Technische Universität München

Fakultät für Mathematik

Efficient Inversion Methods for Constrained Parameter Identification in Full-Waveform Seismic Tomography

Christian Thomas Böhm

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereichten Dissertation.

Vorsitzende:	Univ.-Prof. Dr. C. Lasser
Prüfer der Dissertation:	1. Univ.-Prof. Dr. M. Ulbrich
	2. Univ.-Prof. Dr. A. Schiela Universität Bayreuth
	3. Prof. Dr. G. Stadler New York University / USA (schriftliche Beurteilung)

Die Dissertation wurde am 15.10.2014 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 11.02.2015 angenommen.

To my parents and Katharina

Acknowledgements

Completing this thesis would not have been possible without the help and support of many people. First of all, I would like to thank Michael Ulbrich for giving me the opportunity to work on this interesting topic at the interface of applied mathematics and geophysics and for supervising my thesis. His ideas and remarks were extremely valuable to establish the results of this thesis.
Seismic tomography is a fascinating and also very challenging topic that requires people from different fields to work on. I highly enjoyed being part of the interdisciplinary MAC project B3 and would like to thank Michael Bader, Heiner Igel, Tobias Pfaffelmoser, Simon Stähler and Ran Zhang for the inspiring atmosphere we shared within our team.
I am very grateful for the financial support by the Munich Centre of Advanced Computing and the International Graduate School for Science and Engineering at the Technische Universität München. In particular, the funding enabled my research visit to the University of Texas at Austin in Winter 2011/12. Related to that I would like to express my deep gratitude to Omar Ghattas for hosting me at the Institute of Computational Engineering Sciences as well as for many interesting discussions and valuable insights into large-scale inverse problems. Moreover, I would like to thank Georg Stadler for making my time in Austin so enjoyable.
I would also like to thank all my current and former colleagues at M1, in particular, Andre, Boris, Dennis, Florian, Florian, Martin, Moritz, and Sebastian for the great time we have spent together.
Lastly, and most importantly, I thank my family for their continuous support throughout the years.

Abstract

This thesis is about efficient methods for constrained parameter identification in full-waveform seismic inversion. Seismic tomography is a technique to determine the material structure of the Earth's subsurface based on the observation of waves that are excited by earthquakes. This can be stated as a nonlinear PDE-constrained optimization problem governed by the elastic wave equation.
The main purpose of this thesis is to contribute to theoretical and practical aspects in full-waveform seismic tomography in the time domain. In particular, we study semismooth Newton-type methods that can handle additional constraints on the material parameters in a function space setting. To this end, results on the differentiability of the parameter-to-state operator are established and a semismooth Newton-PCG trust-region method that uses a Moreau-Yosida regularization is proposed. Furthermore, we utilize ideas from stochastic programming and employ randomized source sampling techniques with inexact Hessian approximations to efficiently gather information from a large number of seismic events. In addition, we develop strategies to enhance the practical performance of the proposed methods such as a multi-frequency inversion and regularization-by-discretization.
Numerical results are presented for inverse problems in geophysical exploration on reservoir-scale in both, solid and fluid domain. As part of this thesis, a matrix-free MPI-parallelized implementation that relies on the adjoint-based computation of the gradient and Hessian-vector products has been developed.

Zusammenfassung

Die vorliegende Dissertation beschäftigt sich mit seismischer Wellenform-Tomographie. Es handelt sich hierbei um ein Verfahren, das die Materialstruktur der Erde auf Grundlage von Messungen von Erdbeben erzeugter seismischer Wellen zu bestimmen versucht. Dies kann als nichtlineares Optimierungsproblem mit einer partiellen Differentialgleichung als Nebenbedingung formuliert werden. Hierbei ist das physikalische Modell durch die elastische Wellengleichung gegeben.
Ziel dieser Arbeit ist es, einen Beitrag in theoretischen und praktischen Aspekten zur seismischen Wellenform-Tomographie zu leisten. Wir untersuchen ein semiglattes Newton-artiges Verfahren im Funktionenraum, das zusätzliche Nebenbedingungen an die Materialparameter erlaubt. Dazu weisen wir die Differenzierbarkeit des Lösungsoperators nach und entwickeln ein semiglattes Newton-PCG Verfahren mit einer Trust-Region Globalisierung und Moreau-Yosida Regularisierung. Weiterhin nutzen wir Ideen aus dem Bereich der stochastischen Optimierung und verwenden randomisierte Sampling-Techniken in Kombination mit einer inexakten Approximation des Hesse-Operators, um effizient die Daten von einer Vielzahl seismischer Quellen zu nutzen. Darüberhinaus entwickeln wir Strategien, um die Effizienz der vorgestellten Verfahren in der Praxis zu verbessern. Dies umfasst unter anderem eine Multi-Frequenz Inversion und Regularisierung durch Diskretisierung.
Numerische Beispiele für inverse Probleme im Bereich der geophysikalischen Exploration innerhalb eines Reservoirs mit festem und fluidem Medium werden präsentiert. Dazu wurde im Rahmen der Dissertation eine Matrix-freie Methode implementiert, welche die adjungiertenbasierte Berechnung des Gradienten und von Hesse-Vektor-Produkten erlaubt und Parallelisierung mit MPI unterstützt.

Contents

Chapter 1

Introduction

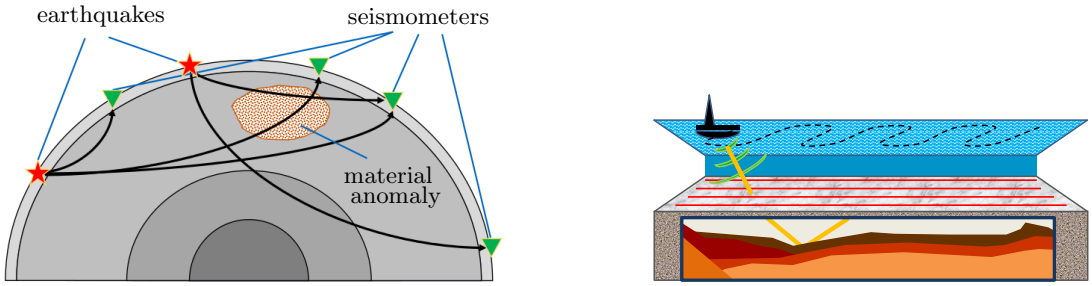
No human being has ever seen the Earth's interior. Nevertheless, nowadays we have a quite strong belief in the tectonic structure, the division into core, mantle and crust and the location of huge reservoirs of natural resources. So where does our knowledge of the subsurface come from? Direct ways of measurement are very limited as the world's deepest borehole Sakhalin-I in the Okhotsk Sea with about 12.7 km depth [50] reaches less than 0.2% of the distance to the center of the core. Hence, indirect methods have to be used in order to gain insights of the Earth's subsurface and seismic tomography is probably the most important one of them. Here, the basic principle is similar to magnetic resonance tomography in medical imaging where measurements of electromagnetic waves are used to reconstruct the structure of tissues and bones inside the human body. Seismic tomography means to infer the material structure of the Earth's interior based on the observation of waves that spread through the subsurface. When an earthquake occurs, seismic waves are emitted at its hypocenter and can be recorded in form of seismograms at locations far away from the source. The received signal carries information about the Earth's interior structure as the velocity of the travelling waves depends on the material and reflections occur at the transition of different layers of rock. For a general overview on seismic tomography we refer to [101, 108, 130].

Detailed knowledge of the structure of the Earth is required to understand the dynamics of the planet and to explain its history and evolution. There are many fields that benefit from an accurate image of the Earth's interior, to which new insights from seismic tomography can contribute. For instance, it can be used to explain geodynamic processes that happen in the Earth's mantle and core. On a regional scale, areas of potential geologic hazards like volcanoes or landslides might be identified. Furthermore, tomography problems of this kind are solved to support the search for natural resources, especially for oil and gas. Moreover, a detailed knowledge of the ocean bottom and shallow subsurface can help to improve reliable Tsunami warning systems and to monitor adherence to the Comprehensive Nuclear-Test-Ban Treaty[1]. Figure 1.1 sketches two examples of use, namely global seismic tomography and marine geophysical exploration.

Efficient inversion methods that are capable of dealing with a huge amount of observed data and a large number of computationally expensive numerical simulations are required to promote future progress in seismic tomography. This work intends to contribute to this long-term endeavor in theoretical and practical aspects.

[1] `http://www.ctbto.org/the-treaty/`

One of the main challenges in seismic tomography is the uneven distribution of the locations of seismic sources (earthquakes) and observatories (receivers) on the globe. On the one hand, earthquakes usually occur only in a few regions of increased seismic activity and in comparatively shallow depth. On the other hand, measurements of the wavefield are almost exclusively available at the Earth's surface. Furthermore, inaccessible regions like oceans make it hard or even impossible to observe data with a dense array of receivers around the globe.



(a) (b)

Figure 1.1: Two examples of seismic tomography problems on different scales. (a) Global tomography: Earthquakes with a high magnitude can be recorded by seismometers thousands of kilometers away from the hypocenter. This data is used to reveal deep structures of the Earth's interior. (b) Marine geophysical exploration: A research vessel equipped with an airgun emits pressure waves into the ocean. A dense array of geophones (red lines) records the response of reflected and refracted waves at the seafloor. The purpose is to identify deposits of natural resources.

In an abstract setting, seismic tomography can be formulated as the equation

$$F(m) = d, \tag{1.1}$$

where d denotes the measurements (e.g., seismograms), the variable m parameterizes the material and F is the physical model that maps the material to observable data, see also Figure 1.2. As a simple physical model, we can, for instance, approximate seismic waves as ray paths between the locations of sources and receivers. This approach is called *traveltime tomography*, since the observable quantity is the arrival time of waves propagating between both points and the physical model is given by the integral of the material velocity along the ray path. Here, pioneering work goes back to the 70s of the last century [3, 20]. More sophisticated models describe the propagation of waves by a partial differential equation (PDE) which can either be the Helmholtz equation in the frequency domain or variants of the acoustic or the elastic wave equation in the time domain. This is called *full-waveform seismic tomography* since the whole seismogram is used instead of only a few arrival times of particular wavefronts. There is also an intermediate approach using the eikonal equation [89]. To the best of our knowledge, full-waveform inversion emerged from the early work of Tarantola [119] and was further pursued in [105]. All physical models have in common that the resulting tomography problem is a so-called *inverse problem*, since the inverse operator F^{-1} is not directly available or does not even exist. Inverse problems have the characteristic property of being "ill-posed" in contrast to Hadamard's classification of well-posedness, cf. [47], which

essentially demands that the problem (1.1) has a unique solution for any "reasonable" data d and that this solution is stable with respect to small perturbations of d. Due to the ill-posedness, inverse problems are challenging to solve and require special treatment in form of regularization. For a comprehensive introduction to inverse problems we refer to [47, 120, 131].

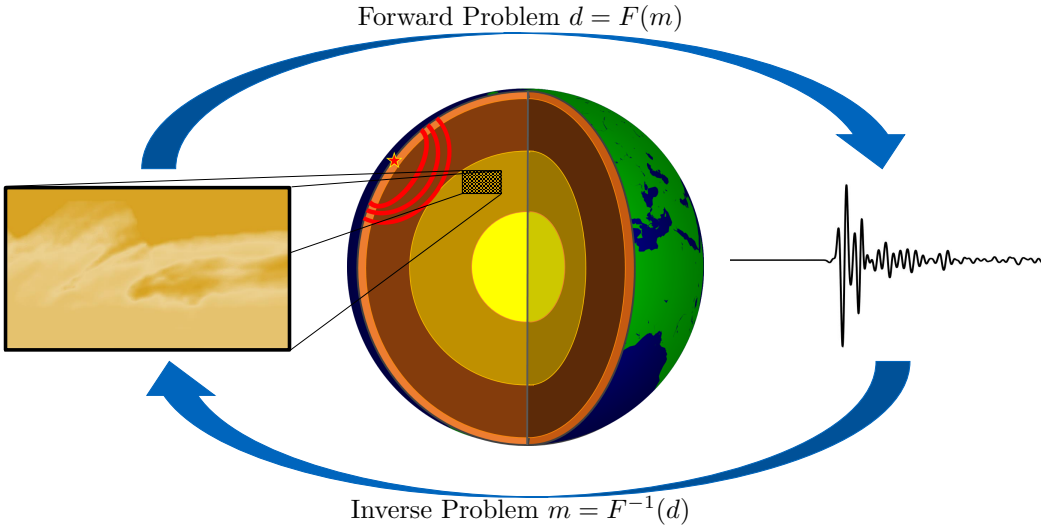


Figure 1.2: The forward problem maps parameters of the material structure to seismograms. The inverse problem seeks to identify the material properties of the Earth based on the observed seismograms and is considered to be ill-posed.

Since this thesis focuses on full-waveform tomography in the time domain, we would like to comment on the problem setting and relevant literature in this area in more detail. We consider seismic tomography as an optimization problem with PDE constraints, where the unknown material parameters are assumed to be spatially heterogeneous and constant in time. Depending on whether the medium is solid or fluid, the governing equation for the propagation of seismic waves is either given by the elastic or the acoustic wave equation, respectively. From the early beginnings in the 80s of the last century, full-waveform inversion has only become computationally feasible with the availability of high-performance computing clusters in recent years. Iterative inversion methods based on first-order information have been applied to 2d and 3d datasets on both, regional and continental scale [52, 53, 76, 118, 124, 135]. A Newton-CG method for the unconstrained parameter identification problem was presented in [48], see also [26]. Alternative approaches work in the frequency domain and involve the Helmholtz equation [22, 61, 95, 104]. All of this work has in common that the focus is on the discretized problem. To the best of our knowledge, little attention has been paid to the analysis of the infinite-dimensional problem. In [88] some results on the differentiability of the wavefield with respect to the wave velocity are given for the acoustic wave equation in 1d. One of the main contributions of this thesis is the rigorous treatment of the problem in a function space setting. We establish results on the existence and uniqueness of solutions to

the governing equations as well as the differentiability of the material-to-solution operator in the given setup. Recently and without the author's knowledge, similar results for the forward problem have been established independently in [16] for hyperbolic evolution equations and in [78] for the acoustic wave equation. We note, however, that the results of this thesis require weaker assumptions on the regularity of the seismic sources. We will comment on that in more detail in Chapter 2.

Most approaches in full waveform tomography do not incorporate additional constraints on the material parameters in the formulation of the inverse problem. However, there naturally exist physical bounds, for instance, the non-negativity of wave velocities or the coercivity of the elliptic operator. Moreover, the problem is ill-posed and also non-convex with several local minimizers likely to exist. Therefore, from a modeling perspective, we would like to include as much prior knowledge as possible into the problem formulation. One way of doing this is to impose additional constraints on the material parameters based on a priori available information. This could comprise, for instance, lower and upper bounds on the velocities of compressional and shear waves. It is also particularly interesting and useful for the joint inversion for both Lamé coefficients λ and μ which is known to be a challenging problem [106]. Here, our approach allows to control the deviation of both parameter fields by imposing bounds on the Poisson's ratio. Moreover, this methodology can be extended to impose constraints on other physical quantities, for instance, the total mass for problems in global seismology.

We apply the Moreau-Yosida regularization to handle the constraints on the material parameters. This penalty method leads to an optimality system involving a semismooth operator equation. For its solution, we propose a semismooth Newton-CG method and a trust-region globalization strategy. Semismooth Newton-type methods for optimization problems in function spaces have been studied extensively in [67, 125, 126] and have been applied to various types of applications, see, for instance, [86] for an optimal control problem governed by the wave equation and [69, 70, 71, 72] for problems involving the Moreau-Yosida regularization. Estimates on the constraint violation have been established in [68, 126].

One may argue, however, that the additional constraints are usually not expected to be active at the global minimizer. In this case, the Moreau-Yosida regularization comes in handy, as the penalized problem coincides with the original problem within the feasible region. Furthermore, in case of soft constraints, the penalty term might actually be interpreted as a different form of regularization which does not necessitate driving the penalty parameter to infinity.

Another challenge that arises particularly in geophysical exploration is the huge amount of data, which can consist of tens of thousands of sources [107], such that the observed seismic traces pile up to several terabytes of data. Hence, it is computationally intractable to incorporate all seismic sources individually and different strategies have to be developed. We build upon ideas from stochastic programming [113], machine learning [28] and related approaches that have been applied successfully in the frequency domain [15, 61, 97]. Preliminary work also exists for the acoustic wave equation [112]. The key idea is to trigger several seismic events simultaneously and to use random weights to accumulate data. We present different algorithms to tackle this problem, which are based on a sample average approximation. One of the contributions of this thesis is to accelerate the solution process by using inexact Hessian information based on mini-batches of the samples.

Efficient inversion methods also rely on a scalable code for the simulation of the wave equation. During the course of this thesis, a simulation code for variants of the wave equation in two and three dimensions has been developed. Additionally, we consider wave propagation in a

coupled system of solid and fluid media. The C++ implementation is parallelized, utilizes MPI-communication and works matrix-free. Moreover, it is not required to solve a linear system during the simulation, since an explicit Newmark time-stepping scheme [74] and a diagonal mass matrix are used. Forward and adjoint simulations are carried out to efficiently compute the reduced gradient and reduced Hessian-vector products that are required by the Newton-CG method. The code has been tested on up to 4096 compute cores on a Linux cluster.

After giving a short summary of each of the following chapters, we will continue this introduction with a brief derivation of the governing equations and summarize important notations.

Chapter 2
In this chapter, we analyze the governing equations. In order to allow for a unified treatment of both, the elastic and the acoustic wave equation, we start by considering general linear hyperbolic equations. After recalling some well-known results on the existence, uniqueness and regularity of solutions, we consider a coefficient-dependent differential operator and establish continuity and differentiability of the solution to the hyperbolic equations with respect to these coefficients.

Chapter 3
We continue with the analysis of the parameter identification problem for seismic tomography and prove the existence of a solution to the regularized inverse problem based on the results from Chapter 2. In the absence of further restrictions, the governing equation might not be well-defined on the whole space of material parameters. Therefore, we present a cutoff strategy involving a superposition operator to circumvent this problem which leads to an unconstrained variant of the inverse problem. In order to explicitly include additional constraints on the material, on the other hand, we consider the Moreau-Yosida regularization. We propose a semismooth Newton method with a trust-region globalization for the regularized problem and, furthermore, we establish estimates for the constraint violation of the Moreau-Yosida regularized solution.

Chapter 4
Next, we describe the discretization of the problem and the implementation of a parallelized wave propagation code. Parallel scaling statistics show the applicability of the implementation to tackle large-scale inverse problems. Furthermore, numerical examples on reservoir scale problems in 2d and 3d are presented.

Chapter 5
In this chapter, we discuss strategies to substantially reduce the computational effort by considering simultaneous seismic sources instead of individual events. In particular, we discuss a sample average approximation model that is accelerated by using inexact Hessian information based on mini-batches of the samples.

Chapter 6
We conclude the thesis with a case study on wave propagation at a solid-fluid interface. Problems of this kind arise, for instance, in marine geophysical exploration of undersea oilfields. The governing equations are given by a coupled system of the elastic and the acoustic wave equation with interface conditions between both media. Numerical results are presented for an elastic inversion of a 3d dataset with a thin layer of water on the top of the domain. In addition, we discuss a multi-frequency inversion and a regularization-by-discretization approach.

Parts of this thesis have been published or submitted to peer-reviewed journals. The relevant articles are:

C. Boehm and M. Ulbrich. "Newton-CG Method for Full-Waveform Inversion in a Coupled Solid-Fluid System." In: *Advanced Computing*. Ed. by M. Bader, H.-J. Bungartz, and T. Weinzierl. Vol. 93. Lecture Notes in Computational Science and Engineering. Springer Berlin Heidelberg, 2013, pp. 99–117.

C. Boehm and M. Ulbrich. "A Semismooth Newton-CG Method for Constrained Parameter Identification in Seismic Tomography.", in revision, 2014.

1.1 Seismic Wave Propagation

In this section we briefly summarize the geophysical background of this work and derive the governing equations for wave propagation in solid and fluid media. Here, we follow [2, 39, 101] and refer to these references for detailed explanations and a thorough analysis.
In what follows, we consider a d-dimensional domain with $d = 2$ or $d = 3$. Let P be a particle that is initially located at $x \in \mathbb{R}^d$ at some reference time t_0 and then moved to P'. We introduce a time-dependent vector field $u = u(x, t)$ which describes the shift in position at time t compared to the original location at time t_0. u is called the *displacement field*. For the moment, we ignore the time parameter and look only at a fixed time t. Now, consider a second point Q, initially located at $x + \delta x$ in the vicinity of P.

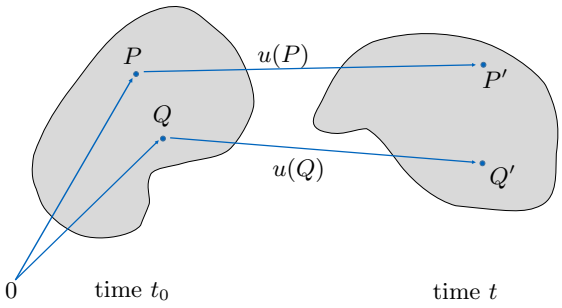


Figure 1.3: Deformation of a body described by the displacement field u.

Its new position Q' at time t is then given by $x+\delta x+u(x+\delta x)$. Assuming that the displacement is small we could use a first order expansion and obtain

$$u(x + \delta x) = u(x) + \nabla u(x)^T \delta x + O(\|\delta x\|^2),$$

where $(\nabla u(x))_{ij} = \partial u_j / \partial x_i$ is the transposed Jacobian of the displacement field. We note that a constant u would describe a simple translation of the body. Since we are rather interested in its deformation, we take a closer look at the derivatives $\partial u_j / \partial x_i$. It is important to note, however, that nonzero derivatives are not necessarily caused by a deformation, but can also come from a rotation of the body. Hence, we have to split up ∇u in order to separate rotational components from deformation. To this end, we define the *strain tensor* ε and the *rotation tensor* Θ by:

$$\varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j}\right), \quad \text{i.e.,} \quad \varepsilon = \frac{1}{2}(\nabla u + \nabla u^T),$$
$$\Theta_{ij} = \frac{1}{2}\left(\frac{\partial u_j}{\partial x_i} - \frac{\partial u_i}{\partial x_j}\right).$$

Hence, we obtain $\nabla u = \varepsilon + \Theta$, where the symmetric strain tensor ε describes the change of relative position of two points and the rotation tensor Θ corresponds to rigid-body rotations, cf. section 2.1 in [2].
In the next step, we are interested in a description of forces acting mutually between particles. Here, *traction* is a vector defined as the force per unit area that acts on the surface of a volume

element. Traction is oriented with the unit normal pointing outwards of the surface and is by definition positive for extensional forces and negative for compressional forces. Since the force acting on an arbitrary internal surface is a linear function of the orientation of that surface, we may relate the traction $\tau(\vec{n})$ to the *stress tensor* $\bar{\sigma}$ by

$$\tau(\vec{n})_i = \sum_{j=1}^{d} \bar{\sigma}_{ij}\vec{n}_j, \quad i = 1, \dots, d. \tag{1.2}$$

Here, $\vec{n}$ is the unit normal to the surface. In Cartesian coordinates with unit vectors denoted by e_i, $i = 1, \dots, d$, $\bar{\sigma}_{ij}$ is the force acting in direction e_i on the surface with normal vector e_j. The diagonal entries of $\bar{\sigma}$ are compressional forces and the off-diagonal elements describe shear forces. Hence, if the stress tensor $\bar{\sigma}(x)$ at a point x is known, the traction on an arbitrary surface through x can be computed by (1.2). Stress is defined in force per unit area, i.e., in N/m^2 or Pa. Furthermore, the stress tensor is symmetric, for a derivation we refer again to [2], section 2.1.

In addition to the contact forces at which particles on both sides of the surface act on each other, we introduce an external body force f that will describe the seismic source. Furthermore, we assume a heterogeneous density ρ. Using Newton's second law, which states that the force is equal to mass times acceleration, we can now state the first version of the elastodynamic equations for a compact volume element $\mathcal{V}$ with smooth surface $\mathcal{S}$ and any time $t > t_0$. Note that x does not depend on t, i.e., the particle velocity is given by u_t and its acceleration by u_{tt}. We obtain:

$$\int_{\mathcal{V}} \rho u_{tt}\, d\mathcal{V} = \int_{\mathcal{V}} f\, d\mathcal{V} + \int_{\mathcal{S}} \tau(\vec{n})\, d\mathcal{S}. \tag{1.3}$$

By Gauss's divergence theorem, we deduce that (1.3) can be reformulated to

$$\int_{\mathcal{V}} \rho u_{tt}\, d\mathcal{V} = \int_{\mathcal{V}} f\, d\mathcal{V} + \int_{\mathcal{V}} \nabla \cdot \bar{\sigma}\, d\mathcal{V}. \tag{1.4}$$

Finally, we have to relate stresses to the displacements. This can be done by applying the generalized Hooke's law which states a linear dependence of the stress tensor on the components of the strain tensor, see, for instance [2], section 2.2. With the help of the fourth-order elastic tensor Ψ, this can be written as

$$\bar{\sigma}_{ij} = \sum_{k,l=1}^{d} \Psi_{ijkl}\varepsilon(u)_{kl}. \tag{1.5}$$

For a compact notation, we write $\bar{\sigma} = \Psi : \varepsilon(u)$ and refer to a body that obeys (1.5) as *linearly elastic*. Although we omitted to state the dependence on location x and time t explicitly, it should be clear from the derivation above that quantities like displacement, strain, stress, traction and external forces are functions of both, space and time. The elastic tensor Ψ and the density ρ only depend on the spatial location.

Now, we have gathered all ingredients to state the elastic wave equation. Here, we consider a bounded domain $\Omega \subset \mathbb{R}^d$ (again $d = 2, 3$) with a smooth boundary. The time interval is denoted by $I := (0, T)$ with $T > 0$. For simplicity and without loss of generality, we always

assume initial time $t_0 = 0$. The elastic wave equation is then given by

$$\begin{cases} \rho u_{tt} - \nabla \cdot (\Psi : \varepsilon(u)) = f & \text{on } \Omega \times I, \\ u(0) = 0 & \text{on } \Omega, \\ u_t(0) = 0 & \text{on } \Omega, \\ (\Psi : \varepsilon(u)) \cdot \vec{n} = 0 & \text{on } \partial\Omega \times I. \end{cases} \tag{1.6}$$

We recall that the external force induced by the seismic source is denoted by f. Furthermore, we assume that the system is at rest at initial time and impose homogeneous initial conditions on u and u_t. At the free surface, the traction is zero which is stated in the fourth line of (1.6). Note that in global seismology we can consider Ω to be the whole globe, i.e., $\partial\Omega$ is the surface of the Earth and the boundary conditions are valid. However, for problems where the domain Ω is truncated to a subset of the Earth, artificial boundaries are introduced and appropriate absorbing boundary conditions have to be imposed. This discussion, however, is postponed to section 4.3.

Parameterization of the Material

The material structure that we seek to determine by solving the tomography problem is characterized by the elastic tensor Ψ and the density ρ. However, due to the interdependencies of ρ and Ψ, we will keep the density fixed and invert for Ψ only. Therefore, we take a closer look at its properties. First of all, we can exploit the symmetry properties $\Psi_{ijkl} = \Psi_{jikl} = \Psi_{klij}$, which yields at most 21 independent components for $d = 3$ (instead of 81). Due to these symmetries, we directly obtain:

$$\bar{\sigma}_{ij} = \sum_{k,l=1}^{d} \Psi_{ijkl} \varepsilon(u)_{kl} = \sum_{k,l=1}^{d} \frac{1}{2} \Psi_{ijkl} \left((\nabla u)_{kl} + (\nabla u)_{lk} \right) = \sum_{k,l=1}^{d} \Psi_{ijkl} \frac{\partial u_l}{\partial x_k}. \tag{1.7}$$

A compact representation of the coefficients Ψ_{ijkl} is often given by the right upper triangular part of the so-called Voigt matrix, cf. [101], p. 291, that groups the components in the following way:

$$\Psi^V = \begin{pmatrix} \Psi_{1111} & \Psi_{1122} & \Psi_{1133} & \Psi_{1123} & \Psi_{1113} & \Psi_{1112} \\ \Psi_{2211} & \Psi_{2222} & \Psi_{2233} & \Psi_{2223} & \Psi_{2213} & \Psi_{2212} \\ \Psi_{3311} & \Psi_{3322} & \Psi_{3333} & \Psi_{3323} & \Psi_{3313} & \Psi_{3312} \\ \Psi_{2311} & \Psi_{2322} & \Psi_{2333} & \Psi_{2323} & \Psi_{2313} & \Psi_{2312} \\ \Psi_{1311} & \Psi_{1322} & \Psi_{1333} & \Psi_{1323} & \Psi_{1313} & \Psi_{1312} \\ \Psi_{1211} & \Psi_{1222} & \Psi_{1233} & \Psi_{1223} & \Psi_{1213} & \Psi_{1212} \end{pmatrix}. \tag{1.8}$$

Examples for Ψ^V will be given in section 2.4. The general form of Ψ allows the treatment of anisotropic material. An important special case, however, is a perfectly elastic, isotropic medium. Here, the tensor simplifies to

$$\Psi_{ijkl} = \lambda \delta_{ij} \delta_{kl} + \mu(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}), \tag{1.9}$$

with the Lamé parameters λ and μ. Alternatively, it can be parameterized in terms of the bulk modulus $\kappa = \lambda + \frac{2}{3}\mu$ which yields

$$\Psi_{ijkl} = (\kappa - \frac{2}{3}\mu)\delta_{ij}\delta_{kl} + \mu(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}). \tag{1.10}$$

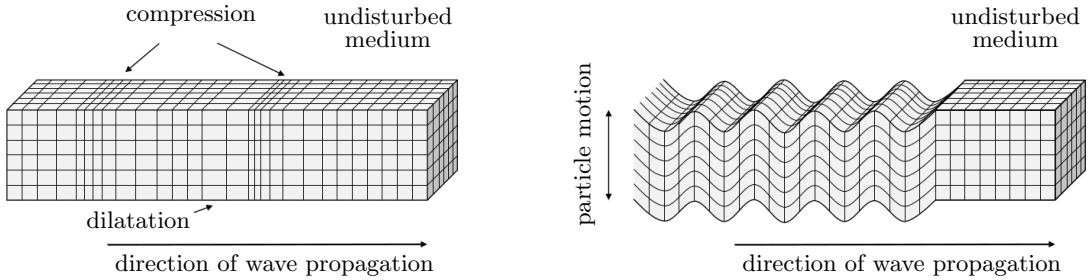
Using (1.9) in (1.7) we obtain

$$\bar{\sigma}_{ij} = \lambda \delta_{ij} (\nabla \cdot u) + 2\mu \varepsilon(u)_{ij}.$$

Thus, the isotropic elastic wave equation is given by:

$$\begin{cases} \rho u_{tt} - \nabla \cdot (2\mu\varepsilon(u) + \lambda(\nabla \cdot u)I) = f & \text{on } \Omega \times I, \\ u(0) = 0 & \text{on } \Omega, \\ u_t(0) = 0 & \text{on } \Omega, \\ (2\mu\varepsilon(u) + \lambda(\nabla \cdot u)I) \cdot \vec{n} = 0 & \text{on } \partial\Omega \times I. \end{cases} \tag{1.11}$$

In computational seismology, one is often rather interested in the velocities of certain types of body waves instead of the Lamé coefficients. Here, we distinguish between compressional and shear waves (see Figure 1.4) and give a short motivation.



(a) (b)

Figure 1.4: (a) Compressional waves (P-waves): Particles move parallel to the direction of the wave. (b) Shear waves (S-waves): Particles move orthogonal to the direction of the wave.

For simplicity, consider a homogeneous isotropic material, i.e., constant ρ, λ and μ and the absence of any external force f. Now we introduce $y = \nabla \cdot u$ and by taking the divergence in the first line of (1.11), we obtain after a short calculation

$$\frac{\rho}{\lambda + 2\mu} y_{tt} - \Delta y = 0, \tag{1.12}$$

which is a scalar wave equation with wave speed $c = (\rho/(\lambda + 2\mu))^{-1/2}$. Since the components of y correspond to the diagonal entries of the strain tensor $\bar{\sigma}$, this quantity describes compressional deformations. Consequently, solutions to (1.12) are called *compressional waves* or P-waves and the velocity is denoted by

$$v_p = \sqrt{\frac{\lambda + 2\mu}{\rho}}. \tag{1.13}$$

Similarly, by denoting $y = \nabla \times u$ and by taking the curl in the first line of (1.11), we obtain

$$\frac{\rho}{\mu} y_{tt} - \Delta y = 0. \tag{1.14}$$

Solutions to (1.14) are called *shear waves* or S-waves and the velocity is denoted by

$$v_s = \sqrt{\frac{\mu}{\rho}}. \tag{1.15}$$

In addition to body waves, which travel through the Earth's interior and can be characterized by the P- and S-wave velocity, there are surface waves that propagate along the Earth's surface. Important types are, for instance, Love waves and Rayleigh waves. The interested reader is referred to [2], section 7, or [101], section 10, for a rigorous analysis. In heterogeneous media, reflections occur at material interfaces which will convert compressional waves into shear waves and vice versa. Distinguishing the different types of waves in a seismogram is a lot more complicated in this case.

Wave Propagation in Fluid Media

In an inviscid fluid medium, the governing equation reduces to an acoustic wave equation since shear stresses are zero and the remaining stress is isotropic. Modeling a fluid medium is important for several reasons. On the one hand, the outer core of the Earth is liquid. On the other hand, for problems in marine geophysical exploration it is necessary to incorporate the ocean layer into the model. Moreover, since the simulation of the acoustic wave equation is much cheaper than the elastic counterpart, it is often used as the governing equation even for tomography problems in solid media [12, 57, 112]. Despite some limitations [99], it can at least be used prior to an elastic inversion to improve the starting model. We assume a constant density $\rho_F > 0$ and denote the displacement potential by u and a pressure source by f. Note that although it is uncommon to denote the displacement potential by u, we want to discuss the acoustic and the elastic wave equation in a unified way later on and, therefore, seek for a general "state variable" u. Furthermore, we have a scalar parameter field m that denotes the squared velocity of compressional waves in this case. The acoustic wave equation is now given by

$$\begin{cases} \rho_F u_{tt} - \rho_F \nabla \cdot (m \nabla u) = f & \text{on } \Omega \times I, \\ u(0) = 0 & \text{on } \Omega, \\ u_t(0) = 0 & \text{on } \Omega, \\ u = 0 & \text{on } \partial\Omega \times I. \end{cases} \tag{1.16}$$

Again, we assume that the system is at rest at initial time and impose homogeneous initial conditions. At the free surface, we work with homogeneous Dirichlet boundary conditions which give the fourth line in (1.16).
We conclude this section with two remarks.

Remark 1.1.1.
For certain applications, e.g., in global seismic tomography, it can be advantageous to consider the elastic wave equation in spherical coordinates. This is not the focus of this thesis but we refer the interested reader to [55] for its derivation.

Remark 1.1.2.
Other approaches work in the frequency domain, see for instance [130] and the references therein. In this case, the physical model is given by the Helmholtz equation. A major difference in tackling tomography problems in the frequency domain is that the PDE operator can often be factorized by a direct solver. As a consequence, multiple seismic events can be incorporated at much lower costs than in the time domain.

1.2 Notation

Here, we provide a short list of notations that are frequently used throughout this thesis. Note, in particular, that we employ the standard notation of computational seismology which results in a state variable u and a parameter variable m which is different from common notations in the context of optimal control or PDE-constrained optimization.

General Notation

$\|.\|_p$	p-norm in $\mathbb{R}^n$ with $p \in [1, \infty]$
$\|.\|_X$	norm of a Banach space X
X^*	dual space of a Banach space X
$(.,.)_X$	inner product of a Hilbert space X
$\langle .,. \rangle_{X^*,X}$	dual pairing of a Banach space X and its dual X^*
$X \hookrightarrow Y$	Banach space X is continuously embedded in Banach space Y
$X \hookrightarrow\hookrightarrow Y$	Banach space X is compactly embedded in Banach space Y
$B_\varepsilon(x)$	closed ball with radius ε centered at $x \in X$
$\mathcal{L}(X, Y)$	space of bounded linear operators from the Banach space X to the Banach space Y
$\mathcal{L}^i(X; Y)$	space of bounded multilinear mappings $X \times X \times \ldots \times X \to Y$ for Banach spaces X and Y
$a \otimes b$	(outer) product of second order tensors a and b, $(a \otimes b)_{ijkl} = a_{ij} b_{kl}$
$A :: B$	(inner) product of fourth order tensors A and B, $A :: B = \sum_{ijkl} A_{ijkl} B_{ijkl}$

Inverse Problem

d	spatial dimension of the state equation, usually $d = 2, 3$
u	state variable (e.g., displacements in the case of the elastic wave equation)
$m \in M$	variable that parameterizes the unknown material
M_{ad}	admissible set of materials
n	dimension of the parameter field
I	time interval (if not explicitly mentioned otherwise, we consider $I = (0, T)$)
$\Omega \subset \mathbb{R}^d$	bounded domain, usually a nonempty, bounded and open set with smooth boundary
$\partial\Omega$	boundary of domain Ω

Physical Quantities

$\varepsilon(u)$	stress tensor defined by $\varepsilon(u) = \frac{1}{2}(\nabla u + \nabla u^T)$
λ, μ	Lamé coefficients in isotropic elastic material
Ψ	fourth-order tensor of elastic moduli
ρ	density
$\bar{\sigma}$	stress tensor
v_p, v_s	velocities of compressional and, resp., shear waves

Function Spaces

$L^p(\Omega)$	Lebesgue space with functions $v : \Omega \to \mathbb{R}$, $p \in [1, \infty]$
$W^{k,p}(\Omega)$	Sobolev space of functions $v \in L^p(\Omega)$ with weak derivatives in $L^p(\Omega)$ up to order k, $k \geq 0$, $p \in [1, \infty]$
$H^k(\Omega)$	short form of the Hilbert space $W^{k,2}(\Omega)$, $k \geq 0$
$C(\bar{\Omega})$	space of continuous functions $v : \bar{\Omega} \to \mathbb{R}$
$C_c^\infty(\Omega)$	space of infinitely differentiable functions $v : \Omega \to \mathbb{R}$ with compact support in Ω
$C^{k,\beta}(\bar{\Omega})$	space of k-times differentiable functions $u : \bar{\Omega} \to \mathbb{R}$ with β-Hölder continuous derivatives, $k \geq 0$, $\beta \in (0, 1]$
$W_0^{k,p}(\Omega)$	closure of $C_c^\infty(\Omega)$ in $W^{k,p}(\Omega)$
$L^p(I; X)$	Bochner space of functions on interval I with values in Banach space X, $p \in [1, \infty]$, analogously: $C^k(\bar{I}; X)$, $H^k(I; X)$
F_0^k	subset of $H^k(I; V^*)$ with $f = 0$ near $t = 0$, V as defined in Chapter 2

Furthermore, $C > 0$ denotes a generic constant.

Chapter 2

Existence and Regularity of Solutions to the State Equation

In this chapter, results on the existence and regularity of solutions to the governing equations are discussed. We start by recalling some general results for linear hyperbolic equations that can be found, for instance, in [49, 87, 91]. The reader is also referred to [90] for optimal control problems governed by hyperbolic equations. However, since the material parameters enter the state equation as coefficients of an elliptic operator, the forthcoming analysis is more involved and we cannot directly refer to standard results from the literature cited above. To this end, we consider linear hyperbolic equations with a parameterized elliptic operator and establish the continuity and Fréchet differentiability of the solution operator that maps these parameters to the unique solution of the corresponding PDE. Parts of this section have been published in [19] for the elastic wave equation. Afterwards, we extend the results to a more general setting that involves a superposition operator. This enables us, on the one hand, to treat various different parameterizations of the material and, on the other hand, to ensure that all values of the coefficients that enter the state equation are physically reasonable. In the last part of this chapter, we apply the findings to the elastic and the acoustic wave equation.

2.1 Linear Hyperbolic Equations of Second Order

We start by briefly recalling some general results for the existence and the regularity of solutions to linear second-order hyperbolic PDEs. We consider a bounded domain $\Omega \subset \mathbb{R}^d$ $(d = 2, 3)$ with a smooth boundary and denote the time interval by $I := (0, T)$, $T > 0$. Let V and H be real Hilbert spaces, V dense in H, that form a Gelfand triple

$$V \hookrightarrow H = H^* \hookrightarrow V.$$

Later on, we will choose

$$V = \{v \in H^1(\Omega)^k : v|_{\Gamma_D} = 0\} \quad \text{and} \quad H = L^2(\Omega)^k \quad \text{with} \quad k \in \{1, d\}.$$

Here, $\Gamma_D \subset \partial\Omega$ denotes the part of the boundary with Dirichlet-type conditions and $\Gamma_D = \emptyset$ corresponds to Neumann-type conditions on every part of the boundary. Let a denote a continuous time-dependent bilinear form

$$a : I \times V \times V \to \mathbb{R},$$

that has the following properties

$$\forall\, v, w \in V \quad t \mapsto a(t; v, w) \text{ is continuously differentiable on } \bar{I}, \tag{2.1}$$

$$\exists\, \beta_1, \beta_2 > 0 \text{ such that } a(t; v, v) + \beta_1 \|v\|_H^2 \geq \beta_2 \|v\|_V^2 \quad \forall\, v \in V,\ \forall\, t \in \bar{I}, \tag{2.2}$$

$$a(t; v, w) = a(t; w, v) \quad \forall\, v, w \in V, \forall\, t \in I. \tag{2.3}$$

For fixed $t \in I$, we identify a with the operator $A(t) \in \mathcal{L}(V, V^*)$ by

$$a(t; v, w) = \langle A(t)v, w\rangle_{V^*,V} \quad \forall\, v, w \in V$$

and consider equations

$$u_{tt}(t) + A(t)u(t) = f(t), \tag{2.4}$$

with initial data

$$u(0) = u_0, \quad u_t(0) = u_1. \tag{2.5}$$

Let (2.1) - (2.3) hold. Now, we consider the variational form of (2.4), (2.5) and seek a solution u satisfying

$$\langle u_{tt}(t), v\rangle_{V^*,V} + a(t; u, v) = \langle f(t), v\rangle_{V^*,V} \quad \forall\, v \in V$$

a.e. in $(0, T)$ and $u(0) = u_0$, $u_t(0) = u_1$.
Note, if $f \in L^2(I; H)$, we continue to write $\langle f(t), v\rangle_{V^*,V}$ which is permissible by the embedding $H \hookrightarrow V^*$.

Theorem 2.1.1.
Let (2.1) - (2.3) *hold. Furthermore, let* $u_0 \in V$, $u_1 \in H$ *and* $f \in L^2(I; H)$. *Then there exists a unique solution* $(u, u_t) \in L^2(I; V) \times L^2(I; H)$ *that satisfies* (2.4) *and* (2.5) *and the mapping*

$$L^2(I; H) \times V \times H \to L^2(I; V) \times L^2(I; H), \quad (f, u_0, u_1) \mapsto (u, u_t)$$

is linear and continuous.

Proof. See Chapter IV, Theorem 1.1 in [90]. □

In fact, we actually obtain a higher temporal regularity of the solution.

Theorem 2.1.2.
Let (2.1) - (2.3) *hold. Furthermore, let* $u_0 \in V$, $u_1 \in H$ *and* $f \in L^2(I; H)$. *Then, after a possible modification on a set of measure zero, there exists a unique solution* $(u, u_t) \in C(\bar{I}; V) \times C(\bar{I}; H)$ *that satisfies* (2.4) *and* (2.5). *The mapping*

$$L^2(I; H) \times V \times H \to C(\bar{I}; V) \times C(\bar{I}; H), \quad (f, u_0, u_1) \mapsto (u, u_t)$$

is linear and continuous.

Proof. See Chapter 3, Theorem 8.2 in [91]. □

Remark 2.1.3.
Due to the improved regularity of $(u, u_t) \in C(\bar{I}; V) \times C(\bar{I}; H)$, the initial conditions (2.5) are well-defined.

If we admit a less regular right-hand side $f \in L^2(I;V^*)$ and initial data $u_0 \in H$, $u_1 \in V^*$, we obtain the following result.

Theorem 2.1.4.
Let (2.2) *and* (2.3) *hold and let* $t \mapsto a(t;v,w)$ *be twice continuously differentiable on* $\bar{I}$ *for all* $v, w \in V$. *Furthermore, let* $u_0 \in H$, $u_1 \in V^*$ *and* $f \in L^2(I;V^*)$. *Then, after a possible modification on a set of measure zero, there exists a unique solution* $(u, u_t) \in C(\bar{I};H) \times C(\bar{I};V^*)$ *that satisfies* (2.4) *and* (2.5). *The mapping*

$$L^2(I;V^*) \times H \times V^* \to C(\bar{I};H) \times C(\bar{I};V^*), \quad (f, u_0, u_1) \mapsto (u, u_t)$$

is linear and continuous.

Proof. See Chapter 3, Theorem 9.3, Theorem 9.4 and Remark 9.11 in [91]. □

By exploiting a higher temporal regularity of the right-hand side and the initial data, we are able to improve the temporal regularity of the solution.

Theorem 2.1.5 (Improved temporal regularity)**.**
Let (2.1) - (2.3) *hold and assume that the operator* A *in* (2.4) *does not depend on time* t. *Furthermore, we denote*

$$\begin{aligned} &\tilde{u}_0 = u_0, \quad \tilde{u}_1 = u_1, \quad \textit{and} \\ &\tilde{u}_l = \frac{\partial^{l-2}}{\partial t^{l-2}} f(0) - A\tilde{u}_{l-2}, \quad \textit{recursively for } l \geq 2. \end{aligned}$$

Now, let $k \geq 1$, $f \in H^{k-1}(I;H)$ *and the following compatibility conditions hold*

$$\tilde{u}_j \in V, \quad 0 \leq j \leq k-1, \quad \textit{and} \quad \tilde{u}_k \in H.$$

Then there exists a unique solution $u \in H^{k-1}(I;V) \cap H^k(I;H) \cap H^{k+1}(I;V^*)$ *to* (2.4) *and* (2.5).

Proof. See Satz 30.1 in [134]. □

Remark 2.1.6 (Improved spatial regularity)**.**
Alternatively, also the spatial regularity of the solution can be improved by assuming a higher spatial regularity of the right-hand side f and initial data u_0 and u_1. Similar compatibility conditions as in Theorem 2.1.5 are required in this case, cf. Chapter 7, Theorem 6 in [49].

With the help of Theorem 2.1.5 we can improve the temporal regularity of the solution if we admit a higher temporal regularity of the right-hand side. However, this result does not improve the spatial regularity of u compared to Theorem 2.1.1. On the other hand, following Remark 2.1.6, a higher spatial regularity of u can be achieved by assuming a higher spatial regularity of the right-hand side. For the forthcoming analysis, however, we seek for a setting that yields $u(t) \in V$ when $f(t)$ is only in V^*. This will be continued in the next section.

2.2 Extensions to a Parameterized Elliptic Operator

In the next step, we generalize the results to the case where the elliptic operator A depends on coefficients denoted by σ. To avoid confusion, we emphasize that σ is not related to the stress tensor $\bar{\sigma}$ from section 1.1.
For notational simplicity and also with the application of seismic tomography in mind, we restrict the analysis to time-independent coefficients and homogeneous initial conditions. We continue to work with hyperbolic equations of the following form. Let $\hat{n} \in \mathbb{N}$ and $\mathcal{S} \subset L^\infty(\Omega)^{\hat{n}}$ be a nonempty, open and bounded set. We consider a family of continuous bilinear forms $a(\sigma) : V \times V \to \mathbb{R}$ and, similar as before, we identify $a(\sigma)$ with the operator $A(\sigma) \in \mathcal{L}(V, V^*)$ by

$$a(\sigma)(v, w) = \langle A(\sigma)v, w\rangle_{V^*,V} \quad \forall\, v, w \in V.$$

Additionally, we introduce a function $\rho : \Omega \to \mathbb{R}$ and consider equations of the following form:

$$\rho u_{tt}(t) + A(\sigma)u(t) = f(t), \tag{2.6}$$

with initial data

$$u(0) = 0, \quad u_t(0) = 0. \tag{2.7}$$

For the further analysis, we compactly write (2.6) and (2.7) with the help of the operator $\hat{E}(u, \sigma) := \rho u_{tt} + A(\sigma)u$ as:

$$\hat{E}(u, \sigma) = f, \quad u(0) = 0, \quad u_t(0) = 0. \tag{2.8}$$

The variational form of (2.8) is given as follows:

$$\langle \rho u_{tt}(t), v\rangle_{V^*,V} + a(\sigma)(u(t), v) = \langle f(t), v\rangle_{V^*,V} \quad \forall\, v \in V \text{ and a.e. } t \text{ in } I. \tag{2.9}$$

We continue to work with the following assumptions on A and ρ.

Assumption 2.2.1.

(A1) $\mathcal{S} \subset L^\infty(\Omega)^{\hat{n}}$ *is a nonempty, open, and bounded set and* $A : L^\infty(\Omega)^{\hat{n}} \to \mathcal{L}(V, V^*)$ *is a bounded linear operator.*

(A2) There exist $\beta_1, \beta_2 > 0$ *independent of* σ *such that*

$$a(\sigma)(v, v) + \beta_1 \|v\|_H^2 \geq \beta_2 \|v\|_V^2 \quad \forall\, v \in V,\ \forall\, \sigma \in \mathcal{S}.$$

(A3) For all $\sigma \in \mathcal{S}$, $a(\sigma)$ *is symmetric, i.e.,* $a(\sigma)(v, w) = a(\sigma)(w, v) \quad \forall\, v, w \in V, \forall \sigma \in \mathcal{S}$.

(A4) $\rho : \Omega \to \mathbb{R}$ *is sufficiently smooth such that* $\rho v \in H^1(\Omega)$ *for all* $v \in H^1(\Omega)$.

(A5) There exist $\rho^a, \rho^b \in \mathbb{R}$ *with* $0 < \rho^a \leq \rho(x) \leq \rho^b$ *a.e. on* Ω.

Remark 2.2.2.
For $d = 2$ or $d = 3$ Assumption 2.2.1 (A4) is satisfied, for instance, if $\rho \in W^{1,3}(\Omega) \cap L^\infty(\Omega)$. In this case, we have $\nabla(\rho v) = v\nabla\rho + \rho\nabla v \in L^2(\Omega)$ by the embedding $H^1(\Omega) \hookrightarrow L^6(\Omega)$ and the generalized Hölder's inequality with $\frac{1}{p} + \frac{1}{q} = \frac{1}{2}$ and $p = 6$, $q = 3$ or, resp., $p = \infty$, $q = 2$.

Theorem 2.2.3.
Let Assumption 2.2.1 hold and let $\sigma \in \mathcal{S}$. Then the problem given by (2.8) possesses a unique solution in the following settings:

(i) For every $f \in L^2(I;H)$ there exists a unique solution

$$(u, u_t) \in C(\bar{I};V) \times C(\bar{I};H).$$

The mapping $L^2(I;H) \to C(\bar{I};V) \times C(\bar{I};H)$, $f \mapsto (u, u_t)$ is linear and bounded.

(ii) For every $f \in L^2(I;V^*)$ there exists a unique solution

$$(u, u_t) \in C(\bar{I};H) \times C(\bar{I};V^*).$$

The mapping $L^2(I;V^*) \to C(\bar{I};H) \times C(\bar{I};V^*)$, $f \mapsto (u, u_t)$ is linear and bounded.

Note that in order to achieve continuity of the solution in time, a suitable modification on a set of measure zero might be necessary.

Proof. The properties of ρ and σ ensure the applicability of general results for hyperbolic equations. In fact, we have $(\rho v, v)_H \geq \rho_a \|v\|_H^2$ for all $v \in V$ and thus, the usual energy estimates can be used. Furthermore, due to Assumption 2.2.1 (A2) and (A3), the existence and regularity of a unique solution as well as the dependence on the right-hand side f follows as in the proofs of Theorem 2.1.1 (resp., Theorem 2.1.2) and Theorem 2.1.4. For the sake of completeness, we show how to derive the energy estimates in the case of $f \in L^2(I;H)$. This is an easy generalization of the energy estimates derived in the proof of Chapter 3, Theorem 8.1 in [91]. See also Satz 29.1 in [134] or Chapter 7, Theorem 2 in [49] for similar results. The following derivation requires to consider an approximate solution which is sufficiently regular to give meaning to the dual pairings and time derivatives. Since this is analogously done in all the Theorems cited above, we omit the details and, for simplicity, denote the approximate solution again by u. Now, we can (formally) test (2.9) with $u_t(t) \in V$ and obtain for every fixed $t \in I$

$$\langle \rho u_{tt}(t), u_t(t) \rangle_{V^*,V} + a(\sigma)(u(t), u_t(t)) = (f(t), u_t(t))_H.$$

Hence, by the symmetry of a (Assumption 2.2.1 (A3)), we obtain

$$\frac{d}{dt}\left[(\rho u_t(t), u_t(t))_H + a(\sigma)(u(t), u(t))\right] = 2(f(t), u_t(t))_H.$$

Integrating from 0 to t and using the homogeneous initial data yields

$$\begin{aligned}(\rho u_t(t),&u_t(t))_H + a(\sigma)(u(t), u(t)) \\ &= (\rho u_t(0), u_t(0))_H + a(\sigma)(u(0), u(0)) + 2\int_0^t (f(\tau), u_t(\tau))_H \, d\tau \\ &= 2\int_0^t (f(\tau), u_t(\tau))_H \, d\tau.\end{aligned}$$

With Assumption 2.2.1 (A2) and $(\rho u_t(t), u_t(t))_H \geq \rho_a \|u_t(t)\|_H^2$, we further deduce

$$\rho_a \|u_t(t)\|_H^2 + \beta_2 \|u(t)\|_V^2 \leq \beta_1 \|u(t)\|_H^2 + 2\int_0^t \|f(\tau)\|_H \, \|u_t(\tau)\|_H \, d\tau. \tag{2.10}$$

Furthermore, we have

$$\|u(t)\|_H \leq \|u(0)\|_H + \int_0^t \|u_t(\tau)\|_H \, d\tau = \int_0^t \|u_t(\tau)\|_H \, d\tau.$$

Hence, it holds

$$\|u(t)\|_H^2 \leq C \int_0^t \|u_t(\tau)\|_H^2 \, d\tau. \tag{2.11}$$

From (2.10), we obtain with Young's inequality and (2.11)

$$\|u_t(t)\|_H^2 + \|u(t)\|_V^2 \leq C \int_0^t \|f(\tau)\|_H^2 \, d\tau + C \int_0^t \|u_t(\tau)\|_H^2 \, d\tau.$$

Finally, from Gronwall's Lemma we deduce the a priori estimate:

$$\|u_t(t)\|_H^2 + \|u(t)\|_V^2 \leq C\|f\|_{L^2(I;H)}^2.$$

□

We proceed with the main result of this section, which shows that the regularity of the solution can be improved if the right-hand side admits a higher temporal regularity.

Theorem 2.2.4.
Let Assumption 2.2.1 hold and let $\sigma \in \mathcal{S}$. We consider $f \in L^2(I;V^*)$ with $f = 0$ near $t = 0$ and, additionally, $f_t \in L^2(I;V^*)$. Then there exists a unique solution $u = u(\sigma)$ to the problem (2.8) that satisfies

$$u \in C(\bar{I};V), \quad u_t \in C(\bar{I};H), \quad u_{tt} \in C(\bar{I};V^*),$$

and the mapping

$$H^1(I;V^*) \to C(\bar{I};V) \times C(\bar{I};H) \times C(\bar{I};V^*), \quad f \mapsto (u, u_t, u_{tt}) \tag{2.12}$$

is linear and bounded. For fixed f, the solutions $u(\sigma)$ are uniformly bounded in $C(\bar{I};V) \cap C^1(\bar{I};H)$ for all $\sigma \in \mathcal{S}$.

Proof. The proof follows the lines of the proof of Chapter 3, Theorem 9.3 for the existence of very weak solutions in [91]. However, the crucial difference is the improved regularity in space that is obtained by utilizing the higher temporal regularity of the right-hand side. We consider

$$f_k \in H^1(I;H) \quad \text{with} \quad f_k \to f \quad \text{in } H^1(I;V^*),$$

and the problem

$$\rho(u_k)_{tt}(t) + A(\sigma)u_k(t) = f_k(t), \quad u_k(0) = 0, \quad (u_k)_t(0) = 0. \tag{2.13}$$

By formally differentiating with respect to t and substituting $(u_k)_t$ by w_k, we obtain

$$\rho(w_k)_{tt}(t) + A(\sigma)w_k(t) = (f_k)_t(t), \quad w_k(0) = 0, \quad (w_k)_t(0) = 0. \tag{2.14}$$

Here, we used that f is zero near $t = 0$. Since $(f_k)_t \in L^2(I;H)$, there exists a unique solution $w_k \in C(\bar{I};V) \cap C^1(\bar{I};H)$ to (2.14) by Theorem 2.2.3. We define the auxiliary function $v_k(t) := \int_0^t w_k(\tau) \, d\tau$ and obtain

$$v_k \in C^1(\bar{I};V), \quad (v_k)_t = w_k \in C(\bar{I};V), \quad (v_k)_{tt} = (w_k)_t \in C(\bar{I};H).$$

Integrating (2.14) in time and inserting the homogeneous initial data yields

$$0 = \int_0^t \rho(w_k)_{tt}(\tau) + A(\sigma)w_k(\tau) - (f_k)_t(\tau)\,d\tau = \rho(v_k)_{tt}(t) + A(\sigma)v_k(t) - f_k(t). \tag{2.15}$$

By subtracting the original equation (2.13) from (2.15), we obtain

$$\rho(v_k - u_k)_{tt}(t) + A(\sigma)(v_k - u_k)(t) = 0, \quad (v_k - u_k)(0) = 0, \quad (v_k - u_k)_t(0) = 0.$$

Hence, by the uniqueness of the solution due to Theorem 2.2.3, we obtain $u_k = v_k$ and $(u_k)_t = w_k$ in $[0, T]$. Thus, we have shown that v_k is the unique solution to (2.13).
The improved regularity in space remains to be shown. In [91], Chapter 3, proof of Theorem 9.3, the energy estimates

$$\|(u_k)_t(t)\|_H^2 + \|(u_k)_{tt}(t)\|_{V^*}^2 \le C\,\|(f_k)_t\|_{L^2(I;V^*)}^2\,, \tag{2.16}$$

$$\|(u_k)(t)\|_H^2 + \|(u_k)_t(t)\|_{V^*}^2 \le C\,\|(f_k)\|_{L^2(I;V^*)}^2 \tag{2.17}$$

are derived for a.a. $t \in I$. Furthermore, using Assumption 2.2.1 (A2), (A4) and (A5), we deduce

$$\begin{aligned}
\beta_2\,\|u_k(t)\|_V^2 &\le a(\sigma)(u_k(t), u_k(t)) + \beta_1\,\|u_k(t)\|_H^2 \\
&= \langle f_k(t) - \rho(u_k)_{tt}(t), u_k(t)\rangle_{V^*,V} + \beta_1\,\|u_k(t)\|_H^2 \\
&\le \|f_k(t)\|_{V^*}\,\|u_k(t)\|_V + |\langle (u_k)_{tt}(t), \rho u_k(t)\rangle_{V^*,V}| + \beta_1\,\|u_k(t)\|_H^2 \\
&\le (\|f_k(t)\|_{V^*} + C_\rho\,\|(u_k)_{tt}(t)\|_{V^*})\,\|u_k(t)\|_V + \beta_1\,\|u_k(t)\|_H^2 \\
&\le \left(\|f_k(t)\|_{V^*} + c_1\,\|(f_k)_t\|_{L^2(I;V^*)}\right)\|u_k(t)\|_V + c_2\,\|(f_k)\|_{L^2(I;V^*)}^2\,,
\end{aligned}$$

where we used (2.16) and (2.17) in the last inequality. Next, we use Young's inequality to obtain

$$\left(\|f_k(t)\|_{V^*} + c_1\,\|(f_k)_t\|_{L^2(I;V^*)}\right)\|u_k(t)\|_V \le$$
$$\frac{1}{2\beta_2}\left(\|f_k(t)\|_{V^*} + c_1\,\|(f_k)_t\|_{L^2(I;V^*)}\right)^2 + \frac{\beta_2}{2}\,\|u_k(t)\|_V^2\,.$$

Hence,

$$\frac{\beta_2}{2}\,\|u_k(t)\|_V^2 \le \frac{1}{2\beta_2}\left(\|f_k(t)\|_{V^*} + c_1\,\|(f_k)_t\|_{L^2(I;V^*)}\right)^2 + c_2\,\|(f_k)\|_{L^2(I;V^*)}^2\,. \tag{2.18}$$

Now, by the embedding $H^1(I;V^*) \hookrightarrow C(\bar{I};V^*)$, we can estimate $\|f_k(t)\|_{V^*} \le \|f_k\|_{C(\bar{I};V^*)} \le C\,\|f_k\|_{H^1(I;V^*)}$ and finally obtain by combining (2.16) and (2.18):

$$\sup_{0\le t\le T}\left(\|u_k(t)\|_V^2 + \|(u_k)_t(t)\|_H^2 + \|(u_k)_{tt}(t)\|_{V^*}^2\right) \le C\,\|f_k\|_{H^1(I;V^*)}^2\,. \tag{2.19}$$

Thus, u_k (resp. $(u_k)_t$, $(u_k)_{tt}$) remains in a bounded set of $L^2(I;V)$ (resp. $L^2(I;H)$, $L^2(I;V^*)$). We can therefore extract a weakly convergent subsequence $u_\kappa \rightharpoonup \tilde{u}$ in $L^2(I;V)$ as well as

$u_\kappa \rightharpoonup \tilde{u}$ in $H^1(I;H)$ and $u_\kappa \rightharpoonup \tilde{u}$ in $H^2(I;V^*)$. By the embeddings $H^1(I;H) \hookrightarrow\hookrightarrow C(\bar{I};H)$ and $H^2(I;V^*) \hookrightarrow\hookrightarrow C^1(\bar{I};V^*)$, we obtain $u_\kappa \to \tilde{u}$ in $C(\bar{I};H)$ and $(u_\kappa)_t \to \tilde{u}_t$ in $C(\bar{I};V^*)$. Therefore, $0 = u_\kappa(0) \to \tilde{u}(0)$ in H, i.e., $\tilde{u}(0) = 0$, and $0 = (u_\kappa)_t(0) \to \tilde{u}_t(0)$ in V^*, i.e., $\tilde{u}_t(0) = 0$. Hence, by passing to the limits in (2.13) we find that $\tilde{u}$ is a solution to (2.8). The continuous dependence in (2.12) follows from (2.19), cf. Remark 9.11 in [91]. By Assumption 2.2.1 (A2), all constants are independent of σ. Thus, the set of unique solutions $u(\sigma) \in C(\bar{I};V) \cap C^1(\bar{I};H)$ is uniformly bounded for all $\sigma \in \mathcal{S}$ and fixed $f \in H^1(I;V^*)$. □

Theorem 2.2.4 can easily be generalized if the right-hand side admits an even higher temporal regularity.

Corollary 2.2.5.
Let Assumption 2.2.1 hold. We denote $F^k := H^k(I;V^*)$, $k \geq 1$, and $F_0^k \subset F^k$ as the subset of source functions with $f = 0$ near $t = 0$. Then, for every $f \in F_0^k$ and fixed $\sigma \in \mathcal{S}$, there exists a unique solution u to the problem (2.8). The mapping

$$F_0^k \to C^{k-1}(\bar{I};V) \cap C^k(\bar{I};H) \cap C^{k+1}(\bar{I};V^*), \quad f \mapsto u$$

is linear and bounded. Furthermore, for fixed $f \in F_0^k$, the solutions $u(\sigma)$ are uniformly bounded in $C^{k-1}(\bar{I};V) \cap C^k(\bar{I};H)$ for all $\sigma \in \mathcal{S}$.

Proof. We only have to show the improved temporal regularity of the solution, the rest follows analogously to Theorem 2.2.4. The regularity of u is obtained by induction and Theorem 2.2.4, which gives the case $k = 1$. Now, assume the statement is true for $k \in \mathbb{N}$ and let $f \in F_0^{k+1}$. Similar as in the proof of Theorem 2.2.4 we consider

$$\rho(w)_{tt}(t) + A(\sigma)w(t) = f_t(t), \quad w(0) = 0, \quad w_t(0) = 0. \tag{2.20}$$

By the induction hypothesis, there exists a unique solution $w \in C^{k-1}(\bar{I};V) \cap C^k(\bar{I};H) \cap C^{k+1}(\bar{I};V^*)$ to (2.20). Using the auxiliary function $v \in C^k(\bar{I};V) \cap C^{k+1}(\bar{I};H) \cap C^{k+2}(\bar{I};V^*)$ defined by $v(t) = \int_0^t w(\tau)\,d\tau$ and integrating (2.20) in time, we find that v is the unique solution to

$$\rho v_{tt}(t) + A(\sigma)v(t) = f(t), \quad v(0) = 0, \quad v_t(0) = 0.$$

which concludes the induction step. □

Remark 2.2.6.
General hyperbolic equations would require stronger compatibility conditions (cf. Theorem 2.1.5 and Chapter 7, Theorem 6 in [49]) to establish similar results as in Theorem 2.2.4 and Corollary 2.2.5. Since we work with homogeneous initial data, these requirements are implicitly given by the assumption that f is zero near $t = 0$.

In the next step, we establish the continuity and Fréchet differentiability of the solution $u = u(\sigma)$ with respect to the coefficients σ.

Theorem 2.2.7.
Let Assumption 2.2.1 hold. For all $\sigma \in \mathcal{S}$ and $f \in F_0^{k+1}$, $k \geq 1$, the problem given by (2.8) possesses a unique solution $u(\sigma) \in C^{k-1}(\bar{I};V) \cap C^k(\bar{I};H)$ and the mapping

$$\mathcal{S} \to C^{k-1}(\bar{I};V) \cap C^k(\bar{I};H) \quad \sigma \mapsto u(\sigma)$$

is Lipschitz continuous.

Proof. By Corollary 2.2.5 there exists a unique solution $u(\sigma) \in C^k(\bar{I};V) \cap C^{k+1}(\bar{I};H)$ to (2.8). Let $s \in L^\infty(\Omega)^{\hat{n}}$ such that $\sigma + s \in \mathcal{S}$. Then there also exists a unique solution $\tilde{u}(s) := u(\sigma+s) \in C^k(\bar{I};V) \cap C^{k+1}(\bar{I};H)$ to $\hat{E}(\tilde{u}(s), \sigma+s) = f$ with homogeneous initial data and $\|\tilde{u}(s)\|_{C^k(\bar{I};V)} \leq C$ uniformly for all s with $\|s\|_{L^\infty(\Omega)^{\hat{n}}} \leq \delta$ and δ sufficiently small. We denote the difference by $h := \tilde{u}(s) - u(\sigma)$ and obtain:

$$\begin{aligned}
0 &= \hat{E}(\tilde{u}(s), \sigma+s) - \hat{E}(u(\sigma), \sigma) \\
&= \rho(\tilde{u}_{tt}(s) - u_{tt}(\sigma)) + A(\sigma+s)\tilde{u}(s) - A(\sigma)u(\sigma) \\
&= \rho h_{tt} + A(\sigma)h + (A(\sigma+s) - A(\sigma))\tilde{u}(s) \\
&= \rho h_{tt} + A(\sigma)h + A(s)\tilde{u}(s),
\end{aligned}$$

where the last step follows by Assumption 2.2.1 (A1), i.e., the linear dependence of A on the coefficients. Furthermore, due to $f \in F_0^{k+1}$ and the homogeneous initial conditions, we also have that $\tilde{u}(s)$ is zero near $t = 0$. Hence, we obtain that h satisfies

$$\hat{E}(h, \sigma) = -A(s)\tilde{u}(s), \quad h(0) = 0, \quad h_t(0) = 0. \tag{2.21}$$

Since $\tilde{u}(s) \in C^k(\bar{I};V) \cap C^{k+1}(\bar{I};H)$, we obtain $-A(s)\tilde{u}(s) \in C^k(\bar{I};V^*) \subset H^k(I;V^*)$. Hence, $-A(s)\tilde{u}(s) \in F_0^k$ and we can apply Corollary 2.2.5 to obtain the estimate

$$\begin{aligned}
\|h\|_{C^{k-1}(\bar{I};V)} + \|h_t\|_{C^{k-1}(\bar{I};H)} &\leq C_1 \|A(s)\tilde{u}(s)\|_{H^k(I;V^*)} \\
&\leq C_2 \|s\|_{L^\infty(\Omega)^{\hat{n}}} \|\tilde{u}(s)\|_{C^k(\bar{I};V)} \\
&\leq C_3 \|s\|_{L^\infty(\Omega)^{\hat{n}}}.
\end{aligned}$$

□

Remark 2.2.8.
Note that it is necessary to exploit a higher temporal regularity of the source term in order to show the continuity (and differentiability) of the state $u(\sigma)$ with respect to the coefficients σ. In fact, consider $f \in L^2(I;H)$ and $h := \tilde{u}(s) - u(\sigma)$ as in the proof of Theorem 2.2.7. We obtain that h satisfies

$$\hat{E}(h, \sigma) = -A(s)\tilde{u}(s), \quad h(0) = 0, \quad h_t(0) = 0,$$

and $\|A(s)\tilde{u}(s)\|_{L^2(I;V^*)} \leq C\|s\|_{L^\infty(\Omega)^{\hat{n}}}$. Thus, from the estimate in Theorem 2.2.3 we would only get that

$$\sigma \in \mathcal{S} \mapsto (u(\sigma), u(\sigma)_t) \in C(\bar{I};H) \times C(\bar{I};V^*)$$

is continuous, although $(u(\sigma), u(\sigma)_t)$ is bounded in the stronger space $C(\bar{I};V) \times C(\bar{I};H)$. For $f \in L^2(I;V^*)$ the situation is even more critical, because in this case we have $\tilde{u}(s) \in C(\bar{I};H)$, but $A(s) \in \mathcal{L}(V, V^*)$.

In the next step, we prove that the solution operator $\sigma \mapsto u(\sigma)$ is Fréchet differentiable if we admit an even higher temporal regularity of the right-hand side.

Theorem 2.2.9.
Let Assumption 2.2.1 hold and consider $f \in F_0^{k+2}$ with $k \geq 1$. Then the solution operator $\mathcal{S} \to C^{k-1}(\bar{I}; V) \cap C^k(\bar{I}; H)$, $\sigma \mapsto u(\sigma)$ is Fréchet differentiable.

Proof. Let $\sigma \in \mathcal{S}$ and $s \in L^\infty(\Omega)^{\hat{n}}$ with $\|s\|_{L^\infty(\Omega)^{\hat{n}}} \leq \delta$ for δ sufficiently small and such that $\sigma + s \in \mathcal{S}$. Similar as in the proof of Theorem 2.2.7 we define $\tilde{u}(s)$ as the solution to $\hat{E}(\tilde{u}(s), \sigma + s) = f$ with homogeneous initial conditions. In order to show the Fréchet differentiability of the solution operator, we consider the solution $d(\sigma, s)$ to the linearized state equation

$$\hat{E}_u(u(\sigma), \sigma))d(\sigma, s) = -\hat{E}_\sigma(u(\sigma), \sigma)\, s,$$

i.e., $d(\sigma, s)$ satisfies

$$\rho\, d_{tt}(\sigma, s) + A(\sigma)d(\sigma, s) = -A(s)u(\sigma), \quad d(\sigma, s)(0) = 0, \quad d_t(\sigma, s)(0) = 0. \tag{2.22}$$

In particular, $s \in L^\infty(\Omega)^{\hat{n}} \mapsto d(\sigma, s) \in C^k(\bar{I}; V) \cap C^{k+1}(\bar{I}; H)$ is linear and bounded by Corollary 2.2.5. Let $r := \tilde{u}(s) - u(\sigma) - d(\sigma, s)$ denote the remainder term of the Fréchet derivative. By combining the hyperbolic equations for $\tilde{u}(s)$, $u(\sigma)$ and $d(\sigma, s)$ we deduce

$$\begin{aligned} 0 &= \rho\, r_{tt} + A(\sigma + s)\tilde{u}(s) - A(\sigma)(u(\sigma) + d(\sigma, s)) - A(s)u(\sigma) \\ &= \rho\, r_{tt} + A(\sigma)r + A(s)(\tilde{u}(s) - u(\sigma)). \end{aligned}$$

Hence, r satisfies $\hat{E}(r, \sigma) = -A(s)(\tilde{u}(s) - u(\sigma))$ with homogeneous initial data. With $h = \tilde{u}(s) - u(\sigma)$, we obtain by Theorem 2.2.7 that h is zero near $t = 0$, $h \in C^k(\bar{I}; V) \cap C^{k+1}(\bar{I}; H)$ and $\|h\|_{C^k(\bar{I};V)} \leq C\|s\|_{L^\infty(\Omega)^{\hat{n}}}$. Thus, we also have $-A(s)h \in F_0^k$ and applying Corollary 2.2.5 yields $r \in C^{k-1}(\bar{I}; V) \cap C^k(\bar{I}; H)$ as well as the estimate

$$\begin{aligned} \|r\|_{C^{k-1}(\bar{I};V)} + \|r_t\|_{C^{k-1}(\bar{I};H)} &\leq C_1 \|A(s)h\|_{H^k(I;V^*)} \\ &\leq C_2 \|s\|_{L^\infty(\Omega)^{\hat{n}}} \|h\|_{C^k(\bar{I};V)} \\ &\leq C_3 \|s\|^2_{L^\infty(\Omega)^{\hat{n}}}. \end{aligned} \tag{2.23}$$

Hence, the remainder term r satisfies

$$\|r\|_{C^{k-1}(\bar{I};V)} + \|r_t\|_{C^{k-1}(\bar{I};H)} = o\left(\|s\|_{L^\infty(\Omega)^{\hat{n}}}\right) \quad \text{for} \quad \|s\|_{L^\infty(\Omega)^{\hat{n}}} \to 0.$$

□

Remark 2.2.10.
Note that the Lipschitz continuity as well as the Fréchet differentiability of the solution operator comes with a loss of temporal regularity of the solution. Indeed, for $f \in F_0^{k+2}$ we obtain by Corollary 2.2.5 that there exists a unique solution $u \in C^{k+1}(\bar{I}; V)$. However, by the preceding Theorems, the mapping $\sigma \mapsto u(\sigma)$ is only Lipschitz-continuous from $\mathcal{S}$ to $C^k(\bar{I}; V)$ and Fréchet differentiable as a map from $\mathcal{S}$ to $C^{k-1}(\bar{I}; V)$.

By further pursuing the techniques from Theorem 2.2.9, we can also establish higher-order Fréchet differentiability of the solution operator under suitable assumptions on the right-hand side. We proceed as follows. First, we create a candidate for the k-th-order derivative by differentiating the state equation k times. Next, we show that this directional derivative itself

satisfies a hyperbolic equation of the same kind and finally, we apply the regularity results that we established previously and prove that the sensitivity is actually the Fréchet derivative of order k. Here, we will require the following notation. Let $\mathbf{s} = (s_1, \dots, s_i)$ be a tuple of length i with $s_j \in L^\infty(\Omega)^{\hat{n}}$, $j = 1, \dots, i$. For $1 \le j \le i$ we denote by $\mathbf{s}_{-j} = (s_1, \dots, s_{j-1}, s_{j+1}, \dots, s_i)$ the $(i-1)$-tuple of all elements of $\mathbf{s}$ except s_j.
We proceed with the following Lemma.

Lemma 2.2.11.
Let Assumption 2.2.1 hold and let $k \ge 1$, $f \in F_0^{k+1}$ and $\sigma \in \mathcal{S}$. Consider $1 \le i \le k$ and $\mathbf{s} = (s_1, \dots, s_i)$ with $s_j \in L^\infty(\Omega)^{\hat{n}}$ and $\sigma + s_j \in \mathcal{S}$, $j = 1, \dots, i$. Furthermore, we set $d^0(\sigma) := u(\sigma)$ and define $d^i := d^i(\sigma, \mathbf{s})$, as the solution of

$$\hat{E}(d^1, \sigma) = -A(s_1)\, d^0(\sigma), \qquad d^1(0) = 0, \qquad d_t^1(0) = 0 \qquad (i = 1), \tag{2.24}$$

$$\hat{E}(d^i, \sigma) = -\sum_{j=1}^{i} A(s_j)\, d^{i-1}(\sigma, \mathbf{s}_{-j}), \qquad d^i(0) = 0, \qquad d_t^i(0) = 0 \qquad (2 \le i \le k). \tag{2.25}$$

Then $d^i \in C^{k-i}(\bar{I}; V) \cap C^{k-i+1}(\bar{I}; H)$ is uniquely determined by (2.24), (2.25) and

$$\left\|d^i\right\|_{C^{k-i}(\bar{I};V)} + \left\|d_t^i\right\|_{C^{k-i}(\bar{I};H)} \le C \prod_{j=1}^{i} \|s_j\|_{L^\infty(\Omega)^{\hat{n}}}, \quad 1 \le i \le k, \tag{2.26}$$

with a constant $C > 0$ that depends on f and can be chosen uniformly on $\mathcal{S}$.

Proof. For $i = 1$ and $s^1 = (s_1)$ the first part in (2.24) reads as follows:

$$\rho\, d_{tt}^1(\sigma, s^1) + A(\sigma)\, d^1(\sigma, s^1) + A(s_1)\, u(\sigma) = 0,$$

which is the linearized state equation that was obtained in (2.22). By Corollary 2.2.5, $u(\sigma) \in C^k(\bar{I}; V)$ and since $f \in F_0^{k+1}$ also $u(\sigma)$ is zero near $t = 0$. Thus, $A(s_1)\, u(\sigma) \in F_0^k$ and we deduce the existence of a unique solution d^1 to (2.24) as well as the estimate

$$\left\|d^1\right\|_{C^{k-1}(\bar{I};V)} + \left\|d_t^1\right\|_{C^{k-1}(\bar{I};H)} \le C \|s_1\|_{L^\infty(\Omega)^{\hat{n}}},$$

where the constant $C > 0$ depends on f and can be chosen uniformly on $\mathcal{S}$. Furthermore, d^1 is zero near $t = 0$. Note that for $i = 1$ and $\mathbf{s} = (s_1)$, $\mathbf{s}_{-1}$ is empty and by setting $d^0(\sigma, \emptyset) = d^0(\sigma) = u(\sigma)$ the right-hand sides in (2.24) and (2.25) coincide. Assume now that the statement is true for $i - 1$ with $1 < i \le k$. By the induction hypothesis, the right-hand side in (2.25) is in $F_0^{k-(i-1)}$. Hence, applying Corollary 2.2.5 again gives the existence of a unique solution $d^i \in C^{k-i}(\bar{I}; V) \cap C^{k-i+1}(\bar{I}; H)$ to (2.25) and

$$\begin{aligned}\left\|d^i\right\|_{C^{k-i}(\bar{I};V)} + \left\|d_t^i\right\|_{C^{k-i}(\bar{I};H)} &\le \sum_{j=1}^{i} C \|s_j\|_{L^\infty(\Omega)^{\hat{n}}} \prod_{p=1, p\ne j}^{i} \|s_p\|_{L^\infty(\Omega)^{\hat{n}}} \\ &= i\, C \prod_{j=1}^{i} \|s_j\|_{L^\infty(\Omega)^{\hat{n}}},\end{aligned}$$

which yields (2.26) for d^i. □

Lemma 2.2.12.
Let Assumption 2.2.1 hold and consider $1 \leq i \leq k$, $f \in F_0^{k+1}$, $\sigma \in \mathcal{S}$, $\mathbf{s} = (s_1, \dots, s_i)$ *with* $s_j \in L^\infty(\Omega)^{\hat{n}}$ *and* $\sigma + s_j \in \mathcal{S}$, $j = 1, \dots, i$. *Furthermore, let* $s \in L^\infty(\Omega)^{\hat{n}}$ *with* $\sigma + s \in \mathcal{S}$. *We define* $h^i(\sigma, s, \mathbf{s}) := d^i(\sigma + s, \mathbf{s}) - d^i(\sigma, \mathbf{s})$, *where* d^i *is defined by* (2.24) *(resp.* (2.25)*), and, furthermore,* $h^0(\sigma, s, \emptyset) := h^0(\sigma, s) := u(\sigma + s) - u(\sigma)$.
Moreover, let $r^i(\sigma, \mathbf{s}) := d^{i-1}(\sigma + s_i, \mathbf{s}_{-i}) - d^{i-1}(\sigma, \mathbf{s}_{-i}) - d^i(\sigma, \mathbf{s})$.
Then $h^i = h^i(\sigma, s, \mathbf{s}) \in C^{k-i}(\bar{I}; V) \cap C^{k-i+1}(\bar{I}; H)$ *satisfies:*

$$\hat{E}(h^i, \sigma) = -\sum_{j=1}^{i} A(s_j) h^{i-1}(\sigma, s, \mathbf{s}_{-j}) - A(s) d^i(\sigma + s, \mathbf{s}), \tag{2.27}$$

and $h^i(0) = 0$, $h_t^i(0) = 0$.
Furthermore, $r^i = r^i(\sigma, \mathbf{s}) \in C^{k-i}(\bar{I}; V) \cap C^{k-i+1}(\bar{I}; H)$ *satisfies:*

$$\hat{E}(r^i, \sigma) = -\sum_{j=1}^{i-1} A(s_j) r^{i-1}(\sigma, \mathbf{s}_{-j}) - A(s_i) h^{i-1}(\sigma, s_i, \mathbf{s}_{-i}), \tag{2.28}$$

and $r^i(0) = 0$, $r_t^i(0) = 0$.

Proof. By inserting $d^i(\sigma + s, \mathbf{s})$ and $d^i(\sigma, \mathbf{s})$ into (2.24) (resp. (2.25)) and subtracting both equations, we obtain

$$\begin{aligned} 0 &= \rho\, h_{tt}^i(\sigma, s, \mathbf{s}) + A(\sigma + s) d^i(\sigma + s, \mathbf{s}) - A(\sigma) d^i(\sigma, \mathbf{s}) \\ &\quad + \sum_{j=1}^{i} A(s_j) \left(d^{i-1}(\sigma + s, \mathbf{s}_{-j}) - d^{i-1}(\sigma, \mathbf{s}_{-j}) \right) \\ &= \hat{E}(h^i, \sigma) + A(s) d^i(\sigma + s, \mathbf{s}) + \sum_{j=1}^{i} A(s_j)\, h^{i-1}(\sigma, s, \mathbf{s}_{-j}). \end{aligned}$$

Next, we note that $r^i(\sigma, \mathbf{s}) = h^{i-1}(\sigma, s_i, \mathbf{s}_{-i}) - d^i(\sigma, \mathbf{s})$. Since the definition of h^0 coincides with h defined in the proof of Theorem 2.2.7, we obtain from (2.21) and (2.24)

$$\hat{E}(r^1, \sigma) = -A(s_1) u(\sigma + s_1) + A(s_1) u(\sigma) = -A(s_1) h^0(\sigma, s_1).$$

Furthermore, for $1 < i \leq k$ we obtain from (2.25) for $d^i = d^i(\sigma, \mathbf{s})$ and from (2.27) for $h^{i-1} = h^{i-1}(\sigma, s_i, \mathbf{s}_{-i})$:

$$\begin{aligned} \hat{E}(r^i, \sigma) &= \hat{E}(h^{i-1}, \sigma) - \hat{E}(d^i, \sigma) \\ &= -\sum_{j=1}^{i-1} A(s_j)\, r^{i-1}(\sigma, \mathbf{s}_{-j}) - A(s_i) \left(d^{i-1}(\sigma + s_i, \mathbf{s}_{-i}) - d^{i-1}(\sigma, \mathbf{s}_{-i}) \right) \\ &= -\sum_{j=1}^{i-1} A(s_j)\, r^{i-1}(\sigma, \mathbf{s}_{-j}) - A(s_i)\, h^{i-1}(\sigma, s_i, \mathbf{s}_{-i}). \end{aligned}$$

□

Now, we can prove the main statement. Here, we apply the same technique as in the proof of Theorem 2.2.9.

Theorem 2.2.13.
Let Assumption 2.2.1 hold and consider $f \in F_0^{k+2}$ with $k \geq 1$. Then the solution operator $\mathcal{S} \to C(\bar{I};V) \cap C^1(\bar{I};H)$, $\sigma \mapsto u(\sigma)$ is k-times Lipschitz continuously Fréchet differentiable.

Proof. Let $1 \leq i \leq k$, $\sigma \in \mathcal{S}$ and $s_j \in L^\infty(\Omega)^{\hat{n}}$ with $\sigma + s_j \in \mathcal{S}$, $j = 1, \dots, i+1$. In the first step, we inductively derive $h^i = h^i(\sigma, s_{i+1}, \mathbf{s})$ with $\mathbf{s} = (s_1, \dots, s_i)$ is zero near $t = 0$ as well as the estimate

$$\left\|h^i\right\|_{C^{k-i}(\bar{I};V)} + \left\|h^i_t\right\|_{C^{k-i}(\bar{I};H)} \leq C \prod_{j=1}^{i+1} \|s_j\|_{L^\infty(\Omega)^{\hat{n}}}\,, \tag{2.29}$$

where the constant $C > 0$ depends on f and can be chosen uniformly on $\mathcal{S}$.
For $i = 0$, Theorem 2.2.7 yields $h^0 = u(\sigma + s_1) - u(\sigma) \in C^k(\bar{I};V)$ and

$$\left\|h^0\right\|_{C^k(\bar{I};V)} + \left\|h^0_t\right\|_{C^k(\bar{I};H)} \leq C \|s_1\|_{L^\infty(\Omega)^{\hat{n}}}\,.$$

Note that here the temporal regularity is shifted by one, since we have $f \in F_0^{k+2}$ instead of F_0^{k+1} this time. Furthermore, since $u(\sigma)$ and $u(\sigma + s_1)$ are zero near $t = 0$, the same holds for h^0.
Assume now that the induction hypothesis is true for $i - 1$. By Lemma 2.2.11, we obtain $A(s_{i+1})d^i \in F_0^{k-i+1}$ with $d^i = d^i(\sigma + s_{i+1}, \mathbf{s})$ as well as the estimate

$$\left\|d^i\right\|_{C^{k-i+1}(\bar{I};V)} + \left\|d^i_t\right\|_{C^{k-i+1}(\bar{I};H)} \leq C \prod_{j=1}^{i} \|s_j\|_{L^\infty(\Omega)^{\hat{n}}}\,, \tag{2.30}$$

where C is independent of s_{i+1}. Note that also the temporal regularity of d^i has been shifted by one since $f \in F_0^{k+2}$. Thus, the right-hand side in (2.27) is in F_0^{k-i+1} and by Corollary 2.2.5 we deduce

$$\begin{aligned}
&\left\|h^i\right\|_{C^{k-i}(\bar{I};V)} + \left\|h^i_t\right\|_{C^{k-i}(\bar{I};H)} \\
&\quad \leq \sum_{j=1}^{i} C\ \|s_j\|_{L^\infty(\Omega)^{\hat{n}}} \left\|h^{i-1}(\sigma, s_{i+1}, \mathbf{s}_{-j})\right\|_{C^{k-i+1}(\bar{I};V)} \\
&\qquad + C \|s_{i+1}\|_{L^\infty(\Omega)^{\hat{n}}} \left\|d^i(\sigma + s_{i+1}, \mathbf{s})\right\|_{C^{k-i+1}(\bar{I};V)} \leq C \prod_{j=1}^{i+1} \|s_j\|_{L^\infty(\Omega)^{\hat{n}}}\,.
\end{aligned}$$

Again, since the right-hand side in (2.27) is zero near $t = 0$, the same holds for h^i.
Finally, we note that $r^i = r^i(\sigma, \mathbf{s})$ as defined in (2.28) is the remainder term of the i-th Fréchet derivative and show by induction the following estimate

$$\left\|r^i\right\|_{C^{k-i}(\bar{I};V)} + \left\|r^i_t\right\|_{C^{k-i}(\bar{I};H)} \leq C \|s_i\|^2_{L^\infty(\Omega)^{\hat{n}}} \prod_{j=1}^{i-1} \|s_j\|_{L^\infty(\Omega)^{\hat{n}}}\,. \tag{2.31}$$

For $i = 1$, we have already shown in (2.23) that

$$\left\|r^1\right\|_{C^{k-1}(\bar{I};V)} + \left\|r^1_t\right\|_{C^{k-1}(\bar{I};H)} \leq C \|s_1\|^2_{L^\infty(\Omega)^{\hat{n}}}\,.$$

Assume now (2.31) is true for $i-1$. Then we obtain by Lemma 2.2.12 and the previous estimate for h^{i-1} that the right-hand side in (2.28) is in F_0^{k-i+1} and

$$\begin{aligned}
&\left\|r^i\right\|_{C^{k-i}(\bar{I};V)} + \left\|r_t^i\right\|_{C^{k-i}(\bar{I};H)} \\
&\quad\leq C_1 \sum_{j=1}^{i-1} \|s_j\|_{L^\infty(\Omega)^{\hat{n}}} \|r^{i-1}(\sigma, \mathbf{s}_{-j})\|_{C^{k-i+1}(\bar{I};V)} \\
&\qquad + C_2 \|s_i\|_{L^\infty(\Omega)^{\hat{n}}} \|h^{i-1}(\sigma, s_i, \mathbf{s}_{-i})\|_{C^{k-i+1}(\bar{I};V)} \\
&\quad\leq C_3 \sum_{j=1}^{i-1} \|s_j\|_{L^\infty(\Omega)^{\hat{n}}} \|s_i\|^2_{L^\infty(\Omega)^{\hat{n}}} \prod_{\substack{p=1 \\ p\neq j}}^{i-1} \|s_p\|_{L^\infty(\Omega)^{\hat{n}}} + C_4 \|s_i\|_{L^\infty(\Omega)^{\hat{n}}} \prod_{j=1}^{i} \|s_j\|_{L^\infty(\Omega)^{\hat{n}}} \\
&\quad\leq C \|s_i\|^2_{L^\infty(\Omega)^{\hat{n}}} \prod_{j=1}^{i-1} \|s_j\|_{L^\infty(\Omega)^{\hat{n}}} .
\end{aligned}$$

Hence, the remainder term of the i-th derivative satisfies

$$\left\|r^i\right\|_{C^{k-i}(\bar{I};V)} = o\left(\|s_i\|_{L^\infty(\Omega)^{\hat{n}}}\right) \quad \text{for} \quad \|s_i\|_{L^\infty(\Omega)^{\hat{n}}} \to 0, \quad 1 \leq i \leq k.$$

The Lipschitz continuity of the derivative follows from (2.29). □

There are two immediate generalizations of Theorem 2.2.13 that follow from Corollary 2.2.5 and are stated below.

Corollary 2.2.14.
Let Assumption 2.2.1 hold and consider $k \geq 1$, $l \geq 0$ and $f \in F_0^{k+l+2}$. Then it holds:

1. *The solution operator $\sigma \mapsto u(\sigma)$ is k-times Lipschitz continuously Fréchet differentiable as a map from $\mathcal{S}$ to $C^l(\bar{I}; V)$.*
2. *The solution operator $\sigma \mapsto u(\sigma)$ is $(k+1)$-times Lipschitz continuously Fréchet differentiable as a map from $\mathcal{S}$ to $C^l(\bar{I}; H)$.*

Remark 2.2.15.
Very recently and without the author's knowledge, similar results as in this section have been established independently in [16] and [78]. While the first paper considers general hyperbolic evolution equations, the latter is restricted to the acoustic wave equation. Both have in common that they work with stronger assumptions on the right-hand sides, namely f in $H^k(I; H)$ or even $C^k(\bar{I}, H)$. Since the seismic sources we consider are modeled as smoothed Dirac measures, requiring only $f \in H^k(I; V^*)$ is clearly advantageous.

2.3 Generalizations Involving Superposition Operators

In this section, we seek to further generalize the dependence of the elliptic operator A on the coefficients. Specifically, we want to replace the coefficients σ by a (nonlinear) superposition operator that acts on some material parameters m before they enter the elliptic operator. The purpose of this is twofold. On the one hand, we want to be able to treat various parameterizations of different physical quantities in a unified way and, on the other hand, we will use a superposition operator to ensure the uniform coercivity of A on the whole space of material parameters without the need of explicit constraints. In order to clarify the notation and to distinguish it from the preceding subsection, we call m parameters instead of coefficients.
We consider superposition operators of the following form:

Definition 2.3.1 (Nemytskii or superposition operator)**.**
Let $\Omega \subset \mathbb{R}^d$ be a bounded measurable set with finite measure and $\varphi : \mathbb{R}^n \to \mathbb{R}^{\hat{n}}$ continuous. The operator Φ defined by $\Phi(m)(x) := \varphi(m(x))$ that assigns a function $\Phi(m) : \Omega \to \mathbb{R}^{\hat{n}}$ to every $m : \Omega \to \mathbb{R}^n$ is called Nemytskii or superposition operator.

Note that a more general definition of Nemytskii operators can, for instance, be found in [122], section 4.3. In particular, it would be possible to consider functions $\varphi : \Omega \times \mathbb{R}^n \to \mathbb{R}^{\hat{n}}$ that depend on x and m, but this is not necessary for our purposes.
In what follows, we will collect some properties on the differentiability of the superposition operator. Similar results can be found, for instance, in [122] (section 4.3), [109] (section 10.3), or the monographs [5, 44].

Lemma 2.3.2.
Let $\varphi : \mathbb{R}^n \to \mathbb{R}^{\hat{n}}$ be Lipschitz continuous. Then the superposition operator Φ defined by $\Phi(m)(x) := \varphi(m(x))$ is Lipschitz continuous from $L^\infty(\Omega)^n$ to $L^\infty(\Omega)^{\hat{n}}$.

Proof. For every $m \in L^\infty(\Omega)^n$ there exists a $C > 0$ such that $\|m(x)\|_\infty \leq C$ for almost all $x \in \Omega$, where $\|.\|_\infty$ is the infinity norm in $\mathbb{R}^{\hat{n}}$. Hence, we obtain

$$\begin{aligned}\|\Phi(m)\|_{L^\infty(\Omega)^{\hat{n}}} &= \|\varphi(m(.))\|_{L^\infty(\Omega)^{\hat{n}}} \leq \|\varphi(0)\|_\infty + \|\varphi(m(.)) - \varphi(0)\|_{L^\infty(\Omega)^{\hat{n}}} \\ &\leq \|\varphi(0)\|_\infty + L\,\|m\|_{L^\infty(\Omega)^n},\end{aligned}$$

where L denotes the Lipschitz constant of φ. Hence, $\Phi(m) \in L^\infty(\Omega)^{\hat{n}}$. Furthermore, for arbitrary $m_1, m_2 \in L^\infty(\Omega)^n$ we obtain

$$\|\Phi(m_1) - \Phi(m_2)\|_{L^\infty(\Omega)^{\hat{n}}} = \|\varphi(m_1(.)) - \varphi(m_2(.))\|_{L^\infty(\Omega)^{\hat{n}}} \leq L\|m_1 - m_2\|_{L^\infty(\Omega)^n}.$$

Thus, Φ maps Lipschitz continuously from $L^\infty(\Omega)^n$ to $L^\infty(\Omega)^{\hat{n}}$. □

In the same way, differentiability properties of φ can be carried over to Φ.

Lemma 2.3.3.
Let $\varphi : \mathbb{R}^n \to \mathbb{R}^{\hat{n}}$ be k-times Lipschitz continuously differentiable. Then the superposition operator $\Phi : L^\infty(\Omega)^n \to L^\infty(\Omega)^{\hat{n}}$, $\Phi(m)(x) := \varphi(m(x))$ is k-times Lipschitz continuously Fréchet differentiable with

$$\left(\Phi^{(k)}(m)(h_1, \ldots, h_k)\right)(x) = \varphi^{(k)}(m(x))(h_1(x), \ldots, h_k(x)) \quad \forall\ h_1, \ldots, h_k \in L^\infty(\Omega)^n. \quad (2.32)$$

Proof. A similar proof can be found in [122], Lemma 4.12 and Lemma 4.13, for the case $n = \hat{n} = 1$. We only show the case $k = 1$ but comment on the extension to derivatives of higher order at the end of the proof, see also Proposition 6.9 in [44] and Satz 4.22 in [122].
Similar as before, for every $m \in L^\infty(\Omega)^n$ there exists a $C > 0$ such that $\|m(x)\|_\infty \leq C$ for almost all $x \in \Omega$ and we obtain for every $h \in L^\infty(\Omega)^n$

$$\begin{aligned}\|\varphi'(m(.))h(.)\|_{L^\infty(\Omega)^{\hat{n}}} &= \|(\varphi'(0) + \varphi'(m(.)) - \varphi'(0))h(.)\|_{L^\infty(\Omega)^{\hat{n}}} \\ &\leq \|\varphi'(0)h(.)\|_{L^\infty(\Omega)^{\hat{n}}} + \|(\varphi'(m(.)) - \varphi'(0))h(.)\|_{L^\infty(\Omega)^{\hat{n}}} \\ &\leq (C + L_1\|m\|_{L^\infty(\Omega)^n})\|h\|_{L^\infty(\Omega)^n},\end{aligned}$$

where L_1 denotes the Lipschitz constant of φ'. Hence, $G(m) : L^\infty(\Omega)^n \to L^\infty(\Omega)^{\hat{n}}$, $h \mapsto \varphi'(m(.))h$ satisfies $G(m) \in \mathcal{L}(L^\infty(\Omega)^n, L^\infty(\Omega)^{\hat{n}})$.
Furthermore, $G : L^\infty(\Omega)^n \to \mathcal{L}(L^\infty(\Omega)^n, L^\infty(\Omega)^{\hat{n}})$ is Lipschitz continuous, because we obtain for arbitrary $m_1, m_2 \in L^\infty(\Omega)^n$

$$\begin{aligned}\|G(m_1) - G(m_2)\|_{\mathcal{L}(L^\infty(\Omega)^n, L^\infty(\Omega)^{\hat{n}})} &= \sup_{\|h\|_{L^\infty(\Omega)^n}=1} \|(G(m_1) - G(m_2))h\|_{L^\infty(\Omega)^{\hat{n}}} \\ &= \sup_{\|h\|_{L^\infty(\Omega)^n}=1} \|(\varphi'(m_1(.)) - \varphi'(m_2(.)))h(.)\|_{L^\infty(\Omega)^{\hat{n}}} \\ &\leq L_1\|m_1 - m_2\|_{L^\infty(\Omega)^n}.\end{aligned}$$

It remains to be shown that $G(m)$ is the Fréchet derivative of Φ. We define

$$r(m, h) := \Phi(m + h) - \Phi(m) - G(m)h.$$

The remainder term can be expressed by

$$\begin{aligned}r(m, h)(x) &= \varphi(m(x) + h(x)) - \varphi(m(x)) - \varphi'(m(x))h(x) \\ &= \int_0^1 \left(\varphi'(m(x) + \tau h(x)) - \varphi'(m(x))\right) h(x)\, d\tau.\end{aligned} \tag{2.33}$$

Hence,

$$\begin{aligned}\|r(m, h)(x)\|_\infty &\leq \int_0^1 \|(\varphi'(m(x) + \tau h(x)) - \varphi'(m(x)))h(x)\|_\infty\, d\tau \\ &\leq \int_0^1 \tau L_1\|h(x)\|_\infty^2\, d\tau \leq \frac{1}{2}L_1\|h\|_{L^\infty(\Omega)^n}^2.\end{aligned} \tag{2.34}$$

Thus, we have $\|r(m, h)\|_{L^\infty(\Omega)^{\hat{n}}} = o\left(\|h\|_{L^\infty(\Omega)^n}\right)$ for $\|h\|_{L^\infty(\Omega)^n} \to 0$ and Φ is Lipschitz continuously Fréchet differentiable with derivative $\Phi'(m) = \varphi'(m(.))$.
Now consider $k \geq 2$ and let (2.32) be true for $k - 1$. In the same way as before, we obtain that $\varphi^{(k)}(m(.))$ is bounded and that $m \mapsto \varphi^{(k)}(m(.))$ is Lipschitz continuous due to the Lipschitz continuity of the derivative $\varphi^{(k)}$. Next, we consider the bounded multilinear form $G^{(k)}(m) : L^\infty(\Omega)^n \times \ldots \times L^\infty(\Omega)^n \to L^\infty(\Omega)^{\hat{n}}$, $(h_1, \ldots, h_k) \mapsto \varphi^k(m(.))(h_1, \ldots, h_k)$. Setting

$$r^k(m, h_1, \ldots, h_k) := \left(\Phi^{k-1}(m + h_k) - \Phi^{k-1}(m)\right)(h_1, \ldots, h_{k-1}) - G^{(k)}(m)(h_1, \ldots, h_k)$$

and repeating (2.33) and (2.34) yields the required estimate for the remainder term of the Fréchet derivative. □

Remark 2.3.4.
From the proofs of Lemma 2.3.2 and Lemma 2.3.3 we easily observe that similar results can be obtained for locally Lipschitz continuous φ and Φ.

We will now consider a nonempty, open and convex set $\mathcal{D} \subset L^\infty(\Omega)^n$ and a superposition operator $\Phi : \mathcal{D} \to \mathcal{S} \subset L^\infty(\Omega)^{\hat{n}}$ that maps the parameters, which we seek to determine in the inverse problem, to coefficients, which enter the hyperbolic equation. To this end, we consider a family of continuous bilinear forms $a(\Phi(m)) : V \times V \to \mathbb{R}$ and, similar as before, we identify $a(\Phi(m))$ with the operator $A(\Phi(m)) \in \mathcal{L}(V, V^*)$ by

$$a(\Phi(m))(v, w) = \langle A(\Phi(m))v, w\rangle_{V^*,V} \quad \forall\ v, w \in V.$$

Now, we consider equations

$$\rho u_{tt}(t) + A(\Phi(m))u(t) = f(t),$$

with initial data

$$u(0) = 0, \quad u_t(0) = 0,$$

and compactly write them with the help of the operator $E(u, m) := \rho u_{tt} + A(\Phi(m))u$ as:

$$E(u, m) = f, \quad u(0) = 0, \quad u_t(0) = 0. \tag{2.35}$$

The variational form of (2.35) reads as follows: For all $v \in V$ and a.e. in I:

$$\langle \rho u_{tt}(t), v\rangle_{V^*,V} + a(\Phi(m))(u(t), v) = \langle f(t), v\rangle_{V^*,V}.$$

Now, we have to adjust Assumption 2.2.1 to the modified setting and continue to work with the following assumptions on Φ, A and ρ.

Assumption 2.3.5.

(A1) $\mathcal{D} \subset L^\infty(\Omega)^n$ *is a nonempty, open and convex set and* $\Phi : L^\infty(\Omega)^n \to L^\infty(\Omega)^{\hat{n}}$ *is a superposition operator such that there exists an open and bounded neighborhood* $\mathcal{S} \subset L^\infty(\Omega)^{\hat{n}}$ *of* $\Phi(\mathcal{D})$*. Furthermore, we assume that* $A : L^\infty(\Omega)^{\hat{n}} \to \mathcal{L}(V, V^*)$ *is a bounded linear operator.*

(A2) Assumption 2.2.1 (A2) - (A5) holds for the neighborhood $\mathcal{S}$*.*

Remark 2.3.6.
Note that the major difference between Assumption 2.3.5 and Assumption 2.2.1 is the fact that $\mathcal{S}$ is required to be bounded while this is not the case for $\mathcal{D}$ anymore. Hence, the superposition operator Φ offers a handy alternative to ensure the well-posedness of the governing equation for the whole space $L^\infty(\Omega)^n$ by mapping all parameters to physically reasonable coefficients. Here, we do not necessarily have to require that $\mathcal{D}$ is convex. However, since $\mathcal{D}$ will be associated with the feasible region of the parameter identification problem in the next chapter, this is a suitable assumption. On the other hand, if $\mathcal{D}$ is convex, we can establish the Lipschitz continuity of the derivatives with some convenient assumptions on Φ, see Theorem 2.3.9.

Corollary 2.3.7.
Let Assumption 2.3.5 hold. Then for every $f \in F_0^k$ with $k \geq 1$ and fixed $m \in \mathcal{D}$ there exists a unique solution u to the problem (2.35)*. The mapping*

$$F_0^k \to C^{k-1}(\bar{I}; V) \cap C^k(\bar{I}; H) \cap C^{k+1}(\bar{I}; V^*), \quad f \mapsto u$$

is linear and bounded and for fixed $f \in F_0^k$, the solutions $u(m)$ are uniformly bounded in $C^{k-1}(\bar{I}; V) \cap C^k(\bar{I}; H)$ for all $m \in \mathcal{D}$.

Proof. This is a direct consequence of Corollary 2.2.5 and Assumption 2.3.5. □

In the next step, we seek to establish the continuity and Fréchet differentiability of the state u with respect to the parameters m instead of the coefficients σ. By imposing suitable assumptions on the continuity and Fréchet differentiability of the superposition operator Φ, this is a straightforward extension utilizing the results and techniques from the previous section and applying the chain rule.

Theorem 2.3.8.
Let Assumption 2.3.5 hold and let Φ be Lipschitz continuous. Then, for all $f \in F_0^{k+1}$, $k \geq 1$, the mapping

$$\mathcal{D} \to C^{k-1}(\bar{I}; V) \cap C^k(\bar{I}; H), \quad m \mapsto u(m)$$

is Lipschitz continuous, where $u(m)$ denotes the unique solution to (2.35)*.*

Proof. The existence of a unique solution $u(m)$ is guaranteed even in the stronger space $C^k(\bar{I}; V) \cap C^{k+1}(\bar{I}; H)$ by Corollary 2.3.7. Consider $m_1, m_2 \in \mathcal{D}$ and let $\sigma_1 := \Phi(m_1)$ and $\sigma_2 := \Phi(m_2)$. We denote the solutions to (2.8) for σ_1, σ_2 by $\hat{u}(\sigma_1)$, $\hat{u}(\sigma_2)$ and obtain by Theorem 2.2.7

$$\begin{aligned}
&\|u(m_1) - u(m_2)\|_{C^{k-1}(\bar{I};V)} + \|u_t(m_1) - u_t(m_2)\|_{C^{k-1}(\bar{I};H)} \\
&\quad = \|\hat{u}(\sigma_1) - \hat{u}(\sigma_2)\|_{C^{k-1}(\bar{I};V)} + \|\hat{u}_t(\sigma_1) - \hat{u}_t(\sigma_2)\|_{C^{k-1}(\bar{I};H)} \\
&\quad \leq L_\sigma \|\sigma_1 - \sigma_2\|_{L^\infty(\Omega)^{\hat{n}}} = L_\sigma \|\Phi(m_1) - \Phi(m_2)\|_{L^\infty(\Omega)^{\hat{n}}} \\
&\quad \leq L_\sigma L_\Phi \|m_1 - m_2\|_{L^\infty(\Omega)^n},
\end{aligned}$$

where L_σ and L_Φ denote the Lipschitz constants from Theorem 2.2.7 and Φ, respectively. □

We can now apply the chain rule to derive the differentiability of the parameter-to-solution operator.

Theorem 2.3.9.
Let Assumption 2.3.5 hold and let Φ be Lipschitz continuous differentiable with Φ' uniformly bounded on $\mathcal{D}$, i.e., there exists a $K > 0$ such that

$$\|\Phi'(m)\|_{\mathcal{L}(L^\infty(\Omega)^n, L^\infty(\Omega)^{\hat{n}})} \leq K \quad \forall\, m \in \mathcal{D}.$$

Then, for all $f \in F_0^{k+2}$, $k \geq 1$, the parameter-to-solution operator of (2.35)

$$\mathcal{D} \to C^{k-1}(\bar{I}; V) \cap C^k(\bar{I}; H), \quad m \mapsto u(m)$$

is Lipschitz continuously differentiable.

Proof. Again, we denote the solution operator of (2.8) by

$$\hat{u} : L^\infty(\Omega)^{\hat{n}} \supset \mathcal{S} \to C^{k-1}(\bar{I}; V) \cap C^k(\bar{I}; H), \quad \sigma \mapsto \hat{u}(\sigma),$$

which is Lipschitz continuously differentiable by Theorem 2.2.9. Now, we write $u(m) = \hat{u}(\Phi(m))$ and applying the chain rule (cf. Theorem 2.2.1 in [31]) yields that $m \mapsto u(m)$ is continuously Fréchet differentiable as a map from $\mathcal{D}$ to $\mathcal{U} := C^{k-1}(\bar{I}; V) \cap C^k(\bar{I}; H)$ with derivative given by

$$u'(m) = \hat{u}'(\Phi(m))\Phi'(m).$$

Moreover, for arbitrary $m_1, m_2 \in \mathcal{D}$, we have

$$\begin{aligned} &\|u'(m_1) - u'(m_2)\|_{\mathcal{L}(L^\infty(\Omega)^n, \mathcal{U})} \\ &= \|\hat{u}'(\Phi(m_1))\Phi'(m_1) - \hat{u}'(\Phi(m_2))\Phi'(m_2)\|_{\mathcal{L}(L^\infty(\Omega)^n, \mathcal{U})} \\ &\leq \|\hat{u}'(\Phi(m_1))(\Phi'(m_1) - \Phi'(m_2)) + (\hat{u}'(\Phi(m_1)) - \hat{u}'(\Phi(m_2)))\Phi'(m_2)\|_{\mathcal{L}(L^\infty(\Omega)^n, \mathcal{U})}. \end{aligned} \tag{2.36}$$

The Lipschitz continuity of $\hat{u}'$ (cf. Corollary 2.2.14) together with the boundedness of $\mathcal{S}$ implies that $\hat{u}'(\sigma) \in \mathcal{L}(L^\infty(\Omega)^{\hat{n}}, \mathcal{U})$ is uniformly bounded on $\mathcal{S}$. Hence, together with the Lipschitz continuity of Φ', we obtain

$$\|\hat{u}'(\Phi(m_1))(\Phi'(m_1) - \Phi'(m_2))\|_{\mathcal{L}(L^\infty(\Omega)^n, \mathcal{U})} \leq C\|m_1 - m_2\|_{L^\infty(\Omega)^n}. \tag{2.37}$$

Furthermore, the boundedness of Φ' implies the Lipschitz continuity of Φ, cf. Theorem 3.3.2 in [31]. Thus, by the Lipschitz continuity of $\hat{u}'$, we deduce

$$\begin{aligned} \|(\hat{u}'(\Phi(m_1)) - \hat{u}'(\Phi(m_2)))\Phi'(m_2)\|_{\mathcal{L}(L^\infty(\Omega)^n, \mathcal{U})} &\leq K\, L_{\hat{u}'}\|\Phi(m_1) - \Phi(m_2)\|_{L^\infty(\Omega)^{\hat{n}}} \\ &\leq K\, L_{\hat{u}'}\, L_\Phi \|m_1 - m_2\|_{L^\infty(\Omega)^n}, \end{aligned} \tag{2.38}$$

where $L_{\hat{u}'}$ and L_Φ are the Lipschitz constants of $\hat{u}'$ and Φ. Hence, we conclude from (2.36), (2.37) and (2.38)

$$\|u'(m_1) - u'(m_2)\|_{\mathcal{L}(L^\infty(\Omega)^n, \mathcal{U})} \leq C\|m_1 - m_2\|_{L^\infty(\Omega)^n}.$$

□

The analysis can be extended to higher-order derivatives by applying the chain rule several times. As an alternative, we provide a direct proof of the differentiability of the parameter-to-solution operator $m \mapsto u(m)$ and derive a representation of the k-th Fréchet derivative as a solution to (2.35) with a different right-hand side. The technique is very similar to the analysis in the preceding section. However, due to the nonlinearity of Φ the recursive formula is more involved. To this end, we slightly change the notation. Let $\mathbf{s} = (s_1, \dots, s_i)$ be a tuple of length i with $s_j \in L^\infty(\Omega)^n$, $j = 1, \dots, i$. For any subset $\mathcal{I} \subseteq \{1, \dots, i\}$, we denote by $\mathbf{s}_\mathcal{I}$ the tuple of length $|\mathcal{I}|$ that contains only the s_j with $j \in \mathcal{I}$. Furthermore, the complement of $\mathcal{I}$ in $\{1, \dots, i\}$ is denoted by $\mathcal{I}^c$. Hence, $\mathbf{s}_{\mathcal{I}^c}$ is a tuple of length $i - |\mathcal{I}|$ with elements s_j and $j \in \{1, \dots, i\} \setminus \mathcal{I}$. We also allow for empty tuples if $|\mathcal{I}| = 0$ or $|\mathcal{I}^c| = 0$, respectively. Moreover, for $\tau \in L^\infty(\Omega)^n$, let $(\mathbf{s}, \tau) := (s_1, \dots, s_i, \tau)$ denote the $(i+1)$-tuple of $\mathbf{s}$ concatenated with τ. Lastly, the remainder term of the i-th Fréchet derivative of Φ is denoted by

$$R^i(m, s) := \Phi^{i-1}(m + s) - \Phi^{i-1}(m) - \Phi^i(m)s.$$

In what follows, we will denote the i-th derivative of the solution operator $m \mapsto u(m)$ by $u^{(i)}(m)$. Applying this operator to a tupel $\mathbf{s}$ is denoted by $u^{(i)}(m)[\mathbf{s}]$. In order to include the case $i = 0$ and to shorten the notation, we use a tuple of length zero and write $u^{(0)}(m)[\emptyset] := u^{(0)}(m) = u(m)$.

Theorem 2.3.10.
Let Assumption 2.3.5 hold and consider $k \geq 1$ and $f \in F_0^{k+2}$. Furthermore, let Φ be k-times Lipschitz continuously Fréchet differentiable with $\Phi^{(i)}(m) \in \mathcal{L}^i(\mathcal{D}; L^\infty(\Omega)^{\hat{n}})$, $i = 1, \dots, k$, uniformly bounded on $\mathcal{D}$. Then the solution operator $\mathcal{D} \to C(\bar{I}; V) \cap C^1(\bar{I}; H)$, $m \mapsto u(m)$ is k-times Lipschitz continuously Fréchet differentiable.
Furthermore, let $\mathbf{s} = (s_1, \dots, s_i)$ with $s_j \in L^\infty(\Omega)^n$ and $m + s_j \in \mathcal{D}$, $j = 1, \dots, i$. Then the i-th derivative applied to $\mathbf{s}$ is recursively defined by $d^i = u^{(i)}(m)[\mathbf{s}]$ as the unique solution to

$$E(d^i, m) = -\sum_{j=1}^{i} \sum_{\substack{\mathcal{I} \subseteq \{1,\dots,i\} \\ |\mathcal{I}| = j}} A(\Phi^{(j)}(m)[\mathbf{s}_{\mathcal{I}}])\, u^{(i-j)}(m)[\mathbf{s}_{\mathcal{I}^c}], \quad d^i(0) \;\; = 0, \quad d_t^i(0) = 0. \tag{2.39}$$

Proof. The proof is divided into three parts.
First, we show that $d^i \in C^{k+1-i}(\bar{I}; V) \cap C^{k+2-i}(\bar{I}; H)$ is uniquely determined by (2.39). In a second step, we show that $m \mapsto u^i(m)$ is Lipschitz continuous from $\mathcal{D}$ to $\mathcal{L}^i(\mathcal{D}; C^{k-i}(\bar{I}; V) \cap C^{k-i+1}(\bar{I}; H))$. Finally, we show that d^i is the Fréchet derivative by providing an estimate for the remainder term. Since the technique is very similar to the previous section, we slightly shorten the presentation.
Part I.
Since the existence of higher-order derivatives follows only from the complete proof, we replace $u^{(i-j)}$ in (2.39) by d^{i-j} and show by induction that $d^i \in C^{k+1-i}(\bar{I}; V) \cap C^{k+2-i}(\bar{I}; H)$ is uniquely determined by

$$E(d^i, m) = -\sum_{j=1}^{i} \sum_{\substack{\mathcal{I} \subseteq \{1,\dots,i\} \\ |\mathcal{I}| = j}} A(\Phi^{(j)}(m)[\mathbf{s}_{\mathcal{I}}])\, d^{i-j}(m)[\mathbf{s}_{\mathcal{I}^c}], \quad d^i(0) \;\; = 0, \quad d_t^i(0) = 0. \tag{2.40}$$

and

$$\left\|d^i\right\|_{C^{k+1-i}(\bar{I};V)} + \left\|d_t^i\right\|_{C^{k+1-i}(\bar{I};H)} \leq C \prod_{j=1}^{i} \|s_j\|_{L^\infty(\Omega)^n}, \quad 1 \leq i \leq k, \tag{2.41}$$

with a constant $C > 0$ that can be chosen uniformly on $\mathcal{D}$. Moreover, we recursively obtain that d^i is zero near $t = 0$ and depends linearly on every s_j, $j = 1, \dots, i$, by Corollary 2.3.7.
For $i = 1$ and $\mathbf{s} = (s_1)$, (2.40) reads as follows:

$$\rho\, d_{tt}^1 + A(\Phi(m))\, d^1 = -A(\Phi'(m)s_1)\, u(m), \quad d^1(0) = 0, \quad d_t^1(0) = 0.$$

Again by Corollary 2.3.7, we deduce $u(m) \in C^{k+1}(\bar{I}; V)$ and $A(\Phi'(m)s_1)\, u(m) \in F_0^{k+1}$. Furthermore, with Assumption 2.3.5 and the uniform boundedness of Φ' we deduce

$$\begin{aligned} \|A(\Phi'(m)s_1)\, u(m)\|_{C^{k+1}(\bar{I};V^*)} &\leq \|A(\Phi'(m)s_1)\|_{\mathcal{L}(V,V^*)} \|u(m)\|_{C^{k+1}(\bar{I};V)} \\ &\leq C \|\Phi'(m)s_1)\|_{L^\infty(\Omega)^{\hat{n}}} \|u(m)\|_{C^{k+1}(\bar{I};V)} \\ &\leq C \|s_1\|_{L^\infty(\Omega)^n}, \end{aligned}$$

where the constant $C > 0$ can be chosen uniformly on $\mathcal{D}$. Hence, by Corollary 2.3.7 there exists a unique solution d^1 to (2.40) which satisfies the estimate

$$\left\|d^1\right\|_{C^k(\bar{I};V)} + \left\|d^1_t\right\|_{C^k(\bar{I};H)} \leq C \left\|s_1\right\|_{L^\infty(\Omega)^n}.$$

Moreover, $s \in L^\infty(\Omega)^n \mapsto d^1 \in C^k(\bar{I};V) \cap C^{k+1}(\bar{I};H)$ is linear and d^1 is zero near $t = 0$. Assume now that the statement is true for $i-1$ with $1 < i \leq k$. By the induction hypothesis, $d^{i-j} \in C^{k+1-(i-j)}(\bar{I};V)$, $j = 1, \dots, i$. I.e., the right-hand side in (2.40) is in F_0^{k+2-i}. By the boundedness of $\Phi^{(j)}$ and the induction hypothesis for (2.41), we obtain for any $\mathcal{I} \subseteq \{1, \dots, i\}$

$$\begin{aligned}
&\|A(\Phi^{(j)}(m)[\mathbf{s}_\mathcal{I}])\, d^{i-j}(m)[\mathbf{s}_{\mathcal{I}^c}]\|_{C^{k+2-i}(\bar{I};V^*)} \\
&\qquad \leq C\|\Phi^{(j)}(m)[\mathbf{s}_\mathcal{I}]\|_{L^\infty(\Omega)^{\hat{n}}}\|d^{i-j}(m)[\mathbf{s}_{\mathcal{I}^c}]\|_{C^{k+2-i}(\bar{I};V)} \\
&\qquad \leq C \prod_{j\in\mathcal{I}} \|s_j\|_{L^\infty(\Omega)^n} \prod_{j\in\mathcal{I}^c} \|s_j\|_{L^\infty(\Omega)^n} = C \prod_{j=1}^{i} \|s_j\|_{L^\infty(\Omega)^n}.
\end{aligned} \tag{2.42}$$

Thus, Corollary 2.3.7 yields again the existence of a unique solution $d^i \in C^{k+1-i}(\bar{I};V) \cap C^{k+2-i}(\bar{I};H)$ to (2.39) and from (2.42) we deduce

$$\left\|d^i\right\|_{C^{k+1-i}(\bar{I};V)} + \left\|d^i_t\right\|_{C^{k+1-i}(\bar{I};H)} \leq C \prod_{j=1}^{i} \|s_j\|_{L^\infty(\Omega)^n},$$

which is (2.41) for d^i.

Part II.

Now, we consider $h^i(m,s) := d^i(m+s) - d^i(m)$ for $s \in L^\infty(\Omega)^n$ with $m + s \in \mathcal{D}$. With the previously used notation, we have $h^0(m,s)[\emptyset] = u(m+s) - u(m)$. For $\mathbf{s} = (s_1, \dots, s_i)$ with $s_j \in L^\infty(\Omega)^n$, $m + s_j \in \mathcal{D}$, we insert $d^i(m+s)$ and, resp., $d^i(m)$ into (2.40), take the difference and obtain that $h^i = h^i(m,s)[\mathbf{s}]$ satisfies $h^i(0) = 0$, $h^i_t(0) = 0$ and

$$\begin{aligned}
&E(h^i, m) \\
&= E(d^i(m+s)[\mathbf{s}], m+s) - E(d^i(m)[\mathbf{s}], m) - A(\Phi(m+s) - \Phi(m))d^i(m+s)[\mathbf{s}] \\
&= -\sum_{j=1}^{i} \sum_{\substack{\mathcal{I}\subseteq\{1,\dots,i\} \\ |\mathcal{I}|=j}} A(\Phi^{(j)}(m+s)[\mathbf{s}_\mathcal{I}])\, d^{i-j}(m+s)[\mathbf{s}_{\mathcal{I}^c}] - A(\Phi^{(j)}(m)[\mathbf{s}_\mathcal{I}])\, d^{i-j}(m)[\mathbf{s}_{\mathcal{I}^c}] \\
&\quad - A(\Phi(m+s) - \Phi(m))d^i(m+s)[\mathbf{s}] \\
&= -\sum_{j=0}^{i} \sum_{\substack{\mathcal{I}\subseteq\{1,\dots,i\} \\ |\mathcal{I}|=j}} A((\Phi^{(j)}(m+s) - \Phi^{(j)}(m))[\mathbf{s}_\mathcal{I}])d^{i-j}(m+s)[\mathbf{s}_{\mathcal{I}^c}] \\
&\quad - \sum_{j=1}^{i} \sum_{\substack{\mathcal{I}\subseteq\{1,\dots,i\} \\ |\mathcal{I}|=j}} A(\Phi^{(j)}(m)[\mathbf{s}_\mathcal{I}])h^{i-j}(m,s)[\mathbf{s}_{\mathcal{I}^c}].
\end{aligned} \tag{2.43}$$

Next, we show by induction that h^i is uniquely determined by (2.43) with homogeneous initial conditions and satisfies

$$\left\|h^i\right\|_{C^{k-i}(\bar{I};V)} + \left\|h^i_t\right\|_{C^{k-i}(\bar{I};H)} \leq C \left\|s\right\|_{L^\infty(\Omega)^n} \prod_{j=1}^{i} \|s_j\|_{L^\infty(\Omega)^n}, \quad 1 \leq i \leq k, \tag{2.44}$$

where $C > 0$ can be chosen uniformly on $\mathcal{D}$.

For $i = 1$ and $\mathbf{s} = (s_1)$, (2.43) reads as

$$\begin{aligned} E(h^1, m) = &- A((\Phi(m+s) - \Phi(m)))d^1(m+s)[s_1] \\ &- A((\Phi'(m+s) - \Phi'(m))[s_1])u(m+s) \\ &- A(\Phi'(m)[s_1])(u(m+s) - u(m)). \end{aligned} \tag{2.45}$$

We note that every term on the right-hand side of (2.45) is bounded in the $C^k(\bar{I}; V^*)$-norm by $C\|s\|_{L^\infty(\Omega)^n}\|s_1\|_{L^\infty(\Omega)^n}$ with a constant C independent of m and s. Here, we have to employ (i) the Lipschitz continuity of Φ and (2.41), (ii) the Lipschitz continuity of Φ' and the boundedness of u and (iii) the boundedness of Φ' and Theorem 2.3.8. Hence, the right-hand side in (2.45) is in F_0^k and Corollary 2.3.7 gives (2.44) for $i = 1$. Furthermore, since $d^1(m+s)[s_1]$, $u(m+s)$ and $u(m)$ are zero near $t = 0$, the same holds true for h^1.

Assume now that the statement is true for some i with $1 \le i < k$. From (2.41) and using the Lipschitz continuity of $\Phi^{(j)}$ we deduce for any $\mathcal{I} \subseteq \{1, \dots, i\}$, $j \in \{0, \dots, i\}$ and $\iota = k + 1 - i$

$$\|A((\Phi^{(j)}(m+s) - \Phi^{(j)}(m))[\mathbf{s}_\mathcal{I}])\, d^{i-j}(m)[\mathbf{s}_{\mathcal{I}^c}]\|_{C^{\iota+j}(\bar{I};V^*)} \le C\|s\|_{L^\infty(\Omega)^n} \prod_{j=1}^{i} \|s_j\|_{L^\infty(\Omega)^n}. \tag{2.46}$$

By the induction hypothesis and the boundedness of $\Phi^{(j)}(m)$, we also obtain for any $\mathcal{I} \subseteq \{1, \dots, i\}$ and $j \in \{1, \dots, i\}$

$$\|A(\Phi^{(j)}(m)[\mathbf{s}_\mathcal{I}])h^{i-j}(m, s)[\mathbf{s}_{\mathcal{I}^c}]\|_{C^{\iota+j-1}(\bar{I};V^*)} \le C\|s\|_{L^\infty(\Omega)^n} \prod_{j=1}^{i} \|s_j\|_{L^\infty(\Omega)^n}. \tag{2.47}$$

Hence, every term of the right-hand side in (2.43) is in $C^\iota(\bar{I}; V^*) \subset H^{k+1-i}(I; V^*)$ and, together with the initial homogeneous conditions, Corollary 2.3.7 gives the existence of a unique solution $h^i \in C^{k-i}(\bar{I}; V) \cap C^{k-i+1}(\bar{I}; H)$ to (2.43). Furthermore, we deduce the estimate (2.44) from (2.46) and (2.47). In particular, $m \mapsto d^i(m)$ is Lipschitz continuous from $\mathcal{D}$ to $\mathcal{L}^i(\mathcal{D}; C^{k-i}(\bar{I}; V) \cap C^{k-i+1}(\bar{I}; H))$.

Part III.
Finally, we consider $r^i(m, s) := h^{i-1}(m, s) - d^i(m)s$ for $s \in L^\infty(\Omega)^n$ with $m + s \in \mathcal{D}$. Hence, r^i is the remainder term of the i-th derivative. For $\mathbf{s} = (s_1, \dots, s_{i-1})$ with $s_j \in L^\infty(\Omega)^n$, $m + s_j \in \mathcal{D}$, we have $(d^i(m)s)[\mathbf{s}] = d^i(m)[(\mathbf{s}, s)]$. Thus, inserting $d^i(m)[(\mathbf{s}, s)]$ into (2.40) and

rearranging the summations with regard to whether $s \in \mathcal{I}$ or $s \in \mathcal{I}^c = \{1, \dots, i-1\} \setminus \mathcal{I}$ yields

$$\begin{aligned} E(d^i(m)[(\mathbf{s}, s)], m) = &- \sum_{j=1}^{i-1} \sum_{\substack{\mathcal{I} \subseteq \{1,\dots,i-1\} \\ |\mathcal{I}|=j}} A(\Phi^{(j)}(m)[\mathbf{s}_{\mathcal{I}}])\, d^{(i-j)}(m)[(\mathbf{s}_{\mathcal{I}^c}, s)] \\ &- \sum_{j=1}^{i} \sum_{\substack{\mathcal{I} \subseteq \{1,\dots,i-1\} \\ |\mathcal{I}|=j-1}} A(\Phi^{(j)}(m)[(\mathbf{s}_{\mathcal{I}}, s)])\, d^{(i-j)}(m)[\mathbf{s}_{\mathcal{I}^c}] \\ = &- \sum_{j=1}^{i-1} \sum_{\substack{\mathcal{I} \subseteq \{1,\dots,i-1\} \\ |\mathcal{I}|=j}} A(\Phi^{(j)}(m)[\mathbf{s}_{\mathcal{I}}])\, d^{(i-j)}(m)[(\mathbf{s}_{\mathcal{I}^c}, s)] \\ &- \sum_{j=0}^{i-1} \sum_{\substack{\mathcal{I} \subseteq \{1,\dots,i-1\} \\ |\mathcal{I}|=j}} A(\Phi^{(j+1)}(m)[(\mathbf{s}_{\mathcal{I}}, s)])\, d^{(i-1-j)}(m)[\mathbf{s}_{\mathcal{I}^c}]. \end{aligned} \tag{2.48}$$

Now, we insert $h^{i-1}(m, s)$ into (2.43), subtract (2.48) and obtain after some rearrangements that $r^i = r^i(m, s)[\mathbf{s}]$ satisfies $r^i(0) = 0$, $r^i_t(0) = 0$ and

$$\begin{aligned} E(r^i, m) = &E(h^{i-1}(m, s)[\mathbf{s}], m) - E(d^i(m)[(\mathbf{s}, s)], m) \\ = &- \sum_{j=0}^{i-1} \sum_{\substack{\mathcal{I} \subseteq \{1,\dots,i-1\} \\ |\mathcal{I}|=j}} A((\Phi^{(j)}(m+s) - \Phi^{(j)}(m))[\mathbf{s}_{\mathcal{I}}]) d^{i-1-j}(m+s)[\mathbf{s}_{\mathcal{I}^c}] \\ &- \sum_{j=1}^{i-1} \sum_{\substack{\mathcal{I} \subseteq \{1,\dots,i-1\} \\ |\mathcal{I}|=j}} A(\Phi^{(j)}(m)[\mathbf{s}_{\mathcal{I}}]) h^{i-1-j}(m, s)[\mathbf{s}_{\mathcal{I}^c}] \\ &- \sum_{j=1}^{i-1} \sum_{\substack{\mathcal{I} \subseteq \{1,\dots,i-1\} \\ |\mathcal{I}|=j}} A(\Phi^{(j)}(m)[\mathbf{s}_{\mathcal{I}}])\, d^{(i-j)}(m)[(\mathbf{s}_{\mathcal{I}^c}, s)] \\ &- \sum_{j=0}^{i-1} \sum_{\substack{\mathcal{I} \subseteq \{1,\dots,i-1\} \\ |\mathcal{I}|=j}} A(\Phi^{(j+1)}(m)[(\mathbf{s}_{\mathcal{I}}, s)])\, d^{(i-1-j)}(m)[\mathbf{s}_{\mathcal{I}^c}] \\ = &- \sum_{j=1}^{i} \sum_{\substack{\mathcal{I} \subseteq \{1,\dots,i-1\} \\ |\mathcal{I}|=j-1}} A(R^j(m, s)[\mathbf{s}_{\mathcal{I}}])\, d^{(i-j)}(m)[\mathbf{s}_{\mathcal{I}^c}] \\ &- \sum_{j=0}^{i-1} \sum_{\substack{\mathcal{I} \subseteq \{1,\dots,i-1\} \\ |\mathcal{I}|=j}} A((\Phi^{(j)}(m+s) - \Phi^{(j)}(m))[\mathbf{s}_{\mathcal{I}}]) h^{i-1-j}(m+s)[\mathbf{s}_{\mathcal{I}^c}] \\ &- \sum_{j=1}^{i-1} \sum_{\substack{\mathcal{I} \subseteq \{1,\dots,i-1\} \\ |\mathcal{I}|=j}} A(\Phi^{(j)}(m)[\mathbf{s}_{\mathcal{I}}]) r^{i-j}(m, s)[\mathbf{s}_{\mathcal{I}^c}]. \end{aligned} \tag{2.49}$$

It remains to be shown that r^i is uniquely determined by (2.49) with homogeneous initial conditions and that it satisfies for $1 \leq i \leq k$

$$\left\|r^i\right\|_{C^{k-i}(\bar{I};V)} + \left\|r^i_t\right\|_{C^{k-i}(\bar{I};H)} = o\left(\|s\|_{L^\infty(\Omega)^n}\right) \prod_{j=1}^{i-1} \|s_j\|_{L^\infty(\Omega)^n} \quad \text{for} \quad \|s\|_{L^\infty(\Omega)^n} \to 0. \tag{2.50}$$

For $i = 1$ we have $r^1(m,s) = h^0(m,s) - d^1(m)s = u(m+s) - u(m) - d^1(m)s$ and (2.49) simplifies to

$$-A(R^1(m,s))\,u(m) - A((\Phi(m+s) - \Phi(m))(u(m+s) - u(m))$$

Now, using the Fréchet differentiability of Φ and the boundedness of $u(m) \in C^{k+1}(\bar{I};V)$ on $\mathcal{D}$ gives

$$\|A(R^1(m,s))\,u(m)\|_{C^{k+1}(\bar{I};V^*)} = o\left(\|s\|_{L^\infty(\Omega)^n}\right) \quad \text{for} \quad \|s\|_{L^\infty(\Omega)^n} \to 0. \tag{2.51}$$

Furthermore, using the Lipschitz continuity of both, Φ and the solution operator, we obtain

$$\|A((\Phi(m+s) - \Phi(m))(u(m+s) - u(m))\|_{C^k(\bar{I};V^*)} \leq C\|s\|^2_{L^\infty(\Omega)^n}. \tag{2.52}$$

Hence, combining (2.51) and (2.52) and applying Corollary 2.3.7 yields (2.50) for $i = 1$.
Assume now, that (2.50) is true for some i with $1 \leq i < k$. Let $\mathcal{I} \subseteq \{1,\dots,i-1\}$ and $\iota = k+1-i$. From the Fréchet differentiability of Φ and (2.41), we obtain for $j \in \{1,\dots,i\}$

$$\|A(R^j(m,s)[\mathbf{s}_\mathcal{I}])\,d^{(i-j)}(m)[\mathbf{s}_{\mathcal{I}^c}]\|_{C^{\iota+j}(\bar{I};V^*)} = o\left(\|s\|_{L^\infty(\Omega)^n}\right) \prod_{j=1}^{i-1} \|s_j\|_{L^\infty(\Omega)^n}. \tag{2.53}$$

Furthermore, for $j \in \{0,\dots,i-1\}$, the Lipschitz continuity of $\Phi^{(j)}$ and (2.44) yields

$$\begin{aligned}\|A((\Phi^{(j)}(m+s) - \Phi^{(j)}(m))[\mathbf{s}_\mathcal{I}])h^{i-1-j}(m+s)[\mathbf{s}_{\mathcal{I}^c}]\|_{C^{\iota+j}(\bar{I};V^*)} \\ \leq C\|s\|^2_{L^\infty(\Omega)^n} \prod_{j=1}^{i-1} \|s_j\|_{L^\infty(\Omega)^n}.\end{aligned} \tag{2.54}$$

Lastly, for $j \in \{1,\dots,i-1\}$, we obtain from the boundedness of $\Phi^{(j)}$ and the induction hypothesis

$$\|A(\Phi^{(j)}(m)[\mathbf{s}_\mathcal{I}]))r^{i-j}(m,s)[\mathbf{s}_{\mathcal{I}^c}]\|_{C^{\iota+j-1}(\bar{I};V^*)} = o\left(\|s\|_{L^\infty(\Omega)^n}\right) \prod_{j=1}^{i-1} \|s_j\|_{L^\infty(\Omega)^n}. \tag{2.55}$$

Hence, every term of the right-hand side in (2.49) is in $C^\iota(\bar{I};V^*) \subset H^{k+1-i}(I;V^*)$ and, together with the initial homogeneous conditions, Corollary 2.3.7 gives the existence of a unique solution $r^i \in C^{k-i}(\bar{I};V) \cap C^{k-i+1}(\bar{I};H)$ to (2.49). Furthermore, by combining (2.53), (2.54) and (2.55), we obtain (2.50) which concludes the proof. □

Again, we can also use a higher temporal regularity of the right-hand side to increase the temporal regularity of the solution.

Corollary 2.3.11.
Consider the same Assumptions as in Theorem 2.3.10 and, in addition, let $l \geq 0$ and $f \in F_0^{k+l+2}$. Then, we have:

1. *The solution operator $m \mapsto u(m)$ is k-times Lipschitz continuously Fréchet differentiable as a map from $\mathcal{D}$ to $C^l(\bar{I}; V)$.*
2. *The solution operator $m \mapsto u(m)$ is $(k+1)$-times Lipschitz continuously Fréchet differentiable as a map from $\mathcal{D}$ to $C^l(\bar{I}; H)$.*

The results of this section provide the suitable framework for problems in seismic tomography in both, fluid and solid medium. Therefore, we consider the elastic and the acoustic wave equation in the next step and verify the assumptions of this section. It should be emphasized, though, that the result are of greater generality and apply to linear hyperbolic equations of second order with parameters that obey Assumption 2.3.5.

2.4 Parameterization of the Elastic Wave Equation

Now, we return to the discussion of the elastic wave equation and analyze the existence and regularity of solutions. We set $V = H^1(\Omega)^d$ and $H = L^2(\Omega)^d$. From (1.6) we recall the strong form of the elastic wave equation that is given by

$$\begin{cases} \rho u_{tt} - \nabla \cdot (\Psi : \varepsilon(u)) = f & \text{on } \Omega \times I, \\ u(0) = 0 & \text{on } \Omega, \\ u_t(0) = 0 & \text{on } \Omega, \\ (\Psi : \varepsilon(u)) \cdot \vec{n} = 0 & \text{on } \partial\Omega \times I. \end{cases} \tag{2.56}$$

Now, we have to establish a suitable function space setting for the material parameters. After reducing symmetry relations, the most general case with $\Psi \in L^\infty(\Omega)^{d^4}$ still yields 21 independent components in 3d (cf. section 1.1). However, the ill-posedness of the inverse problem and the small number of observations that are typically available will not allow for an accurate identification of this high number of individual components. Consequently, a simpler model with less unknowns is required. Depending on the parameterization of the governing equations, the unknown material parameters can be the Lamé parameters, the velocity of compressional and shear waves, further elasticity parameters like the bulk modulus or variables that characterize anisotropy. In either case, the unknown parameter field is heterogeneous in space and does not depend on time. The number of unknowns that we consider in a simplified model will be denoted by n. The parameterization of the material can now be described with the help of a superposition operator $\Psi : L^\infty(\Omega)^n \to L^\infty(\Omega)^{\hat{n}}$ with $\hat{n} = d^4$.

Example 2.4.1 (Perfectly elastic isotropic material)**.**
An important special case of the parameterization was already given in (1.9) with perfectly elastic, isotropic material that is characterized by the Lamé coefficients λ and μ. Hence, we could set $n = 2$ and $m = (\lambda, \mu)$ with $\lambda, \mu \in L^\infty(\Omega)$ and define the superposition operator $\Psi : L^\infty(\Omega)^2 \to L^\infty(\Omega)^{d^4}$ by

$$\Psi_{ijkl}(m) = \lambda \delta_{ij}\delta_{kl} + \mu(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}).$$

Alternatively, we could work with the isotropic elastic wave equation (1.11) and define $\lambda(m)$ and $\mu(m)$ themselves as superposition operators.

Example 2.4.2 (Anisotropy)**.**
Another relevant parameterization includes anisotropy with Cartesian symmetry. Here, in addition to $\lambda = \lambda(m)$ and $\mu = \mu(m)$ there are three more components a_1, a_2, a_3 (for $d = 3$) that characterize anisotropy along the different coordinate axes. According to [7], the Voigt-matrix, is given by

$$\Psi^V = \begin{pmatrix} \lambda+2\mu & \lambda+a_3 & \lambda+a_3 & 0 & 0 & 0 \\ \lambda+a_3 & \lambda+2\mu+a_1 & \lambda+a_1 & 0 & 0 & 0 \\ \lambda+a_3 & \lambda+a_1 & \lambda+2\mu+a_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu+a_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu+a_2 \end{pmatrix}, \tag{2.57}$$

which completely determines Ψ using (1.8). Here, we assume that a_1, a_2 and a_3 do not depend on m. Inverting for anisotropic parameters is beyond the scope of this thesis, but could in principle be incorporated into our framework as well.

Now, we want to apply the results from the previous section to establish the existence of a unique solution to the elastic wave equation as well as the Fréchet differentiability of the parameter-to-solution operator. In order to derive the weak form of the elastic wave equation we consider $\Psi = \Psi(m)$ for a fixed $m \in L^\infty(\Omega)^n$ and apply Gauss' Divergence Theorem. Integration by parts and exploiting the symmetries of Ψ like in (1.7) yield for fixed t (and sufficiently regular test functions v and w)

$$\begin{aligned}
&-\int_\Omega (\nabla\cdot(\Psi(x):\varepsilon(v)(x,t)))^T w(x,t)\,dx + \int_\Gamma (\Psi(x):\varepsilon(v)(x,t))\cdot\vec{n}(x))^T w(x,t)\,dS(x)\\
=\;& \sum_{i,j,k,l}\int_\Omega \left[\Psi_{ijkl}(x)\left[\varepsilon(v)(x,t)\right]_{kl}\right]\frac{\partial}{\partial x_j}w^i(x,t)\,dx\\
=\;& \sum_{i,j,k,l}\int_\Omega \frac{1}{2}\Psi_{ijkl}(x)\left(\frac{\partial}{\partial x_k}v^l(x,t)+\frac{\partial}{\partial x_l}v^k(x,t)\right)\frac{\partial}{\partial x_j}w^i(x,t)\,dx\\
=\;& \sum_{i,j,k,l}\int_\Omega \frac{1}{4}\Psi_{ijkl}(x)\left(\frac{\partial}{\partial x_k}v^l(x,t)+\frac{\partial}{\partial x_l}v^k(x,t)\right)\left(\frac{\partial}{\partial x_i}w^j(x,t)+\frac{\partial}{\partial x_j}w^i(x,t)\right)dx\\
=\;& (\Psi:\varepsilon(v),\varepsilon(w))_{L^2(\Omega)^{d\times d}}.
\end{aligned}$$

Hence, for fixed $m \in L^\infty(\Omega)^n$, we define the operator $A(\Psi(m))$ by

$$\langle A(\Psi(m))v, w\rangle_{V^*,V} = (\Psi(m):\varepsilon(v),\varepsilon(w))_{L^2(\Omega)^{d\times d}} \quad \forall\, v, w \in V. \tag{2.58}$$

Obviously, $A(\Psi(m))$ is bounded and linear in v and w, i.e., $A(\Psi(m)) \in \mathcal{L}(V, V^*)$. In the case of isotropic material (2.58) simplifies to: $\forall\, v, w \in V$:

$$\langle A(\Psi(m))v, w\rangle_{V^*,V} = (\lambda(m)\,\nabla\cdot v, \nabla\cdot w)_{L^2(\Omega)} + 2\,(\mu(m)\,\varepsilon(v),\varepsilon(w))_{L^2(\Omega)^{d\times d}}\,. \tag{2.59}$$

Reusing the notation from the previous section, we compactly write the elastic wave equation in the following way

$$E(u,m) = \rho u_{tt} + A(\Psi(m))u, \quad u(0) = 0, \quad u_t(0) = 0, \tag{2.60}$$

and the variational form of (2.60) reads as follows: For all $v \in V$ and a.e. in I:

$$\langle \rho u_{tt}(t), v\rangle_{V^*,V} + a(\Psi(m))(u(t), v) = \langle f(t), v\rangle_{V^*,V}, \tag{2.61}$$

where $a(\Psi(m)) : V \times V \to \mathbb{R}$ is defined by the identity

$$a(\Psi(m))(v, w) = \langle A(\Psi(m))v, w\rangle_{V^*,V} \quad \forall\, v, w \in V.$$

Now, we have to verify the requirements of Assumption 2.3.5. Due to the symmetry $\Psi_{ijkl} = \Psi_{klij}$, the bilinear form a that is identified with (2.58) (resp. (2.59)) is symmetric. We keep the conditions on the density ρ and modify the first part of Assumption 2.3.5 as follows:

Assumption 2.4.3.

(A1) $\mathcal{D} \subset L^\infty(\Omega)^n$ *is a nonempty, open and convex set and* $\Psi : L^\infty(\Omega)^n \to L^\infty(\Omega)^{\hat{n}}$ *is a superposition operator such that there exists an open and bounded neighborhood* $\mathcal{S} \subset L^\infty(\Omega)^{\hat{n}}$ *of* $\Psi(\mathcal{D})$ *that satisfies for all* $\sigma \in \mathcal{S}$*:*

$$a(\sigma)(v, v) + \beta_1 \|v\|_H^2 \geq \beta_2 \|v\|_V^2 \quad \forall\, v \in V, \tag{2.62}$$

with constants $\beta_1, \beta_2 > 0$ *independent of* σ*.*

(A2) Assumption 2.2.1 (A4) and (A5) hold.

We can give a simpler condition for (2.62) in the isotropic case. To this end, we require Korn's inequality:

Proposition 2.4.4 (Korn's inequality)**.**
Let $\Omega \subset \mathbb{R}^d$*,* $d = 2, 3$*, be an open, bounded set with piecewise smooth boundary. Then there exists a constant* $C > 0$ *such that*

$$(\varepsilon(v), \varepsilon(v))_{L^2(\Omega)^{d\times d}} + \|v\|_{L^2(\Omega)^d}^2 \geq C \|v\|_{H^1(\Omega)^d}^2 \quad \forall\, v \in H^1(\Omega)^d.$$

Proof. See Chapter III, Theorem 3.1 in [45]. □

Lemma 2.4.5.
Let $\mathcal{D} \subset L^\infty(\Omega)^n$ *be a nonempty, open and convex set and let* $\lambda : L^\infty(\Omega) \to L^\infty(\Omega)$*,* $\mu : L^\infty(\Omega) \to L^\infty(\Omega)$ *be superposition operators such that there exist* λ_a*,* λ_b*,* μ_a*,* $\mu_b \in \mathbb{R}$ *with* $0 < \lambda_a < \lambda_b$*,* $0 < \mu_a < \mu_b$ *and* $\lambda_a \leq \lambda(m)(x) \leq \lambda_b$*,* $\mu_a \leq \mu(m)(x) \leq \mu_b$ *for almost all* $x \in \Omega$ *and all* $m \in \mathcal{D}$*. Then Assumption 2.4.3 (A1) holds.*

Proof. By construction there exists an open and bounded neighborhood $\mathcal{S} \subset L^\infty(\Omega)^2$ of $\lambda(\mathcal{D}) \times \mu(\mathcal{D})$ such that for all $(\hat{\lambda}, \hat{\mu}) \in \mathcal{S}$ it holds $\hat{\lambda}(x) \geq \lambda_a/2$, $\hat{\mu}(x) \geq \mu_a/2$ a.e. in Ω. Thus, by Proposition 2.4.4, we obtain for all $(\hat{\lambda}, \hat{\mu}) \in \mathcal{S}$:

$$\begin{aligned} a((\hat{\lambda}, \hat{\mu}))(v, v) &= (\hat{\lambda}\nabla \cdot v, \nabla \cdot v)_{L^2(\Omega)} + (2\hat{\mu}\,\varepsilon(v), \varepsilon(v))_{L^2(\Omega)^{d\times d}} \\ &\geq \frac{1}{2}\lambda_a (\nabla \cdot v, \nabla \cdot v)_{L^2(\Omega)} + \mu_a (\varepsilon(v), \varepsilon(v))_{L^2(\Omega)^{d\times d}} \\ &\geq \mu_a \left(C \|v\|_V^2 - \|v\|_H^2 \right) \qquad\qquad \forall\, v \in V, \end{aligned}$$

where C is the constant from Korn's inequality. □

In order to validate (2.62) in the anisotropic case, further physical properties of the elastic tensor have to be exploited. For a proof we refer to [40].
Using the results from section 2.3, we directly obtain the central result for the elastic wave equation.

Theorem 2.4.6.
Let Assumption 2.4.3 hold and consider $k \geq 1$, $l \geq 0$ and $f \in F_0^{k+l+2}$. Furthermore, let $\Psi : L^\infty(\Omega)^n \to L^\infty(\Omega)^{\hat{n}}$ be k-times Lipschitz continuously differentiable on $\mathcal{D}$ and with uniformly bounded $\Psi^{(k)}$. Then the elastic wave equation (2.60) possesses a unique solution $u(m) \in C^{k+l+1}(\bar{I}; V) \cap C^{k+l+2}(\bar{I}; H)$ for every $m \in \mathcal{D}$ and the solutions $u(m)$ are uniformly bounded in $C^{k+l+1}(\bar{I}; V) \cap C^{k+l+2}(\bar{I}; H)$ for all $m \in \mathcal{D}$. Furthermore:

1. *The solution operator $m \mapsto u(m)$ is k-times Lipschitz continuously Fréchet differentiable as a map from $\mathcal{D}$ to $C^l(\bar{I}; V)$.*
2. *The solution operator $m \mapsto u(m)$ is $(k+1)$-times Lipschitz continuously Fréchet differentiable as a map from $\mathcal{D}$ to $C^l(\bar{I}; H)$.*

Proof. This follows directly from Corollary 2.3.7 and Corollary 2.3.11. □

2.5 Parameterization of the Acoustic Wave Equation

We now turn to the acoustic wave equation that governs the propagation of seismic waves in a fluid medium. Here, the results of section 2.3 can be carried over as well without greater difficulty. Note that this also extends previous results for the one dimensional acoustic wave equation presented in [78, 88].
From (1.16) we recall the strong form of the acoustic wave equation

$$\begin{cases} \rho_F u_{tt} - \rho_F \nabla \cdot (m \nabla u) = f & \text{on } \Omega \times I, \\ u(0) = 0 & \text{on } \Omega, \\ u_t(0) = 0 & \text{on } \Omega, \\ u = 0 & \text{on } \partial\Omega \times I. \end{cases} \tag{2.63}$$

Here, we set $V = H_0^1(\Omega)$ and $H = L^2(\Omega)$ such that $V \hookrightarrow H = H^* \hookrightarrow V$ is a Gelfand triple. For fixed $m \in L^\infty(\Omega)$ we define the operator $A(m)$ by

$$\langle A(m)v, w\rangle_{V^*,V} = \rho_F \, (m \nabla v, \nabla w)_{L^2(\Omega)^d} \quad \forall \, v, w \in V, \tag{2.64}$$

and compactly write the acoustic wave equation as

$$E(u, m) = \rho_F u_{tt} + A(m)u, \quad u(0) = 0, \quad u_t(0) = 0. \tag{2.65}$$

We define $a(m) : V \times V \to \mathbb{R}$ by the identity

$$a(m)(v, w) = \langle A(m)v, w\rangle_{V^*,V} \quad \forall \, v, w \in V.$$

and obtain the variational form of the acoustic wave equation as follows:

$$\langle \rho_F u_{tt}(t), v\rangle_{V^*,V} + a(m)(u(t), v) = \langle f(t), v\rangle_{V^*,V} \quad \forall \, v \in V \text{ and a.a. } t \in I. \tag{2.66}$$

Again, we can deduce the existence and uniqueness of solutions to the acoustic wave equation as well as continuity and differentiability of the material-to-solution operator. Due to the simpler structure, we do not have to invoke a superposition operator.

Theorem 2.5.1.
Let $\rho_F > 0$ and consider an open, convex and bounded set $\mathcal{D} \subset L^\infty(\Omega)$ such that there exists $m_a > 0$ with $m(x) \geq m_a$ for almost all $x \in \Omega$. Then, for $k \geq 1$, $l \geq 0$ and $f \in F_0^{k+l+2}$ the acoustic wave equation (2.65) possesses a unique solution $u = u(m) \in C^{k+l+1}(\bar{I}; V) \cap C^{k+l+2}(\bar{I}; H)$ for every $m \in \mathcal{D}$ and the set of solutions $\{u(m) : m \in \mathcal{D}\} \subset C^{k+l+1}(\bar{I}; V) \cap C^{k+l+2}(\bar{I}; H)$ is bounded. Furthermore,

1. *The solution operator $m \mapsto u(m)$ is k-times Lipschitz continuously Fréchet differentiable as a map from $\mathcal{D}$ to $C^l(\bar{I}; V)$.*
2. *The solution operator $m \mapsto u(m)$ is $(k+1)$-times Lipschitz continuously Fréchet differentiable as a map from $\mathcal{D}$ to $C^l(\bar{I}; H)$.*

Proof. First, we note

$$\begin{aligned} |\langle A(m)v, w\rangle_{V^*,V}| &= \rho_F \, |\, (m\,\nabla v, \nabla w)_{L^2(\Omega)^d} \,| \leq \rho_F \, \|m\|_{L^\infty(\Omega)} \|\nabla v\|_{L^2(\Omega)^d} \|\nabla w\|_{L^2(\Omega)^d} \\ &\leq C\|v\|_V \|w\|_V \qquad \forall\ v, w \in V. \end{aligned}$$

By Poincaré's inequality, we further deduce

$$\langle A(m)v, v\rangle_{V^*,V} \geq \rho_F m_a \|\nabla v\|^2_{L^2(\Omega)^d} \geq C\rho_F m_a \|v\|^2_V \quad \forall\ v \in V.$$

Hence, the acoustic wave equation satisfies Assumption 2.2.1 with $\mathcal{S} = \mathcal{D}$ and the statement follows from Corollary 2.2.5 and Corollary 2.2.14. □

Remark 2.5.2.
Instead of the restrictions on the set $\mathcal{D}$ we could alternatively use a superposition operator $\Phi : L^\infty(\Omega) \to L^\infty(\Omega)$ to ensure $\Phi(m)(x) \geq m_a > 0$ for all $m \in L^\infty(\Omega)$. We then introduce

$$\langle A(\Phi(m))v, w\rangle_{V^*,V} = \rho_F \ (\Phi(m)\,\nabla v, \nabla w)_{L^2(\Omega)^d} \quad \forall\ v, w \in V. \tag{2.67}$$

and can work with a setting as in Assumption 2.4.3 to employ the results from section 2.3. Note that the constant density in the acoustic case can obviously be considered as a special case of Assumption 2.3.5 (A4) and (A5).

2.6 Further Perspectives

We conclude this chapter with some remarks on possible extensions and generalizations of the results. The preceding analysis showed that we rely on a higher temporal regularity of the right-hand side in order to establish the continuity and differentiability of the parameter-to-solution operator. The assumptions on the source terms and the regularity of the material parameters are suitable for the kind of tomography problems that we consider in this thesis, cf. Remark 3.2.4.
Nevertheless, it would be interesting to investigate possibilities to weaken the assumptions on the spatial regularity of the seismic source. While it suffices for our purposes to consider V^*, a more accurate modeling of point sources can be preferable for certain applications. In particular, if we want to invert for the seismic source instead of - or in addition to - the identification of the material structure, a different setting might be necessary. Related work with point sources and optimal control problems in measure space exists for elliptic and parabolic problems, see, for instance, [32, 35, 58, 66].
In a different direction, one could also consider a weaker topology of the set of material parameters. We refer to [16] for results on the continuity and differentiability of the solution operator for hyperbolic equations involving a weaker topology on the parameter set.

Chapter 3

Seismic Inverse Problem

In this chapter, we analyze the parameter identification problem for seismic tomography. Here, we start by showing that there exists a solution to the inverse problem and derive the adjoint-based representation of the derivatives utilizing the results from the previous chapter. In a second step, we consider additional constraints on the material parameters and utilize a Moreau-Yosida regularization which is well known and very often used for problems with state constraints [69, 70, 71, 72, 102]. This penalty method leads to an optimality system involving a semismooth operator equation and requires appropriate solution methods. Semismooth Newton-type methods for optimization problems in function spaces have been studied extensively in [67, 125, 126] and have been applied to various types of applications, see, for instance, [86] for an optimal control problem governed by the wave equation. We incorporate the generalized derivatives into a Newton-PCG method with a trust-region globalization. Furthermore, estimates on the constraint violation that are based on results from [68, 126] can be established.
The analysis of this chapter is not limited to problems in seismic tomography, but hold in greater generality for problems in optimal control or parameter identification governed by linear hyperbolic equations with unknown coefficients. Therefore, we carry out most of the work based on the results from section 2.3 and use the term *state equation*. Nevertheless, we give concrete examples tailored to problems in seismic tomography during the course of this chapter.
Parts of this chapter have been published in [19] for the elastic wave equation.

3.1 Parameterization of the Material

The preceding analysis of the differentiability of the solution operator has been carried out with parameters and coefficients in $L^\infty(\Omega)$ which is suitable because the material structure can in general be quite irregular. On the other hand, we will require a higher regularity for a suitable regularization to treat the ill-posedness of the inverse problem. In order to overcome this tradeoff, we split the material properties into a reference model $\overline{\Psi} \in L^\infty(\Omega)^{\hat{n}}$ that is based on a priori knowledge and the parameter variable $m \in L^\infty(\Omega)^n$ that characterizes smooth variations from this reference model. Furthermore, we choose a Hilbert space $M \hookrightarrow\hookrightarrow L^\infty(\Omega)^n$, i.e., M is compactly embedded in $L^\infty(\Omega)^n$ and work with a parameterization given as follows

$$\Psi(m) = \overline{\Psi} + \Phi(m). \tag{3.1}$$

Here, the superposition operator $\Phi : L^\infty(\Omega)^n \to L^\infty(\Omega)^{\hat{n}}$ plays the same role as before and only a slight modification is introduced due to the reference model $\overline{\Psi}$. However, this constant offset does obviously not affect any of the results regarding Lipschitz continuity or differentiability and, furthermore, it holds $\Psi^{(k)}(m) = \Phi^{(k)}(m)$ if the k-th derivative of Φ exists.

Remark 3.1.1.
For problems in global seismology, a suitable reference model based on a priori knowledge is available, e.g., the Preliminary Reference Earth Model (PREM) [46]. Also, it is a common approach in seismology to parameterize for smooth variations from a reference model that often only varies in depth [54, 118, 130].

Building upon the results from section 2.3, we consider the following assumption on the state equation and the involved superposition operator throughout this chapter:

Assumption 3.1.2.
Let Assumption 2.3.5 hold with the parameterization given by (3.1)*. Furthermore, with* $k \geq 1$*, let* $\Phi : L^\infty(\Omega)^n \to L^\infty(\Omega)^{\hat{n}}$ *be* k*-times Lipschitz continuously Fréchet differentiable with* $\Phi^{(i)}(m)$*,* $i = 1, \dots, k$*, uniformly bounded on* $\mathcal{D}$*.*

We recall the set $\mathcal{D}$ that was introduced in the previous chapter when dealing with the existence of a unique solution to the state equation. In general, Assumption 2.3.5 (A1) will only hold true for a subset $\mathcal{D} \subsetneq L^\infty(\Omega)^n$. Consequently, even in the absence of explicit constraints, we have to restrict the analysis of the inverse problem to the set $M_\mathcal{D} := M \cap \mathcal{D}$ which requires some care. To overcome this difficulty, a strategy that allows to choose $\mathcal{D} = L^\infty(\Omega)^n$ by utilizing properties of the superposition operator will be discussed in section 3.2.4. In the meantime, we continue to work with a convex and closed set of admissible material $M_{\text{ad}} \subseteq M_\mathcal{D}$. Furthermore, and without loss of generality, we assume $0 \in M_{\text{ad}}$ and $\Phi(0) = 0$. This implies, in particular, that the reference model is a physically reasonable material, which is certainly a natural assumption.

3.2 Analysis of the Seismic Inverse Problem

In order to treat tomography problems in solid and fluid media in a unified way, we set $\hat{n} = d^4$ for the elastic and $\hat{n} = 1$ for the acoustic case and use the spaces V and H as defined in section 2.4 or, respectively, section 2.5. Furthermore, we define

$$U := L^2(I;V) \cap H^1(I;H) \cap H^2(I;V^*)$$

and consider the state equation

$$E : U \times M \to L^2(I;V^*), \quad E(u,m) := \rho u_{tt} + A(\Psi(m))u, \quad u(0) = 0, \quad u_t(0) = 0. \tag{3.2}$$

We recall that for the elastic wave equation, $A(\Psi(m))$ is defined by (2.58) and, respectively, by (2.67) for the acoustic wave equation.

Remark 3.2.1.
Following the analysis of hyperbolic equations in the previous section, we recall that for a source $f \in F_0^{l+1}$ with some $l \geq 0$, we obtain $u \in C^l(\bar{I};V) \cap C^{l+1}(\bar{I};H) \cap C^{l+2}(\bar{I};V^*) \subset U$, i.e., the initial conditions make sense.

In seismic tomography data is usually taken from several seismic events. We denote the number of seismic sources by n_s and assume to have a set of source functions f_i, $i = 1, \dots, n_s$. Specific assumptions on the regularity of f_i will be made later in this section. For every event there exist measurements on a subdomain $\Omega_i^\delta \times I$, $\Omega_i^\delta \subset \Omega$ and we denote the data by $u_i^\delta \in L^2(I; L^2(\Omega_i^\delta)^{\hat{d}}) \cap H^2(I; (H^1(\Omega_i^\delta)^{\hat{d}})^*)$ with $\hat{d} = d$ for the elastic wave equation and $\hat{d} = 1$ for the acoustic wave equation. The seismic inverse problem can now be stated as follows

$$\min_{\mathbf{u} \in \mathbf{U},\, m \in M_{\text{ad}}} J(\mathbf{u}, m) \tag{3.3}$$

$$\text{s.t.} \quad E(u_i, m) = f_i, \quad u_i(0) = 0, \quad (u_i)_t = 0, \quad i = 1, \dots, n_s,$$

where $\mathbf{u} = (u_1, \dots, u_{n_s})^T \in \mathbf{U} := U^{n_s}$ is a vector of states for every event, e.g., displacement fields for the elastic wave equation. Furthermore, the cost functional $J : \mathbf{U} \times M \to \mathbb{R}$ has the special structure

$$J(\mathbf{u}, m) = \sum_{i=1}^{n_s} J_{\text{fit},i}(u_i) + \alpha J_{\text{reg}}(m),$$

that consists of the accumulated misfit from all seismic sources and a regularization term $J_{\text{reg}} : M \to \mathbb{R}$ with parameter $\alpha > 0$. The misfit functional of event i is denoted by $J_{\text{fit},i} : U \to \mathbb{R}$ and involves data u_i^δ.

Remark 3.2.2.
For notational simplicity, we assume that all seismic events are defined on the same domain $\Omega \times I$. Without any difficulty, however, we could work with event-dependent domains $\Omega_i \subseteq \Omega$ and different time intervals $I_i = [0, T_i]$, $i = 1, \dots, n_s$. As a consequence, this would require different spaces U_i, e.g., $U_i = L^2(I_i, H^1(\Omega_i)^{\hat{d}}) \cap H^1(I_i, L^2(\Omega_i)^{\hat{d}}) \cap H^2(I_i, (H^1(\Omega_i)^{\hat{d}})^*)$. Additionally, we could also consider seismic sources f_i and data u_i^δ of different regularity. For a better readability of the following results, however, we continue to treat all seismic events in a unified way.

3.2.1 Existence of a Solution

We prove the existence of a solution to the seismic inverse problem (3.3) in the following setting.

Assumption 3.2.3.
In addition to Assumption 2.3.5, we require:

(A1) M is a Hilbert space with compact embedding $M \hookrightarrow\hookrightarrow L^\infty(\Omega)^n$.

(A2) For $i = 1, \dots, n_s$: $J_{\text{fit},i} \geq 0$, convex and, with $l \geq 0$, there holds:

- either $f_i \in F_0^l$ and $C^l(\bar{I}; H) \ni u_i \mapsto J_{\text{fit},i}(u_i)$ is continuous,
- or $f_i \in F_0^{l+1}$ and $C^l(\bar{I}; V) \ni u_i \mapsto J_{\text{fit},i}(u_i)$ is continuous.

(A3) $J_{\text{reg}} : M \to \mathbb{R}$ is convex, lower semicontinuous and $J_{\text{reg}}(m) \to \infty$ for $\|m\|_M \to \infty$.

Remark 3.2.4.
In seismic tomography it is valid assumption to have homogeneous initial conditions and a force term that has compact support in $(0,T)$. This means that the simulation starts with the system at rest and the excitations at the hypocenter last only for a limited time that is significantly shorter than the traveltimes to the receiver locations in the far field. The assumption of a higher temporal regularity of the forcing term can be validated since wavelets are most commonly used to model the time evolution of the seismic source. Furthermore, we consider point sources in space which require only a slight smoothing in $V^* = (H^1(\Omega)^{\hat{d}})^*$.

Theorem 3.2.5.
Let Assumption 3.2.3 hold. Then the seismic inverse problem (3.3) *possesses a solution.*

Proof. In a first step we show that the feasible region can be restricted to a convex and closed set that is bounded in M. Since $0 \in M_{\mathrm{ad}} \subset \mathcal{D}$, the state equation possesses a solution for $m = 0$ and $(\mathbf{u}(0), 0) \in \mathbf{U} \times M_{\mathrm{ad}}$ is feasible for (3.3). By Assumption 3.2.3 (A3), there exists $\beta > 0$ such that $J_{\mathrm{reg}}(m) > J_0 := J(\mathbf{u}(0), 0)$ for all $m \in M$ with $\|m\|_M > \beta$. We denote $B_\beta(0) := \{m \in M : \|m\|_M \leq \beta\}$ and deduce that if a minimizer exists, it must be contained in $M_\beta := M_{\mathrm{ad}} \cap B_\beta(0)$ which is a nonempty, convex, closed and bounded subset of M.
Now, let $(\mathbf{u}(m_k), m_k)_{k\in\mathbb{N}} \subset \mathbf{U} \times M_\beta$ be a minimizing sequence with

$$J(\mathbf{u}(m_k), m_k) \to \inf_{m\in M_{\mathrm{ad}}} J(\mathbf{u}(m), m) =: \bar{J}.$$

By Assumption 3.2.3 and Corollary 2.3.7, the solutions $\mathbf{u}(m)$ are uniformly bounded in $\mathbf{U}$ for all $m \in M_{\mathrm{ad}} \subseteq \mathcal{D}$. Furthermore, $(m_k)_{k\in\mathbb{N}} \subset M_\beta$, i.e., $(m_k)_{k\in\mathbb{N}}$ is bounded in M. Hence, since $\mathbf{U}$ and M are reflexive, there exists a weakly convergent subsequence, for simplicity again denoted by $(\mathbf{u}(m_k), m_k)_{k\in\mathbb{N}}$, with weak limit point $(\bar{\mathbf{u}}, \bar{m}) \in \mathbf{U} \times M$. Next, we show that the state equation is weakly sequentially continuous. Here, it suffices to consider only one component of $\mathbf{u}(m_k)$. For simplicity, we denote $u_k = u(m_k)$ and the right-hand side by f (i.e., neglecting the index of the seismic event). For arbitrary $v \in L^2(I;V)$, we obtain

$$\begin{aligned}
&|\langle E(u_k, m_k), v\rangle_{L^2(I;V^*),L^2(I;V)} - \langle E(\bar{u}, \bar{m}), v\rangle_{L^2(I;V^*),L^2(I;V)}| \\
&\quad= \left| \int_0^T \langle \rho(u_k)_{tt} + A(\Psi(m_k))u_k(t), v(t)\rangle_{V^*,V} - \langle \rho\bar{u}_{tt} + A(\Psi(\bar{m}))\bar{u}(t), v(t)\rangle_{V^*,V}\, dt \right| \\
&\quad= \left| \int_0^T \langle \rho((u_k)_{tt} - \bar{u}_{tt}(t)), v(t)\rangle_{V^*,V} + \langle A(\Psi(\bar{m}))(u_k(t) - \bar{u}(t)), v(t)\rangle_{V^*,V} \right. \\
&\qquad\qquad \left. + \langle (A(\Psi(m_k)) - A(\Psi(\bar{m})))u_k(t), v(t)\rangle_{V^*,V}\, dt \right| \\
&\quad\leq \left|\langle (u_k)_{tt} - \bar{u}_{tt}, \rho v\rangle_{L^2(I;V^*),L^2(I;V)}\right| + \left|\langle A(\Psi(\bar{m}))v, u_k - \bar{u}\rangle_{L^2(I;V^*),L^2(I;V)}\right| \\
&\qquad\qquad + C\,\|m_k - \bar{m}\|_{L^\infty(\Omega)^n}\,\|u_k\|_{L^2(I;V)}\,\|v\|_{L^2(I;V)}\,.
\end{aligned} \tag{3.4}$$

Here, in the last step we used the Lipschitz continuity of Ψ, resp. Φ, for the estimate

$$\begin{aligned}
\langle (A(\Psi(m_k)) - A(\Psi(\bar{m})))u_k(t), v(t)\rangle_{V^*,V} &= \langle A(\Psi(m_k) - \Psi(\bar{m}))u_k(t), v(t)\rangle_{V^*,V} \\
&= \langle A(\Phi(m_k) - \Phi(\bar{m}))u_k(t), v(t)\rangle_{V^*,V} \\
&\leq C\|m_k - \bar{m}\|_{L^\infty(\Omega)^n}\|u_k(t)\|_V\|v(t)\|_V.
\end{aligned}$$

Since $u_k \rightharpoonup \bar{u} \in U$, the first two terms tend to zero as $u_k \rightharpoonup \bar{u}$. Moreover, due to the compact embedding $M \hookrightarrow\hookrightarrow L^\infty(\Omega)^n$, we obtain $\|m_k - \bar{m}\|_{L^\infty(\Omega)^n} \to 0$. Since u_k is uniformly bounded in U, also the last term in (3.4) converges to zero and thus, $E(u_k, m_k) \rightharpoonup E(\bar{u}, \bar{m})$. Since $E(u_k, m_k) = f$ for all k, we obtain $E(\bar{u}, \bar{m}) = f$. Furthermore, by the embeddings $H^1(I; H) \hookrightarrow\hookrightarrow C(\bar{I}; H)$ and $H^2(I; V^*) \hookrightarrow\hookrightarrow C^1(\bar{I}; V^*)$, we obtain $u_k \to \bar{u}$ in $C(\bar{I}; H)$ and $(u_k)_t \to \bar{u}_t$ in $C(\bar{I}; V^*)$. Therefore, $0 = u_k(0) \to \bar{u}(0)$ in H, i.e., $\bar{u}(0) = 0$, and $0 = (u_k)_t(0) \to \bar{u}_t(0)$ in V^*, i.e., $\bar{u}_t(0) = 0$. Hence, we obtain $\bar{u} = u(\bar{m})$. Since M_{ad} is closed and convex in M, it is weakly sequentially closed and thus $\bar{m} \in M_{\mathrm{ad}}$. Now, by returning to the full vector, we also obtain $\bar{\mathbf{u}} = \mathbf{u}(\bar{m})$. Furthermore, by Assumption 3.2.3 (A2), $J_{\mathrm{fit},i}$ is convex and continuous, i.e., weakly lower semicontinuous, for every i. Thus, by Assumption 3.2.3 (A3), also J is weakly lower semicontinuous and we obtain

$$\bar{J} = \lim_{k\to\infty} J(\mathbf{u}(m_k), m_k) = \liminf_{k\to\infty} J(\mathbf{u}(m_k), m_k) \geq J(\bar{\mathbf{u}}, \bar{m}).$$

Hence, $(\bar{\mathbf{u}}, \bar{m})$ is a solution to (3.3). □

Due to the unique solvability of the state equation for every $m \in M_{\mathcal{D}}$, we can replace (3.3) by the reduced problem:

$$\min_{m\in M_{\mathrm{ad}}} j(m) := J(\mathbf{u}(m), m) \tag{P}$$

where the displacements $\mathbf{u}(m) = (u_1(m), \ldots, u_{n_s}(m))^T$ solve

$$E(u_i, m) = f_i, \quad u_i(0) = 0, \quad (u_i)_t(0) = 0, \quad i = 1, \ldots, n_s.$$

Thus, the reduced cost functional j is defined by

$$j : M_{\mathcal{D}} \to \mathbb{R}, \quad m \mapsto \sum_{i=1}^{n_s} J_{\mathrm{fit},i}(u_i(m)) + \alpha J_{\mathrm{reg}}(m). \tag{3.5}$$

The restriction of j to $M_{\mathcal{D}}$ instead of M is required, because $u_i(m)$ is not necessarily well-defined on M. Again, we postpone the discussion on the extension of the reduced cost functional to the whole space M to section 3.2.4. Clearly, with the assumptions of Theorem 3.2.5, there exists also a solution to the reduced seismic inverse problem (P). Fréchet differentiability of the reduced cost functional can be established under the following assumptions. Note that in this case the embedding $M \hookrightarrow L^\infty(\Omega)^n$ is not required to be compact.

Assumption 3.2.6.
Let Assumption 3.1.2 hold with some $k \geq 1$. Furthermore, we require:

(A1) M is a Hilbert space with continuous embedding $M \hookrightarrow L^\infty(\Omega)^n$.

(A2) With $l \geq 0$, there holds for $i = 1, \ldots, n_s$:

- either $f_i \in F_0^{k+l+1}$ and $C^l(\bar{I}; H) \ni u_i \mapsto J_{\mathrm{fit},i}(u_i)$ is k-times continuously Fréchet differentiable,
- or $f_i \in F_0^{k+l+2}$ and $C^l(\bar{I}; V) \ni u_i \mapsto J_{\mathrm{fit},i}(u_i)$ is k-times continuously Fréchet differentiable.

(A3) $J_{\mathrm{reg}} : M \to \mathbb{R}$ is k-times continuously Fréchet differentiable.

Theorem 3.2.7.
Let Assumption 3.2.6 be satisfied for given $k \geq 1$. Then the reduced cost functional j defined in (3.5) is k-times continuously Fréchet differentiable on $M_{\mathcal{D}}$.

Proof. Assumption 3.2.6 ensures the applicability of Theorem 2.3.10 which gives the unique solvability of the state equation and the differentiability of the solution operator from $\mathcal{D}$ to $C^l(\bar{I}; V)$ or, respectively, $C^l(\bar{I}; H)$. Thus, due to the embedding $M \hookrightarrow L^\infty(\Omega)^n$, the solution operator is also differentiable as a map with preimage space $M_{\mathcal{D}}$ and the differentiability properties of $J_{\mathrm{fit},i}$ and J_{reg} can be carried over to j. □

Next, we give a specific example for a typical problem setup in seismic tomography and verify Assumption 3.2.3 and Assumption 3.2.6.

Example 3.2.8.
Let $M = (H^2(\Omega) \cap H_0^1(\Omega))^n$, i.e., we have $M \hookrightarrow\hookrightarrow L^\infty(\Omega)^n$ for $d = 2, 3$. Furthermore, we assume homogeneous boundary data, since the material at the surface is usually known and we do not want to update the material parameters at the artificial boundaries of the computational domain. For notational simplicity, we consider only one seismic event with a source given by

$$f(x,t) := s(t)\, F(x), \tag{3.6}$$

where the time evolution is modeled by the Ricker wavelet $s(t)$ centered at t_0 and with dominant source frequency ω. We assume a point source located at $x_s \in \Omega$ and model the geometry of the source by a time-invariant moment tensor applied to a smoothed Dirac measure in V^*, cf. [92]. The corresponding force vector is denoted by $F \in V^*$. In order to ensure that f is zero near $t = 0$, more precisely, that $f \in F_0^3$, we smoothly connect s to zero for $|t - t_0|$ larger than a threshold. Data is observed around receiver locations $x_{r_1}, \ldots, x_{r_p}$ and we set $\Omega^\delta = \bigcup_{j=1}^p B_{r_j}$ where B_{r_j} is a small ball with center x_{r_j}. The cost function is defined by

$$J_{\mathrm{fit}}(u) := \frac{1}{2} \int_0^T g(t) \|u(t) - u^\delta(t)\|^2_{L^2(\Omega^\delta)^{\hat{d}}}\, dt \quad \text{and} \quad J_{\mathrm{reg}}(m) = \frac{1}{2} \|m\|^2_M\,. \tag{3.7}$$

Here, $g : [0,T] \to \mathbb{R}$ with $g \in C_c^\infty(I)$ and $g(t) \geq 0$ is a smooth weighting function that ensures $g(\tau) = g(T - \tau) = 0$ for $\tau \in [0, \epsilon)$ and some small ϵ, i.e., the misfit is only computed in a subinterval of I. With this choice of f and J, Assumption 3.2.3 and Assumption 3.2.6 are clearly satisfied with $k = 2$ and $l = 0$. Note that with $l > 0$, we can also apply the quadratic misfit to u_t or even u_{tt} provided that u_i^δ is sufficiently regular.

For the rest of this chapter, we will continue to work with the cost function given by (3.7) and $f \in F_0^3$. In particular, this guarantees that Theorem 3.2.7 is satisfied with $k \geq 2$, i.e., that the reduced cost functional is at least twice continuously Fréchet differentiable.

3.2.2 Adjoint-based Representation of the Derivatives

Having established the differentiability of the reduced cost functional, we can use the adjoint approach to efficiently compute the first and second derivatives of the reduced cost functional with respect to m. A general derivation can be found, for instance, in [73], section 1.6.
For $i = 1, \dots, n_s$ let $\tilde{L}^i : U \times M \times U \times H \times H \to \mathbb{R}$ be defined by

$$\begin{aligned}\tilde{L}^i(u, m, z, z', z'') :=& J_{\text{fit},i}(u) + \langle E(u, m) - f_i, z\rangle_{L^2(I;V^*),L^2(I;V)} \\ &+ (u(0), z')_H + (u_t(0), z'')_H.\end{aligned}$$

Then the Lagrangian $\tilde{\mathbf{L}} : \mathbf{U} \times M \times \mathbf{U} \times H^{n_s} \times H^{n_s} \to \mathbb{R}$ of the full problem can we written as

$$\tilde{\mathbf{L}}(\mathbf{u}, m, \mathbf{z}, \mathbf{z}', \mathbf{z}'') := \alpha J_{\text{reg}}(m) + \sum_{i=1}^{n_s} \tilde{L}^i(u_i, m, z_i, z_i', z_i'').$$

Note that actually $E : U \times M \to L^2(I; V^*) \subset U^*$, but we will validate the higher regularity of the Lagrange multiplier $z_i \in U$ instead of $L^2(I; V)$ using the results of Chapter 2 in Lemma 3.2.9. The adjoint equation can be stated as

$$\langle \tilde{\mathbf{L}}_{\mathbf{u}}(\mathbf{u}, m, \mathbf{z}, \mathbf{z}', \mathbf{z}''), \hat{\mathbf{u}}\rangle_{\mathbf{U}^*,\mathbf{U}} = 0 \quad \forall\ \hat{\mathbf{u}} \in \mathbf{U}. \tag{3.8}$$

In the next step, we will show that $\mathbf{z}'$ and $\mathbf{z}''$ can be expressed with the help of $\mathbf{z}$. Furthermore, we outline that the adjoint equation can be interpreted as a variant of the state equation backwards in time with a different right-hand side. To this end, we note that the states for different seismic events i can be separated in (3.3). Likewise, the adjoint states corresponding to different events can be computed independently. Hence, (3.8) is equivalent to

$$\langle \tilde{L}^i_u(u_i, m, z_i, z_i', z_i''), \hat{u}\rangle_{U^*,U} = 0 \quad \forall\ \hat{u} \in U, \qquad i = 1, \dots, n_s,$$

which gives after expanding the operator:

$$\begin{aligned}0 = \quad & \int_0^T g(t)(u_i(t) - u_i^\delta(t), \hat{u}(t))_{L^2(\Omega^\delta)^{\hat{d}}}\, dt \\ & + \int_0^T \langle \rho \hat{u}_{tt}(t), z_i(t)\rangle_{V^*,V} + \langle A(\Psi(m))\hat{u}(t), z_i(t)\rangle_{V^*,V}\, dt \\ & + (\hat{u}(0), z_i')_H + (\hat{u}_t(0), z_i'')_H \qquad \forall\ \hat{u} \in U, \qquad\qquad i = 1, \dots, n_s.\end{aligned} \tag{3.9}$$

Therefore, we restrict the further analysis to a single event and drop the index i to improve the readability. Using the embeddings $H^1(I; H) \hookrightarrow C(\bar{I}; H)$, $H^2(I; V^*) \hookrightarrow C^1(\bar{I}; V^*)$ we can apply integration by parts and obtain for $v, w \in U$ (cf. [90], p. 283):

$$\begin{aligned}&\int_0^T \langle \rho v_{tt}(t), w(t)\rangle_{V^*,V}\, dt \\ &\quad = -\int_0^T (\rho v_t(t), w_t(t))_H\, dt + (\rho v_t(T), w(T))_H - (\rho v_t(0), w(0))_H \\ &\quad = \int_0^T \langle \rho w_{tt}(t), v(t)\rangle_{V^*,V}\, dt + (\rho v_t(T), w(T))_H - (\rho v_t(0), w(0))_H \\ &\qquad - (\rho v(T), w_t(T))_H + (\rho v(0), w_t(0))_H.\end{aligned} \tag{3.10}$$

Furthermore, for all $v, w \in V$, we have by the symmetry of $a(\Psi(m))$

$$\langle A(\Psi(m))v, w\rangle_{V^*,V} = a(\Psi(m))(v,w) = a(\Psi(m))(w,v) = \langle A(\Psi(m))w, v\rangle_{V^*,V}. \tag{3.11}$$

Hence, we obtain from (3.9), (3.10) and (3.11) for all $\hat{u} \in U$

$$\begin{aligned} 0 = \quad & \int_0^T g(t)(u(t) - u^\delta(t), \hat{u}(t))_{L^2(\Omega^\delta)^{\hat{d}}} + \langle \rho z_{tt}(t) + A(\Psi(m))z(t), \hat{u}(t)\rangle_{V^*,V}\, dt \\ & + (z' + \rho z_t(0), \hat{u}(0))_H + (z'' - \rho z(0), \hat{u}_t(0))_H \\ & + (\rho z(T), \hat{u}_t(T))_H - (\rho z_t(T), \hat{u}(T))_H. \end{aligned} \tag{3.12}$$

Taking test functions $\hat{u} = \varphi v$ with $v \in V$ and $\varphi \in C_c^\infty(I)$ then gives on $(0,T)$:

$$\langle \rho z_{tt}(t), v\rangle_{V^*,V} + \langle A(\Psi(m))z(t), v\rangle_{V^*,V} = -g(t)(u(t) - u^\delta(t), v)_{L^2(\Omega^\delta)^{\hat{d}}}, \quad \forall v \in V.$$

Furthermore, considering all $\varphi \in C^\infty(\bar{I})$ with $\varphi(T) = \varphi_t(T) = \varphi_t(0) = 0$ gives $z' = -\rho z_t(0)$ and, similarly, considering $\varphi \in C^\infty(\bar{I})$ with $\varphi(T) = \varphi_t(T) = \varphi(0) = 0$ yields $z'' = \rho z(0)$. Likewise, by considering all $\varphi \in C^\infty(\bar{I})$ with $\varphi(0) = \varphi_t(0) = \varphi(T) = 0$, we obtain $\rho z(T) = 0$ and test functions with $\varphi(0) = \varphi_t(0) = \varphi_t(T) = 0$ finally yields $\rho z_t(T) = 0$. We note that due to Assumption 2.3.5 we have $\rho v = 0 \Leftrightarrow v = 0$ for all $v \in V$. Summing up, we can eliminate $\mathbf{z}'$ and $\mathbf{z}''$ and interpret the adjoint equation as a vector of state equations backwards in time with right-hand sides given by $-J'_{\mathrm{fit},i}(u_i)$ and write

$$E(z_i, m) = -J'_{\mathrm{fit},i}(u_i), \quad z_i(T) = 0, \quad (z_i)_t(T) = 0, \quad i = 1, \dots, n_s. \tag{3.13}$$

Thus, we can apply the results from Chapter 2 to establish existence, uniqueness and regularity of the adjoint states.

Lemma 3.2.9.
Let $m \in M_{\mathcal{D}}$, $u_i^\delta \in L^2(I; L^2(\Omega_i^\delta)^{\hat{d}}) \times H^2(I; (H^1(\Omega_i^\delta)^{\hat{d}})^*)$, $f_i \in F_0^3$, $i = 1, \dots, n_s$, and the misfit functional given by (3.7). Then there exists a unique adjoint state $z_i(m) \in C^1(\bar{I}; V) \cap C^2(\bar{I}; H)$. Furthermore, $\{z_i(m) : m \in M_{\mathcal{D}}\} \subset C^1(\bar{I}; V) \cap C^2(\bar{I}; H)$ is bounded and the mapping $m \mapsto z_i(m)$ is Lipschitz continuous from $M_{\mathcal{D}}$ to $C(\bar{I}; V) \cap C^1(\bar{I}; H)$.

Proof. We use a time transformation $\tau := T - t$ in order to work with initial time conditions and drop the index i. By Corollary 2.3.7, we obtain $u(m) \in C^2(I; V)$, hence $(u(m) - u^\delta)\mathbf{1}_{\Omega^\delta} \in H^2(I; V^*)$, where $\mathbf{1}_{\Omega^\delta}$ denotes the indicator function on $\Omega^\delta \subseteq \Omega$. Furthermore, g ensures that the adjoint right-hand side is in F_0^2. Again by Corollary 2.3.7, we deduce the existence of a unique adjoint state $z(m) \in C^1(\bar{I}; V) \cap C^2(\bar{I}; H)$. Furthermore, since $u(m) \in U$ is uniformly bounded on $M_{\mathcal{D}} \subseteq \mathcal{D}$, the adjoint right-hand side is uniformly bounded in $H^2(I; V^*)$ for all $m \in M_{\mathcal{D}}$. Hence, also $\{z(m) : m \in M_{\mathcal{D}}\} \subset C^1(\bar{I}; V) \cap C^2(\bar{I}; H)$ is bounded. Similar to the techniques in the previous sections, we obtain for sufficiently small $s \in M$ that $h := z(m+s) - z(m)$ satisfies

$$E(h, m) = \hat{f}, \quad h(0) = 0, \quad h_t(0) = 0, \tag{3.14}$$

with

$$\hat{f} = -\left(A(\Psi(m+s)) - A(\Psi(m))\right) z(m+s) - g\left(u(m+s) - u(m)\right)\mathbf{1}_{\Omega^\delta}.$$

With the linearity of A, the Lipschitz continuity of the superposition operator and the uniform boundedness of $z(m+s) \in C^1(\bar{I};V) \cap C^2(\bar{I};H)$ we obtain

$$\begin{aligned}&\| \left(A(\Psi(m+s)) - A(\Psi(m))\right) z(m+s)\|_{C^1(\bar{I};V^*)} \\ &\quad = \| \left(A(\Phi(m+s) - \Phi(m))\right) z(m+s)\|_{C^1(\bar{I};V^*)} \leq C\|s\|_M.\end{aligned}$$

Furthermore, the Lipschitz continuity of the solution operator (cf. Theorem 2.3.8) yields

$$\|g\left(u(m+s) - u(m)\right) \mathbf{1}_{\Omega^d}\|_{C^1(\bar{I};V)} \leq C\| \left(u(m+s) - u(m)\right) \|_{C^1(\bar{I};V)} \leq C\|s\|_M.$$

Hence, the right-hand side in (3.14) is in F_0^1 and bounded by $C\|s\|_M$ and we deduce the Lipschitz continuity of the mapping $m \mapsto z(m)$ from $M_{\mathcal{D}}$ to $C(\bar{I};V) \cap C^1(\bar{I};H)$ by Corollary 2.3.7. □

We will continue to denote the unique adjoint state for a given m by $\mathbf{z}(m)$ or, respectively, $z(m)$. Furthermore, in order to simplify the notation we eliminate $\mathbf{z}'$ and $\mathbf{z}''$ and abbreviate

$$\mathbf{L}(\mathbf{u}, m, \mathbf{z}) := \tilde{\mathbf{L}}(\mathbf{u}, m, \mathbf{z}, -\rho\mathbf{z}_t(0), \rho\mathbf{z}(0))$$

and, respectively,

$$L^i(u, m, z) := \tilde{L}^i(u, m, z, -\rho z_t(0), \rho z(0)).$$

With the help of the adjoint state $\mathbf{z}$, the first derivative of j is equal to

$$j'(m) = \mathbf{L}_m(\mathbf{u}(m), m, \mathbf{z}(m)).$$

Hence, with $u_i = u_i(m)$ and $z_i = z_i(m)$, we obtain

$$\langle j'(m), \hat{m}\rangle_{M^*,M} = \alpha\langle J'_{\text{reg}}(m), \hat{m}\rangle_{M^*,M} + \sum_{i=1}^{n_s} \int_0^T \langle A(\Phi'(m)\hat{m})u_i(t), z_i(t)\rangle_{V^*,V}\, dt \quad \forall\ \hat{m} \in M.$$

We introduce the form $D : M \times L^2(I;V) \times L^2(I;V) \to M^*$ defined by

$$D(m, v, w)(\hat{m}) = \int_0^T \langle A(\Phi'(m)\hat{m})v(t), w(t)\rangle_{V^*,V}\, dt \quad \forall\ \hat{m} \in M. \tag{3.15}$$

Using the Riesz representation for the derivative of the regularization term, the first derivative of j at a given $m \in M$ can then be expressed as

$$\langle j'(m), \hat{m}\rangle_{M^*,M} = \alpha(m, \hat{m})_M + \sum_{i=1}^{n_s} D(m, u_i(m), z_i(m))(\hat{m}) \quad \forall\ \hat{m} \in M. \tag{3.16}$$

Now we proceed with specific examples of the first derivatives for the elastic wave equation.

Example 3.2.10.
We consider the elastic wave equation with the full material tensor and A as defined in (2.58). Then the operator D is given by

$$\begin{aligned}D(m, v, w)(\hat{m}) &= \int_0^T \left((\Phi'(m)\hat{m}) : \varepsilon(v)(t), \varepsilon(w)(t)\right)_{L^2(\Omega)^{d\times d}}\, dt \\ &= \int_0^T \int_\Omega (\varepsilon(v)(x,t) \otimes \varepsilon(w)(x,t)) :: \left(\left(\Phi'(m)\hat{m}\right)(x)\right)\, dx\, dt,\end{aligned}$$

where we used the usual notation for tensor products, $(a \otimes b)_{ijkl} = a_{ij}b_{kl}$ and $A :: B = \sum_{ijkl} A_{ijkl}B_{ijkl}$.
In the isotropic case, we consider A defined as in (2.59) and obtain

$$D(m,v,w)(\hat{m}) = \int_0^T \int_\Omega \left(\sum_{i=1}^d \frac{\partial}{\partial x_i} u^i(x,t) \, \frac{\partial}{\partial x_i} z^i(x,t) \right) \cdot \left(\lambda'(m)\hat{m}\right)(x) + 2 \left(\sum_{i,j=1}^d [\varepsilon(u)(x,t)]_{ij} \, [\varepsilon(z)(x,t)]_{ij} \right) \cdot \left(\mu'(m)\hat{m}\right)(x) \quad dx\,dt.$$

Note that the superscript i in the formulas above refers to the components of the forward and, respectively, the adjoint wavefield and not to a seismic event.

Later on, it will be helpful to utilize that the derivative of the reduced cost functional is bounded on bounded subsets of $M_{\mathcal{D}}$.

Lemma 3.2.11.
Let the assumptions of Lemma 3.2.9 hold. Then, for any subset $\mathcal{G} \subseteq M_{\mathcal{D}}$ that is bounded in M, $j'(m) \in M^*$ is uniformly bounded on $\mathcal{G}$.

Proof. We use the adjoint-based representation of the derivative as given in (3.16). In fact, for every m in $\mathcal{G}$ we obtain for arbitrary $\hat{m} \in M$

$$\begin{aligned} |\langle j'(m), \hat{m} \rangle_{M^*,M}| &\leq \alpha \|m\|_M \|\hat{m}\|_M + \sum_{i=1}^{n_s} c \|u_i(m)\|_{L^2(I;V)} \|z_i(m)\|_{L^2(I;V)} \|\hat{m}\|_{L^\infty(\Omega)^n} \\ &\leq (\alpha \|m\|_M + C) \|\hat{m}\|_M, \end{aligned} \tag{3.17}$$

since we have already shown that $u_i(m)$ and $z_i(m)$ are uniformly bounded on $M_{\mathcal{D}}$. Hence, $j'(m) \in M^*$ is uniformly bounded on any bounded subset $\mathcal{G} \subseteq M_{\mathcal{D}}$. □

Note that the boundedness of $\mathcal{G}$ is only required for the derivative of the regularization term. Alternatively, we could have derived that $j'(m) \in M^*$ is uniformly bounded by utilizing the Lipschitz continuity of the derivative (Theorem 3.2.7) and the boundedness of $\mathcal{G}$.
Next, we turn to the adjoint-based representation of the second derivatives. To this end, we require the second derivatives of L^i which are given as follows:

$$\begin{aligned} \langle L^i_{uu}(u,m,z)\hat{u}_1, \hat{u}_2 \rangle_{U^*,U} &= \left(g\hat{u}_1, \hat{u}_2\right)_{L^2(I;L^2(\Omega^\delta)^{\hat{d}})}, \\ \langle L^i_{um}(u,m,z)\hat{m}, \hat{u} \rangle_{U^*,U} &= D(m,\hat{u},z)(\hat{m}), \\ \langle L^i_{mu}(u,m,z)\hat{u}, \hat{m} \rangle_{M^*,M} &= D(m,\hat{u},z)(\hat{m}), \\ \langle L^i_{mm}(u,m,z)\hat{m}_1, \hat{m}_2 \rangle_{M^*,M} &= \int_0^T \langle A(\Phi''(m)(\hat{m}_1,\hat{m}_2))u(t), z(t) \rangle_{V^*,V} \, dt. \end{aligned}$$

Now, operator-vector products $j''(m)s$ for a given perturbation $s \in M$ can be computed at the cost of two additional simulations per seismic event. Again, we refer to [73], section 1.6, for a detailed derivation and just summarize the required steps, which are given as follows.

For every $i = 1, \dots, n_s$:

1. Compute a perturbed forward wavefield $\delta_s u_i$ by solving

$$E(\delta_s u_i, m) = -A(\Phi'(m)s)u_i(m), \quad \delta_s u_i(0) = 0, \quad (\delta_s u_i)_t(0) = 0. \tag{3.18}$$

2. Compute a perturbed adjoint wavefield $\delta_s z_i$ by solving

$$E(\delta_s z_i, m) = -g\,\delta_s u_i \mathbf{1}_{\Omega^\delta} - A(\Phi'(m)s)z_i(m),\ \delta_s z_i(T) = 0,\ (\delta_s z_i)_t(T) = 0. \tag{3.19}$$

Then, $j''(m)s$ is given by: $\forall\ \hat{m} \in M$

$$\langle j''(m)s, \hat{m}\rangle_{M^*,M} = \alpha(s,\hat{m})_M + \sum_{i=1}^{n_s} \Big(D(m, \delta_s u_i, z_i(m))(\hat{m}) + D(m, u_i(m), \delta_s z_i)(\hat{m}) + \int_0^T \langle A(\Phi''(m)(s,\hat{m}))u_i(m)(t), z_i(m)(t)\rangle_{V^*,V}\, dt \Big).$$

Note that the results from Chapter 2 and Lemma 3.2.9 can be applied to deduce that $\delta_s u_i$ and $\delta_s z_i$ are uniquely determined and bounded in U.

3.2.3 Optimality Conditions

Now, we turn to the discussion of first-order optimality conditions for the reduced problem (P). We recall the assumptions from the previous sections that j is twice continuously differentiable on $M_{\mathcal{D}}$ and $M_{\text{ad}} \subseteq M_{\mathcal{D}}$ is a closed and convex set with $0 \in M_{\text{ad}}$. Then the classical result for necessary first order conditions can be stated:

Theorem 3.2.12.
Let $\bar{m} \in M$ be a local solution of (P)*. Then the following optimality condition holds:*

$$\bar{m} \in M_{ad}, \quad \langle j'(\bar{m}), \hat{m} - \bar{m}\rangle_{M^*,M} \geq 0 \quad \forall\ \hat{m} \in M_{ad}. \tag{3.20}$$

Proof. See Theorem 1.46 in [73]. □

Note that the necessary optimality conditions (3.20) implicitly also contain the state and adjoint equation in $j'(m)$. Therefore, we can state an alternative formulation with the help of the Lagrangian, cf. Corollary 1.3 in [73].

Corollary 3.2.13.
Let $(\bar{\mathbf{u}}, \bar{m}) \in \mathbf{U} \times M_{ad}$ be a local solution of (3.3)*. Then there exists an adjoint state $\bar{\mathbf{z}} = (\bar{z}_1, \dots, \bar{z}_{n_s}) \in \mathbf{U}$ such that the following optimality conditions hold*

$$\begin{aligned} \langle L_z^i(\bar{u}_i, \bar{m}, \bar{z}_i), \hat{z}\rangle_{U^*,U} &= 0 \quad \forall\ \hat{z} \in U, \quad i = 1, \dots, n_s, \\ \langle L_u^i(\bar{u}_i, \bar{m}, \bar{z}_i), \hat{u}\rangle_{U^*,U} &= 0 \quad \forall\ \hat{u} \in U, \quad i = 1, \dots, n_s, \\ \bar{m} \in M_{ad}, \quad \langle \mathbf{L}_m(\bar{\mathbf{u}}, \bar{m}, \bar{\mathbf{z}}), \hat{m} - \bar{m}\rangle_{M^*,M} &\geq 0 \quad \forall\ \hat{m} \in M_{ad}. \end{aligned} \tag{3.21}$$

Using the fact that M_{ad} is a closed and convex set and M is a Hilbert space, we can also rewrite the necessary optimality conditions with the help of a projection operator.

Corollary 3.2.14.
Let $\bar{m} \in M$ be a local solution of (P) *and let $P_{M_{ad}} : M \to M_{ad}$ denote the projection onto M_{ad} with respect to $\|.\|_M$, i.e., $\forall\ m \in M$:*

$$P_{M_{ad}}(m) \in M_{ad}, \quad \|P_{M_{ad}}(m) - m\|_M = \min_{\hat{m} \in M_{ad}} \|\hat{m} - m\|_M. \tag{3.22}$$

Then $\bar{m}$ satisfies the following optimality condition:

$$\bar{m} = P_{M_{ad}}(\bar{m} - \theta \nabla j(\bar{m})),$$

where $\nabla j(\bar{m}) \in M$ denotes the Riesz representation of $j'(\bar{m}) \in M^*$ and $\theta > 0$ is arbitrary, but fixed.

Proof. See Lemma 1.11 in [73]. □

Remark 3.2.15.
Clearly, if a local solution $\bar{m}$ is in the interior of M_{ad}, i.e., if there exists an open neighborhood B of $\bar{m}$ with $B \subset M_{\text{ad}}$, then the optimality conditions (3.20) simplify and $\bar{m}$ satisfies $j'(\bar{m}) = 0$ or, respectively, $\nabla j(\bar{m}) = 0$ In particular, this holds true for every local solution if $M_{\text{ad}} = M_{\mathcal{D}} = M$.

If $M_{\text{ad}} = M_{\mathcal{D}}$, the optimality conditions (3.20) or (3.21) already give the starting point for Newton-type optimization methods. In the presence of additional constraints more work is required and we will analyze the Moreau-Yosida regularization to handle the feasible set M_{ad}. In both cases, it will be helpful if the reduced cost functional is well-defined on the whole space M. Therefore, we proceed with a strategy to ensure this before we turn to the discussion of suitable optimization methods.

3.2.4 Evaluating the Reduced Cost Functional on M

In this section, we describe a strategy how the reduced cost functional can be extended to M. What complicates matters and led to the definition of the set $M_{\mathcal{D}}$ is the fact that, in general, Assumption 2.3.5 (A1) is only satisfied on a subset $\mathcal{D} \subsetneq L^\infty(\Omega)^n$. This means that even in the absence of explicit constraints, we have to restrict the analysis to the set $M_{\mathcal{D}} = M \cap \mathcal{D}$ instead of seeking a solution to the inverse problem in M. Consequently, we have to work with an admissible set $M_{\text{ad}} \subseteq M$ that satisfies $M_{\text{ad}} \subseteq M_{\mathcal{D}}$. A natural way to define M_{ad} would be to consider constraints on the material parameters which will be dealt with in section 3.3. However, even if we have explicit constraints on the material, we still cannot ensure a priori that m stays within reasonable bounds, i.e., within $M_{\mathcal{D}}$, during the sequence of the Moreau-Yosida regularized problems. This may cause two problems. On the one hand, this can yield an operator $A(\Psi(m))$ violating Assumption 2.3.5 (A1) such that $u(m)$ might not be well-defined. On the other hand, difficulties regarding the discretization of the problem arise since the resolution of the computational grid is based on the source frequency and the wave velocities. Thus, an unreasonable parameter model might require a very fine mesh.
This motivates a strategy that employs suitable properties of the superposition operator Φ to circumvent these problems. In particular, we want to ensure that Assumption 2.3.5 (A1) is satisfied for $\mathcal{D} = L^\infty(\Omega)^n$ which gives $M_{\mathcal{D}} = M$. To this end, we propose a smooth cutoff function φ which guarantees that the parameters always remain within a certain range. In

order to simplify the notation, we present the one-dimensional case, but the extension to multiple parameter fields is straightforward. Consider $\bar{x}_l, x_l, x_u, \bar{x}_u \in \mathbb{R}$ with $\bar{x}_l < x_l < x_u < \bar{x}_u$ and $\varphi : \mathbb{R} \to \mathbb{R}$ that satisfies the following properties:

- φ is monotonically increasing and at least twice continuously differentiable,
- $\varphi \equiv \mathrm{id}$ on $[x_l, x_u]$ and
- $\varphi(\mathbb{R}) = [\varphi(\bar{x}_l), \varphi(\bar{x}_u)]$.

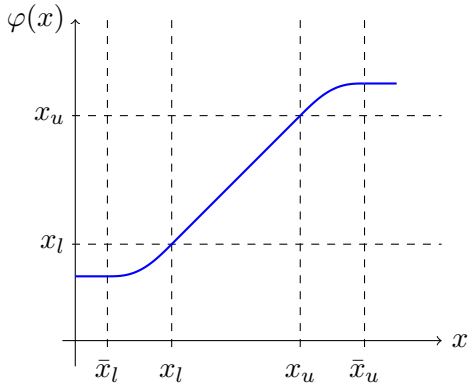


Figure 3.1: Smooth cutoff function to ensure that the parameters remain within physically reasonable bounds. A fourth order polynomial is used for the smooth transition in $[\bar{x}_l, x_l]$ and $[x_u, \bar{x}_u]$.

We now provide a specific example of such a cutoff function that is also depicted in Figure 3.1.

Example 3.2.16.
A cutoff function that has the properties outlined above can be defined as follows:

$$\varphi(x) := \begin{cases} \tilde{\varphi}_l(\bar{x}_l), & x < \bar{x}_l, \\ \tilde{\varphi}_l(x), & \bar{x}_l \le x < x_l, \\ x, & x_l \le x \le x_u, \\ \tilde{\varphi}_u(x), & x_u < x \le \bar{x}_u, \\ \tilde{\varphi}_u(\bar{x}_u), & x > \bar{x}_u, \end{cases} \tag{3.23}$$

with fourth order polynomials $\tilde{\varphi}_l$ and $\tilde{\varphi}_u$ that satisfy the following interpolation conditions:

$$\begin{array}{lllll} \tilde{\varphi}_l(x_l) = x_l, & \tilde{\varphi}_l'(x_l) = 1, & \tilde{\varphi}_l'(\bar{x}_l) = 0, & \tilde{\varphi}_l''(x_l) = 0, & \tilde{\varphi}_l''(\bar{x}_l) = 0, \\ \tilde{\varphi}_u(x_u) = x_u, & \tilde{\varphi}_u'(x_u) = 1, & \tilde{\varphi}_u'(\bar{x}_u) = 0, & \tilde{\varphi}_u''(x_u) = 0, & \tilde{\varphi}_u''(\bar{x}_u) = 0. \end{array} \tag{3.24}$$

Hence, $\tilde{\varphi}_l$ and $\tilde{\varphi}_u$ are uniquely determined by (3.24) and we obtain

$$\tilde{\varphi}_l(x) = \frac{1}{2(x_l - \bar{x}_l)^3}\left(-x^4 + 2(x_l + \bar{x}_l)x^3 - 6x_l\bar{x}_l x^2 + 2(3x_l - \bar{x}_l)\bar{x}_l^2 x + x_l^3(x_l - 2\bar{x}_l)\right)$$
$$\tilde{\varphi}_u(x) = \frac{1}{2(x_u - \bar{x}_u)^3}\left(-x^4 + 2(x_u + \bar{x}_u)x^3 - 6x_u\bar{x}_u x^2 + 2(3x_u - \bar{x}_u)\bar{x}_u^2 x + x_u^3(x_u - 2\bar{x}_u)\right).$$

Now, we can use the cutoff function within the superposition operator Φ in the parameterization of A. Thus, we ensure that Assumption 2.3.5 is satisfied with $\mathcal{D} = L^\infty(\Omega)^n$, i.e., $M_\mathcal{D} = M$, and that the reduced cost function is well-defined and differentiable on M. Here, the inner bounds x_l and x_u should be chosen large enough to cover all models that are physically reasonable. In particular, the solution to the inverse problem should only have values within $[x_l, x_u]$ such that the cutoff function does not artificially influence the reconstruction. The outer bounds should be chosen such that $[\varphi(\bar{x}_l), \varphi(\bar{x}_u)]$ covers only models for which the state equation is well-defined.

Remark 3.2.17.
Note that the proposed cutoff function ensures that only reasonable values for m enter into the elliptic operator in the state equation. On the other hand, this does not affect the regularization term. Thus, while the misfit term of the objective function will not increase for parameter models that exceed the outer bounds, the regularization term J_reg still penalizes large deviations from the reference model.

Remark 3.2.18.
The cutoff function is only a valid approach if the bounds induced by the set $\mathcal{D}$ are "soft constraints" in the sense that we do not expect that these bounds are active in the global minimum of the seismic inverse problem (P). Other bounds should be treated properly by explicit constraints.

3.3 Constrained Parameter Identification Problem

In this section, we consider the case where additional constraints on the parameters are explicitly available. These constraints can be based on a priori knowledge and help to address the ill-posedness and non-convexity of the problem. In order to simplify the notation, we assume that for $n > 1$ every parameter field has the same regularity, i.e., we consider $M = M_1^n$ with a Hilbert space $M_1 \hookrightarrow\hookrightarrow L^\infty(\Omega)$.

Remark 3.3.1.
From an application point of view, the assumption $M = M_1^n$ is reasonable since the different components of m describe similar quantities, e.g., the velocities of compressional and shear waves, the bulk modulus or the Lamé coefficients. However, this restriction is mainly done for notational convenience and we will comment on possible generalizations in Remark 3.3.18.

Now, we introduce additional constraints on the set of feasible parameters. To this end, let $m^a, m^b \in M$ with $m^a \leq 0 < m^b$. Furthermore, let $p \in \mathbb{N}_0$, $g^a, g^b \in M_1^p$ with $g^a \leq 0 < g^b$ and $B \in \mathbb{R}^{p \times n}$. We define pointwise constraints on the parameters by

$$M_\text{ad}^\infty = \left\{ m \in L^\infty(\Omega)^n \,:\, m^a \leq m \leq m^b,\ g_i^a \leq \sum_{j=1}^n b_{ij} m_j \leq g_i^b,\ i = 1, \ldots, p \right\}. \tag{3.25}$$

By construction, the set M_ad^∞ is convex and bounded. Furthermore, it is nonempty since $0 \in M_\text{ad}^\infty$. In addition, we assume throughout this section that the bounds are given such that $M_\text{ad}^\infty \subset \mathcal{D}$, hence $M_\text{ad} = M_\text{ad}^\infty \cap M_\mathcal{D}$.

Remark 3.3.2.
We do not necessarily require that $0 \in M_{\text{ad}}^{\infty}$ and could in principle allow for more general bounds $m^a < m^b$ and $g^a < g^b$. However, recalling the parameterization (3.1), it seems reasonable to assume a feasible reference model, i.e., zero should be a feasible parameter variation.

From the definition of the feasible set it is obvious that simple box constraints can be imposed by choosing $B = 0$ or $p = 0$. Note, however, that (3.25) allows for more general constraints that act on other physical quantities as well. The following examples show, how lower and upper bounds on the Poisson's ratio of the material can be enforced or the difference between the velocities of P- and S-waves can be limited.

Example 3.3.3.
For isotropic material, the Poisson's ratio ν can be expressed in terms of the Lamé coefficients as $\nu(x) = \lambda(x)/(2(\lambda(x)+\mu(x)))$, see [101]. Since λ and μ are positive, lower and upper bounds in the form $\nu^a \leq \nu(x) \leq \nu^b$ can be rewritten as $2\nu^a\mu + (2\nu^a - 1)\lambda \leq 0$ and $-2\nu^b\mu + (1 - 2\nu^b)\lambda \leq 0$. If we jointly invert for both Lamé coefficients, we can define $\lambda(m) = \bar{\lambda} + m_1$ and $\mu(m) = \bar{\mu} + m_2$ with a reference model $(\bar{\lambda}, \bar{\mu})$. Let $\bar{\nu}$ denote the Poisson's ratio of the reference model. Then the inequalities can be rearranged to

$$\begin{aligned} (2\nu^a - 1)m_1 + 2\nu^a m_2 &\leq \bar{\lambda}\,(1 - \frac{\nu^a}{\bar{\nu}}), \\ (1 - 2\nu^b)m_1 - 2\nu^b m_2 &\leq \bar{\lambda}(\frac{\nu^b}{\bar{\nu}} - 1). \end{aligned} \tag{3.26}$$

Thus, we set $p = 2$ and

$$B = \begin{pmatrix} (2\nu^a - 1) & 2\nu^a \\ (1 - 2\nu^b) & -2\nu^b \end{pmatrix}, \quad g^b = \begin{pmatrix} \bar{\lambda}\,(1 - \frac{\nu^a}{\bar{\nu}}) \\ \bar{\lambda}(\frac{\nu^b}{\bar{\nu}} - 1) \end{pmatrix}.$$

Assuming that the reference model has a strictly feasible Poisson's ratio, we obtain $g^b > 0$. Note that we only have upper bounds in (3.26), but using the fact that $\nu \in (0, \frac{1}{2}]$, we could easily add an (artificial) lower bound by setting

$$g^a = \begin{pmatrix} (2\nu^a - 1)m_1^b + 2\nu^a m_2^a \\ (1 - 2\nu^b)m_1^a - 2\nu^b m_2^b \end{pmatrix}.$$

Example 3.3.4.
Let v_p and v_s denote the velocities of compressional and, respectively, shear waves. Recalling (1.12) and (1.14), we obtain

$$\mu = \rho\, v_s^2, \quad \lambda + 2\mu = \rho\, v_p^2.$$

Thus, if we again consider $\lambda(m) = \bar{\lambda} + m_1$ and $\mu(m) = \bar{\mu} + m_2$ and seek to limit the difference of the squared velocities of compressional and shear waves by fixed bounds v^a, v^b, i.e.,

$$v^a \leq v_p^2 - v_s^2 \leq v^b,$$

we can define $p = 1$, $B = (1 \quad 1)$ and

$$g^a = \rho v^a - (\bar{\lambda} + \bar{\mu}), \quad g^b = \rho v^b - (\bar{\lambda} + \bar{\mu}).$$

In order to facilitate the notation for the forthcoming analysis, we define $n_c := n + p$, $M_c := M_1^{n_c}$, $\psi^a := (m^a, g^a)^T$, $\psi^b := (m^b, g^b)^T$ and compactly write the set M_{ad}^∞ defined in (3.25) with the help of a linear operator $S \in \mathcal{L}(M, M_c)$ as

$$M_{\text{ad}}^\infty = \left\{ m \in L^\infty(\Omega)^n \,:\, \psi^a \leq Sm \leq \psi^b \right\}. \tag{3.27}$$

3.3.1 Moreau-Yosida Regularization

In this section, we discuss solution strategies for the constrained parameter identification problem. The constraints induced by M_{ad}^∞ are handled by the Moreau-Yosida regularization. This method is commonly used for state-constrained problems, see e.g. [69, 70, 71]. In particular, we can apply a semismooth Newton method to the penalized problem and establish estimates on the constraint violation for an increasing penalty parameter. Before we start the analysis, the assumptions required in this section are summarized for convenience.

Assumption 3.3.5.
In addition to Assumption 3.1.2, we require:

(A1) Assumption 2.3.5 (A1) is satisfied for $\mathcal{D} = L^\infty(\Omega)^n$.

(A2) The space of material perturbations is given by $M = M_1^n$ with a Hilbert space $M_1 \hookrightarrow\hookrightarrow L^\infty(\Omega)$.

(A3) The feasible set M_{ad} is given by $M_{\text{ad}}^\infty \cap M$ with $M_{\text{ad}}^\infty \subset \mathcal{D}$ as defined in (3.27).

(A4) J_{fit} and J_{reg} are given as in Example 3.2.8.

As has been pointed out before, Assumption 3.3.5 (A1) can, for instance, be ensured with the help of a superposition operator. Note, however, that the superposition operator is only used in the weak form of the state equation and does not appear in the constraints.
For a fixed $\gamma \in (0, \infty)$ we define the penalized problem

$$\min_{m \in M_{\mathcal{D}}} \quad j_\gamma(m) := j(m) + \gamma\phi(m), \tag{P_γ}$$

with the penalty function

$$\phi(m) := \frac{1}{2}\left(\|[Sm - \psi^b]^+\|^2_{L^2(\Omega)^{n_c}} + \|[\psi^a - Sm]^+\|^2_{L^2(\Omega)^{n_c}} \right). \tag{3.28}$$

Here, $[.]^+$ is a vector defined pointwise by $([v(x)]^+)_i = \max\{v_i(x), 0\}$, $i = 1, \ldots, n_c$.
Following section 3.2.3, we can state the optimality conditions for the Moreau-Yosida regularized problem. To this end, let $\bar{m}_\gamma$ be a local solution of (P_γ). Then the following first order optimality conditions hold:

$$j'(\bar{m}_\gamma) + \gamma S^* \left([S\bar{m}_\gamma - \psi^b]^+ - [\psi^a - S\bar{m}_\gamma]^+ \right) = 0 \quad \text{in } M^*. \tag{3.29}$$

Equivalently, we obtain in variational form

$$\langle j'(\bar{m}_\gamma), \hat{m} \rangle_{M^*, M} + \gamma \left([S\bar{m}_\gamma - \psi^b]^+ - [\psi^a - S\bar{m}_\gamma]^+, S\hat{m} \right)_{L^2(\Omega)^{n_c}} = 0 \quad \forall\, \hat{m} \in M. \tag{3.30}$$

We will return to the optimality conditions in section 3.3.2 Before, we want to achieve three things. First, we show that there exists a solution to the penalized problem (P_γ). Afterwards, we prove that for a sequence of penalty parameters $\gamma \to \infty$ and solutions $(m_\gamma)_\gamma$ to (P_γ), every weak limit point will be a solution of (P). Finally, we deduce estimates on the constraint violation depending on γ. Note that the constraints are formulated in L^∞ while the Moreau-Yosida regularization (3.28) penalizes violation of the constraints in L^2. Since $\|v_k\|_{L^2(\Omega)} \to 0$ does not imply $\|v_k\|_{L^\infty(\Omega)} \to 0$, we will utilize an interpolation inequality between L^2 and a stronger space.

Assumption 3.3.6.

(A1) Either $M_1 \hookrightarrow W^{1,q'}(\Omega)$ holds for some $q' > d$,
or $M_1 \hookrightarrow C^{0,\beta}(\bar{\Omega})$ with $0 < \beta < 1$ and $m = 0$ on $\partial\Omega$ for all $m \in M$.

(A2) There exist $\bar{\psi} \in \mathbb{R}$ and $\hat{m} \in M$ such that $\psi_i^a(x) \leq 0 < \bar{\psi} \leq (S\hat{m})_i(x) < \psi_i^b(x)$ a.e. in Ω for all $i = 1, \dots, n_c$.

Lemma 3.3.7 (Interpolation between L^q and $W^{1,q'}$)**.**
Let $\Omega \subset \mathbb{R}^d$ be open, bounded and with Lipschitz boundary. Furthermore, let $q' > d$ and $q \in [1, \infty]$. Then

$$\|v\|_{L^\infty(\Omega)} \leq C \, \|v\|^\theta_{W^{1,q'}(\Omega)} \|v\|^{1-\theta}_{L^q(\Omega)} \quad \textit{with} \quad \theta = \frac{dq'}{dq' - (q'-d)q}$$

and a constant $C = C(d, q', q)$ that is independent of v and Ω.

Proof. This is a special case of Theorem 5.9 in [1]. □

Lemma 3.3.8 (Interpolation between L^2 and $C^{0,\beta}$)**.**
Let $\Omega \subset \mathbb{R}^d$ be open, bounded and $v \in C^{0,\beta}(\bar{\Omega})$ with $0 < \beta < 1$, $v \in L^2(\Omega)$. Furthermore, let $v \geq 0$, and $v = 0$ on $\partial\Omega$. Then

$$\|v\|_{L^\infty(\Omega)} \leq C \, \|v\|^\theta_{L^2(\Omega)} \|v\|^{1-\theta}_{C^{0,\beta}(\bar{\Omega})} \quad \textit{with} \quad \theta = \frac{2\beta}{2\beta + d}$$

and a constant $C = C(d, \beta)$ that is independent of v and Ω.

Proof. A proof can be found in [68], Proposition 2.11. □

Theorem 3.3.9.
Let Assumption 3.3.5 hold. Then the Moreau-Yosida regularized problem (P_γ) has a solution m_γ for all $\gamma > 0$. Furthermore, $(m_\gamma)_{\gamma>0}$ is bounded in M. If, in addition, Assumption 3.3.6 (A1) holds, then all weak limit points of $(m_\gamma)_{\gamma>0}$ solve the seismic inverse problem (P).

Proof. Similar as in the proof of Theorem 3.2.5, we have to consider a convex and closed set that is bounded in M $(= M_{\mathcal{D}})$. Since $0 \in M_{\text{ad}}$, we have

$$j_\gamma(0) = \frac{1}{2} \sum_{i=1}^{n_s} J_{\text{fit},i}(u_i(0)) =: J_0 \geq 0.$$

and with $\varepsilon := \sqrt{2J_0/\alpha}$ and $B_\epsilon(0) := \{m \in M : \|m\|_M \leq \epsilon\}$, we obtain

$$j_\gamma(m) = j(m) + \gamma\phi(m) \geq \frac{\alpha}{2}\|m\|_M^2 > J_0 = j_\gamma(0) \quad \forall\, m \in M \setminus B_\varepsilon(0).$$

Thus, if a minimizer of (P_γ) exists, it can only be contained in $B_\varepsilon(0)$.
Next, we note that ϕ is convex and continuous, hence, j_γ is weakly lower semi-continuous and there exists a solution $m_\gamma \in B_\varepsilon(0)$ to the regularized problem (P_γ) by standard arguments. Now, let $\bar{m}$ be a solution to the seismic inverse problem (P). By the optimality of m_γ for (P_γ) we obtain:

$$\frac{\alpha}{2}\|m_\gamma\|_M^2 \leq j(m_\gamma) \leq j(m_\gamma) + \gamma\phi(m_\gamma) = j_\gamma(m_\gamma) \leq j_\gamma(\bar{m}) = j(\bar{m}). \tag{3.31}$$

Since $(m_\gamma)_{\gamma>0}$ is bounded in M, there exist weak limit points. Moreover, (3.31) yields that $\gamma\phi(m_\gamma)$ is uniformly bounded for all $\gamma > 0$ and, hence, $\phi(m_\gamma) \to 0$ for $\gamma \to \infty$. Now, consider a weak limit point m^* and a sequence $(\gamma_k)_{k\in\mathbb{N}}$ with $m_{\gamma_k} \rightharpoonup m^*$. Due to the compact embedding $M \hookrightarrow\hookrightarrow L^\infty(\Omega)^n$, we have $m_{\gamma_k} \to m^*$ in $L^\infty(\Omega)^n$ (and also strong convergence in $L^2(\Omega)^n$).
Let $v_{\gamma_k}^+ := [Sm_{\gamma_k} - \psi^b]^+ + [\psi^a - Sm_{\gamma_k}]^+$ denote the constraint violation. Since $\phi(m_{\gamma_k}) \to 0$ we deduce $\|v_{\gamma_k}^+\|_{L^2(\Omega)^{n_c}} \to 0$. Moreover, $(Sm_{\gamma_k})_{k\in\mathbb{N}}$ is bounded in M_c and, by Assumption 3.3.6 (A1), also bounded in $W^{1,q'}(\Omega)^{n_c}$ with $q' > d$ or, respectively, in $C^{0,\beta}(\bar{\Omega})^{n_c}$ with $0 < \beta < 1$. Hence, $v_{\gamma_k}^+$ is either bounded in $W^{1,q'}(\Omega)^{n_c}$ or in $C^{0,\beta}(\bar{\Omega})^{n_c}$. Now, an interpolation inequality between $L^2(\Omega)$ and either $W^{1,q'}(\Omega)$ (by Lemma 3.3.7) or $C^{0,\beta}(\bar{\Omega})$ (by Lemma 3.3.8) yields $\|v_{\gamma_k}^+\|_{L^\infty(\Omega)^{n_c}} \to 0$. Hence, m^* is feasible for (P).
It remains to be shown that m^* is a solution to (P). To this end, we note that the sequence of optimal function values of the penalized problem is monotonically increasing, since

$$j_{\gamma_k}(m_{\gamma_k}) \leq j_{\gamma_k}(m_{\gamma_{k+1}}) \leq j_{\gamma_{k+1}}(m_{\gamma_{k+1}}).$$

Together with (3.31) this implies that $(j_{\gamma_k}(m_{\gamma_k}))_{k\in\mathbb{N}}$ converges. By the lower semicontinuity of j, we obtain

$$j(m^*) \leq \liminf_{k\to\infty} j(m_{\gamma_k}) \leq \liminf_{k\to\infty} j_{\gamma_k}(m_{\gamma_k}) = \lim_{k\to\infty} j_{\gamma_k}(m_{\gamma_k}) \leq j(\bar{m}).$$

Due to the optimality of $\bar{m}$, all inequalities above are satisfied with equality. Hence, $j(m^*) = j(\bar{m})$ and m^* solves (P). □

In the following, m_γ always denotes a solution to (P_γ). Concerning the rate of convergence, we continue to use the notation $v_\gamma^+ = [Sm_\gamma - \psi^b]^+ + [\psi^a - Sm_\gamma]^+$ and state the estimate:

Theorem 3.3.10.
Let Assumption 3.3.5 hold and, additionally, let Assumption 3.3.6 (A1) be satisfied with the embedding $M_1 \hookrightarrow W^{1,q'}(\Omega)$. Furthermore, let $(m_{\gamma_k})_{k\in\mathbb{N}} \subset M$ be a weakly convergent subsequence with $m_{\gamma_k} \rightharpoonup m^* \in M$. Then the infeasibility of solutions is bounded by

$$\|v_{\gamma_k}^+\|_{L^2(\Omega)^{n_c}} = o\left({\gamma_k}^{-\frac{1}{2}}\right) \quad (\gamma_k \to \infty)$$

and

$$\|v_{\gamma_k}^+\|_{L^\infty(\Omega)^{n_c}} = o\left(\gamma_k^{-\eta}\right) \quad \textit{with} \quad \eta = \frac{q'-d}{q'd + 2(q'-d)}.$$

Proof. From (3.31) we obtain

$$j_{\gamma_k}(m_{\gamma_k}) = j(m_{\gamma_k}) + \frac{\gamma_k}{2}\|v_{\gamma_k}^+\|^2_{L^2(\Omega)^{n_c}} \leq j(\bar{m}),$$

where $\bar{m}$ solves (P). Hence,

$$\|v_{\gamma_k}^+\|^2_{L^2(\Omega)^{n_c}} \leq \frac{2}{\gamma_k}\left(j(\bar{m}) - j(m_{\gamma_k})\right).$$

By Theorem 3.3.9, m^* solves (P), hence, $j(m_{\gamma_k}) \to j(\bar{m})$ which shows

$$\|v_{\gamma_k}^+\|_{L^2(\Omega)^{n_c}} = o\left({\gamma_k}^{-\frac{1}{2}}\right).$$

Since $(m_{\gamma_k})_{k\in\mathbb{N}}$ is bounded in M, $(Sm_{\gamma_k})_{k\in\mathbb{N}}$ is bounded in M_c and by Assumption 3.3.6 (A1) also bounded in $W^{1,q'}(\Omega)^{n_c}$. Thus, an interpolation inequality yields the L^∞-estimate similar to Lemma 8.26 in [126]. □

Example 3.3.11.
We consider $M_1 = (H^2(\Omega) \cap H_0^1(\Omega))$. Then we have $M_1 \hookrightarrow\hookrightarrow L^\infty(\Omega)$ and $M_1 \hookrightarrow W^{1,q'}(\Omega)$ for all q' with $1 \leq q' \leq 6$ $(d = 2, 3)$. Thus, for $d = 3$ we obtain the estimate

$$\|v_{\gamma_k}^+\|_{L^\infty(\Omega)^{n_c}} = o\left(\gamma_k^{-\frac{1}{8}}\right). \tag{3.32}$$

Following the derivation in [68], we obtain alternative estimates using interpolation between $L^1(\Omega)$ and $C^{0,\beta}(\bar{\Omega})$.

Theorem 3.3.12.
Let Assumption 3.3.5 hold and, additionally, let Assumption 3.3.6 be satisfied with the embedding $M_1 \hookrightarrow C^{0,\beta}(\bar{\Omega})$ and $m = 0$ on $\partial\Omega$ for all $m \in M$. Furthermore, let $(m_{\gamma_k})_{k\in\mathbb{N}}$ be a weakly convergent subsequence with $m_{\gamma_k} \rightharpoonup m^* \in M$. Then we obtain the following estimate on the constraint violation:

$$\|v_{\gamma_k}^+\|_{L^\infty(\Omega)^{n_c}} \leq C\gamma_k^{-\eta} \quad \textit{with } \eta = \frac{\beta}{\beta + d}. \tag{3.33}$$

Proof. The proof follows from [68], Corollary 2.6, but it requires that $\gamma_k v_{\gamma_k}^+$ is uniformly bounded in $L^1(\Omega)^{n_c}$ for $\gamma_k \to \infty$. In order to show this, we use $\bar{\psi}$ and $\hat{m}$ from Assumption 3.3.6 (A2) and define $w := m_{\gamma_k} - \frac{1}{2}\hat{m}$. Now, we observe that for $i = 1, \ldots, n_c$:

$$(Sm_{\gamma_k})_i(x) \geq \psi_i^b(x) \quad \Rightarrow \quad \frac{2}{\bar{\psi}}(Sw)_i(x) = \frac{2(Sm_{\gamma_k})_i(x) - (S\hat{m})_i(x)}{\bar{\psi}} > \frac{(S\hat{m})_i(x)}{\bar{\psi}} \geq 1,$$

and

$$(Sm_{\gamma_k})_i(x) \leq \psi_i^a(x) \quad \Rightarrow \quad \frac{2}{\bar{\psi}}(Sw)_i(x) \leq -\frac{(S\hat{m})_i(x)}{\bar{\psi}} \leq -1,$$

where we used $\psi^a \leq 0$ in the second part. By testing (3.30) with w we obtain

$$\begin{aligned}
&\|\gamma_k v_{\gamma_k}^+\|_{L^1(\Omega)^{n_c}} \\
&\quad = \sum_{i=1}^{n_c} \int_\Omega \gamma_k \left(([Sm_{\gamma_k}(x) - \psi^b(x)]^+)_i + ([\psi^a(x) - Sm_{\gamma_k}(x)]^+)_i \right) dx \\
&\quad \leq \frac{2}{\bar{\psi}} \sum_{i=1}^{n_c} \int_\Omega \gamma_k \left(([Sm_{\gamma_k}(x) - \psi^b(x)]^+)_i - ([\psi^a(x) - Sm_{\gamma_k}(x)]^+)_i \right) (Sw)_i(x)\, dx \qquad (3.34) \\
&\quad = \frac{2}{\bar{\psi}} \left([Sm_{\gamma_k} - \psi^b]^+ - [\psi^a - Sm_{\gamma_k}]^+, Sw \right)_{L^2(\Omega)^{n_c}} \\
&\quad = -\frac{2}{\bar{\psi}} \langle j'(m_{\gamma_k}), w \rangle_{M^*,M}.
\end{aligned}$$

Now, with $B_\varepsilon(0)$ as defined in the proof of Theorem 3.3.9, we have $m_{\gamma_k} \in B_\varepsilon(0)$. Hence, $j'(m_{\gamma_k})$ is uniformly bounded on $B_\varepsilon(0)$ by Lemma 3.2.9, and we conclude

$$\left| -\frac{2}{\bar{\psi}} \langle j'(m_{\gamma_k}), w \rangle_{M^*,M} \right| \leq c\|w\|_M \leq c \left(\|m_{\gamma_k}\|_M + \frac{1}{2}\|\hat{m}\|_M \right) \leq C. \qquad (3.35)$$

The rest of the proof follows from Corollary 2.6 in [68]. □

Example 3.3.13.
Again, we consider $M_1 = (H^2(\Omega) \cap H_0^1(\Omega))$. Hence, we have the embedding $M_1 \hookrightarrow C^{0,\beta}(\Omega)$ with $\beta = \frac{1}{2}$ for $d = 2, 3$, and for $d = 3$ we get the estimate

$$\|v_{\gamma_k}^+\|_{L^\infty(\Omega)^n} \leq C\gamma_k^{-\frac{1}{7}}. \qquad (3.36)$$

Assumption 3.3.6 (A1) in conjunction with the interpolation inequalities can also be used to weaken the requirement $M_\mathcal{D} = M$ in Assumption 3.3.5. Here, the following alternative result can be stated:

Theorem 3.3.14.
Let Assumption 3.3.6 (A1) and Assumption 3.3.5 but not necessarily $\mathcal{D} = L^\infty(\Omega)^n$ hold. Furthermore, let $\mathcal{G}^\infty \subset \mathcal{D}$ be a closed, convex set containing a neighborhood of M_{ad}^∞. Then, there exists $\gamma_0 > 0$ such that the Moreau-Yosida regularized problem (P_γ) has a solution m_γ for all $\gamma > \gamma_0$. Furthermore, $(m_\gamma)_{\gamma \geq \gamma_0}$ is bounded in M and all weak limit points of $(m_\gamma)_{\gamma \geq \gamma_0}$ solve the seismic inverse problem (P).

Proof. Studying the proof of Theorem 3.3.9 reveals that we only have to show that there exists a closed, convex and bounded set $\mathcal{G} \subset M_\mathcal{D}$ with $\hat{m} \in \mathcal{G}$ and $\gamma_0 > 0$ such that for all $\gamma \geq \gamma_0$:

$$j_\gamma(m) > j_\gamma(\hat{m}) \quad \forall\, m \in M_\mathcal{D} \setminus \mathcal{G}. \qquad (3.37)$$

This will ensure the existence of a solution m_γ to (P_γ) in $\mathcal{G}$ for all $\gamma \geq \gamma_0$ as well as the boundedness of $(m_\gamma)_{\gamma \geq \gamma_0}$ in M.
First, we note that if we choose $\hat{m} \in M_{\text{ad}}$, we only have to show (3.37) for $\gamma = \gamma_0$ since this then implies for all $\gamma \geq \gamma_0$:

$$j_\gamma(m) \geq j_{\gamma_0}(m) > j_{\gamma_0}(\hat{m}) = j(\hat{m}) = j_\gamma(\hat{m}) \quad \forall\, m \in M_\mathcal{D} \setminus \mathcal{G}.$$

Now, we set $\mathcal{G} := \mathcal{G}^\infty \cap B_\varepsilon(0)$ with $B_\varepsilon(0)$ as defined previously. Next, we prove that there exists $\gamma_0 > 0$ such that $j_{\gamma_0}(m) > j_{\gamma_0}(0) = J_0$ for all $m \in M_\mathcal{D} \setminus \mathcal{G}$. If this does not hold, then there exist sequences $(\gamma_k)_{k\in\mathbb{N}}$ with $\gamma_k \to \infty$ and $(m_k)_{k\in\mathbb{N}} \subset B_\epsilon(0) \setminus \mathcal{G}^\infty$ with

$$\frac{\alpha}{2}\|m_k\|_M^2 + \gamma_k\phi(m_k) \leq j_{\gamma_k}(m_k) \leq J_0 \quad \forall\, k \in \mathbb{N}, \tag{3.38}$$

where we already used the fact that $j_{\gamma_k}(m) \geq \frac{\alpha}{2}\|m\|_M^2 > J_0$ for $m \notin B_\epsilon(0)$. Thus, $(m_k)_{k\in\mathbb{N}}$ is bounded in M. Furthermore, (3.38) yields $\phi(m_k) \to 0$ and for $v_k^+ := [Sm_k - \psi^b]^+ + [\psi^a - Sm_k]^+$, we deduce $\left\|v_k^+\right\|_{L^2(\Omega)^{n_c}} \to 0$. Similar as before, we can now utilize an interpolation inequality to deduce $\left\|v_k^+\right\|_{L^\infty(\Omega)^{n_c}} \to 0$. Thus, $m_k \in G^\infty$ for k sufficiently large. This is a contradiction. □

Summing up the previous results, we need to solve a sequence of penalized problems (P_γ) with increasing penalty parameters γ in order to compute a solution to (P) . Algorithm 1 describes this outer loop for solving (3.3) by the Moreau-Yosida penalty method.

Algorithm 1 Penalty Method

1: Choose $\gamma_0 > 0$, an initial model m_{init} and $\epsilon > 0$.
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: Solve (P_γ) using initial model m_{init} to a specified tolerance and obtain solution m_{γ_k}.
4: **if** $\left\|v_{\gamma_k}^+\right\|_{L^2(\Omega)^{n_c}}^2 < \epsilon$ **then**
5: Stop with $\bar{m} = m_{\gamma_k}$.
6: **else**
7: Choose $\gamma_{k+1} > \gamma_k$ and set initial model $m_{\text{init}} = m_{\gamma_k}$.
8: **end if**
9: **end for**

Note that in our numerical experiments, we update the penalty parameter quite aggressively instead of solving (P_γ) for a fixed γ to a high accuracy. In particular, γ is increased when the current iterate is infeasible with some specified tolerances on the L^2- and L^∞-norm and when the last steps provided a good progress towards optimality of (P_γ). Here, we usually choose a reduction of the norm of the gradient by half an order of magnitude as criterion. This works well for the type of constrained problems we are dealing with in section 4.3 and Chapter 5. We emphasize, however, that more sophisticated strategies on updating γ exist [70]. This would be an interesting field for future research.

Remark 3.3.15.
It should be emphasized that due to the non-convexity of the problem, we cannot expect to attain global solutions of (P_γ). The same holds true for (P). Here, we have to rely on a good starting point and a suitable regularization parameter.

3.3.2 Optimality Conditions for the Regularized Problem

We recall the optimality conditions for the penalized problem given by 3.29:

$$j'(\bar{m}_\gamma) + \gamma S^*\left([S\bar{m}_\gamma - \psi^b]^+ - [\psi^a - S\bar{m}_\gamma]^+\right) = 0 \quad \text{in } M^*. \tag{3.39}$$

Due to the maximum operator in $[.]^+$, the operator equation (3.39) is not differentiable. To overcome this difficulty, the concept of generalized derivatives and semismooth operator equations can be applied. We work with the following definition, cf. [73], Definition 2.1.

Definition 3.3.16 (Semismoothness)**.**
Let X, Y be Banach spaces and $G : X \to Y$ continuous. Furthermore, consider a set-valued mapping $\partial G : X \rightrightarrows \mathcal{L}(X, Y)$. If

$$\sup_{H \in \partial G(x+s)} \|G(x+s) - G(x) - Hs\|_Y = o(\|s\|_X) \quad \text{for} \quad \|s\|_X \to 0,$$

then G is called semismooth with generalized derivative ∂G (sometimes also called ∂G-semismooth).

The classical choice of the generalized derivative for the operator $[.]^+$ is given in the next Lemma.

Lemma 3.3.17.
Let Assumption 3.3.5 hold. Then the optimality condition (3.29) *is a semismooth operator equation, which can be stated as* $G(m) = 0$ *with*

$$G : M \to M^*, \quad G(m) := j'(\bar{m}) + \gamma S^* \left([S\bar{m} - \psi^b]^+ - [\psi^a - S\bar{m}]^+ \right),$$

and with a generalized derivative given by $j''(\bar{m}) + \gamma S^* \partial D(\bar{m}) S$ *and* $\partial D(\bar{m})$ *defined by*

$$\partial D_i(\bar{m})(x) \begin{cases} = & 0 & \psi_i^a(x) < (S\bar{m})_i(x) < \psi_i^b(x), \\ = & 1 & (S\bar{m})_i(x) < \psi_i^a(x) \text{ or } (S\bar{m})_i(x) > \psi_i^b(x), \\ \in & [0,1] & (S\bar{m})_i(x) = \psi_i^a(x) \text{ or } (S\bar{m})_i(x) = \psi_i^b(x), \end{cases} \tag{3.40}$$

for $i = 1, \ldots, n_c$.

Proof. We only have to consider the second part since $j'(\bar{m})$ is smooth. By definition, we have $S : M \to M_c \hookrightarrow L^\infty(\Omega)^{n_c}$. Furthermore, $[.]^+$ is semismooth from $L^q(\Omega)^{n_c}$ to $L^2(\Omega)^{n_c}$ for any $q > 2$ with a generalized derivative given by ∂D (cf. Proposition 4.1 in [67] with a straightforward extension to $n_c > 1$). Since $S^* \in \mathcal{L}(M_c{}^*, M^*)$ and $L^2(\Omega)^{n_c} \hookrightarrow M_c{}^*$, we deduce that (3.29) is semismooth. □

We conclude this section with a remark on possible extensions of the problem setting.

Remark 3.3.18.
It would be possible to work with different spaces M_i for every component of the parameters as long as every one of them is compactly embedded into $L^\infty(\Omega)$ and, furthermore, every M_i is embedded into $W^{1,q'}(\Omega)$ or $C^{0,\beta}(\bar{\Omega})$ for applying the interpolation inequalities.

3.4 Trust-Region Newton Method

Now we turn to the discussion of obtaining points satisfying the necessary first-order optimality conditions (3.29). Following the analysis in the previous sections, we will only consider the case $M_{\mathcal{D}} = M$. Note that in the absence of additional constraints, the optimality conditions

are given by $j'(\bar{m}) = 0$, which can be treated as a special case of (3.29). With the adjoint-based representation of the first and second derivatives as well as the generalized derivative that comes from the penalty term, we have everything at hand that is required to apply a semismooth Newton method to this nonlinear operator equation. The general iterative scheme - yet without a globalization strategy - is outlined in Algorithm 2.

Algorithm 2 Semismooth Newton Method

1: Choose $m^0 \in M$.
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: Choose $H_k \in \partial G$.
4: Obtain s^k by solving $H_k s^k = -G(m^k)$.
5: Update $m^{k+1} = m^k + s^k$.
6: **end for**

A superlinear rate of convergence of the semismooth Newton method can be established, provided that m^0 is chosen sufficiently close to a solution $\bar{m}$ and with the textbook regularity condition

$$\|H^{-1}\|_{\mathcal{L}(M^*,M)} \leq C \quad \forall\, H \in \partial G(m) \quad \forall\, m \in B_\delta(\bar{m}), \tag{3.41}$$

with constants $C, \delta > 0$, cf. Theorem 2.12 in [73]. However, this regularity condition (3.41) is hard to verify in practice.

Remark 3.4.1.
It is important to note that the superlinear convergence rate can be maintained if the generalized Newton system is solved inexactly. To this end, we might interpret the inexact solution as the exact solution to a system with a perturbed operator $\tilde{H}_k$ instead of H_k, cf. Algorithm 3.16 and Theorem 3.18 in [126]. Hereby, we require the Dennis-Moré conditions to be satisfied for the perturbed operator, see Assumption 3.14 in [126] and [42].

To ensure global convergence, we propose a trust-region scheme. This method is widely used for globalization of Newton-type methods in finite and infinite dimensions. For a detailed analysis, we refer to [37, 126] and limit the forthcoming presentation to the general ideas.
We consider the quadratic model

$$\begin{aligned} \min_{s \in M} \quad & q_k(s) := \left\langle j'_\gamma(m^k), s \right\rangle_{M^*,M} + \tfrac{1}{2} \left\langle (j''(m^k) + \gamma S^* \partial D(m^k) S)s, s \right\rangle_{M^*,M} \\ & \text{s.t. } \|s\|_M \leq \Delta_k. \end{aligned} \tag{3.42}$$

Here m^k denotes the current iterate and Δ_k the trust-region radius in iteration k. $\partial D(m^k)$ is a generalized derivative as defined in Lemma 3.3.17. Hereby, the first derivatives $j'(m^k)$ and operator-vector products $j''(m^k)s$ are computed using the adjoint representation outlined in section 3.2.2. Note that $j''(m^k) + \gamma S^* \partial D(m^k) S$ in the second term of q_k can be replaced by suitable approximations H_k, for instance, quasi-Newton approximations.
The update of the trust-region radius Δ_k and the acceptance of a computed step is controlled by the ratio of actual reduction, i.e., the decrease of j_γ, and predicted ratio, i.e., the decrease of the quadratic model function q_k. Therefore, we define

$$\text{ared}_k(s) := j_\gamma(m_k) - j_\gamma(m_k + s) \quad \text{and} \quad \text{pred}_k(s) := -q_k(s), \tag{3.43}$$

as well as the reduction ratio $\varrho_k = \varrho_k(s^k)$ with

$$\varrho_k(s) := \frac{\mathrm{ared}_k(s)}{\mathrm{pred}_k(s)}. \tag{3.44}$$

More details on the specific implementation of steering the trust-region radius and deciding whether to accept or reject step s^k will be given in the next section.
Proofs for global convergence and the transition to fast local convergence, i.e., a superlinear rate of convergence, can be found in [126], chapter 7, in great generality and for different variants of the trust-region algorithm. Here, we require that the norms of the approximated Hessians H_k are uniformly bounded, i.e., there exists a constant C_H such that $\|H_k\|_{\mathcal{L}(M,M^*)} \leq C_H$ for all k.
Approximate solutions to the trust-region subproblem can be efficiently computed by the Steihaug-CG method [116], which will be explained in the next section.

3.5 Discretized Problem

In this section, we highlight some important considerations to solve the discretized problem. Let M_h denote the discretized finite element space for M equipped with a nodal basis such that elements of $m^h \in M_h$ can be represented as coefficient vectors $\mathbf{m} \in \mathbb{R}^{n_m}$. Furthermore, let $\mathbf{M} \in \mathbb{R}^{n_m \times n_m}$ denote the symmetric and positive definite matrix that is associated with the inner product, i.e., the inner product of two coefficient vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^{n_m}$ is given by

$$(\mathbf{v}, \mathbf{w})_{\mathbf{M}} := \mathbf{v}^T \mathbf{M} \mathbf{w}, \tag{3.45}$$

and, likewise, the corresponding norm is denoted by

$$\|\mathbf{v}\|_{\mathbf{M}} = \left(\mathbf{v}^T \mathbf{M} \mathbf{v}\right)^{\frac{1}{2}}. \tag{3.46}$$

The specific finite element spaces for the discretization of the material parameters will be given in section 4.1. Since the focus of this section is the finite-dimensional character of the trust-region subproblem, we continue to work directly with the reduced problem and postpone details on the discretization of the state equation. For the moment, it is sufficient to consider $E^h : \mathbb{R}^{n_u} \times \mathbb{R}^{n_m} \to \mathbb{R}^{n_u}$ as the discretized version of the state equation. Furthermore, let $j^h : \mathbb{R}^{n_m} \to \mathbb{R}$ denote the discretized reduced cost functional and consider the discrete version of (P_γ)

$$\min_{\mathbf{m} \in \mathbb{R}^{n_m}} \; j_\gamma^h(\mathbf{m}) \tag{P_γ^h}$$

Now, we obtain the discretized trust-region subproblem

$$\begin{aligned} \min_{\mathbf{s} \in \mathbb{R}^{n_m}} \quad & q_k^h(\mathbf{s}) := \nabla j_\gamma^h(\mathbf{m}^k)^T \mathbf{s} + \tfrac{1}{2} \mathbf{s}^T \mathbf{H}_k \mathbf{s} \\ & \text{s.t. } \|\mathbf{s}\|_{\mathbf{M}} \leq \Delta_k, \end{aligned} \tag{3.47}$$

where $\mathbf{H}_k$ is an approximation of the generalized derivative of $j_\gamma^{h'}(m^h)$ in Euclidean coordinates. For a proper scaling, we have to formulate the trust-region subproblem in terms of coefficient vectors and matrices. Let $\mathbf{g}^k := \mathbf{M}^{-1} \nabla j_\gamma^h(\mathbf{m}^k)$ denote the transformation of the discrete gradient to the M_h inner product. Similarly, the Hessian approximation $\mathbf{H}_k$ has to be

transformed by $\tilde{\mathbf{H}}_k = \mathbf{M}^{-1}\mathbf{H}_k$. We compute an approximate solution $\mathbf{s}^k$ to (3.47) by the Steihaug preconditioned conjugate gradient method [116] that is outlined in Algorithm 3. Here, $\mathbf{M}^{-1}$ is used as a preconditioner, which might be interpreted as performing the conjugate gradient iterations in M_h. It is important to note that every CG iteration requires to solve two additional discretized state equations per seismic source to compute a Hessian-vector product, see section 3.2.2. This vastly dominates the computational costs of Algorithm 3. In particular, although in all numerical examples the matrix $\mathbf{M}$ will not be diagonal, the costs of computing $\mathbf{M}^{-1}\mathbf{v}$ are negligible compared to solving a single state equation.
Algorithm 3 terminates with one of the following cases:

1. negative curvature is encountered (line 8)
2. the next iterate would be outside the trust-region (line 13)
3. the residual is below the specified tolerance (line 17)
4. a maximum number of CG iterations is reached.

Algorithm 3 Trust-region Newton-CG

1: **Input:** $\mathbf{m}^k$, $\mathbf{H}_k$ to set up model q_k^h.
2: Choose $\epsilon_k > 0$, $\max_{\text{cg}} \in \mathbb{N}$.
3: Set $i = 0$, $\mathbf{p}^0 = 0$, $\mathbf{r}^0 = -\nabla j_\gamma^h(\mathbf{m}^k)$, $\mathbf{z}^0 = \mathbf{M}^{-1}\mathbf{r}^0$, $\mathbf{d}^0 = \mathbf{z}^0$.
4: **for** $i = 0, 1, 2, \ldots, \max_{\text{cg}}$ **do**
5: Set $\eta = (\mathbf{d}^i, \mathbf{H}_k\mathbf{d}^i)$.
6: **if** $\eta \leq 0$ **then**
7: Find $\tau > 0$ such that $\|\mathbf{p}^i + \tau\mathbf{d}^i\|_{\mathbf{M}} = \Delta_k$.
8: **return** $\mathbf{s}^k = \mathbf{p}^i + \tau\mathbf{d}^i$.
9: **end if**
10: Set $\alpha = (\mathbf{r}_i, \mathbf{z}_i)/\eta$, $\mathbf{p}^{i+1} = \mathbf{p}^i + \alpha\mathbf{d}^i$
11: **if** $\|\mathbf{p}^{i+1}\|_{\mathbf{M}} \geq \Delta_k$ **then**
12: Find τ such that $\|\mathbf{p}^i + \tau\mathbf{d}^i\|_{\mathbf{M}} = \Delta_k$.
13: **return** $\mathbf{s}^k = \mathbf{p}^i + \tau\mathbf{d}^i$.
14: **end if**
15: Set $\mathbf{r}^{i+1} = \mathbf{r}^i - \alpha\mathbf{H}_k\mathbf{d}^i$.
16: **if** $\|\mathbf{r}_{i+1}\|_{\mathbf{M}}/\|\mathbf{r}_i\|_{\mathbf{M}} < \varepsilon_k$ **or** $i == \max_{cg}$ **then**
17: **return** $\mathbf{s}^k = \mathbf{p}^{i+1}$.
18: **end if**
19: Set $\mathbf{z}^{i+1} = \mathbf{M}^{-1}\mathbf{r}^{i+1}$, $\beta = (\mathbf{r}^{i+1}, \mathbf{z}^{i+1})/(\mathbf{r}^i, \mathbf{z}^i)$, $\mathbf{d}^{i+1} = \mathbf{z}^{i+1} + \beta\mathbf{d}^i$.
20: **end for**

In [116], Theorem 2.1, it is shown that $q_k^h(\mathbf{p}^i)$ is strictly decreasing. Furthermore, it can be shown that $\|\mathbf{p}^i\|_{\mathbf{M}}$ is increasing and thus, the second termination criterion is reasonable. Since the problems we consider are non-convex, the Hessian might be indefinite, thus an early termination of the CG method is crucial, when negative curvature is encountered. In any case, Algorithm 3 terminates with a descent direction $\mathbf{s}^k$. As has been pointed out in the previous section, the decision, whether $\mathbf{s}^k$ is accepted for updating $\mathbf{m}^k$ or not, depends on the ratio of actual and predicted reduction ϱ_k. This procedure is described in Algorithm 4.

Unless explicitly stated otherwise, all numerical examples use the following configuration in Algorithm 4: $\Delta_{\min} = 10^{-12}$, $\eta_0 = 0.001$, $\eta_1 = 0.25$, $\eta_2 = 0.75$ and $\tau = 0.9$. In particular, we emphasize that the trust-region radius is only increased, if the current step is sufficiently close to the trust-region boundary which proved advantageous in our numerical experiments.

Algorithm 4 Update of the Trust-region Radius and Acceptance of Steps

Input: trial step $\mathbf{s}^k$.
Compute ϱ_k using (3.44).
Choose $\Delta_{\min} \geq 0$, $0 < \eta_0 \leq \eta_1 < \eta_2 < 1$, $0 < \beta_0 < 1 < \beta_1$ and $\tau \in (0, 1)$
if $\varrho_k \leq \eta_1$ **then**
  Set $\Delta_{k+1} = \max\{\beta_0 \Delta_k, \Delta_{\min}\}$.
else if $\varrho_k > \eta_2$ **and** $\|s^k\|_M > \tau \Delta_k$ **then**
  Set $\Delta_{k+1} = \beta_1 \Delta_k$.
else
  Set $\Delta_{k+1} = \Delta_k$.
end if
if $\varrho_k > \eta_0$ **then**
  Accept step $\mathbf{s}^k$.
else
  Reject step $\mathbf{s}^k$.
end if

Since the Steihaug-CG method requires many simulations of the state equation, it is highly desirable to avoid rejecting steps without reusing the information that has already been computed previously. To this end, we propose two strategies to cope with this situation. Both methods require additional solutions to the state equation only to compute the objective function (either for computing the actual reduction or checking Armijo's condition).

In the first case, we assume that all previously computed steps can be stored in memory. If Algorithm 4 returns with the rejection of the step, we successively reduce the trust-region radius, reuse the Steihaug-CG steps to compute a new trial step and invoke Algorithm 4 again. Alternatively, we can continue to use the computed direction $\mathbf{s}^k$ but perform backtracking steps to find a point that satisfies Armijo's condition and update the trust-region radius thereafter to the norm of the update. This is illustrated in Algorithm 5. Note that in all numerical examples, we will use $\gamma_A = 0.001$.

Algorithm 5 Handling Rejected Steps

Input: rejected step $\mathbf{s}^k$ from Steihaug-CG method, $\gamma_A \in (0, 1)$
Set $\sigma_A = 1$.
while $j_\gamma^h(\mathbf{m}^k + \sigma_A \mathbf{s}^k) - j_\gamma^h(\mathbf{m}^k) > \gamma_A \sigma_A (\mathbf{g}^k, \mathbf{s}^k)_{\mathbf{M}}$ **do**
  $\sigma_A = \sigma_A / 2$.
end while
Set $\Delta_{k+1} = \sigma_A \|\mathbf{s}^k\|_{\mathbf{M}}$.
return $\mathbf{s}^k = \sigma_A \mathbf{s}^k$.

3.6 Discussion and Further Perspectives

In this chapter, we presented the seismic inverse problem and suitable solution strategies. Restrictions on the admissible material have been introduced in two steps. Initially, we utilized a superposition operator to guarantee that only "reasonable" material enters the state equation, i.e., coefficients for which the elliptic operator is uniformly coercive and the theory on existence and uniqueness of solutions can be applied. In a second step, we introduced explicit constraints on the material parameters and treated them by the Moreau-Yosida regularization. This offers a powerful framework to solve the constrained parameter identification problem. We want so emphasize the regularizing character of the penalty method. For many problems in seismic tomography, the constraints are usually not expected to be active in the global minimum. Nevertheless, constraints can add prior knowledge to the problem formulation and prevent convergence towards "physically unreasonable" local minima. This can be viewed as a robustification of the inversion process. On the other hand, this means that we do not necessarily have to drive the penalty parameter to infinity and the penalty term is only invoked if infeasibilities are encountered during the inversion process.
On a different note, the analysis is not limited to the elastic or the acoustic wave equation and the results can be applied to other problems in optimal control governed by linear hyperbolic equations as well.

We conclude this chapter with some remarks on possible extensions. As mentioned before, this thesis focuses on tomography problems that seek to determine structural parameters within a domain of interest. In fact, however, for earthquake data neither the source time function nor the hypocenter (especially, the depth) is known exactly. Typically, the inversion for source parameters is done in a preprocessing step using simpler physical models that are mostly based on traveltime tomography [115]. Nevertheless, it would be interesting to pursue an all-at-once approach that simultaneously inverts for source and structural parameters. It will be challenging, however, to establish a suitable parameterization of the source and appropriate regularization, because the seismic sources typically only have a small spatial support.
Regarding the minimization procedure, it would be desirable to work with more sophisticated preconditioners, especially for the misfit term. A possible approach would reuse the gradients from the previous iterations and set up an L-BFGS preconditioner [98]. Similar ideas have been applied to tomography problems in [48]. In addition, there is also related work regarding efficient preconditioners for Moreau-Yosida regularized problems, see [102].

Chapter 4

Implementation and Numerical Results

In this chapter, we discuss the discretization of the inverse problem and the implementation of a matrix-free MPI-parallelized code for simulating seismic wave propagation and solving tomography problems. In the second part, we consider examples for problems on reservoir scale in 2d and 3d. In particular, we discuss adjoint sources, statistics on the parallel performance of the code and, most importantly, the solution of seismic inverse problems. Parts of this chapter have been published in [19]. Note that further numerical results follow in Chapter 5 and Chapter 6.

4.1 Discretization of the Wave Equation

We apply a continuous high-order finite element method for the spatial discretization of the state and an explicit time-stepping scheme. This approach is commonly used in seismic applications, cf. [51, 103, 123], see also [133] for the analysis of a high-order discontinuous Galerkin method.
We use different spatial meshes for the discretization of the state and the material parameters. This is motivated by the fact that the data contains only a limited amount of information and thus a coarser mesh in the parameter space prevents an over-parameterization. Additionally, the parameter mesh might be adaptively refined based on goal-oriented error estimates [14, 18] or prior knowledge to acknowledge the varying amount of information in the data for different regions of the domain. Note that using different grids for the state and the parameters requires interpolating the parameter values onto the finer state mesh before every simulation.

Spatial Discretization of the State Space

We consider shape-regular meshes consisting of quadrilateral or, respectively, hexahedral cells K that cover the computational domain Ω. Let $\mathbb{T}_h = \{K\}$ denote the finite element mesh and h the discretization parameter. Furthermore, let Q_s denote the space of polynomials of degree s in each variable x_i, $i = 1, \dots, d$, defined on the reference cell $K_{\text{ref}} = [-1, 1]^d$. We use the Lagrange polynomials of degree s with the collocation points given by the Gauss-Lobatto-Legendre (GLL) quadrature rule [74] as basis of Q_s. This yields a nodal basis for

the numerical representation of the elements of Q_s. To this end, let ξ_i, $i = 0, \dots, s$, denote the collocation points of the GLL rule on the interval $[-1, 1]$. Furthermore, let l_i denote the Lagrange polynomials associated with the points ξ_i, i.e.,

$$l_i(\xi) = \prod_{j=0, j\neq i}^{s} \frac{\xi - \xi_j}{\xi_i - \xi_j}.$$

We obtain the polynomial basis on the reference cell by tensorization of the 1d bases, i.e., for a multi-index $\iota \in \{0, \dots, s\}^d$ we define

$$\varphi^\iota : K_{\text{ref}} \to \mathbb{R}, \quad \varphi^\iota(x) := \prod_{i=1}^{d} l_{\iota_i}(x_i).$$

By definition, the Lagrange polynomials vanish at all but one of the collocation points. Hence, the Lagrange interpolant of a function $v : K_{\text{ref}} \to \mathbb{R}$ is given by

$$\mathcal{I}v(x) = \sum_{\iota \in \{0,\dots,s\}^d} v(\xi^\iota)\varphi^\iota(x) \quad \text{with} \quad \xi^\iota = (\xi_{\iota_1}, \dots, \xi_{\iota_d})^T.$$

Integrals over the reference cell are approximated by the GLL quadrature rule. To this end, let w_i, $i = 0, \dots, s$ denote the quadrature weights associated with ξ_i in 1d. We obtain

$$\int_{K_{\text{ref}}} v(x)\, dx \quad \approx \sum_{\iota \in \{0,\dots,s\}^d} w^\iota v(\xi^\iota) \quad \text{with weights} \quad w^\iota = \prod_{i=1}^{d} w_{\iota_i},$$

which is exact if $v \in Q_{2s-1}$.
Now, we introduce the finite element subspaces $V_h^s \subset V$ by

$$V_h^s = \left\{ v^h \in C(\bar{\Omega})^{\hat{d}} \;\middle|\; v_h|_K \in \hat{Q}_s(K)^{\hat{d}} \;\; \forall\, K \in \mathbb{T}_h \right\},$$

where in every component, $\hat{Q}_s$ is obtained by bi- or trilinear transformations of the nodal basis defined on the reference cell. We recall that $\hat{d} = d$ for the elastic wave equation and $\hat{d} = 1$ in the acoustic case. In all numerical tests presented in this thesis we use $s = 4$.
Detailed derivations of the spatial discretization of the elastic wave equation using this particular choice of test functions and quadrature rule can be found multiple times in the literature, see, for instance, [34, 54, 79, 123]. Therefore, we just summarize the results.
By replacing V by V_h^s in (2.61) or (2.66), respectively, we obtain the Galerkin approximation for the polynomial basis and compute the integrals with the GLL quadrature rule. With $N := \dim(V_h^s)$ and a time-dependent coefficient vector $\mathbf{u}(t) \in C^2(\bar{I})^N$, the spatially semi-discrete formulation of the wave equation is a system of linear ordinary differential equations which can be written in the following form:

$$\bar{\mathbf{M}}\mathbf{u}_{tt}(t) + \bar{\mathbf{K}}\mathbf{u}(t) = \bar{\mathbf{F}}(t). \tag{4.1}$$

Here, $\bar{\mathbf{M}} \in \mathbb{R}^{N\times N}$ denotes the mass matrix (weighted by the density to include ρ), $\bar{\mathbf{K}} \in \mathbb{R}^{N\times N}$ is the stiffness matrix and $\bar{\mathbf{F}}(t) \in \mathbb{R}^N$ is the semi-discrete force vector. Most importantly, the quadrature rule in combination with the interpolation nodes of the Lagrange polynomials yields a diagonal matrix $\bar{\mathbf{M}}$, which enables an explicit time-stepping scheme. On the other hand, this introduces an integration error, because the GLL quadrature rule is only exact for polynomials up to degree $2s - 1$.

Temporal Discretization of the State Space

Next, we turn to the temporal discretization of the state equation. Similar to [103], we apply an explicit Newmark time-stepping scheme to solve the semi-discrete system (4.1). Let

$$0 = t_0 < t_1 < \ldots < t_{n_t} = T$$

be a partition of the interval $\bar{I}$ and let $\Delta t_k = t_k - t_{k-1},\ k = 1, \ldots, n_t$ denote the time increment. Now, we introduce a set of independent variables $\mathbf{u}_{k,0}, \mathbf{u}_{k,1}, \mathbf{u}_{k,2}$ to approximate $\mathbf{u}(t_k), \mathbf{u}_t(t_k)$ and $\mathbf{u}_{tt}(t_k)$, respectively. Furthermore, let $\mathbf{F}_k$ denote the time-discrete version of $\mathbf{F}(t_k)$.
The family of Newmark schemes for the fully discrete system is given by the update formulas

$$\begin{aligned}
\mathbf{u}_{k+1,2} &= -\mathbf{M}^{-1}\left(\mathbf{K}\mathbf{u}_{k+1,0} - \mathbf{F}_{k+1}\right), \\
\mathbf{u}_{k+1,0} &= \mathbf{u}_{k,0} + \Delta t_k \mathbf{u}_{k,1} + \frac{1}{2}\Delta t_k^2 \left((1-2\beta)\mathbf{u}_{k,2} + 2\beta \mathbf{u}_{k+1,2}\right), \\
\mathbf{u}_{k+1,1} &= \mathbf{u}_{k,1} + \Delta t_k \left((1-\theta)\mathbf{u}_{k,2} + \theta \mathbf{u}_{k+1,2}\right).
\end{aligned} \tag{4.2}$$

This scheme is second-order accurate and conditionally stable for $\theta = 1/2$, see [74], Chapter 9. Furthermore, for $\beta = 0$, we obtain an explicit time stepping scheme. Note that $\bar{\mathbf{M}}$ and $\bar{\mathbf{K}}$ are time-invariant since the material parameters do not depend on time and we do not change the state mesh during the simulation. While $\bar{\mathbf{M}}$ is diagonal and can easily be hold in memory, matrix-vector products $\bar{\mathbf{K}}\mathbf{u}_{k+1,0}$ are computed on the fly without assembling the matrix $\bar{\mathbf{K}}$.
A severe drawback of explicit time-stepping schemes is the limitation of the step-size by the Courant-Friedrichs-Lewy (CFL) condition:

$$\Delta t \leq C_{\mathrm{CFL}} \frac{h}{v_p}. \tag{4.3}$$

Here, h denotes the spatial mesh width and v_p the wave speed. Loosely speaking, the CFL condition states that the time step must be chosen smaller than the time it takes the wave to travel through the cell. When using a global time step, we have to consider the minimum ratio over all cells in (4.3). Hence, it is desirable to choose the local mesh size according to the local wave velocity.
Despite this limitation, parallelization can be carried out much easier with the explicit scheme as we do not have to solve a linear system in every time step due to the diagonal mass matrix. The explicit Newmark time stepping scheme outlined in (4.2) with $\beta = 0$ and $\theta = 1/2$ is widely used for numerical simulations of seismic wave propagation [34, 54, 79]. Note, however, that other (explicit) time stepping schemes can be used as well. For instance, a five-stage fourth-order low-storage Runge-Kutta method is applied in [133].

Spatial Discretization of the Parameter Space

As previously outlined, we separate the discretization of the state and the parameter and treat both meshes independently. For the parameter mesh, we use a continuous Galerkin finite element discretization and introduce the finite element subspace

$$M_h^s = \left\{ m^h \in C(\bar{\Omega})^n \;\middle|\; m_h|_K \in \hat{Q}_s(K)^n \ \ \forall\ K \in \mathbb{T}_h^M \right\}.$$

Here, $\mathbb{T}_h^M$ denotes the decomposition of $\bar{\Omega}$ for the parameter space. In general, $\mathbb{T}_h^M$ consists of larger cells than $\mathbb{T}_h$, otherwise the small-scale heterogeneities in the material will not be seen by the wavefield. In all numerical tests we will use a polynomial degree of $s = 1$, i.e., bi- or trilinear elements. Note that instead of using the GLL quadrature rule for the parameter space we apply an exact integration rule, which results in a non-diagonal mass matrix $\mathbf{M}$.

Remark 4.1.1.
We recall the assumptions of Chapter 3, where we require $M \hookrightarrow\hookrightarrow L^\infty(\Omega)$. In order to justify the choice of bi- or trilinear elements, we point out the regularizing effect of the discretization and the equivalence of all norms for the finite dimensional problem. Furthermore, the numerical results neither show oscillating solutions nor undesirable artifacts in the reconstruction, which justifies the choice of the discretization of the parameter space.

4.2 Implementation

Efficient inversion methods rely on a scalable code for the simulation of the elastic-acoustic wave equation. As part of this thesis, a software package written in C++ for seismic inversion has been developed and we briefly highlight some of its features. The implementation aims to offer a flexible framework to simulate the propagation of seismic waves in solid and fluid media and to solve related tomography problems. There are three main components: a container to store seismic sources and observed data, a wave simulation code and optimization routines to solve the inverse problem. Figure 4.1 shows the individual parts and how interaction is intended on the different levels.
Seismic data is currently assumed to stem from point-sources, which are accurate enough to model seismic waves in the far field, and observations in form of seismograms at a set of receiver locations.
The code has been developed to simulate the acoustic and the elastic wave equation in two and three dimensions. We also consider the case of wave propagation in a coupled system of fluid and acoustic media in 3d, see Chapter 6. The implementation is parallelized, utilizes MPI-communication and works matrix-free. Due to the similar discretization scheme, the implementation is inspired by the SPECFEM3D code [13, 103]. There are, however, significant differences concerning an object-oriented structure of the code as well as functional extensions to efficiently compute the gradient and Hessian-vector products. We utilize the Epetra package, which is part of the Trilinos library [63, 64], to handle the distributed data structures, in particular, `Epetra_FEVector`, `Epetra_Map` and `Epetra_Export`.
Parallelization is carried out in two stages. Trivially, different seismic events can be simulated in parallel and communication is only required during a post-processing step to add up the individual contributions to the cost functional and its derivatives. Moreover, the implementation allows to solve a single event on multiple cores using a spatial partitioning of the computational domain and communication with MPI. To this end, we employ two layers of MPI communicators, one acting globally to run the inversion and a local one to manage the simulation of the wave equation for a single event. On the global layer, the material has to be distributed across all processors and, in reverse direction, the individual misfits and gradients have to be accumulated. On the second layer, every seismic event has its own MPI communicator. Here, every processor needs to pass the wavefield at the boundaries of the local domain to its neighbors in every time step. The second group of communicators is derived from the

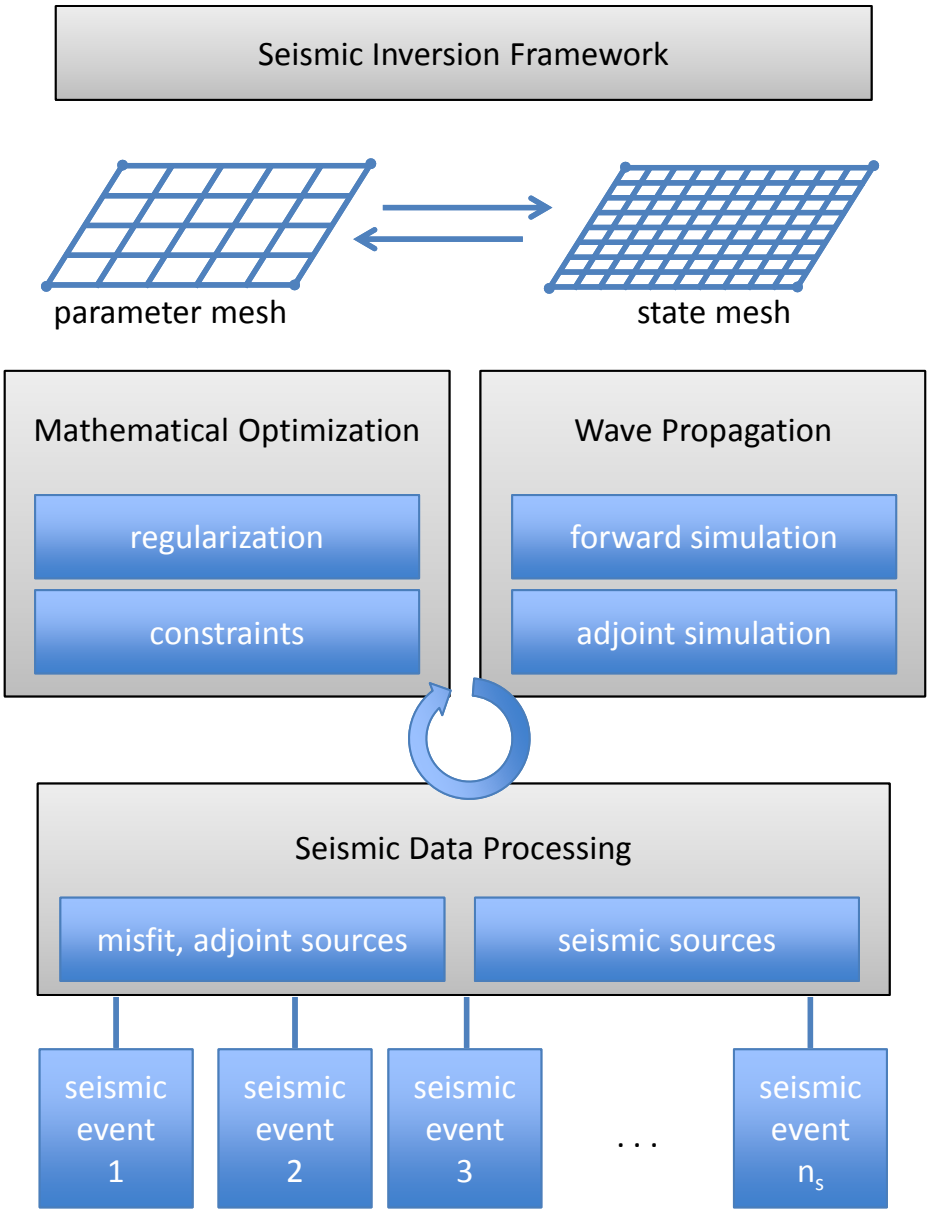


Figure 4.1: Schematic overview of the different components of the seismic inversion toolbox.

global one by `MPI::COMM_WORLD.Split`. Thus, both communication patterns can be handled simultaneously by invoking one of the two layers.
The optimization routines implement the previously described trust-region Newton-PCG method, which can handle additional constraints on the material parameters by utilizing the Moreau-Yosida regularization. Forward and adjoint simulations are carried out to efficiently compute the reduced gradient and reduced Hessian-vector products. The computation of the first derivatives requires the forward displacement field at all time-steps in reverse order. Moreover, forward and adjoint states need to be accessed again for computing Hessian-vector products. Note that we do not have to store the perturbed wavefields $\delta_s u$ and $\delta_s z$, cf. (3.18) and (3.19), because the contributions to the derivatives can be computed on the fly. Currently, we store both wavefields on disk, however, if memory poses a bottleneck, checkpointing techniques [59] could be applied. For an alternative strategy, see section 4.4.
The material mesh uses data structures of the deal.ii library [9, 10], which provides various tools for finite elements. In particular, we utilize functions

- to compute the regularization term and its derivatives,
- to compute the inner product as well as the preconditioner for the Steihaug-CG method,
- to process data for visualization purposes,
- to adaptively refine the grid (see Chapter 6).

As pointed out above, deal.ii is not used for solving the state equation.
Note that in order to offer more flexibility, state and parameter mesh are uncoupled. This modular structure allows to easily interchange either the simulation code or the optimization method. Using different meshes for the material and the state requires frequent mapping of elements from one mesh onto the other. This is necessary in both directions: the material is mapped onto the state mesh for carrying out the forward and adjoint simulations. In return, the gradient is computed on the state mesh first, requiring forward and adjoint state, and then mapped back to the parameter space.
Further information on the implementation and the user interface can be found in the documentation [17].

4.3 Numerical Results

In this section, we present numerical examples for the wave propagation code and seismic tomography problems. Because the inverse problem is not convex, we can usually not expect to converge to a global minimum. To this end, we use the term *reconstruction* instead of *solution*. On the other hand, since all examples are based on synthetic data and we use a moderate noise level and/or a sufficiently strong regularization, the results look promising. Before we start with the examples, we collect a few important remarks.
In all numerical examples, we choose the regularization term as the weighted sum of the L^2-norm and the H^1-seminorm, i.e., the discrete representation of

$$\alpha_1 \|m\|^2_{L^2(\Omega)} + \alpha_2 \|\nabla m\|^2_{L^2(\Omega)^d}.$$

Note that we use the same α_1/α_2-ratio to compute the $\|.\|_{\mathbf{M}}$-norm of discrete coefficient vectors. Typically, α_2 is a few magnitudes larger than α_1 as the L^2-regularization would often

yield oscillating reconstructions. In the rest of this section, all norms on the parameter mesh are computed using (3.46) and we drop the subscript **M**. With a slight abuse of notation, we write $\nabla j_\gamma(m^k)$ for the discrete gradient that is transformed to the M_h inner product, cf. section 3.5.
Concerning the size of the computational domain and the frequency of the seismic sources, all forthcoming examples refer to tomography problems on reservoir scale. This requires, in particular, to truncate the computational domain artificially at all borders but the surface and we have to impose absorbing boundary conditions to reduce reflections from non-physical boundaries. Here, we follow ideas from [80] and apply dampers that relate the traction to the velocity in the solid domain:

$$(\Psi : \varepsilon(u)) \cdot \vec{n} = v_p \rho \, (u_t \cdot \vec{n}) \, \vec{n} + v_s \rho \, (u_t - (u_t \cdot \vec{n}) \, \vec{n}) \quad \text{on} \quad \Gamma^{\text{abs}} \times I.$$

For the acoustic wave equation, we apply a Sommerfeld-like condition [121] in the fluid domain

$$\nabla u \cdot \vec{n} = -c^{-1} u_t \quad \text{on} \quad \Gamma^{\text{abs}} \times I.$$

Here, $\Gamma^{\text{abs}} \subset \partial\Omega$ denotes the artificial boundaries. It is important to note that these absorbing boundary conditions do not completely eliminate reflections from the boundaries and more sophisticated strategies exist, for instance, perfectly matched layers [96].

4.3.1 Adjoint and Gradient Computations

In this section, we illustrate the interplay of forward and adjoint sources, wavefields and the gradient of the material parameters. Here, we choose a similar setup as in the examples of [124]. We consider a two-dimensional solid domain of 20km × 12km, a single source located at $(5[km], 6[km])$ and one receiver at $(15[km], 6[km])$. In order to eliminate reflections from artificial boundaries, we extend the computational domain for the simulation. The expansion spans 12.5km in lateral directions and 18km in depth. The experimental setup is sketched in Figure 4.2. We use a Ricker wavelet with dominant frequency of 1Hz. The material is homogeneous with a P-wave velocity of 2500m/s and a constant Poisson's ratio of 0.25.

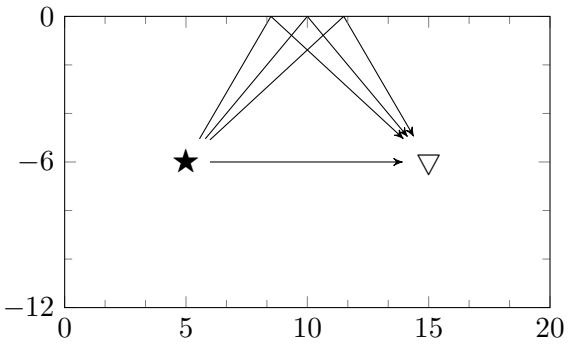


Figure 4.2: Experimental setup for the example in section 4.3.1. The source location is indicated by ★, the receiver location by ▽. Some ray paths of the direct wave as well as reflections from the free surface are illustrated.

By assuming a constant Poisson's ratio, we only have one unknown parameter field left for the inversion. To keep matters simple, we use $\|u(m)(x_r, .)\|^2_{L^2(I)^2}$ as cost functional. Obviously, this is not a useful misfit criterion, but it helps to highlight the regions that the particular wavefronts are sensitive to. Note that in this case the adjoint source is given by a subinterval of the recorded signal. Figure 4.3 shows snapshots of the forward and the adjoint displacement field as well as the interaction of both wavefields.

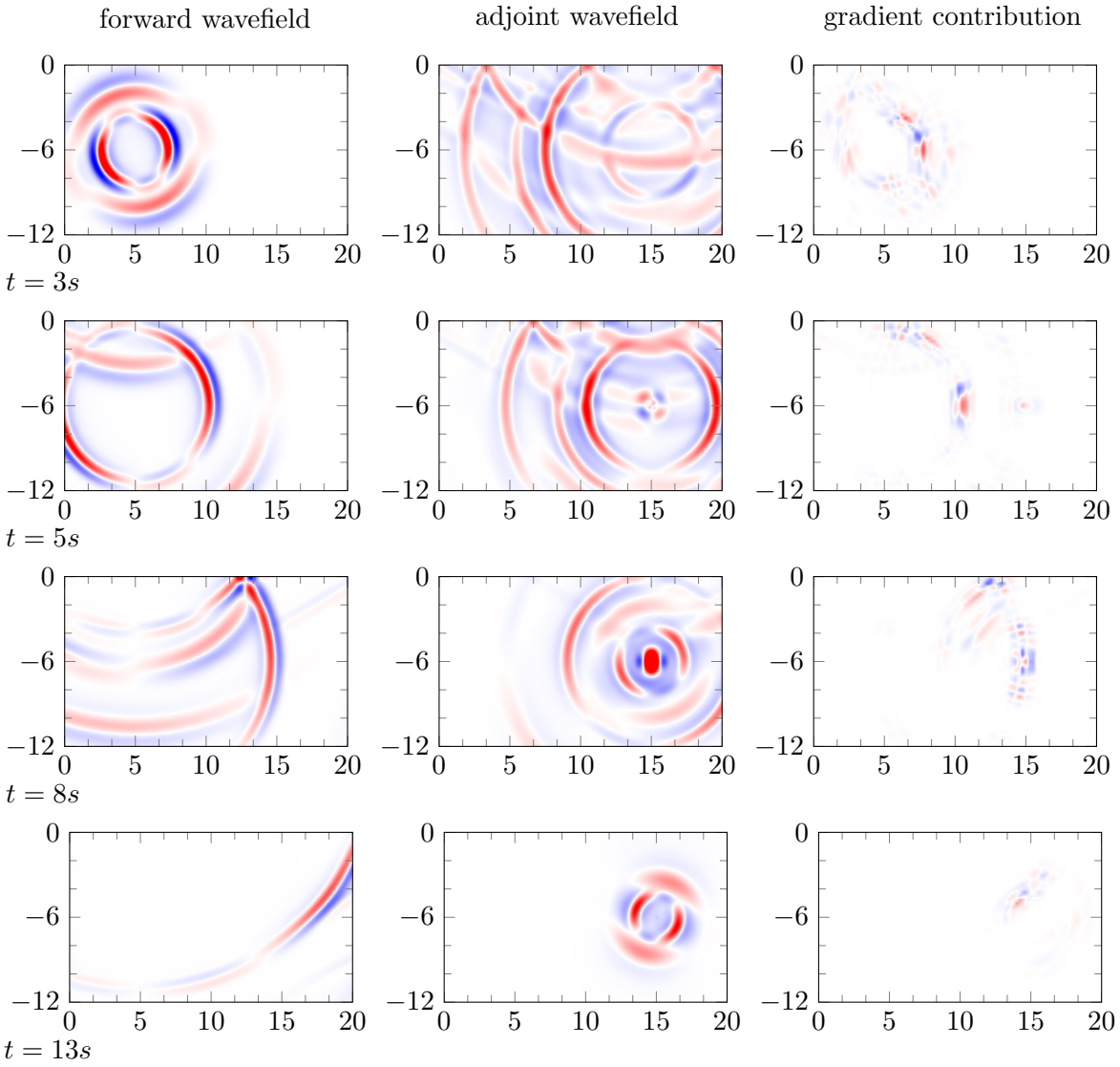


Figure 4.3: Snapshots of the forward (left) and the adjoint wavefield (middle) for different time steps. The contribution to the material gradient from the current time step is depicted in the right column. Snapshots are taken at 3s, 5s, 8s and 13s.

The paths along which the waves propagate on the way from the source to the receiver define the part of the domain, where the data is sensitive to changes of the material parameters. In other words, the gradient of the misfit with respect to the material coefficients is nonzero in this area. In Figure 4.4, we depict the gradient for three different scenarios, where either the whole seismogram or only a subinterval is used to compute the misfit. Hence, we are able to separate waves that propagate on a direct line from the source to the receiver, from the wavefronts that are reflected from the free surface. Such preprocessing of the data can be a powerful tool to consider distinct waveforms and certain parts of the domain. Because of their shape, the gradient is also called "banana-doughnut kernel" [124, 127] or Fréchet kernel [101] in geophysical literature.

4.3.2 Parallel Scaling

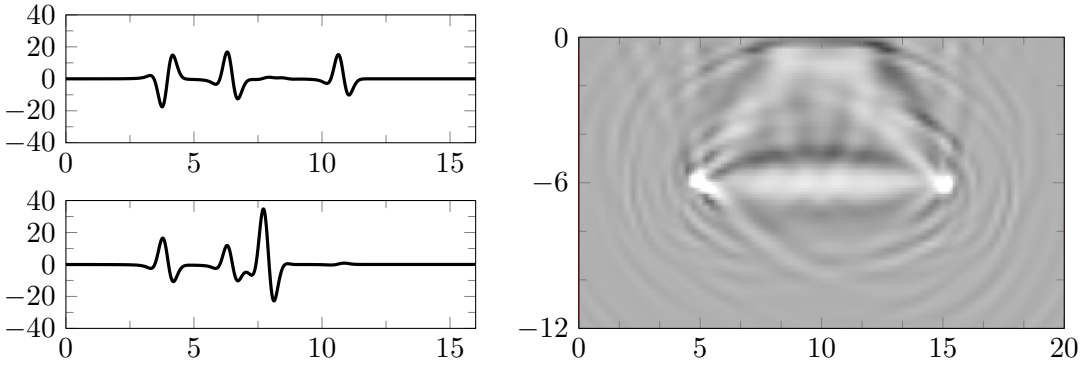
In this section, we present results on the parallel efficiency of the implementation. The computations have been carried out on two different architectures:

(I) AMD Opteron based 32-way nodes,
Linux Cluster at the Leibniz Supercomputing Centre (LRZ),

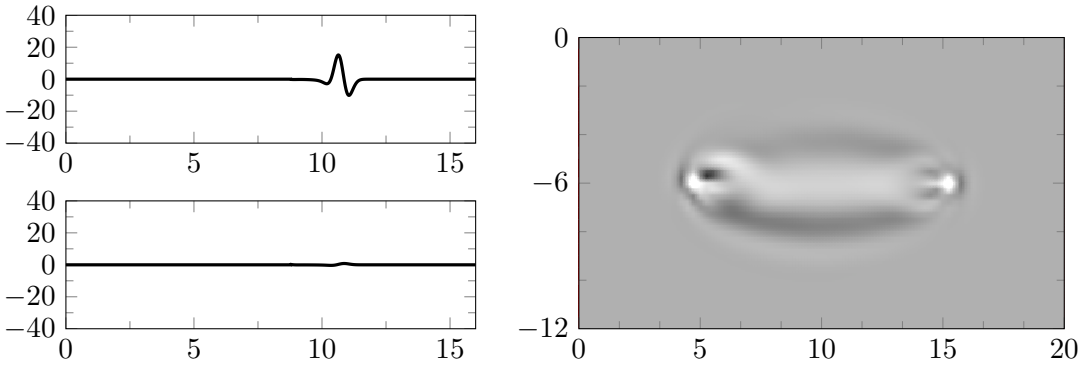
(II) Cray XC30 supercomputer based on Intel® Xeon® E5 processors,
Piz Daint at the Swiss National Supercomputing Center (CSCS).

All of the following statistics are based on computations using the second configuration.

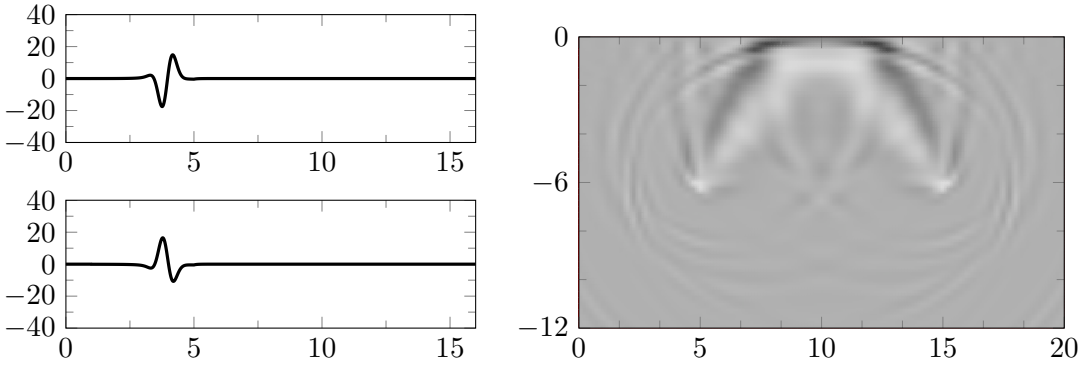
As a first example, we consider a 2d elastic simulation using the Marmousi model with 32 seismic sources. More details on the model will be given in section 5.2. The computational grid consists of 197,633 degrees of freedom in space and 6,000 time steps for every source. Note that by "degrees of freedom" we always refer to the number of grid points. Hence, for vector-valued problems this number has to be multiplied by the number of individual components to obtain the total number of unknowns. For instance, for the elastic wave equation, we have d components of the state and if we jointly invert for λ and μ, we also have two components in the parameter mesh. The reference configuration uses 32 cores in total and one core per source to simultaneously solve the elastic wave equation for every event. The number of cores is then increased up to 1024 cores. At the same time, the number of cores per source is increased from 1 to 32. The results are given in Table 4.1, indicating a reasonably strong parallel performance, with a parallel efficiency of 88.3% on 1024 cores.
Next, we present weak parallel scaling of a 3d elastic simulation. Here, we consider a domain of 4km $\times$ 4km $\times$ 4km and a homogeneous medium. We use a fixed number of 1,000 elements per core, which amounts in 68,921 degrees of freedom for each core, and increase the number of processors from 8 to 4096. The number of time steps is fixed to 1,000 for every scenario. Due to the CFL condition, this requires to successively reduce the simulation time interval by a factor of 0.5. The weak scaling statistic is given in Table 4.2. In this example, we define "scaling efficiency" of N cores as the ratio of the total run-time on 8 cores and the total run-time on N cores. The results indicate only a gradual growth of the computing time when the number of cores is increased.
As a final example, we present a strong parallel scaling statistic for a forward simulation of the elastic wave equation. Again, we consider a domain of 4km $\times$ 4km $\times$ 4km with a homogeneous



(a) full seismogram



(b) primary wave



(c) reflected wave

Figure 4.4: Sensitivity of the material with respect to certain time windows of the observed data. The left image shows the adjoint source in both components. The right column illustrates the gradient. In (a) the full seismogram is considered to compute the misfit, in (b) only the P-wave arrival (with time window $t \leq 7s$) is used and in (c) the sensitivity with respect to the reflected waves from the free surface is computed (with time window $10.5s \leq t \leq 15s$).

#cores	32	64	128	256	512	1024
#sources	32	32	32	32	32	32
#cores / source	1	2	4	8	16	32
total time (s)	265.3	130.1	66.6	33.7	18.5	9.4
speed-up	1.0	2.04	3.99	7.88	14.36	28.27
par. efficiency	1.000	1.020	0.996	0.985	0.898	0.883

Table 4.1: Parallel scaling statistics for a 2d elastic simulation of 32 sources. Discretization: 12,288 elements with 4th-order shape functions, 197,633 degrees of freedom, 6,000 time steps.

#cores	8	64	512	4096
#elements	8,000	64,000	512,000	4,096,000
#elements / core	1,000	1,000	1,000	1,000
total time (s)	16.7	17.0	17.3	17.9
scaling efficiency	1.000	0.979	0.963	0.935

Table 4.2: Weak scaling statistic for a simulation of the elastic wave equation in 3d. Discretization: 1,000 elements per core with 4th order shape functions, 68,921 degrees of freedom per core, 1,000 time steps (for all configurations).

medium and 4,000 time steps. The spatial discretization uses a mesh size of 200m, which yields 8,000 elements and 531,441 degrees of freedom. The results are given in Table 4.3. Note that the comparison to a single core is slightly unfair, because the single process can use the total memory of the node. Nevertheless, the parallel efficiency is very good and we observe nearly perfect scaling from 8 to 64 nodes, i.e., 1 to 4 full nodes.

#cores	1	2	4	8	16	32	64
#elements / core	8,000	4,000	2,000	1,000	500	250	125
#dofs / core	531,441	269,001	136,161	68,921	35,301	18,081	9,261
total time (s)	458.7	233.4	122.6	67.6	33.4	16.7	8.29
par. efficiency	1	0.982	0.935	0.848	0.859	0.859	0.865

Table 4.3: Strong scaling statistic for a simulation of the elastic wave equation in 3d. Discretization: 8,000 elements in total with 4th-order shape functions, 531,441 degrees of freedom, 4,000 time steps.

Summing up, the results indicate the strong parallel performance of the implementation, which shows the capability to tackle large-scale seismic inverse problems. We assume the parallel efficiency can be improved further if non-blocking MPI communication is utilized to full extend.

4.3.3 Joint Inversion for Both Lamé Coefficients

Now, we turn to the first inverse problem. The results have also been published in [19]. In this example, we invert for both Lamé coefficients, λ and μ, simultaneously. Here, we use additional constraints on the Poisson's ratio of the material to relate both parameter fields to each other and to ensure that this quantity remains within reasonable bounds. We recall from Example 3.3.3 that the Poisson's ratio ν is defined by

$$\nu(x) = \frac{1}{2}\frac{\lambda(x)}{\lambda(x)+\mu(x)}, \quad x \in \Omega. \tag{4.4}$$

By rearranging the terms, we can impose a pointwise lower and an upper bound on the Poisson's ratio $\nu^a \leq \nu \leq \nu^b$ by

$$2\nu^a\mu + (2\nu^a - 1)\lambda \leq 0, \quad -2\nu^b\mu + (1 - 2\nu^b)\lambda \leq 0. \tag{4.5}$$

We refer to Example 3.3.3 for a different representation of (4.5) which includes the reference model.

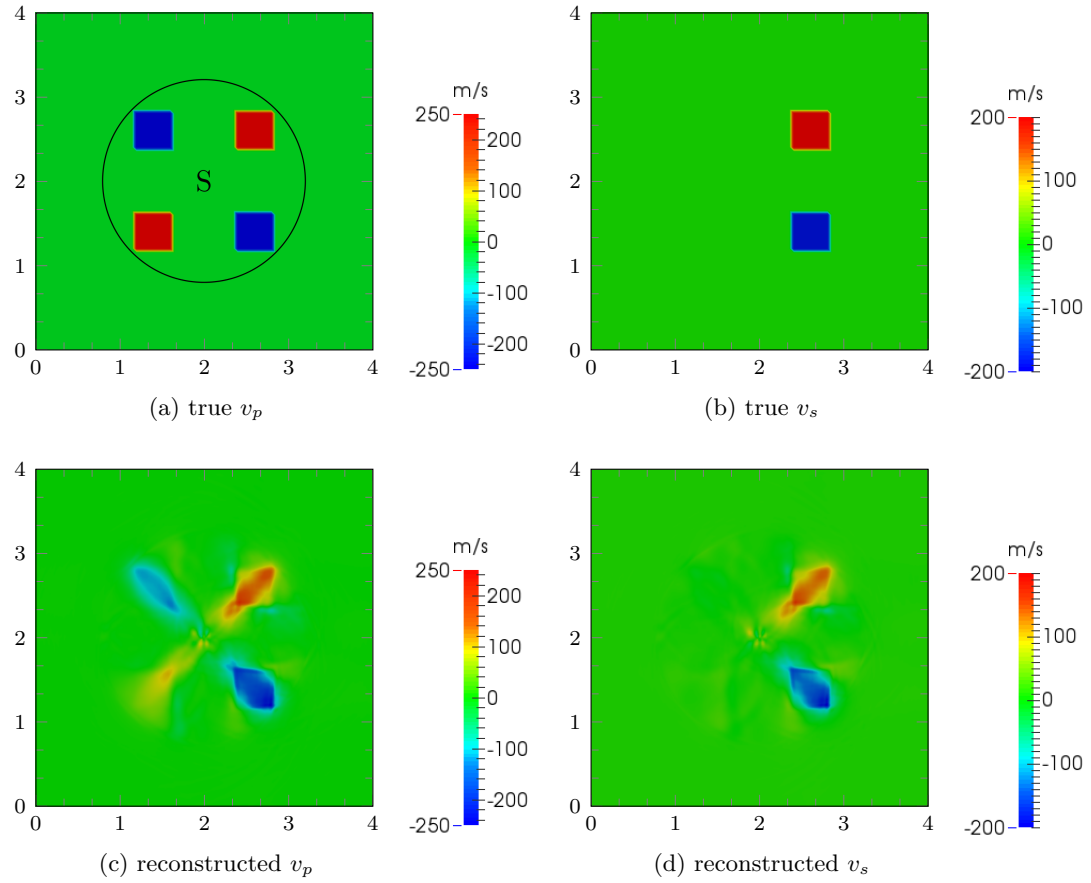
In this test setup, we consider a time interval of 2.5s and a rectangular domain of 4km $\times$ 4km with a single source in the center of the domain. Furthermore, there are 360 receivers on a sphere in 1.2km distance from the source. The reference material has a P-wave velocity of 2500m/s and a constant Poisson's ratio of 0.25. There are four block perturbations of the material with a P-wave velocity of either 2750m/s or 2250m/s. These perturbations are created by modifying either λ or μ, but not both (see first row of Figure 4.5). Thereby, the Poisson's ratio varies from 0.15 to 0.31. Data is generated by a simulation using this material model and adding 2% Gaussian noise. The source is modeled by a Ricker wavelet with a dominant frequency of 10Hz. The discretized problem has 103,041 spatial grid points for the parameter and state mesh and 3,000 time steps. Note, however, that the material is parameterized with bilinear shape functions while the state uses 4th-order polynomials. The initial model is homogeneous with a P-wave velocity of 2500m/s and a constant Poisson's ratio of 0.25.

We impose constraints on the Poisson's ratio for the inverse problem and restrict ν to $[0.15, 0.31]$. The regularization parameters are chosen $\alpha_1 = 10^{-10}$ and $\alpha_2 = 10^{-6}$.

The reconstruction is shown in the second row of Figure 4.5. Here, we show v_p and v_s computed from the reconstructed λ and μ and observe that the reconstruction matches the true material very well. Figure 4.6 shows a shotgather for all receivers before and after the inversion. Here, the amplitude of the signal is visualized as a function of receiver location (on the horizontal axis) and time (on the vertical axis). The order of the receivers is clock-wise. The first arrival around 0.6s shows the P-wave and the second arrival at roughly 1s is the S-wave. While the arrival times are the same at all locations for the homogeneous initial material, the waves arrive delayed or premature when the true material is used . In particular, the P-wave arrival time is affected by all four block perturbations, while the S-wave arrival time is only sensitive to the first two. We observe a good match of synthetic and observed data. In particular, the misfit has been reduced by 95% compared to the initial material. Note that the constraints never become active during the inversion, hence γ is not increased and j and j_γ coincide for every iterate.

In this example, we do not limit the number of CG iterations in Algorithm 3 and choose a tolerance

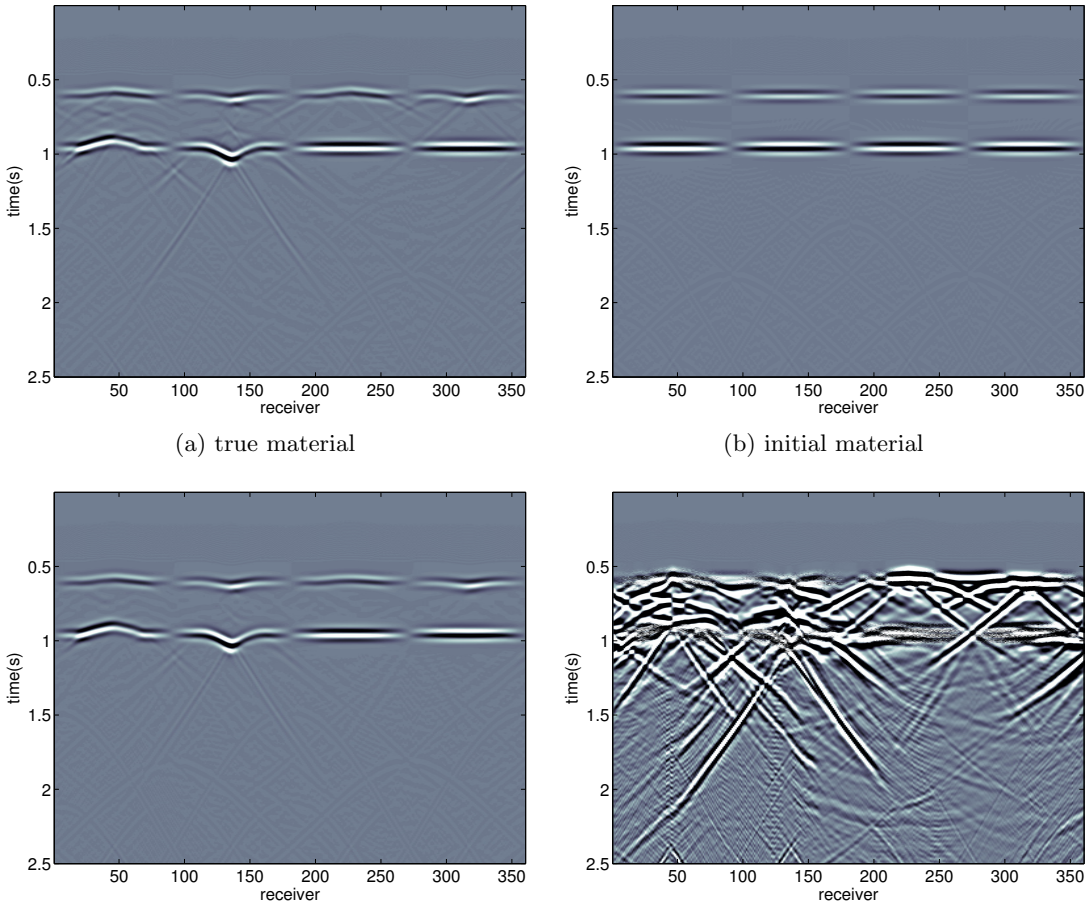
$$\varepsilon_k = 0.01 * \min\{1, \|\nabla j_\gamma(m^k)\|/\|\nabla j_\gamma(m^0)\|\}.$$



(a) true v_p (b) true v_s

(c) reconstructed v_p (d) reconstructed v_s

Figure 4.5: Joint inversion for both Lamé coefficients. All images show the deviation from the reference material. (a) true P-wave velocity (generated by modifying λ on the left half and μ on the right). (b) true S-wave velocity (changes only due to μ) (c) reconstruction of P-wave velocity (d) reconstruction of S-wave velocity.



(a) true material

(b) initial material

(c) reconstructed material

(d) error (amplified by 100)

Figure 4.6: Shotgather of all receivers using different material. (a) data signal for the true material, (b) data signal simulated with initial material model, (c) data signal simulated with the reconstructed material and (d) difference between the data signals for true and reconstructed material, amplified by a factor of 100.

The optimization process is shown in Table 4.4. After 22 iterations the norm of the gradient has been reduced by more than 12 orders of magnitude and we observe a superlinear convergence rate. However, the number of CG iterations increases significantly for the last 7 iterations. To investigate a possible improvement, we test different strategies regarding the maximum number of CG iterations. In order to compare the different configurations, we use a reduction of the norm of the gradient by 6 orders of magnitude as stopping criterion, which is sufficiently accurate for inverse problems. The computational effort is summarized in Table 4.5. With a maximum number of 20 CG iterations, 37 TR-Newton iterations are required and the total number of simulations is higher than in the previous case. Limiting the number of CG iterations to 40 provides a good tradeoff. Here, only two additional Newton iterations are required compared to the unlimited case and the total number of PDEs is about 10% less. From Figure 4.7 we see the fast local rate of convergence in the unlimited case, which is at least superlinear. This can not be observed for the other two cases. Nevertheless, the computational costs are dominated by the simulation of the elastic wave equation. Hence, the limit of 40 CG iterations gives the best result in terms of computing time.

it	$j_\gamma(m^k)$	$\frac{\|\nabla j_\gamma(m^k)\|}{\|\nabla j_\gamma(m^0)\|}$	cg it	Δ
0	1.85586	1.00e+00		
1	1.64239	1.15e+00	1	1.00e+06
2	1.37926	1.49e+00	2	2.00e+06
3	1.01413	1.26e+00	3	2.00e+06
4	0.77855	5.58e-01	3	2.00e+06
5	0.63894	6.08e-01	5	4.00e+06
6	0.61081	6.19e-01	9	8.00e+06
7	0.42257	3.94e-01	8	4.00e+06
8	0.36327	2.95e-01	11	4.00e+06
9	0.29691	2.09e-01	12	4.00e+06
10	0.25126	2.31e-01	15	8.00e+06
11	0.22447	1.97e-01	13	4.00e+06
12	0.19257	2.73e-01	18	8.00e+06
13	0.16579	1.82e-01	15	4.00e+06
14	0.14928	3.81e-01	16	8.00e+06
15	0.11612	1.74e-01	19	4.00e+06
16	0.09775	1.77e-01	62	8.00e+06
17	0.09306	2.75e-02	48	8.00e+06
18	0.09245	2.81e-02	68	8.00e+06
19	0.09240	9.88e-04	52	8.00e+06
20	0.09240	7.13e-05	72	8.00e+06
21	0.09240	2.74e-08	74	8.00e+06
22	0.09240	2.24e-13	169	8.00e+06

Table 4.4: Iteration tableau for the test without a limit on the number of CG iterations. Note that j and j_γ coincide in every iteration and we can use the reduction of the norm of the gradient as stopping criterion. Δ denotes the trust-region radius.

max cg	it Newton	avg. cg it	# PDEs
20	37	15.9	1370
40	23	20.4	1060
∞	21	25.4	1162

Table 4.5: Comparison of the computational effort for different configurations of the Steihaug-CG method. While the number of Newton iterations decreases when more CG iterations are allowed, the second configuration achieves the best results in terms of PDE simulations.

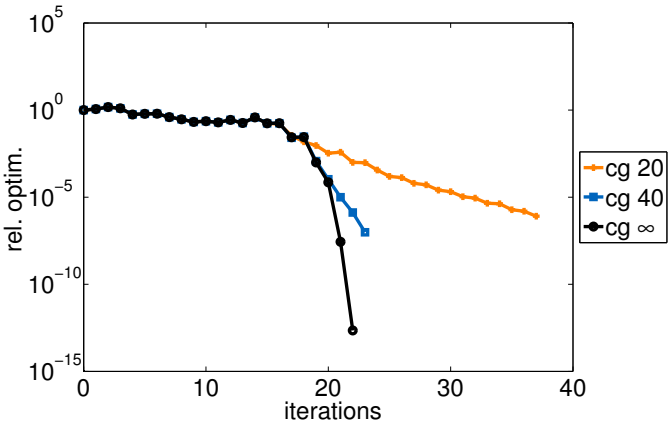
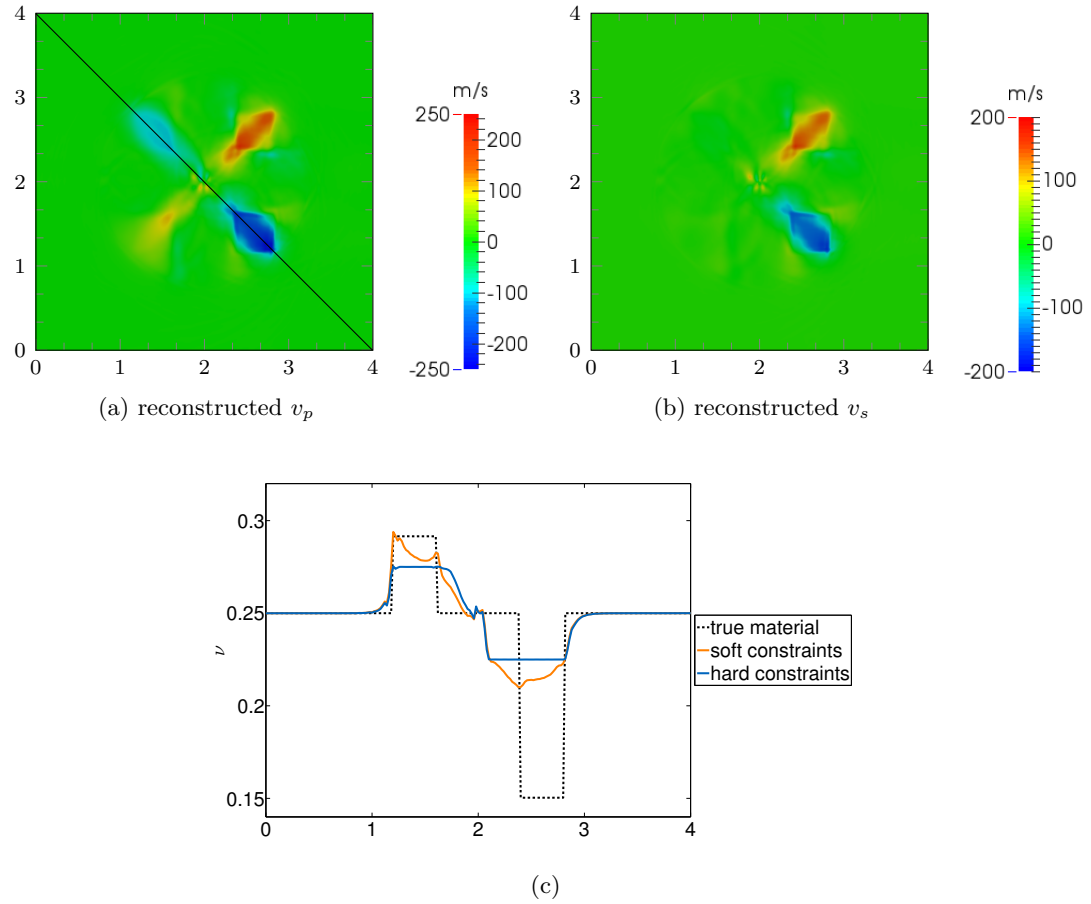


Figure 4.7: The different lines depict the relative norm of the gradient in every iteration. Fast local convergence can be observed if we do not impose a limit on the number of CG iterations.

In order to analyze the effect of constraints to more extend, we modify the problem formulation and restrict ν to values in $[0.225, 0.275]$. Hence, the true material is infeasible and the test setup is rather of academic nature. We use the previous reconstruction as initial model and restart the inversion which required 6 iterations and included 2 updates of the penalty parameter. Since the true model is unattainable, the bounds on the Poisson's ratio are active in the reconstruction. This is shown in Figure 4.8. Interestingly, however, the reconstructed P- and S-wave velocities still look very similar. The final misfit is 0.0378 and, therefore, slightly larger than in the previous case with 0.0329. Interestingly, however, the reconstructed P- and S-wave velocities still look very similar with a maximum pointwise difference of 29m/s for v_p and 41m/s for v_s. This shows that constraints can be used to add prior knowledge to the formulation of the inverse problem in order to restrict physical quantities that cannot be resolved by the measurements.

4.3.4 Borehole Tomography in 3D

In this example, which has also been published in [19], we consider a three-dimensional domain of 4km $\times$ 4km $\times$ 4km and a time interval of 6s. There is one seismic source with a dominant source frequency of 2.5Hz located in the lateral and longitudinal center at 3.75km depth. Furthermore, there are four boreholes near the corners of the domain equipped with receivers that measure data every 200m. In addition there is an array of 441 stations near the surface



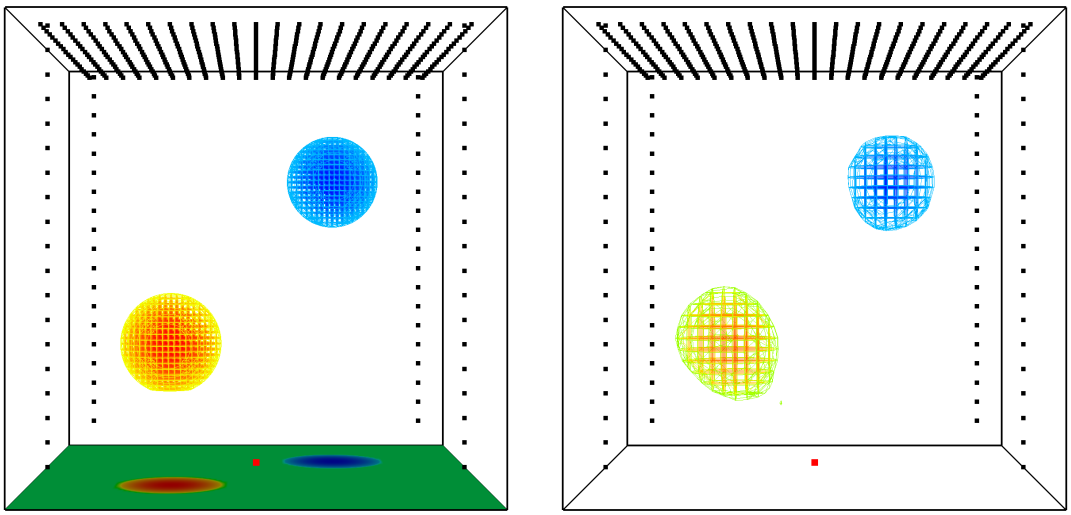
(a) reconstructed v_p

(b) reconstructed v_s

(c)

Figure 4.8: Joint inversion for both Lamé coefficients with unattainable true material model. (a) and (b) depict the reconstructed P-wave and S-wave velocity. (c) indicates the Poisson's ratio along the diagonal line depicted in (a). The plot shows the true material with the dotted black line, the previous reconstruction in orange and the reconstruction with "hard constraints" in blue. Both, lower and upper bounds are active in parts of the domain.

with 21 receivers each in lateral and longitudinal directions and a 175m spacing. Similar as in the previous example, the "true" material has a homogeneous P-wave velocity of 2500m/s with two ball-shaped perturbations of either 2700m/s or 2250 m/s. The material model as well as the locations of sources and receivers are shown in Figure 4.9(a). Here, we assume a constant Poisson's ratio of 0.25 and invert only for λ. Again, the initial model is homogeneous with a P-wave velocity of 2500m/s. We use the lower and upper bounds of the true material as constraints for the absolute value of λ, which gives $\lambda \in [3.375 \cdot 10^9, 5.042 \cdot 10^9]$.
For the spatial discretization of the elastic wave equation, we use 531,441 grid points and 4,000 time steps. The parameter mesh has 68,921 degrees of freedom and is discretized by 41 $\times$ 41 $\times$ 41 grid points. Figure 4.9(b) shows the reconstruction, see also Figure 4.10 for slices through the domain.

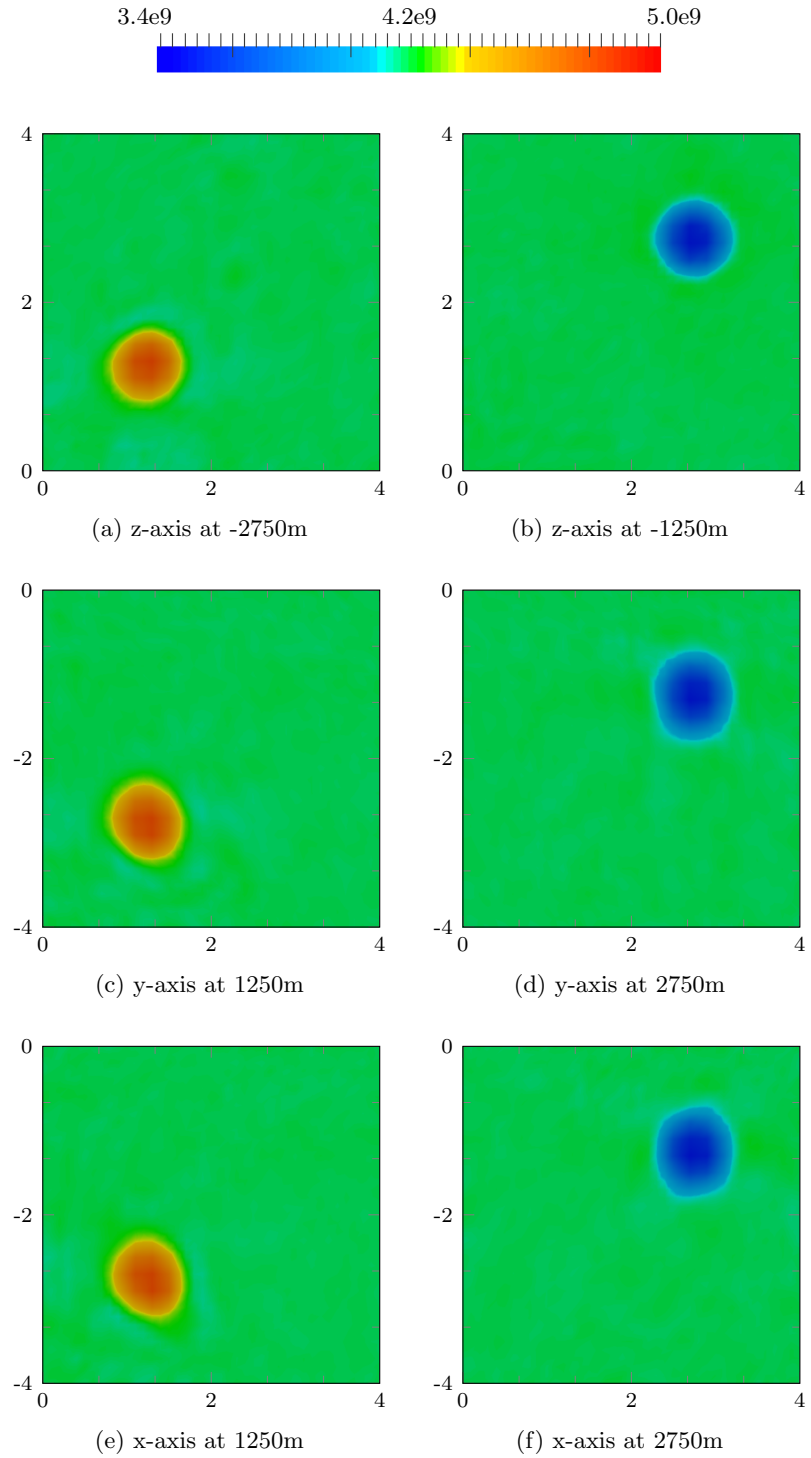


(a) true material and setup

(b) reconstruction

Figure 4.9: Borehole Tomography. The test setup is shown in (a) with black dots representing the locations of the receivers and the red dot indicating the position of the seismic source. Furthermore, the two perturbations of the true material are visualized and projected to the bottom for a better visibility. (b) shows the reconstruction which captures the perturbations very well.

Only 6 Newton iterations with a maximum of 40 CG steps are required to solve the problem to a relative tolerance of 10^{-6}. Hereby, the misfit has been reduced by more than 99%. In total, 396 PDEs are solved during the inversion. Note that this problem is considerably easier to solve than the example in section 4.3.3, which has mainly two reasons. On the one hand, there is only one parameter field to determine instead of both Lamé coefficients. On the other hand, there is a good coverage of the domain by the receivers at the surface and inside the boreholes.
Summing up, the numerical experiments of this section show the capability of the previously discussed inversion methods. As expected, the quality of the reconstruction and the computational effort to obtain a solution depends on the amount of information that is included in



(a) z-axis at -2750m (b) z-axis at -1250m

(c) y-axis at 1250m (d) y-axis at 2750m

(e) x-axis at 1250m (f) x-axis at 2750m

Figure 4.10: Borehole tomography reconstruction. The images (a) - (f) depict slices through the domain for every plane and show the absolute value of λ.

the data. Further numerical examples can be found in Chapter 5 and Chapter 6. We conclude this part with some remarks on possible extensions to enhance the practical performance with regard to robustness and memory requirements.

4.4 Further Perspectives

Misfit Functional

The proper choice of the misfit functional is crucial for the success of the inversion. In our numerical experiments, we use the L^2-misfit which is suitable for the case of synthetic data and a small amount of Gaussian noise. However, more sophisticated misfit criteria have been tested and are required to quantify the difference between simulated and observed seismograms for real data. For instance, in [128] a correlation-based misfit criterion that uses a weighted norm of the convolution of both signals is proposed, which is very suitable if both signals differ by a phase shift or a phase rotation. Time-frequency misfits [82, 83]give another alternative, where phase and amplitude information of the seismograms are separated. See also [23] for an analysis on robust misfit measures in the frequency domain. A comparative study is beyond the scope of this thesis, but we note that different criteria can easily be incorporated into the inversion framework, as long as they are sufficiently smooth, cf. Assumption 3.2.6.

Regularization

In all numerical examples, we consider a Tikhonov-type regularization term, which is often the classical choice for inverse problems. On the other hand, seismic tomography is strongly related to problems in image reconstruction which suggests using other types of regularization that are commonly used in this field. This can include, for instance, a total variation regularization [48] or exploring sparsity in a curvelet space [29, 65]. While there are limitations with regard to the infinite-dimensional framework of Chapter 3, it might be a very interesting enhancement for practical purposes.

Wavefield Compression

For large problems in seismic tomography, I/O operations and memory requirements become increasingly important. This is due to the fact that forward and adjoint wavefields have to be processed asynchronously in time. In particular, the forward wavefield has to accessed during the adjoint simulation for computing the gradient and, furthermore, both forward and adjoint wavefield are required to compute Hessian-vector products.
Even for mid-size problems this can amount in several terabytes of data. It is well-known that memory requirements can be traded for additional simulations using checkpointing techniques [59]. However, this might increase the computing time significantly. A promising alternative is proposed in [41] where the wavefield is locally compressed using different floating point accuracies. It would be an interesting field for future research to tailor these methods to the computation of gradients or Hessian-vector products.

Chapter 5

Randomized Source Sampling

The computational costs of solving the seismic inverse problem are vastly dominated by the frequent simulations of the wave equation. We recall that n_s denotes the number of seismic events from which data is considered. The conventional formulation of the inverse problem, which was discussed in Chapter 3, requires n_s simulations of the wave equation in order to evaluate the objective function for a given material model. By using adjoint techniques, additional n_s PDEs have to be solved to compute the gradient. Furthermore, we require $2\,n_s$ simulations per CG iteration during the Newton-CG method. Let $\mathrm{N}_{\mathrm{newt}}$ denote the number of Newton iterations and N_{acg} the average number of CG steps per Newton iteration. Then the total number of PDEs that have to be solved is at least $2n_s\, N_{\mathrm{newt}}(1 + N_{acg})$. Additional simulations might be necessary if steps are rejected due to a poor ratio of actual and predicted reduction, cf. Algorithm 4. Moreover, the numbers above implicitly assume that it is possible to store the whole forward and adjoint wavefields for all sources. If this is not feasible due to memory restrictions, checkpointing techniques [59] have to be employed and the number of PDEs per iteration increases further. Hence, an approach that reduces the amount of required simulations is highly desirable.

In seismic tomography, data is usually taken from a large number of seismic events which is in the order of tens to a few hundreds for problems in global seismology [110, 118] and even 10^5 to 10^6 for problems in geophysical exploration on industrial data sets [107]. Here, the number of required simulations would be computationally intractable if every source was considered independently. A promising approach, which has been applied successfully to tomography problems in the frequency domain [61], is to exploit the linearity of the elastic wave operator with respect to the state and trigger several sources simultaneously as so-called *super-shots*. We also refer to [112] for a simplified version in the time domain using the acoustic wave equation. Note that similar methods are sometimes also called *source stacking* [30] or *source encoding* [15, 81] in the geophysical community.

Building upon ideas from [61], we present an extension that is tailored to Newton-type methods, where the computational costs can be reduced further by working with an approximation of the Hessian of the sample average. Note that some of the analysis is restricted to the finite-dimensional parameter space, but we omit the superscript h to improve the readability. The numerical examples in this chapter underline the applicability of the proposed method for tomography problems in the time domain.

5.1 Source Stacking

In this section, the data is assumed to have the following properties: We consider n_s seismic events with sources denoted by by f_i. For each source there exist measurements u_i^δ in $\Omega_{\text{obs}} \times I$, $\Omega_{\text{obs}} \subset \Omega$. Note that it is essential for the following analysis that all events are observed at the same locations. We refer to Remark 5.3.5 for a possible extension.

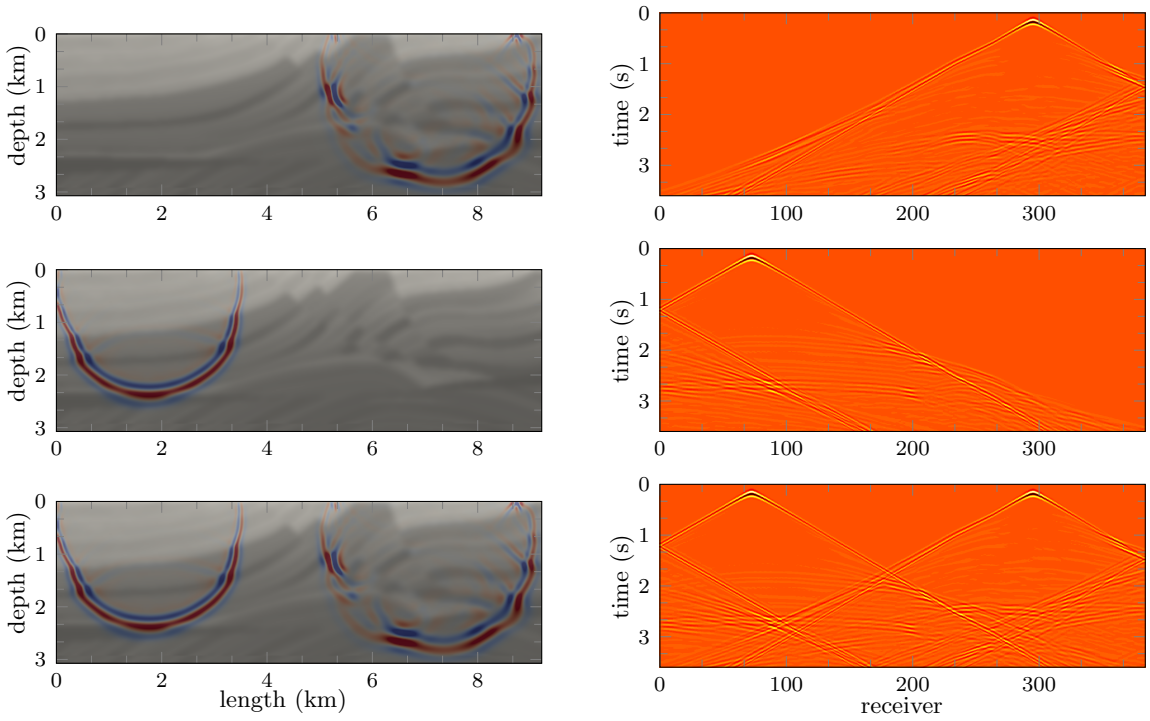


Figure 5.1: Illustration of the source stacking approach. The left column shows snapshots of the wavefield for two individual sources as well as the wavefield for the simultaneously triggered sources in the bottom row. The common shotgathers for the individual and the stacked sources are depicted in the right column. The traces start to overlap at roughly 1.6s.

As pointed out above, the key idea is to trigger different seismic sources simultaneously. This is shown in Figure 5.1 for a simple case with two sources. Note that the wavefronts as well as the recorded signals overlap in this approach, which can be seen from the common shotgather for time steps greater than 1.6s. Now, we turn to the general case with a large number of sources. Here, we combine the individual events by building the weighted sum of the corresponding right-hand sides. For $w \in \mathbb{R}^{n_s}$, we define $\mathbf{u}(m, w)$ as the unique solution to:

$$E(\mathbf{u}, m) = \sum_{i=1}^{n_s} w_i f_i, \quad \mathbf{u}(0) = 0, \quad \mathbf{u}_t(0) = 0.$$

Due to the linear dependence of the elastic wave operator on the right-hand side, we have

$$\mathbf{u}(m, w) = \sum_{i=1}^{n_s} w_i u_i, \quad \text{where } u_i \text{ solves} \quad E(u_i, m) = f_i, \quad u_i(0) = (u_i)_t(0) = 0. \tag{5.1}$$

Consequently, we now compare the seismograms generated by $\mathbf{u}(m, w)$ with the weighted sum of the measured data. To this end, we slightly change the notation of the misfit functional and allow for two variables, the simulated wavefield and the observed data. Thus, we define $J_{\text{fit}} : U \times L^2(I; L^2(\Omega_{\text{obs}})^{\hat{d}}) \to \mathbb{R}$ and, furthermore,

$$J(m; w) := J_{\text{fit}} \left(\mathbf{u}(m, w), \sum_{i=1}^{n_s} w_i u_i^\delta \right).$$

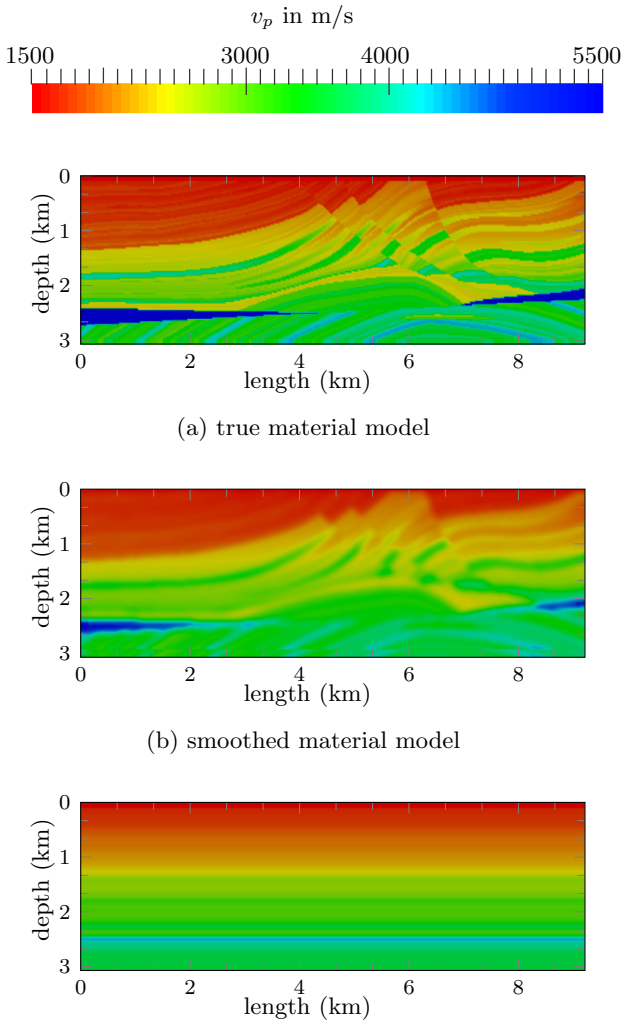
It is important to note that, in general, we have $J(m; w) \neq \sum_{i=1}^{n_s} w_i J_{\text{fit}}(u_i(m), u_i^\delta)$. Hence, there is a potential loss of information using the super-shot approach instead of individual events. From (5.1) it can be seen that this is solely caused by the nonlinearity of the misfit functional.

Remark 5.1.1.
The misfit functional has to be chosen carefully such that comparing the accumulated data still provides a meaningful value. This might be challenging for noisy measurements. In particular, we assume only a small level of Gaussian noise and use a least squares misfit in our numerical examples. Note, however, that the same ideas can be applied to various statistics of the measurement errors - including distributions with large outliers - if the misfit criteria is chosen appropriately, see [6].

We will continue with different strategies that incorporate the idea of the super-shots into the inversion process. Beforehand, we introduce the data set that is considered throughout this section.

5.2 The Marmousi Model

Our study is based on the Marmousi data set provided by the Institut Français du Pétrole Energies Nouvelles [21, 129]. It consists of a rectangular domain of 9,216m × 3,072m. The wave velocities are highly heterogeneous and the material contains a series of normal faults and resulting tilted blocks. The P-wave velocity ranges from 1500m/s to 5500m/s. Note that the original data set is an acoustic model, however, we generate an elastic model by using the P-wave velocities of the acoustic model and assuming a constant Poisson's ratio of 0.25. This gives the relation $\lambda = \mu$ and we invert for λ only. Due to the constant Poisson's ratio the S-wave velocity is given by $v_s = (1/\sqrt{3})v_p$. The P-wave velocity profile is depicted in Figure 5.2. In addition, we show two other material models that are considered for our tests.
For setting up the experiment, we place 191 seismic sources at 36m depth. The signal is recorded by 384 receivers that are distributed equidistantly on a horizontal line at 100m depth. This setup mimics a marine seismic exploration with sources located in a thin layer of water at the top of the domain and geophones placed at the bottom of the ocean. Note, however, that we do not explicitly model the fluid layer by the acoustic wave equation in this example.
Figure 5.3 depicts snapshots of the elastic wavefield of a single seismic source. The heterogeneous media generates a lot of reflections and refractions of the waves. The parameter mesh is uniformly refined with a 24m spacing which results in 49,665 degrees of freedom.



(a) true material model

(b) smoothed material model

(c) initial material model

Figure 5.2: P-wave velocity profile of the Marmousi data set. (a) "true" Marmousi model (b) "smooth" Marmousi model (both can be obtained from `http://www.reproducibility.org/RSF/book/data/marmousi/paper_html/`). In order to generate a more realistic initial model, we assume a laterally constant velocity for fixed depth and assign the average value of the true Marmousi model to it. This leads to the velocity profile in (c).

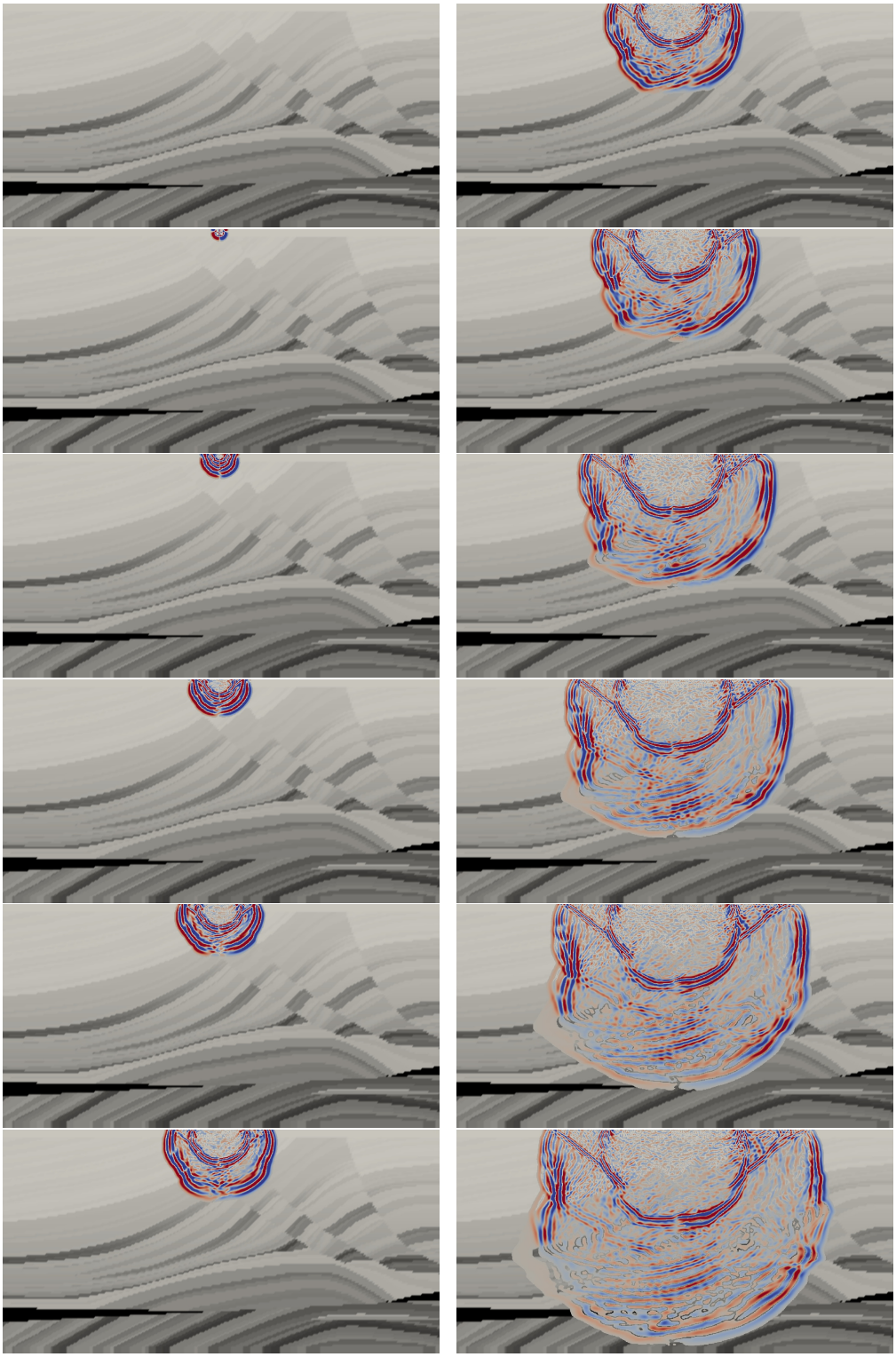


Figure 5.3: Snapshots of the vertical component of the displacement field for the elastic wave equation. Time flows by column from top to bottom.

5.3 Sample Average Approximation

Stacking all the data into just one super-shot will probably lead to a considerable loss of information, for instance, due to interference and cancellation of waves. However, by choosing several different weight vectors w^k and corresponding super-shots, we hope to diminish this effect, on the one hand, and still be able to reduce the computational costs significantly compared to individual right-hand sides, on the other hand. Now, consider $K \in \mathbb{N}$, $\mathcal{W} = \{w^1, \dots, w^K\}$ and the problem

$$\min_{m \in M_{\text{ad}}} j(m; \mathcal{W}) := \frac{1}{|\mathcal{W}|} \sum_{w \in \mathcal{W}} J(m; w) + \alpha J_{\text{reg}}(m). \tag{5.2}$$

Throughout this section, we work with the L^2 misfit.

Remark 5.3.1.
The formulation of the inverse problem (P) in Chapter 3 is a special case of (5.2). Indeed, for the choice $K = n_s$ and $w^k = e^k$ - where e^k is the k-th unit vector - the two cost functions coincide if the regularization parameter is scaled with $1/n_s$.

Remark 5.3.2.
On the other hand, for fixed $\mathcal{W}$, (5.2) might also be interpreted as a special case of (P) with K seismic events, sources $\hat{f}_k := \sum_{i=1}^{n_s} w_i^k f_i$ and measurements $\hat{u}_k^\delta := \sum_{i=1}^{n_s} w_i^k u_i^\delta$. Hence, we can directly deduce - even in the infinite-dimensional case - the existence of a solution to (5.2) for any fixed set $\mathcal{W}$ by Theorem 3.2.5. However, it is not clear whether the optimal solutions of (P) and (5.2) are related.

We may also interpret problem (5.2) from a stochastic programming perspective. To this end, we assume the weights w^k are i.i.d. samples from a distribution W with support on a subset of $\mathbb{R}^{n_s}$. In this context, the first term of the cost function in (5.2) is then called sample average approximation (see [113], Chapter 5) providing an estimate for the expected misfit:

$$\mathbb{E}\left[J(m; w)\right] \approx \frac{1}{K} \sum_{k=1}^{K} J(m; w^k). \tag{5.3}$$

Now the questions arise, whether for fixed m the sample average approximation converges to $\mathbb{E}[J(m; w)]$ for $K \to \infty$ and, furthermore, whether the solutions to (5.2) converge for $K \to \infty$. To this end, we define

$$\varphi(m) := \mathbb{E}[J(m; w)] + \alpha J_{\text{reg}}(m), \quad \varphi_K(m) := \frac{1}{K} \sum_{k=1}^{K} \Big(J(m; w^k) + \alpha J_{\text{reg}}(m) \Big)$$

and denote

$$\bar{\varphi} := \inf_{m \in M_{\text{ad}}} \varphi(m) \quad \text{and} \quad \bar{\varphi}_K := \min_{m \in M_{\text{ad}}} \varphi_K(m).$$

Furthermore, the sets of optimal solutions - if they exist - are denoted by $\mathcal{M}$ and $\mathcal{M}_K$.
In the following, we choose the probability space W to be Rademacher's distribution, i.e., $w^k \in \{-1, 1\}^{n_s}$ are i.i.d. samples with $P(w_i^k = 1) = P(w_i^k = -1) = 0.5$. This has been proposed in [61] as a suitable strategy to choose the weights. For a finite-dimensional parameter space M_{ad} in (5.2), we obtain the following result:

Theorem 5.3.3.
Let $M_{ad} = M_{\mathcal{D}} = \mathbb{R}^{n_m}$ and w^k i.i.d. samples of Rademacher's distribution. Furthermore, we define $\varepsilon := \sqrt{2J_0/\alpha}$ with

$$J_0 := \max_{w\in\{-1,1\}^{n_s}} J(0;w).$$

Then $\varphi_K(m)$ converges to $\varphi(m)$ with probability one uniformly on $B_\varepsilon(0)$. Moreover, it holds $\bar{\varphi}_K \to \bar{\varphi}$ and $\mathbb{D}(\mathcal{M}_K, \mathcal{M}) \to 0$ with probability one as $K \to \infty$, where the deviation $\mathbb{D}(A,B)$ of two sets $A, B \subset \mathbb{R}^{n_m}$ is defined by

$$\mathbb{D}(A,B) := \sup_{a\in A} \mathit{dist}(a,B).$$

Proof. For an arbitrary sample w of Rademacher's distribution, we observe using (5.1)

$$\|u(m;w)\|_U \le \sum_{i=1}^{n_s} |w_i| \cdot \|u_i(m)\|_U = \sum_{i=1}^{n_s} \|u_i(m)\|_U \le C,$$

since $u_i(m)$ is uniformly bounded on M_{ad} by Corollary 2.3.7. Furthermore,

$$\begin{aligned}\|u(m;w) - \sum_{i=1}^{n_s} w_i u_i^\delta\|_{L^2(I;L^2(\Omega)^{\hat{d}})} &= \|\sum_{i=1}^{n_s} w_i(u_i(m) - u_i^\delta)\|_{L^2(I;L^2(\Omega)^{\hat{d}})} \\ &\le \sum_{i=1}^{n_s} \|(u_i(m) - u_i^\delta)\|_{L^2(I;L^2(\Omega)^{\hat{d}})}.\end{aligned}$$

Thus, there exist $C > 0$ independent of w and m such that

$$|J(m;w) + \alpha J_{\mathrm{reg}}(m)| \le C \quad \forall\, m \in B_\varepsilon(0), \quad \forall\, w \in W. \tag{5.4}$$

Hence, the first statement follows by the uniform law of large numbers, cf. Theorem 7.48 in [113]. Furthermore, W is finite, thus minimizing φ might itself be interpreted as a special case of (5.2). In particular, $\mathcal{M}$ and $\mathcal{M}_K$ are nonempty and contained in $B_\varepsilon(0)$. Thus, the convergence of the optimal values as well as the convergence of the sets of optimal solutions is a special case of a more general setting that is proven in Theorem 5.3 in [113]. □

By employing Rademacher's distribution, we obtain a finite set W which facilitates the requirements of the cited Theorems considerably. However, Theorem 5.3 and Theorem 7.48 in [113] can also applied to general distributions with some additional assumptions. Moreover, the analysis can be extended to the derivatives of φ and φ_K.

Remark 5.3.4.
Despite the result from Lemma 5.3.3, it is hard to quantify the quality of the sample average approximation and the amount of preserved information in practice. We refer to [61] for a strategy to estimate $\bar{\varphi}_K - \bar{\varphi}$ that relies on solving a few instances of (5.2) for different weights.

We will now assess the practical performance of the sample average approximation by some numerical tests. First, we consider the acoustic version of the Marmousi model and a source frequency of 10Hz. We simulate a period of 3.6s and use a mesh with 197,633 grid points in space and 6,000 time steps. Figure 5.4 shows the sample average for up to 32 super-shots. Again, we use Rademacher's distribution for the probability space W.

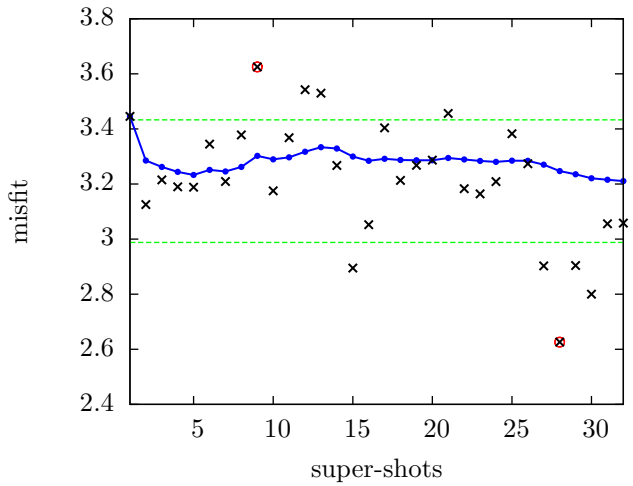
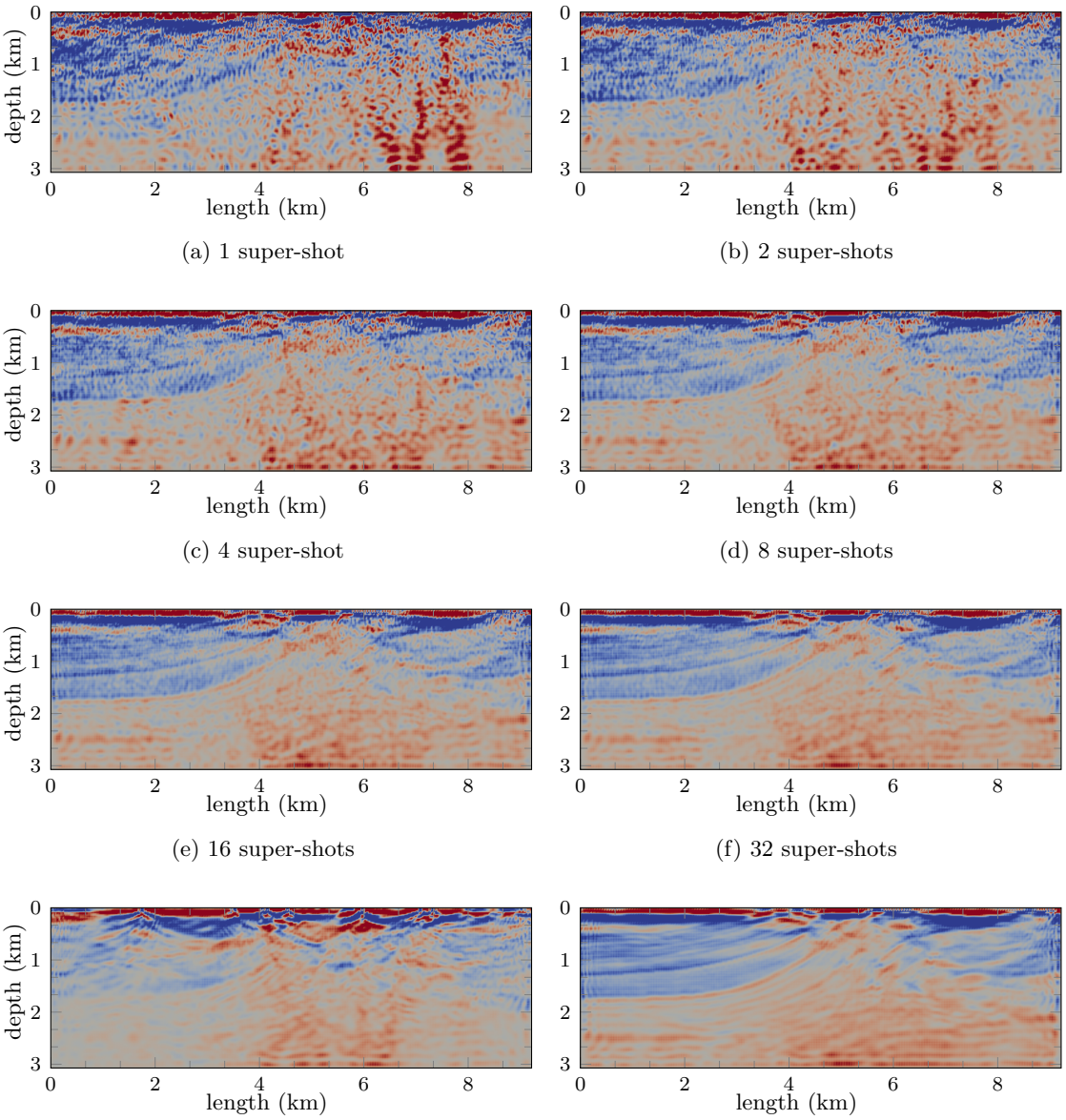


Figure 5.4: Sample average of the misfit for up to 32 super-shots. The black crosses denote the misfits of single super-shots and the blue line depicts the sample average of all preceding super-shots. The maximum and minimum values are highlighted with a red circle and the green dashed lines indicate the interval of mean ± empirical standard deviation for 32 super-shots.

In order to clarify the gain of information using several super-shots instead of only one, we illustrate the gradient for different choices of K in Figure 5.5. To this end, data is generated with the true material model and the gradient is computed for the smoothed material (see Figure 5.2(b). We compare the gradient for 1 to 32 super-shots in (a) - (f). Clearly, the gradient becomes smoother the more super-shots are considered. We also depict the gradient computed with data from 8 individual seismic events in Figure 5.5(g). While the computational effort is the same as for 8 super-shots, using super-shots instead of individual sources seems to provide a significantly better update. In addition, we compute the full gradient of 191 individual events and visualize the result in Figure 5.5(h). Keeping in mind that the computational costs are proportional to the number of sources (individual or simultaneous), we observe that a small number K seems to be a good tradeoff between preserving information and efficiency of computation.

Remark 5.3.5.
There are some limitations of the sample average approach regarding the application to real data. On the one hand, data might not be observed at the same locations for all sources. This is usually the case for earthquake sources but also applies to certain experimental design setups in geophysical exploration where the array of receivers moves with the source. To overcome this problem, strategies on extrapolating the wavefield are discussed in [61] and the references therein. On the other hand, corrupted data, e.g., due to a defective device, can have a stronger influence with the super-shot approach than with individual sources. Here, either an appropriate misfit criterion or an effective preprocessing step to filter out corrupted data is necessary.



(a) 1 super-shot

(b) 2 super-shots

(c) 4 super-shot

(d) 8 super-shots

(e) 16 super-shots

(f) 32 super-shots

(g) 8 individual sources

(h) all individual sources

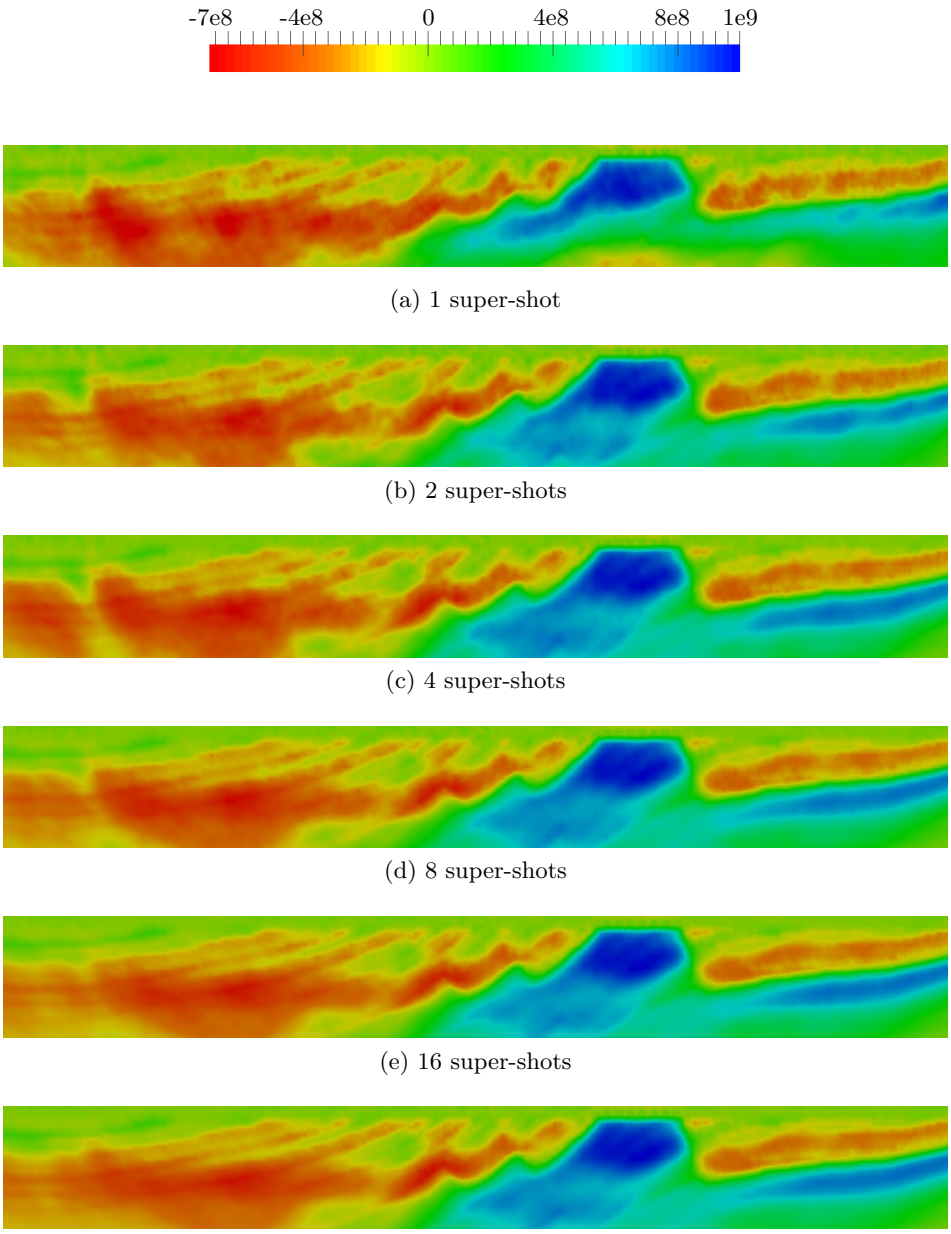
Figure 5.5: Gradient of the sample average for the acoustic Marmousi model. In (a)-(f) the gradient of the sample average misfit using 1, 2, 4, 8, 16, and 32 super-shots is shown. For a comparison, (g) shows the gradient computed by 8 randomly selected individual events and (h) displays the full gradient with 191 independent sources. The computational effort increases by a factor of 2 at a time from (a) to (f) and again by almost a factor of 6 from (f) to (h).

Now, we return to problem (5.2) and observe that is has the same structure as (P) for fixed $\mathcal{W}$. Hence, we can readily use the methods from Chapter 3 to solve the inverse problem. As main advantage of the sample average approximation, the computational costs for evaluating the objective function or gradient, or performing a conjugate gradient iteration now scale with factor $K = |\mathcal{W}|$ instead of n_s. Provided that neither the number of iterations nor the quality of the reconstruction change considerably for $K \ll n_s$, this reduces the number of simulations significantly. In addition, the memory requirements for storing forward and adjoint wavefields during the CG iterations are also reduced by K/n_s.
The most important question is certainly, how the quality of the reconstruction is affected by K and, furthermore, what the computational effort to obtain a solution is. To this end, we perform a second test using the elastic version of the Marmousi model and the initial model depicted in Figure 5.2(c). This is a realistic scenario for a starting model as no further information apart from a single depth profile is required. The source time function at all locations is a Ricker wavelet with dominant source frequency of 5Hz. 1% Gaussian noise is added to the data generated with the true Marmousi material. Again, we use i.i.d. samples of Rademacher's distribution for the weights of the super-shots and we do not impose constraints on the material in this example.
The reconstructions for different numbers of super-shots are shown in Figure 5.6. Note that we only consider depths up to 1km as the reconstruction becomes less accurate at greater depths and the error would be dominated by those regions. Compared to 8 super-shots, the reconstruction using only one super-shot is rather noisy. On the other hand, it is quite remarkable how much information is preserved already by a single super-shot.
In order to quantify the quality of the reconstruction, we compare it with the true material model and with respect to three different criteria in Table 5.1. Again, the comparison considers only up to 1km of depth. In particular, we specify the maximum pointwise error relatively to the true model and utilize two criteria that are commonly used for the purpose of image comparison. Here, we compute the 2D correlation coefficient as well as the structural similarity index (SSIM), see [132]. While the different criteria rank the reconstructions in slightly different order, there is a clear indication that the quality of the reconstruction improves with the number of super-shots.

	1	2	4	8	16
max. error	0.5880	0.5411	0.5618	0.5365	0.5371
corr2	0.8906	0.9008	0.9052	0.9031	0.9042
SSIM	0.3952	0.4072	0.4183	0.4197	0.4213

Table 5.1: Quantitative assessment of the reconstruction. The table indicates the maximum pointwise error (the smaller the better), the 2D correlation coefficient (corr2, the higher the better) and the structural similarity index (SSIM, the higher the better).

Table 5.2 compares the accumulated misfit computed with the initial as well as the final model for the different scenarios. In all cases, the total misfit has been reduced by around 92%.
The computational effort for solving this problem with different numbers of super-shots is given in Table 5.3. Slightly more Newton iterations are required for smaller number of super-shots. Afterwards, the number of iterations seem to stabilize around 30, which suggests that



(a) 1 super-shot

(b) 2 super-shots

(c) 4 super-shots

(d) 8 super-shots

(e) 16 super-shots

(f) 32 super-shots

Figure 5.6: Reconstruction of λ for an increasing number of super-shots in the sample average approximation. The images show the deviation from the initial model. The reconstructed material improves with increasing number of super-shots. On the other hand, very few super-shots already yield a good result.

	1	2	4	8	16
initial misfit	4.6689	4.7902	4.8081	4.8435	4.9401
final misfit	0.3441	0.4089	0.4306	0.4383	0.4404

Table 5.2: Comparison of the initial and final accumulated misfits for different numbers of super-shots K.

the sample average approach yields a similar behavior of the optimization process for an improved approximation of the expected value (5.3). As stopping criterion we use a relative reduction of the norm of the gradient by either 10^{-3} or 10^{-6}. The CG iterations are terminated after at most 40 iterations or if the relative residual is less than 0.01.

K	tol = 10^{-3}			tol = 10^{-6}		
	it Newton	avg. cg it	# PDEs	it Newton	avg. cg it	# PDEs
1	24	14.3	810	41	25.0	2255
2	25	15.3	1784	46	26.6	5354
4	28	18.8	4796	38	24.4	8196
8	24	17.0	7504	30	21.6	11584
16	23	15.6	13376	31	21.9	24256

Table 5.3: Comparison of the computational effort for different number of super-shots and different tolerances for the optimality conditions. The table states the number of Newton iterations, the average number of CG iterations per Newton iteration as well as the total number of PDE simulations for different numbers of super-shots K.

In the next step, we present an extension to the sample average approach that is tailored to the previously proposed Newton-type methods and is based on mini-batches of the samples. To this end, let $\mathcal{W} = \{w^1, \dots, w^K\}$ be fixed. From the structure of the sample average approximation there naturally arises the question, whether subsets (or mini-batches) of the samples can be used to compute the objective function and/or its derivatives. Note that the gradient and Hessian of $j(m; \mathcal{W})$ share the same structure:

$$\nabla j(m; \mathcal{W}) = \frac{1}{|\mathcal{W}|} \sum_{w \in \mathcal{W}} \left(\nabla J(m; w) + \alpha \nabla J_{\text{reg}}(m) \right),$$
$$\nabla^2 j(m; \mathcal{W}) = \frac{1}{|\mathcal{W}|} \sum_{w \in \mathcal{W}} \left(\nabla^2 J(m; w) + \alpha \nabla^2 J_{\text{reg}}(m) \right).$$

Now, we propose a trust-region Newton-CG method using subsamples of $\mathcal{W}$. Similar ideas have been applied to problems in machine learning [27]. We formulate Algorithm 6 in a general way and allow for different subsets of the samples for the objective function and gradient, on the one hand, as well as for the Hessian vector products, on the other hand. Here, we only consider the unconstrained and finite-dimensional case, but it can also be applied to the Moreau-Yosida penalized objective.

Algorithm 6 Inexact Newton with Mini-Batches

1: Choose initial model m^0 and samples $\mathcal{H}$, $\mathcal{H}_0$ with $\emptyset \neq \mathcal{H} \subseteq \mathcal{H}_0 \subseteq \mathcal{W}$.
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: Evaluate $j(m^k; \mathcal{H}_0)$ and $\nabla j(m^k; \mathcal{H}_0)$.
4: Invoke Algorithm 3 with $H_k = \nabla^2 j(m^k; \mathcal{H})$ and retrieve trial step s^k.
5: Invoke Algorithm 4 (and Algorithm 5, if necessary).
6: Choose new samples $\emptyset \neq \mathcal{H} \subseteq \mathcal{H}_0 \subseteq \mathcal{W}$.
7: **end for**

The computational costs of applying the Steihaug-CG method in Algorithm 6 are proportional to $|\mathcal{H}|$ instead of $|\mathcal{W}|$. Likewise, the memory requirements are reduced. Note that the condition $\mathcal{H} \subseteq \mathcal{H}_0$ is crucial, because for every $w \in \mathcal{H}$ the forward and adjoint wavefields are required to compute $\nabla^2 j(m; w)$. Thus, the corresponding state and adjoint equations have to be solved anyways. Note that the optimality conditions for (5.2) can not be checked using only a subset $\mathcal{H}_0 \subset \mathcal{W}$. Hence, including a convergence criterion into Algorithm 6 necessitates to compute $\nabla j(m; \mathcal{W})$.

The set $\mathcal{W}$ in Algorithm 6 might be interpreted in two different ways. On the one hand, we can choose $\mathcal{W} = \{e^1, \ldots, e^{n_s}\}$ and apply the method to the conventional problem formulation. This is similar to problems in machine learning, where parameters of a statistical model are estimated based on a small training set. Following the original idea of the super-shots, on the other hand, $\mathcal{W}$ is not necessarily fixed a priori, as we can choose an arbitrary number of samples of a probability distribution W. From this point of view, Algorithm 6 is similar to stochastic descent methods that are outlined in section 5.4.

Now, we evaluate the practical performance of the proposed method by reconsidering the elastic Marmousi example and using the same setup. Algorithm 6 offers a lot of flexibility to choose the mini-batches. We consider either $K = 8$ or $K = 16$ super-shots and choose $\mathcal{H}_0 = \mathcal{W}$ in every iteration. The Hessian-vector products during the CG iterations are computed using only a single super-shot. Here, a different super-shot is chosen cyclical from 1 to K in every Newton iteration. In both cases, the iterates converge to the same points as before. Table 5.4 summarizes the computational effort for 8 and 16 super-shots. Due to the approximation of the Hessian the number of Newton iteration increases, which is expected. Furthermore, the number of rejected steps during the trust-region iteration increases from 2 to 8 ($K = 8$) and from 1 to 6 ($K = 16$). However, and most importantly, compared to the "full Hessian" the number of PDE simulations reduces considerably, even for a higher tolerance. Here, we achieve savings of 56% ($K = 8$) and 75% ($K = 16$), respectively. In a second test, we increase the number of super-shots considered for computing the objective and gradient along the way and continue to approximate the Hessian using only one super-shot. Here, the results do not further improve. In contrast, increasing K from 1 to 8 results in a total number of 7984 PDEs (compared to 5058) and increasing K from 8 to 16 requires 8088 simulations (compared to 6112).

Due to the same structure of problem (P) and (5.2), it is straightforward to incorporate additional constraints. To this end, we perform another test with the Marmousi model and add lower bounds on the P-wave velocity. In particular, we impose $\lambda \geq 1.4 \cdot 10^9$ in the whole domain. Due to the constant Poisson's ratio this is equivalent (with minor rounding)

K	tol = 10^{-3}			tol = 10^{-6}		
	it Newton	avg. cg it	# PDEs	it Newton	avg. cg it	# PDEs
8	45	17.8	3090	65	23.2	5058
16	42	17.2	3700	62	22.5	6112

Table 5.4: Comparison of the computational effort for different number of super-shots and different tolerances on the optimality conditions. The Hessian is approximated by only one super-shot as outlined in Algorithm 6.

K ($\|\mathcal{H}\|$)	tol = 10^{-3}			tol = 10^{-6}		
	it Newton	avg. cg it	# PDEs	it Newton	avg. cg it	# PDEs
8 (8)	25	17.9	8264	33	23.3	13704
8 (1)	50	19.6	3440	66	24.0	5092
16 (1)	46	20.7	4384	65	24.4	6358

Table 5.5: Comparison of the computational effort for the Marmousi model with lower bounds. The number in brackets in the first column indicate the number of super-shots that are used to compute the Hessian-vector products of the misfit functional.

to $v_p \geq 1450[m/s]$. This bound is only relevant near the surface as the velocities increase with depth. Note that the true Marmousi model is feasible with respect to this bound.

In a first scenario, we solve the constrained inverse problem using 8 super-shots for the approximation of the objective, gradient and Hessian-vector products. In addition, we use Algorithm 6 with the same setting as before and compute solutions to the constrained problem with $K = 8$ and $K = 16$. In shallow depth up to 100m there is a slight improvement of the reconstruction for 8 super-shots compared to the unconstrained case, where the maximum pointwise error is reduced from 0.22094 to 0.22068. In the rest of the domain the solutions are the same. In addition, there is a small improvement of the final misfit to 0.438278. The regularization parameters were chosen identically in all scenarios with $\alpha_1 = 10^{-8}$ and $\alpha_1 = 5 \cdot 10^{-5}$. For $K = 16$ the reconstructions in the constrained and unconstrained cases are identical, although 29 of the iterates are infeasible material models which triggers an increase of γ.

Finally, Table 5.5 summarizes the computational effort to solve the constrained problems using the Moreau-Yosida regularization. The number of required PDE simulations increased only about 0.7% ($K = 8$) or 4% ($K = 16$), respectively, compared to the unconstrained case.

Summing up, this test shows a nice application of constraints where the solution can be slightly improved by adding prior knowledge to the problem formulation. More importantly, the computational effort for solving the constrained problem does not increase significantly. This is also visualized in Figure 5.7, which shows that in both test cases the mini-batch Hessian based on a single super-shot is more efficient - even to get from tolerance 10^{-3} to 10^{-6}.

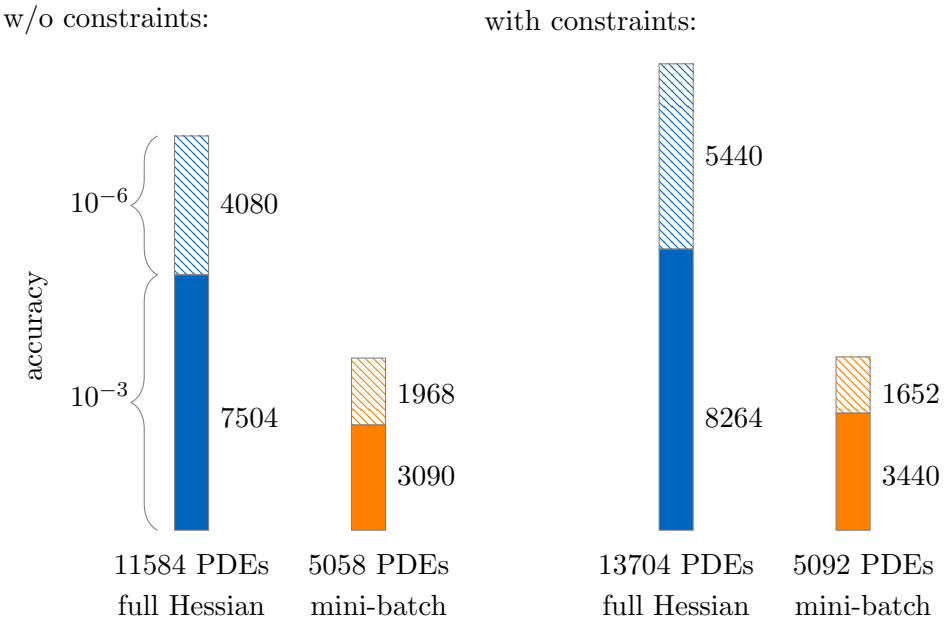


Figure 5.7: Summary of the computational effort to solve the unconstrained and constrained inverse problem for $K = 8$. Even for higher accuracies, the inexact Hessian based on a single super-shot is more efficient.

5.4 Related Approaches and Further Perspectives

Stochastic Descent Methods

If K and the random weights are fixed, the sample average approximation approach yields a deterministic optimization problem that can be solved with any choice of nonlinear optimization method, for instance Algorithm 6. However, since $\mathcal{W}$ is not necessarily fixed a priori and samples can be chosen randomly in every iteration, the proposed method is closely related to stochastic descent methods.
In particular, the stochastic gradient descent method (also known as Robbins-Monro method [111]) is based on the update rule

$$m^{k+1} = m^k - \sigma \nabla j(m^k; \mathcal{W}_k),$$

where $\sigma > 0$ is a step size and $\mathcal{W}_k$ is a small number of samples drawn from the distribution W that changes in every iteration. Again, we can also allow for randomly picked individual sources. We refer to [113], section 5.9, for a general overview on the stochastic gradient descent method. Note, however, that most of the theoretical findings are limited to convex problems. However, even in the non-convex case, the method often yields good results in practical applications [61]. A key observation is that the performance of stochastic descent methods can be accelerated by using a preconditioner, i.e., by modifying the update rule to

$$m^{k+1} = m^k - \sigma B^{-1} \nabla j(m^k; \mathcal{W}_k),$$

with a symmetric positive definite matrix B approximating the Hessian at the solution $\bar{m}$. Recently, limited-memory BFGS (L-BFGS) methods [97] as well as variants using Hessian-vector

products based on mini-batches to construct the L-BFGS approximation [28] for accelerating the stochastic gradient descent method have been proposed.
Despite being a very interesting topic, a rigorous comparison of stochastic descent methods and its relation to the sample average approximation is beyond the scope of this thesis and left for future research.

Source Stacking with Time Shifts

So far, all sources (possibly with weight zero) are triggered simultaneously. However, we can also introduce a temporal delay and trigger the sources with a time shift. To this end, we assume that all receiver locations observe continuously in time, i.e., for every event there exist measurements $u_i^\delta(\theta)$ for every $\theta \in \mathbb{R}$ with time θ referring to the global time scale. Furthermore, let $f_i : [0,T] \to V^*$, $i = 1, \dots, n_s$, denote the source time functions for the single events. Here, the length of the interval T indicates the duration for which significant measurements are observed. For simplicity, we assume that this duration is the same for all events. Note that f_i introduces an individual time scale for event i and initial time zero on this local scale does not mean that all events happen simultaneously. Instead, let θ_i denote the initial time of event i on the global time scale. Then, on the global time scale, the period with significant measurements for event i corresponds to $u_i^\delta(\theta)$ with $\theta \in [\theta_i, \theta_i + T]$.
Now, we extend the source stacking approach by allowing for additional time shifts. To this end, let $w \in \mathbb{R}^{n_s}$ and $\tau \in [0, (n_s - 1)T]^{n_s}$ denote weight and time shift for every source. We set

$$\bar{T} := T + \max_{i=1,\dots,n_s} \tau_i$$

and define the super-shot right-hand side with time shifts as

$$F_{w,\tau} : [0, \bar{T}] \to V^*, \quad F_{w,\tau}(t) = \sum_{i=1}^{n_s} \mathbf{1}_{\{\tau_i \le t \le \tau_i + T\}} w_i f_i(t - \tau_i).$$

Here, $\mathbf{1}$ denotes the indicator function. We recall the assumptions from the previous chapters that every seismic source f_i is continuous in time and has compact support in $[0,T]$. Hence, these properties also hold for $F_{w,\tau}$ on $[0,\bar{T}]$. The misfit is now computed using measurements $u_i^\delta(\theta)$ with $\theta \in [\theta_i - \tau_i, \theta_i + (\bar{T} - \tau_i)]$.
Clearly, the sample average approach presented previously can be interpreted as a special case of this strategy with the choice $\tau_i = 0$ for all i. Compared to the conventional approach with individual events, the number of time steps would be reduced by a fraction of $\bar{T}/(n_s T)$. Since the number of time steps are roughly proportional to the overall costs of simulating all seismic events, the same fraction applies to the reduction of the total costs.
In summary, allowing for time shifts might help to reduce the loss of information by shifting the arrivals of particular wavefronts and still provide significant savings of computational time.

Optimal Experimental Design to Determine the Weights

In all the numerical examples presented above, the weights were chosen as i.i.d. samples of Rademacher's distribution, which is motivated by the fact that this distribution is optimal in the sense that it minimizes the variance of $\mathbb{E}[J(m;w)]$ among all distributions with mean zero and the identity matrix as covariance [75]. Despite this observation, it is not clear, however,

whether a different choice of the weights could either provide a better descent direction or result in fewer iterations to solve the problem. More generally, while the computational effort using super-shots can be reduced substantially, it is also unknown, how much information is lost by the sample average approximation, especially in the presence of very noisy measurements. Hence, it is desirable to determine the weights such that as much information as possible is preserved. Therefore, criteria to quantify the quality of information are required. Related literature exists for linear tomography problems [38, 60] as well as a preliminary study for the acoustic wave equation [117]. Moreover, general considerations on optimal experimental design of inverse problems can be found in [62]. Many of these criteria depend on the spectrum of the Hessian of the sample average approximation. In order to apply these methods, representing the second derivatives by adjoint techniques will be a crucial ingredient.

Chapter 6

Wave Propagation at a Solid-Fluid Interface

Complementing the results from the previous chapters, we conclude this thesis with a case study on tomography problems with a domain that has both fluid and solid media. To this end, we present a model that consists of a coupled system of the elastic and the acoustic wave equation. Hereby, a fundamental assumption is that the location of the interface between both media is known a priori. Afterwards, we apply the framework for solving tomography problems that we have developed in the previous chapters. However, a rigorous analysis of the coupled system in a function space setting is beyond the scope of this section.

Most of the results of this chapter have been published in [18].

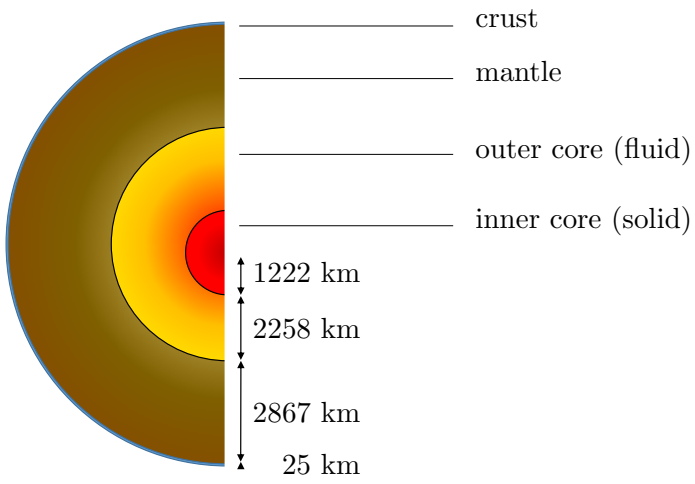


Figure 6.1: Sketch of the inner structure of the Earth with a solid mantle and inner core and a fluid outer core [46].

6.1 Motivation

So far, we have treated fluid and solid media separately, considering either the elastic or the acoustic wave equation. However, in several applications the domain of interest has both solid and fluid regions which requires appropriate modeling. These problems occur on very different scales, for instance, in global seismic tomography where the liquid structure of the Earth's outer core has to be considered, see Figure 6.1.
On a much smaller scale, problems in marine geophysical exploration require an accurate model of the ocean. An important application is the search for natural resources, especially oil reservoirs that are located below the sea. Here, measurements are generated by a research vessel that emits pressure waves with an airgun. Figure 6.2 sketches the setup for the collection of reflection and refraction surveys. Geophones buried into the shallow seafloor record the seismic response from the subsurface. The Valhall oil field in the North Sea is an example for such an industrial data set with ongoing research [107, 112].

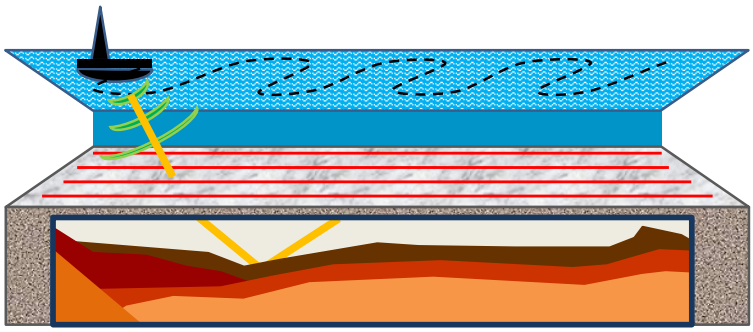


Figure 6.2: Sketch of marine geophysical exploration: A research vessel equipped with an airgun cruises the sea and emits pressure waves. The seismic response of reflected and refracted waves is recorded by a dense array of geophones located at the bottom of the ocean or buried into the shallow seafloor (shown as red lines).

6.2 Modeling and Inverse Problem

We consider a domain that consists of a solid and a fluid layer. The solid and fluid regions are denoted by Ω_S and Ω_F, respectively, and we set $\Omega = \Omega_S \cup \Omega_F \subset \mathbb{R}^d$ with $d = 2, 3$. Throughout this chapter, we assume that Ω_S and Ω_F are bounded domains with smooth boundaries and interface Γ_{int}. The remaining parts of the boundaries are denoted by $\Gamma_F := \partial\Omega_F \setminus \Gamma_{\text{int}}$ and $\Gamma_S := \partial\Omega_S \setminus \Gamma_{\text{int}}$, respectively. Figure 6.3 shows the geometry of the domain with both layers.

As has been discussed in the previous chapters, the propagation of waves in the solid medium is governed by the elastic wave equation and, respectively, by the acoustic wave equation in the fluid domain. In addition, continuity of traction and continuity of the normal displacement have to be ensured at the interface, cf. [34]. In the solid medium, we assume a heterogeneous and positive density ρ_S and a linear elastic rheology and continue to denote the displacement field by u. In the fluid domain, we consider an inviscid fluid medium with a homogeneous

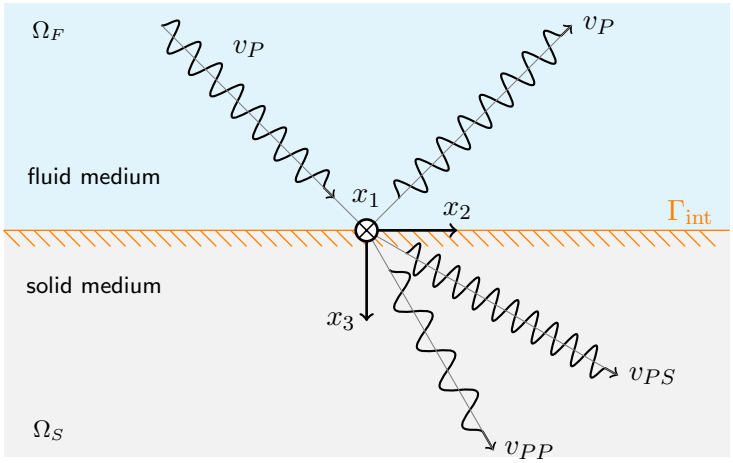


Figure 6.3: Sketch of a solid-fluid interface. Parts of the incoming waves are reflected to the acoustic medium while other parts are transmitted to the elastic medium.

density $\rho_F > 0$. In order to better distinguish both regions, we introduce a new state variable χ to denote the displacement potential. The speed of compressional waves in the fluid domain is assumed to be constant and denoted by $c > 0$. Furthermore, the normal vector pointing outwards of the fluid domain is denoted by $\vec{n}_F$ and the normal vector pointing outwards of the solid domain by $\vec{n}_S$, respectively.
In order to cover the more general case we place seismic sources f_F and f_S into both media. Note that sources in the acoustic medium are pressure sources. The complete coupled system in strong form is now given by:

$$\begin{cases} \rho_F \chi_{tt} - \rho_F c^2 \Delta \chi &= f_F & \text{on } \Omega_F \times I, \\ \chi(0) = 0, \quad \chi_t(0) &= 0 & \text{on } \Omega_F, \\ \chi &= 0 & \text{on } \Gamma_F \times I, \\ \\ \rho_S u_{tt} - \nabla \cdot (\Psi(m) : \varepsilon(u)) &= f_S & \text{on } \Omega_S \times I, \\ u(0) = 0, \quad u_t(0) &= 0 & \text{on } \Omega_S, \\ (\Psi(m) : \varepsilon(u)) \cdot \vec{n}_S &= 0 & \text{on } \Gamma_S \times I, \\ \\ (\Psi(m) : \varepsilon(u)) \cdot \vec{n}_S &= -\rho_F \chi_{tt}\, \vec{n}_F & \text{on } \Gamma_{\text{int}} \times I, \\ -\vec{n}_F \cdot \nabla \chi &= \vec{n}_S \cdot u & \text{on } \Gamma_{\text{int}} \times I. \end{cases} \tag{6.1}$$

Here, we work with the same parameterization as in Chapter 3, i.e.,

$$\Psi(m) = \overline{\Psi} + \Phi(m).$$

Remark 6.2.1.
Since we focus on marine geophysical exploration, we assume that the fluid medium is known with a constant density and a constant velocity of compressional waves. Note however, that we could easily extend the problem formulation using the results from section 2.5 to invert for the acoustic material as well. This would be interesting for problems in global seismic tomography that seek to identify the structure of the Earth's outer core.

Remark 6.2.2.
Note that the model presented above requires exact knowledge of the location of the interface between the solid and the fluid medium. Again, this might not be the case for problems in global seismic tomography. We refer to [36] for a discussion of techniques to identify the location of such interfaces. In fact, this is an inverse problem itself.

In the next step, we turn to the weak form of (6.1). Let $V_S := H^1(\Omega_S)^d$ and, furthermore, let $V_F := H_0^1(\Gamma_F, \Omega_F)$ denote the space of all H^1-functions that vanish on Γ_F. To shorten the notation, we define the following bilinear forms:

$$\begin{aligned}
&a_F : V_F \times V_F \to \mathbb{R}, && a_F(v,w) := \rho_F c^2 \left(\nabla v, \nabla w\right)_{L^2(\Omega_F)^d}, \\
&a_S(\Psi(m)) : V_S \times V_S \to \mathbb{R}, && a_S(\Psi(m))(v,w) := \left(\Psi(m) : \varepsilon(v), \varepsilon(w)\right)_{L^2(\Omega_S)^{d\times d}}, \\
&a_{\text{int}} : L^2(\Omega_F) \times L^2(\Omega_S)^d \to \mathbb{R}, && a_{\text{int}}(v,w) := \rho_F \int_{\Gamma_{\text{int}}} v(x)\ w(x)\cdot \vec{n}_S(x)\, dS(x).
\end{aligned}$$

Here, the normal vector $\vec{n}_S$ in the definition of a_{int} is pointing outwards of the solid and into the fluid domain. The variational form of (6.1) can be written as:
$\forall\ v \in V_F, \forall\ w \in V_S$ and a.a. $t \in I$:

$$\begin{aligned}
\rho_F \langle \chi_{tt}(t), v\rangle_{V_F^*, V_F} + a_F(\chi(t), v) + c^2\, a_{\text{int}}(v, u(t)) &= \langle f_F(t), v\rangle_{V_F^*, V_F}, \\
\langle \rho_S\, u_{tt}(t), w\rangle_{V_S^*, V_S} + a_S(m)(u(t), w) - a_{\text{int}}(\chi_{tt}(t), w) &= \langle f_S(t), w\rangle_{V_S^*, V_S}.
\end{aligned} \tag{6.2}$$

Analogously to the analysis in Chapter 3, we define the spaces

$$\begin{aligned}
\mathcal{X} &:= L^2(I; V_F) \cap H^1(I; L^2(\Omega_F)) \cap H^2(I; V_F^*), \\
U &:= L^2(I; V_S) \cap H^1(I; L^2(\Omega_S)^d) \cap H^2(I; V_S^*).
\end{aligned}$$

Furthermore, we set $Y = \mathcal{X} \times U$ and let $y = (\chi, u)$ denote the state that consists of the displacement potential in the fluid domain and the displacement field in the solid domain. The following notation will be useful. For an arbitrary element $y \in Y$ we split $y = (y^1, y^2)$ with $y^1 \in \mathcal{X}$ and $y^2 \in U$. With $f = (f_F, f_S)^T$, the weak form of the coupled system (6.1) can be written as:

$$E(y,m) = f \quad :\Leftrightarrow (y,m) \text{ satisfies (6.2) a.e. in } I,$$

and, additionally, the initial conditions $y(0) = 0$ and $y_t(0) = 0$ have to be satisfied.
Now, we state the seismic inverse problem for the coupled system. To this end, we consider n_s seismic events with sources $f_i = (f_F)_i$ in the fluid domain. Then, the coupled inverse problem can be stated as follows:

$$\min_{\mathbf{y}\in\mathbf{Y},\ m\in M} J(\mathbf{y}, m) \quad \text{s.t.} \quad E(y_i, m) = \begin{pmatrix} f_i \\ 0 \end{pmatrix} \quad 1 \le i \le n_s, \quad \begin{pmatrix} \mathbf{y}(0) \\ \mathbf{y}_t(0) \end{pmatrix} = 0. \tag{6.3}$$

Here, $\mathbf{y} = (y_i)_{1\le i\le n_s}$ denotes a vector of states for different seismic events and $\mathbf{Y} := Y^{n_s}$. Note that the state y_i only enters into the i-th component of E, while the parameters m are the same for all components. Clearly, (6.3) has the same structure as the seismic inverse problem from Chapter 3. We consider cost functions $J : \mathbf{Y} \times M \to \mathbb{R}$ of the form:

$$J(\mathbf{y}, m) = \sum_{i=1}^{n_s} J_{\text{fit},i}(y_i) + \alpha J_{\text{reg}}(m),$$

i.e., we allow for observations in both solid and fluid media. In the numerical example in this section, however, we will consider data observations u_i^δ only in the solid domain, i.e., on $\Omega_i \times I$ with $\Omega_i \subset \Omega_S$. Hence, the misfit term $J_{\text{fit},i}$ and the regularization term J_{reg} are given by

$$J_{\text{fit},i}(y_i) = \frac{1}{2}\|u_i - u_i^\delta\|^2_{L^2(\Omega_i \times I)}, \quad J_{\text{reg}}(m) = \frac{1}{2}\|m\|^2_M.$$

Note that a rigorous analysis of the coupled system regarding existence and uniqueness of solutions is beyond the scope of this thesis. Hence, we follow a discretize-then-optimize approach and consider only the finite-dimensional version of (6.3).

6.3 Adjoint Equation

Similar as before, the state variables for different seismic events can be separated in (6.3). Hence, the same holds true for the adjoint states and we restrict this section to the analysis of a single seismic event and drop the index i.
We formally derive the adjoint equation and define $L(y, m, z) : Y \times M \times Y \to \mathbb{R}$ with the adjoint state $z = (z^1, z^2)$ as:

$$\begin{aligned} L(y,m,z) \;:=\; & J(y,m) + \int_0^T \rho_F \langle y^1_{tt}(t), z^1(t)\rangle_{V_F^*,V_F}\, dt + \int_0^T a_F(y^1(t), z^1(t))\, dt \\ & + \int_0^T \langle \rho_S\, y^2_{tt}(t), z^2(t)\rangle_{V_S^*,V_S}\, dt + \int_0^T a_S(\Psi(m))(y^2(t), z^2(t))\, dt \\ & - \int_0^T a_{\text{int}}(y^1_{tt}(t), z^2(t))\, dt + \int_0^T c^2\, a_{\text{int}}(z^1(t), y^2(t))\, dt \\ & - \int_0^T \langle f_S(t), z^2(t)\rangle_{V_S^*,V_S}\, dt - \int_0^T \langle f_F(t), z^1(t)\rangle_{V_F^*,V_F}\, dt \\ & - \rho_F(y^1(0), z^1_t(0))_{L^2(\Omega_F)} + \rho_F(y^1_t(0), z^1(0))_{L^2(\Omega_F)} \\ & - (\rho_S\, y^2(0), z^2_t(0))_{L^2(\Omega_S)^d} + (\rho_S\, y^2_t(0), z^2(0))_{L^2(\Omega_S)^d}. \end{aligned} \tag{6.4}$$

Thus, the adjoint equation is given by

$$L_y(y(m), m, z) = 0, \tag{6.5}$$

where $y(m)$ denotes the state corresponding to m. By considering the variational form of (6.5) and carefully integrating by parts with respect to time, we obtain the adjoint equation for given m and $y(m)$ as

$$E^{\text{ad}}(z, m) = -J_y(y(m), m), \quad z(T) = 0, \quad z_t(T) = 0. \tag{6.6}$$

With the assumption of sufficient regularity, the adjoint equation E^{ad} in strong form can be interpreted as a coupled system like (6.1) backwards in time with final time instead of initial conditions, interchanged interface conditions and a different right-hand-side. In case of measurements in both fluid and solid domain, the adjoint source $-J_y(y(m), m)$ will have support in both media as well.
Now, we turn to the adjoint-based representation of first and second derivatives. To this end, let $z(m)$ denote the adjoint state for given m, which we assume to be uniquely determined by

(6.6). It is important to note that the material parameters m do not appear in the interface conditions. Now, we introduce the form $D_{SF} : M \times Y \times Y \to \mathcal{L}(M, \mathbb{R})$ analogously to section 3.2.2 by

$$D_{FS}(m, v, w)(s) = \int_0^T \int_{\Omega_S} \left(\varepsilon(v^1)(x,t) \otimes \varepsilon(w^1)(x,t)\right) :: \left(\left(\Phi'(m)s\right)(x)\right) \, dx \, dt \quad \forall \, s \in M,$$

which enables us to express the first derivatives as

$$j'(m) = D_{FS}(m, y(m), z(m)) + \alpha J'_{\text{reg}}(m). \tag{6.7}$$

Furthermore, we note that the weak form (6.2) depends linearly on y and m. Thus, the derivation of the second derivatives for the coupled system is straightforward. We obtain:

$$\begin{aligned}
L_{yy}(y, m, z)(\hat{y}_1, \hat{y}_2) &= J_{yy}(y, m)(\hat{y}_1, \hat{y}_2), \\
L_{ym}(y, m, z)(\hat{m}, \hat{y}) &= \int_0^T \left((\Psi'(m)\hat{m}) : \varepsilon(\hat{y}^2)(t), \varepsilon(z^2)(t)\right)_{L^2(\Omega_S)^{d\times d}} \, dt, \\
L_{my}(y, m, z)(\hat{y}, \hat{m}) &= \int_0^T \left((\Psi'(m)\hat{m}) : \varepsilon(\hat{y}^2)(t), \varepsilon(z^2)(t)\right)_{L^2(\Omega_S)^{d\times d}} \, dt, \\
L_{mm}(y, m, z)(\hat{m}_1, \hat{m}_2) &= \int_0^T \left((\Psi''(m)(\hat{m}_1, \hat{m}_2)) : \varepsilon(y^2(t)), \varepsilon(z^2(t))\right)_{L^2(\Omega_S)^{d\times d}} \, dt.
\end{aligned}$$

In particular, we note that the derivatives have a very similar structure as in the purely elastic case. Therefore, operator-vector products $j''(m)s$ for a given perturbation $s \in M$ can be computed at the cost of two additional simulations of the coupled system.
With the adjoint-based representation of the first derivatives and operator-vector products representing the second derivatives applied to a search direction, we have everything at hand to apply the trust-region Newton-PCG method presented in section 3.5.

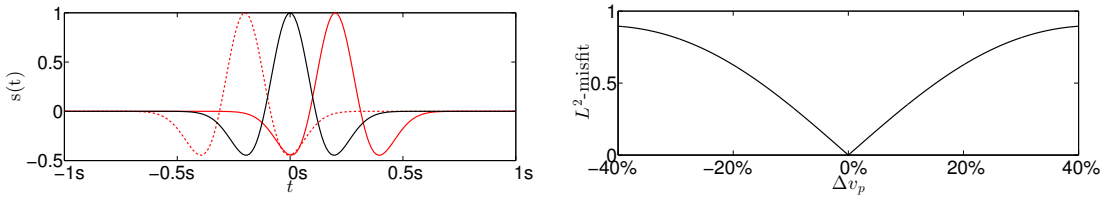
Remark 6.3.1.
Following the analysis from Chapter 3, it would be straightforward to include constraints on the material as well. However, we restrict the presentation to the unconstrained case and utilize the superposition operator Φ to ensure positive wave velocities during the inversion.

We continue with a strategy to deal with the difficulty of having several local minima which is caused by the non-convexity of the problem. This poses a severe challenge when solving the inverse problem and has not been addressed comprehensively so far in this thesis.

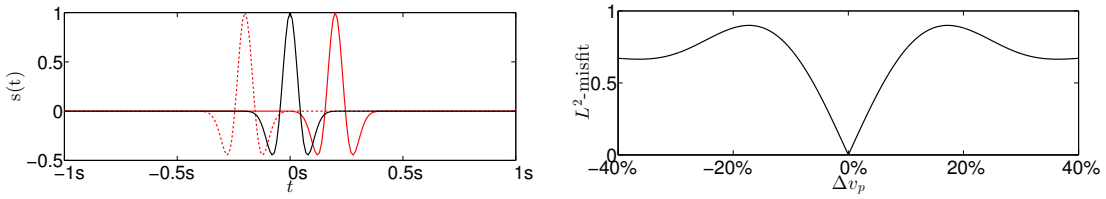
6.4 Multi-Frequency Inversion and Goal-Oriented Adaptivity

High-resolution reconstructions of the material properties require high-frequency information in the observed data. This introduces several challenges. For increasing frequencies the data is more prone to errors induced by noisy measurements. Furthermore, it can easily be observed that the threat of several local minima increases with frequencies. This is shown in Figure 6.4 which considers a Ricker wavelet recorded at 5km distance from the source. We use different source frequencies of 2Hz, 5Hz and 10Hz. The arrival of the signal might be premature or

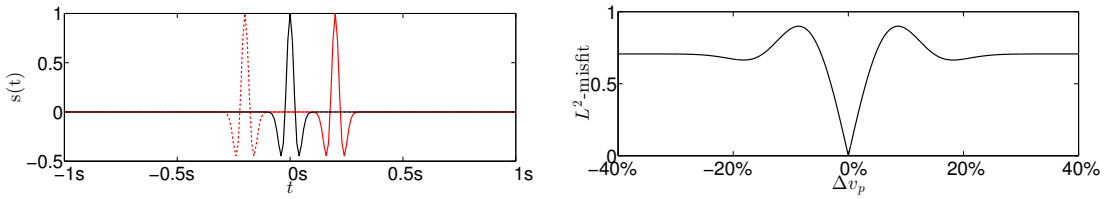
delayed if the underlying material has a higher or lower velocity, respectively. We depict the L^2-misfit as a function of the difference in P-wave velocity in a homogeneous medium, indicating that the basis of attraction around the global minimum shrinks with higher frequencies. In other words, incorporating higher frequency data during the inversion necessitates an initial model of increasing quality.



(a) source frequency 2Hz



(b) source frequency 5Hz



(c) source frequency 10Hz

Figure 6.4: Frequency-dependence of the misfit. The images on the left show the Ricker wavelet for three different dominant frequencies (2Hz, 5Hz, 10Hz) in black as well as a shifted arrivals of $\pm$ 0.2s in red. On the right-hand side, we depict the resulting L^2-misfit as a function of the difference in P-wave velocity Δv_p. Clearly, the basin of attraction is larger for lower source frequencies.

Even when more sophisticated misfit criteria are used, this problem is not resolved to full extend. This observation suggests a multi-frequency approach that sequentially inverts for increasing source frequencies and reuses the reconstructions for lower frequencies as improved initial material model for the subsequent inversions. Hereby, a bandpass filter can be used to down-sample the observed measurements to lower frequencies. This approach, which is sometimes also referred to as multi-scale inversion, has been proven effective in empirical studies [25, 54], see also [48].
As another important advantage of this approach, inverting for lower source frequencies can be carried out on a coarser mesh, since the discretization of the state depends on the wavelengths that should be resolved. Due to the CFL condition, this affects spatial and temporal discretization and provides significant savings.

The procedure works as follows:

0. Choose an initial parameter mesh.

For a sequence of increasing source frequencies $\omega_1 \leq \omega_2 \leq \ldots$

1. Choose the state mesh based on the dominant frequency ω_i of the seismic source and the wave velocities.

2. Solve the discretized problem to a specified tolerance and reuse the reconstruction as the next initial model.

While the state mesh depends on the velocities and frequency, it is not clear, how the parameter mesh should be discretized during the multi-frequency inversion. For a general discussion of discretization techniques for inverse problems and adaptive mesh refinement, we refer to [11]. One option is to follow a regularization-by-discretization strategy and combine the multi-frequency inversion approach with an adaptive grid refinement based on goal-oriented error estimates. To this end, we add this third step to the procedure above:

3. Adaptively refine the parameter mesh using goal-oriented error estimators.

The adaptively refined grid allows to reduce the number of optimization variables without a loss of resolution in the reconstruction. In order to capture the material heterogeneities, the cells of the parameter mesh cannot be smaller than the cells of the state mesh. The process is sketched in Figure 6.5.

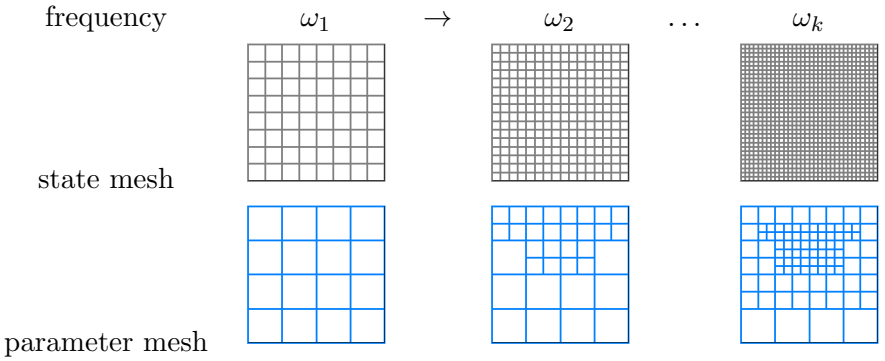


Figure 6.5: Multi-frequency inversion with goal-oriented adaptivity. For a sequence of increasing source frequencies the parameter mesh is adaptively refined. The state mesh is uniformly refined based on the source frequency and wave velocities.

In the following, we briefly summarize the main ideas of a posteriori error estimation and goal-oriented adaptivity following the comprehensive analysis in [14]. In addition, we motivate why the grid refinement can be carried out already for an inexact solution to the coarse problem. Further studies on goal-oriented adaptivity for problems governed by parabolic or hyperbolic PDEs can be found, for instance, in [85, 94]. The main result for a posteriori error estimation of the discretization error is given as follows, cf. [14]:

Theorem 6.4.1.
Let X be a Banach space and $L : X \to \mathbb{R}$ a three times Gâteaux differentiable functional. Furthermore, let $X_1 \subset X$ be a subspace of X and $\bar{x}_1 \in X_1$ a stationary point of L on X_1, i.e.,

$$L'(\bar{x}_1)(v) = 0 \quad \forall\, v \in X_1.$$

Now, consider an approximation of this equation by a Galerkin method in a subspace $X_2 \subset X$ and a stationary point $\bar{x}_2 \in X_2$ that satisfies

$$L'(\bar{x}_2)(v) = 0 \quad \forall\, v \in X_2.$$

If, in addition, $\bar{x}_1$ satisfies

$$L'(\bar{x}_1)(\bar{x}_2) = 0, \tag{6.8}$$

then the difference $L(\bar{x}_1) - L(\bar{x}_2)$ can be represented in the form

$$L(\bar{x}_1) - L(\bar{x}_2) = \frac{1}{2} L'(\bar{x}_2)(\bar{x}_1 - \hat{x}_2) + R$$

with an arbitrary $\hat{x}_2 \in X_2$ and a cubic remainder term R given by

$$R = \frac{1}{2} \int_0^1 L'''(\bar{x}_2 + se)(e, e, e) s(s-1)\, ds,$$

where $e := \bar{x}_1 - \bar{x}_2$.

Proof. See [14], Proposition 2.1. □

Remark 6.4.2.
If $X_2 \subset X_1$, then (6.8) is obviously satisfied by the stationarity of x_1 in X_1.

Following the lines of the proof of Theorem 6.4.1 in [14], Proposition 2.1, we directly obtain the following variant.

Theorem 6.4.3.
Let X be a Banach space and $L : X \to \mathbb{R}$ a three times Gâteaux differentiable functional. Furthermore, let $X_1 \subset X$ be a subspace of X and $\bar{x}_1 \in X_1$ a stationary point of L on X_1, i.e.,

$$L'(\bar{x}_1)(v) = 0 \quad \forall\, v \in X_1.$$

Furthermore, we consider a subspace $X_2 \subset X$ with $x_2 \in X_2$ and assume $\bar{x}_1$ satisfies

$$L'(\bar{x}_1)(x_2) = 0.$$

Then the difference $L(\bar{x}_1) - L(x_2)$ can be represented in the form

$$L(\bar{x}_1) - L(x_2) = \frac{1}{2} L'(x_2)(\bar{x}_1 - x_2) + R$$

with a cubic remainder term R given by

$$R = \frac{1}{2} \int_0^1 L'''(x_2 + se)(e, e, e) s(s-1)\, ds,$$

where $e := \bar{x}_1 - x_2$.

Proof. With $e := \bar{x}_1 - x_2$, we obtain

$$\begin{aligned} L(\bar{x}_1) - L(x_2) &= \int_0^1 L'(x_2 + se)(e)\, ds \\ &= \int_0^1 L'(x_2 + s(\bar{x}_1 - x_2))(e)\, ds + \frac{1}{2}L'(x_2)(e) - \frac{1}{2}L'(x_2)(e) - \frac{1}{2}L'(\bar{x}_1)(e), \end{aligned}$$

where we used

$$L'(\bar{x}_1)(e) = L'(\bar{x}_1)(\bar{x}_1) - L'(\bar{x}_1)(x_2) = 0.$$

Now, we can apply the trapezoidal rule for integration to obtain the remainder term:

$$\begin{aligned} &\int_0^1 L'(x_2+s(\bar{x}_1 - x_2))(e)\, ds - \frac{1}{2}L'(x_2)(e) - \frac{1}{2}L'(\bar{x}_1)(e) \\ &\qquad = \int_0^1 s(s-1)\; L'''(x_2 + se)(e,e,e)\, ds, \end{aligned}$$

which yields the error representation above. □

Remark 6.4.4.
The only difference between Theorem 6.4.1 and Theorem 6.4.3 is that the first derivative of L is applied to $\bar{x}_1 - x_2$ instead of $\bar{x}_1 - \hat{x}_2$ with an arbitrary $\hat{x}_2 \in X_2$. In return, x_2 does not have to be a stationary point of L in X_2.

In the context of optimal control or parameter identification problems, this error representation is applied to the Lagrangian function (6.4). To this end, consider $Y_h \subset Y$ and $M_h \subset M$ and let $(\bar{y}, \bar{m}, \bar{z}) \in Y \times M \times Y$ and $(\bar{y}^h, \bar{m}^h, \bar{z}^h) \in Y_h \times M_h \times Y_h$ be stationary points of the Lagrangian L for the continuous and discretized problem. Due to the stationarity of $(\bar{y}, \bar{m}) \in Y \times M$ and, respectively, $(\bar{y}^h, \bar{m}^h) \in Y_h \times M_h$ we have

$$J(\bar{y}, \bar{m}) = L(\bar{y}, \bar{m}, \bar{z}), \quad J(\bar{y}^h, \bar{m}^h) = L(\bar{y}^h, \bar{m}^h, \bar{z}^h).$$

Thus, applying Theorem 6.4.1 to L yields the estimate

$$\begin{aligned} J(\bar{y}, \bar{m}) - J(\bar{y}^h, \bar{m}^h) \approx{}& \frac{1}{2}L_y(\bar{y}^h, \bar{m}^h, \bar{z}^h)(\bar{y} - \hat{y}^h) \\ &+ \frac{1}{2}L_m(\bar{y}^h, \bar{m}^h, \bar{z}^h)(\bar{m} - \hat{m}^h) \\ &+ \frac{1}{2}L_z(\bar{y}^h, \bar{m}^h, \bar{z}^h)(\bar{z} - \hat{z}^h), \end{aligned} \tag{6.9}$$

with arbitrary $\hat{y}^h \in Y_h, \hat{m}^h \in M_h$ and $\hat{z}^h \in Y_h$.
The framework of goal-oriented adaptivity is typically applied to refine the mesh for the discretization of the state and adjoint equation based on the estimate (6.9). This has been carried out successfully for forward simulations of the wave equation [8] as well as for optimal

control problems involving the elastic wave equation [84, 85]. Note, however, that in both cases the material was homogeneous. In seismic tomography, the potential savings are probably smaller as the inhomogeneous medium results in many reflected waves with amplitudes at very different magnitudes that might be equally important for different misfit criteria. On a different note, integrating an adaptively refined spatial mesh, which changes over time, into the MPI parallelized code would be a lot more involved. Furthermore, a local time stepping scheme where the length of the time step is adjusted to the spatial size of the cell might be required.

For these reasons, we only consider the parameter mesh and assume that state and adjoint equation are solved exactly for $Y_h = Y$. Hence, we estimate the difference in the reduced cost functional as

$$j(\bar{m}) - j(\bar{m}^h) \approx \frac{1}{2} L_m(u(\bar{m}^h), \bar{m}^h, z(\bar{m}^h))(\bar{m} - m^h).$$

Here, the first two terms in (6.9) vanish by the assumption $Y_h = Y$.

The remaining difficulty is to approximate the unknown solution $\bar{m}$ to compute $\bar{m} - m^h$. Typically, this is done by a local interpolation in a higher-order finite element space. Assuming that the parameter mesh has a patch structure, we combine 2^d neighboring cells and define $\mathcal{I}_{2h}^2$ as the bi-/tri-quadratic interpolation on the macro cells. This is visualized in Figure 6.6 for 1d. The difference $\bar{m} - m^h$ is then approximated by $(\mathcal{I}_{2h}^2 - \mathrm{id})m^h$. Although there is no theoretical justification, this heuristic has been proven effective in many applications [8, 14].

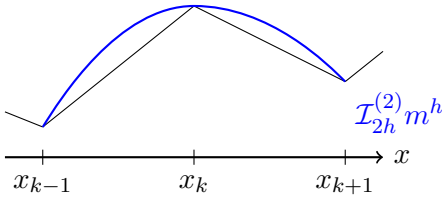


Figure 6.6: Linear and quadratic interpolation on element patches in 1d.

In particular, we observe that this heuristic does not utilize the fact that m^h can be chosen arbitrarily, which motivates to solve the reduced problem for M_h only inexactly and to apply Theorem 6.4.3 with the approximate stationary point m^h. Loosely speaking, this strategy assumes that the discretization error dominates the non-stationarity.

The considerations outlined above enable us to adaptively refine the parameter mesh during the multi-frequency inversion. The grid refinement is implemented using deal.ii [10].

It should be emphasized that there is no theoretical justification of the effectiveness of the refinement process. Moreover, since the state mesh is not affected, the computational savings are rather small. Therefore, we choose a high ratio of cells to be refined during the multi-frequency inversion. On the other hand, the error estimates are cheap to compute, especially as they do not require the solution of a PDE and involve only the parameter mesh. Furthermore, we expect some robustification in addition to the regularization term in the objective and hope to avoid artifacts in the reconstruction that are caused by an over-parameterization of the material, but not explained by the data.

6.5 Implementation

Due to the similar structure of the weak form as well as the adjoint equation and the representation of the derivatives, most of the implementation presented in Chapter 4 can be reused. In particular, we use fourth-order Lagrange-polynomials as shape functions and apply the Gauß-Lobatto-Legendre quadrature. In addition to mass matrix $\bar{\mathbf{M}}_{S/F}$, stiffness matrix $\bar{\mathbf{K}}_{S/F}$ and force vector $\bar{\mathbf{F}}_{S/F}$ in both media, we obtain an extra term from the interface conditions which acts on the set of elements that are adjacent to the interface. Similar as for the stiffness terms, these contributions are computed on the fly, but for notational convenience we denote this by a matrix $\bar{\mathbf{B}}_{S/F}$. The subscripts are used to indicate solid and fluid domain. For a complete derivation of the spatial discretization we refer to [33].
For the temporal discretization, we use the Newmark scheme as outlined in (4.2) with $\beta = 0$ and $\theta = 1/2$. The main important difference is introduced by the interface conditions that have to be carefully incorporated into the time-stepping scheme. Following [100], it can still be carried out fully explicit. To this end, let $\mathbf{u}_{k,0}, \mathbf{u}_{k,1}, \mathbf{u}_{k,2}$ denote the time-discrete approximations of $\mathbf{u}(t_k), \mathbf{u}_t(t_k)$ and $\mathbf{u}_{tt}(t_k)$ and, respectively, $\mathbf{x}_{k,0}, \mathbf{x}_{k,1}, \mathbf{x}_{k,2}$ the time-discrete approximation of $\mathbf{x}$, the coefficient vector of the nodal basis representation of χ^h. Then the Newmark update formulas are given as follows:

$$\begin{aligned}
\mathbf{x}_{k+1,0} &= \mathbf{x}_{k,0} + \Delta t_k \mathbf{x}_{k,1} + \frac{1}{2}\Delta t_k^2 \mathbf{x}_{k,2}, \\
\mathbf{u}_{k+1,0} &= \mathbf{u}_{k,0} + \Delta t_k \mathbf{u}_{k,1} + \frac{1}{2}\Delta t_k^2 \mathbf{u}_{k,2}, \\
\mathbf{x}_{k+1,2} &= -\bar{\mathbf{M}}_F^{-1}\left(\bar{\mathbf{K}}_F \mathbf{x}_{k+1,0} + \bar{\mathbf{B}}_F \mathbf{u}_{k+1,0} - \bar{\mathbf{F}}_{F,k+1}\right), \\
\mathbf{u}_{k+1,2} &= -\bar{\mathbf{M}}_S^{-1}\left(\bar{\mathbf{K}}_S \mathbf{u}_{k+1,0} + \bar{\mathbf{B}}_S \mathbf{x}_{k+1,2} - \bar{\mathbf{F}}_{S,k+1}\right), \\
\mathbf{x}_{k+1,1} &= \mathbf{x}_{k,1} + \frac{1}{2}\Delta t_k \left(\mathbf{x}_{k,2} + \mathbf{x}_{k+1,2}\right), \\
\mathbf{u}_{k+1,1} &= \mathbf{u}_{k,1} + \frac{1}{2}\Delta t_k \left(\mathbf{u}_{k,2} + \mathbf{u}_{k+1,2}\right).
\end{aligned} \tag{6.10}$$

Hereby, the order of the update is important, as $\mathbf{x}_{k+1,2}$ is required for updating $\mathbf{u}_{k+1,2}$. The order of $\mathbf{u}$ and $\mathbf{x}$ has to be interchanged for the adjoint equation.

6.6 Numerical Example

Now, we present a numerical example that is inspired by the Valhall oil field in the North Sea. Prior work on this field can be found in [22, 107, 114]. This example has also been published in [18]. The geometry is given by a rectangular domain of 8km × 8km × 4km. There is a layer of water on top of the solid domain with a constant depth of 400m. We use 36 seismic sources that are triggered simultaneously in the fluid region at 200m depth. The source time function for all sources is a Ricker wavelet with dominant frequency of 2.5Hz. There are 441 seismic receivers buried into the seafloor at 50m depth that form a dense array of 16km^2 in the center of the domain. In the fluid domain, we set $\rho_F = 1000\text{kg/m}^3$ and $c = 1500\text{m/s}$. In the solid domain, we assume a constant density of 2300kg/m^3 and a constant Poisson's ratio of 0.25, i.e., we have the relation $v_p = \sqrt{3}v_s$ for the velocity of compressional and shear waves and only one parameter field to invert for. The synthetic target model has P-wave velocities that range from 1400m/s to 3400m/s, see Figure 6.7 for a vertical profile. Data is generated

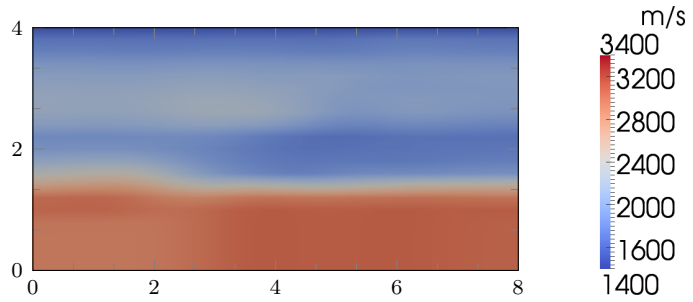
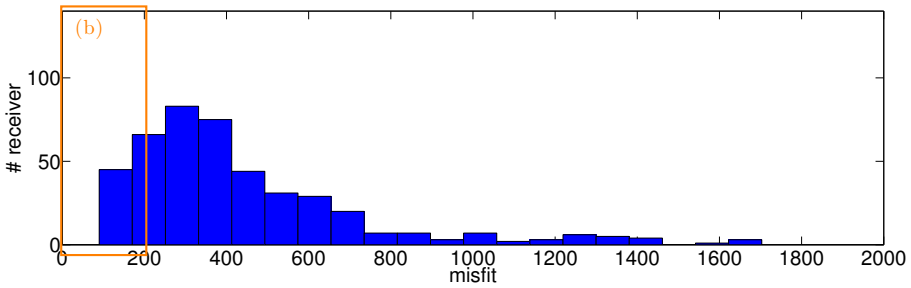


Figure 6.7: Vertical cross section of the P-wave velocity of the target model.

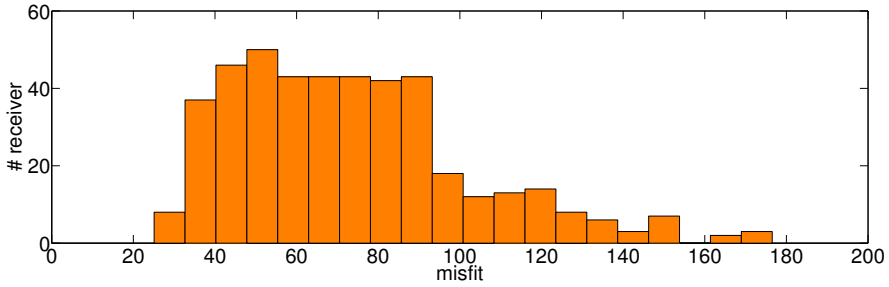
by running a simulation with the target model. In order to deal with the difficulties in using synthetic data and their potential for committing inverse crimes [77], we use a finer mesh for both, the parameter and the state space and add 2% Gaussian noise to the seismograms. The reference model varies only vertically and for every fixed depth, we use the average value of the target model in the horizontal plane as the reference value.
Similar as before, we have to impose absorbing boundary conditions at the artificial boundaries $\Gamma_S^{\text{abs}} \subset \partial\Omega_S$ and $\Gamma_F^{\text{abs}} \subset \partial\Omega_F$ of the computational domain:

$$\begin{aligned} \nabla\chi \cdot \vec{n} &= -c^{-1}\chi_t && (x,t) \in \Gamma_F^{\text{abs}} \times I, \\ (\Psi : \varepsilon(u)) \cdot \vec{n} &= v_p \rho_S \left(u_t \cdot \vec{n}\right) \vec{n} + v_s \rho_S \left(u_t - \left(u_t \cdot \vec{n}\right) \vec{n}\right) && (x,t) \in \Gamma_S^{\text{abs}} \times I. \end{aligned}$$

Furthermore, we enforce $\Phi(m) = 0$ on Γ_S^{abs}, i.e., the parameter model is not updated on the artificial boundaries, in order to avoid artifacts in the reconstruction.
We solve the seismic inverse problem by sequentially inverting for source frequencies of 0.625Hz, 1.25Hz and 2.5Hz. On the finest level, we obtain a parameter mesh with roughly 100,000 degrees of freedom. The state mesh has approximately 300,000 spatial grid points and 1,000 time steps.
Figure 6.8 shows histograms for the misfit at all receiver locations before and after the optimization. The accumulated misfit has been reduced by more than 83%, i.e., there is a good match between observed and reconstructed data. Figure 6.9 compares the initial, target and reconstructed parameter model at a vertical cross section through the domain. The inversion output looks reasonable, especially near the surface. As expected the reconstruction becomes less accurate at greater depths. Figure 6.11 shows horizontal snapshots of the P-wave velocity for the reconstructed and the target model. The images show the deviation from the reference model that is homogeneous for every fixed depth.
The adaptively refined parameter meshes are illustrated in Figure 6.10. Table 6.1 summarizes the optimization process on the different frequency levels. We use a relative reduction of the norm of the gradient by 3 orders of magnitude as stopping criterion. The computations were carried out on a Linux cluster using 32 processors.

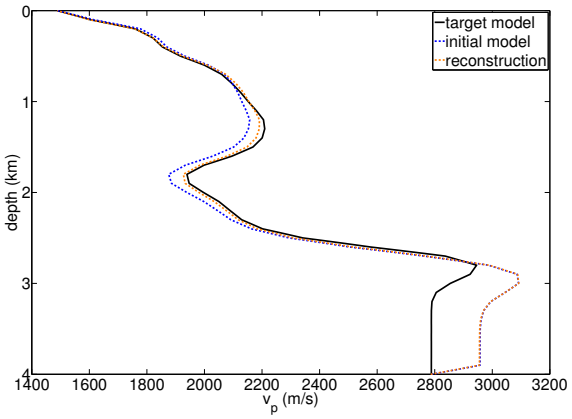


(a) initial misfits

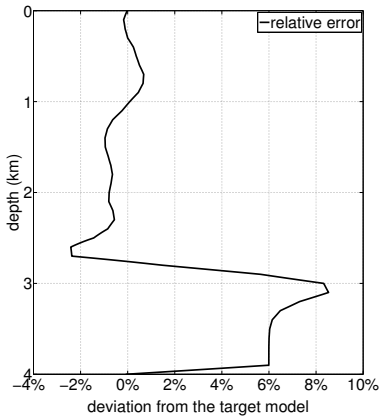


(b) final misfits

Figure 6.8: (a) Histogram of the misfit at all receiver locations using the initial model. (b) Histogram obtained with the reconstructed model. Note the different scaling of the x-axis. The misfit has been reduced by more than 83%. All three components of the seismograms are used to compute the misfit.



(a)



(b)

Figure 6.9: (a) P-wave velocity for a vertical cross section through the domain at the center of the x-y-plane. (b) Relative error between the reconstruction and the target model.

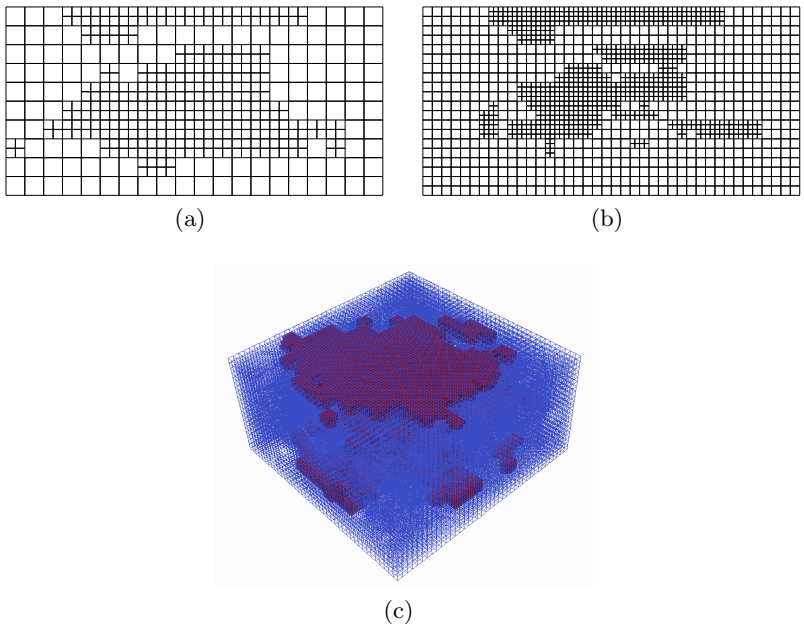
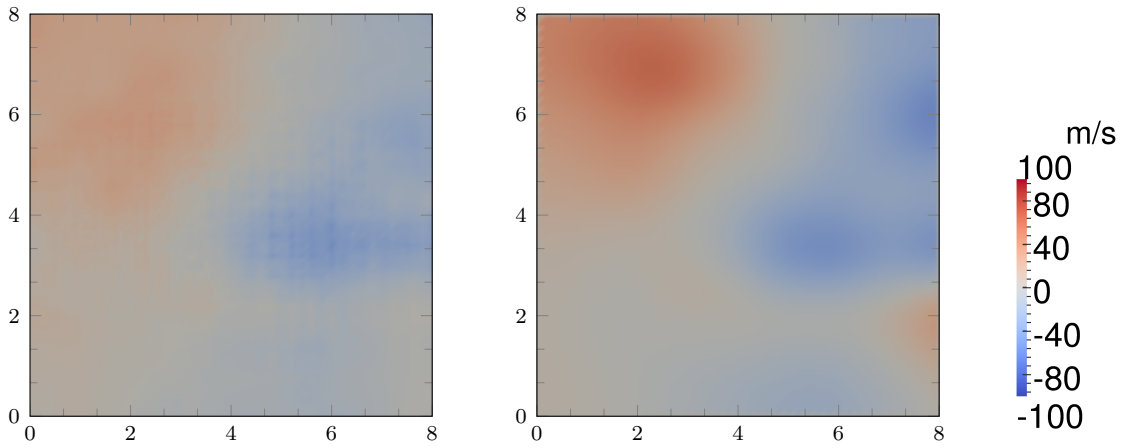


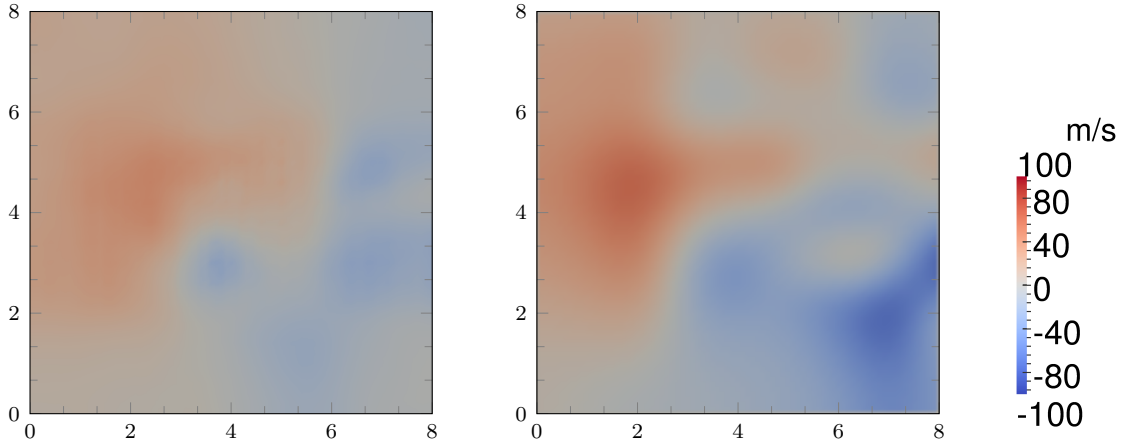
Figure 6.10: Adaptively refined parameter meshes. (a) and (b) show vertical profiles of the mesh in the center of the domain after the first and second refinement. (c) depicts the whole grid after the second refinement. Cells that are colored in dark red indicate the smallest elements.

frequency	# dof	it Newton	#PDE solves
0.625Hz	4851	22	886
1.25Hz	15949	15	550
2.5Hz	101015	13	384

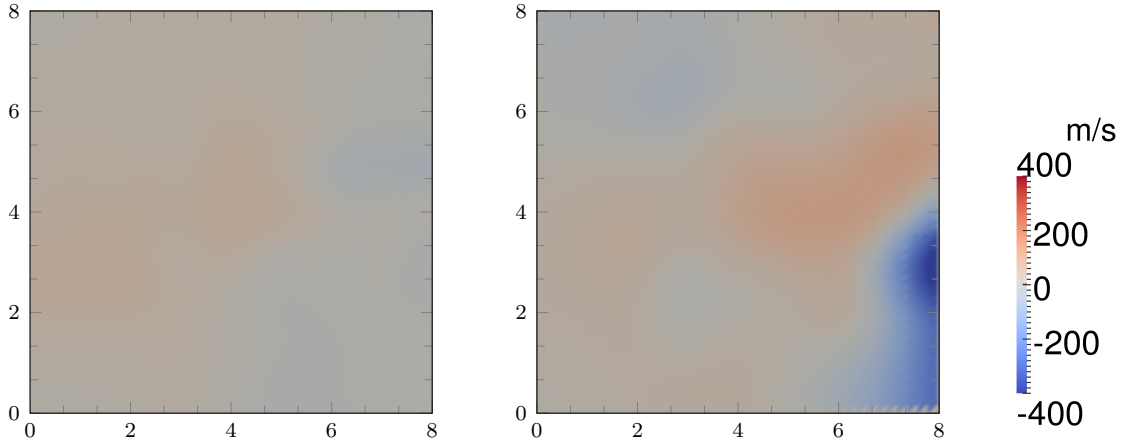
Table 6.1: Summary of the optimization process. The second column lists the degrees of freedom of the parameter mesh. On every frequency level, the algorithm was terminated, after the norm of the gradient had been reduced by three orders of magnitude.



(a) v_p at 150m depth



(b) v_p at 800m depth



(c) v_p at 1500m depth

Figure 6.11: Horizontal snapshots of v_p at different depths for the reconstruction (left) and target model (right). The images show the deviation from the reference model.

6.7 Further Perspectives

In this chapter, we discussed seismic wave propagation at a solid-fluid interface and focused on modeling the coupled system and the numerical realization of the inverse problem. To this end, we proposed some practical enhancements like a multi-frequency inversion and goal-oriented adaptive mesh refinement of the parameter grid.
Dealing with existence and uniqueness of solutions to the coupled system of PDEs is beyond the scope of this thesis but would be an interesting field for future research.
Due to the similar structure, the wave propagation code described in Chapter 4 requires only small changes to incorporate the interface conditions. However, an efficient decomposition of the computational domain is more challenging, because the fluid elements require significantly less effort than the solid ones. This has not been exploited to full extend in the current implementation and could be investigated in the future to improve the performance.

Chapter 7

Conclusion and Outlook

In this thesis, we considered problems in seismic tomography governed by the elastic and/or the acoustic wave equation. As some of the main contributions, we rigorously analyzed the infinite-dimensional problem, provided results on the differentiability of the solution operator and proved the existence of solutions to the regularized inverse problem. Despite being a field of very active research, the focus of related literature is often limited to the discretization and the infinite-dimensional problem has yet not been addressed in full extend.
In addition, the commonly used formulation of the inverse problem without constraints has been extended. Here, we established a framework that allows for additional restrictions on the material parameters and utilized the Moreau-Yosida regularization as a proper solution method. Building upon existing work for state-constrained problems, error estimates for the violation of the constraints have been established. It is important to note that the results from Chapter 2 and Chapter 3 are of greater validity and can be applied to general problems governed by linear hyperbolic PDEs of second order that depend on unknown coefficients.
The numerical results focus on the application to parameter identification in seismic tomography. As part of this thesis, an MPI-parallelized code for simulating wave propagation in solid and fluid media as well as for solving the related inverse problems has been developed. This software is capable of running on large-scale computing clusters and has been tested on up to 4096 cores. The framework features the adjoint-based computation of the gradient and Hessian-vector products as well as a matrix-free algorithm to solve the discretized problem with a trust-region Newton-PCG method.
The costs of conventional approaches for full-waveform tomography scale proportionally with the number of seismic sources. Here, randomized source sampling techniques that trigger different sources simultaneously have been proven to be a successful tool to trade a small loss of information for huge savings of computational time to solve the inverse problem. In particular, we accelerated a sample average approximation model by using inexact Hessian information based on mini-batches of the samples.

Possible extensions and further related questions have already been pointed out at the end of each chapter. In more general terms, there are two driving forces that can guide future directions to continue the research of this thesis.
First of all, the optimization methods that have been presented in this work are suitable to identify local minima. Since the problem is non-convex this might be a severe limitation in cases where a sufficiently good initial material model is not available. In order to improve this

situation and to enlarge the basin of attraction, we proposed a multi-frequency inversion and commented on different misfit criteria. Nevertheless, there is a general trend towards *statistical inverse problems* [77, 120], which seek to identify a probability distribution quantifying the likelihood of all material models, opposed to *deterministic inverse problems* considered in this thesis, which select only one specific model. This has already been applied in the context of seismic tomography [24, 56, 93] and it would be very interesting to combine ideas from randomized source sampling and the constrained parameter identification problem with a stochastic Newton Markov chain Monte Carlo method.
Another direction for future research relates to the many new challenges that will arise when real-world problems using actual recordings from earthquakes or geophysical exploration are considered. The numerical experiments in this thesis have been carried out with synthetic data. While the results look very promising, the presented methods yet have to prove their suitability for application to real data.

Appendix A

Preliminaries from Functional Analysis

In this appendix, we collect some background material from functional analysis. For the sake of brevity, we restrict the presentation to settings that are relevant for this thesis. It should be emphasized, however, that many of the subsequently stated results can be generalized, for instance, with respect to weakened assumptions on Ω. The interested reader is referred to [1, 4, 31, 43, 49].

Proposition A.1 (Poincaré's inequality)**.**
Let $\Omega \subset \mathbb{R}^d$ be open and bounded. Then there exists a constant $C > 0$ such that

$$|u|_{H^1(\Omega)} \leq \|u\|_{H^1(\Omega)} \leq C|u|_{H^1(\Omega)} \quad \forall\ u \in H_0^1(\Omega).$$

Proof. A proof can be found, for instance, in [43], Satz 6.13 . □

Proposition A.2 (Generalized Hölder's inequality)**.**
Let $p_i \in [1, \infty]$, $i = 1, \dots, k$, and $q \in [1, \infty]$ with

$$\sum_{i=1}^{k} \frac{1}{p_i} = \frac{1}{q}.$$

Then, for all $v_i \in L^{p_i}(\Omega)$, there holds $v = v_1 v_2 \dots v_k \in L^q(\Omega)$ and

$$\|v\|_{L^q(\Omega)} \leq \prod_{i=1}^{k} \|v_i\|_{L^{p_i}(\Omega)}.$$

Proof. A proof can be found, for instance, in [4], Lemma 1.18. □

Proposition A.3 (Young's inequality)**.**
Let $a, b \geq 0$, $\epsilon > 0$ and $p, q \in (1, \infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$. Then, with $0^p = 0$, there holds

$$ab \leq \frac{\epsilon}{p} a^p + \frac{\epsilon^{-q/p}}{q} b^q.$$

Proof. A proof can be found, for instance, in [49], Appendix B. Note that the variant including ϵ is a straightforward extension by setting $\tilde{a} = \epsilon^{1/p} a$ and $\tilde{b} = \epsilon^{-1/p} b$. □

Lemma A.4 (Gronwall's inequality)**.**
Let $I = [0, T]$ and consider continuous, real-valued and non-negative functions f, g, h defined on I which satisfy for almost all $t \in I$ the integral inequality

$$f(t) \leq g(t) + \int_0^t f(\tau)\, h(\tau)\, d\tau.$$

Then,

$$f(t) \leq g(t) + \int_0^t g(\tau)\, h(\tau)\, \exp\left(\int_\tau^t h(s)\, ds\right)\, d\tau$$

for almost all $t \in I$.

Proof. A proof can be found, for instance, in [49], Appendix B. □

Theorem A.5 (Weak Gauss-Green theorem)**.**
Let $\Omega \subset \mathbb{R}^d$ be open, bounded and with Lipschitz boundary and $p, q \in [1, \infty]$ with $\frac{1}{p} + \frac{1}{q} = 1$. Then for all $v \in W^{1,p}(\Omega)$ and $w \in W^{1,q}(\Omega)$ the following integration by parts formula holds for $i = 1, \ldots, d$:

$$\int_\Omega \frac{\partial}{\partial x_i} v(x)\, w(x)\, dx = -\int_\Omega v(x)\, \frac{\partial}{\partial x_i} w(x)\, dx + \int_{\partial\Omega} v(x)\, w(x)\, \vec{n}_i(x)\, dS(x).$$

Proof. A proof can be found, for instance, in [4], Theorem A6.8. □

In this thesis, we only require the special case $p = q = 2$. Note that the extension to vector-valued functions is straightforward.

Definition A.6 (Gâteaux and Fréchet differentiability)**.**
Let X, Y be Banach spaces, $\mathcal{S} \subset X$ a nonempty open set and $F : \mathcal{S} \to Y$.

(i) F is called Gâteaux differentiable at $x \in \mathcal{S}$ if the directional derivative

$$dF(x, h) = \lim_{t \searrow 0^+} \frac{F(x + th) - F(x)}{t} \quad \in Y$$

exists for all $h \in X$ and, furthermore, if $F'(x) : X \to Y$, $h \mapsto dF(x, h)$ is bounded and linear.

(ii) F is called Fréchet differentiable at $x \in \mathcal{S}$ if F is Gâteaux differentiable at x and the following estimate for the remainder term holds:

$$\|F(x + h) - F(x) - F'(x)h\|_Y = o(\|h\|_X) \quad \text{for } \|h\|_X \to 0.$$

Furthermore, if F is Gâteaux (Fréchet) differentiable at every x in an open neighborhood of $\mathcal{S}$, then F is called Gâteaux (Fréchet) differentiable on $\mathcal{S}$. F is called continuously differentiable if the mapping $x \mapsto F'(x)$ is continuous from X to $\mathcal{L}(X, Y)$.

Derivatives of higher order can be defined in the same way. With regard to this thesis, we focus on Fréchet differentiable functions.

Definition A.7 (Second derivatives)**.**
Let X, Y be Banach spaces, $\mathcal{S} \subset X$ a nonempty open set and $F : \mathcal{S} \to Y$ Fréchet differentiable. F is called twice Fréchet differentiable at $x \in \mathcal{S}$ if $F' : X \to \mathcal{L}(X, Y)$ is differentiable at x. The second derivative is denoted by $F''(x) \in \mathcal{L}(X, \mathcal{L}(X, Y))$. To shorten the notation, we write $\mathcal{L}^2(X; Y)$ for the Banach space of bilinear mappings $F''(x) : X \times X \to Y$.

Higher order derivatives can be defined analogously and lead to continuous multilinear mappings $F^{(k)}(x) \in \mathcal{L}^k(X; Y)$. Thus, the approximation condition of the remainder term of the k-th Fréchet derivative can be expressed as
For all $h_1, \dots, h_{k-1} \in X$:

$$\left\| (F^{(k-1)}(x+h) - F^{(k-1)}(x) - F^{(k)}(x)h)(h_1, \dots, h_{k-1}) \right\|_Y = o\left(\|h\|_X\right) \quad \text{for} \quad \|h\|_X \to 0.$$

The following result states that the order of h_i is arbitrary.

Theorem A.8.
Let X, Y be Banach spaces, $\mathcal{S} \subset X$ a nonempty open set and $F : \mathcal{S} \to Y$. If F is k-times Fréchet differentiable at $x \in \mathcal{S}$ then the derivative $F^{(k)}(x) \in \mathcal{L}^k(X; Y)$ is a multilinear symmetric mapping $X \times X \times \ldots \times X \to Y$, i.e., for $h_1, \dots, h_k \in X$ and any permutation σ of $\{1, \dots, k\}$, it holds

$$F^{(k)}(x)(h_1, \dots, h_k) = F^{(k)}(x)(h_{\sigma(1)}, \dots, h_{\sigma(k)}).$$

Proof. See Theorem 5.3.1 in [31]. □

We frequently require continuous and compact embeddings of Sobolev spaces. For this purpose, we recall the famous theorem:

Theorem A.9 (Sobolev embedding theorem)**.**
Let $\Omega \subset \mathbb{R}^d$ be open, bounded and with Lipschitz boundary. Furthermore, let $k_1, k_2 \in \mathbb{N}_0$, $1 \le p_1 < \infty$, $1 \le p_2 \le \infty$ and $\beta \in [0, 1]$.

(i) If $k_1 \ge k_2$ and $k_1 - \frac{d}{p_1} \ge k_2 - \frac{d}{p_2}$, then $W^{k_1,p_1}(\Omega)$ is continuously embedded into $W^{k_2,p_2}(\Omega)$, i.e.,

$$W^{k_1,p_1}(\Omega) \hookrightarrow W^{k_2,p_2}(\Omega).$$

If $k_1 > k_2$ and $k_1 - \frac{d}{p_1} > k_2 - \frac{d}{p_2}$, then the embedding is compact, i.e.,

$$W^{k_1,p_1(\Omega)} \hookrightarrow\hookrightarrow W^{k_2,p_2}(\Omega).$$

(ii) If $k_1 - \frac{d}{p_1} \ge k_2 + \beta$ and $\beta \in (0, 1)$, then W^{k_1,p_1} is continuously embedded into $C^{k_2,\beta}(\bar{\Omega})$, i.e.,

$$W^{k_1,p_1}(\Omega) \hookrightarrow C^{k_2,\beta}(\bar{\Omega}).$$

If $k_1 - \frac{d}{p_1} > k_2 + \beta$, then the embedding is compact, i.e.,

$$W^{k_1,p_1}(\Omega) \hookrightarrow\hookrightarrow C^{k_2,\beta}(\bar{\Omega}).$$

Proof. A proof can be found, for instance, in [1], Theorem 4.12, or, respectively, in [4], Theorem 8.9 and Theorem 8.13. □

Example A.10.
Here are some important special cases of Sobolev embeddings:

(i) For $d = 1$ and arbitrary $k \in \mathbb{N}$, we obtain $H^k(\Omega) \hookrightarrow\hookrightarrow C^{k-1}(\bar{\Omega})$.

(ii) With $k_2 = 0$, we obtain embeddings into $L^{p_2}(\Omega)$, for instance, $H^2(\Omega) \hookrightarrow\hookrightarrow L^{p_2}(\Omega)$ for $1 \leq p_2 \leq \infty$ and $d \in \{1, 2, 3\}$.

(iii) For $d \in \{1, 2, 3\}$, the embedding $H^1(\Omega) \hookrightarrow L^6(\Omega)$ is continuous and the embedding $H^2(\Omega) \hookrightarrow\hookrightarrow C(\bar{\Omega})$ is compact.

List of Figures

List of Algorithms

List of Tables

Bibliography

[1] R. A. Adams and J. J. F. Fournier. *Sobolev Spaces*. Vol. 140. Pure and Applied Mathematics. Amsterdam: Elsevier, 2003.

[2] K. Aki and P. G. Richards. *Quantitative seismology: Theory and methods*. 2nd edition. Mill Valley, CA: University Science Books, 2002.

[3] K. Aki and W. H. K. Lee. "Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes: 1. A homogeneous initial model". In: *Journal of Geophysical Research* 81.23 (1976), pp. 4381–4399.

[4] H. W. Alt. *Lineare Funktionalanalysis*. Vol. 6. Berlin: Springer, 2012.

[5] J. Appell and P. Zabrejko. *Nonlinear Superposition Operators*. Cambridge: Cambridge University Press, 1990.

[6] A. Aravkin, M. Friedlander, F. Herrmann, and T. van Leeuwen. "Robust inversion, dimensionality reduction, and randomized sampling". In: *Mathematical Programming* 134.1 (2012), pp. 101–125.

[7] V. Babuska and M. Cara. *Seismic anisotropy in the Earth*. Vol. 10. Modern Approaches in Geophysics. Dordrecht: Kluwer, 1991.

[8] W. Bangerth, M. Geiger, and R. Rannacher. "Adaptive Galerkin finite element methods for the wave equation". In: *Computational Methods in Applied Mathematics* 10.1 (2010), pp. 3–48.

[9] W. Bangerth, R. Hartmann, and G. Kanschat. "deal.II – a General Purpose Object Oriented Finite Element Library". In: *ACM Trans. Math. Softw.* 33.4 (2007), pp. 24/1–24/27.

[10] W. Bangerth and G. Kanschat. `deal.II` *Differential Equations Analysis Library, Technical Reference*. `http://www.dealii.org`.

[11] W. Bangerth. "A framework for the adaptive finite element solution of large inverse problems". In: *SIAM Journal on Scientific Computing* 30 (2008), pp. 2965–2989.

[12] C. Barnes and M. Charara. "The domain of applicability of acoustic full-waveform inversion for marine seismic data". In: *GEOPHYSICS* 74.6 (2009), WCC91–WCC103.

[13] P. Basini, C. Blitz, E. Bozdag, E. Casarotti, M. Chen, H. Gharti, V. Hjörleifsdóttir, S. Kientz, D. Komatitsch, J. Labarta, et al. *SPECFEM 3D user manual*. Version 2.0. Princeton University / California Institute of Technology and University of Pau / CNRS / INRIA. 2011.

[14] R. Becker and R. Rannacher. “An optimal control approach to a posteriori error estimation in finite element methods”. In: *Acta Numerica 2001* 10.1 (2001), pp. 1–102.

[15] H. Ben-Hadj-Ali, S. Operto, and J. Virieux. “An efficient frequency-domain full waveform inversion method using simultaneous encoded sources”. In: *GEOPHYSICS* 76.4 (2011), R109–R124.

[16] K. D. Blazek, C. Stolk, and W. W. Symes. “A mathematical framework for inverse wave problems in heterogeneous media”. In: *Inverse Problems* 29.6 (2013), p. 065001.

[17] C. Boehm. *MACESOPT: User Guide*. Tech. rep. Technische Universität München, 2014.

[18] C. Boehm and M. Ulbrich. “A Newton-CG Method for Full-Waveform Inversion in a Coupled Solid-Fluid System”. In: *Advanced Computing*. Ed. by M. Bader, H.-J. Bungartz, and T. Weinzierl. Vol. 93. Lecture Notes in Computational Science and Engineering. Berlin: Springer, 2013, pp. 99–117.

[19] C. Boehm and M. Ulbrich. *A Semismooth Newton-CG Method for Constrained Parameter Identification in Seismic Tomography*. In revision. 2014.

[20] P. Bois, M. La Porte, M. Lavergne, and G. Thomas. “Well-to-well seismic measurements”. In: *GEOPHYSICS* 37.3 (1972), pp. 471–480.

[21] A. Bourgeois, M. Bourget, P. Lailly, M. Poulet, P. Ricarte, and R. Versteeg. “Marmousi, model and data”. In: *The Marmousi Experience, Proceedings of the 1990 EAEG Workshop on Practical Aspects of Seismic Data Inversion*. Ed. by G. Grau and R. Versteeg. EAEG, Zeist, 1991, pp. 5–16.

[22] R. Brossier, S. Operto, and J. Virieux. “Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion”. In: *GEOPHYSICS* 74.6 (2009), WCC105–WCC118.

[23] R. Brossier, S. Operto, and J. Virieux. “Which data residual norm for robust elastic frequency-domain full waveform inversion?” In: *GEOPHYSICS* 75.3 (2010), R37–R46.

[24] T. Bui-Thanh, O. Ghattas, J. Martin, and G. Stadler. “A Computational Framework for Infinite-Dimensional Bayesian Inverse Problems Part I: The Linearized Case, with Application to Global Seismic Inversion”. In: *SIAM Journal on Scientific Computing* 35.6 (2013), A2494–A2523.

[25] C. Bunks, F. Saleck, S. Zaleski, and G. Chavent. “Multiscale seismic waveform inversion”. In: *GEOPHYSICS* 60.5 (1995), pp. 1457–1473.

[26] C. Burstedde and O. Ghattas. “Algorithmic strategies for full waveform inversion: 1D experiments”. In: *GEOPHYSICS* 74.6 (2009), WCC37–WCC46.

[27] R. Byrd, G. Chin, W. Neveitt, and J. Nocedal. “On the Use of Stochastic Hessian Information in Optimization Methods for Machine Learning”. In: *SIAM Journal on Optimization* 21.3 (2011), pp. 977–995.

[28] R. H. Byrd, S. Hansen, J. Nocedal, and Y. Singer. “A Stochastic Quasi-Newton Method for Large-Scale Optimization”. In: *arXiv preprint arXiv:1401.7020* (2014).

[29] E. Candès, L. Demanet, D. Donoho, and L. Ying. “Fast Discrete Curvelet Transforms”. In: *Multiscale Modeling & Simulation* 5.3 (2006), pp. 861–899.

[30] Y. Capdeville, Y. Gung, and B. Romanowicz. “Towards global earth tomography using the spectral element method: a technique based on source stacking”. In: *Geophysical Journal International* 162.2 (2005), pp. 541–554.

[31] H. Cartan. *Differential calculus*. Vol. 1. Paris: Hermann, 1971.

[32] E. Casas, C. Clason, and K. Kunisch. “Parabolic Control Problems in Measure Spaces with Sparse Solutions”. In: *SIAM Journal on Control and Optimization* 51.1 (2013), pp. 28–63.

[33] E. Chaljub, Y. Capdeville, and J.-P. Vilotte. “Solving elastodynamics in a fluid–solid heterogeneous sphere: a parallel spectral element approximation on non-conforming grids”. In: *Journal of Computational Physics* 187.2 (2003), pp. 457 –491.

[34] E. Chaljub, D. Komatitsch, J.-P. Vilotte, Y. Capdeville, B. Valette, and G. Festa. “Spectral-element analysis in seismology”. In: *Advances in Wave Propagation in Heterogenous Earth*. Ed. by V. M. Ru-Shan Wu and R. Dmowska. Vol. 48. Advances in Geophysics. Elsevier, 2007, pp. 365 –419.

[35] C. Clason and K. Kunisch. “A measure space approach to optimal source placement”. English. In: *Computational Optimization and Applications* 53.1 (2012), pp. 155–171.

[36] A. Colombi, T. Nissen-Meyer, L. Boschi, and D. Giardini. “Seismic waveform sensitivity to global boundary topography”. In: *Geophysical Journal International* 191.2 (2012), pp. 832–848.

[37] A. R. Conn, N. I. M. Gould, and P. L. Toint. *Trust Region Methods*. Philadelphia, PA: SIAM, 2000.

[38] A. Curtis. “Theory of model-based geophysical survey and experimental design: Part 1—Linear problems”. In: *The Leading Edge* 23.10 (2004), pp. 997–1004.

[39] F. Dahlen and J. Tromp. *Theoretical global seismology*. Princeton, NJ: Princeton university press, 1998.

[40] S. Dain. “Generalized Korn’s inequality and conformal Killing vectors”. In: *Calculus of Variations and Partial Differential Equations* 25.4 (2006), pp. 535–540.

[41] F. R. Dalmau, M Hanzich, J de la Puente, and N Gutierrez. “Lossy Data Compression with DCT Transforms”. In: *EAGE Workshop on High Performance Computing for Upstream*. 2014.

[42] J. Dennis Jr. and J. Moré. “Quasi-Newton Methods, Motivation and Theory”. In: *SIAM Review* 19.1 (1977), pp. 46–89.

[43] M. Dobrowolski. *Angewandte Funktionalanalysis*. Vol. 2. Heidelberg: Springer, 2010.

[44] R. M. Dudley and R. Norvaiša. *Concrete functional calculus*. New York, NY: Springer, 2010.

[45] G. Duvaut and J. Lions. *Inequalities in Mechanics and Physics*. Vol. 170. Berlin: Springer, 1976.

[46] A. Dziewonski and D. Anderson. “Preliminary reference Earth model”. In: *Physics of the earth and planetary interiors* 25.4 (1981), pp. 297–356.

[47] H. W. Engl. *Regularization of inverse problems*. Dordrecht: Kluwer, 2000.

[48] I. Epanomeritakis, V. Akçelik, O. Ghattas, and J. Bielak. "A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion". In: *Inverse Problems* 24.3 (2008), p. 034015.

[49] L. Evans. *Partial Differential Equations*. Graduate Studies in Mathematics. Providence, RI: American Mathematical Society, 2010.

[50] Exxon Neftegas Limited. *Sakhalin-1 Drilling Program Marks 10th Anniversary with World Records*. 2013, accessed May 02, 2014. URL: `http://www.sakhalin-1.com/Sakhalin/Russia-English/Upstream/Files/10-year_drilling_ENG.pdf`.

[51] A. Fichtner. *SES3D version 2.1: Programme Description and Mathematical Background.* Tech. rep. Ludiwg-Maximilians-Universität München, 2009.

[52] A. Fichtner, B. Kennett, H. Igel, and H. Bunge. "Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods". In: *Geophysical Journal International* 179.3 (2009), pp. 1703–1725.

[53] A. Fichtner, B. Kennett, H. Igel, and H. Bunge. "Theoretical background for continental- and global-scale full-waveform inversion in the time–frequency domain". In: *Geophysical Journal International* 175.2 (2008), pp. 665–685.

[54] A. Fichtner. *Full seismic waveform modelling and inversion*. Berlin Heidelberg: Springer, 2011.

[55] A. Fichtner. "Full waveform inversion for structural and source parameters". PhD thesis. Ludwig-Maximilians-Universität, München, 2010.

[56] A. Fichtner and J. Trampert. "Resolution analysis in full waveform inversion". In: *Geophysical Journal International* 187.3 (2011), pp. 1604–1624.

[57] Y Gholami, R Brossier, S Operto, V Prieux, A Ribodetti, and J. Virieux. "Which parametrization is suitable for acoustic VTI full-waveform inversion? Part 2: application to Valhall". In: *GEOPHYSICS* 78.2 (2013), R81–R105.

[58] W. Gong, M. Hinze, and Z. Zhou. "A Priori Error Analysis for Finite Element Approximation of Parabolic Optimal Control Problems with Pointwise Control". In: *SIAM Journal on Control and Optimization* 52.1 (2014), pp. 97–119.

[59] A. Griewank. "Achieving logarithmic growth of temporal and spatial complexity in reverse automatic differentiation". In: *Optimization Methods and Software* 1.1 (1992), pp. 35–54.

[60] T. Guest and A. Curtis. "On standard and optimal designs of industrial-scale 2-D seismic surveys". In: *Geophysical Journal International* 186.2 (2011), pp. 825–836.

[61] E. Haber, M. Chung, and F. Herrmann. "An Effective Method for Parameter Estimation with PDE Constraints with Multiple Right-Hand Sides". In: *SIAM Journal on Optimization* 22.3 (2012), pp. 739–757.

[62] E. Haber, L. Horesh, and L. Tenorio. "Numerical methods for experimental design of large-scale linear ill-posed inverse problems". In: *Inverse Problems* 24.5 (2008), p. 055012.

[63] M. A. Heroux, R. A. Bartlett, V. E. Howle, R. J. Hoekstra, J. J. Hu, T. G. Kolda, R. B. Lehoucq, K. R. Long, R. P. Pawlowski, E. T. Phipps, A. G. Salinger, H. K. Thornquist, R. S. Tuminaro, J. M. Willenbring, A. Williams, and K. S. Stanley. "An overview of the Trilinos project". In: *ACM Trans. Math. Softw.* 31.3 (2005), pp. 397–423.

[64] M. A. Heroux and J. M. Willenbring. *Trilinos Users Guide*. Tech. rep. SAND2003-2952. Sandia National Laboratories, 2003.

[65] F. J. Herrmann, P. Moghaddam, and C. C. Stolk. "Sparsity- and continuity-promoting seismic image recovery with curvelet frames". In: *Applied and Computational Harmonic Analysis* 24.2 (2008). Special Issue on Mathematical Imaging – Part {II}, pp. 150 –173.

[66] R. Herzog, G. Stadler, and G. Wachsmuth. "Directional Sparsity in Optimal Control of Partial Differential Equations". In: *SIAM Journal on Control and Optimization* 50.2 (2012), pp. 943–963.

[67] M. Hintermüller, K. Ito, and K. Kunisch. "The Primal-Dual Active Set Strategy as a Semismooth Newton Method". In: *SIAM Journal on Optimization* 13.3 (2002), pp. 865–888.

[68] M. Hintermüller, A. Schiela, and W. Wollner. "The Length of the Primal-Dual Path in Moreau–Yosida-Based Path-Following Methods for State Constrained Optimal Control". In: *SIAM Journal on Optimization* 24.1 (2014), pp. 108–126.

[69] M. Hintermüller and M. Hinze. "Moreau–Yosida Regularization in State Constrained Elliptic Control Problems: Error Estimates and Parameter Adjustment". In: *SIAM Journal on Numerical Analysis* 47.3 (2009), pp. 1666–1683.

[70] M. Hintermüller and K. Kunisch. "Feasible and Noninterior Path-Following in Constrained Minimization with Low Multiplier Regularity". In: *SIAM Journal on Control and Optimization* 45.4 (2006), pp. 1198–1221.

[71] M. Hintermüller and K. Kunisch. "Path-following Methods for a Class of Constrained Minimization Problems in Function Space". In: *SIAM Journal on Optimization* 17.1 (2006), pp. 159–187.

[72] M. Hintermüller and K. Kunisch. "PDE-Constrained Optimization Subject to Pointwise Constraints on the Control, the State, and Its Derivative". In: *SIAM Journal on Optimization* 20.3 (2010), pp. 1133–1156.

[73] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE constraints*. Berlin: Springer, 2008.

[74] T. Hughes. *The finite element method*. Mineola, NY: Dover Publications, 2000.

[75] M. Hutchinson. "A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines". In: *Communications in Statistics - Simulation and Computation* 19.2 (1990), pp. 433–450.

[76] H. Igel, H. Djikpéssé, and A. Tarantola. "Waveform inverstion of marine reflection seismograms for P impedance and Poisson's ratio". In: *Geophysical Journal International* 124.2 (1996), pp. 363–371.

[77] J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*. Vol. 160. New York, NY: Springer, 2004.

[78] A. Kirsch and A. Rieder. “On the linearization of operators related to the full waveform inversion in seismology”. In: *Mathematical Methods in the Applied Sciences* (2013). first published online, doi: 10.1002/mma.3037.

[79] D. Komatitsch, C. Barnes, and J. Tromp. “Wave propagation near a fluid-solid interface: A spectral-element approach”. In: *GEOPHYSICS* 65.2 (2000), pp. 623–631.

[80] D. Komatitsch and J.-P. Vilotte. “The spectral element method: An efficient tool to simulate the seismic response of 2D and 3D geological structures”. In: *Bulletin of the Seismological Society of America* 88.2 (1998), pp. 368–392.

[81] J. R. Krebs, J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse. “Fast full-wavefield seismic inversion using encoded sources”. In: *GEOPHYSICS* 74.6 (2009), WCC177–WCC188.

[82] M. Kristekova, J. Kristek, P. Moczo, and S. Day. “Misfit criteria for quantitative comparison of seismograms”. In: *Bulletin of the seismological Society of America* 96.5 (2006), pp. 1836–1850.

[83] M. Kristekova, J. Kristek, and P. Moczo. “Time-frequency misfit and goodness-of-fit criteria for quantitative comparison of time signals”. In: *Geophysical Journal International* 178.2 (2009), pp. 813–825.

[84] A. Kröner. “Numerical Methods for Control of Second Order Hyperbolic Equations”. PhD thesis. Technische Universität München, 2011.

[85] A. Kröner. “Adaptive finite element methods for optimal control of second order hyperbolic equations”. In: *Computational Methods in Applied Mathematics* 11.2 (2011), pp. 214–240.

[86] A. Kröner, K. Kunisch, and B. Vexler. “Semismooth Newton Methods for Optimal Control of the Wave Equation with Control Constraints”. In: *SIAM Journal on Control and Optimization* 49.2 (2011), pp. 830–858.

[87] I. Lasiecka and R. Triggiani. “Sharp regularity theory for second order hyperbolic equations of Neumann type”. In: *Annali di Matematica Pura ed Applicata* 157.1 (1990), pp. 285–367.

[88] R. M. Lewis and W. W. Symes. “On the relation between the velocity coefficient and boundary value for solutions of the one-dimensional wave equation”. In: *Inverse Problems* 7.4 (1991), p. 597.

[89] F.-C. Lin, M. H. Ritzwoller, and R. Snieder. “Eikonal tomography: surface wave tomography by phase front tracking across a regional broad-band seismic array”. In: *Geophysical Journal International* 177.3 (2009), pp. 1091–1110.

[90] J. Lions. *Optimal control of systems governed by partial differential equations*. Vol. 170. Berlin: Springer, 1971.

[91] J. Lions and E. Magenes. *Problèmes aux limites non homogenes et applications, Vol. 1*. Paris: Dunod, 1968.

[92] R. Madariaga. “Seismic Source Theory”. In: *Treatise on Geophysics*. Ed. by G. Schubert. Amsterdam: Elsevier, 2007, pp. 59 –82.

[93] J. Martin, L. Wilcox, C. Burstedde, and O. Ghattas. “A Stochastic Newton MCMC Method for Large-Scale Statistical Inverse Problems with Application to Seismic Inversion”. In: *SIAM Journal on Scientific Computing* 34.3 (2012), A1460–A1487.

[94] D. Meidner and B. Vexler. “Adaptive Space-Time Finite Element Methods for Parabolic Optimization Problems”. In: *SIAM Journal on Control and Optimization* 46.1 (2007), pp. 116–142.

[95] L. Métivier, R. Brossier, J. Virieux, and S. Operto. “Full Waveform Inversion and the Truncated Newton Method”. In: *SIAM Journal on Scientific Computing* 35.2 (2013), B401–B437.

[96] K. C. Meza-Fajardo and A. S. Papageorgiou. “A Nonconvolutional, Split-Field, Perfectly Matched Layer for Wave Propagation in Isotropic and Anisotropic Elastic Media: Stability Analysis”. In: *Bulletin of the Seismological Society of America* 98.4 (2008), pp. 1811–1836.

[97] P. Moghaddam, H. Keers, F. Herrmann, and W. Mulder. “A new optimization approach for source-encoding full-waveform inversion”. In: *GEOPHYSICS* 78.3 (2013), R125–R132.

[98] J. Morales and J. Nocedal. “Automatic Preconditioning by Limited Memory Quasi-Newton Updating”. In: *SIAM Journal on Optimization* 10.4 (2000), pp. 1079–1096.

[99] W. Mulder and R.-E. Plessix. “Exploring some issues in acoustic full waveform inversion”. In: *Geophysical Prospecting* 56.6 (2008), pp. 827–841.

[100] T. Nissen-Meyer, A. Fournier, and F. A. Dahlen. “A 2-D spectral-element method for computing spherical-earth seismograms - II. Waves in solid-fluid media”. In: *Geophysical Journal International* 174.3 (2008), pp. 873–888.

[101] G. Nolet. *A breviary of seismic tomography*. Cambridge: Cambridge University Press, 2008.

[102] J. W. Pearson, M. Stoll, and A. J. Wathen. “Preconditioners for state-constrained optimal control problems with Moreau–Yosida penalty function”. In: *Numerical Linear Algebra with Applications* 21.1 (2014), pp. 81–97.

[103] D. Peter, D. Komatitsch, Y. Luo, R. Martin, N. Le Goff, E. Casarotti, P. Le Loher, F. Magnoni, Q. Liu, C. Blitz, T. Nissen-Meyer, P. Basini, and J. Tromp. “Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes”. In: *Geophysical Journal International* 186.2 (2011), pp. 721–739.

[104] R. G. Pratt, C. Shin, and G. J. Hick. “Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion”. In: *Geophysical Journal International* 133.2 (1998), pp. 341–362.

[105] R. G. Pratt and M. H. Worthington. “Inverse theory applied to multi-source cross-hole tomography”. In: *Geophysical Prospecting* 38.3 (1990), pp. 287–310.

[106] V. Prieux, R. Brossier, S. Operto, and J. Virieux. “Multiparameter full waveform inversion of multicomponent OBC data from valhall. part 2 : imaging compressional and shearwave velocities”. In: *Geophysical Journal International* (2013).

[107] V. Prieux, R. Brossier, Y. Gholami, S. Operto, J. Virieux, O. I. Barkved, and J. H. Kommedal. "On the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study". In: *Geophysical Journal International* 187.3 (2011), pp. 1495–1515.

[108] N. Rawlinson, S. Pozgay, and S. Fishwick. "Seismic tomography: A window into deep Earth". In: *Physics of the Earth and Planetary Interiors* 178.3-4 (2010), pp. 101 –135.

[109] M. Renardy and R. C. Rogers. *An introduction to partial differential equations*. Vol. 2. New York, NY: Springer, 2004.

[110] F. Rickers, A. Fichtner, and J. Trampert. "The Iceland–Jan Mayen plume system and its impact on mantle dynamics in the North Atlantic region: Evidence from full-waveform inversion". In: *Earth and Planetary Science Letters* 367 (2013), pp. 39–51.

[111] H. Robbins and S. Monro. "A stochastic approximation method". In: *The annals of mathematical statistics* (1951), pp. 400–407.

[112] A. Schiemenz and H. Igel. "Accelerated 3-D full-waveform inversion using simultaneously encoded sources in the time domain: application to Valhall ocean-bottom cable data". In: *Geophysical Journal International* 195.3 (2013), pp. 1970–1988.

[113] A. Shapiro, D. Dentcheva, and A. Ruszczynski. *Lectures on Stochastic Programming*. Philadelphia, PA: SIAM, 2009.

[114] L. Sirgue, O. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. Kommedal. "Full waveform inversion: the next leap forward in imaging at Valhall". In: *First Break* 28 (2010).

[115] S. C. Stähler and K. Sigloch. "Fully probabilistic seismic source inversion, Part 1: Efficient parameterisation". In: *Solid Earth Discussions* 5.2 (2013), pp. 1125–1162.

[116] T. Steihaug. "The Conjugate Gradient Method and Trust Regions in Large Scale Optimization". In: *SIAM Journal on Numerical Analysis* 20.3 (1983), pp. 626–637.

[117] W. Symes. "Source synthesis for waveform inversion". In: *SEG Expanded Abstracts*. Vol. 29. 1. 2010, pp. 1018–1022.

[118] C. Tape, Q. Liu, A. Maggi, and J. Tromp. "Seismic tomography of the southern California crust based on spectral-element and adjoint methods". In: *Geophysical Journal International* 180.1 (2010), pp. 433–462.

[119] A. Tarantola. "Inversion of seismic reflection data in the acoustic approximation". In: *GEOPHYSICS* 49.8 (1984), pp. 1259–1266.

[120] A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia, PA: SIAM, 2005.

[121] L. L. Thompson and P. M. Pinsky. "A space-time finite element method for structural acoustics in infinite domains part 1: Formulation, stability and convergence". In: *Computer Methods in Applied Mechanics and Engineering* 132.3–4 (1996), pp. 195 – 227.

[122] F. Tröltzsch. *Optimale Steuerung partieller Differentialgleichungen*. Vol. 2. Wiesbaden: Vieweg Teubner, 2009.

[123] J. Tromp, D. Komattisch, and Q. Liu. "Spectral-element and adjoint methods in seismology". In: *Communications in Computational Physics* 3.1 (2008), pp. 1–32.

[124] J. Tromp, C. Tape, and Q. Liu. “Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels”. In: *Geophysical Journal International* 160.1 (2005), pp. 195–216.

[125] M. Ulbrich. “Semismooth Newton Methods for Operator Equations in Function Spaces”. In: *SIAM Journal on Optimization* 13.3 (2002), pp. 805–841.

[126] M. Ulbrich. *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*. Philadelphia, PA: SIAM, 2011.

[127] R. D. Van Der Hilst and M. V. De Hoop. “Banana-doughnut kernels and mantle tomography”. In: *Geophysical Journal International* 163.3 (2005), pp. 956–961.

[128] T. Van Leeuwen and W. A. Mulder. “A correlation-based misfit criterion for wave-equation traveltime tomography”. In: *Geophysical Journal International* 182.3 (2010), pp. 1383–1394.

[129] R. Versteeg. “The Marmousi experience: Velocity model determination on a synthetic complex data set”. In: *The Leading Edge* 13.9 (1994), pp. 927–936.

[130] J. Virieux and S. Operto. “An overview of full-waveform inversion in exploration geophysics”. In: *GEOPHYSICS* 74.6 (2009), WCC1–WCC26.

[131] C. R. Vogel. *Computational Methods for Inverse Problems*. Philadelphia, PA: SIAM, 2002.

[132] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. “Image quality assessment: from error visibility to structural similarity”. In: *Image Processing, IEEE Transactions on* 13.4 (2004), pp. 600–612.

[133] L. Wilcox, G. Stadler, C. Burstedde, and O. Ghattas. “A high-order discontinuous Galerkin method for wave propagation through coupled elastic-acoustic media”. In: *Journal of Computational Physics* 229.24 (2010), pp. 9373 –9396.

[134] J. Wloka. *Partielle Differentialgleichungen: Sobolevräume und Randwertaufgaben*. Stuttgart: Teubner, 1982.

[135] H. Zhu, E. Bozdağ, D. Peter, and J. Tromp. “Structure of the European upper mantle revealed by adjoint tomography”. In: *Nature Geoscience* 5.7 (2012), pp. 493–498.