{u}_n)_n \subset V_0$ be an extremal sequence, i.e.,

$$\Phi_{\underline{\eta}}(\mathbf{u}_n) \xrightarrow{n \rightarrow \infty} \inf_{\mathbf{v} \in V_0} \Phi_{\underline{\eta}}(\mathbf{v}). \quad (3.32)$$

Using Korn's inequality (Proposition 2.4) for arbitrary $\mathbf{v} \in V_0$ and the non-negativity of \mathcal{G} , we obtain

$$\begin{aligned} \Phi_{\underline{\eta}}(\mathbf{v}) &= \underline{\eta} \|\dot{\boldsymbol{\varepsilon}}(\mathbf{v})\|_{L^2}^2 + \mathcal{G}(\dot{\boldsymbol{\varepsilon}}(\mathbf{v})) - \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} \\ &\geq \underline{\eta} C_K^2 \|\mathbf{v}\|_{H^1}^2 - \|\mathbf{f}\|_{V'_0} \|\mathbf{v}\|_{H^1} \end{aligned} \quad (3.33)$$

and thus

$$\Phi_{\underline{\eta}}(\mathbf{v}) \rightarrow +\infty \quad \text{for } \|\mathbf{v}\|_{H^1} \rightarrow +\infty \quad (3.34)$$

since the constants are positive and the quadratic term dominates.

Due to (3.34), the sequence $(\mathbf{u}_n)_n$ is bounded, and since the Hilbert space V_0 is reflexive, there is $\mathbf{u} \in V_0$ and a weakly convergent subsequence $(\mathbf{u}_{n_k})_k \subseteq (\mathbf{u}_n)_n$ such that

$$\mathbf{u}_{n_k} \xrightarrow{k \rightarrow \infty} \mathbf{u} \quad \text{in } V_0. \quad (3.35)$$

The strain rate tensor $\dot{\boldsymbol{\varepsilon}}$ is continuous from V_0 to $L^2(\Omega)^{d \times d}$, implying continuity of the function $\mathbf{v} \mapsto \underline{\eta} \|\dot{\boldsymbol{\varepsilon}}(\mathbf{v})\|_{L^2}^2$. Since the linear function $\mathbf{v} \mapsto \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x}$ is bounded for

$\mathbf{f} \in V'_0$, Lemma 3.2 yields continuity of $\Phi_{\underline{\eta}}$. Due to its (strict) convexity, $\Phi_{\underline{\eta}}$ is also weakly lower semi-continuous (see Ekeland and Temam, 1999, Chapter I, Corollary 2.2). Combining this with (3.32) and (3.35), we obtain

$$\inf_{\mathbf{v} \in V_0} \Phi_{\underline{\eta}}(\mathbf{v}) \leq \Phi_{\underline{\eta}}(\mathbf{u}) \leq \liminf_{k \rightarrow \infty} \Phi_{\underline{\eta}}(\mathbf{u}_{n_k}) = \inf_{\mathbf{v} \in V_0} \Phi_{\underline{\eta}}(\mathbf{v}), \quad (3.36)$$

and thus $\mathbf{u} = \arg \min_{\mathbf{v} \in V_0} \Phi_{\underline{\eta}}(\mathbf{v})$.

The stability estimate follows from the minimizing property of \mathbf{u} and (3.33):

$$0 = \Phi_{\underline{\eta}}(\mathbf{0}) \geq \Phi_{\underline{\eta}}(\mathbf{u}) \geq \underline{\eta} C_K^2 \|\mathbf{u}\|_{H^1}^2 - \|\mathbf{f}\|_{V'_0} \|\mathbf{u}\|_{H^1}, \quad (3.37)$$

which immediately implies (3.31). \square

Corollary 3.4. *For arbitrary $\mathbf{f} \in V' \subset V'_0$, there is a unique weak solution $\mathbf{u} \in V_0$ of the visco-plastic Stokes equations (3.11).*

In complete analogy to Chapters 1 and 2, existence of a unique pressure field follows from the validity of the inf-sup condition on $V \times Q$ (see Section 1.2.1 or Girault and Raviart, 1986, Chapter I, Theorem 3.7). Consequently, we obtain the following:

Corollary 3.5. *For arbitrary $\mathbf{f} \in V'$, there is a unique solution $(\mathbf{u}, p) \in V \times Q$ of the mixed weak formulation (3.9). Furthermore, the velocity field \mathbf{u} satisfies the stability estimate (3.31) and for the pressure field we obtain*

$$\|p\|_{L^2} \leq C^{-1} \left[1 + 2 \left(\underline{\eta} + \|\eta_r\|_{\infty} \right) \underline{\eta}^{-1} C_K^{-2} \right] \|\mathbf{f}\|_{V'}, \quad (3.38)$$

where $C, C_K > 0$ are the constants from the inf-sup condition and Korn's inequality, respectively.

PROOF. It only remains to show the stability estimate for the pressure field. Due to the inf-sup condition, the weak form (3.9) using $q \equiv 0$, and the stability estimate (3.31) for the velocity field \mathbf{u} , we obtain

$$\begin{aligned} \|p\|_{L^2} &\leq C^{-1} \sup_{0 \neq \mathbf{v} \in V} \frac{-\int_{\Omega} p \operatorname{div} \mathbf{v} \, d\mathbf{x}}{\|\mathbf{v}\|_{H^1}} \\ &= C^{-1} \sup_{0 \neq \mathbf{v} \in V} \left(\|\mathbf{v}\|_{H^1}^{-1} \left[\int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} - \int_{\Omega} 2\underline{\eta} \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) : \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \, d\mathbf{x} \right. \right. \\ &\quad \left. \left. - \int_{\Omega} \min \left(2\underline{\eta}_r, \frac{\sqrt{2}\tau_y}{|\dot{\boldsymbol{\varepsilon}}(\mathbf{u})|} \right) \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) : \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \, d\mathbf{x} \right] \right) \\ &\leq C^{-1} \sup_{0 \neq \mathbf{v} \in V} \frac{\|\mathbf{f}\|_{V'} \|\mathbf{v}\|_{H^1} + 2\underline{\eta} \|\mathbf{u}\|_{H^1} \|\mathbf{v}\|_{H^1} + 2 \|\eta_r\|_{\infty} \|\mathbf{u}\|_{H^1} \|\mathbf{v}\|_{H^1}}{\|\mathbf{v}\|_{H^1}} \\ &\leq C^{-1} \left[1 + 2 \left(\underline{\eta} + \|\eta_r\|_{\infty} \right) \underline{\eta}^{-1} C_K^{-2} \right] \|\mathbf{f}\|_{V'}. \end{aligned} \quad \square$$

3.3 Ideal Visco-plastic Rheology: $\underline{\eta} = 0$

In this section, we assume $\underline{\eta} = 0$ and consider the unregularized case of an ideal visco-plastic rheology. Recalling the proofs of the previous section, several issues arise for $\underline{\eta} = 0$. First, the objective function is no longer strictly convex but only convex since the strictly convex term in $\Phi_{\underline{\eta}}$ vanishes; a unique solution is thus not guaranteed. However, the situation is even worse: The quadratic term in the coercivity estimate (3.33) does not exist anymore, and bounding \mathcal{G} from below by its linear part, we would instead obtain

$$\begin{aligned} \Phi_0(\mathbf{v}) &= \mathcal{G}(\dot{\boldsymbol{\varepsilon}}(\mathbf{v})) - \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \, d\mathbf{x} \\ &\geq \sqrt{2}\tau_{y,\min} \|\dot{\boldsymbol{\varepsilon}}(\mathbf{v})\|_{L^1} - \left\| \tau_y^2/2\eta_t \right\|_{L^1} - \|\mathbf{f}\|_{W'} \|\mathbf{v}\|_W, \end{aligned} \quad (3.39)$$

where $\tau_{y,\min} > 0$ denotes the essential infimum of $\tau_y \in L_+^\infty(\Omega)$ and W denotes some ansatz space for the solution. Since the only positive term in (3.39) involves the L^1 norm of $\dot{\boldsymbol{\varepsilon}}(\mathbf{v})$, none of the terms bounds the H^1 norm of \mathbf{v} such that the objective function can remain bounded for $\|\mathbf{v}\|_{H^1} \rightarrow \infty$; it is apparent that we need a larger space for solutions in this case. One might think that $W = W^{1,1}(\Omega)^d$ is an appropriate ansatz space to obtain coercivity of the objective function, however, it is known that Korn's inequality does not hold on $W^{1,1}(\Omega)^d$ (see e.g. Ornstein, 1962), which suggests that the regularity assumptions on the solution need to be relaxed even further for existence of a solution of the unregularized problem. This leads us to the space of *functions of bounded deformation*, where the components of the strain rate (or deformation) tensor are not in L^1 but only finite signed measures.

3.3.1 Functions of Bounded Deformation and a Solution Space

In the following, we define functions of bounded deformation and review the results that are most important for our purposes. The presentation follows Temam (1985, Chapter II, Sections 2 and 3) and Kohn (1979, Part II, Sections 2 and 3), and we refer the reader to these sources for proofs and more details.

Let $\mu = (\mu_{ij})_{i,j=1}^d$ be a *finite signed measure* on Ω with values in $\mathbb{R}^{d \times d}$, i.e., μ_{ij} is a finite signed measure on Ω for every $i, j \in \{1, \dots, d\}$.¹ We denote the space of all such measures μ by $\mathcal{M}(\Omega; \mathbb{R}^{d \times d})$, which is a Banach space for the *total variation* norm $\|\cdot\|_{\mathcal{M}}$:

$$\|\mu\|_{\mathcal{M}} := \sup_{\substack{\mathbf{E} \in \mathcal{C}_0(\Omega; \mathbb{R}^{d \times d}) \\ |\mathbf{E}| \leq 1 \text{ on } \Omega}} \langle \mathbf{E}, \mu \rangle. \quad (3.40)$$

Here, $\mathcal{C}_0(\Omega; \mathbb{R}^{d \times d})$ denotes the space of matrix-valued continuous functions $\mathbf{E} = (e_{ij})_{i,j=1}^d : \Omega \rightarrow \mathbb{R}^{d \times d}$ that vanish on $\partial\Omega$, and

$$\langle \mathbf{E}, \mu \rangle := \int_{\Omega} \mathbf{E}(\mathbf{x}) : \mu(d\mathbf{x}) = \sum_{i,j=1}^d \int_{\Omega} e_{ij}(\mathbf{x}) \mu_{ij}(d\mathbf{x}). \quad (3.41)$$

¹Strictly speaking, the measures are defined on the Borel sigma algebra on Ω with respect to the Euclidean topology.

3.3 Ideal Visco-plastic Rheology: $\eta = 0$

Since $\mathcal{C}_0(\Omega; \mathbb{R}^{d \times d})' = \mathcal{M}(\Omega; \mathbb{R}^{d \times d})$ (e.g. Elstrodt, 2011, Chapter VIII, Theorem 2.26), the suggestive notation $\langle \cdot, \cdot \rangle$ indeed defines a dual pairing and (3.40) is none other than the dual norm. The space $L^1(\Omega)^{d \times d}$ can be seen as a subspace of $\mathcal{M}(\Omega; \mathbb{R}^{d \times d})$ if one identifies $\mathbf{E} \in L^1(\Omega)^{d \times d}$ with the measure $\mu_{\mathbf{E}}$ given by

$$\mu_{\mathbf{E}}(A) = \int_A \mathbf{E}(\mathbf{x}) \, d\mathbf{x} \quad (3.42)$$

for every measurable set $A \subseteq \Omega$. In this case, we refer to \mathbf{E} as the *Lebesgue density* of $\mu_{\mathbf{E}}$.

The space of *functions of bounded deformation* is defined as

$$BD(\Omega) := \left\{ \mathbf{u} \in L^1(\Omega)^d \mid \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) \in \mathcal{M}(\Omega; \mathbb{R}^{d \times d}) \right\}, \quad (3.43)$$

and it is a Banach space when equipped with the norm

$$\|\mathbf{u}\|_{BD} := \|\mathbf{u}\|_{L^1} + \|\dot{\boldsymbol{\varepsilon}}(\mathbf{u})\|_{\mathcal{M}}, \quad \mathbf{u} \in BD(\Omega). \quad (3.44)$$

Definition (3.43) is to be understood in the following sense: The strain rate tensor $\dot{\boldsymbol{\varepsilon}}(\mathbf{u})$ is a priori a distribution, i.e., a real-valued continuous linear map defined on $\mathcal{C}_c^\infty(\Omega; \mathbb{R}^{d \times d})$, the space of infinitely differentiable matrix-valued functions with compact support in Ω . By “ $\dot{\boldsymbol{\varepsilon}}(\mathbf{u}) \in \mathcal{M}(\Omega; \mathbb{R}^{d \times d})$ ”, we mean that there exists a (necessarily unique) $\mu \in \mathcal{M}(\Omega; \mathbb{R}^{d \times d})$ such that

$$\dot{\boldsymbol{\varepsilon}}(\mathbf{u})(\mathbf{E}) = \int_{\Omega} \mathbf{E} : \mu(d\mathbf{x}), \quad \mathbf{E} \in \mathcal{C}_c^\infty(\Omega; \mathbb{R}^{d \times d}). \quad (3.45)$$

We then identify $\dot{\boldsymbol{\varepsilon}}(\mathbf{u})$ with μ , so that in particular the total variation norm in (3.44) is well-defined.

Note that $BD(\Omega)$ shows similarities to the well-known space $BV(\Omega)$ of scalar functions of bounded variation (e.g. Cesari, 1956; Oleinik, 1963; Ambrosio, Fusco, and Pallara, 2000), where instead of the components of the strain rate tensor *all* (distributional) first partial derivatives are finite signed measures. It can be shown that $BV(\Omega)^d$ is a subset of $BD(\Omega)$ but the inclusion is strict, i.e., the spaces do not coincide.

Theorem 3.6. *Let $\Omega \subset \mathbb{R}^d$, $d \geq 2$, be a bounded domain with Lipschitz boundary $\partial\Omega$. There is a linear and bounded trace operator*

$$\gamma_0: BD(\Omega) \rightarrow L^1(\partial\Omega)^d \quad (3.46)$$

satisfying $\gamma_0(\mathbf{u}) = \mathbf{u}|_{\partial\Omega}$ for all $\mathbf{u} \in BD(\Omega) \cap \mathcal{C}(\bar{\Omega})^d$. Furthermore, for all $\mathbf{u} \in BD(\Omega)$, $v \in C^1(\bar{\Omega})$, and $i, j \in \{1, \dots, d\}$, Green's formula holds true:

$$\frac{1}{2} \int_{\partial\Omega} v (u_i \nu_j + u_j \nu_i) \, ds = \frac{1}{2} \int_{\Omega} \left(u_i \frac{\partial v}{\partial x_j} + u_j \frac{\partial v}{\partial x_i} \right) \, d\mathbf{x} + \langle v, \dot{\boldsymbol{\varepsilon}}_{ij}(\mathbf{u}) \rangle. \quad (3.47)$$

In particular, setting $i = j$ and summing up over $i = 1, \dots, d$, we obtain

$$\int_{\partial\Omega} v \mathbf{u} \cdot \boldsymbol{\nu} \, ds = \int_{\Omega} \mathbf{u} \cdot \nabla v \, d\mathbf{x} + \langle v, \operatorname{div} \mathbf{u} \rangle. \quad (3.48)$$

Theorem 3.7. *Let $\Omega \subset \mathbb{R}^d$, $d \geq 2$, be a bounded domain with Lipschitz boundary $\partial\Omega$.*

- (a) *For all $1 \leq r \leq d/(d-1)$, the space $BD(\Omega)$ is contained in $L^r(\Omega)^d$, and there is a continuous injection of the former into the latter.*
- (b) *If $1 \leq r < d/(d-1)$, the injection of $BD(\Omega)$ into $L^r(\Omega)^d$ is compact.*

In contrast to the solution space V_0 in the regularized case, $BD(\Omega)$ is *not* reflexive. Hence, we have to work with a different topology when employing the direct method in the calculus of variations. In particular, a bounded sequence $(\mathbf{u}_n)_n \subset BD(\Omega)$ contains a subsequence $(\mathbf{u}_{n_k})_k \subseteq (\mathbf{u}_n)_n$ that converges to some $\mathbf{u} \in BD(\Omega)$ in the following sense:

$$\mathbf{u}_{n_k} \xrightarrow{k \rightarrow \infty} \mathbf{u} \quad \text{in } L^1(\Omega)^d, \quad (3.49a)$$

$$\dot{\boldsymbol{\varepsilon}}(\mathbf{u}_{n_k}) \xrightarrow{k \rightarrow \infty, *} \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) \quad \text{in } \mathcal{M}(\Omega; \mathbb{R}^{d \times d}), \quad (3.49b)$$

i.e., the velocity fields converge strongly in $L^1(\Omega)^d$ and the strain rate tensors converge weakly* in $\mathcal{M}(\Omega; \mathbb{R}^{d \times d})$, where the latter means that

$$\langle \mathbf{E}, \dot{\boldsymbol{\varepsilon}}(\mathbf{u}_{n_k}) \rangle \xrightarrow{k \rightarrow \infty} \langle \mathbf{E}, \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) \rangle \quad (3.50)$$

for all $\mathbf{E} \in \mathcal{C}_0(\Omega; \mathbb{R}^{d \times d})$. We refer to the topology defined through (3.49) as the *weak* topology* of $BD(\Omega)$.² Note that a weakly* convergent sequence is bounded in $BD(\Omega)$ and we can therefore, due to Theorem 3.7, always obtain a subsequence such that additionally

$$\mathbf{u}_{n_k} \xrightarrow{k \rightarrow \infty} \mathbf{u} \quad \text{in } L^r(\Omega)^d \quad \text{for all } 1 \leq r \leq \frac{d}{d-1}, \quad (3.51a)$$

$$\mathbf{u}_{n_k} \xrightarrow{k \rightarrow \infty} \mathbf{u} \quad \text{in } L^r(\Omega)^d \quad \text{for all } 1 \leq r < \frac{d}{d-1}. \quad (3.51b)$$

Similar to the regularized case $\underline{\eta} > 0$, we will seek minimizers in a subspace of $BD(\Omega)$, namely

$$U_0 := \{ \mathbf{u} \in BD(\Omega) \mid \operatorname{div} \mathbf{u} = 0 \text{ in } \Omega, \mathbf{u} \cdot \boldsymbol{\nu} = 0 \text{ on } \partial\Omega \}, \quad (3.52)$$

equipped with the norm $\|\cdot\|_{BD}$. The functions in U_0 are divergence-free in the sense of distributions. However, care must be taken when dealing with the boundary condition since the trace in $BD(\Omega)$ is not continuous with respect to the weak* topology (Temam, 1985, Chapter II, Section 3). Hence, it is not clear a priori whether weak* limits of sequences in U_0 are indeed in U_0 , which would cause issues when proving existence of minimizers using the direct method. On the other hand, due to the continuous inclusion of $BD(\Omega)$ in $L^r(\Omega)^d$, $1 \leq r \leq d/(d-1)$, and since the zero function trivially lies in any L^r space, U_0 is contained in

$$H^r(\operatorname{div}; \Omega) := \{ \mathbf{u} \in L^r(\Omega)^d \mid \operatorname{div} \mathbf{u} \in L^r(\Omega) \} \quad (3.53)$$

²Note that this topology is sometimes referred to as the weak topology in the reference literature.

3.3 Ideal Visco-plastic Rheology: $\eta = 0$

for arbitrary $1 < r \leq d/(d-1)$. The space $H^r(\text{div}; \Omega)$ is a Banach space when equipped with the norm

$$\|\mathbf{u}\|_{H^r(\text{div}; \Omega)} := (\|\mathbf{u}\|_{L^r}^r + \|\text{div } \mathbf{u}\|_{L^r}^r)^{1/r}, \quad \mathbf{u} \in H^r(\text{div}; \Omega). \quad (3.54)$$

We refer to Sohr (2001, Chapter II, Section 1.2) for an introduction of this function space including Theorem 3.8 stated below. Due to the inclusion $U_0 \subset H^r(\text{div}; \Omega)$ for $1 < r \leq d/(d-1)$, we can take advantage of the properties of the generalized trace for the normal component defined on $H^r(\text{div}; \Omega)$:

Theorem 3.8. *Let $\Omega \subset \mathbb{R}^d$, $d \geq 2$, be a bounded Lipschitz domain with boundary $\partial\Omega$, and define $r' := r/(r-1)$ for $1 < r < \infty$. There is a linear and bounded generalized trace operator for the normal component,*

$$\gamma_{\boldsymbol{\nu}}: H^r(\text{div}; \Omega) \rightarrow W^{-\frac{1}{r}, r}(\partial\Omega), \quad (3.55)$$

such that $\gamma_{\boldsymbol{\nu}}(\mathbf{u})$ coincides with the functional

$$g \mapsto \int_{\partial\Omega} g \mathbf{u} \cdot \boldsymbol{\nu} \, ds, \quad g \in W^{1-\frac{1}{r'}, r'}(\partial\Omega), \quad (3.56)$$

if $\mathbf{u} \in C^\infty(\bar{\Omega})^d$. Furthermore, for all $v \in W^{1, r'}(\Omega)$ and $\mathbf{u} \in H^r(\text{div}; \Omega)$, Green's formula holds true:

$$\langle v, \mathbf{u} \cdot \boldsymbol{\nu} \rangle_{\partial\Omega} = \int_{\Omega} \mathbf{u} \cdot \nabla v \, d\mathbf{x} + \int_{\Omega} v \text{div } \mathbf{u} \, d\mathbf{x}, \quad (3.57)$$

where $\langle v, \mathbf{u} \cdot \boldsymbol{\nu} \rangle_{\partial\Omega}$ is well-defined because $v|_{\partial\Omega} \in W^{1-\frac{1}{r'}, r'}(\partial\Omega)$.

This allows us to prove the following:

Proposition 3.9. *The space U_0 is weakly* sequentially closed.*

PROOF. Let $(\mathbf{u}_n)_n \subset U_0$, $\mathbf{u} \in BD(\Omega)$ with

$$\mathbf{u}_n \xrightarrow{n \rightarrow \infty} \mathbf{u} \quad \text{in } L^1(\Omega)^d, \quad (3.58a)$$

$$\dot{\boldsymbol{\varepsilon}}(\mathbf{u}_n) \xrightarrow{n \rightarrow \infty} \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) \quad \text{in } \mathcal{M}(\Omega; \mathbb{R}^{d \times d}). \quad (3.58b)$$

As mentioned above, we can extract a subsequence $(\mathbf{u}_{n_k})_k \subseteq (\mathbf{u}_n)_n$ such that additionally

$$\mathbf{u}_{n_k} \xrightarrow{k \rightarrow \infty} \mathbf{u} \quad \text{in } L^r(\Omega)^d \quad \text{for all } 1 \leq r < \frac{d}{d-1}. \quad (3.59)$$

Let $1 < r < d/(d-1)$. For arbitrary $\varphi \in C_c^\infty(\Omega)$, i.e., $\varphi: \Omega \rightarrow \mathbb{R}$ is an infinitely differentiable function with compact support in Ω , we have

$$\begin{aligned} 0 &= \int_{\Omega} \varphi \text{div } \mathbf{u}_{n_k} \, d\mathbf{x} \\ &= - \int_{\Omega} \mathbf{u}_{n_k} \cdot \nabla \varphi \, d\mathbf{x} \xrightarrow{k \rightarrow \infty} - \int_{\Omega} \mathbf{u} \cdot \nabla \varphi \, d\mathbf{x} = \int_{\Omega} \varphi \text{div } \mathbf{u} \, d\mathbf{x}, \end{aligned} \quad (3.60)$$

where the convergence follows from (3.58a). Therefore, $\operatorname{div} \mathbf{u} = 0$ in Ω and in particular $\mathbf{u} \in H^r(\operatorname{div}; \Omega)$. Now, since the occurring velocity fields are divergence-free,

$$\begin{aligned} \|\mathbf{u}_{n_k} - \mathbf{u}\|_{H^r(\operatorname{div}; \Omega)} &= (\|\mathbf{u}_{n_k} - \mathbf{u}\|_{L^r}^r + \|\operatorname{div}(\mathbf{u}_{n_k} - \mathbf{u})\|_{L^r}^r)^{1/r} \\ &= \|\mathbf{u}_{n_k} - \mathbf{u}\|_{L^r} \xrightarrow{k \rightarrow \infty} 0 \end{aligned} \quad (3.61)$$

due to (3.59) and thus

$$0 = \gamma_\nu(\mathbf{u}_{n_k}) \xrightarrow{k \rightarrow \infty} \gamma_\nu(\mathbf{u}) \quad \text{in } W^{-\frac{1}{r}, r}(\partial\Omega), \quad (3.62)$$

implying $\gamma_\nu(\mathbf{u}) = 0$ and therefore $\mathbf{u} \in U_0$. \square

Note that convergence in $BD(\Omega)$ implies weak* convergence. Due to Proposition 3.9, $(U_0, \|\cdot\|_{BD})$ is thus a Banach space as a closed subspace of $BD(\Omega)$.

3.3.2 Generalizing the Objective Function

Recalling the definition of the objective function,

$$\Phi_0(\mathbf{u}) = \mathcal{G}(\dot{\boldsymbol{\varepsilon}}(\mathbf{u})) - \int_{\Omega} \mathbf{f} \cdot \mathbf{u} \, dx \quad (3.63)$$

with \mathcal{G} of (3.14), we have seen that $\mathcal{G}(\dot{\boldsymbol{\varepsilon}}(\mathbf{u}))$ is well-defined for $\dot{\boldsymbol{\varepsilon}}(\mathbf{u}) \in L^1(\Omega)^{d \times d}$ but it is not immediately clear how to interpret the functional if $\dot{\boldsymbol{\varepsilon}}(\mathbf{u}) \in \mathcal{M}(\Omega; \mathbb{R}^{d \times d})$. In order to make sense of the expression, we take advantage of *convex conjugates* and *biconjugates* (e.g. Ekeland and Temam, 1999, Chapter I, Section 4) as explained below.

Since $\mathcal{G}(\mathbf{E}) \geq 0$ for all $\mathbf{E} \in L^1(\Omega)^{d \times d}$ and $\mathcal{G}(\mathbf{0}) = 0$, \mathcal{G} is proper.³ Due to Proposition 3.1 and Lemma 3.2, \mathcal{G} is convex and lower semi-continuous, and thus $\mathcal{G} = \mathcal{G}^{**}$ holds (Ekeland and Temam, 1999, Chapter I, Proposition 4.1), i.e., \mathcal{G} coincides with its *biconjugate functional* \mathcal{G}^{**} . Therefore,

$$\mathcal{G}(\mathbf{E}) = \mathcal{G}^{**}(\mathbf{E}) = \sup_{\mathbf{E}^* \in L^\infty(\Omega)^{d \times d}} \left(\langle \mathbf{E}^*, \mathbf{E} \rangle - \mathcal{G}^*(\mathbf{E}^*) \right), \quad \mathbf{E} \in L^1(\Omega)^{d \times d}, \quad (3.64)$$

where the second equality is the definition of the biconjugate functional and \mathcal{G}^* denotes the *convex conjugate* of \mathcal{G} defined via

$$\mathcal{G}^*(\mathbf{E}^*) = \sup_{\mathbf{E} \in L^1(\Omega)^{d \times d}} \left(\langle \mathbf{E}^*, \mathbf{E} \rangle - \mathcal{G}(\mathbf{E}) \right), \quad \mathbf{E}^* \in L^\infty(\Omega)^{d \times d}. \quad (3.65)$$

Using (3.64), we can extend \mathcal{G} to $\mathcal{M}(\Omega; \mathbb{R}^{d \times d})$ if we adjust the dual pairing accordingly. Before doing this, we investigate the convex conjugate \mathcal{G}^* of \mathcal{G} .

Lemma 3.10. *The convex conjugate of the functional \mathcal{G} is given by*

$$\mathcal{G}^*(\mathbf{E}^*) = \begin{cases} \int_{\Omega} \frac{\mathbf{E}^* : \mathbf{E}^*}{4\eta_r} \, dx & \text{if } |\mathbf{E}^*| \leq \sqrt{2}\tau_y \text{ a.e. on } \Omega, \\ +\infty & \text{else,} \end{cases} \quad (3.66)$$

for arbitrary $\mathbf{E}^* \in L^\infty(\Omega)^{d \times d}$.

³This technicality simply means that \mathcal{G} never takes the value $-\infty$ and is not identically $+\infty$.

3.3 Ideal Visco-plastic Rheology: $\eta = 0$

PROOF. We first show that for every $\mathbf{E}^* \in \mathbb{R}^{d \times d}$ and almost every $\mathbf{x} \in \Omega$, the convex conjugate of the function g is given by

$$\begin{aligned} g^*(\mathbf{x}, \mathbf{E}^*) &= \sup_{\mathbf{E} \in \mathbb{R}^{d \times d}} (\mathbf{E}^* : \mathbf{E} - g(\mathbf{x}, \mathbf{E})) \\ &= \begin{cases} \frac{|\mathbf{E}^*|^2}{4\eta_r(\mathbf{x})} & \text{if } |\mathbf{E}^*| \leq \sqrt{2}\tau_y(\mathbf{x}), \\ +\infty & \text{else.} \end{cases} \end{aligned} \quad (3.67)$$

The statement then follows from Ekeland and Temam (1999, Chapter IX, Proposition 2.1), which yields

$$\mathcal{G}^*(\mathbf{E}^*) = \int_{\Omega} g^*(\mathbf{x}, \mathbf{E}^*(\mathbf{x})) \, d\mathbf{x} \quad (3.68)$$

for arbitrary $\mathbf{E}^* \in L^\infty(\Omega)^{d \times d}$.

Note that (3.67) is valid for $\mathbf{E}^* = \mathbf{0}$ since the function g is non-negative and its minimum is 0 for almost every $\mathbf{x} \in \Omega$. Let $\mathbf{E}^* \in \mathbb{R}^{d \times d} \setminus \{\mathbf{0}\}$ and $\mathbf{x} \in \Omega$. Since $g(\mathbf{x}, \mathbf{E})$ only depends on $|\mathbf{E}|$ rather than \mathbf{E} and the inner product $\mathbf{E}^* : \mathbf{E}$ is maximal for a given value of $|\mathbf{E}^*|$ if \mathbf{E}^* and \mathbf{E} are linearly dependent, we assume without loss of generality $\mathbf{E} = c\mathbf{E}^*/|\mathbf{E}^*|$ for some $c > 0$. Defining the function

$$\begin{aligned} h_{\mathbf{x}, \mathbf{E}^*}(c) &:= \mathbf{E}^* : \mathbf{E} - g(\mathbf{x}, \mathbf{E}) \\ &= c|\mathbf{E}^*| - \begin{cases} \eta_r(\mathbf{x})c^2 & \text{if } 2\eta_r(\mathbf{x})c < \sqrt{2}\tau_y(\mathbf{x}), \\ \sqrt{2}\tau_y(\mathbf{x})c - \tau_y^2(\mathbf{x})/2\eta_r(\mathbf{x}) & \text{else,} \end{cases} \end{aligned} \quad (3.69)$$

we rewrite the supremum:

$$g^*(\mathbf{x}, \mathbf{E}^*) = \sup_{\mathbf{E} \in \mathbb{R}^{d \times d}} (\mathbf{E}^* : \mathbf{E} - g(\mathbf{x}, \mathbf{E})) = \sup_{c > 0} h_{\mathbf{x}, \mathbf{E}^*}(c). \quad (3.70)$$

For sufficiently large $c > 0$,

$$h_{\mathbf{x}, \mathbf{E}^*}(c) = c \left(|\mathbf{E}^*| - \sqrt{2}\tau_y(\mathbf{x}) \right) + \tau_y^2(\mathbf{x})/2\eta_r(\mathbf{x}) \quad (3.71)$$

and thus

$$h_{\mathbf{x}, \mathbf{E}^*}(c) \xrightarrow{c \rightarrow \infty} \begin{cases} -\infty & \text{if } |\mathbf{E}^*| < \sqrt{2}\tau_y(\mathbf{x}), \\ \tau_y^2(\mathbf{x})/2\eta_r(\mathbf{x}) & \text{if } |\mathbf{E}^*| = \sqrt{2}\tau_y(\mathbf{x}), \\ +\infty & \text{if } |\mathbf{E}^*| > \sqrt{2}\tau_y(\mathbf{x}), \end{cases} \quad (3.72)$$

which immediately implies

$$g^*(\mathbf{x}, \mathbf{E}^*) = +\infty \quad \text{if } |\mathbf{E}^*| > \sqrt{2}\tau_y(\mathbf{x}). \quad (3.73)$$

We calculate the local extremum of $h_{\mathbf{x}, \mathbf{E}^*}$ to determine the supremum in the case $|\mathbf{E}^*| \leq \sqrt{2}\tau_y(\mathbf{x})$. The parabola takes its maximum for $2\eta_r(\mathbf{x})c - |\mathbf{E}^*| = 0$, i.e., for

$$c = \frac{|\mathbf{E}^*|}{2\eta_r(\mathbf{x})} \in \left(0, \frac{\sqrt{2}\tau_y(\mathbf{x})}{2\eta_r(\mathbf{x})} \right], \quad (3.74)$$

where the range of c follows from the assumption on \mathbf{E}^* . Due to

$$h_{\mathbf{x}, \mathbf{E}^*} \left(\frac{|\mathbf{E}^*|}{2\eta_r(\mathbf{x})} \right) = \frac{|\mathbf{E}^*|}{2\eta_r(\mathbf{x})} |\mathbf{E}^*| - \eta_r(\mathbf{x}) \frac{|\mathbf{E}^*|^2}{4\eta_r^2(\mathbf{x})} = \frac{|\mathbf{E}^*|^2}{4\eta_r(\mathbf{x})} > 0 \quad (3.75)$$

and $h_{\mathbf{x}, \mathbf{E}^*}(0) = 0$, the limit in (3.72) implies that this is a global maximum and thus

$$g^*(\mathbf{x}, \mathbf{E}^*) = \frac{|\mathbf{E}^*|^2}{4\eta_r(\mathbf{x})} \quad \text{if } |\mathbf{E}^*| < \sqrt{2}\tau_y(\mathbf{x}). \quad (3.76)$$

Since the values of (3.72) and (3.75) coincide if $|\mathbf{E}^*| = \sqrt{2}\tau_y(\mathbf{x})$, this concludes the proof. \square

Rewriting \mathcal{G} in terms of its biconjugate \mathcal{G}^{**} allows us to extend the functional to $\mathcal{M}(\Omega; \mathbb{R}^{d \times d})$ simply by adjusting the dual pairing. Therefore, motivated by (3.64) and (3.66), and since $\mathcal{C}_0(\Omega; \mathbb{R}^{d \times d})' = \mathcal{M}(\Omega; \mathbb{R}^{d \times d})$, we define

$$\begin{aligned} \tilde{\mathcal{G}}: \mathcal{M}(\Omega; \mathbb{R}^{d \times d}) &\rightarrow \mathbb{R}, \\ \mu &\mapsto \sup_{\substack{\mathbf{E}^* \in \mathcal{C}_0(\Omega; \mathbb{R}^{d \times d}) \\ |\mathbf{E}^*| \leq \sqrt{2}\tau_y \text{ a.e.}}} \left(\langle \mathbf{E}^*, \mu \rangle - \int_{\Omega} \frac{\mathbf{E}^* : \mathbf{E}^*}{4\eta_r} d\mathbf{x} \right). \end{aligned} \quad (3.77)$$

This is a meaningful extension in the sense that $\tilde{\mathcal{G}}|_{L^1} = \mathcal{G}$ as shown below.

Proposition 3.11. *If $\mu \in \mathcal{M}(\Omega; \mathbb{R}^{d \times d})$ admits a Lebesgue density $\mathbf{E} \in L^1(\Omega)^{d \times d}$, i.e., $\mu = \mu_{\mathbf{E}}$ in the sense of (3.42), then $\tilde{\mathcal{G}}(\mu) = \mathcal{G}(\mathbf{E})$.*

PROOF. Let $\mu \in \mathcal{M}(\Omega; \mathbb{R}^{d \times d})$ and denote by $\mathbf{E} \in L^1(\Omega)^{d \times d}$ its Lebesgue density. Introducing the notation

$$\mathcal{F}(\mathbf{E}^*) := \int_{\Omega} \mathbf{E}^* : \mathbf{E} d\mathbf{x} - \int_{\Omega} \frac{\mathbf{E}^* : \mathbf{E}^*}{4\eta_r} d\mathbf{x}, \quad (3.78)$$

we need to show that

$$\tilde{\mathcal{G}}(\mu) = \sup_{\substack{\mathbf{E}^* \in \mathcal{C}_0(\Omega; \mathbb{R}^{d \times d}) \\ |\mathbf{E}^*| \leq \sqrt{2}\tau_y \text{ a.e.}}} \mathcal{F}(\mathbf{E}^*) = \sup_{\substack{\mathbf{E}^* \in L^\infty(\Omega)^{d \times d} \\ |\mathbf{E}^*| \leq \sqrt{2}\tau_y \text{ a.e.}}} \mathcal{F}(\mathbf{E}^*) = \mathcal{G}(\mathbf{E}). \quad (3.79)$$

Since $\mathcal{C}_0(\Omega; \mathbb{R}^{d \times d}) \subset L^\infty(\Omega)^{d \times d}$, the supremum for \mathcal{G} is taken over a larger set than the supremum for $\tilde{\mathcal{G}}$, and consequently $\tilde{\mathcal{G}}(\mu) \leq \mathcal{G}(\mathbf{E})$ holds.

To prove the other inequality, we will show that for every $\varepsilon > 0$ and $\mathbf{E}^* \in L^\infty(\Omega)^{d \times d}$, there exists $\mathbf{F}^* \in \mathcal{C}_0(\Omega; \mathbb{R}^{d \times d})$ with $|\mathbf{F}^*| \leq |\mathbf{E}^*|$ almost everywhere on Ω and

$$|\mathcal{F}(\mathbf{E}^*) - \mathcal{F}(\mathbf{F}^*)| < \varepsilon. \quad (3.80)$$

In fact, we will show the analogous statement for scalar functions since

$$\mathcal{F}(\mathbf{E}^*) = \sum_{i,j=1}^d \left(\int_{\Omega} e_{ij}^* e_{ij} d\mathbf{x} - \int_{\Omega} \frac{e_{ij}^{*2}}{4\eta_r} d\mathbf{x} \right) \quad (3.81)$$

and (3.80) then follows by applying the scalar statement to every component of \mathbf{E}^* . Note that $|\mathbf{F}^*| \leq |\mathbf{E}^*|$ is true if $|f_{ij}^*| \leq |e_{ij}^*|$ holds for each $i, j \in \{1, \dots, d\}$.

Let $e \in L^1(\Omega)$. The proof is divided into two steps: We first approximate bounded functions with simple functions, then we approximate the value of \mathcal{F} for simple functions with continuous functions that vanish on the boundary. Let $f \in L^\infty(\Omega)$ be arbitrary and denote its positive and negative parts by $f^+ := \max(f, 0)$ and $f^- := \max(-f, 0)$, respectively. Due to Elstrodt (2011, Chapter III, Corollary 4.14), there are sequences of non-negative simple functions $(g_n^+)_n$ and $(g_n^-)_n$ such that $g_n^+ \uparrow f^+$ and $g_n^- \uparrow f^-$ uniformly on Ω . Defining the simple functions $g_n := g_n^+ - g_n^-$, $n \in \mathbb{N}$, we obtain

$$g_n \xrightarrow{n \rightarrow \infty} f^+ - f^- = f \quad \text{uniformly on } \Omega \quad (3.82)$$

and

$$|g_n| = g_n^+ + g_n^- \leq f^+ + f^- = |f| \quad \text{almost everywhere on } \Omega. \quad (3.83)$$

In particular,

$$g_n(\mathbf{x})e(\mathbf{x}) \xrightarrow{n \rightarrow \infty} f(\mathbf{x})e(\mathbf{x}) \quad \text{for almost every } \mathbf{x} \in \Omega, \quad (3.84a)$$

$$\frac{g_n(\mathbf{x})^2}{4\eta_r(\mathbf{x})} \xrightarrow{n \rightarrow \infty} \frac{f(\mathbf{x})^2}{4\eta_r(\mathbf{x})} \quad \text{for almost every } \mathbf{x} \in \Omega, \quad (3.84b)$$

where the convergence in (3.84b) is in fact uniform. Since

$$|g_n e| \leq |f| |e| \leq \|f\|_\infty |e| \in L^1(\Omega), \quad (3.85a)$$

$$\frac{g_n^2}{4\eta_r} \leq \frac{\|f\|_\infty^2}{4\eta_{r,\min}} \in L^1(\Omega), \quad (3.85b)$$

where $\eta_{r,\min} > 0$ denotes the essential infimum of $\eta_r \in L^\infty_+(\Omega)$, Lebesgue's dominated convergence theorem yields⁴

$$\mathcal{F}(g_n) = \int_\Omega g_n e \, d\mathbf{x} - \int_\Omega \frac{g_n^2}{4\eta_r} \, d\mathbf{x} \xrightarrow{n \rightarrow \infty} \int_\Omega f e \, d\mathbf{x} - \int_\Omega \frac{f^2}{4\eta_r} \, d\mathbf{x} = \mathcal{F}(f). \quad (3.86)$$

Now let g be an arbitrary simple function, i.e., there are $m \in \mathbb{N}$ and

$$a_k \in \mathbb{R}, \quad A_k \subseteq \Omega, \quad k \in \{1, \dots, m\}, \quad A_k \cap A_j = \emptyset, \quad k \neq j, \quad \bigcup_{k=1}^m A_k = \Omega, \quad (3.87)$$

such that

$$g = \sum_{k=1}^m a_k \mathbf{1}_{A_k}, \quad (3.88)$$

where

$$\mathbf{1}_{A_k} : \Omega \rightarrow \mathbb{R}, \quad \mathbf{x} \mapsto \begin{cases} 1 & \text{if } \mathbf{x} \in A_k, \\ 0 & \text{else.} \end{cases} \quad (3.89)$$

⁴For simplicity, we use the same notation for \mathcal{F} when defined on scalar functions.

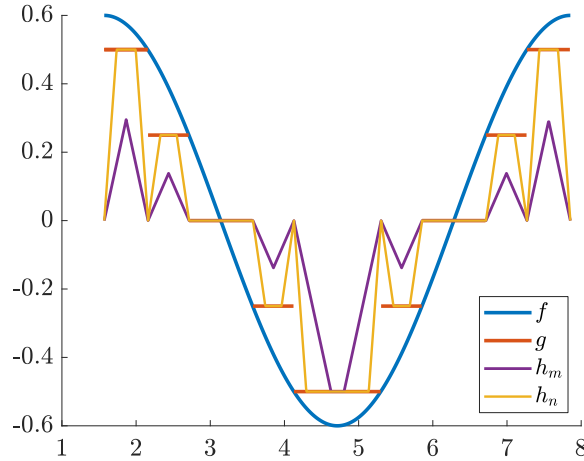


Figure 3.3 Sketch of the approximations in the proof of Proposition 3.11. The L^∞ function f is approximated with a simple function g , which in turn is approximated with C_0 functions. In the above plot, h_m and h_n are two such functions with $n > m$.

For arbitrary $n \in \mathbb{N}$ and $k \in \{1, \dots, m\}$, define

$$a_{k,n}: \Omega \rightarrow \mathbb{R}, \quad \mathbf{x} \mapsto \begin{cases} a_k & \text{if } \mathbf{x} \in A_k, \quad d(\mathbf{x}, \partial A_k) > \frac{1}{n}, \\ a_k n d(\mathbf{x}, \partial A_k) & \text{if } \mathbf{x} \in A_k, \quad d(\mathbf{x}, \partial A_k) \leq \frac{1}{n}, \\ 0 & \text{else,} \end{cases} \quad (3.90)$$

where

$$d(\mathbf{x}, \partial A_k) = \inf_{\mathbf{z} \in \partial A_k} \|\mathbf{x} - \mathbf{z}\|_2 \quad (3.91)$$

is the distance of a point \mathbf{x} to ∂A_k . Since the functions $a_{k,n}$ are continuous and vanish on $\partial\Omega$ for all $n \in \mathbb{N}$, $k \in \{1, \dots, m\}$,

$$h_n: \Omega \rightarrow \mathbb{R}, \quad \mathbf{x} \mapsto \sum_{k=1}^m a_{k,n}(\mathbf{x}) \quad (3.92)$$

lies in $C_0(\Omega)$ for all $n \in \mathbb{N}$ (see Figure 3.3 for an illustration of the approximation). By construction, $|h_n| \leq |g|$ on Ω for every $n \in \mathbb{N}$ and

$$h_n(\mathbf{x}) \xrightarrow{n \rightarrow \infty} g(\mathbf{x}) \quad \text{for almost every } \mathbf{x} \in \Omega. \quad (3.93)$$

Due to $g \in L^\infty(\Omega)$ and the pointwise convergence (3.93), we obtain analogous statements to (3.84) and (3.85) for $(h_n)_n$ and g , i.e., $(g_n)_n$ and f are replaced by $(h_n)_n$ and g , respectively. Once again using Lebesgue's dominated convergence theorem yields

$$\mathcal{F}(h_n) = \int_{\Omega} h_n e \, d\mathbf{x} - \int_{\Omega} \frac{h_n^2}{4\eta_r} \, d\mathbf{x} \xrightarrow{n \rightarrow \infty} \int_{\Omega} g e \, d\mathbf{x} - \int_{\Omega} \frac{g^2}{4\eta_r} \, d\mathbf{x} = \mathcal{F}(g). \quad (3.94)$$

We are now able to combine the above results: Let $f \in L^\infty(\Omega)$ and $\varepsilon > 0$ be arbitrary. Due to (3.86), there is a simple function g with $|g| \leq |f|$ almost everywhere on Ω such that

$$|\mathcal{F}(g) - \mathcal{F}(f)| < \varepsilon/2. \quad (3.95)$$

3.3 Ideal Visco-plastic Rheology: $\eta = 0$

For this simple function, (3.94) guarantees existence of some $h \in \mathcal{C}_0(\Omega)$, $|h| \leq |g| \leq |f|$ almost everywhere on Ω , with

$$|\mathcal{F}(h) - \mathcal{F}(g)| < \varepsilon/2. \quad (3.96)$$

Hence,

$$|\mathcal{F}(h) - \mathcal{F}(f)| \leq |\mathcal{F}(h) - \mathcal{F}(g)| + |\mathcal{F}(g) - \mathcal{F}(f)| < \varepsilon, \quad (3.97)$$

which concludes the proof. \square

Lemma 3.12. *The functionals $\tilde{\mathcal{G}}: \mathcal{M}(\Omega; \mathbb{R}^{d \times d}) \rightarrow \mathbb{R}$ and $\tilde{\mathcal{G}} \circ \dot{\varepsilon}: BD(\Omega) \rightarrow \mathbb{R}$ are convex.*

PROOF. We start by proving convexity of $\tilde{\mathcal{G}}$. For arbitrary $t \in (0, 1)$ and $\mu, \tilde{\mu} \in \mathcal{M}(\Omega; \mathbb{R}^{d \times d})$,

$$\begin{aligned} & \tilde{\mathcal{G}}(t\mu + (1-t)\tilde{\mu}) \\ &= \sup_{\substack{\mathbf{E}^* \in \mathcal{C}_0(\Omega; \mathbb{R}^{d \times d}) \\ |\mathbf{E}^*| \leq \sqrt{2}\tau_y \text{ a.e.}}} \left(\langle \mathbf{E}^*, t\mu + (1-t)\tilde{\mu} \rangle - (t+1-t) \int_{\Omega} \frac{\mathbf{E}^* : \mathbf{E}^*}{4\eta_r} dx \right) \\ &= \sup_{\substack{\mathbf{E}^* \in \mathcal{C}_0(\Omega; \mathbb{R}^{d \times d}) \\ |\mathbf{E}^*| \leq \sqrt{2}\tau_y \text{ a.e.}}} \left(t \langle \mathbf{E}^*, \mu \rangle - t \int_{\Omega} \frac{\mathbf{E}^* : \mathbf{E}^*}{4\eta_r} dx \right. \\ & \quad \left. + (1-t) \langle \mathbf{E}^*, \tilde{\mu} \rangle - (1-t) \int_{\Omega} \frac{\mathbf{E}^* : \mathbf{E}^*}{4\eta_r} dx \right) \\ &\leq t \sup_{\substack{\mathbf{E}^* \in \mathcal{C}_0(\Omega; \mathbb{R}^{d \times d}) \\ |\mathbf{E}^*| \leq \sqrt{2}\tau_y \text{ a.e.}}} \left(\langle \mathbf{E}^*, \mu \rangle - \int_{\Omega} \frac{\mathbf{E}^* : \mathbf{E}^*}{4\eta_r} dx \right) \\ & \quad + (1-t) \sup_{\substack{\mathbf{E}^* \in \mathcal{C}_0(\Omega; \mathbb{R}^{d \times d}) \\ |\mathbf{E}^*| \leq \sqrt{2}\tau_y \text{ a.e.}}} \left(\langle \mathbf{E}^*, \tilde{\mu} \rangle - \int_{\Omega} \frac{\mathbf{E}^* : \mathbf{E}^*}{4\eta_r} dx \right) \\ &= t \tilde{\mathcal{G}}(\mu) + (1-t) \tilde{\mathcal{G}}(\tilde{\mu}). \end{aligned} \quad (3.98)$$

Since the strain rate tensor $\dot{\varepsilon}: BD(\Omega) \rightarrow \mathcal{M}(\Omega; \mathbb{R}^{d \times d})$ is linear, this also proves the convexity of $\tilde{\mathcal{G}} \circ \dot{\varepsilon}$. \square

3.3.3 Existence of Solutions

Through the definition of $\tilde{\mathcal{G}}$, we are able to generalize the objective function Φ_0 and define it for arbitrary functions in $BD(\Omega)$. Consequently, we seek minimizers of

$$\Phi_0: U_0 \longrightarrow \mathbb{R}, \quad \mathbf{v} \mapsto \tilde{\mathcal{G}}(\dot{\varepsilon}(\mathbf{v})) - \int_{\Omega} \mathbf{f} \cdot \mathbf{v} dx \quad (3.99)$$

for a given $\mathbf{f} \in U_0'$. To prove existence of minimizers, we first show validity of Korn's inequality on U_0 .

Proposition 3.13. *Under the assumptions on Ω from Section 3.1, there is a constant $C_K > 0$ such that*

$$C_K \|\mathbf{u}\|_{BD} \leq \|\dot{\boldsymbol{\varepsilon}}(\mathbf{u})\|_{\mathcal{M}} \leq \|\mathbf{u}\|_{BD} \quad (3.100)$$

for all $\mathbf{u} \in U_0$.

PROOF. The upper bound in (3.100) follows immediately from the definition of the norm on $BD(\Omega)$. Recall that

$$R := \left\{ \mathbf{r}: \Omega \rightarrow \mathbb{R}^d \mid \mathbf{r}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{b} \in \mathbb{R}^d, \mathbf{A} \in \mathbb{R}^{d \times d} \text{ with } \mathbf{A} = -\mathbf{A}^\top \right\} \quad (3.101)$$

is the nullspace of $\dot{\boldsymbol{\varepsilon}}$ and consists of rigid displacements. Due to Kohn (1979, Part II, Proposition 3.11), for any continuous and linear map $T: BD(\Omega) \rightarrow R$ with $T|_R = I|_R$, where I denotes the identity, there is a constant $C > 0$ such that

$$C \|\mathbf{u} - T(\mathbf{u})\|_{L^{\frac{d}{d-1}}(\Omega)} \leq \|\dot{\boldsymbol{\varepsilon}}(\mathbf{u})\|_{\mathcal{M}} \quad (3.102)$$

for all $\mathbf{u} \in BD(\Omega)$. If we find such a map T that vanishes on U_0 , the statement follows from the continuous inclusion $L^{\frac{d}{d-1}}(\Omega)^d \subset L^1(\Omega)^d$ and by adding $\|\dot{\boldsymbol{\varepsilon}}(\mathbf{u})\|_{\mathcal{M}}$ to both sides of the inequality.

First note that R is a closed subspace of $BD(\Omega)$: Let $(\mathbf{r}_n)_n \subset R$ and $\mathbf{u} \in BD(\Omega)$ such that

$$\mathbf{r}_n \xrightarrow{n \rightarrow \infty} \mathbf{u} \quad \text{in } BD(\Omega). \quad (3.103)$$

For arbitrary $\varepsilon > 0$, there is $N \in \mathbb{N}$ such that

$$\varepsilon > \|\mathbf{r}_n - \mathbf{u}\|_{BD} \geq \|\dot{\boldsymbol{\varepsilon}}(\mathbf{r}_n - \mathbf{u})\|_{\mathcal{M}} = \|\dot{\boldsymbol{\varepsilon}}(\mathbf{u})\|_{\mathcal{M}} \geq 0 \quad (3.104)$$

for all $n \geq N$, implying $\dot{\boldsymbol{\varepsilon}}(\mathbf{u}) = \mathbf{0}$ and $\mathbf{u} \in \mathcal{N}(\dot{\boldsymbol{\varepsilon}}) = R$.

Furthermore, $U_0 \cap R = \{\mathbf{0}\}$: For arbitrary $\mathbf{r} \in U_0 \cap R$, $\gamma_\nu(\mathbf{r}) = 0$ by definition of the space U_0 , and thus

$$0 = \mathbf{r}(\mathbf{x}) \cdot \boldsymbol{\nu}(\mathbf{x}) = (\mathbf{A}\mathbf{x} + \mathbf{b}) \cdot \boldsymbol{\nu}(\mathbf{x}) \quad (3.105)$$

holds for almost every $\mathbf{x} \in \partial\Omega$ (see Theorem 3.8). From here on, one can proceed analogously to the proof of Lemma 2.3 to see that $\mathbf{r} \equiv \mathbf{0}$.

Since U_0 is a closed subspace of $BD(\Omega)$ due to Proposition 3.9 and R is a finite-dimensional closed subspace of $BD(\Omega)$ with $U_0 \cap R = \{\mathbf{0}\}$, there is a continuous linear projection $T: BD(\Omega) \rightarrow R$ with $T|_R = I|_R$ and $U_0 \subset \mathcal{N}(T)$ (e.g. Alt, 2016, Theorem 9.16), which concludes the proof. \square

Note that the constant C_K from Proposition 3.13 is in general different from the constant from Korn's inequality on V_0 . However, we use the same notation as there will be no confusion due to the context.

With these preparations, we are now able to prove existence of a minimizer of Φ_0 in U_0 under a smallness assumption on the right-hand side forcing term \mathbf{f} .

Theorem 3.14. *Let $r := d/(d-1)$ and denote its Hölder conjugate by $r' := r/(r-1)$. Assume that the volumetric force $\mathbf{f} \in L^{r'}(\Omega)^d \subset U'_0$ satisfies $\|\mathbf{f}\|_{U'_0} < \sqrt{2}\tau_{y,\min}C_K$, where C_K is the constant from Korn's inequality (3.100). There is at least one minimizer $\mathbf{u} \in U_0$ of the functional Φ_0 defined in (3.99). Furthermore, any solution satisfies the stability estimate*

$$\|\mathbf{u}\|_{BD} \leq \left(\sqrt{2}\tau_{y,\min}C_K - \|\mathbf{f}\|_{U'_0} \right)^{-1} \left\| \tau_y^2/2\eta_r \right\|_{L^1}. \quad (3.106)$$

PROOF. Let $(\mathbf{u}_n)_n \subset U_0$ be an extremal sequence, i.e.,

$$\Phi_0(\mathbf{u}_n) \xrightarrow{n \rightarrow \infty} \inf_{\mathbf{v} \in U_0} \Phi_0(\mathbf{v}). \quad (3.107)$$

Using Korn's inequality (Proposition 3.13) and the properties of the functions the supremum is taken over, for arbitrary $\mathbf{v} \in BD(\Omega)$ we obtain:

$$\begin{aligned} \Phi_0(\mathbf{v}) &= \sup_{\substack{\mathbf{E}^* \in \mathcal{C}_0(\Omega; \mathbb{R}^{d \times d}) \\ |\mathbf{E}^*| \leq \sqrt{2}\tau_y \text{ a.e.}}} \left(\langle \mathbf{E}^*, \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \rangle - \int_{\Omega} \frac{\mathbf{E}^* : \mathbf{E}^*}{4\eta_r} dx \right) - \langle \mathbf{f}, \mathbf{v} \rangle \\ &\geq \sup_{\substack{\mathbf{E}^* \in \mathcal{C}_0(\Omega; \mathbb{R}^{d \times d}) \\ |\mathbf{E}^*| \leq \sqrt{2}\tau_y \text{ a.e.}}} \left(\langle \mathbf{E}^*, \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \rangle \right) - \left\| \tau_y^2/2\eta_r \right\|_{L^1} - \|\mathbf{f}\|_{U'_0} \|\mathbf{v}\|_{BD} \\ &\geq \sup_{\substack{\mathbf{E}^* \in \mathcal{C}_0(\Omega; \mathbb{R}^{d \times d}) \\ |\mathbf{E}^*| \leq \sqrt{2}\tau_{y,\min} \text{ a.e.}}} \left(\langle \mathbf{E}^*, \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \rangle \right) - \left\| \tau_y^2/2\eta_r \right\|_{L^1} - \|\mathbf{f}\|_{U'_0} \|\mathbf{v}\|_{BD} \quad (3.108) \\ &= \sqrt{2}\tau_{y,\min} \|\dot{\boldsymbol{\varepsilon}}(\mathbf{v})\|_{\mathcal{M}} - \left\| \tau_y^2/2\eta_r \right\|_{L^1} - \|\mathbf{f}\|_{U'_0} \|\mathbf{v}\|_{BD} \\ &\geq \sqrt{2}\tau_{y,\min}C_K \|\mathbf{v}\|_{BD} - \left\| \tau_y^2/2\eta_r \right\|_{L^1} - \|\mathbf{f}\|_{U'_0} \|\mathbf{v}\|_{BD} \\ &= \|\mathbf{v}\|_{BD} \left(\sqrt{2}\tau_{y,\min}C_K - \|\mathbf{f}\|_{U'_0} \right) - \left\| \tau_y^2/2\eta_r \right\|_{L^1}, \end{aligned}$$

and thus $\Phi_0(\mathbf{v}) \rightarrow \infty$ for $\|\mathbf{v}\|_{BD} \rightarrow \infty$ since $\|\mathbf{f}\|_{U'_0} < \sqrt{2}\tau_{y,\min}C_K$. Hence, the sequence $(\mathbf{u}_n)_n$ is bounded and we can extract a subsequence, without loss of generality using the same indices, such that there is $\mathbf{u} \in U_0$ with

$$\mathbf{u}_n \xrightarrow{n \rightarrow \infty} \mathbf{u} \quad \text{in } L^r(\Omega)^d, \quad (3.109a)$$

$$\dot{\boldsymbol{\varepsilon}}(\mathbf{u}_n) \xrightarrow{n \rightarrow \infty} \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) \quad \text{in } \mathcal{M}(\Omega; \mathbb{R}^{d \times d}). \quad (3.109b)$$

To prove lower semi-continuity of Φ_0 with respect to this topology, we note that (3.109a) implies

$$\langle \mathbf{f}, \mathbf{u}_n \rangle \xrightarrow{n \rightarrow \infty} \langle \mathbf{f}, \mathbf{u} \rangle \quad (3.110)$$

due to $\mathbf{f} \in L^{r'}(\Omega)^d$, and thus it remains to show

$$\liminf_{n \rightarrow \infty} \tilde{\mathcal{G}}(\dot{\boldsymbol{\varepsilon}}(\mathbf{u}_n)) \geq \tilde{\mathcal{G}}(\dot{\boldsymbol{\varepsilon}}(\mathbf{u})). \quad (3.111)$$

Let $\varepsilon > 0$ be arbitrary. Since $\tilde{\mathcal{G}}$ is defined as the supremum over such functions, there is $\mathbf{E}_\varepsilon^* \in \mathcal{C}_0(\Omega; \mathbb{R}^{d \times d})$ with $|\mathbf{E}_\varepsilon^*| \leq \sqrt{2}\tau_y$ almost everywhere such that

$$\tilde{\mathcal{G}}(\dot{\boldsymbol{\varepsilon}}(\mathbf{u})) - \varepsilon \leq \langle \mathbf{E}_\varepsilon^*, \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) \rangle - \int_{\Omega} \frac{\mathbf{E}_\varepsilon^* : \mathbf{E}_\varepsilon^*}{4\eta_r} dx. \quad (3.112)$$

Due to (3.109b),

$$\begin{aligned}
\tilde{\mathcal{G}}(\dot{\mathbf{u}}) - \varepsilon &\leq \lim_{n \rightarrow \infty} \langle \mathbf{E}_\varepsilon^*, \dot{\mathbf{u}}_n \rangle - \int_{\Omega} \frac{\mathbf{E}_\varepsilon^* : \mathbf{E}_\varepsilon^*}{4\eta_r} d\mathbf{x} \\
&= \liminf_{n \rightarrow \infty} \left(\langle \mathbf{E}_\varepsilon^*, \dot{\mathbf{u}}_n \rangle - \int_{\Omega} \frac{\mathbf{E}_\varepsilon^* : \mathbf{E}_\varepsilon^*}{4\eta_r} d\mathbf{x} \right) \\
&\leq \liminf_{n \rightarrow \infty} \sup_{\substack{\mathbf{E}^* \in \mathcal{C}_0(\Omega; \mathbb{R}^{d \times d}) \\ |\mathbf{E}^*| \leq \sqrt{2}\tau_y \text{ a.e.}}} \left(\langle \mathbf{E}^*, \dot{\mathbf{u}}_n \rangle - \int_{\Omega} \frac{\mathbf{E}^* : \mathbf{E}^*}{4\eta_r} d\mathbf{x} \right) \\
&= \liminf_{n \rightarrow \infty} \tilde{\mathcal{G}}(\dot{\mathbf{u}}_n).
\end{aligned} \tag{3.113}$$

Since $\varepsilon > 0$ was arbitrary, (3.111) holds and we obtain

$$\inf_{\mathbf{v} \in U_0} \Phi_0(\mathbf{v}) \leq \Phi_0(\mathbf{u}) \leq \liminf_{n \rightarrow \infty} \Phi_0(\mathbf{u}_n) = \inf_{\mathbf{v} \in U_0} \Phi_0(\mathbf{v}), \tag{3.114}$$

yielding $\mathbf{u} \in \arg \min_{\mathbf{v} \in U_0} \Phi_0(\mathbf{v})$.

Analogously to the previous section, the stability estimate follows from (3.108) and the minimizing property of \mathbf{u} :

$$0 = \Phi_0(\mathbf{0}) \geq \Phi_0(\mathbf{u}) \geq \|\mathbf{u}\|_{BD} \left(\sqrt{2}\tau_{y,\min} C_K - \|\mathbf{f}\|_{U'_0} \right) - \|\tau_y^2/2\eta_r\|_{L^1}. \tag{3.115}$$

Rearranging yields (3.106). \square

Remark 3.15. Note that Lemma 3.12 only implies convexity of Φ_0 , but not strict convexity. Hence, uniqueness of a solution cannot be guaranteed. Formally, this is expected in the purely plastic limit: Assume there is a solution \mathbf{u} that lies in the plastic regime everywhere in Ω , and assume additionally that the solution is sufficiently regular such that it satisfies the first-order necessary conditions. In this case, these conditions read

$$\begin{aligned}
\int_{\Omega} \mathbf{f} \cdot \mathbf{v} d\mathbf{x} &= \int_{\Omega} \min \left(2\eta_r, \frac{\sqrt{2}\tau_y}{|\dot{\mathbf{e}}(\mathbf{u})|} \right) \dot{\mathbf{e}}(\mathbf{u}) : \dot{\mathbf{e}}(\mathbf{v}) d\mathbf{x} \\
&= \int_{\Omega} \sqrt{2}\tau_y \frac{\dot{\mathbf{e}}(\mathbf{u})}{|\dot{\mathbf{e}}(\mathbf{u})|} : \dot{\mathbf{e}}(\mathbf{v}) d\mathbf{x}
\end{aligned} \tag{3.116}$$

for all \mathbf{v} (cf. (3.11)). For arbitrary constants $c > 1$, $c\mathbf{u}$ still lies in the plastic regime everywhere in Ω and therefore $c\mathbf{u}$ also satisfies (3.116) since the constant just cancels out in the integrand.

3.4 Conclusion and Outlook

We have shown well-posedness of Stokes systems with a regularized visco-plastic rheology that is widely used in geodynamic applications. For the unregularized case, we have extended the energy functional to the space of functions of bounded deformations and have shown existence of minimizers in a subspace of divergence-free velocity fields. Additionally, we have provided stability estimates for solutions. However, uniqueness of minimizers cannot be guaranteed in general for $\underline{\eta} = 0$.

The visco-plastic rheology presented here uses a linear reference viscosity in the viscous regime. It is possible to generalize this model by replacing the reference viscosity with the power-law viscosity from Chapter 2 such that both viscous and plastic deformations are nonlinear. The ideas from this chapter can be combined with the ones from the previous chapter to argue well-posedness for $\underline{\eta} > 0$ and existence of solutions for the unregularized case. Similarly, different models for plasticity can be used, e.g., pressure-dependent ones like Drucker-Prager plasticity, however, the approaches employed here do not directly carry over to these models (cf. Lanzendörfer, 2011).

The regularized and unregularized problems of Sections 3.2 and 3.3 are connected through the fact that the energy functionals coincide if the strain rate tensor lies in $L^1(\Omega)^{d \times d}$. An interesting extension of our work would be to study the behavior of the regularized solution when $\underline{\eta}$ tends to zero. In particular, given a null sequence of regularization parameters, one could investigate numerically and theoretically whether the corresponding solutions contain a subsequence that converges to a solution of the unregularized problem with respect to some topology in $BD(\Omega)$. Potential issues can be expected for $\underline{\eta} \rightarrow 0$ since unique solutions cannot be guaranteed for $\underline{\eta} = 0$.

4 CHAPTER

Adjoint-based Inversion of Principal Stress Directions

After discussing well-posedness of the Stokes system with different rheologies in Chapters 1 to 3, we are now ready to approach selected computational inverse problems for the identification of geometric structures and rheological characteristics inside the Earth in Chapters 4 and 5. During the course of this PhD project, the work presented in the current chapter has already been published in a peer-reviewed journal (Reuber et al., 2020).

4.1 Motivation

Geodynamic inverse methods aim to reproduce the shape (e.g., Worthen et al., 2014) and the material parameters (e.g., Reuber et al., 2018) of subsurface structures based on physical observations, such as plate velocities or principal stress directions, measured on or close to the Earth's surface. These measurements are usually expensive to obtain and cover the area of the Earth only sparsely. Additionally, considering the slow movement of plates, the observations only represent instantaneous snapshots. Inferring subsurface structures and material parameters merely from surface data is thus a significant challenge.

It is therefore beneficial to investigate the computational effort and the effect of incorporating as many of the available observation types as possible into the inverse problem. Performing a direct grid-based search to minimize the data misfit (e.g. Baumann, Kaus, and Popov, 2014) is a straightforward method since the underlying equation systems do not change. However, the number of required forward simulations scales in a nonlinear manner with the number of inversion parameters, which makes the approach impractical for a large amount of unknown parameters. Gradient-based methods can be more efficient, particularly if the initial guess is close to the true solution and when combined with an adjoint approach to compute the gradient of the objective function with respect to the material parameters. The adjoint method is widely used in seismology (e.g. Tarantola, 2005; Tromp, Tape, and Liu, 2005; Fichtner, 2011) and has previously been applied to global scale mantle convection problems to fit global plate velocities (e.g. Ismail-Zadeh et al., 2003; Horbach, Bunge, and Oeser, 2014; Ratnaswamy, Stadler, and Gurnis, 2015). Yet, there have been some unresolved issues when incorporating other (derived) observation types such as stresses in an adjoint-based inversion.

In this work, we present an adjoint-based framework for efficient computation of

gradients with principal stress directions as data, which might come from measurements in boreholes or focal mechanism data determined from earthquakes. We only consider simplified problems in the numerical computations to build the basis for a future use of this method in combination with other data types in larger frameworks and real-world applications.

4.2 Governing Equations

4.2.1 Rheological Model

Since the inverse problems we address in this chapter occur in the upper mantle and lithosphere, the slow deformation of rocks can be modeled as a highly viscous shear-thinning fluid. Therefore, we consider a power-law form for the effective viscosity:

$$\eta(\mathbf{u}) = \eta_0 \left(\frac{\dot{\epsilon}_{\text{II}}(\mathbf{u})}{\dot{\epsilon}_0} \right)^{\frac{1}{n}-1}. \quad (4.1)$$

Here, η_0 is the reference viscosity, $\dot{\epsilon}_0$ is the reference strain rate, $n \geq 1$ is the power-law exponent, and $\dot{\epsilon}_{\text{II}}(\mathbf{u}) = \sqrt{\frac{1}{2}\dot{\boldsymbol{\epsilon}}(\mathbf{u}) : \dot{\boldsymbol{\epsilon}}(\mathbf{u})}$ is the effective strain rate (square root of the second invariant of the strain rate tensor). Note that the power-law viscosity (4.1) is a special case of the rheological model discussed in Chapter 2. Since we consider two different ways of obtaining adjoint-based gradients, we briefly state both the strong and weak form of the forward problem in the next section.

4.2.2 Well-Posedness of the Forward Problem

As in Chapter 2, we choose the domain $\Omega \subset \mathbb{R}^3$ to be bounded with polyhedral boundary $\partial\Omega$. Assuming that gravity is the only body force applied to the fluid, we obtain the Stokes system corresponding to the power-law viscosity (4.1) as follows:

$$-\operatorname{div} \boldsymbol{\sigma} = \rho \mathbf{g}, \quad (4.2)$$

$$\operatorname{div} \mathbf{u} = 0, \quad (4.3)$$

$$\boldsymbol{\sigma}(\mathbf{u}, p) = \boldsymbol{\tau}(\mathbf{u}) - p \mathbf{I}, \quad (4.4)$$

where $\mathbf{u} = (u_x, u_y, u_z)^\top$ is the velocity vector, $\boldsymbol{\sigma}$ is the Cauchy stress tensor, ρ is the density, and $\mathbf{g} = g(0, 0, -1)^\top \in \mathbb{R}^3$ is the gravity acceleration vector with $g \in \mathbb{R}_+$. Furthermore,

$$\boldsymbol{\tau}(\mathbf{u}) = \begin{pmatrix} \tau_{xx} & \tau_{xy} & \tau_{xz} \\ \tau_{xy} & \tau_{yy} & \tau_{yz} \\ \tau_{xz} & \tau_{yz} & \tau_{zz} \end{pmatrix} = 2\eta(\mathbf{u}) \begin{pmatrix} \dot{\epsilon}_{xx} & \dot{\epsilon}_{xy} & \dot{\epsilon}_{xz} \\ \dot{\epsilon}_{xy} & \dot{\epsilon}_{yy} & \dot{\epsilon}_{yz} \\ \dot{\epsilon}_{xz} & \dot{\epsilon}_{yz} & \dot{\epsilon}_{zz} \end{pmatrix} = 2\eta(\mathbf{u})\dot{\boldsymbol{\epsilon}}(\mathbf{u}) \quad (4.5)$$

is the viscous (or deviatoric) stress tensor, p is the pressure (positive in compression), and \mathbf{I} is the second-order identity tensor. Recall that we impose a homogeneous normal flow and free slip boundary condition on $\partial\Omega$:

$$\mathbf{u} \cdot \boldsymbol{\nu} = 0, \quad (4.6)$$

$$\mathbf{T}(\boldsymbol{\sigma}\boldsymbol{\nu}) = \mathbf{0}, \quad (4.7)$$

4.3 Inverse Problem Using Principal Stress Directions

where $\boldsymbol{\nu}$ is the unit-length outward normal vector and $\mathbf{T} := \mathbf{I} - \boldsymbol{\nu}\boldsymbol{\nu}^\top$ is the tangential projection.

For given $n \geq 1$, let $r = \left(\frac{1}{n} + 1\right) \in (1, 2]$ and denote its Hölder conjugate by $r' = (1 + n) \in [2, \infty)$. The corresponding weak form of the Stokes equations reads as follows: Given $\rho \in L^{r'}(\Omega)$, $\eta_0 \in L_+^\infty(\Omega)$, and $\dot{\epsilon}_0 \in \mathbb{R}_+$, find $(\mathbf{u}, p) \in V \times Q$ such that

$$\int_{\Omega} 2\eta(\mathbf{u}) \dot{\epsilon}(\mathbf{u}) : \dot{\epsilon}(\mathbf{v}) \, d\mathbf{x} - \int_{\Omega} p \operatorname{div} \mathbf{v} \, d\mathbf{x} - \int_{\Omega} q \operatorname{div} \mathbf{u} \, d\mathbf{x} - \int_{\Omega} \rho \mathbf{g} \cdot \mathbf{v} \, d\mathbf{x} = 0 \quad (4.8)$$

holds for all $(\mathbf{v}, q) \in V \times Q$, where

$$V := \left\{ \mathbf{u} \in W^{1,r}(\Omega)^d \mid \mathbf{u} \cdot \boldsymbol{\nu} = 0 \text{ on } \partial\Omega \right\}, \quad (4.9)$$

$$Q := L_0^{r'}(\Omega) := \left\{ p \in L^{r'}(\Omega) \mid \int_{\Omega} p \, d\mathbf{x} = 0 \right\}. \quad (4.10)$$

As seen in Chapter 2, the weak form (4.8) of the Stokes system is well-posed. In the following, we approach computational inverse problems governed by the above equations. For this purpose, we focus on the formal development of an adjoint-based inversion method and investigate the framework computationally rather than from a theoretical point of view.

4.3 Inverse Problem Using Principal Stress Directions

Although the effective physical and rheological parameters of the lithosphere and mantle (e.g., ρ , η_0 , n) cannot generally be measured directly, they can still be estimated by solving an inverse problem. Here, we take the approach of formulating the geodynamic inverse problem as a constrained optimization problem (see e.g., Ismail-Zadeh et al., 2003; Ratnaswamy, Stadler, and Gurnis, 2015; Reuber, Popov, and Kaus, 2017), where the difference between the observed data and simulation results is to be minimized under the constraint that the Stokes equations are satisfied. The solution to such an inverse problem is one parameter vector and does not include information about uncertainties, which can be particularly useful if there are many parameter vectors that fit the data similarly well. In Chapter 5, we will instead employ the Bayesian approach to inverse problems (e.g. Kaipio and Somersalo, 2005; Stuart, 2010), where the solution describes the probabilities of different parameter vectors given the data and some prior probability distribution for the material parameters. However, the main purpose of the current chapter is to introduce principal stress directions as observations in a geodynamic inversion and to provide a framework for efficient computation of gradients of the objective function, enabling the future use of this data type and the corresponding gradients, alone or in combination with other data types, both in deterministic and Bayesian inversions.

A common type of observations used in geodynamic inversions is the surface plate velocity field. However, there is an important pool of data that has, thus far, not been incorporated in geodynamic inversions, namely stresses. In practice, the exact entries of the stress tensor are difficult to measure. Nonetheless, it is possible to acquire measurements of the *principal stress directions (PSD)*, which are related to the eigenvectors of the viscous stress tensor and can be obtained from either borehole

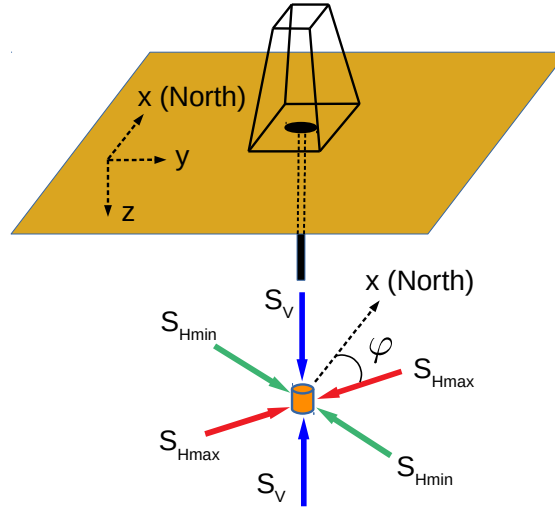


Figure 4.1 Principal stress orientation around a borehole according to Anderson's assumption. One of the principal stresses is assumed to be vertical (S_v), and the other two are horizontal (S_{Hmin} and S_{Hmax}). The angle φ describes the direction of maximum horizontal compressive stress (S_{Hmax}) measured from the North.

breakouts or seismic focal mechanism inversions (see e.g. Zoback, 2010). A typical assumption that is made about the orientation of principal stresses in the lithosphere, i.e., the eigenvectors of the viscous stress tensor, is a so-called *Anderson's model* illustrated in Figure 4.1. One of the principal stresses is assumed to be vertical (S_v), and the other two are horizontal (S_{Hmin} and S_{Hmax}). The magnitudes of the principal stresses, i.e., the eigenvalues of the viscous stress tensor, are rarely available. The only parameter that can be used as a constraint for the geodynamic inversion is the orientation of the maximum horizontal compressive stress (S_{Hmax}), denoted as φ in Figure 4.1. Note that S_{Hmax} is the eigenvector corresponding to the smaller eigenvalue of the projection of the viscous stress tensor to the xy -plane.

Assuming Anderson's model, the orientation angle of the maximum horizontal compressive stress can be computed using three components of the deviatoric stress tensor (e.g. Gere, 2004):

$$\varphi := \varphi(\eta, \mathbf{u}) := \frac{1}{2} \tan^{-1} \left(\frac{2\tau_{xy}}{\tau_{xx} - \tau_{yy}} \right). \quad (4.11)$$

For the power-law viscous rheology with isotropic constitutive properties employed in this study, it is possible to express the orientation angle in terms of the strain rate tensor since it has the same unit direction, i.e.,

$$\varphi := \varphi(\mathbf{u}) := \frac{1}{2} \tan^{-1} \left(\frac{2\dot{\epsilon}_{xy}}{\dot{\epsilon}_{xx} - \dot{\epsilon}_{yy}} \right). \quad (4.12)$$

Note that for an elastic rather than a viscous rheology, the stress history becomes important and the deviatoric stress tensor deviates from the unit direction of the strain rate.

4.3 Inverse Problem Using Principal Stress Directions

In the following, we extend the geodynamic inversion framework to incorporate PSD as data. We formulate the inverse problem as a regularized infinite-dimensional PDE-constrained nonlinear least-squares optimization problem:

$$\min_{\mathbf{u}, \mathbf{m}} \left(\frac{1}{2} \int_{\Omega_{\text{obs}}} [\varphi(\mathbf{u}(\mathbf{m})) - \varphi_{\text{obs}}]^2 d\mathbf{x} + \frac{\lambda}{2} \int_{\Omega} (\mathbf{m} - \mathbf{m}_0) \cdot \mathcal{W}(\mathbf{m} - \mathbf{m}_0) d\mathbf{x} \right), \quad (4.13)$$

where φ_{obs} denotes the stress directions measured on $\Omega_{\text{obs}} \subseteq \Omega$ and $\varphi(\mathbf{u}(\mathbf{m}))$ is given by (4.12) through the solution of the nonlinear forward Stokes problem (4.8) corresponding to $\mathbf{m} = (\rho, \eta_0, n)^\top$. Furthermore, the second term in (4.13) serves as a regularization to cope with the ill-posedness of the inverse problem.¹ Here, \mathcal{W} is a self-adjoint² positive definite weighting operator, \mathbf{m}_0 is a guess for the true parameter, and $\lambda \geq 0$ controls the influence of the penalty term in the objective function. In the simplest case, $\mathbf{m}_0 = \mathbf{0}$ and \mathcal{W} is the identity such that one obtains a standard L^2 Tikhonov regularization. For more information about regularizing inverse problems, we refer the reader to Engl, Hanke, and Neubauer (1996). Note that the material parameters ρ , η_0 , and n do not need to be treated as constants but can be considered as spatially varying functions $\mathbf{m}(\mathbf{x}) = (\rho(\mathbf{x}), \eta_0(\mathbf{x}), n(\mathbf{x}))^\top$.

To solve the minimization problem (4.13), we employ a quasi-Newton method that requires computation of the gradient $\nabla \mathcal{J}(\mathbf{m})$ of the objective function

$$\mathcal{J}(\mathbf{m}) := \frac{1}{2} \int_{\Omega_{\text{obs}}} [\varphi(\mathbf{u}(\mathbf{m})) - \varphi_{\text{obs}}]^2 d\mathbf{x} + \frac{\lambda}{2} \int_{\Omega} (\mathbf{m} - \mathbf{m}_0) \cdot \mathcal{W}(\mathbf{m} - \mathbf{m}_0) d\mathbf{x}, \quad (4.14)$$

which we approach in the next section.

Remark 4.1. The definition of φ in (4.11) and (4.12) is imprecise. In fact, it is only guaranteed that φ is the angle corresponding to $S_{H_{\text{min}}}$ or $S_{H_{\text{max}}}$. However, since $S_{H_{\text{min}}}$ and $S_{H_{\text{max}}}$ are eigenvectors of the projection of the viscous stress tensor $\boldsymbol{\tau}$ onto the xy -plane and this projected matrix is symmetric, the eigenvectors are orthogonal and the sought angle is thus either φ or $\varphi + \pi/2$. The measured angle belongs to the smaller eigenvalue of the projected viscous stress tensor and can therefore be determined numerically through an eigendecomposition. As long as the two eigenvalues do not coincide, the principal stress direction is differentiable since eigenvalues depend in a differentiable way on the entries of a matrix. Furthermore, the derivatives are equal since the angles are only shifted by $\pi/2$. Nevertheless, if the eigenvalues coincide, the angle might jump between φ and $\varphi + \pi/2$ when the entries of the matrix are perturbed and the principal stress direction is not differentiable. Hence, we assume the eigenvalues to be distinct inside the measurement domain Ω_{obs} such that the derivative-based inversion proposed in the next section is well-defined. For real-world applications, this is not a restrictive assumption: The (projected) stress tensor is a multiple of the identity if the two eigenvalues coincide, and the corresponding compressional stresses are thus equal in x - and y -direction. While this scenario can theoretically occur at isolated points inside the Earth, it is very unlikely to be observed and would additionally be apparent from the experimental results.

¹Note that this refers to regularization of inverse problems as explained in Chapter 1 and is different from the viscosity regularization we used in Chapter 3.

²Note that we implicitly assume that \mathcal{W} is linear and bounded.

4.3.1 Adjoint Equations and Adjoint-based Gradients

The adjoint method is an efficient way to compute gradients for optimization problems with partial differential equations as constraints. To derive these adjoint-based gradients, we will use the *formal Lagrangian approach* (see Bertsekas, 1982; Tröltzsch, 2010). This requires the definition of the so-called Lagrangian functional and the computation of its derivatives. The Lagrangian functional is obtained by adding the objective function to the weak form of the forward problem:

$$\begin{aligned}
 \mathcal{L}(\mathbf{u}, p; \mathbf{v}, q; \mathbf{m}) := & \frac{1}{2} \int_{\Omega_{\text{obs}}} (\varphi(\mathbf{u}) - \varphi_{\text{obs}})^2 d\mathbf{x} \\
 & + \frac{\lambda}{2} \int_{\Omega} (\mathbf{m} - \mathbf{m}_0) \cdot \mathcal{W}(\mathbf{m} - \mathbf{m}_0) d\mathbf{x} \\
 & + \int_{\Omega} 2\eta(\mathbf{u}, \mathbf{m}) \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) : \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) d\mathbf{x} \\
 & - \int_{\Omega} p \operatorname{div} \mathbf{v} d\mathbf{x} \\
 & - \int_{\Omega} q \operatorname{div} \mathbf{u} d\mathbf{x} \\
 & - \int_{\Omega} \rho \mathbf{g} \cdot \mathbf{v} d\mathbf{x}.
 \end{aligned} \tag{4.15}$$

Here, the test functions \mathbf{v} and q take the role of Lagrange multipliers for the PDE constraint, which turn out to be the adjoint velocity and adjoint pressure, respectively (see (4.17) below). Since the PDE constraint is enforced by the Lagrange multipliers in the Lagrangian approach, the forward problem solution \mathbf{u} and p , the adjoint variables \mathbf{v} and q , and the parameter vector \mathbf{m} are treated as independent variables. Minimizing (4.15) over all five variables is thus formally an unconstrained optimization problem that is equivalent to the constrained optimization problem (4.13). Therefore, minimizers of (4.13) are characterized by the first-order necessary conditions for (4.15), which we derive next.

Taking the directional derivative of \mathcal{L} with respect to the variables \mathbf{v} and q , and requiring it to vanish in the direction of all admissible test functions $\hat{\mathbf{v}}$ and \hat{q} , yields

$$\mathcal{L}_{(v,q)}(\mathbf{u}, p; \mathbf{v}, q; \mathbf{m})(\hat{\mathbf{v}}, \hat{q}) \stackrel{!}{=} 0. \tag{4.16}$$

As mentioned above, one simply recovers the weak form of the nonlinear Stokes equations (4.8) here, which is used to obtain the forward solution \mathbf{u} and p .

Similarly, setting the directional derivatives of \mathcal{L} with respect to the forward solution variables \mathbf{u} and p in the direction of all admissible test functions $\hat{\mathbf{u}}$ and \hat{p} to

zero, we obtain

$$\begin{aligned}
 0 &\stackrel{!}{=} \mathcal{L}_{(\mathbf{u}, p)}(\mathbf{u}, p; \mathbf{v}, q; \mathbf{m})(\hat{\mathbf{u}}, \hat{p}) \\
 &= \int_{\Omega_{\text{obs}}} \varphi'(\mathbf{u}) \hat{\mathbf{u}} (\varphi(\mathbf{u}) - \varphi_{\text{obs}}) d\mathbf{x} \\
 &\quad + \int_{\Omega} (\mathbb{D}(\mathbf{u}, \mathbf{m}) \dot{\varepsilon}(\hat{\mathbf{u}})) : \dot{\varepsilon}(\mathbf{v}) d\mathbf{x} \\
 &\quad - \int_{\Omega} \hat{p} \operatorname{div} \mathbf{v} d\mathbf{x} \\
 &\quad - \int_{\Omega} q \operatorname{div} \hat{\mathbf{u}} d\mathbf{x}.
 \end{aligned} \tag{4.17}$$

Here,

$$\mathbb{D}(\mathbf{u}, \mathbf{m}) = 2\eta(\mathbf{u}, \mathbf{m}) \left(\mathbb{I} + \left(\frac{1}{n} - 1 \right) \frac{\dot{\varepsilon}(\mathbf{u}) \otimes \dot{\varepsilon}(\mathbf{u})}{2\dot{\varepsilon}_{\mathbb{I}}(\mathbf{u})^2} \right) \tag{4.18}$$

with the fourth-order identity tensor \mathbb{I} and the outer product \otimes , and

$$\begin{aligned}
 \varphi'(\mathbf{u}) \hat{\mathbf{u}} &= \frac{\hat{\varepsilon}_{xy}(\dot{\varepsilon}_{xx} - \dot{\varepsilon}_{yy}) - \dot{\varepsilon}_{xy}(\hat{\varepsilon}_{xx} - \hat{\varepsilon}_{yy})}{(\dot{\varepsilon}_{xx} - \dot{\varepsilon}_{yy})^2 + 4\dot{\varepsilon}_{xy}^2} \\
 &= \frac{1}{(\dot{\varepsilon}_{xx} - \dot{\varepsilon}_{yy})^2 + 4\dot{\varepsilon}_{xy}^2} \frac{1}{2} (\mathbf{E} + \mathbf{E}^{\top}) : \hat{\varepsilon}
 \end{aligned} \tag{4.19}$$

with

$$\hat{\varepsilon} = \dot{\varepsilon}(\hat{\mathbf{u}}), \tag{4.20}$$

$$\mathbf{E} = \mathbf{E}(\mathbf{u}) = \begin{pmatrix} -\dot{\varepsilon}_{xy} & \dot{\varepsilon}_{xx} & 0 \\ -\dot{\varepsilon}_{yy} & \dot{\varepsilon}_{xy} & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{4.21}$$

Equation (4.17) is referred to as the weak form of the *adjoint equation*³ and has to be solved for \mathbf{v} and q , while \mathbf{u} and p are fixed solutions to the forward problem for given parameters \mathbf{m} . Note that this equation is linear since the fourth-order tensor $\mathbb{D}(\mathbf{u}, \mathbf{m})$ only depends on the forward solution and not on the adjoint variables \mathbf{v} and q .

To complete the first-order necessary conditions, the Lagrangian functional \mathcal{L} is differentiated with respect to each material parameter. The weak form of the gradient of the objective function \mathcal{J} is given through the relation

$$\mathcal{L}_{\mathbf{m}}(\mathbf{u}, p; \mathbf{v}, q; \mathbf{m}) \hat{\mathbf{m}} = \int_{\Omega} \hat{\mathbf{m}} \cdot \nabla \mathcal{J}(\mathbf{m}) d\mathbf{x}, \tag{4.22}$$

which is required to vanish in all test directions $\hat{\mathbf{m}}$ for minimizers of the objective function. Taking derivatives with respect to the parameters ρ , η_0 and n , the strong form of the gradient evaluated at $\mathbf{m} = (\rho, \eta_0, n)^{\top}$ can be obtained from (4.22):

$$\nabla \mathcal{J}(\mathbf{m}) = \begin{pmatrix} -\mathbf{g} \cdot \mathbf{v} \\ \eta_0^{-1} \boldsymbol{\tau}(\mathbf{u}, \mathbf{m}) : \dot{\varepsilon}(\mathbf{v}) \\ n^{-2} [\log(\dot{\varepsilon}_0) - \log(\dot{\varepsilon}_{\mathbb{I}}(\mathbf{u}))] \boldsymbol{\tau}(\mathbf{u}, \mathbf{m}) : \dot{\varepsilon}(\mathbf{v}) \end{pmatrix} + \lambda \mathcal{W}(\mathbf{m} - \mathbf{m}_0). \tag{4.23}$$

³The name originates from the relation between the gradient of the objective function and the adjoint of the Fréchet derivative of the observable. Informally, $\nabla \mathcal{J}(\mathbf{m})$ can be identified with $\varphi'(\mathbf{m})^*(\varphi(\mathbf{m}) - \varphi_{\text{obs}})$.

In this expression, (\mathbf{u}, p) denotes the solution of the forward Stokes equation (4.8) corresponding to the parameters \mathbf{m} , and (\mathbf{v}, q) is the solution of the adjoint equation (4.17) corresponding to (\mathbf{u}, p) and \mathbf{m} . Note that the first row of $\nabla \mathcal{J}(\mathbf{m})$ is the partial derivative of the objective function with respect to ρ , the second row corresponds to η_0 , and the third row to n . The gradient expression (4.23) is defined on function spaces and therefore gives a parameter update in the form of a function that can be evaluated at any point in the domain, independent of the discretization.

In addition to solving the nonlinear forward problem (4.8), it is thus only required to solve one *linear* PDE, i.e., the adjoint system (4.17), and evaluate the expression (4.23) to compute the gradient of the objective function \mathcal{J} . Since the former is significantly more expensive computationally and is required for evaluation of the data misfit in (4.14) anyway, the adjoint method provides gradients at negligible numerical cost. Furthermore, the additional effort of solving the adjoint system is independent of the number of inversion parameters, which makes the approach particularly attractive for large numbers of parameters.

Regardless of the parameter structure, equations (4.8) and (4.17) represent weak forms and can be readily employed for a finite element discretization. Alternatively, the strong form of the adjoint system can be derived and discretized by finite differences. After isolating the test functions in the weak form using Green's first identity, the resulting equations are:

$$\operatorname{div} \mathbf{v} = 0, \quad (4.24)$$

$$-\operatorname{div} \hat{\boldsymbol{\sigma}} = \begin{cases} \operatorname{div} \left(\frac{\varphi(\mathbf{u}) - \varphi_{\text{obs}}}{(\dot{\epsilon}_{xx} - \dot{\epsilon}_{yy})^2 + 4\dot{\epsilon}_{xy}^2} \frac{1}{2} (\mathbf{E} + \mathbf{E}^T) \right) & \text{on } \Omega_{\text{obs}}, \\ \mathbf{0} & \text{else,} \end{cases} \quad (4.25)$$

where

$$\hat{\boldsymbol{\sigma}} = \hat{\boldsymbol{\sigma}}(\mathbf{v}, q) = \mathbb{D}(\mathbf{u}, \mathbf{m}) \dot{\boldsymbol{\epsilon}}(\mathbf{v}) - q \mathbf{I} \quad (4.26)$$

and the boundary conditions are

$$\mathbf{v} \cdot \boldsymbol{\nu} = 0, \quad (4.27)$$

$$\mathbf{T}(\hat{\boldsymbol{\sigma}} \boldsymbol{\nu}) = \mathbf{0}. \quad (4.28)$$

The overall solution scheme employing the strong form of the adjoint equations (4.24) to (4.28) to compute the gradient (4.23) follows exactly the same path as the weak form solution outlined above. Hence, the computational advantages apply here as well.

4.4 Discretization

4.4.1 Forward Problem

We discretize the mass and momentum equations (4.2) and (4.3) using a staggered-grid finite difference scheme (Harlow and Welch, 1965; Gerya and Yuen, 2007) implemented in the 3D Stokes solver LaMEM (Kaus et al., 2016), which can handle visco-elasto-plastic materials. LaMEM has already been used as a forward problem

solution tool in sampling-based inversion (Baumann, Kaus, and Popov, 2014) and gradient-based inversion (Reuber et al., 2018). For a given set of material parameters \mathbf{m} , the discretized nonlinear equations are written in the form of

$$\mathbf{0} = \mathbf{r}(\mathbf{m}, \mathbf{z}(\mathbf{m})) := \begin{pmatrix} \mathbf{K}(\mathbf{m}, \mathbf{z}) & \mathbf{G} \\ \mathbf{D} & \mathbf{0} \end{pmatrix} \mathbf{z} - \mathbf{f}(\mathbf{m}), \quad (4.29)$$

where $\mathbf{z} = (\mathbf{u}, \mathbf{p})^\top$ denotes the solution components, \mathbf{r} is the associated residual vector, and \mathbf{f} the discretized right-hand side of (4.2) and (4.3). Furthermore, \mathbf{K} denotes the stiffness matrix associated with the velocity variable, \mathbf{D} is the velocity divergence matrix, and \mathbf{G} is the pressure gradient matrix.⁴ The residual vector is thus a direct result of discretizing the conservation equations (4.2) and (4.3) together with the boundary conditions (4.6) and (4.7) using the staggered-grid finite difference method.

The nonlinear system (4.29) is solved for \mathbf{z} using the SNES component of PETSc (Balay et al., 2023). In particular, we employ a Newton method with line search:

$$\mathbf{J}(\mathbf{z}_k) \delta \mathbf{z}_k = -\mathbf{r}(\mathbf{z}_k), \quad (4.30)$$

$$\mathbf{z}_{k+1} = \mathbf{z}_k + \alpha \delta \mathbf{z}_k, \quad (4.31)$$

where $\mathbf{J} = \partial \mathbf{r} / \partial \mathbf{z}$ denotes the Jacobian matrix, $\delta \mathbf{z}$ is the iterative correction vector, k is the iteration index, and α the line-search step length. Since we use a Krylov method for solving the linear systems (4.30), only the action of \mathbf{J} on arbitrary vectors is required. Furthermore, the linear systems (4.30) are preconditioned with \mathbf{A}_k^{-1} , where

$$\mathbf{A}_k = \begin{pmatrix} \mathbf{K}(\mathbf{z}_k) & \mathbf{G} \\ \mathbf{D} & \mathbf{C}(\mathbf{z}_k) \end{pmatrix} \quad (4.32)$$

and the blocks \mathbf{K} , \mathbf{D} , and \mathbf{G} are obtained by discretizing the conservation equations (4.2) and (4.3) using the effective viscosity corresponding to the current iterate \mathbf{z}_k (Picard linearization). The remaining block \mathbf{C} is a diagonal matrix whose entries depend on the current inverse effective viscosity (see e.g. May and Moresi, 2008). To approximately invert the preconditioning matrix \mathbf{A}_k , we employ a custom set of multigrid operators specifically designed for the staggered grid (Cai et al., 2014). For more details, the reader is referred to Kaus et al., 2016.

4.4.2 Discrete Adjoint-based Gradient

In this work, we partition the domain into a set of non-overlapping subdomains occupied by different materials, which we refer to as phases. Within each subdomain, we assume a constant approximation for the material parameters (see Figure 4.2). Note that the number of entries in the discrete gradient vector is controlled by the number of parameters used to discretize the material parameters (one entry per parameter, which is 6 for the case shown in Figure 4.2).

For a finite difference discretization as employed in this work, it is not guaranteed that the gradient equation (4.23) is automatically consistent with the finite difference

⁴If the pressure discretization coincides with the discretization of the velocity divergence, $\mathbf{D} = \mathbf{G}^\top$. However, this is not the case for the staggered-grid finite difference discretization employed here.

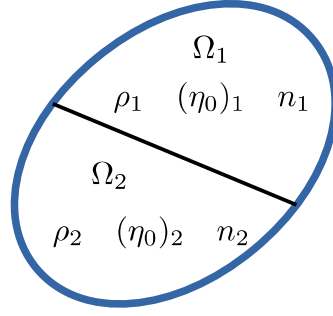


Figure 4.2 Material parameters discretization scheme. Material phases are occupying non-overlapping subdomains. All phase properties are assumed to be constant within each phase subdomain.

stencil used for the forward problem, e.g., due to the required interpolation if stencil and phases are not aligned. When inverting for a spatially dependent parameter field, equation (4.23) can be used directly. However, other types of parameter discretizations, such as piecewise constants (Figure 4.2), require computing a weighted sum of derivatives similar to (4.23) for each discrete material parameter. The choice of these weights depends on the particular discretization and must be made carefully in order to be consistent with interpolation and parameter discretization. To avoid these complications, it becomes convenient to select the discretization first, and then derive the adjoint equations depending on this discretization. This alternative way to obtain adjoint-based gradients is described in the following section.

4.4.3 Discrete Adjoint Equations

After selecting an appropriate discretization of the forward problem and material parameter field, the adjoint equations can also be derived purely algebraically in a finite-dimensional setting. Denoting the solution of the discrete forward problem by \mathbf{z} , which depends on the parameter vector \mathbf{m} , one can define an objective function in a general form:

$$J = J(\mathbf{m}, \mathbf{z}(\mathbf{m})). \quad (4.33)$$

In this work, we minimize the regularized misfit between the computed principal strain directions $\boldsymbol{\varphi}$ and the measured ones $\boldsymbol{\varphi}_{\text{obs}}$:

$$J = J(\mathbf{m}, \boldsymbol{\varphi}(\hat{\boldsymbol{\varepsilon}}(\mathbf{u}(\mathbf{m})))) = \frac{1}{2} |\boldsymbol{\varphi} - \boldsymbol{\varphi}_{\text{obs}}|^2 + \frac{\lambda}{2} |\mathbf{m} - \mathbf{m}_0|_{\mathbf{W}}^2, \quad (4.34)$$

where $|\cdot|$ denotes the Euclidean norm and $|\cdot|_{\mathbf{W}} = |\mathbf{W}^{1/2} \cdot|$. Note that the second term in (4.34) is a regularization analogous to the one used in Section 4.3, with a symmetric positive definite weighting matrix \mathbf{W} and $\lambda \geq 0$.

The gradient of the objective function with respect to the model parameters can

be computed using the following two-step procedure (e.g. Giles and Pierce, 2000):

$$\mathbf{J}^T \boldsymbol{\psi} = \left(\frac{\partial J}{\partial \mathbf{z}} \right)^\top, \quad (4.35)$$

$$\frac{dJ}{d\mathbf{m}} = -\boldsymbol{\psi}^\top \frac{\partial \mathbf{r}}{\partial \mathbf{m}} + \lambda \mathbf{W}(\mathbf{m} - \mathbf{m}_0). \quad (4.36)$$

Here, \mathbf{J} denotes the Jacobian matrix that is already used by the Newton forward solver and can be reused for the adjoint equation (4.35) (see e.g. Worthen et al., 2014), which makes this approach particularly appealing computationally.

Remark 4.2. At first sight, it might seem artificial to solve (4.35) for $\boldsymbol{\psi}$ when the desired gradient is given by

$$\frac{dJ}{d\mathbf{m}} = -\frac{\partial J}{\partial \mathbf{z}} \mathbf{J}^{-1} \frac{\partial \mathbf{r}}{\partial \mathbf{m}} + \lambda \mathbf{W}(\mathbf{m} - \mathbf{m}_0). \quad (4.37)$$

The whole point of the above approach is to avoid assembling the inverse of \mathbf{J} or solving the system

$$\mathbf{J} \boldsymbol{\Psi} = \frac{\partial \mathbf{r}}{\partial \mathbf{m}}, \quad (4.38)$$

which boils down to solving as many linear equation systems as there are entries in the parameter vector \mathbf{m} . In contrast to this, the computational effort of (4.35) and (4.36) amounts to solving only *one* sparse linear system and computing two matrix-vector products. Since the effort is essentially independent of the dimension of the parameter space, this is particularly advantageous for a large number of parameters, e.g., when inverting for a spatially dependent function.

The partial derivative of the residual with respect to the parameters can be computed analytically, or approximated by finite differences (at the cost of accuracy and efficiency):

$$\frac{\partial \mathbf{r}}{\partial m_i} \approx \frac{\mathbf{r}(\mathbf{m} + h \mathbf{e}_i) - \mathbf{r}(\mathbf{m})}{h}. \quad (4.39)$$

In this equation, h is a small perturbation parameter and \mathbf{e}_i is the i -th unit vector. Since every material parameter, in general, only influences very few residual entries, it can be practical to handle the discrete gradient (4.36) as a sparse matrix-vector product, which can be computed on the fly in a matrix-free manner.

The remaining derivative of the objective function with respect to the forward solution vector can be expanded using the chain rule as follows:

$$\frac{\partial J}{\partial \mathbf{z}} = \frac{\partial J}{\partial \boldsymbol{\varphi}} \frac{\partial \boldsymbol{\varphi}}{\partial \dot{\boldsymbol{\varepsilon}}} \frac{\partial \dot{\boldsymbol{\varepsilon}}}{\partial \mathbf{u}}, \quad (4.40)$$

where

$$\frac{\partial J}{\partial \boldsymbol{\varphi}} = \boldsymbol{\varphi} - \boldsymbol{\varphi}_{\text{obs}}, \quad (4.41)$$

and

$$\frac{\partial \boldsymbol{\varphi}}{\partial \dot{\boldsymbol{\varepsilon}}} := \frac{\partial \boldsymbol{\varphi}}{\partial (\dot{\varepsilon}_{xx}, \dot{\varepsilon}_{yy}, \dot{\varepsilon}_{xy})} = \frac{1}{(\dot{\varepsilon}_{xx} - \dot{\varepsilon}_{yy})^2 + 4\dot{\varepsilon}_{xy}^2} (-\dot{\varepsilon}_{xy}, \dot{\varepsilon}_{xy}, \dot{\varepsilon}_{xx} - \dot{\varepsilon}_{yy}). \quad (4.42)$$

Phase	ρ	η_0	$\dot{\epsilon}_0$	n
Matrix	1	1	10^{-6}	2
Block 1 and 2	2	100	10^{-6}	2

Table 4.1 “True” material parameters. In the linear case, $n = 1$ such that $\eta = \eta_0$ (see (4.1)).

The last term $\partial\dot{\epsilon}/\partial\mathbf{u}$ is discretization-dependent; in the appendix of this chapter, the computation is summarized for the staggered-grid finite difference discretization employed here.

In complete analogy to the infinite-dimensional derivation in Section 4.3.1, the computational steps include solving the nonlinear forward problem (4.30) and (4.31), followed by one linear adjoint solve (4.35), and calculating the gradient according to (4.36). The overall numerical cost is essentially independent of the number of material parameters if an analytical expression for $\partial\mathbf{r}/\partial\mathbf{m}$ is used in (4.36).

The computed gradient can be used in any gradient-based inversion framework. In LaMEM, we can use both the TAO optimization library, which is provided through PETSc, and a simple steepest descent algorithm. The TAO BLMVM optimization tool applies a BFGS approximation of the Hessian \mathbf{H} of J and the Moré-Thuente line search (Munson et al., 2012) to update the parameter vector with a quasi-Newton scheme:

$$\mathbf{m}_{l+1} = \mathbf{m}_l - \tilde{\alpha}\mathbf{H}_l^{-1} \frac{dJ}{d\mathbf{m}}(\mathbf{m}_l), \quad (4.43)$$

where l is the inversion iteration counter and $\tilde{\alpha}$ is a line-search parameter.

4.5 Numerical Results

4.5.1 Model Setup

We test the proposed adjoint method with PSD constraints using a synthetic falling block setup. The model consists of a cubic block, with a side length of 0.5, placed in the center of a unit cube domain (see Figure 4.3). The cubic block is surrounded by a lower viscosity and lighter matrix, which causes the block to sink. The domain is discretized by $32 \times 32 \times 32$ cells of the staggered-grid finite difference scheme. As described in Section 4.2.2, we employ a homogeneous normal flow and free slip boundary condition on all sides of the model domain.

We test both linear and nonlinear (power-law) rheology scenarios. The non-dimensional “true” material parameters are listed for both linear and nonlinear cases in Table 4.1. The stiff block inclusion is divided into two equally sized parts, in the following referred to as Block 1 and Block 2, as shown in Figure 4.3. To mimic the sparsity of observations in natural applications, we only use one PSD measurement at a single near-surface point, which is referred to as P in the following sections and has coordinates $\mathbf{x}_P = (0.6, 0.5, 0.99)^\top$.

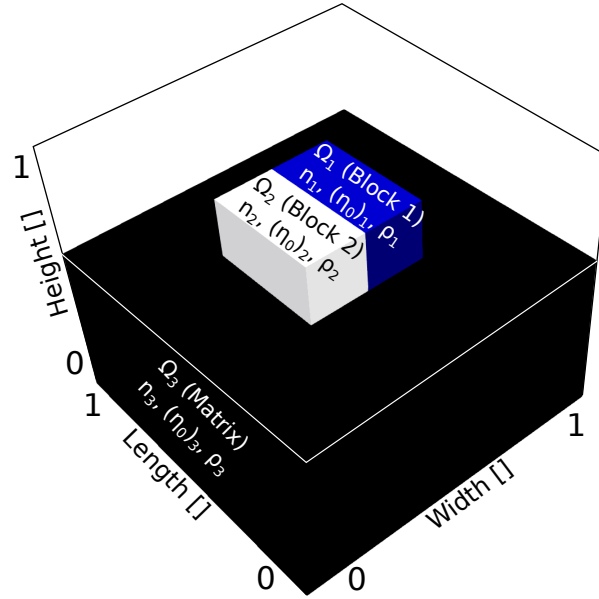


Figure 4.3 Model setup: A cube sinks into the surrounding matrix. The cube is centered in the middle of the domain and consists of two equally sized parts, Ω_1 and Ω_2 . The less viscous and less dense matrix is denoted by Ω_3 . Each phase is described by a constant power-law exponent n_i , constant reference viscosity $(\eta_0)_i$ and a constant density ρ_i . The synthetic data for the simulations in Section 4.5.3 is taken from a near-surface observation point P with coordinates $\mathbf{x}_P = (0.6, 0.5, 0.99)^\top$. In all 3D plots, width, length and height refer to the x -, y - and z -axis, respectively.

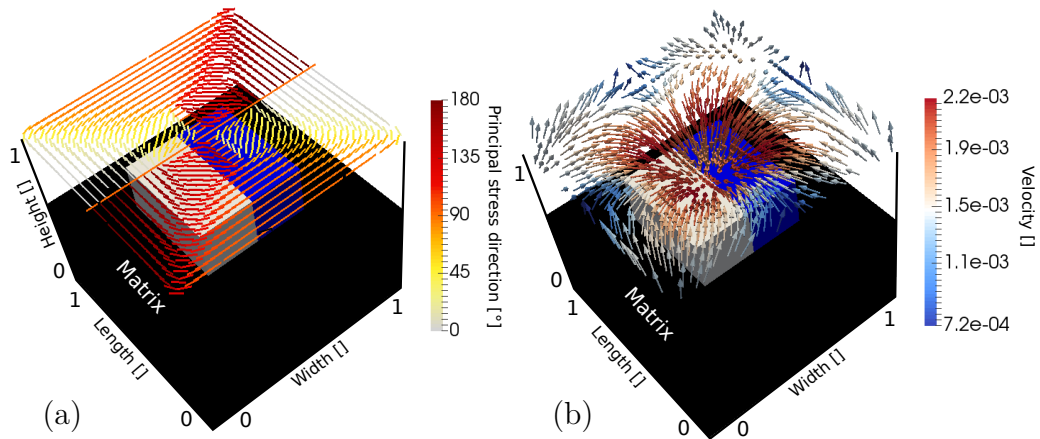


Figure 4.4 Forward problem results for the homogeneous cube: (a) principal stress directions with respect to x -axis, (b) corresponding velocity field.

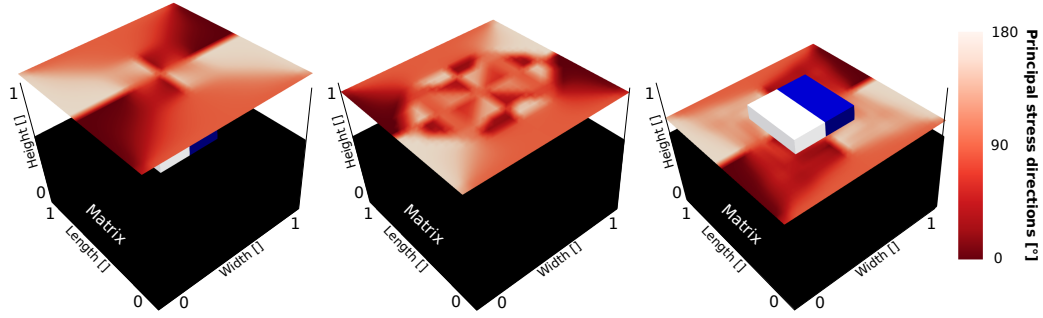


Figure 4.5 Stress direction field in three planes at different heights (from left to right: 1, 0.8, 0.6) in case of a homogeneous cube.

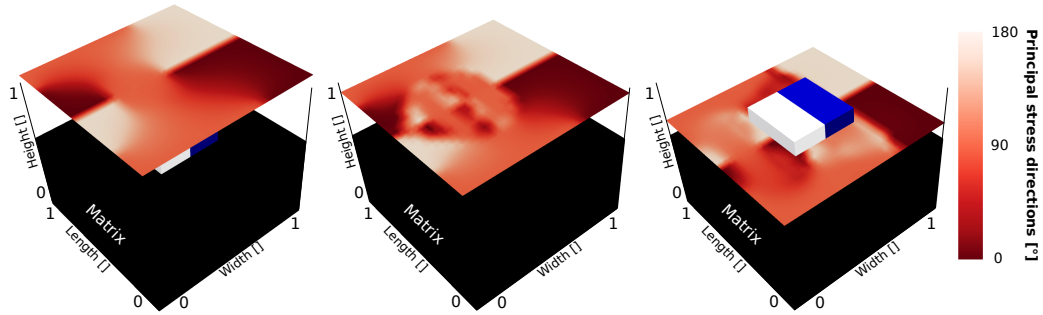


Figure 4.6 Stress direction field in three planes at different heights (from left to right: 1, 0.8, 0.6) in the heterogeneous case, where Block 1 and Block 2 have different rheologies.

4.5.2 Forward Simulations

Figures 4.4 to 4.6 show typical solutions of the forward problem. For a homogeneous block rheology, i.e., the material parameters of Block 1 and 2 are equal, the principal stress orientation forms a concentrically symmetric pattern around a vertical axis passing through the block center (Figure 4.4a). The computed near-surface Stokes velocity field shows the pattern of a downward block motion (Figure 4.4b). The principal stress orientation visualized as a scalar angle field is shown at different depths in Figure 4.5 for the homogeneous rheology case, and in Figure 4.6 for the heterogeneous rheology case (i.e., Block 1 has a different rheology compared to Block 2). As expected, the homogeneous case is characterized by the symmetric PSD orientation pattern at all depths, whereas this symmetry is lost in the heterogeneous case.

4.5.3 Inversion Results

Here, we present a set of simple simulations that illustrate the behavior of the adjoint-based inversion using PSD constraints. All computations start with a synthetic PSD orientation field created with the “true” parameter values given in Table 4.1. In all inversion tests, we keep the rheology of the matrix and Block 2 fixed and invert for one of the following parameters: density ρ , linear viscosity η ($= \eta_0$ in this case), or power-law reference viscosity η_0 of Block 1. The principal stress direction is observed

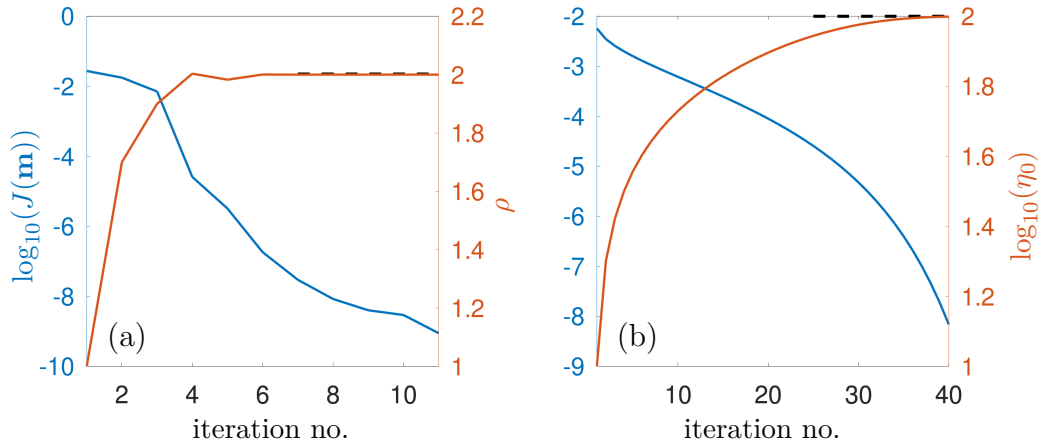


Figure 4.7 Inversion results for linear forward problem ($n = 1$ given). The dashed black line visualizes the “true” synthetic value, the x -axis indicates the number of iterations, the left y -axis shows the logarithm of the objective function value $J(\mathbf{m})$ in blue, and the right y -axis shows the inversion parameter in red: (a) density ρ , (b) linear viscosity $\eta_0 = \eta$.

at a single point, P . Since we only consider simplified models with one observation and one inference parameter, we do not regularize the objective function, i.e., we choose $\lambda = 0$ in (4.34).

Figure 4.7 shows the convergence of the gradient-based inversion for the linear viscous case, i.e., $n = 1$ is fixed. We perform two separate single-parameter inversions, one for viscosity and one for density. In both cases, the inversion parameter converges to the “true” value within a few inverse iterations.

Similar to the previous case, we perform two inversions for a nonlinear power-law rheology. Figure 4.8 shows the inversion results for density and power-law reference viscosity. As in the linear viscous case, convergence is achieved very rapidly in both cases.

4.6 Conclusion and Outlook

We have formulated the adjoint method for efficient computation of gradients using principal stress directions as observations. These gradients can be used in inversion methods to infer material parameters inside the Earth’s crust and lithosphere, such as density or viscosity. Constraining these only from surface observations results in a severely ill-posed inverse problem, which can potentially be improved by including more observations of different data types, which we will also see in Chapter 5. Besides the potential benefit from simply increasing the number of observations, principal stress directions are additionally attractive due to their measurement location inside the Earth. While PSD are measured at comparably low depth, they still provide constraints in different regions of the Earth compared to plate velocities observed on the Earth’s surface. Even though PSD measurements are costly and difficult to obtain and thus only sparsely available (e.g. Heidbach et al., 2018), they can enhance

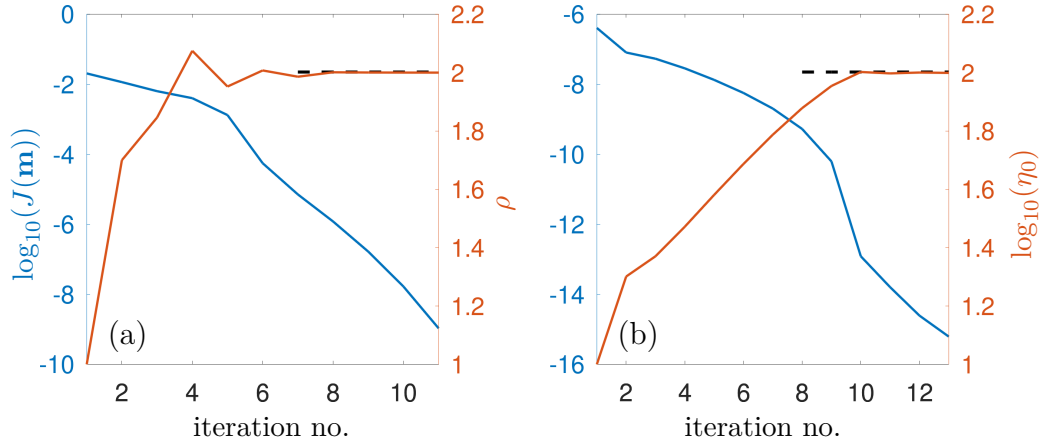


Figure 4.8 Inversion results for nonlinear forward problem. The dashed black line visualizes the “true” synthetic value, the x -axis indicates the number of iterations, the left y -axis shows the logarithm of the objective function value $J(\mathbf{m})$ in blue, and the right y -axis shows the inversion parameter in red: (a) density ρ , (b) power-law reference viscosity η_0 .

an inversion framework that is solely based on GPS velocities (e.g. Kreemer, Holt, and Haines, 2003).

We have shown two ways of computing the adjoint-based gradients. The first approach is formulated on function spaces and thus formally independent of the discretization of the forward and the adjoint problem, whereas the second approach derives the adjoint-based gradients purely algebraically in a finite-dimensional setting, after the conservation equations and parameter fields have been discretized appropriately. The different approaches are sometimes referred to as *optimize-then-discretize (OTD)* and *discretize-then-optimize (DTO)*, respectively (Hinze et al., 2008). While the OTD approach guarantees that the correct infinite-dimensional system is found, the discretizations have to be chosen carefully to ensure consistency between discretized forward problem and adjoint-based gradients. This consistency is automatically guaranteed in the DTO approach since the calculations are based on the discretized forward problem. On the other hand, the adjoint-based gradients obtained here are not necessarily consistent with the corresponding infinite-dimensional system. Furthermore, the discretization of the adjoint system is determined by the discretization of the forward problem in the DTO approach. Although this can have some computational advantages (e.g., possibility to reuse the Jacobian of the Newton solver for the forward problem in the adjoint system, cf. Section 4.4.3), it reduces the flexibility since other discretizations might be more appropriate for the adjoint system. In general, deciding between OTD and DTO depends on both the specific application and the available computational resources.

Although we restricted the numerical simulations to piecewise constant material parameters in fixed subdomains, the framework presented here also allows for inversion of spatially varying fields using adjoint-based gradients. Since spatial subsurface structures are typically not known a priori, this is arguably an important feature for

the application to real-world models. However, due to the severe ill-posedness of this inverse problem, the inferred parameters often suffer from blurry edges even if the data stem from discontinuous parameter fields (Worthen et al., 2014). While the results can be improved using regularization methods like total variation regularization, which promotes sharp transitions between different materials, there can still be many different shapes that fit the data similarly well. In Chapter 5, we will therefore formulate geometric inverse problems for the Stokes equations that guarantee sharp edges and additionally allow us to quantify uncertainties in the inferred subsurface structures by using the Bayesian approach to inverse problems.

In general, the preliminary results based on the synthetic simulations suggest that incorporation of principal stress directions in the computational framework may help improve the robustness and quality of geodynamic inversions, which is encouraging for future applications to large-scale real-world problems.

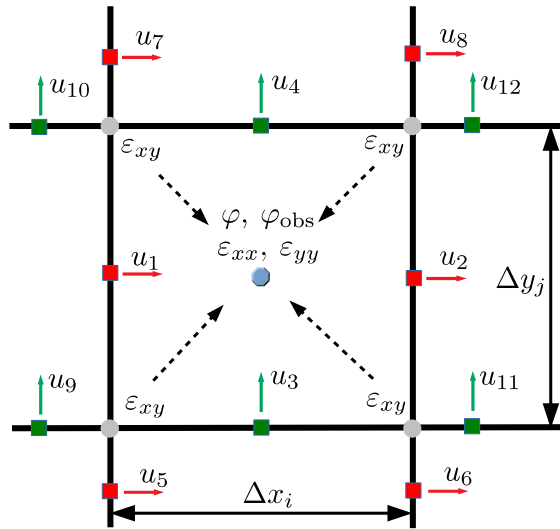


Figure 4.9 Staggered-grid finite difference discretization stencil for the $\partial \hat{\boldsymbol{\varepsilon}} / \partial \mathbf{u}$ term. Shown is a cross-section of a 3D cell in the horizontal (xy -)plane. Circles represent middle points of the edges, squares represent centers of the faces.

5 CHAPTER

Identifying Geometric Structures and Rheological Properties Inside the Earth: A Bayesian Level Set Method

Most of this chapter has already been published in the course of this PhD project in a peer-reviewed journal (Holbach, Gurnis, and Stadler, 2023). Compared to the published version, sections about the well-posedness of the forward as well as the Bayesian inverse problems have been added.

5.1 Motivation

Imaging structures at depth, e.g. in the crust, lithosphere or mantle, and constraining their mechanical properties is a major challenge in geodynamics. In Chapter 4, we addressed this problem by introducing principal stress directions as a novel data type for geodynamic inverse problems. However, while the approach is feasible for the identification of constant material properties in predefined regions, there is little hope in recovering geometric structures inside the Earth when inverting for these parameters as fields as discussed in the last chapter. Therefore, the focus of this chapter is on the development of methods for geometric inverse problems and their application in geodynamics.

While seismic tomography provides images of the lithosphere and mantle, they tend to be blurry with no direct translation from reconstructed velocity anomalies to rheology. However, incorporating seismic images as prior knowledge into models for the mechanical long-term behavior of the mantle along with consistency between model predictions and observational data is a promising approach. By modeling mantle flow using the Stokes equations, we aim at simultaneously inferring subsurface structures and material parameters from surface measurements of plate velocities and normal tractions while incorporating seismic information as prior knowledge.

Optimization of geometric structures is challenging, in particular when the topology of the approximation can change between subsequent iterations. For many physics simulations, level set and phase field methods have been established as flexible and efficient tools for this purpose (e.g. Osher and Sethian, 1988; Boettinger et al., 2002). Inferring spatial structures from observational data is particularly challenging when access to data is limited and if one is additionally interested in quantifying uncertainties in the inferred parameters. Various approaches for geometric inverse problems have been explored, including (i) inverting for fields, e.g., Worthen et al.

(2014), (ii) inferring constant parameters in predefined regions, e.g., Baumann and Kaus (2015), (iii) explicitly parameterizing subdomains, e.g., Carpio, Iakunin, and Stadler (2020), and (iv) using level-set parameterizations, e.g., Kadu, Leeuwen, and Mulder (2017) and Muir and Tsai (2019).

When aiming to identify high-dimensional parameter fields in geodynamics, the results often suffer from blurry edges due to the ill-posedness of the problem (Worthen et al., 2014). Therefore, one may resort to predefined regions in which one infers spatially restricted constants (Baumann and Kaus (2015); Chapter 4). While this approach does not allow one to infer geometric features, it reduces the dimension and thus enables quantification of uncertainties using sampling methods. Alternatively, one can rely on parameterizations of geometrical structures. These include explicit parameterizations of geometric interfaces (Carpio, Iakunin, and Stadler, 2020), for which a moderate number of parameters may be sufficient. However, this approach limits geometric flexibility. Using an implicit parameterization such as describing geometric structures as level sets of an auxiliary function has become a particularly popular tool for geometric inverse problems due to the shape flexibility it provides (Santosa, 1996).

In the context of seismic imaging and full waveform inversion, parametric level set methods have been employed to identify subsurface salt bodies (Kadu, Leeuwen, and Mulder, 2017), whereas a non-parametric level set method has been used to improve image resolution in travel-time tomography (Muir and Tsai, 2019). Additionally, the level-set technique has been applied to joint inversion of density contrast and seismic slowness using gravity and travel-time data (Li and Qian, 2016). The above papers deal with a purely deterministic setting and do not allow for quantification of uncertainties. Recently, methods to estimate geometric uncertainties have been applied to level-set based travel-time tomography (Muir et al., 2022). In general, Bayesian inference is a powerful tool to infer geometric uncertainties. In this approach, prior beliefs are stated in terms of probabilities and the solution to the inverse problem is not one particular shape but instead a probability distribution for the shapes. On the other hand, the curse of dimensionality typically forces one to either work in low dimensions or use dimension-robust inference methods, which is the approach taken here.

Some geodynamic inverse studies have assumed the spatial structure of the present-day subsurface (e.g., mantle structure) to be known and used adjoints of the Stokes equations while employing surface velocity measurements (Ratnaswamy, Stadler, and Gurnis, 2015; Rudi, Gurnis, and Stadler, 2022). These studies have proven to be efficient at recovering rheological parameters and the covariances between them. However, the models are limited in their ability to decipher the trade-offs between rheology and density or assessing the role of uncertain interfaces between different materials. Other studies have attempted to recover geometrical structures in the geological past through an integration of the convection equations backward in time (Bunge, Hagelberg, and Travis, 2003; Liu, Spasojevic, and Gurnis, 2008).

Here, we build on Iglesias, Lu, and Stuart, 2016 and use flexible geometric priors for binary functions with parameter-to-observable maps that are expensive to evaluate. After reviewing well-posedness of the relevant forward problem in Section 5.2, we formulate the inverse problem using a level-set parameterization for uncertain

subsurface structures and material parameters in Section 5.3. By applying the theory of Iglesias, Lu, and Stuart, 2016, in Section 5.4 we show that the corresponding Bayesian inverse problem is well-posed, i.e., the posterior distribution is well-defined and locally Lipschitz continuous in the data with respect to a suitable metric. In Section 5.5, we propose a heuristic approach for constructing tailored level-set priors with adjusted mean and variance representing seismic knowledge, building a bridge between seismic and mechanical models. Numerical methods and examples are presented in Sections 5.6 and 5.7. In particular, we study the trade-offs between geometric structures, density, and viscosity using dimension-robust sampling methods to approximate the posterior distribution and quantify uncertainties in the inferred shapes and material parameters. Finally, we investigate how informative different observation types are and demonstrate the benefits of combining them.

5.2 Well-Posedness of the Forward Problem

While the Bayesian level set method we describe in the following sections is also applicable to Stokes equations with various nonlinear rheologies, here we focus on linear Stokes equations with Newtonian rheology. Furthermore, although the method readily extends to higher dimensions, we only consider two-dimensional domains for easier visualization. The model we use is thus the one discussed in Section 1.2. To provide a cleaner structure, we review the results here.

Denote $\Omega \subset \mathbb{R}^2$ as the bounded domain with boundary $\partial\Omega$. For illustration, we assume Ω to be a rectangle in this chapter. Let $\eta \in L_+^\infty(\Omega)$ be the spatially varying effective viscosity. The weak form of the Stokes equations considered here consists of finding $(\mathbf{u}, p) \in V \times Q$ such that

$$\int_{\Omega} 2\eta \dot{\boldsymbol{\varepsilon}}(\mathbf{u}) : \dot{\boldsymbol{\varepsilon}}(\mathbf{v}) \, d\mathbf{x} - \int_{\Omega} p \operatorname{div} \mathbf{v} \, d\mathbf{x} - \int_{\Omega} q \operatorname{div} \mathbf{u} \, d\mathbf{x} = \int_{\Omega} \rho \mathbf{g} \cdot \mathbf{v} \, d\mathbf{x} \quad (5.1)$$

for all $(\mathbf{v}, q) \in V \times Q$, where

$$V := \left\{ \mathbf{u} = (u_1, u_2)^\top \in H^1(\Omega)^2 \mid \mathbf{u} \cdot \boldsymbol{\nu} = 0 \text{ on } \partial\Omega \right\}, \quad (5.2)$$

$$Q := L_0^2(\Omega) := \left\{ p \in L^2(\Omega) \mid \int_{\Omega} p \, d\mathbf{x} = 0 \right\}. \quad (5.3)$$

Note that we assume that gravity is the only body force applied to the fluid, i.e., $\mathbf{f} = \rho \mathbf{g}$, where $\rho \in L^2(\Omega)$ is the fluid's density and $\mathbf{g} = g(0, -1)^\top \in \mathbb{R}^2$ with the gravitational acceleration $g \in \mathbb{R}_+$. Using this notation, we state the well-posedness of the forward problem discussed in Section 1.2 tailored to the purposes of this chapter.

Theorem 5.1. *Let $\Omega \subset \mathbb{R}^2$ be an open rectangle, $\eta \in L_+^\infty(\Omega)$, and $\rho \in L^2(\Omega)$.*

- (a) *There is a unique weak solution $(\mathbf{u}, p) \in V \times Q$ to the Stokes equations (5.1). Furthermore, the stress tensor $\boldsymbol{\sigma}(\mathbf{u}, p)$ lies in $H(\operatorname{div}; \Omega)$ and we have the following stability estimates for some constants $C_i = C_i(\|\eta\|_\infty, \eta_{\min}) > 0$, $i \in \{1, 2, 3\}$:*

$$\|\mathbf{u}\|_{H^1} \leq C_1 \|\rho\|_{L^2}, \quad (5.4)$$

$$\|p\|_{L^2} \leq C_2 \|\rho\|_{L^2}, \quad (5.5)$$

$$\|\boldsymbol{\sigma}\|_{H(\operatorname{div}; \Omega)} \leq C_3 \|\rho\|_{L^2}. \quad (5.6)$$

(b) If additionally $\rho \in L^r(\Omega) \subset L^2(\Omega)$ for some $2 < r < \infty$, the solution depends continuously on ρ in $L^2(\Omega)$ and η in $L^s(\Omega)$ for $2 < s := 2r/(r-2) < \infty$. More precisely, denoting $\hat{\mathbf{u}} \in V$, $\hat{p} \in Q$, $\hat{\boldsymbol{\sigma}} \in H(\text{div}; \Omega)$ the unique solution corresponding to the perturbed parameters $\hat{\rho} \in L^r(\Omega) \subset L^2(\Omega)$ and $\hat{\eta} \in L^s_+(\Omega) \subset L^s(\Omega)$, there are constants $C_i > 0$, $i \in \{4, 5, 6\}$, such that

$$\|\hat{\mathbf{u}} - \mathbf{u}\|_{H^1} \leq C_4 (\|\hat{\rho} - \rho\|_{L^2} + \|\hat{\rho}\|_{L^r} \|\hat{\eta} - \eta\|_{L^s}), \quad (5.7)$$

$$\|\hat{p} - p\|_{L^2} \leq C_5 (\|\hat{\rho} - \rho\|_{L^2} + \|\hat{\rho}\|_{L^r} \|\hat{\eta} - \eta\|_{L^s}), \quad (5.8)$$

$$\|\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\|_{H(\text{div}; \Omega)} \leq C_6 (\|\hat{\rho} - \rho\|_{L^2} + \|\hat{\rho}\|_{L^r} \|\hat{\eta} - \eta\|_{L^s}). \quad (5.9)$$

5.3 Formulating the Inverse Problem

Although mantle structures can be inferred with seismic or with electrical and magnetic imaging methods, rheological parameters cannot be directly determined through measurements, and density is poorly determined at regional and smaller scales. Consequently, we estimate rheology and density by solving an inverse problem. More precisely, we are interested in recovering a vector \mathbf{m} of unknown parameters (e.g. density, viscosity, etc.) from a finite number of noisy measurements $\mathbf{y} \in \mathbb{R}^n$, i.e.,

$$\mathbf{y} = \mathcal{G}(\mathbf{m}) + \boldsymbol{\delta}, \quad (5.10)$$

where $\boldsymbol{\delta} \in \mathbb{R}^n$ describes measurement and model errors, and \mathcal{G} is the parameter-to-observable map. In our case, \mathcal{G} involves solving the Stokes equations in their weak form (5.1) given the parameters \mathbf{m} and evaluating the solution at the measurement locations. As mentioned in Chapter 1, inverse problems governed by partial differential equations are typically ill-posed, i.e., they do not have a unique solution or small perturbations in the data may lead to large differences in the reconstructions (e.g. Hanke, 2017). To overcome this issue, we regularize the inverse problem by employing the Bayesian framework in Section 5.4, which has the additional advantage of providing information about uncertainties in the parameters, in contrast to the deterministic optimization approach taken in Chapter 4.

One of our aims is to investigate the effect of different data types on the inversion. In all calculations, the data \mathbf{y} are observed on the upper boundary of the domain, which corresponds to the Earth's surface. We will consider two data types individually and in combination: (i) horizontal velocities u_1 and (ii) normal traction $\sigma_\nu := \boldsymbol{\nu} \cdot \boldsymbol{\sigma}\boldsymbol{\nu}$.

5.3.1 Level Set Parameterization for Uncertain Geometric Structures

The rheological parameters of interest often exhibit large variation over small length scales. Therefore, we assume them to be piecewise continuous functions, allowing for discontinuities in the fields. As we are primarily concerned with inferring geometric structures, we impose as few restrictions as possible on the interfaces between different structural entities, which we refer to as phases. This is achieved by adapting a Bayesian version of the level set method that was introduced in Iglesias, Lu, and Stuart (2016). The main idea is to describe interfaces as level sets of a continuous function and infer this function from observed data. Note that, as opposed to classical level set methods that use the level set function to evolve shapes in some

time-dependent or iterative process, here level set functions are only used to define and sample distributions of geometrical subsurface structures. Next, we detail level set parameterizations using the density field ρ as an example.

We denote the indicator function of a set $A \subseteq \Omega$ with $\mathbb{1}_A$, i.e.,

$$\mathbb{1}_A(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in A, \\ 0, & \text{if } \mathbf{x} \notin A. \end{cases} \quad (5.11)$$

Let

$$\rho(\mathbf{x}) = \sum_{i=1}^m f_i(\mathbf{x}, \rho_i) \mathbb{1}_{\Omega_i}(\mathbf{x}) \quad (5.12)$$

for some subsets

$$\Omega_i \subseteq \Omega, \quad i = 1, \dots, m, \quad \Omega_i \cap \Omega_j = \emptyset \text{ for } i \neq j, \quad \bigcup_{i=1}^m \bar{\Omega}_i = \bar{\Omega}. \quad (5.13)$$

The (known) coefficients f_i are continuous and essentially bounded from above and below in both arguments, i.e., $f_i \in L^{\infty}_+(\Omega \times \mathbb{R}) \cap \mathcal{C}(\Omega \times \mathbb{R})$. In the simplest case, $f_i \equiv \rho_i$ for fixed scalars $\rho_i > 0$ and (5.12) defines a piecewise constant density field (cf. Figure 5.1 (b)). Introducing the functions f_i additionally allows for (nonlinear) transformations of the parameters ρ_i , e.g., to model log-normally distributed variables, as well as spatial dependencies inside the different phases, e.g., to prescribe known smooth transitions in certain regions of the domain. Each coefficient function f_i is parameterized by a single value ρ_i because it is notationally convenient for the applications considered in Section 5.7. However, (5.12) can be generalized to arbitrary continuous and bounded functions f_i .

The sets Ω_i determine the shapes of the different phases. The idea of the level set method is to parameterize these sets with the help of a continuous function $h: \Omega \rightarrow \mathbb{R}$, the so-called *level set function*. For given thresholds

$$-\infty = a_0 < a_1 < \dots < a_m = \infty, \quad (5.14)$$

we define

$$\Omega_i := \Omega_i(h) := \{\mathbf{x} \in \Omega \mid a_{i-1} \leq h(\mathbf{x}) < a_i\}, \quad i = 1, \dots, m, \quad (5.15)$$

such that the shapes of the phases are fully determined by the level set function h . This leaves considerable flexibility in describing the geometric structure of ρ and allows us to change the interfaces implicitly by updating h .

The level set function is naturally linked to the physical parameter of interest ρ through the so-called *level set map* F defined via

$$\begin{aligned} F: \mathcal{C}(\bar{\Omega}) \times \mathbb{R}^m &\rightarrow L^r(\Omega), \\ (h, \boldsymbol{\rho}) &\mapsto \sum_{i=1}^m f_i(\cdot, \rho_i) \mathbb{1}_{\Omega_i(h)} = \rho, \end{aligned} \quad (5.16)$$

where $\boldsymbol{\rho} = (\rho_1, \dots, \rho_m)^\top$ and $1 \leq r < \infty$. This procedure is visualized in Figure 5.1.

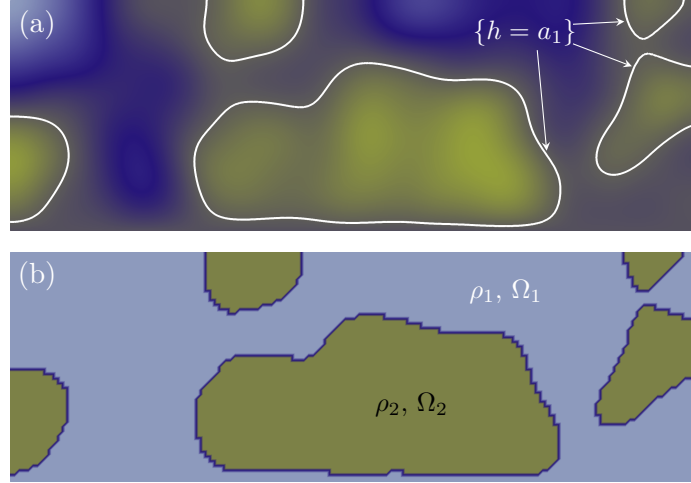


Figure 5.1 Visualization of the level set map F for the simple case with two phases and constant coefficients $f_i(\cdot, \rho_i) \equiv \rho_i$. (a) Continuous level set function h with different colors indicating different values of the function and the white lines indicating the level set threshold a_1 between the two phases. (b) Corresponding piecewise constant density field $\rho = F(h, \rho)$.

Using the above parameterization, we can infer the continuous level set function h rather than the density field ρ directly, i.e., we replace the density parameter ρ in \mathbf{m} by (h, ρ) and invert for both the geometric structure and the specific phase values of the density simultaneously. For simplicity, we abuse notation and refer to the adjusted parameter-to-observable map as \mathcal{G} as well.

Compared to Iglesias, Lu, and Stuart (2016), the level set parameterization (5.16) formulated above allows for varying phase values of the density and inversion of these values as well; in the cited paper, $f_i \equiv \rho_i$ for known constants $\rho_i > 0$. For this setting, they prove that the level set map is continuous at $h \in \mathcal{C}(\overline{\Omega})$ if and only if the Lebesgue measure of the level sets corresponding to the thresholds a_i , $i \in \{1, \dots, m-1\}$, is 0. This allows us to formulate an analogous statement for the present setting:

Proposition 5.2. *Let $(h, \rho) \in \mathcal{C}(\overline{\Omega}) \times \mathbb{R}^m$ and $1 \leq r < \infty$. The level set map $F: \mathcal{C}(\overline{\Omega}) \times \mathbb{R}^m \rightarrow L^r(\Omega)$ defined in (5.16) is continuous at (h, ρ) if and only if $\mathbb{K}(\{h = a_i\}) = 0$ for all $i \in \{1, \dots, m-1\}$.*

PROOF. The statement follows immediately from Iglesias, Lu, and Stuart (2016, Proposition 2.6): It is clear that the condition remains necessary for continuity. On the other hand, the coefficients f_i are continuous in ρ_i and thus F is a composition of continuous functions if $\mathbb{K}(\{h = a_i\}) = 0$ for all $i \in \{1, \dots, m-1\}$. \square

Remark 5.3. Proposition 5.2 is not valid for $r = \infty$ as one can see with simple examples: Let $\Omega = (-1, 1)^2$ and define the level set map via $F: \mathcal{C}(\overline{\Omega}) \times \mathbb{R}^2 \rightarrow L^\infty(\Omega)$, $(h, \rho) \mapsto \mathbb{1}_{\{h \geq 0\}}$. Let $h(x_1, x_2) := x_1$ and $h_n(\mathbf{x}) := h(\mathbf{x}) + 1/n$, $n \in \mathbb{N}$, for $\mathbf{x} =$

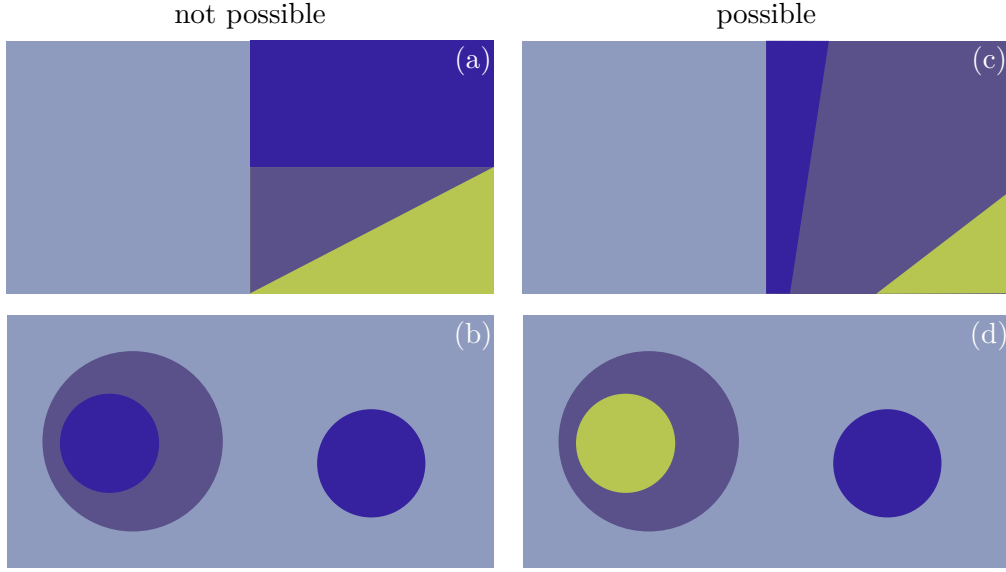


Figure 5.2 Examples of piecewise constant functions, where different colors indicate different values. Fields (a) and (b) cannot be realized using the level set parameterization (5.15) since the level set function h is continuous. On the other hand, fields (c) and (d) show interfaces and corresponding phase values that are possible.

$(x_1, x_2)^\top \in \Omega$. Clearly, $\lambda(\{h = 0\}) = 0$ and

$$\|h_n(\mathbf{x}) - h(\mathbf{x})\|_\infty = \frac{1}{n} \xrightarrow{n \rightarrow \infty} 0, \quad (5.17)$$

but

$$\|F(h_n, \boldsymbol{\rho}) - F(h, \boldsymbol{\rho})\|_\infty = \left\| \mathbf{1}_{\{x_1 + 1/n \geq 0\}} - \mathbf{1}_{\{x_1 \geq 0\}} \right\|_\infty = 1, \quad n \in \mathbb{N}, \quad (5.18)$$

since, e.g., $-1/(2n) < 0$ and $-1/(2n) + 1/n > 0$ for every $n \in \mathbb{N}$. This is problematic because the natural space for the viscosity field, one of the inversion parameters, is $L_+^\infty(\Omega)$. However, Theorem 5.1 provides continuous dependence of the weak solution of the Stokes equations on $\eta \in L_+^\infty(\Omega) \subset L^s(\Omega)$ for some $2 < s < \infty$, which will help overcome these issues in Section 5.4.1.

As stated in Proposition 5.2, the level set map F is discontinuous in general, even in $L^r(\Omega)$, $1 \leq r < \infty$, and hence the adjusted parameter-to-observable map \mathcal{G} is discontinuous as well. This implies that \mathcal{G} is nonlinear even if the data depend linearly on the physical parameters of interest.

As pointed out in Iglesias, Lu, and Stuart (2016), due to the way the different phases Ω_i are defined in (5.15), only phases with subsequent indices can occur next to each other (see Figure 5.2). In their setting, this means that one has to determine the ordering of the (constant) phase values, ρ_i , a priori. Our situation is slightly improved since the coefficients are continuous functions that are part of the inversion, and can thus change their values and therefore the ordering. Nonetheless, our formulation

also prohibits jumps between arbitrary phases. To achieve this, one needs multiple level set functions or a vector-valued level set function (cf. Dorn and Lesselier, 2006; Dorn and Lesselier, 2009).

5.4 Well-Posedness of the Bayesian Inverse Problem

We briefly introduce Bayesian inversion as applied here; more detailed introductions can be found in Kaipio and Somersalo (2005) and Stuart (2010). In this approach, all quantities in (5.10) are treated as random variables and we assume some prior knowledge on our parameters of interest \mathbf{m} in the form of a probability distribution, the *prior distribution* μ_0 , which serves as a regularization. As opposed to deterministic inversion methods, where the error is typically assumed to be either very small or known, uncertainties in the measurements are encoded into the probability distribution for the noise vector $\boldsymbol{\delta}$. This is arguably a more realistic assumption since one might know the average measurement error of some device but not the exact error for any instance of the measurements. Additionally, approximation errors or uncertainties in the model itself can be accounted for in the noise distribution, another difference to deterministic approaches, which usually assume to know the exact model.

The goal of Bayesian inversion is to find and describe the *posterior probability distribution* $\mu^{\mathcal{Y}}$, which is a conditional distribution of the parameters \mathbf{m} given the data \mathbf{y} — contrary to deterministic inversion methods, the solution to a Bayesian inverse problem is thus not one estimate but a probability distribution for the unknown parameters of interest. This allows us to obtain multiple estimates with assigned probabilities and therefore to quantify uncertainties in the inferred parameters, an aspect that can be crucial for practical applications.

Let M be a separable Banach space and (M, \mathcal{B}, μ_0) a complete probability space for the parameter vector \mathbf{m} , where \mathcal{B} is the sigma algebra. We denote the Gaussian prior distribution as $\mu_0 = \mathcal{N}(\mathbf{m}_0, \mathcal{C}_0)$, where \mathbf{m}_0 is the mean and \mathcal{C}_0 the covariance operator. The noise vector $\boldsymbol{\delta}$ and \mathbf{m} are assumed to be mutually independent, and $\boldsymbol{\delta}$ is normally distributed with mean zero and covariance matrix $\boldsymbol{\Gamma} \in \mathbb{R}^{n \times n}$, i.e., $\boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma})$. The posterior distribution $\mu^{\mathcal{Y}}$ is defined through Bayes' law. Since our inversion parameter vector \mathbf{m} contains a function, an infinite-dimensional version of Bayes' law is needed. For continuous parameter-to-observable maps, the posterior distribution is well-defined under mild conditions and given by (e.g. Stuart, 2010)

$$\begin{aligned} \frac{d\mu^{\mathcal{Y}}}{d\mu_0}(\mathbf{m}) &= \frac{\exp(-\mathcal{J}(\mathbf{m}; \mathbf{y}))}{\int_M \exp(-\mathcal{J}(\hat{\mathbf{m}}; \mathbf{y})) \mu_0(d\hat{\mathbf{m}})} \\ &\propto \exp(-\mathcal{J}(\mathbf{m}; \mathbf{y})), \end{aligned} \quad (5.19)$$

where $d\mu^{\mathcal{Y}}/d\mu_0$ denotes the Radon-Nikodym derivative of $\mu^{\mathcal{Y}}$ with respect to μ_0 and

$$\mathcal{J}(\mathbf{m}; \mathbf{y}) := \frac{1}{2} \|\mathbf{y} - \mathcal{G}(\mathbf{m})\|_{\boldsymbol{\Gamma}^{-1}}^2 \quad (5.20)$$

$$= \frac{1}{2} (\mathbf{y} - \mathcal{G}(\mathbf{m}))^\top \boldsymbol{\Gamma}^{-1} (\mathbf{y} - \mathcal{G}(\mathbf{m})). \quad (5.21)$$

Furthermore, $\mu^{\mathbf{y}}$ is locally Lipschitz continuous in the data \mathbf{y} with respect to the Hellinger distance; we give the precise mathematical statement in Section 5.4.1.

Note that \mathcal{J} is the least-squares data misfit weighted by the inverse noise covariance matrix. It can be shown that maximizers of (5.19), referred to as *maximum a posteriori* or *MAP estimators*¹, coincide with minimizers of the prior-regularized least-squares misfit

$$\frac{1}{2} |\mathbf{y} - \mathcal{G}(\mathbf{m})|_{\Gamma^{-1}}^2 + \frac{1}{2} \|\mathbf{m} - \mathbf{m}_0\|_{\mathcal{C}_0^{-1}}^2. \quad (5.22)$$

In the special case $\Gamma = \mathbf{I}$ and $\mathcal{C}_0^{-1} = \lambda\mathcal{W}$ with the regularization parameter λ and weighting operator \mathcal{W} from Chapter 4, (5.22) agrees with the objective function minimized there, linking Bayesian inversion to deterministic approaches to solving inverse problems.

Yet, the variance in the MAP estimator can be large and the function (5.22) can have multiple minima. The Bayesian approach allows us to study the corresponding probabilities and related statistical quantities, not only for the full parameter vector \mathbf{m} but also for individual components (like the density $\boldsymbol{\rho}$ or viscosity $\boldsymbol{\eta}$) through the marginal posterior distributions. Furthermore, choosing a prior distribution is a more explicit way of incorporating knowledge about the parameter into the inversion compared to selecting and tuning a regularization term in a deterministic approach. A prior distribution tailored to the specific model can be critical for realistic inversion results as we will see in Section 5.7.2.

Returning to our level-set setting, in view of Proposition 5.2, the parameter-to-observable map is discontinuous in general and the theory above does not apply directly. However, Iglesias, Lu, and Stuart (2016, Theorem 2.3) show that the continuity assumption can be relaxed. The crucial difference to standard theory for Bayesian inverse problems is that the parameter-to-observable map only needs to be continuous almost surely with respect to the prior measure μ_0 . More precisely, the Bayesian inverse problem is well-posed under the following assumptions:

Assumptions 5.4. Let $\mathcal{J}: M \times \mathbb{R}^n \rightarrow \mathbb{R}$ and μ_0 be the prior distribution on (M, \mathcal{B}) .

1. For every $R > 0$, there is a constant $C_1 = C_1(R) > 0$ such that

$$0 \leq \mathcal{J}(\mathbf{m}; \mathbf{y}) \leq C_1 \quad (5.23)$$

for all $\mathbf{m} \in M$ and all $\mathbf{y} \in \mathbb{R}^n$ with $|\mathbf{y}|_{\Gamma^{-1}} < R$.

2. For any fixed $\mathbf{y} \in \mathbb{R}^n$, the functional $\mathcal{J}(\cdot; \mathbf{y}): M \rightarrow \mathbb{R}$ is continuous μ_0 -almost surely on the complete probability space (M, \mathcal{B}, μ_0) .
3. For every $R > 0$, there is a constant $C_2 = C_2(R) > 0$ such that

$$|\mathcal{J}(\mathbf{m}; \mathbf{y}_1) - \mathcal{J}(\mathbf{m}; \mathbf{y}_2)| \leq C_2 |\mathbf{y}_1 - \mathbf{y}_2|_{\Gamma^{-1}} \quad (5.24)$$

for all $\mathbf{m} \in M$ and all $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^n$ with $\max(|\mathbf{y}_1|_{\Gamma^{-1}}, |\mathbf{y}_2|_{\Gamma^{-1}}) < R$.

This suggests choosing prior measures μ_0 satisfying the following condition:

¹These are in general not the same as the non-Bayesian maximum likelihood (or ML) estimators, which instead minimize the unregularized data misfit (5.20).

Assumption 5.5. For μ_0 -almost every level set function h ,

$$\mathbb{A}(\{h = a_i\}) = 0 \quad \text{for all } i \in \{1, \dots, m-1\}. \quad (5.25)$$

For any Gaussian prior measure μ_0 satisfying Assumption 5.5, in the next section we show that Assumptions 5.4 hold for the inverse problems governed by Stokes flow that we consider in Section 5.7.1.

5.4.1 Linear Stokes Equations with Surface Measurements

We formulate the most general setting that we employ in Section 5.7 such that the arguments for the simpler cases are included. As a first step, we formally define the parameter-to-observable map \mathcal{G} , which we split into three functions:

$$\mathcal{G}: M \xrightarrow{F} L^2(\Omega) \times L^s(\Omega) \xrightarrow{G} V \times H(\text{div}; \Omega) \xrightarrow{\mathcal{O}} \mathbb{R}^n, \quad (5.26)$$

where F is the level set map, G the *parameter-to-state map*, and \mathcal{O} an *observation operator*, all of which will be specified in the following.

Since we are interested in inverting for both the density and the viscosity fields, we introduce individual level set maps for each field. However, we assume that the geometric structure is defined by the same level set function such that only the phase values are different.² Let $f_{1,i}, f_{2,i} \in L_+^\infty(\Omega \times \mathbb{R}) \cap \mathcal{C}(\Omega \times \mathbb{R})$, $i \in \{1, \dots, m\}$, be the continuous coefficients. We define the two level set maps

$$F_1: \mathcal{C}(\bar{\Omega}) \times \mathbb{R}^m \rightarrow L_+^\infty(\Omega), \quad (h, \boldsymbol{\rho}) \mapsto \sum_{i=1}^m f_{1,i}(\cdot, \rho_i) \mathbb{1}_{\Omega_i(h)} =: \rho, \quad (5.27a)$$

$$F_2: \mathcal{C}(\bar{\Omega}) \times \mathbb{R}^m \rightarrow L_+^\infty(\Omega), \quad (h, \boldsymbol{\eta}) \mapsto \sum_{i=1}^m f_{2,i}(\cdot, \eta_i) \mathbb{1}_{\Omega_i(h)} =: \eta, \quad (5.27b)$$

where the sets Ω_i are defined in (5.15). Note that

$$\|F_1(h, \boldsymbol{\rho})\|_\infty \leq \max_{i=1, \dots, m} (\text{ess sup } f_{1,i}) < \infty, \quad (5.28a)$$

$$\|F_1(h, \boldsymbol{\rho})\|_{L^2} \leq \max_{i=1, \dots, m} (\text{ess sup } f_{1,i}) \mathbb{A}(\Omega)^{\frac{1}{2}} < \infty, \quad (5.28b)$$

$$\text{ess inf } F_1(h, \boldsymbol{\rho}) \geq \min_{i=1, \dots, m} (\text{ess inf } f_{1,i}) > 0, \quad (5.28c)$$

and these bounds do not depend on the argument $(h, \boldsymbol{\rho})$. This and the analogous statement for F_2 will become important when checking Assumptions 5.4. Defining the separable Banach space $M := \mathcal{C}(\bar{\Omega}) \times \mathbb{R}^m \times \mathbb{R}^m$, the joint level set map is given by

$$F: M \rightarrow L^2(\Omega) \times L^s(\Omega), \quad (h, \boldsymbol{\rho}, \boldsymbol{\eta}) \mapsto (F_1(h, \boldsymbol{\rho}), F_2(h, \boldsymbol{\eta})). \quad (5.29)$$

Although $\mathcal{R}(F) \subset L_+^\infty(\Omega) \times L_+^\infty(\Omega)$ and $L_+^\infty(\Omega)$ is the natural space for the viscosity η , in order to apply Proposition 5.2 it is necessary to view the codomain of F as

²The well-posedness of the Bayesian inverse problem follows analogously with two distinct level set functions. As it is reasonable to assume the geometrical structures to be coupled in the applications considered, we focus on the case of one level set function to keep the notation as simple as possible.

5.4 Well-Posedness of the Bayesian Inverse Problem

$L^2(\Omega) \times L^s(\Omega)$ for some $2 < s < \infty$ as explained below. Note that it is possible to assume certain values ρ_i or η_i to be known and remove these components from the parameter space M .

The function G now maps the density field $\rho = F_1(h, \boldsymbol{\rho})$ and viscosity field $\eta = F_2(h, \boldsymbol{\eta})$ to the corresponding velocity field \mathbf{u} and stress tensor $\boldsymbol{\sigma}$ that solve the Stokes equations (5.1), i.e.,

$$G: L^2(\Omega) \times L^s(\Omega) \rightarrow V \times H(\operatorname{div}; \Omega), \quad (\rho, \eta) \mapsto (\mathbf{u}, \boldsymbol{\sigma}). \quad (5.30)$$

Due to Theorem 5.1 (b), the parameter-to-state map G is continuous³ if $\rho \in L^r(\Omega)$ for some $r \in (2, \infty)$ and one chooses $s := 2r/(r-2) \in (2, \infty)$. Since $\rho \in L^{\infty}_+(\Omega)$ by definition (5.27a), G is continuous for any choice of $s \in (2, \infty)$.

It remains to define the observation operator. Let $\partial\Omega_t$ denote the upper boundary of the domain Ω , corresponding to the Earth's surface. Two types of surface measurements on $\partial\Omega_t$ are considered: (i) horizontal velocities u_1 , and (ii) normal traction $\sigma_\nu = \boldsymbol{\nu} \cdot \boldsymbol{\sigma} \boldsymbol{\nu}$. In both cases, the measurements are smoothed point observations taken at some $\mathbf{x}_j \in \partial\Omega_t$, $j = 1, \dots, n$. More precisely, using the smoothing functionals

$$l_j: H^{-\frac{1}{2}}(\partial\Omega) \rightarrow \mathbb{R}, \quad v \mapsto \int_{\partial\Omega_t} \frac{1}{\sqrt{2\pi\varepsilon^2}} \exp\left(-\frac{|\mathbf{x} - \mathbf{x}_j|^2}{2\varepsilon^2}\right) v(\mathbf{x}) \, ds, \quad j = 1, \dots, n, \quad (5.31)$$

for some fixed $\varepsilon > 0$, the horizontal velocity and normal traction measurements are given by

$$\mathbf{u}_{1,\text{data}} := [l_j(u_1)]_{j=1}^{n_1} \quad (5.32)$$

and

$$\boldsymbol{\sigma}_{\nu,\text{data}} := \left[l_{n_1+j}(\sigma_\nu) - \frac{1}{n_2} \sum_{i=1}^{n_2} l_{n_1+i}(\sigma_\nu) \right]_{j=1}^{n_2}, \quad (5.33)$$

respectively, where $n_1 + n_2 = n$. In our applications, normal traction data relates to dynamic topography, which describes elevation variations of the Earth's surface induced by mantle flow. Since such measurements are typically taken relative to each other, we enforce a zero mean in (5.33) to account for the loss of one degree of freedom.

Using the trace theorems (Theorems 2.1 and 3.8) with $\mathbf{u} \in V \subset H^1(\Omega)^2$ and $\boldsymbol{\sigma} \in H(\operatorname{div}; \Omega)$, we define the observation operator

$$\mathcal{O}: V \times H(\operatorname{div}; \Omega) \rightarrow \mathbb{R}^n \quad (5.34)$$

by concatenating the measurements (5.32) and (5.33) into one n -dimensional vector. Since the functionals l_j , $j = 1, \dots, n$, as well as the trace operators are linear and bounded, the observation operator \mathcal{O} is linear and bounded as well.

As $\mathcal{G} = \mathcal{O} \circ G \circ F$ and \mathcal{O} and G are continuous, Proposition 5.2 immediately yields:

Corollary 5.6. *Under Assumption 5.5, the parameter-to-observable map \mathcal{G} and thus the potential $\mathcal{J}(\cdot, \mathbf{y}) = \frac{1}{2} \|\mathbf{y} - \mathcal{G}(\cdot)\|_{\Gamma^{-1}}^2$ are continuous μ_0 -almost surely for any fixed $\mathbf{y} \in \mathbb{R}^n$.*

³We equip the space $V \times H(\operatorname{div}; \Omega)$ with the norm $\|(\mathbf{u}, \boldsymbol{\sigma})\|_{V \times H(\operatorname{div}; \Omega)} = \|\mathbf{u}\|_V + \|\boldsymbol{\sigma}\|_{H(\operatorname{div}; \Omega)}$.

With these preparations, we are able to establish well-posedness of the Bayesian inverse problem (5.10) by applying Theorem 2.3 of Iglesias, Lu, and Stuart (2016):

Theorem 5.7. *Let $\mathcal{J}: M \times \mathbb{R}^n \rightarrow \mathbb{R}$, $(\mathbf{m}, \mathbf{y}) \mapsto \frac{1}{2} \|\mathbf{y} - \mathcal{G}(\mathbf{m})\|_{\Gamma^{-1}}^2$. Under Assumption 5.5, the following hold:*

(a) $\mu^{\mathbf{y}} \ll \mu_0^4$ with Radon-Nikodym derivative

$$\frac{d\mu^{\mathbf{y}}}{d\mu_0}(\mathbf{m}) = \frac{\exp(-\mathcal{J}(\mathbf{m}; \mathbf{y}))}{\int_M \exp(-\mathcal{J}(\hat{\mathbf{m}}; \mathbf{y})) \mu_0(d\hat{\mathbf{m}})}, \quad (5.35)$$

where

$$\int_M \exp(-\mathcal{J}(\hat{\mathbf{m}}; \mathbf{y})) \mu_0(d\hat{\mathbf{m}}) > 0 \quad (5.36)$$

for \mathbf{y} almost surely.

(b) *The posterior distribution is locally Lipschitz continuous in the data with respect to the Hellinger distance d_{Hell} : For every $R > 0$, there is a constant $C = C(R) > 0$ such that*

$$\begin{aligned} d_{\text{Hell}}(\mu^{\mathbf{y}_1}, \mu^{\mathbf{y}_2}) &= \left[\frac{1}{2} \int_M \left(\sqrt{\frac{d\mu^{\mathbf{y}_1}}{d\mu_0}}(\mathbf{m}) - \sqrt{\frac{d\mu^{\mathbf{y}_2}}{d\mu_0}}(\mathbf{m}) \right)^2 \mu_0(d\mathbf{m}) \right]^{\frac{1}{2}} \\ &\leq C \|\mathbf{y}_1 - \mathbf{y}_2\|_{\Gamma^{-1}} \end{aligned} \quad (5.37)$$

for all $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^n$ with $\max(\|\mathbf{y}_1\|_{\Gamma^{-1}}, \|\mathbf{y}_2\|_{\Gamma^{-1}}) < R$. This implies that, for every separable Banach space S and every $g \in L^2_{\mu_0}(M; S)$, there is a constant $\hat{C} = \hat{C}(R, g) > 0$ such that⁵

$$\|\mathbb{E}_{\mu^{\mathbf{y}_1}}[g(\mathbf{m})] - \mathbb{E}_{\mu^{\mathbf{y}_2}}[g(\mathbf{m})]\|_S \leq \hat{C} \|\mathbf{y}_1 - \mathbf{y}_2\|_{\Gamma^{-1}} \quad (5.38)$$

for all $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^n$ with $\max(\|\mathbf{y}_1\|_{\Gamma^{-1}}, \|\mathbf{y}_2\|_{\Gamma^{-1}}) < R$.

PROOF. We only need to check whether the first and third condition of Assumptions 5.4 hold; due to Corollary 5.6, the second condition is valid.

First note that for arbitrary $\mathbf{m} = (h, \boldsymbol{\rho}, \boldsymbol{\eta}) \in M$, since \mathcal{O} is linear and bounded and due to Theorem 5.1 (a),

$$\begin{aligned} \|\mathcal{G}(\mathbf{m})\|_{\Gamma^{-1}} &= \|\mathcal{O} \circ G(F(\mathbf{m}))\|_{\Gamma^{-1}} \\ &\leq \tilde{C} \|G(F(\mathbf{m}))\|_{V \times H(\text{div}; \Omega)} \\ &\leq \hat{C} \left(\|F_2(h, \boldsymbol{\eta})\|_{\infty}, \text{ess inf } F_2(h, \boldsymbol{\eta}) \right) \|F_1(h, \boldsymbol{\rho})\|_{L^2} \\ &\leq C, \end{aligned} \quad (5.39)$$

⁴This means that $\mu^{\mathbf{y}}$ is *absolutely continuous* with respect to μ_0 , i.e., every null set of μ_0 is a null set of $\mu^{\mathbf{y}}$.

⁵Here, $\mathbb{E}_{\mu^{\mathbf{y}_i}}[g(\mathbf{m})] = \int_M g(\mathbf{m}) \mu^{\mathbf{y}_i}(d\mathbf{m})$.

5.5 Prior Distribution for the Level Set Function

for some constants \tilde{C} , $\hat{C}(\|F_2(h, \boldsymbol{\eta})\|_\infty, \text{ess inf } F_2(h, \boldsymbol{\eta}))$, $C > 0$, and the latter is independent of \mathbf{m} due to (5.28). Using estimate (5.39), it is straightforward to establish Assumptions 5.4: For arbitrary $R > 0$, $\mathbf{m} \in M$ and $\mathbf{y} \in \mathbb{R}^n$ with $|\mathbf{y}|_{\Gamma^{-1}} < R$,

$$0 \leq \frac{1}{2} |\mathbf{y} - \mathcal{G}(\mathbf{m})|_{\Gamma^{-1}}^2 \leq \frac{1}{2} (|\mathbf{y}|_{\Gamma^{-1}} + |\mathcal{G}(\mathbf{m})|_{\Gamma^{-1}}) < \frac{1}{2} (R + C) \quad (5.40)$$

and thus the first condition is satisfied. Furthermore, for arbitrary $R > 0$, $\mathbf{m} \in M$, and $\mathbf{y}_1, \mathbf{y}_2 \in \mathbb{R}^n$ with $\max(|\mathbf{y}_1|_{\Gamma^{-1}}, |\mathbf{y}_2|_{\Gamma^{-1}}) < R$,

$$\begin{aligned} 2 |\mathcal{J}(\mathbf{m}; \mathbf{y}_1) - \mathcal{J}(\mathbf{m}; \mathbf{y}_2)| &= \left| |\mathbf{y}_1|_{\Gamma^{-1}}^2 - |\mathbf{y}_2|_{\Gamma^{-1}}^2 - 2 \langle \mathbf{y}_1 - \mathbf{y}_2, \mathcal{G}(\mathbf{m}) \rangle_{\Gamma^{-1}} \right| \\ &\leq (|\mathbf{y}_1|_{\Gamma^{-1}} + |\mathbf{y}_2|_{\Gamma^{-1}}) (|\mathbf{y}_1|_{\Gamma^{-1}} - |\mathbf{y}_2|_{\Gamma^{-1}}) \\ &\quad + 2 |\langle \mathbf{y}_1 - \mathbf{y}_2, \mathcal{G}(\mathbf{m}) \rangle_{\Gamma^{-1}}| \\ &\leq 2R |\mathbf{y}_1 - \mathbf{y}_2|_{\Gamma^{-1}} + 2 |\mathbf{y}_1 - \mathbf{y}_2|_{\Gamma^{-1}} |\mathcal{G}(\mathbf{m})|_{\Gamma^{-1}} \\ &\leq 2(R + C) |\mathbf{y}_1 - \mathbf{y}_2|_{\Gamma^{-1}}, \end{aligned} \quad (5.41)$$

establishing the third condition. Hence, Assumptions 5.4 hold and the statement of the theorem follows from Iglesias, Lu, and Stuart (2016, Theorem 2.3). \square

In the next section, we discuss the prior distributions we choose for the simulations in Section 5.7 and the validity of Assumption 5.5.

5.5 Prior Distribution for the Level Set Function

The prior distribution for the parameter vector $\mathbf{m} = (h, \boldsymbol{\rho}, \boldsymbol{\eta})$ is assumed to be Gaussian; the parameter space M is thus a complete probability space. Additionally, we assume the level set function h and the other components of the parameter vector $(\boldsymbol{\rho}, \boldsymbol{\eta})$ to be independent. Since Assumption 5.5 only depends on h , we focus on the Gaussian prior measure of the level set function, $\mathcal{N}(h_0, \mathcal{C}_{0,h})$, where h_0 denotes the mean and $\mathcal{C}_{0,h}$ the covariance operator. In the function space setting, it is common to choose $\mathcal{C}_{0,h}$ to be a power of an inverse differential operator. This allows exploitation of fast PDE solvers for sampling from the distribution and yields control over the samples' regularity (Bui-Thanh et al., 2013; Stuart, 2010).

In particular, we define the covariance operator $\mathcal{C}_{0,h}$ on the domain Ω in terms of the PDE operator \mathcal{A} as

$$\mathcal{C}_{0,h} := \mathcal{A}^{-\alpha} := (\tau^2 I - \Delta)^{-\alpha}, \quad (5.42)$$

where I is the identity, Δ the Laplace operator, and \mathcal{A} is equipped with the boundary conditions defined below. The use of inverse elliptic operators as covariance operators is motivated by their relationship to Matérn covariance functions, and an empirical relation between τ and α to the correlation length of the resulting random fields can be found in Lindgren, Rue, and Lindström, 2011. In particular, the constant $\tau \geq 0$ controls the (inverse) correlation length, and $\alpha > 0$ determines the regularity of the samples. For example, increasing τ leads to small-scale features in realizations of the random field, whereas increasing α yields smoother samples. In the level set context,

the latter results in smoother interfaces between the different phases in physical space.

For the PDE operator \mathcal{A} , we employ homogeneous Dirichlet boundary conditions on the bottom part of the boundary, $\partial\Omega_b$, and homogeneous Neumann boundary conditions everywhere else:

$$h = 0 \quad \text{on } \partial\Omega_b, \quad (5.43)$$

$$\nabla h \cdot \boldsymbol{\nu} = 0 \quad \text{on } \partial\Omega \setminus \partial\Omega_b. \quad (5.44)$$

The implications of these prior modeling choices can be seen in Figure 5.1: When the threshold parameter satisfies $a_1 > 0$, the subdomain Ω_2 cannot reach the bottom boundary. This is desirable in our applications where we are typically only interested in the upper part of a physical domain and the bottom boundary is artificial and only included for computational reasons.

Defining $H := \{h \in H^1(\Omega) \mid h = 0 \text{ on } \partial\Omega_b\}$, the weak solution h of $\mathcal{A}h = e$ for arbitrary $e \in L^2(\Omega)$ is defined as follows: Find $h \in H$ such that

$$\int_{\Omega} \tau^2 h k \, d\mathbf{x} + \int_{\Omega} \nabla h \cdot \nabla k \, d\mathbf{x} = \int_{\Omega} e k \, d\mathbf{x} \quad (5.45)$$

for all $k \in H$. Existence of a unique solution of (5.45) follows immediately from the Lax-Milgram theorem (e.g. Evans, 1998, Chapter 6, Theorem 1).

Sampling from the prior distribution requires access to the square root of the covariance operator, $\mathcal{C}_{0,h}^{1/2} = \mathcal{A}^{-\alpha/2}$. This is realized by solving systems of the form $\mathcal{A}h = e$ subsequently $\alpha/2$ times. Hence, we limit the choices of $\alpha \geq 2$ to even integers.

Lemma 5.8. *Let $\mu_0 = \mathcal{N}(h_0, \mathcal{C}_{0,h})$ and $h_0 \in \mathcal{R}(\mathcal{C}_{0,h}^{1/2}) \cap \mathcal{C}(\overline{\Omega})$. The following holds for μ_0 -almost every h :*

- (a) $h \in \mathcal{C}(\overline{\Omega})$,
- (b) $\mathbb{1}(\{h = a_i\}) = 0$ for all $i \in \{1, \dots, m-1\}$.

PROOF. Let h be a sample of $\mathcal{N}(h_0, \mathcal{C}_{0,h})$, i.e., $h = h_0 + h_1$ for a sample h_1 of $\mathcal{N}(0, \mathcal{C}_{0,h})$. According to Dashti and Stuart (2017, Theorem 12), h_1 is almost surely continuous. Since $h_0 \in \mathcal{C}(\overline{\Omega})$, this is also true for $h = h_0 + h_1$, and (a) is established.

Following the approach outlined in Iglesias, Lu, and Stuart (2016, Appendix 2), condition (b) is satisfied for $\mathcal{N}(0, \mathcal{C}_{0,h})$ -almost every h . Due to $h_0 \in \mathcal{R}(\mathcal{C}_{0,h}^{1/2})$ and Stuart (2010, Theorem 6.13), the Gaussian measures $\mathcal{N}(0, \mathcal{C}_{0,h})$ and $\mu_0 = \mathcal{N}(h_0, \mathcal{C}_{0,h})$ are equivalent, i.e., their sets of measure zero coincide. This implies that condition (b) also holds for μ_0 -almost every h . \square

In this way, μ_0 is a Gaussian measure on $\mathcal{C}(\overline{\Omega})$ that satisfies Assumption 5.5. Therefore, the Bayesian inverse problems considered in Section 5.7.1 are well-posed in the sense of Theorem 5.7. Note that requiring the mean h_0 to be continuous is a natural assumption since the parameter space for the level set function is $\mathcal{C}(\overline{\Omega})$.

5.5.1 Incorporating Seismic Images as Prior Knowledge

Inverting for subsurface structures from surface observations is challenging. Thus, it is advantageous to incorporate any available information about these structures into the prior distribution. For example, seismic images can provide insight into the location of interfaces between different materials inside the Earth a priori. Since we do not invert for the physical fields directly, the question arises of how to translate this knowledge into prior information for the level set function. In this section, we propose a heuristic to construct the mean h_0 for the prior distribution of the level set function given a seismic image. For the presentation of this idea, we use a density field ρ with two phases and known values $\boldsymbol{\rho} = (\rho_1, \rho_2)^\top$ as in Figure 5.1.

Assume we have a seismic image of our computational domain Ω . Using this image, estimates of the prior density can be obtained through experimental data or first principle calculations. Let $\hat{\rho}: \Omega \rightarrow [\rho_1, \rho_2]$ denote this estimated density field, which we assume to be continuous — mimicking the idea (and inherent problem) that seismic images tend to be blurry and do not have sharp edges. Our aim is to construct the mean h_0 for the level set function in such a way that the expected value of the corresponding binary functions matches the seismic image $\hat{\rho}$, i.e.,

$$\hat{\rho}(\mathbf{x}) = \mathbb{E}[F_1(h, \boldsymbol{\rho})(\mathbf{x})] \quad \text{for all } \mathbf{x} \in \Omega. \quad (5.46)$$

To derive a computable expression for h_0 from (5.46), we denote the probability density function and cumulative distribution function of the standard normal distribution $\mathcal{N}(0, 1)$ as φ and Φ , respectively. Furthermore, let $c: \Omega \times \Omega \rightarrow \mathbb{R}$ be the covariance function corresponding to the prior operator $\mathcal{C}_{0,h}$. Note that $h(\mathbf{x}) \sim \mathcal{N}(h_0(\mathbf{x}), c(\mathbf{x}, \mathbf{x}))$ for arbitrary $\mathbf{x} \in \Omega$, which can be equivalently stated as

$$h(\mathbf{x}) = h_0(\mathbf{x}) + \sqrt{c(\mathbf{x}, \mathbf{x})}Z, \quad Z \sim \mathcal{N}(0, 1). \quad (5.47)$$

In the case with two phases, the level set map simplifies to

$$F_1(h, \boldsymbol{\rho})(\mathbf{x}) = \begin{cases} \rho_1, & \text{if } h(\mathbf{x}) \leq a_1, \\ \rho_2, & \text{if } h(\mathbf{x}) > a_1, \end{cases} \quad (5.48)$$

where the threshold $a_1 \in \mathbb{R}$ is given. The inequality $h(\mathbf{x}) \leq a_1$ holds if and only if $Z \leq (a_1 - h_0(\mathbf{x}))/\sqrt{c(\mathbf{x}, \mathbf{x})}$. Substituting this and (5.47) into (5.46) yields

$$\hat{\rho}(\mathbf{x}) = \mathbb{E}\left[F_1(h_0(\mathbf{x}) + \sqrt{c(\mathbf{x}, \mathbf{x})}Z, \boldsymbol{\rho})\right] \quad (5.49)$$

$$= \int_{-\infty}^{\frac{a_1 - h_0(\mathbf{x})}{\sqrt{c(\mathbf{x}, \mathbf{x})}}} \rho_1 \varphi(x) dx + \int_{\frac{a_1 - h_0(\mathbf{x})}{\sqrt{c(\mathbf{x}, \mathbf{x})}}}^{\infty} \rho_2 \varphi(x) dx \quad (5.50)$$

$$= \rho_1 \Phi\left(\frac{a_1 - h_0(\mathbf{x})}{\sqrt{c(\mathbf{x}, \mathbf{x})}}\right) + \rho_2 \left(1 - \Phi\left(\frac{a_1 - h_0(\mathbf{x})}{\sqrt{c(\mathbf{x}, \mathbf{x})}}\right)\right), \quad (5.51)$$

which can be rearranged to obtain the desired prior mean of the level set function:

$$h_0(\mathbf{x}) = a_1 + \sqrt{c(\mathbf{x}, \mathbf{x})}\Phi^{-1}\left(\frac{\hat{\rho}(\mathbf{x}) - \rho_2}{\rho_1 - \rho_2}\right) \quad \text{for all } \mathbf{x} \in \Omega. \quad (5.52)$$

Note that, in discretized form, the values $c(\mathbf{x}, \mathbf{x})$ are the diagonal values of the covariance matrix and h_0 can be computed easily.

If $h_0 \in \mathcal{R}(\mathcal{C}_{0,h}^{1/2}) \cap \mathcal{C}(\bar{\Omega})$, Lemma 5.8 is applicable and the Bayesian inverse problem using this prior distribution is well-posed. However, this cannot be guaranteed without further assumptions on the prior covariance operator $\mathcal{C}_{0,h}$ and the estimated density field $\hat{\rho}$. Nonetheless, we will demonstrate the practical use of such tailored level-set priors in the numerical examples of Section 5.7.2.

5.6 Numerical Methods

5.6.1 Discretization and Prior Distribution

The discretizations of the governing equations and the prior distribution are based on the finite element method. In all computations, the mesh consists of triangular elements. Since the domain and local mesh refinement are different for the two models in Sections 5.7.1 and 5.7.2, the exact properties are stated there.

For the discretization of the weak form of the Stokes equations (5.1), we choose continuous piecewise quadratic basis functions for the velocity components and continuous piecewise linear basis functions for the pressure. This choice of basis functions, also known as Taylor-Hood elements, is stable and a standard choice for the Stokes equations (see e.g. Elman, Silvester, and Wathen, 2014). For computational reasons, we use a different approach for the discretization of the pressure space than stated in Section 5.2. Instead of enforcing a zero mean on the pressure field (and thus coupling all pressure degrees of freedom), we remove the excess degree of freedom by setting the pressure to zero at one fixed node, resulting in a shift of the pressure field.

We discretize the weak form (5.45) of the PDE operator underlying the prior covariance using continuous piecewise linear finite elements. More information about the discretization of the relevant inner products can be found in Bui-Thanh et al., 2013. We choose the inverse length scale $\tau = 14$ and the exponent $\alpha = 8$, resulting in rather smooth level set functions with relatively large correlation lengths. Choosing (τ, α) constitutes prior modeling, in which one incorporates all prior knowledge into the prior distribution. Practically, to find appropriate (τ, α) , we visualize multiple samples from the resulting distribution and aim at choosing parameters (τ, α) such that samples represent potential models for the true parameters (see, e.g., the prior samples in Figure 5.5). Alternatively, it is possible to consider (τ, α) as hyper-parameters as in Dunlop, Iglesias, and Stuart, 2016 at the cost of increasing the computational requirements. Our implementation is based on FEniCS (Logg, Mardal, and Wells, 2012), an open-source software library for solving partial differential equations using the finite element method. FEniCS offers Python interfaces to efficient sparse direct solver tools such as MUMPS (Amestoy et al., 2001) or SuperLU (Li, 2005), which we use for the computations in Section 5.7.

5.6.2 Approximating the Posterior Distribution

Since the parameter-to-observable map \mathcal{G} is nonlinear, the posterior distribution is non-Gaussian and does not have a closed form. Using Gaussian approximations to the posterior distribution typically requires derivatives of \mathcal{G} , which are not available here as \mathcal{G} is discontinuous. Therefore, we rely on sampling methods to explore the distribution. The goal is to find a sequence of samples from the posterior measure $\mu^{\mathbf{y}}$ and use them to approximate statistical quantities like expected values, (pointwise) variances, or confidence regions of the inversion parameters. In order to construct such a sequence of samples, we employ the Metropolis-Hastings algorithm, an accept-reject mechanism yielding a Markov chain whose stationary distribution is the posterior distribution (Metropolis et al., 1953; Hastings, 1970). We briefly outline the Metropolis-Hastings algorithm and some basics of Markov chain Monte Carlo (MCMC) sampling following Geyer (2011). For this purpose, we restrict the presentation to a finite-dimensional space for better intuition and to avoid introducing additional notation and concepts.

Assume that the posterior distribution has density $\pi^{\mathbf{y}}$ with respect to the Lebesgue measure and denote the density of the conditional *proposal distribution* given \mathbf{m} as $\pi_{\text{prop}}(\mathbf{m}, \cdot)$. Given the current state $\mathbf{m}^{(i)}$ of the Markov chain, we propose a move to state $\widetilde{\mathbf{m}}^{(i)}$ using the proposal density $\pi_{\text{prop}}(\mathbf{m}^{(i)}, \cdot)$. To ensure that the stationary distribution of the resulting Markov chain is the posterior distribution, this move is only accepted with a certain probability, which is given by

$$a(\mathbf{m}^{(i)}, \widetilde{\mathbf{m}}^{(i)}) = \min \left\{ 1, \frac{\pi^{\mathbf{y}}(\widetilde{\mathbf{m}}^{(i)})\pi_{\text{prop}}(\widetilde{\mathbf{m}}^{(i)}, \mathbf{m}^{(i)})}{\pi^{\mathbf{y}}(\mathbf{m}^{(i)})\pi_{\text{prop}}(\mathbf{m}^{(i)}, \widetilde{\mathbf{m}}^{(i)})} \right\}. \quad (5.53)$$

In practice, we simulate a random number $s \sim \text{unif}(0, 1)$ and set $\mathbf{m}^{(i+1)} = \widetilde{\mathbf{m}}^{(i)}$ if $s < a(\mathbf{m}^{(i)}, \widetilde{\mathbf{m}}^{(i)})$; otherwise we reject the move and set $\mathbf{m}^{(i+1)} = \mathbf{m}^{(i)}$. The procedure is summarized in Algorithm 5.1. Note that the posterior density $\pi^{\mathbf{y}}$ only occurs as a quotient in the acceptance probability (5.53). This is very convenient since this density is only known up to a multiplicative constant and approximating the normalization constant is essentially as difficult as sampling from the posterior distribution (cf. (5.35) and (5.36)).

Once samples $(\mathbf{m}^{(i)})_{i=1}^N$ from the posterior distribution $\mu^{\mathbf{y}}$ have been obtained, statistical quantities of the form

$$\mathbb{E}_{\mu^{\mathbf{y}}}[g(\mathbf{m})], \quad (5.54)$$

where g is a real-valued function defined on the parameter space, can be estimated using a sample average approximation:

$$\mathbb{E}_{\mu^{\mathbf{y}}}[g(\mathbf{m})] = \int_M g(\mathbf{m})\mu^{\mathbf{y}}(d\mathbf{m}) \approx \frac{1}{N} \sum_{i=1}^N g(\mathbf{m}^{(i)}). \quad (5.55)$$

Typical statistical quantities of interest in (5.54) include moments of components of \mathbf{m} , which corresponds to choosing $g(\mathbf{m}) = m_j^k$ for the k -th moment of the j -th component m_j of \mathbf{m} and enables, e.g., estimation of expected values and variances.

Algorithm 5.1. Metropolis-Hastings Markov chain Monte Carlo method

- 1: **Input:** Observations \mathbf{y} , proposal density $\pi_{\text{prop}}(\cdot, \cdot)$, chain length i_{max}
 - 2: **Result:** Markov chain $(\mathbf{m}^{(i)})_i$ that is invariant with respect to $\pi^{\mathbf{y}}$
 - 3: Pick initial guess $\mathbf{m}^{(1)}$
 - 4: **for** $i = 1 : i_{\text{max}}$ **do**
 - 5: Propose $\widetilde{\mathbf{m}}^{(i)} \sim \pi_{\text{prop}}(\mathbf{m}^{(i)}, \cdot)$
 - 6: Calculate $a(\mathbf{m}^{(i)}, \widetilde{\mathbf{m}}^{(i)}) = \min \left\{ 1, \frac{\pi^{\mathbf{y}}(\widetilde{\mathbf{m}}^{(i)})\pi_{\text{prop}}(\widetilde{\mathbf{m}}^{(i)}, \mathbf{m}^{(i)})}{\pi^{\mathbf{y}}(\mathbf{m}^{(i)})\pi_{\text{prop}}(\mathbf{m}^{(i)}, \widetilde{\mathbf{m}}^{(i)})} \right\}$
 - 7: Draw random number $s \sim \text{unif}(0, 1)$
 - 8: **if** $s < a(\mathbf{m}^{(i)}, \widetilde{\mathbf{m}}^{(i)})$ **then**
 - 9: Set $\mathbf{m}^{(i+1)} = \widetilde{\mathbf{m}}^{(i)}$
 - 10: **else**
 - 11: Set $\mathbf{m}^{(i+1)} = \mathbf{m}^{(i)}$
 - 12: **end if**
 - 13: **end for**
-

Under certain conditions, a strong law of large numbers and a central limit theorem hold (e.g. Tierney, 1994; Nummelin, 1984). In particular, the approximation (5.55) converges to the expected value (5.54) and the variance of this estimator is

$$\text{Var} \left[\frac{1}{N} \sum_{i=1}^N g(\mathbf{m}^{(i)}) \right] = \frac{\text{Var}_{\mu^{\mathbf{y}}}[g(\mathbf{m})]}{N} \left(1 + 2 \sum_{i=1}^{N-1} \frac{N-i}{N} \text{Corr} [g(\mathbf{m}^{(1)}), g(\mathbf{m}^{(1+i)})] \right). \quad (5.56)$$

If the samples are independent, the correlations $\text{Corr} [g(\mathbf{m}^{(1)}), g(\mathbf{m}^{(1+i)})]$ vanish and only $N^{-1} \text{Var}_{\mu^{\mathbf{y}}}[g(\mathbf{m})]$ remains. However, subsequent samples of a Markov chain are in general not independent and the error therefore depends on the correlations between elements of the chain. Consequently, it is crucial to investigate these correlations when employing MCMC methods, which we will do in Section 5.7 by estimating the *autocorrelation function*

$$\text{acf} : \mathbb{N}_0 \rightarrow \mathbb{R}, \quad i \mapsto \text{Corr} [g(\mathbf{m}^{(1)}), g(\mathbf{m}^{(1+i)})]. \quad (5.57)$$

Choosing the proposal distribution in Algorithm 5.1 has significant influence on the autocorrelations and thus the efficiency of the sampling method. Returning to our original setting where the parameter space is infinite-dimensional and \mathcal{G} is not differentiable, we need a derivative-free sampling method that is well-defined on function spaces. This is achieved by choosing $\mathcal{N}(\sqrt{1-\beta^2}\mathbf{m}^{(i)}, \beta^2\mathcal{C}_0)$ for some *jump parameter* $\beta \in (0, 1]$, see below, as the proposal distribution, which is referred to as the *preconditioned Crank-Nicolson Markov chain Monte Carlo method* (pCN-MCMC, see Cotter et al., 2013). The pCN-MCMC algorithm satisfies the theoretical requirements stated above, yielding discretization-independent convergence of the Markov chain. In practice, however, one might still need a large number of samples for accurate approximations of the posterior distribution due to the autocorrelations.

For the pCN-MCMC method, the acceptance probability (5.53) simplifies to

$$a(\mathbf{m}^{(i)}, \widetilde{\mathbf{m}}^{(i)}) = \min \left\{ 1, \exp \left(\mathcal{J}(\mathbf{m}^{(i)}; \mathbf{y}) - \mathcal{J}(\widetilde{\mathbf{m}}^{(i)}; \mathbf{y}) \right) \right\}. \quad (5.58)$$

Therefore, one step of the algorithm requires drawing one sample $\widetilde{\mathbf{m}}^{(i)}$ from the proposal distribution $\mathcal{N}(\sqrt{1 - \beta^2}\mathbf{m}^{(i)}, \beta^2\mathcal{C}_0)$ and evaluating the data misfit $\mathcal{J}(\widetilde{\mathbf{m}}^{(i)}; \mathbf{y})$ once to calculate the acceptance probability. For the discretizations and prior parameters we choose, the latter clearly dominates the computational effort of the MCMC method — despite the need to solve four⁶ PDEs for sampling from the proposal distribution due to our choice $\alpha = 8$. However, solving these PDEs can be done very efficiently by precomputing a Cholesky factorization of the discretized prior operator \mathcal{A} and solving the systems using forward and backward substitution. Additionally, the number of unknowns in the discretized Stokes systems is significantly larger than for the prior operator due to our finite-element choices. Hence, solving the Stokes equations remains the bottleneck of the algorithm even for special cases where the Stokes operator does not depend on the inversion parameters and therefore requires only one factorization.

Since subsequent samples of the Markov chain are in general correlated, it is important to choose the jump parameter β in the proposal distribution carefully to obtain a sufficiently large effective sample size. A larger jump parameter β leads to a better exploration of the full space of the posterior distribution but will in turn decrease the acceptance rate of the samples, whereas a smaller jump parameter will increase the acceptance rate but might lead to a poor exploration of the space. Both situations result in a slow convergence and large autocorrelation times of the Markov chain. As a compromise, we adjust the jump parameter adaptively, aiming at a statistically desirable acceptance rate of 15 – 30% (cf. Roberts and Rosenthal, 1998).

5.7 Numerical Experiments

We apply the proposed method to two different scenarios with synthetic data: First, one where the data is generated using two sinkers in an otherwise homogeneous rectangular domain. This problem (discussed in Section 5.7.1) is used to develop intuition for the method. Second, we use a more realistic geophysical setup with a subduction zone as detailed in Section 5.7.2. All quantities are non-dimensional. We visualize the results, interpret them in relation to what we expect from viscous fluid dynamics, and discuss the prospects and limitations of this inference approach.

As explained in Section 5.4.1, in both scenarios the measurements are smoothed point observations taken at some \mathbf{x}_j , $j = 1, \dots, n$, located on the upper part of the boundary $\partial\Omega_t$, i.e., the Earth’s surface. The exact locations of the measurement points \mathbf{x}_j as well as the model setup used to create synthetic data for the inversion are specified for each model separately. For both models, the synthetic velocity and traction data are corrupted by 5% and 10% uncorrelated noise, respectively, to mitigate the “inverse crime” (Kaipio and Somersalo, 2005). More precisely, we define the noise covariance matrix $\mathbf{\Gamma}$ as a diagonal matrix with entries $\sigma_{\text{vel}}^2, \sigma_{\text{trac}}^2 > 0$, which only depend on the data type but not on the individual measurement points, and

⁶Recall that we need to apply the square root of the covariance operator to obtain samples from the distribution.

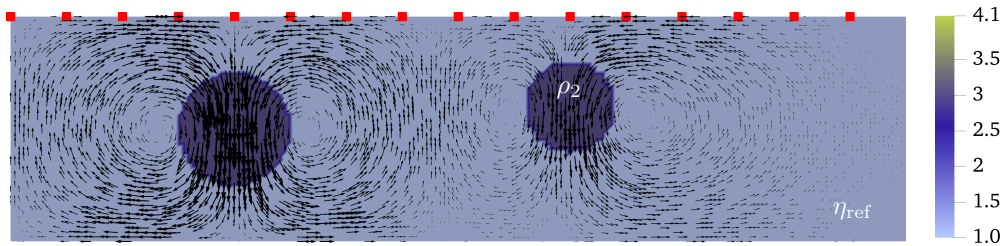


Figure 5.3 Sinker model setup: Density field ρ (in shades of blue) used to generate synthetic observations and corresponding velocity field (arrows). The measurement locations are depicted by red squares \blacksquare . Inversion parameters are the shape of the sinkers, their density, ρ_2 , and the constant background viscosity, η_{ref} .

generate “noise vectors” accordingly. In particular, we choose the standard deviations

$$\begin{aligned}\sigma_{\text{vel}} &= 0.05 \cdot \|\mathbf{u}_{1,\text{data}}\|_{\infty}, \\ \sigma_{\text{trac}} &= 0.1 \cdot \|\boldsymbol{\sigma}_{\nu,\text{data}}\|_{\infty},\end{aligned}\tag{5.59}$$

where $\mathbf{u}_{1,\text{data}}$ and $\boldsymbol{\sigma}_{\nu,\text{data}}$ denote the measurements corresponding to the “true” parameters obtained via (5.32) and (5.33), respectively, and $\|\cdot\|_{\infty}$ denotes the maximum norm. Note that the relative noise is larger for the normal traction data since these measurements are expected to be less precise in reality compared to plate velocities.

5.7.1 Model 1: Sinkers

The first model consists of two higher-density sinkers in the rectangular domain $\Omega = [0, 4] \times [0, 1]$, which is discretized with a uniform triangular mesh constructed from 200 squares in x - and 50 squares in y -direction, each of which is split into two triangles (Figure 5.3). We aim to recover the geometric structure of the density field ρ , the higher density value ρ_2 , and the constant background viscosity η_{ref} . Since the velocity is driven by the density difference between ρ_1 and ρ_2 , we consider $\rho_1 = 1$ to be known. To ensure that the density of the sinkers ρ_2 is larger than ρ_1 , we define

$$\rho_2 := \min(\rho_1 + \exp(\rho_{\log}), \rho_{\max})\tag{5.60}$$

and invert for ρ_{\log} . Using the notation of Section 5.4, this means

$$\begin{aligned}f_{1,1}(\mathbf{x}, \rho_{\log}) &:= \rho_1 = 1, \\ f_{1,2}(\mathbf{x}, \rho_{\log}) &:= \min(\rho_1 + \exp(\rho_{\log}), \rho_{\max}), \\ F_1(h, \rho_{\log}) &:= \rho_1 \mathbb{1}_{\{h < a_1\}} + f_{1,2}(\cdot, \rho_{\log}) \mathbb{1}_{\{h \geq a_1\}},\end{aligned}\tag{5.61}$$

and we choose $a_1 := 3 \cdot 10^{-10}$. The upper bound $\rho_{\max} \in \mathbb{R}_+$ can be chosen arbitrarily large such that it is never reached in practice; it just exists to ensure $f_{1,2} \in L_+^{\infty}(\Omega \times \mathbb{R})$ and the theory of Section 5.4 applies.

For the viscosity, we use

$$\eta_{\text{ref}} := f_{2,1}(\mathbf{x}, \eta_{\log}) := \max(\eta_{\min}, \min(\exp(\eta_{\log}), \eta_{\max}))\tag{5.62}$$

Model 1a: Sinkers

parameter		ρ_{\log}	η_{\log}	
true value		0.693	0	
prior	mean	1.046	4	
	st. dev.	0.325	2	
posterior	V	mean	1.085	1.11
		st. dev.	0.333	0.34
	T	mean	0.747	4.00
		st. dev.	0.310	1.98
	V+T	mean	0.857	0.01
		st. dev.	0.260	0.08

Table 5.1 Means and standard deviations (st. dev.) of the Gaussian prior distributions for the scalar parameters and their “true” values used to create synthetic data for the first sinker model (5.63). The true values of ρ_{\log} and η_{\log} correspond to the physical quantities $\rho_2 = 3$ and $\eta_{\text{ref}} = 1$. Additionally, the means and standard deviations of the posterior marginal distributions for the simulations using only velocity data (V), only traction data (T), and both data types (V+T) are given. Note that the posterior distributions are not Gaussian and therefore the posterior marginals are not fully characterized by the above values.

and invert for η_{\log} , again with some lower and upper bounds $\eta_{\min}, \eta_{\max} \in \mathbb{R}_+$ that can be chosen such that they are not relevant in practice. Note that we assume η_{ref} to be constant in the entire domain Ω and it does not depend on the level set function h .

Summarizing the above definitions, we set $M := \mathcal{C}(\overline{\Omega}) \times \mathbb{R}^2$ and

$$F(h, \rho_{\log}, \eta_{\log}) := (F_1(h, \rho_{\log}), f_{2,1}(\cdot, \eta_{\log})\mathbf{1}_{\Omega}) \quad (5.63)$$

such that the full parameter vector is $\mathbf{m} = (h, \rho_{\log}, \eta_{\log}) \in M$. The prior means and standard deviations of ρ_{\log} and η_{\log} can be found in Table 5.1. They are chosen such that the “true” values of the physical quantities ρ_2 and η_{ref} are reasonably likely under the prior distribution but at the same time model substantial prior uncertainty; e.g., the true density is a bit more than one standard deviation away from the prior mean, whereas the true viscosity is two standard deviations away from the prior mean, and this standard deviation is chosen such that the viscosity can get two orders of magnitude larger and smaller with similar probability (see also Figure 5.8).

In this example, we do not incorporate “seismic” knowledge about the subsurface into the prior mean and choose $h_0 \equiv 0$ as mean for the level set function. Synthetic data is created with the density field shown in Figure 5.3 and the “true” parameters listed in Table 5.1, where the observations are taken at 16 equidistant surface points \mathbf{x}_j indicated by red squares in Figure 5.3.

In all computations, we draw five million samples and discard the first million as burn-in. When using Markov chains to explore the posterior distribution, it is important to ensure that the chain is reasonably close to equilibrium and exhibits sufficient mixing to obtain a good approximation of the distribution. For this purpose, we estimate the autocorrelation function (5.57) for certain quantities of \mathbf{m} after the

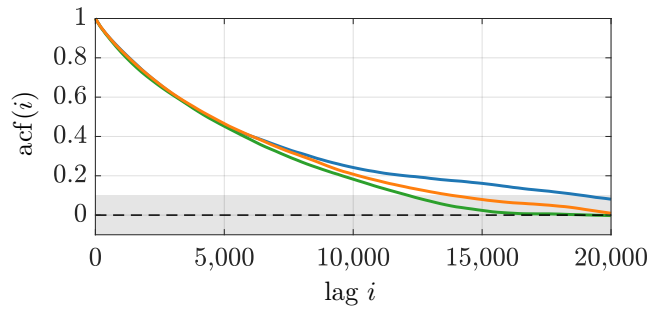


Figure 5.4 Autocorrelation estimates for certain coefficients in an expansion of the level set function h in the basis spanned by the eigenvectors of the prior covariance operator $\mathcal{C}_{0,h}$. The coefficients belong to the first (blue), second (green), and fifth (orange) largest eigenvalues. We consider samples with autocorrelation of 0.1 as statistically independent, which is visualized using the gray region.

burn-in phase. In particular, we calculate several expansion coefficients⁷ of the level set function samples in the orthonormal basis spanned by the eigenvectors of the prior covariance operator $\mathcal{C}_{0,h}$ and compare their autocorrelation functions⁸ (Figure 5.4). Considering the statistical dependence of samples for which autocorrelation drops below 0.1 as negligible, the estimates suggest that roughly every 15,000th sample of the chain is statistically independent, which shows that a large number of samples is needed for accurate approximation of the posterior distribution (cf. (5.56)).

Comparing samples from the prior and posterior distribution of the density field, the horizontal location of the sinkers is recovered well, whereas there is some variance in the vertical location and the sizes of the sinkers (Figure 5.5). This is expected since the measurements at the surface are mostly determined by the sinkers' mass — a larger lower-density sinker excites almost the same surface flow as a smaller higher-density sinker. Furthermore, small artifacts with higher density can appear close to the bottom of the domain. Again, this is due to the observations being taken on the surface and since higher density material leads to less flow when closer to the bottom of the domain due to the no-outflow boundary condition at the bottom boundary.

We study the role of different observation data on the recovered density field comparing three different settings: (i) using only horizontal velocities, (ii) using only normal tractions, and (iii) using both horizontal velocities and normal tractions as data. For cases (i) and (ii), measurements are taken at a refined grid of 32 equidistant points on the surface, which include the 16 points shown in Figure 5.3 that are used for case (iii). Hence, the total number of measurements is the same for all three cases.

In Figure 5.6, we compare the expected values $\mathbb{E}[F_1(h, \rho_{\log})]$ of the density field and its pointwise variance for the prior distribution and the posterior distributions corresponding to the different observation data. The pointwise prior variance shows

⁷These coefficients are sometimes referred to as the *Karhunen-Loève modes* or just *KL modes*.

⁸Note that this corresponds to defining g in (5.57) as the map from level set function to KL modes.

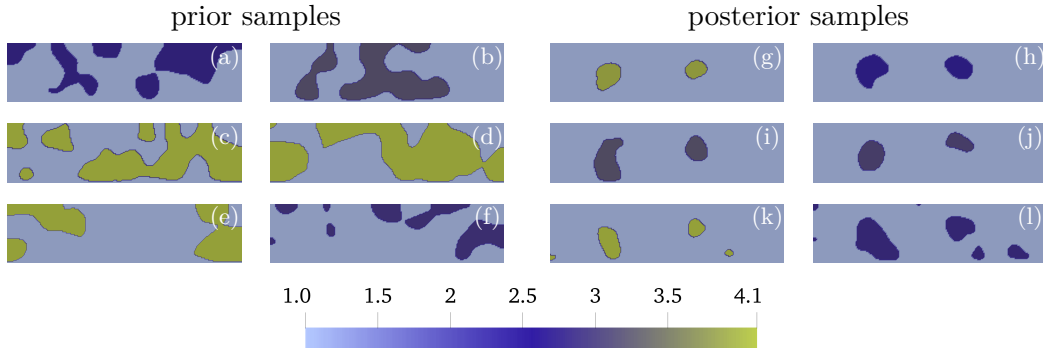


Figure 5.5 Sample density fields $F_1(h, \rho_{\log})$ from the prior distribution (left two columns, (a)–(f)) and from the posterior distribution (right two columns, (g)–(l)). Both data types were used in the inversion.

that there is no information about the location of the sinkers a priori, except near the bottom boundary. This is due to our choice of the prior that assumes sinkers are unlikely to be found there. Without normal traction data, only the horizontal locations of the sinkers are recoverable with their size and depth extent remaining largely uncertain. Parameter recovery is better when only normal traction is used as data, although there are uncertainties mainly in the exact shape of the sinkers. Combining both data types yields the best result. Shape and locations of the sinkers are recovered and the uncertainties are reduced significantly. The advantage of combining both data types is further supported by the prior and posterior distributions of the sinkers’ mass (Figure 5.7), i.e., the sinker area multiplied by its density. The “true” mass is recovered with small uncertainty, particularly compared to the cases with only one type of observation data. Note that this highlights another perk of Bayesian inversion: We are able to approximate probability distributions for quantities of interest like the sinkers’ mass on the fly and are thus able to study uncertainties in the inferred values, which are not accessible in a purely deterministic inversion.

Finally, the two-dimensional joint marginal distributions of the background viscosity η_{ref} and the sinkers’ density ρ_2 (Figure 5.8) are visualized through the prior probability density function (pdf) as well as kernel density estimates of the posterior pdfs. Using both data types, the background viscosity is constrained well (Figure 5.8c). As expected, this uncertainty is larger for the case with just velocity data, in which we also observe a positive correlation between the background viscosity and the density, in accordance with physical expectations (Figure 5.8a). In contrast to these cases, we do not learn much about η_{ref} or ρ_2 when using only normal tractions as data (Figure 5.8b). However, one should keep in mind that this is a two-dimensional marginal distribution of a high-dimensional distribution — in fact, using normal tractions as data allowed us to infer the shapes and locations of the sinkers reasonably well (Figure 5.6e–f).

Note that despite there being substantial uncertainty in the inferred density for all three observation data types (Figure 5.8), we observe that the uncertainty in the overall sinker mass is small (Figure 5.7). This is due to the different sinker shapes

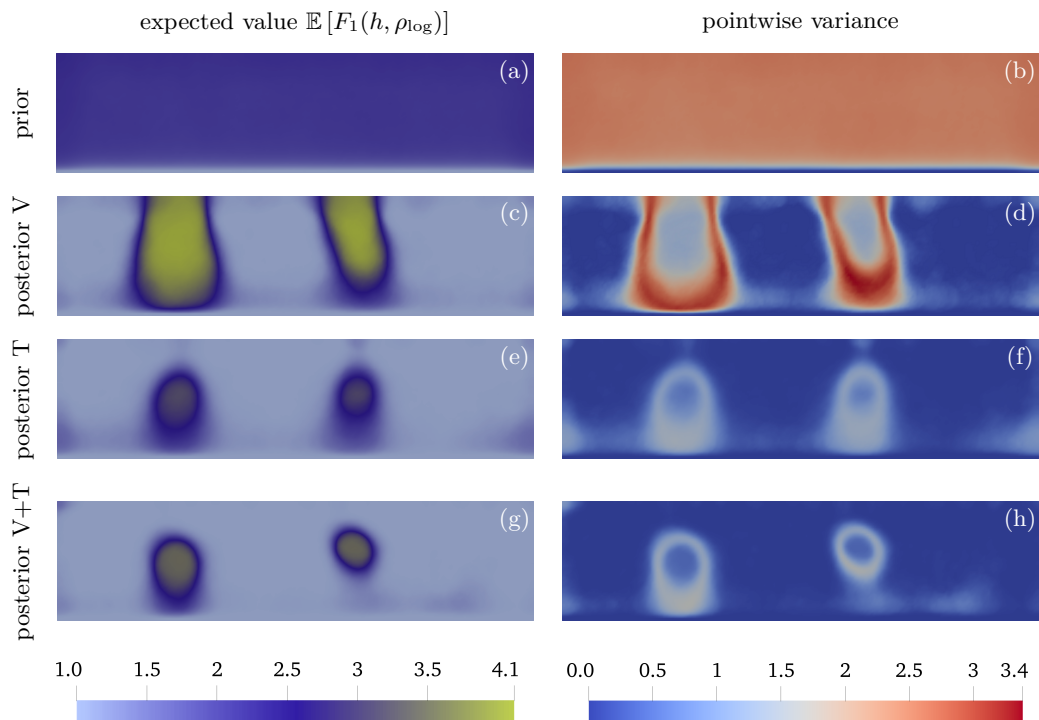


Figure 5.6 Expected value and pointwise variance of the density field's prior and posterior distributions for the first sinker model (5.63). The first row shows the prior (a) mean and (b) pointwise variance, the other rows show the posterior mean and pointwise variance for different data types: (c)–(d) only horizontal velocities (V), (e)–(f) only normal traction (T), (g)–(h) both horizontal velocities and normal traction (V+T).

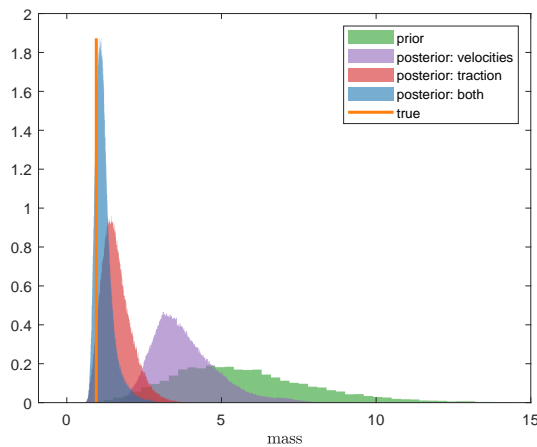


Figure 5.7 Prior and posterior distributions of the sinkers' mass depending on the type of observation data used in the inference.

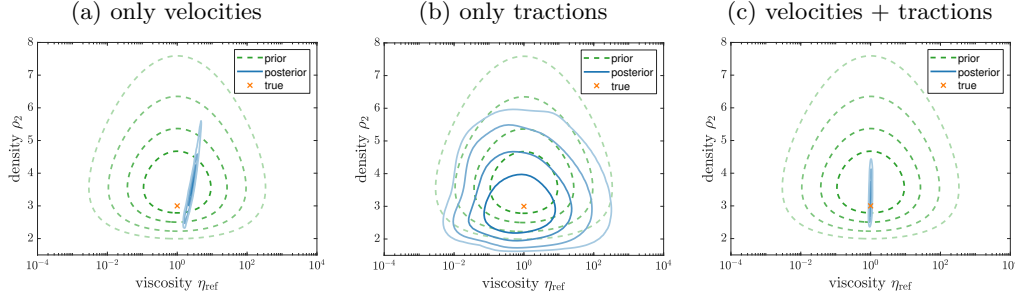


Figure 5.8 Comparison of joint marginal distributions of the viscosity η_{ref} and density ρ_2 for different data types: (a) only horizontal velocities, (b) only normal tractions, and (c) both horizontal velocities and normal tractions. The contour lines indicate highest density regions containing (from dark to light): 10%, 25%, 50%, 75% of the total probability mass.

whose size is anticorrelated to the density, as can also be seen from the posterior samples (Figure 5.5g–l).

Model 1b: Sinkers with Varying Viscosity

In a similar model setup, we perform computations where, in addition to the geometric structure of the density field, the structure of the viscosity field is also unknown. However, their shapes are coupled using one level set function for both fields. We then invert for the sinker viscosity, η_2 , instead of a constant background viscosity. More precisely,

$$F_2(h, \eta_{\log}) := f_{2,1}(\cdot, \eta_{\log}) \mathbb{1}_{\{h < a_1\}} + f_{2,2}(\cdot, \eta_{\log}) \mathbb{1}_{\{h \geq a_1\}}, \quad (5.64)$$

where $\eta_1 = f_{2,1} \equiv 1$ and

$$\eta_2 = f_{2,2}(\mathbf{x}, \eta_{\log}) := 100 \max(\eta_{\min}, \min(\exp(\eta_{\log}), \eta_{\max})), \quad (5.65)$$

similar to η_{ref} in (5.62), such that

$$F(h, \rho_{\log}, \eta_{\log}) := (F_1(h, \rho_{\log}), F_2(h, \eta_{\log})) \quad (5.66)$$

instead of (5.63). The prior distribution of $\mathbf{m} = (h, \rho_{\log}, \eta_{\log})$ is the same as in the previous example. Since the geometrical structures of the density and viscosity fields coincide, the quantities $\mathbb{E}[F_1(h, \rho_{\log})]$ and $\mathbb{E}[F_2(h, \eta_{\log})]$ yield the same qualitative information about the sinkers' shapes, which is why we focus on visualizing $\mathbb{E}[F_1(h, \rho_{\log})]$ in Figure 5.9 for consistency between the simulations. Regardless of the data type used, the locations of the sinkers are well recovered. On the other hand, neither the sinkers' density nor viscosity can be well constrained (Table 5.2), leading to more uncertainty around the edges of the sinkers compared to the other model presented in this section — the pointwise variances in this region are roughly 1.62 in Figure 5.6h and 2.65 in Figure 5.9h. The sinkers' mass, however, is again inferred with low uncertainty.

Model 1b: Sinkers

parameter		ρ_{\log}	η_{\log}	
true value		0.693	0	
prior	mean	1.046	4	
	st. dev.	0.325	2	
posterior	V	mean	1.076	3.82
		st. dev.	0.215	2.17
	T	mean	1.085	3.98
		st. dev.	0.264	2.00
	V+T	mean	1.082	3.53
		st. dev.	0.232	2.10

Table 5.2 Means and standard deviations (st. dev.) of the Gaussian prior distributions for the scalar parameters and their “true” values used to create synthetic data for the second sinker model (5.66). The true values of ρ_{\log} and η_{\log} correspond to the physical quantities $\rho_2 = 3$ and $\eta_2 = 100$. Additionally, the means and standard deviations of the posterior marginal distributions for the simulations using only velocity data (V), only traction data (T), and both data types (V+T) are given. Note that the posterior distributions are not Gaussian and therefore the posterior marginals are not fully characterized by the above values.

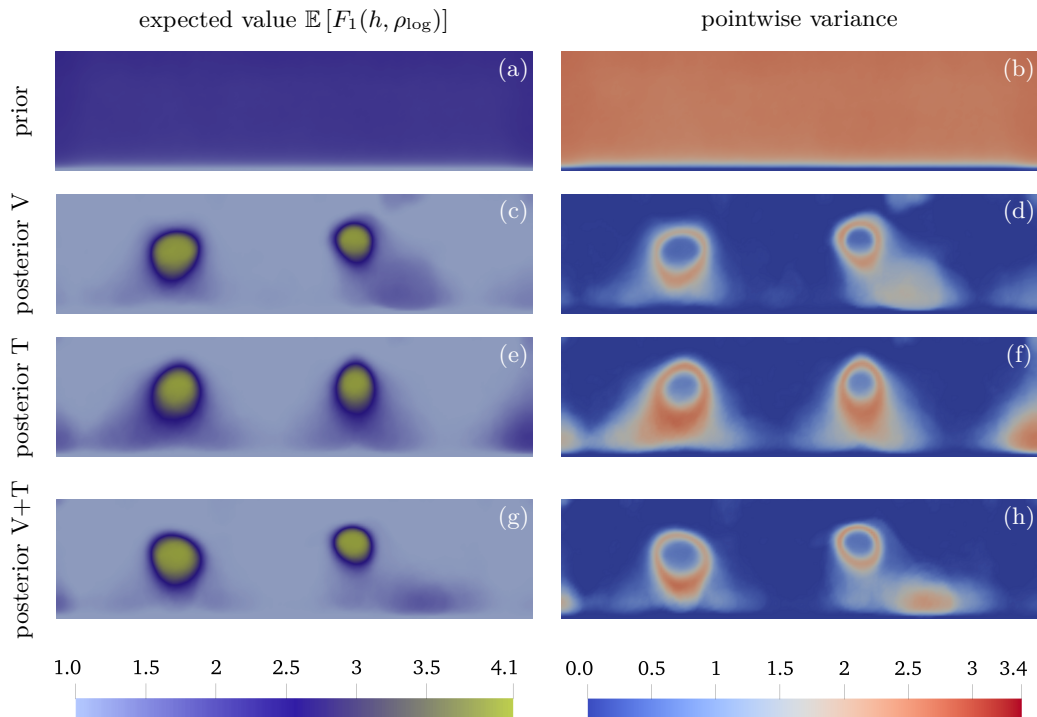


Figure 5.9 Expected value and pointwise variance of the density field’s prior and posterior distributions for the second sinker model (5.66). The first row shows the prior (a) mean and (b) pointwise variance, the other rows show the posterior mean and pointwise variance for different data types: (c)–(d) only horizontal velocities (V), (e)–(f) only normal traction (T), (g)–(h) both horizontal velocities and normal traction (V+T).

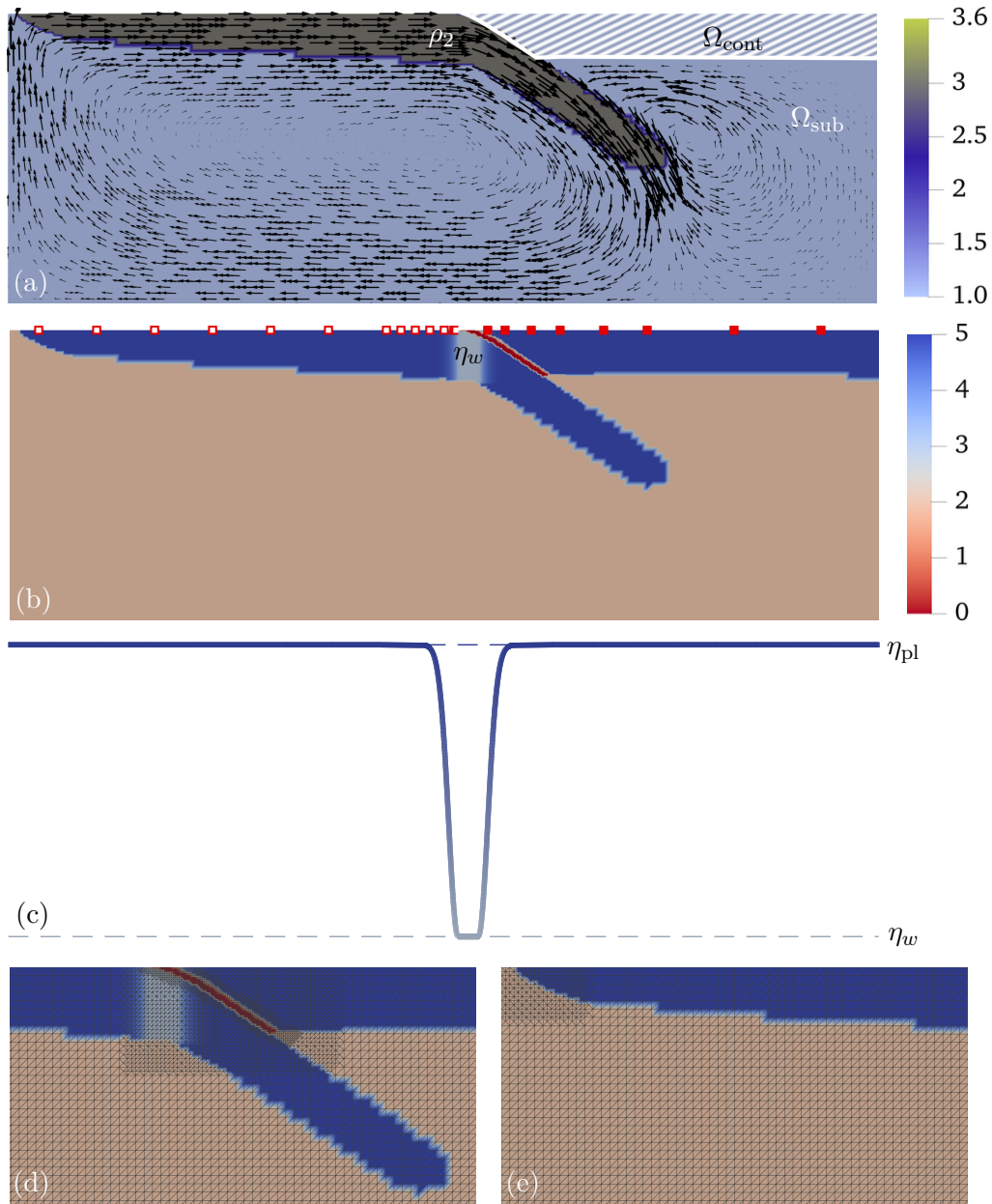


Figure 5.10 Model setup for Section 5.7.2: (a) density field ρ with corresponding velocity field (arrows) and indication of subdomains Ω_{cont} (hatched region) and $\Omega_{\text{sub}} = \Omega \setminus \Omega_{\text{cont}}$, (b) log-viscosity field $\log_{10}(\eta)$ with mesh refinement in (d) the plate collision zone including weak zone, and (e) the ridge (top left corner of domain). The weakening mechanism (5.70) for the subducting plate's viscosity is visualized in (c); the x -axis is aligned with (b) and the colors match the colorbar above. The measurement locations are shown in (b): normal tractions at red squares \blacksquare , horizontal velocities at white squares \square . Inversion parameters are the shape of the subducting plate, the subducting plate's density, ρ_2 , and the weakened viscosity in the hinge zone, η_w .

5.7.2 Model 2: Subduction Zone

The second model describes the collision of two tectonic plates, where a heavier oceanic plate (referred to as the *subducting plate* or *slab*) dives underneath a *continental plate* (sometimes referred to as the *over-riding plate*), mainly caused by negative buoyancy of the subducting plate (Figure 5.10a–b). Due to the large stresses in the collision zone, the material weakens locally — a *weak zone* is formed, modeled by the thin dark red layer between the two plates in Figure 5.10b. Without this weak zone, the highly viscous plates would just “stick” to each other and hardly move at all. Another characteristic feature of tectonic plates is that they move almost like rigid bodies, again due to their extreme viscosities. This behavior can be observed in the velocity field shown in Figure 5.10a, where the magnitudes of the subducting plate’s velocities remain almost the same throughout the entire plate. However, for this effect to occur, the plate needs to “bend down” around the collision zone, meaning that there is another (not as severe) weakening happening in the so-called *hinge zone*, which is visualized in light blue inside the subducting plate in Figure 5.10b. The goal of the inverse problem in this section is to recover the subducting plate’s shape, its density ρ_2 , and the viscosity inside the hinge zone, η_w .

The domain $\Omega = [0, 3] \times [0, 1]$ is discretized with a mesh that is locally refined around the ridge (top left corner of the domain) and hinge zone, and further refined around the weak zone (Figure 5.10d–e). As the locations and shapes of continental plates are typically well constrained, we assume that the properties of the over-riding plate and the weak zone are known. We then invert for both the density field ρ and viscosity field η in the remaining part of the domain, which we refer to as $\Omega_{\text{sub}} \subset \Omega$. The known parts of the density and viscosity fields in the continental plate and weak zone are denoted by $\rho_{\text{cont}} \in L_+^\infty(\Omega_{\text{cont}})$ and $\eta_{\text{cont}} \in L_+^\infty(\Omega_{\text{cont}})$, respectively, where $\Omega_{\text{cont}} = \Omega \setminus \Omega_{\text{sub}}$ (see Figure 5.10a).

We assume that the geometric structures of ρ and η in Ω_{sub} represent the subducting plate, so we use a single level set function h for the shape of both fields. Using the same exponential parameterization of the density in terms of ρ_{\log} as in Section 5.7.1,

$$\rho_2 = f_{1,2}(\mathbf{x}, \rho_{\log}) = \min(\rho_1 + \exp(\rho_{\log}), \rho_{\max}) \quad (5.67)$$

with $\rho_1 = f_{1,1}(\mathbf{x}, \rho_{\log}) \equiv 1$ and $\rho_{\max} \in \mathbb{R}_+$ known, we write

$$F_1(h, \rho_{\log}) := \begin{cases} \rho_1 \mathbb{1}_{\{h < a_1\}} + f_{1,2}(\cdot, \rho_{\log}) \mathbb{1}_{\{h \geq a_1\}} & \text{on } \Omega_{\text{sub}}, \\ \rho_{\text{cont}} & \text{on } \Omega_{\text{cont}}, \end{cases} \quad (5.68)$$

again choosing $a_1 := 3 \cdot 10^{-10}$. Furthermore, we consider the background viscosity

$$\eta_1 := f_{2,1}(\mathbf{x}, w_{\log}) \equiv 10^2 \quad (5.69)$$

to be known and parameterize the subducting plate’s viscosity as

$$\begin{aligned} \eta_2(\mathbf{x}) &:= f_{2,2}(\mathbf{x}, w_{\log}) \\ &:= \max \left[w_{\min}, \min \left(1, 1 - [1 - \exp(w_{\log})] \exp(-\gamma(x_1 - \xi_1)) \right) \right] \eta_{\text{pl}}, \end{aligned} \quad (5.70)$$

where the reference viscosity $\eta_{\text{pl}} = 10^5$ is given and the prescribed (and continuous) function $\gamma \geq 0$ controls the size, location, and smoothness of the hinge zone (cf. Rudi,

parameter		ρ_{\log}	w_{\log}	
true value		0.693	-4.605	
prior	mean	1.046	-3.368	
	st. dev.	0.325	1.112	
posterior	V	mean	1.227	-3.538
		st. dev.	0.301	1.109
	T	mean	0.684	-3.397
		st. dev.	0.134	1.103
	V+T	mean	0.713	-4.348
		st. dev.	0.114	0.554

Table 5.3 Means and standard deviations (st. dev.) of the Gaussian prior distributions for the scalar parameters and their “true” values used to create synthetic data for the plate subduction model. The true values of ρ_{\log} and w_{\log} correspond to the physical quantities $\rho_2 = 3$ and $\eta_w = 1000$. Additionally, the means and standard deviations of the posterior marginal distributions for the simulations using only velocity data (V), only traction data (T), and both data types (V+T) are given. Note that the posterior distributions are not Gaussian and therefore the posterior marginals are not fully characterized by the above values.

2019, Definition 2.5.1). The factor in front of η_{pl} models the weakening of the plate and is thus forced between w_{\min} and 1 for some given minimal weakening factor $0 < w_{\min} < 1$. Furthermore, we choose a smooth transition from plate viscosity η_{pl} to hinge zone viscosity $\eta_w = \min_{\mathbf{x}} \eta_2(\mathbf{x})$ that depends on the distance of the x_1 -coordinate to the (horizontal) center $\xi_1 = 1.585$ of the hinge zone; the spatial weakening and the corresponding viscosity of the subducting plate are visualized in Figure 5.10b–c.

Although the physical parameter of interest is η_w , we invert for w_{\log} to allow for sufficiently large variance in this value, leading to the full parameter vector $\mathbf{m} = (h, \rho_{\log}, w_{\log})$ on the space $M := \mathcal{C}(\bar{\Omega}) \times \mathbb{R}^2$. The level set map is thus defined via

$$F(h, \rho_{\log}, w_{\log}) := (F_1(h, \rho_{\log}), F_2(h, w_{\log})), \quad (5.71)$$

where

$$F_2(h, w_{\log}) := \begin{cases} \eta_1 \mathbf{1}_{\{h < a_1\}} + f_{2,2}(\cdot, w_{\log}) \mathbf{1}_{\{h \geq a_1\}} & \text{on } \Omega_{\text{sub}}, \\ \eta_{\text{cont}} & \text{on } \Omega_{\text{cont}}. \end{cases} \quad (5.72)$$

Data are synthesized using the density and viscosity fields visualized in Figure 5.10 with the “true” scalar parameters of Table 5.3. Normal tractions are measured at all 20 squares depicted on the over-riding and subducting plates, whereas horizontal velocities are observed at 12 open squares located only on the subducting plate (Figure 5.10b) because the continental plate does not move horizontally in our model due to the boundary conditions.

Prior means and standard deviations of the scalar parameters are chosen in a similar way as in Section 5.7.1 and are given in Table 5.3. For this inversion, we construct a non-zero mean h_0 for the prior distribution of the level set function



Figure 5.11 Representative sample of the Markov chain using zero prior mean for the level set function. The density field does not resemble the “true” shape of the subducting plate. Both data types were used for the simulation.

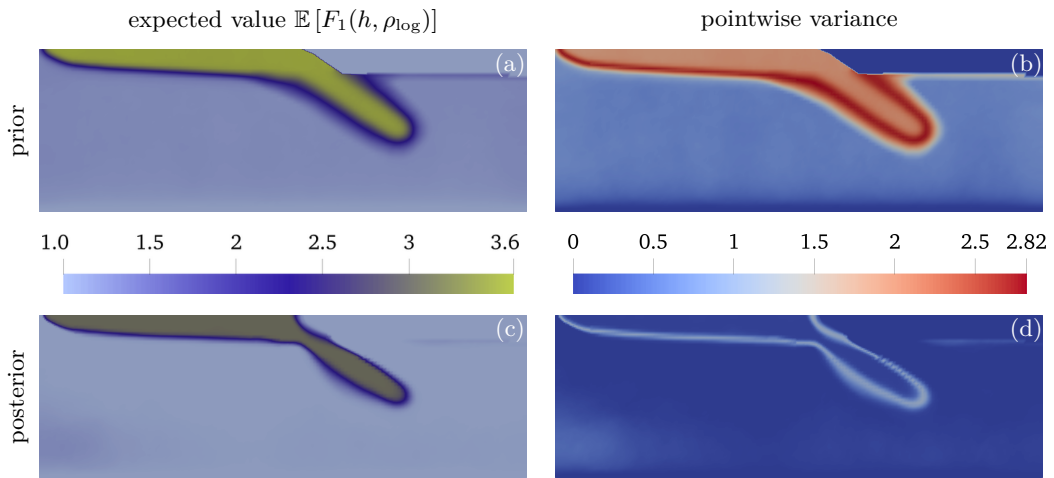


Figure 5.12 Expected value and pointwise variance of the density field’s prior and posterior distribution. The first row shows the prior (a) mean and (b) pointwise variance, the second row shows the posterior (c) mean and (d) pointwise variance. For this simulation, both plate velocities and normal tractions were used as observations.

as described in Section 5.5.1. We also ran simulations with a zero prior mean for comparison. The elements of the chain obtained using the latter approach were not near the “true” parameter values despite fitting the data reasonably well. Figure 5.11 shows a representative sample of this simulation. Clearly, it does not resemble the shape of the “true” subducting plate, supporting the need for a tailored level-set prior mean to infer reasonable values.

In all computations, we draw five million samples and discard the first million as burn-in. Again, several KL modes of the level set function samples are inspected to check for sufficient mixing and to assess whether the chain is close to equilibrium. The estimated autocorrelation time for these quantities is about 15,000, similar to the results in Section 5.7.1.

Comparing prior and posterior mean and pointwise variance of the density field ρ , the shape of the slab as well as its density are recovered well and the pointwise variance is significantly reduced (Figure 5.12). At first sight, the tailored prior

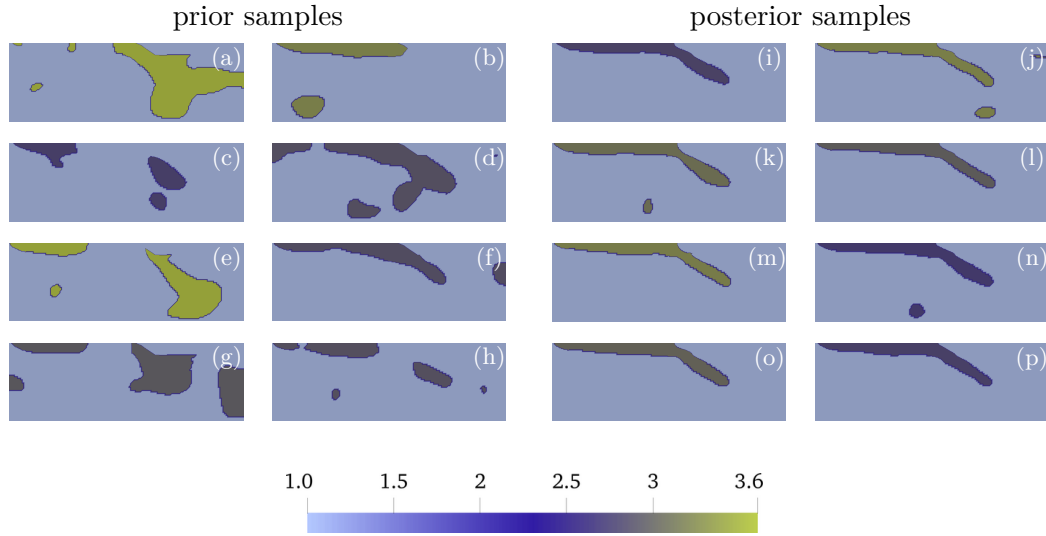


Figure 5.13 Sample density fields $F_1(h, \rho_{\log})$ from the (a)–(h) prior distribution (left two columns) and (i)–(p) posterior distribution (right two columns). Both data types were used for the inversion.

mean and corresponding pointwise variance (Figure 5.12a–b) seem to encode strong assumptions on the geometric structures. However, comparing samples from the prior and posterior distributions (Figure 5.13) shows this is not the case. The prior density samples mostly do not resemble plates and have a large variance in the plate’s density ρ_2 . On the other hand, the shapes of the posterior density samples closely resemble the true subducting plate with less variance in ρ_2 . As observed in the sinker model in Section 5.7.1, higher-density artifacts can appear close to the bottom of the domain because their influence on surface measurements is small.

If just velocity data is used, little information is learned about the scalar parameters ρ and η_w (Figure 5.14a and Table 5.3). However, using only traction data, the density is substantially narrowed to values around the truth (Figure 5.14b). On the other hand, the viscosity of the hinge zone is not much changed from the prior marginal in this case either. By combining the two data types, the recovery of the density (from the traction) allows for the viscosity of the hinge zone to be much better estimated. The final marginal from the combined data is firmly centered on the true value (Figure 5.14c). Nevertheless, the shape of the plate does differ somewhat from the true shape within the hinge zone with a slight down warping of the top side of the slab (Figure 5.12c). In addition, the top side of the slab has a wider zone of higher pointwise variance (Figure 5.12d). This is consistent with earlier work that has investigated the trade-offs between plate velocity, viscosity, and thickness of the plate (Conrad and Hager, 1999).

Finally, we study the influence of different data types in the situation where we prescribe the “true” shape of the subducting plate and the only unknown parameters are η_w and ρ_2 (Figure 5.14d–f). This reduces the dimension of the parameter space drastically, and as expected in all three cases the parameters are better constrained than when geometric shapes are uncertain. Using only velocity data, there is a

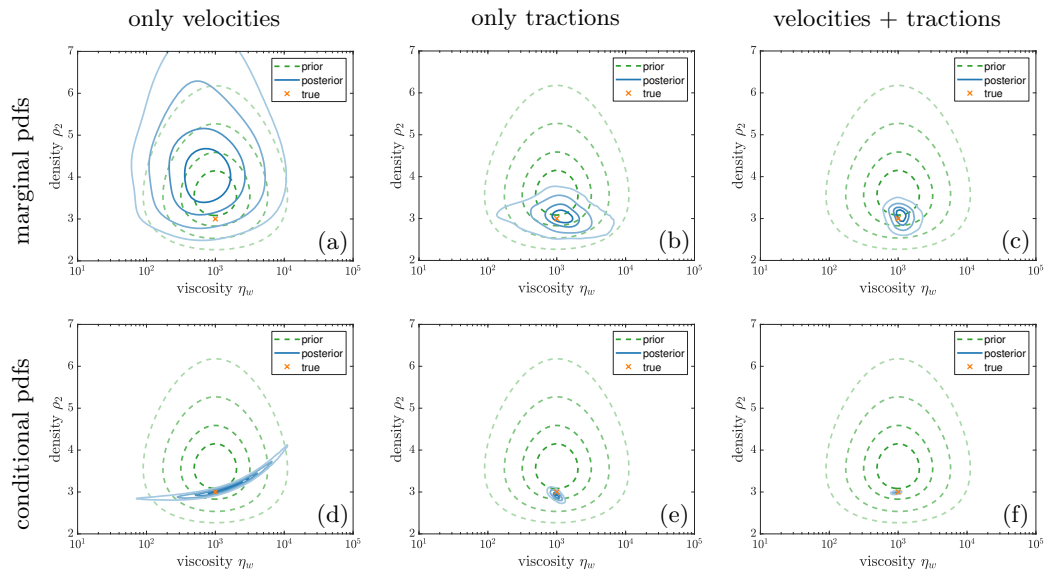


Figure 5.14 Comparison of joint marginal and conditional distributions for different data types. The contour lines indicate highest density regions containing (from dark to light): 10%, 25%, 50%, 75% of the total probability mass. The first row shows the joint marginal distributions of the viscosity η_w in the hinge zone and density ρ_2 of the subducting plate, where the columns indicate the different data types used: (a) only plate velocities, (b) only normal tractions, and (c) both plate velocities and normal tractions. The second row (d)–(f) depicts the conditional distributions of viscosity and density given the “true” shape of the subducting plate.

nonlinear positive correlation between the hinge zone viscosity and subducting plate’s density, which illustrates the trade-off between these parameters well. If the density is higher (and thus a larger force is pulling the plate), the hinge zone viscosity does not need to be as small for the same bending to occur, and vice versa. In the computation with only normal traction data, the parameters are very well constrained. Moreover, the correlation between the parameters changes its sign. As before, combining the data types yields the best result, in this case with rather low uncertainty.

5.8 Conclusion and Outlook

We have demonstrated the potential of using Bayesian level set methods to infer subsurface structures and material parameters in Stokes flow and quantify their uncertainties. Through construction of tailored level-set priors, we have formulated a method to incorporate geometric knowledge from seismic imaging in the inversion while imposing few restrictions on the specific shapes. The method, furthermore, enables inversion for density and rheological parameter fields, complementing seismic imaging methods. In the two numerical examples we showed, the shape of two sinkers as well as the shape of a subducting plate were inferred with their uncertainties, along with densities and viscosity parameters. Additionally, we found trade-offs between material parameters and geometrical subsurface structures that are consistent with earlier work and insight from fluid dynamics. The results of both models demonstrate the benefits of combining two different surface data types, horizontal velocities and normal tractions, which reduced the uncertainties significantly compared to using more measurements of the same data type. For a detailed discussion and interpretation from a geological point of view, we refer to the journal version of this chapter (Holbach, Gurnis, and Stadler, 2023).

In one of the presented model problems we used locally refined meshes to resolve the subduction physics of the setup. This refinement was static, i.e., it was kept the same for all samples. An extension of the present method could be to use dynamic refinement that aims at resolving the level set interface, which changes in each sampling step. However, this would increase the cost per MCMC sample substantially, in particular for the case where the viscosity is constant and thus the Stokes discretization matrix only has to be assembled and factorized *once*, since these factors can be reused for all samples.

A focus on linear Stokes equations allowed for a clearer presentation of the underlying ideas. While the method is applicable to models with nonlinear viscosities, this becomes computationally more challenging due to the significantly higher computational cost per sample — recall that each sample requires solving the Stokes equations. Due to the high-dimensional parameter space and the resulting slow decay of autocorrelation times, many samples are needed to approximate the posterior distribution. More efficient sampling methods can potentially alleviate the effect to some extent. A Gaussian approximation to the posterior distribution can be employed to enable faster sampling or as a direct approximation to the distribution (Pinski et al., 2015; Bui-Thanh et al., 2013). Although derivative information is typically required for these approaches and the non-differentiable parameter-to-observable map would need to be adjusted, a derivative-free method to generate samples from a Gaussian

approximation to the posterior distribution was proposed recently (Garbuno-Inigo et al., 2020). Alternatively, one can employ a multilevel MCMC method (Lykkegaard et al., 2023), which exploits a hierarchy of increasingly complex models for efficient sampling of the complex posterior distribution. This seems particularly promising because it potentially improves two of the above issues: First, one could use a differentiable approximation of the level set map on the coarser levels, which would allow for the use of efficient derivative-based sampling methods like DILI (Cui, Law, and Marzouk, 2016) on these levels. Second, the degree of nonlinearity could also vary in the hierarchy of the models, with a (near-)linear rheology on the coarsest level, additionally reducing the cost per sample.

When inverting for the shape of the subducting plate, it became evident that a tailored level-set prior representing “seismic” knowledge is crucial when dealing with more realistic models, as we were not able to obtain meaningful results without a tailored prior mean. Constraints on the prior mean could come from traditional body wave seismic tomographic images (Xue and Allen, 2007; Zhao, Hasegawa, and Horiuchi, 1992). To construct a target density mean for the prior distribution, an approximate mapping from seismic velocities to densities is required. This is often accomplished through thermodynamic relations determined experimentally or from first principle calculations, such as through the framework in Stixrude and Lithgow-Bertelloni, 2005. Taken together, there are several surface data and seismic controls within individual subduction zones, which is promising for an application of the presented method to problems with real-world data.

List of Symbols

\mathbb{N}	set of all positive integers $\{1, 2, \dots\}$
\mathbb{N}_0	set of all non-negative integers $\{0, 1, 2, \dots\}$
\mathbb{R}	set of all real numbers
\mathbb{R}_+	set of all positive real numbers
$U \times V$	Cartesian product of the two sets U and V
\mathbb{R}^n	n -fold Cartesian product of the real numbers \mathbb{R}
\mathbf{e}_i	i -th standard unit vector of \mathbb{R}^n in Cartesian coordinates
$\mathbb{R}^{n \times m}$	set of all real matrices $\mathbf{E} = (e_{ij})_{i,j=1}^{n,m}$ with n rows and m columns
$\mathbf{x}^\top, \mathbf{E}^\top$	transpose of a vector \mathbf{x} or matrix \mathbf{E} , respectively
$\mathbf{x} \cdot \mathbf{y}$	Euclidean inner product between the vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$
$\mathbf{E} : \mathbf{E}^*$	Frobenius inner product between the matrices $\mathbf{E}, \mathbf{E}^* \in \mathbb{R}^{n \times m}$
$ \cdot $	depending on the argument, either the absolute value $ x = \sqrt{x^2}$ of a scalar $x \in \mathbb{R}$, the Euclidean norm $ \mathbf{x} = \sqrt{\mathbf{x} \cdot \mathbf{x}}$ of a vector $\mathbf{x} \in \mathbb{R}^n$, or the Frobenius norm $ \mathbf{E} = \sqrt{\mathbf{E} : \mathbf{E}}$ of a matrix $\mathbf{E} \in \mathbb{R}^{n \times m}$
$\bar{A}, \partial A$	closure and boundary of the set A with respect to the topology of the space it is contained in
$\text{ess inf } A$	essential infimum of the set $A \subset \mathbb{R}$
$\text{ess sup } A$	essential supremum of the set $A \subset \mathbb{R}$
V'	dual space of the normed vector space V
\rightharpoonup	weak convergence
\rightharpoonup^*	weak* convergence
$\langle u, v \rangle$	dual pairing between $u \in U$ and $v \in V$; if V is a Hilbert space, the notation coincides with the inner product on $V \times V$
$\langle u, v \rangle_{\mathcal{W}}$	weighted inner product $\langle u, v \rangle_{\mathcal{W}} = \langle u, \mathcal{W}v \rangle$, where $(V, \langle \cdot, \cdot \rangle)$ is a Hilbert space and $\mathcal{W}: V \rightarrow V$ is a self-adjoint positive definite operator
$\ v\ _{\mathcal{W}}$	weighted norm $\ v\ _{\mathcal{W}} = \sqrt{\langle v, v \rangle_{\mathcal{W}}}$, $v \in V$, with the same notations as above; if $V = \mathbb{R}^n$ and $\mathcal{W} = \mathbf{W} \in \mathbb{R}^{n \times n}$ is symmetric and positive definite, we always use the notation $ \mathbf{v} _{\mathbf{W}}$ for $\mathbf{v} \in \mathbb{R}^n$ to stress that this is a norm on a finite-dimensional space
$\mathbb{1}_A$	indicator function of the set A
$f _A$	restriction of a function $f: U \rightarrow V$ to a subset $A \subset U$
$f \equiv c$	short for $f(x) = c$ for all x in the domain of f
$\mathcal{N}(f)$	nullspace of the map f
$\mathcal{R}(f)$	range of the map f

List of Symbols

$\arg \min_{x \in U} f(x)$	the set of minimizers of $f: U \rightarrow \mathbb{R}$; interpreted as an element of U if the minimizer is unique
$\partial^k / \partial x_j$	k -th partial derivative with respect to the variable x_j
$\nabla \mathbf{f}$	gradient of the function $\mathbf{f}: \Omega \subseteq \mathbb{R}^d \rightarrow \mathbb{R}^n$ with entries $(\nabla \mathbf{f})_{ij} = \partial f_j / \partial x_i$, i.e., $\nabla \mathbf{f}$ is the transpose of the Jacobian of \mathbf{f} ; analogously for scalar functions
f'	short for the derivative of a function f of only one variable
\mathbb{A}	Lebesgue measure
a.e.	almost everywhere
$\int_{\Omega} \cdots d\mathbf{x}$	Lebesgue integral
$\int_{\partial\Omega} \cdots ds$	surface integral
$\mathcal{C}(\Omega; V)$	set of all continuous functions $f: \Omega \rightarrow V$, where $\Omega \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ or $V \subseteq \mathbb{R}^{l \times m}$; abbreviated to $\mathcal{C}(\Omega)$ if $V = \mathbb{R}$
$\mathcal{C}^k(\Omega; V)$	set of all $f \in \mathcal{C}(\Omega; V)$ that are k times continuously (partially) differentiable, $k \in \mathbb{N}_0 \cup \{\infty\}$, with the convention $\mathcal{C}^0(\Omega; V) = \mathcal{C}(\Omega; V)$; abbreviated to $\mathcal{C}^k(\Omega)$ if $V = \mathbb{R}$
$\mathcal{C}_c^k(\Omega; V)$	set of all $f \in \mathcal{C}^k(\Omega; V)$ with compact support in Ω ; abbreviated to $\mathcal{C}_c^k(\Omega)$ if $V = \mathbb{R}$
$\mathcal{C}_0(\Omega; V)$	set of all $f \in \mathcal{C}(\Omega; V)$ that vanish on the boundary, $\mathcal{C}_0(\Omega; V) = \overline{\mathcal{C}_c(\Omega; V)}^{\ \cdot\ _{\infty}}$, where $\ f\ _{\infty} = \max_{\mathbf{x} \in \Omega} f(\mathbf{x}) $; abbreviated to $\mathcal{C}_0(\Omega)$ if $V = \mathbb{R}$
$L_{\mu}^r(M; S)$	Bochner space of all (equivalence classes of) μ -measurable functions $f: M \rightarrow S$ such that $\ f\ _S^r$ is integrable with respect to the measure μ , where $(M, \ \cdot\ _M)$ and $(S, \ \cdot\ _S)$ are separable Banach spaces and $1 \leq r < \infty$; equipped with the norm $\ f\ _{L_{\mu}^r} = (\int_M \ f(m)\ _S^r \mu(dm))^{1/r}$
$L_{\mu}^{\infty}(M; S)$	Bochner space of all (equivalence classes of) μ -measurable functions $f: M \rightarrow S$ that are essentially bounded; equipped with the norm $\ f\ _{\infty} = \text{ess sup}_{m \in M} \ f(m)\ _S$
$L^r(\Omega)^{m \times n}$	Lebesgue space: alternative notation for $L_{\mu}^r(M; S)$, $1 \leq r \leq \infty$, if $M = \Omega \subseteq \mathbb{R}^d$, $\mu = \mathbb{A}$ is the Lebesgue measure, and $(S, \ \cdot\ _S) = (\mathbb{R}^{m \times n}, \cdot)$; analogously $L^r(\Omega)^n$ for $S = \mathbb{R}^n$ and $L^r(\Omega)$ for $S = \mathbb{R}$
$L_{+}^{\infty}(\Omega)$	subset of all functions $f \in L^{\infty}(\Omega)$ that are essentially bounded from below by a strictly positive constant, i.e., $\text{ess inf}_{\mathbf{x} \in \Omega} f(\mathbf{x}) > 0$
$W^{1,r}(\Omega)^d$	Sobolev space of all functions $\mathbf{f} \in L^r(\Omega)^d$ such that all weak first partial derivatives $\partial f_i / \partial x_j$ exist and lie in $L^r(\Omega)$, where $1 \leq r < \infty$; equipped with the norm $\ \mathbf{f}\ _{W^{1,r}} = (\ \mathbf{f}\ _{L^r}^r + \ \nabla \mathbf{f}\ _{L^r}^r)^{1/r}$
$W_0^{1,r}(\Omega)^d$	subspace of all functions $\mathbf{f} \in W^{1,r}(\Omega)^d$ that vanish on the boundary, $W_0^{1,r}(\Omega)^d = \overline{\mathcal{C}_c^{\infty}(\Omega; \mathbb{R}^d)}^{\ \cdot\ _{W^{1,r}}}$

$H^1(\Omega)^d$	short for $W^{1,2}(\Omega)^d$
$H(\operatorname{div}; \Omega)$	subset of all functions $\mathbf{f} \in L^2(\Omega)^d$ such that $\operatorname{div} \mathbf{f} \in L^2(\Omega)$
$X \sim \mu$	the random variable X is distributed according to the probability distribution μ
$\mathbb{E}[X]$	expected value of the random variable X
$\operatorname{Var}[X]$	variance of the random variable X
$\frac{d\nu}{d\mu}$	Radon-Nikodym derivative of the measure ν with respect to the measure μ

List of Figures

1.1	Examples of objective functions evaluated on a regular grid in the parameter space, one without regularization and one using Tikhonov regularization.	5
2.1	Sketch of a potential polygonal domain $\Omega \subset \mathbb{R}^2$ as in the proof of Lemma 2.3.	24
3.1	Viscous stresses τ_{II} as a function of the strain rates $\dot{\epsilon}_{II}$ at some fixed point.	35
3.2	Graph of $g(\mathbf{x}, \cdot)$ for fixed $\mathbf{x} \in \Omega$ as a function of $ \mathbf{E} $	37
3.3	Sketch of the approximations in the proof of Proposition 3.11.	50
4.1	Principal stress orientation around a borehole according to Anderson's assumption.	60
4.2	Material parameters discretization scheme.	66
4.3	Model setup of a sinking cube for the PSD inversions.	69
4.4	Forward problem results for the homogeneous cube: principal stress directions and velocity field.	69
4.5	Stress direction field in three planes at different heights in case of a homogeneous cube.	70
4.6	Stress direction field in three planes at different heights in the heterogeneous case.	70
4.7	Inversion results for linear forward problem.	71
4.8	Inversion results for nonlinear forward problem.	72
4.9	Staggered-grid finite difference discretization stencil for the strain rate term in the adjoint-based gradient.	75
5.1	Visualization of the level set map for the simple case with two phases and constant coefficients.	82
5.2	Examples of realizable and non-realizable piecewise constant functions using the level set parameterization.	83
5.3	Sinker model setup including measurement locations.	96
5.4	Autocorrelation estimates for certain KL modes of the level set function.	98
5.5	Sample density fields from the prior and posterior distributions for the first sinker model (5.63).	99
5.6	Expected value and pointwise variance of the density field's prior and posterior distributions for the first sinker model (5.63).	100
5.7	Prior and posterior distributions of the sinkers' mass depending on the type of observation data used in the inference.	100

List of Figures

5.8	Comparison of joint marginal distributions of the scalar material parameters in the first sinker model for different data types.	101
5.9	Expected value and pointwise variance of the density field's prior and posterior distributions for the second sinker model (5.66).	102
5.10	Plate subduction model setup including measurement locations.	103
5.11	Representative sample of the Markov chain using zero prior mean for the level set function in the plate subduction model.	106
5.12	Expected value and pointwise variance of the density field's prior and posterior distribution for the plate subduction model.	106
5.13	Sample density fields from the prior and posterior distributions for the plate subduction model.	107
5.14	Comparison of joint marginal and conditional distributions of the scalar material parameters in the plate subduction model for different data types.	108

List of Tables

2.1	Rheological parameters and corresponding function spaces.	20
4.1	“True” material parameters for PSD inversion.	68
5.1	Means and standard deviations of the prior and posterior distributions for the scalar parameters and their “true” values used to create synthetic data for the first sinker model (5.63).	97
5.2	Means and standard deviations of the prior and posterior distributions for the scalar parameters and their “true” values used to create synthetic data for the second sinker model (5.66).	102
5.3	Means and standard deviations of the prior and posterior distributions for the scalar parameters and their “true” values used to create synthetic data for the plate subduction model.	105

Bibliography

- Adams, R. A. and Fournier, J. J. F. (2003). *Sobolev Spaces*. 2nd ed. Vol. 140. Pure and Applied Mathematics. San Diego: Academic Press. ISBN: 0120441438.
- Alisic, L. et al. (2012). “Multi-scale dynamics and rheology of mantle flow with plates”. *Journal of Geophysical Research: Solid Earth* vol. 117, B10402. DOI: 10.1029/2012jb009234.
- Alt, H. W. (2016). *Linear Functional Analysis*. London: Springer. ISBN: 9781447172802.
- Ambrosio, L., Fusco, N., and Pallara, D. (2000). *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. London: Oxford University Press. ISBN: 9780198502456.
- Amestoy, P. R. et al. (2001). “A Fully Asynchronous Multifrontal Solver Using Distributed Dynamic Scheduling”. *SIAM Journal on Matrix Analysis and Applications* vol. 23, pp. 15–41. DOI: 10.1137/S0895479899358194.
- Amrouche, C. and Girault, V. (1994). “Decomposition of vector spaces and application to the Stokes problem in arbitrary dimension”. *Czechoslovak Mathematical Journal* vol. 44, pp. 109–140. URL: <http://eudml.org/doc/31399> (visited on 01/13/2024).
- Amrouche, C. and Rejaiba, A. (2014). “Lp-theory for Stokes and Navier–Stokes equations with Navier boundary condition”. *Journal of Differential Equations* vol. 256, pp. 1515–1547. DOI: 10.1016/j.jde.2013.11.005.
- Balay, S. et al. (2023). *PETSc/TAO Users Manual*. Tech. rep. Argonne National Laboratory. DOI: 10.2172/2205494.
- Batchelor, G. K. (2000). *An Introduction to Fluid Dynamics*. Cambridge: Cambridge University Press. ISBN: 9780511800955.
- Baumann, T. S., Kaus, B. J., and Popov, A. A. (2014). “Constraining effective rheology through parallel joint geodynamic inversion”. *Tectonophysics* vol. 631, pp. 197–211. DOI: 10.1016/j.tecto.2014.04.037.
- Baumann, T. and Kaus, B. J. (2015). “Geodynamic inversion to constrain the non-linear rheology of the lithosphere”. *Geophysical Journal International* vol. 202, pp. 1289–1316. DOI: 10.1093/gji/ggv201.
- Bertsekas, D. (1982). *Constrained optimization and Lagrange multiplier methods*. New York: Academic Press. ISBN: 0120934809.

Bibliography

- Boettinger, W. J. et al. (2002). “Phase-field simulation of solidification”. *Annual review of materials research* vol. 32, pp. 163–194. DOI: 10.1146/annurev.matsci.32.101901.155803.
- Bonito, A., DeVore, R. A., and Nochetto, R. H. (2013). “Adaptive Finite Element Methods for Elliptic Problems with Discontinuous Coefficients”. *SIAM Journal on Numerical Analysis* vol. 51, pp. 3106–3134. DOI: 10.1137/130905757.
- Breit, D. and Diening, L. (2011). “Sharp Conditions for Korn Inequalities in Orlicz Spaces”. *Journal of Mathematical Fluid Mechanics* vol. 14, pp. 565–573. DOI: 10.1007/s00021-011-0082-x.
- Brezis, H. (2011). *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. New York: Springer. ISBN: 9780387709147.
- Bui-Thanh, T. et al. (2013). “A Computational Framework for Infinite-Dimensional Bayesian Inverse Problems Part I: The Linearized Case, with Application to Global Seismic Inversion”. *SIAM Journal on Scientific Computing* vol. 35, A2494–A2523. DOI: 10.1137/12089586X.
- Bunge, H.-P., Hagelberg, C. R., and Travis, B. J. (2003). “Mantle circulation models with variational data assimilation: inferring past mantle flow and structure from plate motion histories and seismic tomography”. *Geophysical Journal International* vol. 152, pp. 280–301. DOI: 10.1046/j.1365-246x.2003.01823.x.
- Bunge, H.-P., Richards, M. A., and Baumgardner, J. R. (1996). “Effect of depth-dependent viscosity on the planform of mantle convection”. *Nature* vol. 379, pp. 436–438. DOI: 10.1038/379436a0.
- Burstedde, C. et al. (2008). “Scalable Adaptive Mantle Convection Simulation on Petascale Supercomputers”. In: *Proceedings of the 2008 ACM/IEEE Conference on Supercomputing*. SC '08. Austin: IEEE Press. ISBN: 9781424428359.
- Cai, M. et al. (2014). “Efficient Variable-Coefficient Finite-Volume Stokes Solvers”. *Communications in Computational Physics* vol. 16, pp. 1263–1297. DOI: 10.4208/cicp.070114.170614a.
- Carpio, A., Iakunin, S., and Stadler, G. (2020). “Bayesian approach to inverse scattering with topological priors”. *Inverse Problems* vol. 36, p. 105001. DOI: 10.1088/1361-6420/abaa30.
- Cesari, L. (1956). *Surface Area. (AM-35)*. Princeton: Princeton University Press. ISBN: 9781400882328.
- Conrad, C. P. and Hager, B. H. (1999). “Effects of plate bending and fault strength at subduction zones on plate dynamics”. *Journal of Geophysical Research: Solid Earth* vol. 104, pp. 17551–17571. DOI: 10.1029/1999JB900149.

- Cotter, S. L. et al. (2013). “MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster”. *Statistical Science* vol. 28, pp. 424–446. DOI: 10.1214/13-sts421.
- Cui, T., Law, K. J., and Marzouk, Y. M. (2016). “Dimension-independent likelihood-informed MCMC”. *Journal of Computational Physics* vol. 304, pp. 109–137. DOI: 10.1016/j.jcp.2015.10.008.
- Dacorogna, B. (2007). *Direct Methods in the Calculus of Variations*. New York: Springer. ISBN: 9780387552491.
- Dashti, M. and Stuart, A. M. (2017). “The Bayesian Approach to Inverse Problems”. In: *Handbook of Uncertainty Quantification*. Ed. by Ghanem, R., Higdon, D., and Owhadi, H. Cham: Springer International Publishing, pp. 311–428. DOI: 10.1007/978-3-319-12385-1_7.
- Dorn, O. and Lesselier, D. (2006). “Level set methods for inverse scattering”. *Inverse Problems* vol. 22, R67–R131. DOI: 10.1088/0266-5611/22/4/R01.
- (2009). “Level set methods for inverse scattering—some recent developments”. *Inverse Problems* vol. 25, p. 125001. DOI: 10.1088/0266-5611/25/12/125001.
- Drucker, D. C. and Prager, W. (1952). “Soil mechanics and plastic analysis or limit design”. *Quarterly of Applied Mathematics* vol. 10, pp. 157–165. URL: <http://www.jstor.org/stable/43633942> (visited on 01/13/2024).
- Dunlop, M. M., Iglesias, M. A., and Stuart, A. M. (2016). “Hierarchical Bayesian level set inversion”. *Statistics and Computing* vol. 27, pp. 1555–1584. DOI: 10.1007/s11222-016-9704-8.
- Ekeland, I. and Temam, R. (1999). *Convex Analysis and Variational Problems*. Philadelphia: Society for Industrial and Applied Mathematics. ISBN: 9781611971088.
- Elman, H., Silvester, D., and Wathen, A. (2014). *Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics*. Oxford: Oxford University Press. ISBN: 9780199678808.
- Elstrodt, J. (2011). *Maß- und Integrationstheorie*. 7th ed. Berlin: Springer. ISBN: 9783642179051.
- Engl, H. W., Hanke, M., and Neubauer, A. (1996). *Regularization of Inverse Problems*. Dordrecht: Springer. ISBN: 9780792341574.
- Evans, L. (1998). *Partial differential equations*. Providence: American Mathematical Society. ISBN: 9780821807729.
- Fichtner, A. (2011). *Full Seismic Waveform Modelling and Inversion*. Berlin: Springer. ISBN: 9783642158070.

Bibliography

- Fullsack, P. (1995). “An arbitrary Lagrangian-Eulerian formulation for creeping flows and its application in tectonic models”. *Geophysical Journal International* vol. 120, pp. 1–23. DOI: 10.1111/j.1365-246X.1995.tb05908.x.
- Garbuno-Inigo, A. et al. (2020). “Interacting Langevin Diffusions: Gradient Structure and Ensemble Kalman Sampler”. *SIAM Journal on Applied Dynamical Systems* vol. 19, pp. 412–441. DOI: 10.1137/19m1251655.
- Gere, J. (2004). *Mechanics of materials*. 6th ed. Melbourne: Thomson Learning. ISBN: 9780534417932.
- Gerya, T. V. and Yuen, D. A. (2007). “Robust characteristics method for modelling multiphase visco-elasto-plastic thermo-mechanical problems”. *Physics of the Earth and Planetary Interiors* vol. 163, pp. 83–105. DOI: 10.1016/j.pepi.2007.04.015.
- Geyer, C. J. (2011). “Introduction to Markov Chain Monte Carlo”. In: *Handbook of Markov Chain Monte Carlo*. Ed. by Brooks, S. et al. Boca Raton: Chapman & Hall/CRC, pp. 3–48. DOI: 10.1201/b10905.
- Geymonat, G., Suquet, P., and Nedelec, J. C. (1986). “Functional Spaces for Norton-Hoff materials”. *Mathematical Methods in the Applied Sciences* vol. 8, pp. 206–222. DOI: 10.1002/mma.1670080113.
- Giga, Y. and Novotný, A., eds. (2018). *Handbook of Mathematical Analysis in Mechanics of Viscous Fluids*. Cham: Springer. ISBN: 9783319133447.
- Giles, M. B. and Pierce, N. A. (2000). “An introduction to the adjoint approach to design”. *Flow, turbulence and combustion* vol. 65, pp. 393–415. DOI: 10.1023/A:1011430410075.
- Girault, V. and Raviart, P.-A. (1986). *Finite element methods for Navier-Stokes equations: theory and algorithms*. Berlin: Springer-Verlag. ISBN: 9783642648885.
- Glowinski, R. and Xu, J. (2011). *Numerical Methods for Non-Newtonian Fluids*. 1st ed. Vol. 16. Handbook of Numerical Analysis. Amsterdam: Elsevier. ISBN: 9780444530479.
- Gurnis, M. and Hager, B. H. (1988). “Controls of the structure of subducted slabs”. *Nature* vol. 335, pp. 317–321. DOI: 10.1038/335317a0.
- Han, W. and Reddy, B. D. (2013). *Plasticity: mathematical theory and numerical analysis*. New York: Springer. ISBN: 9781461459408.
- Hanke, M. (2017). *A Taste of Inverse Problems*. Philadelphia: Society for Industrial and Applied Mathematics. ISBN: 9781611974942.
- Harlow, F. H. and Welch, J. E. (1965). “Numerical Calculation of Time-Dependent Viscous Incompressible Flow of Fluid with Free Surface”. *Physics of Fluids* vol. 8, pp. 2182–2189. DOI: 10.1063/1.1761178.

- Hastings, W. K. (1970). “Monte Carlo sampling methods using Markov chains and their applications”. *Biometrika* vol. 57, pp. 97–109. DOI: 10.1093/biomet/57.1.97.
- Heidbach, O. et al. (2018). “The World Stress Map database release 2016: Crustal stress pattern across scales”. *Tectonophysics* vol. 744, pp. 484–498. DOI: 10.1016/j.tecto.2018.07.007.
- Heister, T. et al. (2017). “High accuracy mantle convection simulation through modern numerical methods – II: realistic models and problems”. *Geophysical Journal International* vol. 210, pp. 833–851. DOI: 10.1093/gji/ggx195.
- Hill, R. (1998). *The mathematical theory of plasticity*. Vol. 11. London: Oxford University Press. ISBN: 9780198503675.
- Hintermüller, M. and Stadler, G. (2006). “An Infeasible Primal-Dual Algorithm for Total Bounded Variation–Based Inf-Convolution-Type Image Restoration”. *SIAM Journal on Scientific Computing* vol. 28, pp. 1–23. DOI: 10.1137/040613263.
- Hinze, M. et al. (2008). *Optimization with PDE constraints*. Mathematical Modelling: Theory and Applications. New York: Springer. ISBN: 9781402088391.
- Holbach, L., Gurnis, M., and Stadler, G. (2023). “A Bayesian level set method for identifying subsurface geometries and rheological properties in Stokes flow”. *Geophysical Journal International* vol. 235, pp. 260–272. DOI: 10.1093/gji/ggad220.
- Horbach, A., Bunge, H.-P., and Oeser, J. (2014). “The adjoint method in geodynamics: derivation from a general operator formulation and application to the initial condition problem in a high resolution mantle circulation model”. *GEM - International Journal on Geomathematics* vol. 5, pp. 163–194. DOI: 10.1007/s13137-014-0061-5.
- Howell, J. S. and Walkington, N. J. (2011). “Inf-sup conditions for twofold saddle point problems”. *Numerische Mathematik* vol. 118, pp. 663–693. DOI: 10.1007/s00211-011-0372-5.
- Huber, P. J. (1964). “Robust Estimation of a Location Parameter”. *The Annals of Mathematical Statistics* vol. 35, pp. 73–101. DOI: 10.1214/aoms/1177703732.
- Hundertmark-Zaušková, A., Lukáčová-Medvidová, M., and Nečasová, Š. (2016). “On the existence of weak solution to the coupled fluid-structure interaction problem for non-Newtonian shear-dependent fluid”. *Journal of the Mathematical Society of Japan* vol. 68, pp. 193–243. DOI: 10.2969/jmsj/06810193.
- Hutter, K. (1983). *Theoretical Glaciology*. Mathematical Approaches to Geophysics. Dordrecht: Kluwer Academic. ISBN: 9789027714732.

Bibliography

- Iglesias, M., Lu, Y., and Stuart, A. (2016). “A Bayesian level set method for geometric inverse problems”. *Interfaces and Free Boundaries* vol. 18, pp. 181–217. DOI: 10.4171/ifb/362.
- Isaac, T., Stadler, G., and Ghattas, O. (2015). “Solution of Nonlinear Stokes Equations Discretized By High-Order Finite Elements on Nonconforming and Anisotropic Meshes, with Application to Ice Sheet Dynamics”. *SIAM Journal on Scientific Computing* vol. 37, B804–B833. DOI: 10.1137/140974407.
- Ismail-Zadeh, A. T. et al. (2003). “Three-dimensional numerical simulation of the inverse problem of thermal convection”. *Computational Mathematics and Mathematical Physics* vol. 43, pp. 581–599.
- Jouvet, G. and Rappaz, J. (2011). “Analysis and Finite Element Approximation of a Nonlinear Stationary Stokes Problem Arising in Glaciology”. *Advances in Numerical Analysis* vol. 2011, pp. 1–24. DOI: 10.1155/2011/164581.
- Kadu, A., Leeuwen, T. van, and Mulder, W. A. (2017). “Salt Reconstruction in Full-Waveform Inversion With a Parametric Level-Set Method”. *IEEE Transactions on Computational Imaging* vol. 3, pp. 305–315. DOI: 10.1109/TCI.2016.2640761.
- Kaipio, J. and Somersalo, E. (2005). *Statistical and Computational Inverse Problems*. New York: Springer. ISBN: 9780387271323.
- Kaus, B. J. P. (2016). “Constraining lithospheric flow”. *Science* vol. 353, pp. 1495–1496. DOI: 10.1126/science.aai8442.
- Kaus, B. et al. (2016). “Forward and inverse modelling of lithospheric deformation on geological timescales”. In: *NIC Symposium 2016 - Proceedings*. Ed. by Binder, K., Müller, M., and Schnurpfeil, A. Vol. 48. NIC Series. Jülich: Forschungszentrum Jülich, pp. 299–307. URL: <http://hdl.handle.net/2128/10411> (visited on 01/13/2024).
- Kaus, B. J. (2010). “Factors that control the angle of shear bands in geodynamic numerical models of brittle deformation”. *Tectonophysics* vol. 484. Quantitative modelling of geological processes, pp. 36–47. DOI: 10.1016/j.tecto.2009.08.042.
- Kohn, R. V. (1979). *New estimates for deformations in terms of their strains: I) Estimates of Wirtinger type for nonlinear strains; II) Functions whose linearized strains are measures*. Dissertation. Princeton University. URL: <https://www.proquest.com/openview/4540c34765832d04f849a6411b67e26f> (visited on 01/13/2024).
- Kreemer, C., Holt, W. E., and Haines, A. J. (2003). “An integrated global model of present-day plate motions and plate boundary deformation”. *Geophysical Journal International* vol. 154, pp. 8–34. DOI: 10.1046/j.1365-246X.2003.01917.x.
- Ladyzhenskaya, O. A. (2003). “Sixth problem of the millennium: Navier-Stokes equations, existence and smoothness”. *Russian Mathematical Surveys* vol. 58, pp. 251–286. DOI: 10.1070/RM2003v058n02ABEH000610.

- Lanzendörfer, M. (2011). *Flows of incompressible fluids with pressure-dependent viscosity (and their application to modelling the flow in journal bearing)*. Dissertation. Univerzita Karlova, Matematicko-fyzikální fakulta. URL: https://msekcce.karlin.mff.cuni.cz/~lanz/documents/LanzM_PhDthesis_2011_mono.pdf (visited on 01/13/2024).
- Li, W. and Qian, J. (2016). “Joint inversion of gravity and traveltime data using a level-set-based structural parameterization”. *GEOPHYSICS* vol. 81, G107–G119. DOI: 10.1190/geo2015-0547.1.
- Li, X. S. (2005). “An Overview of SuperLU: Algorithms, Implementation, and User Interface”. *ACM Transactions on Mathematical Software* vol. 31, pp. 302–325. DOI: 10.1145/1089014.1089017.
- Lindgren, F., Rue, H., and Lindström, J. (2011). “An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* vol. 73, pp. 423–498. DOI: 10.1111/j.1467-9868.2011.00777.x.
- Liu, L., Spasojevic, S., and Gurnis, M. (2008). “Reconstructing Farallon plate subduction beneath North America back to the Late Cretaceous”. *Science* vol. 322, pp. 934–938. DOI: 10.1126/science.1162921.
- Logg, A., Mardal, K.-A., and Wells, G., eds. (2012). *Automated solution of differential equations by the finite element method: The FEniCS book*. Berlin: Springer. ISBN: 9783642230998.
- Lykkegaard, M. B. et al. (2023). “Multilevel Delayed Acceptance MCMC”. *SIAM/ASA Journal on Uncertainty Quantification* vol. 11, pp. 1–30. DOI: 10.1137/22M1476770.
- Málek, J. et al. (1996). *Weak and measure-valued solutions to evolutionary PDEs*. Applied Mathematics. London: CRC Press. ISBN: 9780412577505.
- May, D. and Moresi, L. (2008). “Preconditioned iterative methods for Stokes flow problems arising in computational geodynamics”. *Physics of the Earth and Planetary Interiors* vol. 171, pp. 33–47. DOI: 10.1016/j.pepi.2008.07.036.
- McKenzie, D. (1984). “The Generation and Compaction of Partially Molten Rock”. *Journal of Petrology* vol. 25, pp. 713–765. DOI: 10.1093/petrology/25.3.713.
- Metropolis, N. et al. (1953). “Equation of state calculations by fast computing machines”. *The Journal of Chemical Physics* vol. 21, pp. 1087–1092. DOI: 10.1063/1.1699114.
- Moresi, L. et al. (2007). “Computational approaches to studying non-linear dynamics of the crust and mantle”. *Physics of the Earth and Planetary Interiors* vol. 163. Computational Challenges in the Earth Sciences, pp. 69–82. DOI: 10.1016/j.pepi.2007.06.009.

Bibliography

- Muir, J. B. and Tsai, V. C. (2019). “Geometric and level set tomography using ensemble Kalman inversion”. *Geophysical Journal International* vol. 220, pp. 967–980. DOI: 10.1093/gji/ggz472.
- Muir, J. B. et al. (2022). “Parsimonious Velocity Inversion Applied to the Los Angeles Basin, CA”. *Journal of Geophysical Research: Solid Earth* vol. 127, e2021JB023103. DOI: 10.1029/2021jb023103.
- Munson, T. et al. (2012). *TAO 2.0 User Manual*. Argonne National Laboratory. URL: <http://www.mcs.anl.gov/uploads/cels/papers/ANL-MCS-TM-322.pdf> (visited on 01/13/2024).
- Nečas, J. (2012). *Direct Methods in the Theory of Elliptic Equations*. Berlin: Springer. ISBN: 9783642104558.
- Nummelin, E. (1984). *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge: Cambridge University Press. DOI: 10.1017/CB09780511526237.
- Oleinik, O. A. (1963). “Discontinuous solutions of nonlinear differential equations”. *American Mathematical Society Translations* vol. 26, pp. 95–172. URL: <http://www.ams.org/books/trans2/026> (visited on 01/13/2024).
- (1992). *Mathematical Problems in Elasticity and Homogenization*. Amsterdam: North-Holland. ISBN: 0444884416.
- Ornstein, D. (1962). “A non-inequality for differential operators in the L1 norm”. *Archive for Rational Mechanics and Analysis* vol. 11, pp. 40–49. DOI: 10.1007/bf00253928.
- Osher, S. and Sethian, J. A. (1988). “Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations”. *Journal of Computational Physics* vol. 79, pp. 12–49. DOI: 10.1016/0021-9991(88)90002-2.
- Pinski, F. J. et al. (2015). “Algorithms for Kullback–Leibler Approximation of Probability Measures in Infinite Dimensions”. *SIAM Journal on Scientific Computing* vol. 37, A2733–A2757. DOI: 10.1137/14098171X.
- Ranalli, G. (1995). *Rheology of the Earth*. London: Chapman & Hall. ISBN: 0412546701.
- Räss, L. et al. (2017). “M2Di: Concise and efficient MATLAB 2-D Stokes solvers using the Finite Difference Method”. *Geochemistry, Geophysics, Geosystems* vol. 18, pp. 755–768. DOI: 10.1002/2016gc006727.
- Ratnaswamy, V., Stadler, G., and Gurnis, M. (2015). “Adjoint-based estimation of plate coupling in a non-linear mantle flow model: theory and examples”. *Geophysical Journal International* vol. 202, pp. 768–786. DOI: 10.1093/gji/ggv166.
- Reuber, G. S. et al. (2020). “Inferring rheology and geometry of subsurface structures by adjoint-based inversion of principal stress directions”. *Geophysical Journal International* vol. 223, pp. 851–861. DOI: 10.1093/gji/ggaa344.

- Reuber, G. S., Popov, A. A., and Kaus, B. J. (2017). “Deriving scaling laws in geodynamics using adjoint gradients”. *Tectonophysics* vol. 746, pp. 352–363. DOI: 10.1016/j.tecto.2017.07.017.
- Reuber, G. S. et al. (2018). “Unraveling the Physics of the Yellowstone Magmatic System Using Geodynamic Simulations”. *Frontiers in Earth Science* vol. 6. DOI: 10.3389/feart.2018.00117.
- Roberts, G. O. and Rosenthal, J. S. (1998). “Markov-chain Monte Carlo: some practical implications of theoretical results”. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique* vol. 26, pp. 5–20. DOI: 10.2307/3315667.
- Rudi, J. (2019). *Global convection in Earth’s mantle: Advanced numerical methods and extreme-scale simulations*. Dissertation. The University of Texas at Austin. DOI: 10.26153/TSW/1258.
- Rudi, J., Gurnis, M., and Stadler, G. (2022). “Simultaneous inference of plate boundary stresses and mantle rheology using adjoints: large-scale 2-D models”. *Geophysical Journal International* vol. 231, pp. 597–614. DOI: 10.1093/gji/ggac207.
- Rudi, J., Shih, Y.-h., and Stadler, G. (2020). “Advanced Newton Methods for Geodynamical Models of Stokes Flow With Viscoplastic Rheologies”. *Geochemistry, Geophysics, Geosystems* vol. 21, e2020GC009059. DOI: 10.1029/2020GC009059.
- Rudi, J., Stadler, G., and Ghattas, O. (2017). “Weighted BFBT Preconditioner for Stokes Flow Problems with Highly Heterogeneous Viscosity”. *SIAM Journal on Scientific Computing* vol. 39, S272–S297. DOI: 10.1137/16M108450X.
- Rudi, J. et al. (2015). “An Extreme-Scale Implicit Solver for Complex PDEs: Highly Heterogeneous Flow in Earth’s Mantle”. In: *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. SC ’15. Austin: Association for Computing Machinery. ISBN: 9781450337236. DOI: 10.1145/2807591.2807675.
- Santosa, F. (1996). “A level-set approach for inverse problems involving obstacles”. *ESAIM: Control, Optimisation and Calculus of Variations* vol. 1, pp. 17–33. DOI: 10.1051/cocv:1996101.
- Schubert, G., Turcotte, D. L., and Olson, P. (2001). *Mantle Convection in the Earth and Planets*. Cambridge: Cambridge University Press. ISBN: 9780511612879.
- Sohr, H. (2001). *The Navier-Stokes Equations: An Elementary Functional Analytic Approach*. BAT - Birkhäuser Advanced Texts. Basel: Birkhäuser Verlag AG. ISBN: 9783764365455.
- Spiegel, E. A. and Veronis, G. (1960). “On the Boussinesq Approximation for a Compressible Fluid.” *The Astrophysical Journal* vol. 131, pp. 442–447. DOI: 10.1086/146849.

Bibliography

- Spiegelman, M., May, D. A., and Wilson, C. R. (2016). “On the solvability of incompressible Stokes with viscoplastic rheologies in geodynamics”. *Geochemistry, Geophysics, Geosystems* vol. 17, pp. 2213–2238. DOI: 10.1002/2015gc006228.
- Stadler, G. et al. (2010). “The dynamics of plate tectonics and mantle flow: From local to global scales”. *Science* vol. 329, pp. 1033–1038. DOI: 10.1126/science.1191223.
- Stixrude, L. and Lithgow-Bertelloni, C. (2005). “Thermodynamics of mantle minerals — I. Physical properties”. *Geophysical Journal International* vol. 162, pp. 610–632. DOI: 10.1111/j.1365-246X.2005.02642.x.
- Stuart, A. M. (2010). “Inverse problems: A Bayesian perspective”. *Acta Numerica* vol. 19, pp. 451–559. DOI: 10.1017/S0962492910000061.
- Tackley, P. J. (2008). “Modelling compressible mantle convection with large viscosity contrasts in a three-dimensional spherical shell using the yin-yang grid”. *Physics of the Earth and Planetary Interiors* vol. 171. Recent Advances in Computational Geodynamics: Theory, Numerics and Applications, pp. 7–18. DOI: 10.1016/j.pepi.2008.08.005.
- Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia: Society for Industrial and Applied Mathematics. ISBN: 9780898715729.
- Temam, R. (1985). *Mathematical Problems in Plasticity*. Modern Applied Mathematics S. Paris: Gauthier-Villars. ISBN: 9780471918592.
- Tierney, L. (1994). “Markov Chains for Exploring Posterior Distributions”. *The Annals of Statistics* vol. 22, pp. 1701–1728. DOI: 10.1214/aos/1176325750.
- Tröltzsch, F. (2010). *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*. Vol. 112. Graduate Studies in Mathematics. Providence: American Mathematical Society. ISBN: 9781470476441.
- Tromp, J., Tape, C., and Liu, Q. (2005). “Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels”. *Geophysical Journal International* vol. 160, pp. 195–216. DOI: 10.1111/j.1365-246x.2004.02453.x.
- von Mises, R. (1913). “Mechanik der festen Körper im plastisch- deformablen Zustand”. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* vol. 1913, pp. 582–592. URL: <http://eudml.org/doc/58894> (visited on 01/13/2024).
- Werner, D. (2018). *Funktionalanalysis*. Heidelberg: Springer. ISBN: 9783662554074.
- Worthen, J. et al. (2014). “Towards adjoint-based inversion for rheological parameters in nonlinear viscous mantle flow”. *Physics of the Earth and Planetary Interiors* vol. 234, pp. 23–34. DOI: 10.1016/j.pepi.2014.06.006.

- Xue, M. and Allen, R. M. (2007). “The fate of the Juan de Fuca plate: Implications for a Yellowstone plume head”. *Earth and Planetary Science Letters* vol. 264, pp. 266–276. DOI: 10.1016/j.epsl.2007.09.047.
- Zeidler, E. (1990). *Nonlinear Functional Analysis and its Applications*. New York: Springer. ISBN: 9781461209812.
- Zhao, D., Hasegawa, A., and Horiuchi, S. (1992). “Tomographic imaging of P and S wave velocity structure beneath northeastern Japan”. *Journal of Geophysical Research: Solid Earth* vol. 97, pp. 19909–19928. DOI: 10.1029/92jb00603.
- Zhong, S., Gurnis, M., and Moresi, L. (1998). “Role of faults, nonlinear rheology, and viscosity structure in generating plates from instantaneous mantle flow models”. *Journal of Geophysical Research: Solid Earth* vol. 103, pp. 15255–15268. DOI: 10.1029/98jb00605.
- Zoback, M. (2010). *Reservoir Geomechanics*. Cambridge: Cambridge University Press. ISBN: 9780521146197.

Lebenslauf

removed from online version