A SEMISMOOTH NEWTON-CG METHOD FOR CONSTRAINED PARAMETER IDENTIFICATION IN SEISMIC TOMOGRAPHY

CHRISTIAN BOEHM* AND MICHAEL ULBRICH†

Abstract. Seismic tomography is a technique to determine the material properties of the Earth's subsurface based on the observation of seismograms. This can be stated as a PDE-constrained optimization problem governed by the elastic wave equation. We present a semismooth Newton-PCG method with a trust-region globalization for full-waveform seismic inversion that uses a Moreau-Yosida regularization to handle additional constraints on the material parameters. We establish results on the differentiability of the parameter-to-state operator and analyze the proposed optimization method in a function space setting. The elastic wave equation is discretized by a high-order continuous Galerkin method in space and an explicit Newmark time-stepping scheme. The matrix-free implementation relies on the adjoint-based computation of the gradient and Hessian-vector products and on an MPI-based parallelization. Numerical results are shown for an application in geophysical exploration on reservoir-scale.

Key words. seismic tomography, full-waveform inversion, elastic wave equation, semismooth Newton, Moreau-Yosida regularization

AMS subject classifications. 86-08, 49M15, 90C06, 35L05

1. Introduction.

When earthquakes occur, seismic waves are emitted and can be measured in form of seismograms at receiver locations far away from the source. The velocity of the travelling waves depends on the material and reflections occur at the transition of different layers of rock. Seismic tomography means to infer the material structure of the Earth's subsurface based on the observation of seismic waves that spread through the Earth. This poses a challenging inverse problem since data observations are usually only available for a small part of the domain.

There are many applications requiring an accurate knowledge of the Earth's interior. For instance, new insights from seismic tomography can be used to explain geodynamic processes, to identify potential geologic hazards (e.g. landslides or volcanoes) and to support the search for natural resources.

Seismic tomography can be stated 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 given either by the elastic or the acoustic wave equation, respectively.

A general overview on seismic tomography can be found in [39, 48]. Iterative inversion methods based on first-order information have been applied to 2d and 3d datasets on both regional and continental scale [14, 15, 43]. A Newton-CG method for the unconstrained parameter identification problem was presented in [10]. Alternative approaches work in the frequency domain and involve the Helmholtz equation [34, 37]. To the best of our knowledge, few attention has been paid to the analysis of the infinite-dimensional problem. Recently, results on the differentiability of the parameter-to-solution operator have been established independently in [3] and [27] for the acoustic wave equation.

*ETH Zürich, Institute of Geophysics, Sonneggstrasse 5, 8092 Zürich, Switzerland (christian.boehm@erdw.ethz.ch).

†Technische Universität München, Chair of Mathematical Optimization, Department of Mathematics, Boltzmannstr. 3, 85748 Garching b. München, Germany (mulbrich@ma.tum.de).

Most approaches in full waveform tomography do not include 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, constraints can be used to address the ill-posedness of the problem by incorporating additional prior knowledge into the problem formulation. This is particularly interesting for the case of the joint inversion for both Lamé coefficients λ and μ or, likewise, for the velocities of compressional and shear waves. This is known to be a challenging problem [38] and our approach allows to control the deviation of both parameter fields by imposing bounds on the Poisson's ratio.

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 and requires appropriate optimization methods. To this end, 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 [19, 44, 45] and have been applied to various types of applications, see, for instance, [30] for an optimal control problem governed by the wave equation and [18, 20, 21, 22] for problems involving the Moreau-Yosida regularization.

Efficient inversion methods rely on a scalable code for the simulation of the elastic wave equation. Our parallel implementation 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 [26] 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.

This paper is organized as follows. In section 2 we analyze the governing equations and establish continuity and differentiability of the parameter-to-state operator. Based on these results we prove the existence of a solution to the regularized inverse problem and present the semismooth Newton method in section 3. Furthermore, we establish estimates for the constraint violation of the Moreau-Yosida regularized solution. We continue with the discretization of the problem in section 4 and conclude with numerical examples on reservoir scale problems in 2d and 3d in section 5.

2. Analysis of the elastic wave equation.

2.1. Parameterization of the material coefficients. We consider a bounded domain $\Omega \subset \mathbb{R}^d$ $(d = 2, 3)$ with a smooth boundary and denote the time interval by $I := (0, T)$, $T > 0$. Furthermore, we assume a heterogeneous density ρ and a linear elastic rheology and let f denote the seismic source function. The elastic wave equation is then given by

$$\begin{cases} \rho u_{tt} - \nabla \cdot (\Psi : \varepsilon(u)) = f & \text{on } \Omega \times I, \\ u(0) = 0, \quad u_t(0) = 0 & \text{on } \Omega, \\ (\Psi : \varepsilon(u)) \cdot \vec{n} = 0 & \text{on } \partial\Omega \times I. \end{cases} \tag{2.1}$$

Here, u denotes the displacement field, $\varepsilon(u) = \frac{1}{2}\left(\nabla u + \nabla u^T\right)$ is the strain tensor of u and $\Psi = (\Psi_{ijkl})$ is a fourth-order elastic tensor. We write $\Psi : \varepsilon(u)$ for the matrix with entries

$$(\Psi : \varepsilon(u))_{ij} = \sum_{k,l=1}^{d} \frac{1}{2} \Psi_{ijkl}\, \varepsilon(u)_{kl} = \sum_{k,l=1}^{d} \frac{1}{2} \Psi_{ijkl} \left([u_l]_{x_k} + [u_k]_{x_l} \right).$$

REMARK 2.1. The tensor Ψ of elastic moduli has the symmetry properties $\Psi_{ijkl} = \Psi_{jikl} = \Psi_{klij}$, which yields at most 21 independent components for $d = 3$. This general form allows the treatment of anisotropic material. An important special case, however, is a perfectly elastic, isotropic medium. Here, the tensor simplifies to

$$\Psi_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}), \tag{2.2}$$

with the Lamé parameters λ and μ. In this case we have the relation

$$v_p = \sqrt{(\lambda + 2\mu)/\rho}, \quad v_s = \sqrt{\mu/\rho}, \tag{2.3}$$

where v_p and v_s denote the speed of compressional and shear waves [35]. The isotropic elastic wave equation is thus given by:

$$\begin{cases} \rho u_{tt} - \nabla \cdot (2\mu\varepsilon(u) + \lambda(\nabla \cdot u)I) = f & \text{on } \Omega \times I, \\ u(0) = 0, \quad u_t(0) = 0 & \text{on } \Omega, \\ (2\mu\varepsilon(u) + \lambda(\nabla \cdot u)I) \cdot \vec{n} = 0 & \text{on } \partial\Omega \times I. \end{cases} \tag{2.4}$$

Before we continue with the analysis of the state equation, we have to establish a suitable function space setting for the material parameters. Due to the interdependencies of ρ and Ψ, we keep the density fixed and invert for Ψ only. 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 its components is denoted by n. We then split the parameterization of the material into two parts. On the one hand, we use a reference model $\overline{\Psi} \in L^\infty(\Omega)^{d^4}$ that is based on prior knowledge and captures major discontinuities of the material. On the other hand, we consider smooth variations m from that reference model, which we seek to determine by solving the inverse problem. We assume a stronger regularity of the variations with $m \in M := M_1^n$ and a Hilbert space $M_1 \hookrightarrow\hookrightarrow L^\infty(\Omega)$ that is compactly embedded in $L^\infty(\Omega)$. The complete parameterization is now given by

$$\Psi(m) = \overline{\Psi} + \Phi m, \tag{2.5}$$

with a linear function $\Phi : \mathbb{R}^n \to \mathbb{R}^{d^4}$ that maps m to the full elastic tensor.

REMARK 2.2. Note that we rely on the strong assumption of the compact embedding $M \hookrightarrow\hookrightarrow L^\infty(\Omega)^n$ in order to show the existence of a solution to the regularized inverse problem, see Theorem 3.2. It is a common approach in geophysics to parameterize for smooth variations from a reference model [13, 41, 48] which often only varies in depth. On the other hand, we allow for less regular material in the reference model. For problems in global seismology, a suitable reference model based on a priori knowledge is available, e.g., the Preliminary Reference Earth Model (PREM) [9].

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

$$m^a \leq m \leq m^b \quad \text{and} \quad g_i^a \leq \sum_{j=1}^{n} b_{ij} m_j \leq g_i^b, \quad i = 1, \ldots, p. \tag{2.6}$$

In order to facilitate the notation, we set $n_c := n + p$, $M_c := M_1^{n_c}$, $\psi^a := (m^a, g^a)^T$, $\psi^b := (m^b, g^b)^T$ and compactly write the set defined by (2.6) with the help of a linear operator $S \in \mathcal{L}(M, M_c)$ as

$$M_{\text{ad}}^\infty := \left\{ m \in L^\infty(\Omega)^n \,:\, \psi^a \le Sm \le \psi^b \right\}. \tag{2.7}$$

By construction, the set M_{ad}^∞ is a convex, closed and bounded set. Furthermore, it is nonempty since $0 \in M_{\text{ad}}^\infty$. The admissible set is now given by $M_{\text{ad}} = M \cap M_{\text{ad}}^\infty$.

REMARK 2.3. 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, from an application point of view, it seems reasonable to assume a feasible reference model, i.e., zero should be a feasible parameter variation. On a different note, we could also extend the analysis to a more general superposition operator Φ, but restrict the presentation to the linear case. Moreover, it is possible to work with different spaces M_i for every component of the parameters as long as they are compactly embedded into $L^\infty(\Omega)$.

From the definition of the feasible set, it is obvious that simple box constraints can be imposed by choosing $B = 0$. We conclude this subsection with an example for the set M_{ad}^∞ that imposes additional restrictions on the Poisson's ratio of the material.

EXAMPLE 2.4. For isotropic material, the Poisson's ratio ν can be expressed in terms of the Lamé coefficients as $\nu(x) = \lambda(x)/(2(\lambda(x) + \mu(x)))$ [35]. Since λ and μ are positive, lower and upper bounds in the form $\nu^a \le \nu(x) \le \nu^b$ can be rewritten as $2\nu^a\mu + (2\nu^a - 1)\lambda \le 0$ and $-2\nu^b\mu + (1 - 2\nu^b)\lambda \le 0$. 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

$$(2\nu^a - 1)m_1 + 2\nu^a m_2 \le \bar\lambda\,(1 - \frac{\nu^a}{\bar\nu}), \quad (1 - 2\nu^b)m_1 - 2\nu^b m_2 \quad \le \bar\lambda(\frac{\nu^b}{\bar\nu} - 1). \tag{2.8}$$

Thus, we set $p = 2$ and

$$B = \begin{pmatrix} (2\nu^a - 1) & 2\nu^a \\ (1 - 2\nu^b) & -2\nu^b \end{pmatrix}, \quad g^b = \begin{pmatrix} \bar\lambda\,(1 - \frac{\nu^a}{\bar\nu}) \\ \bar\lambda(\frac{\nu^b}{\bar\nu} - 1) \end{pmatrix}. \tag{2.9}$$

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 (2.8), but using the fact that $\nu \in (0, \frac{1}{2}]$, we could easily add an (artificial) lower bound by setting $g^a = ((2\nu^a - 1)m_1^b + 2\nu^a m_2^a, (1 - 2\nu^b)m_1^a - 2\nu^b m_2^b)^T$.

2.2. Existence, uniqueness and regularity of solutions. Now we turn to the discussion of the elastic wave equation and analyze the existence and regularity of solutions. We set $V = H^1(\Omega)^d$ and $H = L^2(\Omega)^d$ such that $V \hookrightarrow H = H^* \hookrightarrow V$ is a Gelfand triple. For fixed $m \in M$ we define the elliptic operator $A(m) \in \mathcal{L}(V, V^*)$ by

$$\langle A(m)v, w\rangle_{V^*,V} = \left(\Psi(m) : \varepsilon(v), \varepsilon(w)\right)_{L^2(\Omega)^{d\times d}} \quad \forall\, v, w \in V. \tag{2.10}$$

In the case of isotropic material, (2.10) can be simplified to: $\forall\, v, w \in V$:

$$\langle A(m)v, w\rangle_{V^*,V} = \left(\lambda(m)\,\nabla\cdot v, \nabla\cdot w\right)_{L^2(\Omega)} + 2\left(\mu(m)\,\varepsilon(v), \varepsilon(w)\right)_{L^2(\Omega)^{d\times d}}. \tag{2.11}$$

The elastic wave equation can then compactly be written as

$$E(u, m) = \rho u_{tt} + A(m)u, \quad u(0) = 0, \quad u_t(0) = 0. \tag{2.12}$$

The variational form of (2.12) reads as follows: For all $v \in V$ and a.a. $t \in I$:

$$\langle \rho u_{tt}(t), v\rangle_{V^*,V} + \langle A(m)u(t), v\rangle_{V^*,V} = \langle f(t), v\rangle_{V^*,V}. \tag{2.13}$$

For the rest of the paper we require the following assumptions on the material and the parameterization:

ASSUMPTION 1.

(A1.1) The density $\rho : \Omega \to \mathbb{R}$ is sufficiently smooth such that $\rho v \in H^1(\Omega)$ for all $v \in H^1(\Omega)$. Furthermore, ρ satisfies $0 < \rho^a \leq \rho(x) \leq \rho^b$ a.e. on Ω with some bounds $\rho^a, \rho^b \in \mathbb{R}$.

(A1.2) There exists a convex, open and bounded set $G^\infty \subset L^\infty(\Omega)^n$ with $M_{\text{ad}}^\infty \subset G^\infty$ and $\beta_1, \beta_2 > 0$, independent of m, such that

$$\langle A(m)v, v\rangle_{V^*,V} + \beta_1 \|v\|_H^2 \geq \beta_2 \|v\|_V^2 \quad \forall\, v \in V,\ \forall\, m \in G^\infty. \tag{2.14}$$

REMARK 2.5. For $d = 2$ or $d = 3$, Assumption (A1.1) 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 inequality with $\frac{1}{p} + \frac{1}{q} = \frac{1}{2}$ and $p = 6$, $q = 3$ or, resp., $p = \infty$, $q = 2$.

In the case of perfectly elastic, isotropic material, Assumption (A1.2) is satisfied, for instance, if there exist positive lower bounds on both Lamé coefficients.

LEMMA 2.6. *Consider the isotropic elastic wave equation* (2.11) *and let the set* M_{ad}^∞*, the reference model* $(\bar\lambda, \bar\mu)$ *and scalars* $\lambda_{\min}, \mu_{\min} > 0$ *be given, such that for all* $m \in M_{ad}^\infty$ *there holds* $\lambda(m) \geq \lambda_{\min}$ *and* $\mu(m) \geq \mu_{\min}$ *a.e. in* Ω*. Then Assumption (A1.2) is satisfied.*

Proof. By construction, there exists an open and bounded neighborhood G^∞ of M_{ad}^∞ such that $\lambda(m) \geq \lambda_{\min}/2 > 0$, $\mu(m) \geq \mu_{\min}/2 > 0$ a.e. in Ω for all $m \in G^\infty$. Furthermore, by Korn's inequality (cf. [25]), there exists a constant $C > 0$ such that

$$(\varepsilon(v), \varepsilon(v))_{L^2(\Omega)^{d\times d}} + \|v\|_{L^2(\Omega)^d}^2 \geq C \|v\|_{H^1(\Omega)^d}^2 \quad \forall\, v \in V. \tag{2.15}$$

Hence, we obtain for all $m \in G^\infty$ and all $v \in V$:

$$\begin{aligned}
\langle A(m)v, v\rangle_{V^*,V} &= (\lambda(m)\nabla\cdot v, \nabla\cdot v)_{L^2(\Omega)} + (2\mu(m)\varepsilon(v), \varepsilon(v))_{L^2(\Omega)^{d\times d}} \\
&\geq \frac{1}{2}\lambda_{\min} (\nabla\cdot v, \nabla\cdot v)_{L^2(\Omega)} + \mu_{\min} (\varepsilon(v), \varepsilon(v))_{L^2(\Omega)^{d\times d}} \\
&\geq \mu_{\min} \left(C \|v\|_V^2 - \|v\|_H^2 \right).
\end{aligned}$$

□

In order to validate (2.14) in the anisotropic case, further physical properties of the elastic tensor have to be exploited that could be incorporated into the definition of the set M_{ad}^∞. For a proof we refer to [7]. The following analysis is restricted to the set G^∞, because a unique solution to the elastic wave equation does not necessarily exist outside of this set.

THEOREM 2.7. *Let* $m \in G^\infty$*. Then the elastic wave equation* (2.12) *possesses a unique solution in the following settings:*

(i) For every $f \in L^2(I;H)$ *there exists a unique solution* $(u, u_t) \in C(\bar I; V) \times C(\bar I; H)$ *and 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* $(u, u_t) \in C(\bar I; H) \times C(\bar I; V^*)$ *and 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.*

Proof. Since $m \in G^\infty$, $A(m)$ is uniformly coercive by Assumption (A1.2). Furthermore, Assumption (A1.1) ensures the applicability of general results for hyperbolic equations. In fact, (A1.1) yields $(\rho v, v)_H \geq \rho_a \|v\|_H^2$ for all $v \in V$ and the usual energy estimates can be used. Hence, the existence and regularity of a unique solution as well as the dependence on the right-hand side f follows by standard arguments (cf. Chapter 3, Theorem 8.1, 8.2 and, respectively, Theorem 9.3, 9.4 in [32]). ∎

Note that in order to achieve continuity of the solution in time, a modification on a set of measure zero might be necessary. The regularity of the solution can be improved by exploiting a higher temporal regularity of the seismic source.

THEOREM 2.8. *Let $m \in G^\infty$ and $f \in L^2(I;V^*)$ with $f = 0$ near $t = 0$ and, additionally, $f_t \in L^2(I;V^*)$. Then there exists a unique solution u to the elastic wave equation* (2.12) *satisfying $u \in C(\bar{I};V)$, $u_t \in C(\bar{I};H)$, $u_{tt} \in C(\bar{I};V^*)$ and the mapping*

$$H^1(I;V^*) \to C(\bar{I};V) \times C(\bar{I};H) \times C(\bar{I};V^*), \quad f \mapsto (u, u_t, u_{tt}) \tag{2.16}$$

is linear and bounded. For fixed f, the set of solutions $u(m) \in C(\bar{I};V) \cap C^1(\bar{I};H)$ is uniformly bounded for all $m \in G^\infty$.

Proof. The proof follows the lines of the proof of Chapter 3, Theorem 9.3 for the existence of very weak solutions in [32]. 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

$$f_k \in H^1(I;H) \quad \text{with} \quad f_k \to f \quad \text{in } H^1(I;V^*), \tag{2.17}$$

and the problem

$$\rho(u_k)_{tt}(t) + A(m)u_k(t) = f_k(t), \quad u_k(0) = 0, \quad (u_k)_t(0) = 0. \tag{2.18}$$

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

$$\rho(w_k)_{tt}(t) + A(m)w_k(t) = (f_k)_t(t), \quad w_k(0) = 0, \quad (w_k)_t(0) = 0. \tag{2.19}$$

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 (2.19) by Theorem 2.7. We define the auxiliary function $v_k(t) := \int_0^t w_k(\tau)\, d\tau$ and obtain

$$v_k \in C^1(\bar{I};V), \quad (v_k)_t = w_k \in C(\bar{I};V), \quad (v_k)_{tt} = (w_k)_t \in C(\bar{I};H). \tag{2.20}$$

Integrating (2.19) in time and inserting the homogeneous initial data yields

$$0 = \int_0^t \rho(w_k)_{tt}(\tau) + A(m)w_k(\tau) - (f_k)_t(\tau)\, d\tau = \rho(v_k)_{tt}(t) + A(m)v_k(t) - f_k(t). \tag{2.21}$$

By subtracting the original equation (2.18) from (2.21), we obtain

$$\rho(v_k - u_k)_{tt}(t) + A(m)(v_k - u_k)(t) = 0, \quad (v_k - u_k)(0) = 0, \quad (v_k - u_k)_t(0) = 0. \tag{2.22}$$

Hence, by the uniqueness of the solution due to Theorem 2.7, 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 (2.18). The improved regularity in space remains to be shown. In [32], Chapter 3, proof of Theorem 9.3, the energy estimates

$$\|(u_k)_t(t)\|_H^2 + \|(u_k)_{tt}(t)\|_{V^*}^2 \leq C \|(f_k)_t\|_{L^2(I;V^*)}^2, \tag{2.23}$$

$$\|(u_k)(t)\|_H^2 + \|(u_k)_t(t)\|_{V^*}^2 \leq C \|(f_k)\|_{L^2(I;V^*)}^2 \tag{2.24}$$

are derived for a.a. $t \in I$. Using (2.14) and Assumption 1, we deduce

$$\begin{aligned}\beta_2 \|u_k(t)\|_V^2 &\le \langle A(m)u_k(t), u_k(t)\rangle_{V^*,V} + \beta_1 \|u_k(t)\|_H^2 \\ &= \langle f_k(t) - \rho(u_k)_{tt}(t), u_k(t)\rangle_{V^*,V} + \beta_1 \|u_k(t)\|_H^2 \\ &\le (\|f_k(t)\|_{V^*} + C_\rho \|(u_k)_{tt}(t)\|_{V^*}) \|u_k(t)\|_V + \beta_1 \|u_k(t)\|_H^2 \\ &\le \left(\|f_k(t)\|_{V^*} + c_1 \|(f_k)_t\|_{L^2(I;V^*)}\right) \|u_k(t)\|_V + c_2 \|(f_k)\|_{L^2(I;V^*)}^2,\end{aligned} \tag{2.25}$$

where we used (2.23) and (2.24) in the last inequality. Next, we use Young's inequality to obtain

$$\begin{aligned}&\left(\|f_k(t)\|_{V^*} + c_1 \|(f_k)_t\|_{L^2(I;V^*)}\right) \|u_k(t)\|_V \le \\ &\frac{1}{2\beta_2}\left(\|f_k(t)\|_{V^*} + c_1 \|(f_k)_t\|_{L^2(I;V^*)}\right)^2 + \frac{\beta_2}{2}\|u_k(t)\|_V^2.\end{aligned} \tag{2.26}$$

Hence,

$$\frac{\beta_2}{2}\|u_k(t)\|_V^2 \le \frac{1}{2\beta_2}\left(\|f_k(t)\|_{V^*} + c_1 \|(f_k)_t\|_{L^2(I;V^*)}\right)^2 + c_2 \|(f_k)\|_{L^2(I;V^*)}^2. \tag{2.27}$$

Now, we can estimate $\|f_k(t)\|_{V^*} \le \|f_k\|_{C(\bar{I};V^*)} \le c\|f_k\|_{H^1(I;V^*)}$ by the embedding $H^1(I;V^*) \hookrightarrow C(\bar{I};V^*)$ and finally obtain by combining (2.23) and (2.27):

$$\sup_{0\le t\le T}\left(\|u_k(t)\|_V^2 + \|(u_k)_t(t)\|_H^2 + \|(u_k)_{tt}(t)\|_{V^*}^2\right) \le C\|f_k\|_{H^1(I;V^*)}^2. \tag{2.28}$$

Thus, u_k (resp. $(u_k)_t$, $(u_k)_{tt}$) remain 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 (2.18) we find that $\tilde{u}$ is a solution to the elastic wave equation.

The continuous dependence in (2.16) follows from (2.28), cf. Chapter 3, Remark 9.11 in [32]. By Assumption (A1.2), all constants are independent of m. Thus, the set of unique solutions $u(m) \in C(\bar{I};V) \cap C^1(\bar{I};H)$ is uniformly bounded for all $m \in G^\infty$ and fixed $f \in H^1(I;V^*)$. ∎

COROLLARY 2.9. *Let $F^k = H^k(I;V^*)$, $k \ge 1$, and $F_0^k \subset F^k$ denote the subset of source functions with $f = 0$ near $t = 0$. Then, for every $f \in F_0^k$ and fixed $m \in G^\infty$, there exists a unique solution $u \in C^{k-1}(\bar{I};V) \cap C^k(\bar{I};H) \cap C^{k+1}(\bar{I};V^*)$ to the elastic wave equation* (2.12). *The mapping*

$$F_0^k \to C^{k-1}(\bar{I};V) \cap C^k(\bar{I};H) \cap C^{k+1}(\bar{I};V^*), \quad f \mapsto u \tag{2.29}$$

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

Proof. The proof follows by induction and Theorem 2.8 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 2.8 we consider

$$\rho(w)_{tt} + A(m)w = f_t, \quad w(0) = 0, \quad w_t(0) = 0. \tag{2.30}$$

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 (2.30). 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 (2.30) in time, we find that v is the unique solution to

$$\rho v_{tt} + A(m)v = f, \quad v(0)=0, \quad v_t(0)=0. \tag{2.31}$$

which concludes the induction step. ∎

REMARK 2.10. Usually, second-order hyperbolic equations require stronger compatibility conditions on the initial values and the right-hand side (cf. §7.2 Theorem 6 in [11]) to establish similar results as in Theorem 2.8 and Corollary 2.9. Since we work with homogeneous initial data, these requirements are implicitly given by $f=0$ near $t=0$. Moreover, in this case also the solution to the elastic wave equation is zero near $t=0$, which we will frequently use in the following.

In the next step, we establish continuity and Fréchet differentiability of the parameter-to-state operator by utilizing a higher temporal regularity of the source.

THEOREM 2.11. *For all $m \in G^\infty$ and $f \in F_0^{k+1}$, $k \geq 1$, the elastic wave equation defined in* (2.12) *possesses a unique solution $u(m) \in C^k(\bar{I};V)\cap C^{k+1}(\bar{I};H)$ and $m \mapsto u(m)$ is Lipschitz continuous as a map from G^∞ to $C^{k-1}(\bar{I};V)\cap C^k(\bar{I};H)$.*

Proof. By Corollary 2.9, there exists a unique solution $u(m) \in C^k(\bar{I};V)\cap C^{k+1}(\bar{I};H)$ to (2.12). Let $s \in L^\infty(\Omega)^n$ satisfy $m+s \in G^\infty$. Then, again by Corollary 2.9, there exists a unique solution $\hat{u}(s) := u(m+s) \in C^k(\bar{I};V)\cap C^{k+1}(\bar{I};H)$ to $E(\hat{u}(s), m+s) = f$ and $\|\hat{u}(s)\|_{C^k(\bar{I};V)} \leq C$ uniformly for all $s \in L^\infty(\Omega)^n$ with $m+s \in G^\infty$. We denote the difference by $h := \hat{u}(s) - u(m)$ and obtain:

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

Using the notation $A_\Delta(s) \in \mathcal{L}(V, V^*)$ for the operator defined by

$$\langle A_\Delta(s)v, w\rangle_{V^*,V} = ((\Phi s) : \varepsilon(v), \varepsilon(w))_{L^2(\Omega)^{d\times d}} \quad \forall\, v, w \in V, \tag{2.32}$$

we obtain $A_\Delta(s) = A(m+s) - A(m)$ and find that h satisfies an elastic wave equation

$$E(h, m) = -A_\Delta(s)\hat{u}(s), \quad h(0)=0, \quad h_t(0)=0. \tag{2.33}$$

Since $\hat{u}(s) \in C^k(\bar{I};V)\cap C^{k+1}(\bar{I};H)$, we obtain $-A_\Delta(s)\hat{u}(s) \in C^k(\bar{I};V^*) \subset H^k(I;V^*)$ and, furthermore, $-A_\Delta(s)\hat{u}(s) \in F_0^k$, due to $f \in F_0^{k+1}$ and the homogeneous initial conditions. Hence, we can apply Corollary 2.9 and obtain

$$\begin{aligned} \|h\|_{C^{k-1}(\bar{I};V)} + \|h_t\|_{C^{k-1}(\bar{I};H)} &\leq C_1 \|A_\Delta(s)\hat{u}(s)\|_{H^k(I;V^*)} \\ &\leq C_2 \|s\|_{L^\infty(\Omega)^n} \|\hat{u}(s)\|_{C^k(\bar{I};V)} \leq C_3 \|s\|_{L^\infty(\Omega)^n}. \end{aligned} \tag{2.34}$$

∎

REMARK 2.12. 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 $u(m)$ with respect to m. In fact, consider $f \in L^2(I;H)$ and $h := \hat{u}(s) - u(m)$ as in the proof of Theorem 2.11. We obtain that h satisfies

$$E(h, m) = -A_\Delta(s)\hat{u}(s), \quad h(0)=0, \quad h_t(0)=0, \tag{2.35}$$

and $\|A_\Delta(s)\hat{u}(s)\|_{L^2(I;V^*)} \leq C\|s\|_{L^\infty(\Omega)^n}$. Thus, from the estimate for very weak solutions (cf. Theorem 2.7) we would only get that $m \mapsto u(m)$ is continuous as a map from G^∞ to $C(\bar{I};H) \cap C^1(\bar{I};V^*)$, although $u(m)$ is bounded in the stronger space $C(\bar{I};V) \cap C^1(\bar{I};H)$. For $f \in L^2(I;V^*)$ the situation is even more critical, because in this case we have $\hat{u}(s) \in C(\bar{I};H)$, but $A_\Delta(s) \in \mathcal{L}(V,V^*)$.

THEOREM 2.13. *Let $k \geq 1$ and $f \in F_0^{k+2}$. Then the solution operator $G^\infty \to C^{k-1}(\bar{I};V) \cap C^k(\bar{I};H)$, $m \mapsto u(m)$ is Fréchet differentiable.*

Proof. Let $m \in G^\infty$ and $s \in L^\infty(\Omega)^n$ with $m + s \in G^\infty$. Similar as in the proof of Theorem 2.11, we define $\hat{u}(s)$ as the solution to $E(\hat{u}(s), m+s) = f$. In order to show the Fréchet differentiability of the parameter-to-state operator we consider the solution to the linearized state equation $d(m,s)$

$$E_u(u(m),m))d(m,s) = -E_m(u(m),m)\,s, \tag{2.36}$$

i.e., $d(m,s)$ satisfies

$$\rho\, d_{tt}(m,s) + A(m)d(m,s) = -A_\Delta(s)u(m),\ d(m,s)(0) = 0,\ d_t(m,s)(0) = 0, \tag{2.37}$$

where $A_\Delta(s)$ is defined as in (2.32). In particular, $s \in L^\infty(\Omega)^n \mapsto d(m,s) \in C^k(\bar{I};V) \cap C^{k+1}(\bar{I};H)$ is linear and bounded by Corollary 2.9.

Let $r := \hat{u}(s) - u(m) - d(m,s)$ denote the remainder term of the Fréchet derivative. By combining the elastic wave equations for $\hat{u}(s)$, $u(m)$ and $d(m,s)$, we deduce

$$\begin{aligned} 0 &= \rho\, r_{tt} + A(m+s)\hat{u}(s) - A(m)(u(m) + d(m,s)) - A_\Delta(s)u(m) \\ &= \rho\, r_{tt} + A(m)r + A_\Delta(s)(\hat{u}(s) - u(m)). \end{aligned} \tag{2.38}$$

Hence, r satisfies an elastic wave equation $E(r,m) = -A_\Delta(s)(\hat{u}(s) - u(m))$ with homogeneous initial data. With $h := \hat{u}(s) - u(m)$, we obtain $h \in C^{k+1}(\bar{I};V) \cap C^{k+2}(\bar{I};H)$ and $-A_\Delta(s)h \in F_0^{k+1}$. Furthermore, by the Lipschitz continuity of the solution operator as a map from G^∞ to $C^k(\bar{I};V) \cap C^{k+1}(\bar{I};H)$ (cf. Theorem 2.11), we have $\|h\|_{C^k(\bar{I};V)} \leq C\|s\|_{L^\infty(\Omega)^n}$. Thus, applying Corollary 2.9 yields the estimate

$$\begin{aligned} \|r\|_{C^{k-1}(\bar{I};V)} + \|r_t\|_{C^{k-1}(\bar{I};H)} &\leq C_1 \left\|A_\Delta(s)h\right\|_{H^k(I;V^*)} \\ &\leq C_2 \left\|s\right\|_{L^\infty(\Omega)^n} \left\|h\right\|_{C^k(\bar{I};V)} \leq C_3 \left\|s\right\|^2_{L^\infty(\Omega)^n}. \end{aligned} \tag{2.39}$$

In particular, the remainder term r satisfies

$$\|r\|_{C^{k-1}(\bar{I};V)} + \|r_t\|_{C^{k-1}(\bar{I};H)} = o\left(\|s\|_{L^\infty(\Omega)^n}\right) \quad \text{for} \quad \|s\|_{L^\infty(\Omega)^n} \to 0. \tag{2.40}$$

□

By further pursuing this technique, we can also establish higher-order Fréchet differentiability of the solution operator under suitable assumptions on the seismic source. Here, we will require the following notation. Let $\mathbf{s} = (s_1, \ldots, s_i)$ be a tuple of length i with $s_j \in L^\infty(\Omega)^n$, $j = 1, \ldots, i$. Furthermore, for $1 \leq j \leq i$, we denote by $\mathbf{s}_{-j} := (s_1, \ldots, s_{j-1}, s_{j+1}, \ldots, s_i)$ the $(i-1)$-tuple of elements of $\mathbf{s}$ except s_j.

THEOREM 2.14. *Let $k \geq 1$ and $f \in F_0^{k+2}$. Then the solution operator $G^\infty \to C(\bar{I};V) \cap C^1(\bar{I};H)$, $m \mapsto u(m)$ is k-times Lipschitz continuously Fréchet differentiable on G^∞. Furthermore, let $\mathbf{s} = (s_1, \ldots, s_i)$ with $s_j \in L^\infty(\Omega)^n$ and $m + s_j \in G^\infty$, $j = 1, \ldots, i$. Then the i-th derivative applied to $\mathbf{s}$ is recursively defined as the unique*

solution $d^i := d^i(m, \mathbf{s})$ *to*

$$E(d^1, m) = -A_\Delta(s_1)\, u(m), \qquad d^1(0) = 0, \quad d^1_t(0) = 0 \qquad (i = 1), \tag{2.41}$$

$$E(d^i, m) = -\sum_{j=1}^{i} A_\Delta(s_j)\, d^{i-1}(m, \mathbf{s}_{-j}), \quad d^i(0) = 0, \quad d^i_t(0) = 0 \quad (2 \le i \le k). \tag{2.42}$$

Proof. For the sake of brevity and since the same techniques as in the proof of Theorem 2.13 can be applied, we only provide a sketch of the proof.

First, it can be shown inductively that $d^i \in C^{k-i+1}(\bar{I}; V) \cap C^{k-i+2}(\bar{I}; H)$ is uniquely determined by (2.41), (2.42) and satisfies

$$\left\|d^i\right\|_{C^{k-i+1}(\bar{I};V)} + \left\|d^i_t\right\|_{C^{k-i+1}(\bar{I};H)} \le C \prod_{j=1}^{i} \|s_j\|_{L^\infty(\Omega)^n}, \quad 1 \le i \le k, \tag{2.43}$$

with a constant $C > 0$ that depends linearly on f and can be chosen uniformly on G^∞. Here, Theorem 2.13 gives the induction basis for $i = 1$ and the induction step is obtained by applying Corollary 2.9 to the elastic wave equation defined in (2.42).

In a second step, we consider $h^i(m, s_{i+1}, \mathbf{s}) := d^i(m+s_{i+1}, \mathbf{s}) - d^i(m, \mathbf{s})$ with $s_{i+1} \in L^\infty(\Omega)^n$ and $m + s_{i+1} \in G^\infty$. With $h^0(m, s_1, \emptyset) := h^0(m, s_1) := u(m + s_1) - u(m)$, we deduce from (2.41), (2.42) that $h^i = h^i(m, s_{i+1}, \mathbf{s})$ satisfies:

$$E(h^i, m) = -\sum_{j=1}^{i} A_\Delta(s_j) h^{i-1}(m, s_{i+1}, \mathbf{s}_{-j}) - A_\Delta(s_{i+1}) d^i(m + s_{i+1}, \mathbf{s}), \tag{2.44}$$

and $h^i(0) = 0$, $h^i_t(0) = 0$. Again, we can show inductively that $h^i \in C^{k-i}(\bar{I}; V) \cap C^{k-i+1}(\bar{I}; H)$ is uniquely determined by (2.44) and satisfies

$$\left\|h^i\right\|_{C^{k-i}(\bar{I};V)} + \left\|h^i_t\right\|_{C^{k-i}(\bar{I};H)} \le C \prod_{j=1}^{i+1} \|s_j\|_{L^\infty(\Omega)^n}, \quad 0 \le i \le k, \tag{2.45}$$

where the constant $C > 0$ depends linearly on f and can be chosen uniformly on G^∞. In this case, (2.34) gives the induction basis for $i = 0$ and the induction step is obtained by applying Corollary 2.9 to the elastic wave equation defined in (2.44). This shows the Lipschitz continuity of d^i.

As the final step, we have to provide an estimate for the remainder term of the i-th Fréchet derivative. To this end, we set $r^i(m, \mathbf{s}) := h^{i-1}(m, s_i, \mathbf{s}_{-i}) - d^i(m, \mathbf{s})$. By combining (2.42) and (2.44) we obtain the recursive formula for $r^i = r^i(m, \mathbf{s})$:

$$E(r^i, m) = -\sum_{j=1}^{i-1} A_\Delta(s_j) r^{i-1}(m, \mathbf{s}_{-j}) - A_\Delta(s_i) h^{i-1}(m, s_i, \mathbf{s}_{-i}), \tag{2.46}$$

and $r^i(0) = 0$, $r^i_t(0) = 0$. Now, a third induction shows the estimate

$$\left\|r^i\right\|_{C^{k-i}(\bar{I};V)} + \left\|r^i_t\right\|_{C^{k-i}(\bar{I};H)} \le o\left(\|s_i\|_{L^\infty(\Omega)^n}\right) \prod_{j=1}^{i-1} \|s_j\|_{L^\infty(\Omega)^n}, \tag{2.47}$$

for $\|s\|_{L^\infty(\Omega)^n} \to 0$. Here, (2.40) gives the induction basis and subsequently applying Corollary 2.9 to (2.46) yields the induction step. ∎

There are two immediate generalizations of Theorem 2.14:

COROLLARY 2.15. *Let $k \geq 1$, $l \geq 0$ and $f \in F_0^{k+l+2}$. Then on a neighborhood G^∞ of M_{ad}^∞ it holds:*

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

We conclude this section with two remarks.

REMARK 2.16. It is a valid assumption for problems in seismic tomography to have homogeneous initial conditions and a force term that is zero near $t = 0$, which means that the simulation starts with the system at rest. In addition, the excitations at the hypocenter usually last only for a limited time which is significantly shorter than the traveltimes to the receiver locations in the far field. An extension to inhomogeneous initial data is possible, but beyond the scope of this paper. Moreover, the assumption of a higher temporal regularity of the right-hand side 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 $V^* = (H^1(\Omega)^d)^*$. The specific setup will be given in Example 3.4.

REMARK 2.17. The results of this section can be carried over to the acoustic wave equation that describes the propagation of seismic waves in an inviscid fluid medium. Here, the unknown material parameter is given by the squared velocity of compressional waves. Similar results on the Fréchet differentiability of the parameter-to-state operator have been established in [3] and [27]. However, both work with stronger assumptions on the right-hand sides, namely f in $H^k(I; H)$ or even $C^k(\bar{I}, H)$ instead of $H^k(I; V^*)$, which is advantageous for modeling point sources.

3. Semismooth Newton method.

3.1. Analysis of the seismic inverse problem. Having established the continuity and the differentiability of the parameter-to-state operator, we now turn to the formulation of the inverse problem. 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, $i = 1, \ldots, n_s$. Specific assumptions on the regularity of f_i will be made later in this section. Furthermore, for every event we assume to have observations on a subdomain $\Omega_i^\delta \times I$, $\Omega_i^\delta \subset \Omega$ and denote the data by $u_i^\delta \in L^2(I; L^2(\Omega_i^\delta)^d) \cap H^2(I; (H^1(\Omega_i^\delta)^d)^*)$. The seismic inverse problem is given by:

$$\begin{aligned} &\min_{\mathbf{u} \in \mathbf{U},\, m \in M_{\mathrm{ad}}} \quad J(\mathbf{u}, m) \\ &\text{s.t.} \quad E(u_i, m) = f_i, \quad u_i(0) = 0, \quad (u_i)_t = 0, \quad i = 1, \ldots, n_s. \end{aligned} \tag{3.1}$$

Here, $\mathbf{u} = (u_1, \ldots, u_{n_s})^T \in \mathbf{U} = U^{n_s}$ is a vector of displacement fields and we set

$$U := L^2(I; V) \cap H^1(I; H) \cap H^2(I; V^*). \tag{3.2}$$

REMARK 3.1. Following the analysis of the elastic wave equation in the previous section, we recall that for a source $f \in F_0^{l+1}(I; V^*)$ 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.

The cost functional $J : \mathbf{U} \times M \to \mathbb{R}$ has the special structure

$$J(\mathbf{u}, m) = \sum_{i=1}^{n_s} J_{\mathrm{fit}}(u_i, u_i^\delta) + \frac{\alpha}{2} \|m\|_M^2, \tag{3.3}$$

that consists of the accumulated misfit from all seismic sources and a Tikhonov-type regularization term with parameter $\alpha > 0$. The existence of a solution to the seismic inverse problem can be proven in the following setting.

ASSUMPTION 2. In addition to Assumption 1, we require:

(A2.1) M is a Hilbert space with compact embedding $M \hookrightarrow\hookrightarrow L^\infty(\Omega)^n$.

(A2.2) $J_{\text{fit}} \geq 0$, convex and, with $l \geq 0$, there holds for $i = 1, \dots, n_s$:

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

THEOREM 3.2. *Let Assumption 2 hold. Then the seismic inverse problem* (3.1) *possesses a solution.*

Proof. The proof follows by standard arguments, cf. Theorem 1.45 in [24]. The compact embedding $M \hookrightarrow\hookrightarrow L^\infty(\Omega)^n$ is required to show that the elastic wave equation is weakly sequentially continuous. ∎

Due to the unique solvability of the elastic wave equation for fixed m, we replace (3.1) by the reduced problem:

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

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

$$E(u_i, m) = f_i, \quad u_i(0) = 0, \quad (u_i)_t(0) = 0, \quad i = 1, \dots, n_s. \tag{3.4}$$

The reduced cost functional j is defined by

$$j : G \to \mathbb{R}, \quad m \mapsto \sum_{i=1}^{n_s} J_{\text{fit}}(u_i(m), u_i^\delta) + \frac{\alpha}{2} \|m\|_M^2, \tag{3.5}$$

where $G := M \cap G^\infty$. The restriction to G is required, because $u_i(m)$ is not necessarily well-defined on the whole space M. We will outline a strategy to circumvent this difficulty in Remark 3.13.

By Theorem 3.2, also the reduced seismic inverse problem (P) admits a solution. 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. In addition to Assumption 1, we require with $k \geq 1$:

(A3.1) M is a Hilbert space with continuous embedding $M \hookrightarrow L^\infty(\Omega)^n$.

(A3.2) With $l \geq 0$, there holds for $i = 1, \dots, n_s$:

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

THEOREM 3.3. *Let Assumption 3 be satisfied for given $k \geq 1$. Then the reduced cost functional j defined in* (3.5) *is k-times continuously Fréchet differentiable on G.*

Proof. Assumption 3 ensures the applicability of Corollary 2.9 and Corollary 2.15 that provide the unique solvability of the elastic wave equation and the differentiability of the parameter-to-state operator from G^∞ 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 differentiability properties of J_{fit} and the regularization term can be carried over to j. ∎

Next, we give a specific example for a typical problem setup and verify Assumption 2 and Assumption 3 for it.

EXAMPLE 3.4. Let $M_1 = H^2(\Omega) \cap H_0^1(\Omega)$, i.e., we have $M_1 \hookrightarrow\hookrightarrow L^\infty(\Omega)$ for $d = 2, 3$. We assume homogeneous boundary data, since the material at the surface is known and we do not want to update the material parameters at the artificial boundaries of the computational domain (cf. section 5). 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. [33]. 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 $x_{r_1}, \dots, x_{r_p}$ 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 misfit function is given by

$$J_{\text{fit}}(u, u^\delta) = \frac{1}{2} \int_0^T g(t) \|u(t) - u^\delta(t)\|^2_{L^2(\Omega^\delta)^d}\, dt. \tag{3.7}$$

Here, $g : [0,T] \to \mathbb{R}$ 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_{fit}, Assumption 2 and Assumption 3 are clearly satisfied with $k = 2$ and $l = 0$.

REMARK 3.5. The proper choice of the misfit functional is crucial for the success of the inversion. The L^2-misfit is widely used, cf. [10, 34], and suitable for synthetic data with a small amount of amount of Gaussian noise. We emphasize, however, that there exist more sophisticated misfit criteria separating phase and amplitude information of the seismograms, see, for instance, [13, 29, 46]. Note that different criteria can easily be incorporated into the inversion framework, as long as they are sufficiently smooth, cf. Assumption (A3.2).

3.2. 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 with respect to m. Note that the displacement fields for different seismic events can be separated completely in (3.1). Likewise, the adjoint states can be computed independently. Due to the symmetry of A and by carefully integrating by parts, we obtain that the adjoint states $p_i = p_i(m)$ are (very) weak solutions of the adjoint equations

$$\begin{cases} \rho (p_i)_{tt} - \nabla \cdot (\Psi(m) : \varepsilon(p_i)) = -g\,(u_i(m) - u_i^\delta) \mathbb{1}_{\Omega_i^\delta} & \text{on } \Omega \times I, \\ p_i(T) = 0, \quad (p_i)_t(T) = 0 & \text{on } \Omega, \\ (\Psi(m) : \varepsilon(p_i)) \cdot n = 0 & \text{on } \partial\Omega \times I. \end{cases} \tag{3.8}$$

The adjoint equation (3.8) can be interpreted as an elastic wave equation backwards in time with a different right-hand side. Thus, we can apply the results from section 2 to establish existence, uniqueness and regularity of the adjoint states. For the adjoint-based representation of j', we note that $E_m(u(m), m)\, s = A_\Delta(s) u(m)$ and introduce

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

$$\begin{aligned} D(v,w)(s) &= \int_0^T \langle A_\Delta(s)v(t), w(t)\rangle_{V^*,V}\, dt \\ &= \int_0^T \int_\Omega (\varepsilon(v)(x,t) \otimes \varepsilon(w)(x,t)) :: ((\Phi s)(x))\, dx\, dt \quad \forall\ s \in M. \end{aligned} \tag{3.9}$$

Here, we used the notation $(a \otimes b)_{ijkl} = a_{ij}b_{kl}$ and $A :: B = \sum_{ijkl} A_{ijkl}B_{ijkl}$ for tensor products. Then the first derivative of j at a given $m \in G$ can be expressed by

$$\langle j'(m), s\rangle_{M^*,M} = \alpha(m,s)_M + \sum_{i=1}^{n_s} D(u_i(m), p_i(m))(s) \quad \forall\ s \in M, \tag{3.10}$$

where we used the Riesz representation in the first term.

LEMMA 3.6. *Let $m \in G$, $u_i^\delta \in L^2(I; L^2(\Omega_i^\delta)^d) \times H^2(I; (H^1(\Omega_i^\delta)^d)^*)$ and $f_i \in F_0^3$, $i = 1, \dots, n_s$. Then there exists a unique adjoint state $p_i(m) \in C^1(\bar{I}; V) \cap C^2(\bar{I}; H)$ and $p_i(m) \in C^1(\bar{I}; V) \cap C^2(\bar{I}; H)$ is uniformly bounded for all $m \in G$. Moreover, for any subset $G' \subseteq G$ that is bounded in M, $j'(m) \in M^*$ is uniformly bounded on G'.*

Proof. We use a time transformation $\tau := T - t$ in order to work with initial time conditions and drop the index i. By Corollary 2.9, we obtain $(u - u^\delta) \in H^2(I; V^*)$ and the weighting g ensures that the adjoint right-hand side is in F_0^2. and we deduce the existence of a unique adjoint state $p(m) \in C^1(\bar{I}; V) \cap C^2(\bar{I}; H)$. Furthermore, since $u(m) \in U$ is uniformly bounded on G, the adjoint right-hand side is uniformly bounded in $H^2(I; V^*)$ and hence, also $p(m) \in C^1(\bar{I}; V) \cap C^2(\bar{I}; H)$ is uniformly bounded on G. Moreover, for every m in G we obtain for arbitrary $s \in M$

$$\begin{aligned} |\langle j'(m), s\rangle_{M^*,M}| &\le \alpha \|m\|_M \|s\|_M + \sum_{i=1}^{n_s} c \|u_i(m)\|_{L^2(I;V)} \|p_i(m)\|_{L^2(I;V)} \|s\|_{L^\infty(\Omega)^n} \\ &\le (\alpha \|m\|_M + C)\|s\|_M. \end{aligned} \tag{3.11}$$

Thus, $j'(m) \in M^*$ is uniformly bounded on any bounded subset $G' \subseteq G$. ∎

We can also employ the adjoint approach to compute operator-vector products $j''(m)s$ for a given perturbation $s \in M$. This can be carried out at the cost of two additional simulations per seismic event and requires the following steps:

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

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

$$E(\delta_s u_i, m) = -A_\Delta(s)u_i(m), \quad \delta_s u_i(0) = 0, \quad (\delta_s u_i)_t(0) = 0. \tag{3.12}$$

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

$$E(\delta_s p_i, m) = -g\, \delta_s u_i - A_\Delta(s)p_i(m),\ \delta_s p_i(T) = 0,\ (\delta_s p_i)_t(T) = 0. \tag{3.13}$$

Then $j''(m)s$ is given by: $\forall\ v \in M$

$$\langle j''(m)s, v\rangle_{M^*,M} = \alpha(s,v)_M + \sum_{i=1}^{n_s} D(\delta_s u_i, p_i(m))(v) + D(u_i(m), \delta_s p_i)(v). \tag{3.14}$$

Note that the results from section 2 and Lemma 3.6 can be applied to deduce that $\delta_s u_i$ and $\delta_s p_i$ are uniquely determined and bounded in U.

3.3. Moreau-Yosida regularization. Now we turn to the discussion of the optimization method and continue to work with the setting of Example 3.4. 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. [18, 20, 21]. For fixed $\gamma \in (0,\infty)$ we define the penalized problem

$$\min_{m\in G} \; j_\gamma(m) := j(m) + \gamma\phi(m), \tag{P_γ}$$

with the penalty function

$$\phi(m) := \frac{1}{2}\left(\|[Sm-\psi^b]^+\|^2_{L^2(\Omega)^{n_c}} + \|[\psi^a - Sm]^+\|^2_{L^2(\Omega)^{n_c}} \right). \tag{3.15}$$

Here, $[.]^+$ is a vector defined pointwise by $([v(x)]^+)_i = \max\{v_i(x),0\}$, $i=1,\dots,n_c$. The first-order optimality conditions for (P_γ) are given by:

$$j'(\bar m) + \gamma S^*\left([S\bar m-\psi^b]^+ - [\psi^a - S\bar m]^+\right) = 0 \quad \text{in } M^*. \tag{3.16}$$

Equivalently, we obtain in variational form

$$\langle j'(\bar m), v\rangle_{M^*,M} + \gamma\left([S\bar m-\psi^b]^+ - [\psi^a - S\bar m]^+, Sv\right)_{L^2(\Omega)^{n_c}} = 0 \quad \forall\, v\in M. \tag{3.17}$$

LEMMA 3.7. *Let Assumption 3 hold with $k\geq 2$. Then the optimality condition* (3.16) *defines a semismooth operator equation with generalized derivative $j''(\bar m) + \gamma S^*\partial\mathcal{D}(\bar m)S$, where $\partial\mathcal{D}(\bar m)$ is the set of all operators of the form $L^\infty(\Omega)^{n_c} \ni v \mapsto (g_i\, v_i)_{1\leq i\leq n_c}$ with $g_i\in L^\infty(\Omega)$, $i=1,\dots,n_c$, and*

$$g_i(x)\begin{cases} = & 0 & \text{if} & \psi_i^a(x) < (S\bar m)_i(x) < \psi_i^b(x), \\ = & 1 & \text{if} & (S\bar m)_i(x) < \psi_i^a(x) \text{ or } (S\bar m)_i(x) > \psi_i^b(x), \\ \in & [0,1] & \text{if} & (S\bar m)_i(x) = \psi_i^a(x) \text{ or } (S\bar m)_i(x) = \psi_i^b(x). \end{cases} \tag{3.18}$$

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 $(g_i)_{1\leq i\leq n_c}$ (cf. Proposition 4.1 in [19] 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 (3.16) is semismooth. ∎

We require the following additional assumption on the constraints.

ASSUMPTION 4.

(A4.1) Either, $M_1 \hookrightarrow W^{1,q'}(\Omega)$ holds for some $q'>d$,
or $M_1\hookrightarrow C^{0,\beta}(\bar\Omega)$ with $0<\beta\leq 2$ and $m=0$ on $\partial\Omega$ for all $m\in M$.

(A4.2) 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 $i=1,\dots,n_c$.

THEOREM 3.8. *Let Assumption 2 and Assumption 3 hold with $k\geq 2$ and, additionally, let Assumption (A4.1) be satisfied. Furthermore, let $\mathcal{C}^\infty\subset G^\infty$ 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. First, we show that there exists $\gamma_0>0$ and a closed and convex set $\mathcal{C}\subset G$ that is bounded in M with $0\in\mathcal{C}$ and such that for all $\gamma\geq\gamma_0$:

$$j_\gamma(m) \geq j_{\gamma_0}(m) > j_{\gamma_0}(0) = j(0) = j_\gamma(0) \quad \forall\, m\in G\setminus\mathcal{C}.$$

This will guarantee the existence of a solution $m_\gamma \in \mathcal{C}$ to (P_γ) in for all $\gamma \geq \gamma_0$ as well as the boundedness of $(m_\gamma)_{\gamma\geq\gamma_0}$ in M. Since $0 \in M_{\text{ad}}^\infty$, we have

$$j_\gamma(0) = \frac{1}{2}\sum_{i=1}^{n_s} J_{\text{fit}}(u_i(0), u_i^\delta) =: J_0 \geq 0, \tag{3.19}$$

and with $\varepsilon := \sqrt{2J_0/\alpha}$ and $B_\epsilon(0) := \{m \in M : \|m\|_M \leq \epsilon\}$, we set $\mathcal{C} := \mathcal{C}^\infty \cap B_\varepsilon(0)$. Next, we prove that there exists $\gamma_0 > 0$ such that $j_{\gamma_0}(m) > j_{\gamma_0}(0) = J_0$ for all $m \in G \setminus \mathcal{C}$. If this does not hold, then there exist sequences $(\gamma_k)_{k\in\mathbb{N}}$ with $\gamma_k \to \infty$ and $(m_k)_{k\in\mathbb{N}} \subset B_\epsilon(0) \setminus \mathcal{C}^\infty$ with

$$\frac{\alpha}{2}\|m_k\|_M^2 + \gamma_k\phi(m_k) \leq j_{\gamma_k}(m_k) \leq J_0 \quad \forall\, k \in \mathbb{N}, \tag{3.20}$$

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, (3.20) yields $\phi(m_k) \to 0$ and for $v_k^+ := [Sm_k - \psi^b]^+ + [\psi^a - Sm_k]^+$, we deduce $\|v_k^+\|_{L^2(\Omega)^{n_c}} \to 0$. Moreover, $(Sm_k)_{k\in\mathbb{N}}$ is bounded in M_c and by Assumption (A4.1) also bounded in $W^{1,q'}(\Omega)^{n_c}$ with $q' > d$ or, respectively, bounded in $C^{0,\beta}(\bar\Omega)^{n_c}$ with $0 < \beta \leq 2$. Hence, v_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 $W^{1,q'}(\Omega)$ (cf. Theorem 5.10 in [1]) or $C^{0,\beta}(\bar\Omega)$ (cf. Proposition 2.11 in [23]) yields $\|v_k^+\|_{L^\infty(\Omega)^{n_c}} \to 0$. Thus, $m_k \in \mathcal{C}^\infty$ for k sufficiently large. This is a contradiction.

Hence, we can restrict the analysis to the set $\mathcal{C} \subset G$ and $\gamma \geq \gamma_0$ sufficiently large. By construction, $A(m)$ is uniformly coercive for all $m \in \mathcal{C}$. ϕ is convex and continuous, hence, j_γ is weakly lower semi-continuous and the existence of a solution $m_\gamma \in \mathcal{C}$ to the regularized problem (P_γ) follows again 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:

$$\frac{\alpha}{2}\|m_\gamma\|_M^2 \leq j(m_\gamma) \leq j(m_\gamma) + \gamma\phi(m_\gamma) = j_\gamma(m_\gamma) \leq j_\gamma(\bar m) = j(\bar m). \tag{3.21}$$

Since $(m_\gamma)_{\gamma\geq\gamma_0}$ is bounded in M, there exist weak limit points. Moreover, (3.21) 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 $L^2(\Omega)^n$). Since $\phi(m_{\gamma_k}) \to 0$, we obtain again by an interpolation inequality $\|v_k^+\|_{L^\infty(\Omega)^{n_c}} \to 0$, i.e., m^* is feasible for (P). It remains to be shown, that m^* is a solution to (P). The sequence of optimal function values of the penalized problem is monotonically increasing, since

$$j_{\gamma_k}(m_{\gamma_k}) \leq j_{\gamma_k}(m_{\gamma_{k+1}}) \leq j_{\gamma_{k+1}}(m_{\gamma_{k+1}}). \tag{3.22}$$

Together with (3.21) this implies that $(j_{\gamma_k}(m_{\gamma_k}))_{k\in\mathbb{N}}$ converges. By the lower semi-continuity of j, we obtain

$$j(m^*) \leq \liminf_{k\to\infty} j(m_{\gamma_k}) \leq \liminf_{k\to\infty} j_{\gamma_k}(m_{\gamma_k}) = \lim_{k\to\infty} j_{\gamma_k}(m_{\gamma_k}) \leq j(\bar m). \tag{3.23}$$

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

We continue to use the notation $v_\gamma^+ = [Sm_\gamma - \psi^b]^+ + [\psi^a - Sm_\gamma]^+$ and state the following result concerning the rate of convergence.

THEOREM 3.9. *Let Assumption 2 and Assumption 3 hold with $k \geq 2$, and let Assumption (A4.1) be satisfied with the embedding $M_1 \hookrightarrow W^{1,q'}(\Omega)$. Furthermore, let $(m_{\gamma_k})_{k\in\mathbb{N}}$ be a weakly convergent subsequence with $m_{\gamma_k} \rightharpoonup m^* \in M$. Then the infeasibility of solutions m_γ to (P_γ) is bounded by*

$$\|v_{\gamma_k}^+\|_{L^2(\Omega)^{n_c}} = o\left({\gamma_k}^{-\frac{1}{2}}\right) \quad (\gamma_k \to \infty) \quad \textit{and} \tag{3.24}$$

$$\|v_{\gamma_k}^+\|_{L^\infty(\Omega)^{n_c}} = o\left(\gamma_k^{-\eta}\right) \quad \textit{with} \quad \eta = \frac{q'-d}{q'd+2(q'-d)}. \tag{3.25}$$

Proof. From (3.21) we obtain

$$j_{\gamma_k}(m_{\gamma_k}) = j(m_{\gamma_k}) + \frac{\gamma_k}{2}\|v_{\gamma_k}^+\|^2_{L^2(\Omega)^{n_c}} \leq j(\bar{m}), \tag{3.26}$$

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

$$\|v_{\gamma_k}^+\|^2_{L^2(\Omega)^{n_c}} \leq \frac{2}{\gamma_k}\left(j(\bar{m}) - j(m_{\gamma_k})\right). \tag{3.27}$$

By Theorem 3.8, m^* solves (P), hence, $j(m_{\gamma_k}) \to j(\bar{m})$ which shows

$$\|v_{\gamma_k}^+\|_{L^2(\Omega)^{n_c}} = o\left({\gamma_k}^{-\frac{1}{2}}\right). \tag{3.28}$$

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 (A4.1) also bounded in $W^{1,q'}(\Omega)^{n_c}$. Thus, an interpolation inequality yields the L^∞-estimate similar to Lemma 8.26 in [45]. ∎

EXAMPLE 3.10. 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

$$\|v_{\gamma_k}^+\|_{L^\infty(\Omega)^{n_c}} = o\left(\gamma_k^{-\frac{1}{8}}\right). \tag{3.29}$$

Following the derivation in [23] we obtain alternative estimates using interpolation between $L^1(\Omega)$ and $C^{0,\beta}(\bar{\Omega})$.

THEOREM 3.11. *Let Assumption 2 and Assumption 3 hold with $k \geq 2$. Additionally, let Assumption 4 be satisfied with the embedding $M_1 \hookrightarrow C^{0,\beta}(\bar{\Omega})$ and $m = 0$ on $\partial\Omega$ for all $m \in M$. Consider a weakly convergent subsequence $(m_{\gamma_k})_{k\in\mathbb{N}}$ with $m_{\gamma_k} \rightharpoonup m^* \in M$. Then we obtain the following estimate on the constraint violation:*

$$\|v_{\gamma_k}^+\|_{L^\infty(\Omega)^{n_c}} \leq C\gamma_k^{-\eta} \quad \textit{with } \eta = \frac{\beta}{\beta+d}. \tag{3.30}$$

Proof. The proof follows from [23], Corollary 2.6, but 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 (A4.2) and define $w := m_{\gamma_k} - \frac{1}{2}\hat{m}$. Now we observe 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 (3.16) with w we obtain from (3.17):

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

Now, for γ_k sufficiently large, we have $m_{\gamma_k} \in \mathcal{C}$ with $\mathcal{C} \subset G$ as defined in the proof of Theorem 3.8. Hence, $j'(m_{\gamma_k})$ is uniformly bounded on $\mathcal{C}$ by Lemma 3.6, and we conclude

$$\left| -\frac{2}{\bar{\psi}} \langle j'(m_{\gamma_k}), w \rangle_{M^*,M} \right| \leq c\|w\|_M \leq c \left(\|m_{\gamma_k}\|_M + \frac{1}{2}\|\hat{m}\|_M \right) \leq C. \tag{3.31}$$

The rest of the proof follows from Corollary 2.6 in [23]. □

EXAMPLE 3.12. Again, we consider $M_1 = H^2(\Omega) \cap H_0^1(\Omega)$. Hence, we have the embedding $M_1 \hookrightarrow C^{0,\beta}(\bar{\Omega})$ with $\beta = \frac{1}{2}$ for $d = 2, 3$ and for $d = 3$ we get the estimate

$$\|v_{\gamma_k}^+\|_{L^\infty(\Omega)^n} \leq C\gamma_k^{-\frac{1}{7}}. \tag{3.32}$$

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) 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**

In order to solve (3.1), we have to compute solutions to a sequence of penalized problems (P_γ) with increasing penalty parameter γ, which is described in Algorithm 1. It should be emphasized that due to the non-convexity of the problem, we cannot expect to attain global solutions of (P_γ). Here, we rely on a good starting point in the vicinity and a suitable regularization parameter.

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 and when the last step provided a good progress towards optimality of (P_γ). To this end, we 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 5. We emphasize,

however, that more sophisticated strategies on updating γ exist [20]. This would be an interesting field for future research.

REMARK 3.13. The Moreau-Yosida regularization does not ensure that m stays within the set G during the optimization process. This can yield an operator $A(m)$ that is not uniformly coercive, in which case $u(m)$ is not well-defined. To overcome this difficulty, we can replace Φ in (2.5) by a nonlinear superposition operator that utilizes a smooth cutoff function to guarantee that the parameters always remain within a certain range and allows us to choose $G = M$. For details we refer to [4].

3.4. Trust-region Newton-CG method. In the next step, we turn to the discussion of solving the problem (P_γ) for a fixed penalty parameter γ. We apply a trust-region Newton-CG method to solve (P_γ), i.e., we iteratively compute approximate solutions to the trust-region subproblem

$$\begin{aligned} \min_{s\in M} \quad & q_i(s) := \left\langle j'_\gamma(m^i), s\right\rangle_{M^*,M} + \tfrac{1}{2}\left\langle (j''(m^i) + \gamma S^* H_i S)s, s\right\rangle_{M^*,M} \\ & \text{s.t.} \quad \|s\|_M \le \Delta_i. \end{aligned} \tag{3.33}$$

Here m^i denotes the current iterate and Δ_i the trust-region radius in iteration i and q is a quadratic model function with $H_i \in \partial\mathcal{D}(m^i)$ and $\partial\mathcal{D}(m^i)$ as defined in Lemma 3.7. The first derivatives $j'(m^i)$ and operator-vector products $j''(m^i)s$ are computed using adjoint-based techniques as outlined above. Instead of the exact second derivatives in (3.33), approximations of the Hessian, e.g. by a Quasi-Newton method, can also be used. We compute an approximate solution to (3.33) by the Steihaug conjugate gradient method [40]. Here, the inner product induced by the norm of M is used as preconditioner. The CG iterations are early terminated if one of the following stopping criteria is met: a direction of negative curvature is encountered, the trust region radius is exceeded by the current iterate, the relative residual is smaller than a threshold, or a maximum number of CG iterations is reached.

Under a standard regularity condition on H_i in (3.33), local superlinear convergence of the semismooth Newton method can be established, provided that the initial model m^0 is chosen sufficiently close to a solution $\bar{m}$, cf. Theorem 2.12 in [24]. The superlinear rate of convergence can be maintained if the generalized Newton system is solved inexactly, cf. Algorithm 3.16 and Theorem 3.18 in [45]. To this end, we require the Dennis-Moré conditions [8], see also Assumption 3.14 in [45]. The regularity condition that is required to ensure fast local convergence is hard to verify in practice, but we will investigate the rate of convergence in section 5.

4. Discretization. In this section, we outline the discretization of the inverse problem. We apply a continuous high-order finite element method for the spatial discretization of the state and an explicit time-stepping scheme, see section 4.1 and 4.2. This approach is commonly used in seismic applications, cf. [12, 42]. A high-order discontinuous Galerkin method is described in [49]. Furthermore, we use different spatial meshes for the state and the material parameters. This is motivated by the fact that the information on the material properties is limited, 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 [4] or prior knowledge to address the varying amount of information in the data in different regions of the domain. Using different grids for the state and the parameters requires to interpolate the parameter values onto the finer state mesh in every iteration.

4.1. Spatial discretization of the state space. We consider two- or three-dimensional shape-regular meshes consisting of quadrilateral or hexahedral cells K

that cover the computational domain Ω. $\mathbb{T}_h = \{K\}$ denotes the finite element mesh and h the discretization parameter. Let Q_s denote the space of polynomials of degree s in each variable x_i, $i = 1, \dots, d$, on the reference cell $K_{\text{ref}} = [-1, 1]^d$. We use the Lagrange polynomials of degree s with the collocation points of the Gauss-Lobatto-Legendre (GLL) quadrature rule [26] as basis of Q_s. This yields a nodal basis for the numerical representation of the elements of Q_s. 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. 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

$$\varphi^\iota : K_{\text{ref}} \to \mathbb{R}, \quad \varphi^\iota(x) := \prod_{i=1}^{d} l_{\iota_i}(x_i). \tag{4.1}$$

By definition, the Lagrange polynomials vanish at all but one of the collocation points. Integrals over the reference cell are approximated by the GLL quadrature rule, which is exact for integrands in Q_{2s-1}. Now, we introduce the finite element subspaces $V_h^s \subset V$ by

$$V_h^s = \left\{ v^h \in C(\bar{\Omega})^d \;\middle|\; v_h|_K \in \hat{Q}_s(K)^d \;\; \forall\, K \in \mathbb{T}_h \right\}, \tag{4.2}$$

where $\hat{Q}_s$ is obtained by bi- or trilinear transformations of the nodal basis defined on the reference cell in every component. All numerical tests presented in this paper 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, [6, 13, 28, 42]. Therefore, we just summarize the outcome. By replacing V by V_h^s in (2.13), 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 ODEs which can be written in the following form:

$$\bar{\mathbf{M}}\mathbf{u}_{tt}(t) + \bar{\mathbf{K}}\mathbf{u}(t) = \bar{\mathbf{F}}(t), \tag{4.3}$$

Here, $\bar{\mathbf{M}} \in \mathbb{R}^{N\times N}$ denotes the mass matrix (modified 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 allows for 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 $2s - 1$.

4.2. Time discretization. Now, we turn to the temporal discretization of the state equation. Similar to [36], we apply an explicit Newmark time-stepping scheme to solve (4.3). Let

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

be a partition of the interval $\bar{I}$ with constant time increment $\Delta t = t_k - t_{k-1}$, $k = 1, \dots, n_t$. For the Newmark time-stepping scheme 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 fully discrete Newmark system is then given by the update formulas (cf. [6, 26]):

$$\begin{aligned}
\mathbf{u}_{k+1,2} &= -\mathbf{M}^{-1}\left(\mathbf{K}\mathbf{u}_{k+1,0} - \mathbf{F}_{k+1}\right), \\
\mathbf{u}_{k+1,0} &= \mathbf{u}_{k,0} + \Delta t\, \mathbf{u}_{k,1} + \frac{1}{2}\Delta t^2\, \mathbf{u}_{k,2}, \\
\mathbf{u}_{k+1,1} &= \mathbf{u}_{k,1} + \frac{1}{2}\Delta t\, \left(\mathbf{u}_{k,2} + \mathbf{u}_{k+1,2}\right).
\end{aligned} \tag{4.4}$$

This scheme is second-order accurate and conditionally stable, see [26], Chapter 9. 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 $\bar{\mathbf{K}}$.

The explicit Newmark time stepping scheme is widely used for the numerical simulation of seismic wave propagation [6, 13, 28]. 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 [49]. We also refer to [31], where the elastic wave equation is discretized by the Crank-Nicolson scheme. A severe drawback of explicit time-stepping schemes is the limitation of the step-size by the CFL condition. However, due to the diagonal mass matrix, parallelization can be carried out much easier as we do not have to solve a linear system in every time step.

4.3. Spatial discretization of the parameter space. As outlined above, we completely separate the discretization of state and parameter meshes. For the parameter mesh, we also use a continuous Galerkin finite element discretization and introduce the finite element subspace

$$M_h^s = \left\{ m^h \in C(\bar{\Omega})^n \;\middle|\; m_h|_K \in \hat{Q}_s(K)^n \;\; \forall\, K \in \mathbb{T}_h^M \right\}. \tag{4.5}$$

Here, $\mathbb{T}_h^M$ denotes the decomposition of $\bar{\Omega}$ for the parameter space. $\mathbb{T}_h^M$ will generally consists of larger cells than $\mathbb{T}_h$. In all numerical tests, we will use a polynomial degree of $s = 1$, i.e., bi- or trilinear elements. We recall the assumption $M \hookrightarrow\hookrightarrow L^\infty(\Omega)$ that we required to prove the existence of a solution to the inverse problem in the infinite-dimensional case. 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. This motivates that a discrete H^1-type norm suffices in our discrete setting. 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

$$\alpha_1 \|m_i\|^2_{L^2(\Omega)} + \alpha_2 \|\nabla m_i\|^2_{L^2(\Omega)^d}, \quad i = 1, \dots, n.$$

We use the same ratio α_1/α_2 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. It is important to note that the costs of computing $\mathbf{M}^{-1}\mathbf{v}$ for the preconditioner as well as the Riesz representation of the derivative with respect to the $\mathbf{M}$-inner product are negligible compared to solving a single wave equation. Furthermore, the numerical results neither show oscillating solutions nor undesirable artifacts in the reconstruction, which justifies our choice of discretization for the parameter space.

5. Numerical examples. In this section, we discuss some aspects of the parallel implementation and present numerical results for inverse problems in 2d and 3d. In order to prevent artificial reflections from the boundaries of the computational domain, we impose the following absorbing boundary conditions at all boundaries except the free surface, cf. [10],

$$(\Psi(m) : \varepsilon(u)) \cdot \vec{n} = v_p \rho \, (u_t \cdot \vec{n}) \, \vec{n} + v_s \rho \, (u_t - (u_t \cdot \vec{n}) \, \vec{n}) \, . \tag{5.1}$$

Here, $\vec{n}$ denotes the normal vector pointing outwards of the domain. Depending on the parameterization, v_p and v_s can, for instance, be computed using (2.3).

5.1. Implementation. The wave propagation code as well as the optimization routines are implemented in C++. Due to the similarities in the discretization, the implementation is inspired by the SPECFEM code [36]. There are, however, significant differences in the computation of the discrete gradient and Hessian-vector products. We make use of the Epetra data structures of the Trilinos library [17] and utilize the therein provided MPI-communication. 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.

Table 5.1 and Table 5.2 present statistics for strong and weak parallel scaling of a forward simulation of the elastic wave equation in 3d. Here, we consider the same problem setup as in section 5.3. The computations are carried out on a Cray XC30 supercomputer based on Intel® Xeon® E5 processors. Both tables indicate a good parallel performance.

TABLE 5.1
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.

#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	0.982	0.935	0.848	0.859	0.859	0.865

TABLE 5.2
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). "Scaling efficiency" of N cores is defined as the ratio of the total run-time on 8 cores and the total run-time on N cores.

#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	1.0	0.979	0.963	0.935

5.2. Joint inversion for both Lamé coefficients. 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 refer to Example 2.4 for a representation of the constraints.

For the test setup, we consider a time interval of 2.5s and a 2d domain of 4km $\times$ 4km with a single source in the center. 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 column of Figure 5.1(a)). Thereby, the Poisson's ratio varies from 0.15 to 0.31. Data is generated by a simulation with 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 reconstruction is shown in the second column of Figure 5.1. Here, we show v_p and v_s computed from the reconstructed λ and μ. The right column in Figure 5.1 compares shotgathers of all receivers for the true and reconstructed material. Here, the amplitude of the signal is visualized as a function of receiver's location (on the horizontal axis) and time (on the vertical axis). The order of the receivers' locations is clock-wise. The first arrival around 0.6s shows the P-wave and the second arrival at roughly 1s is the S-wave. The wavefronts arrive delayed or premature due to the heterogeneities of the material. 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 between 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. We terminate the Steihaug-CG algorithm if the relative residual is less than

$$\varepsilon_k = 0.01 \cdot \min\{1, \|\nabla j_\gamma(m^i)\|/\|\nabla j_\gamma(m^0)\|\},$$

where m^i denotes the i-th Newton iterate. Note that we do not impose a fixed limit on the number of CG iterations in this example. After 22 iterations the norm of the gradient has been reduced by more than 12 orders of magnitude and we observe a superlinear rate of convergence, see Figure 5.2(b). However, the number of CG iterations increases significantly for the last 7 iterations, see Figure 5.2(a). Therefore, we test different strategies regarding the maximum number of CG iterations and define the stopping tolerance as a reduction of the norm of the gradient by 6 orders of magnitude, which is sufficiently accurate for inverse problems. The computational effort is summarized in Table 5.3. With a maximum number of 20 CG iterations, 37 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. However, a superlinear rate of convergence cannot be observed for these two configurations, see Figure 5.2(b). Nevertheless, simulating the elastic wave equation dominates the computational costs, thus, the limit of 40 CG iterations gives the best result in terms of computing time.

In order to analyze the effect of constraints in more detail, we modify the problem formulation and restrict ν to values in $[0.225, 0.275]$. Hence, the true material is

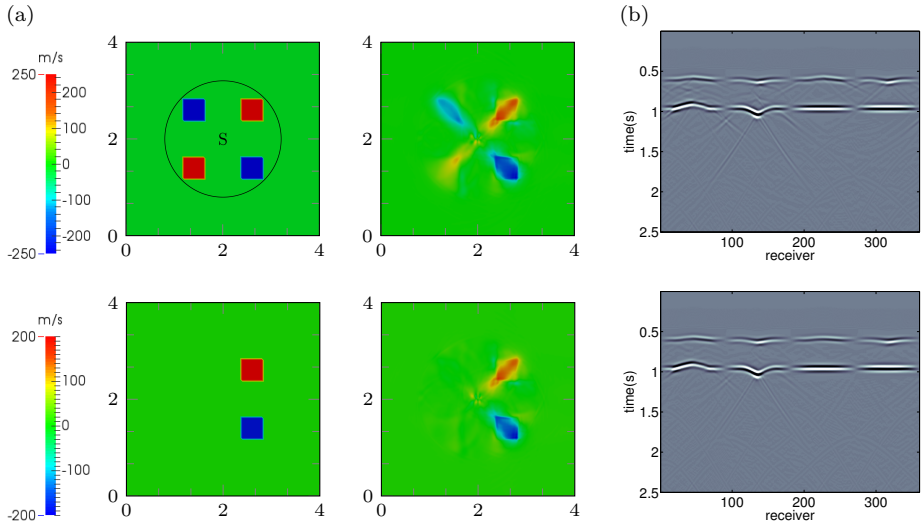

FIG. 5.1. *Joint inversion for both Lamé coefficients. (a) First column: true P-wave velocity (generated by modifying λ on the left half and μ on the right) and true S-wave velocity (changes only due to μ). Second column: reconstruction of P-wave velocity (top) and S-wave velocity (bottom). All images show the deviation from the reference material. (b) depicts shotgathers of the vertical component of all receivers for the true material (top row) and the reconstruction (bottom row).*

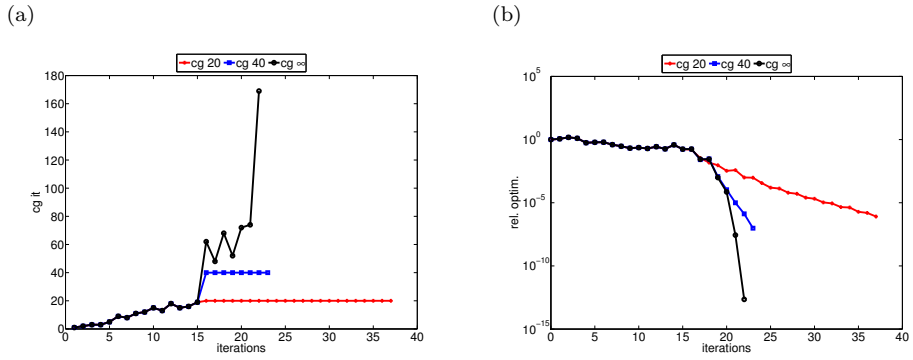

FIG. 5.2. *(a) shows the number of CG steps per iteration for different configurations of the Steihaug-CG solver. (b) depicts the relative norm of the gradient in every iteration. Superlinear convergence can be observed if we do not impose a limit on the number of CG iterations.*

TABLE 5.3

Comparison of the computational effort for different configurations of the Steihaug-CG method. Column 2-4 indicate the number of Newton iterations, the average number of CG iterations per Newton iteration and the total number of PDEs that have to be solved. While the number of Newton iterations decreases when more CG steps are allowed, the second configuration achieves the best results in terms of PDE simulations.

max cg	it Newton	avg. cg it	# PDEs
20	37	15.9	1370
40	23	20.4	1060
∞	21	25.4	1162

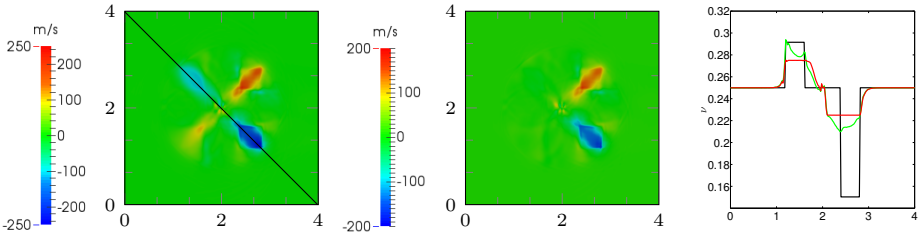

FIG. 5.3. *Joint inversion for both Lamé coefficients with unattainable true material model. The left and middle image depict the reconstructed P-wave (left) and S-wave velocity (middle). The image on the right indicates the Poisson's ratio along the diagonal line depicted in the left image. The plot shows the true material in black, the previous reconstruction in green and the reconstruction with "hard constraints" in red. Both, lower and upper bounds are active in parts of the domain.*

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 become active in the reconstruction and the result is slightly worse. This is shown in Figure 5.3. The final misfit is 0.0378 compared to 0.0329 in the first case. 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. The Poisson's ratio for the true material and both reconstructions is shown in the right image of Figure 5.3.

5.3. Borehole Tomography in 3D. In this example, we consider a 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. 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 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 2250m/s. The material model as well as the locations of sources and receivers are shown in Figure 5.4. 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 on the absolute value of λ, which gives $\lambda \in [3.375, 5.042] \cdot 10^9$.

For the spatial discretization of the elastic wave equation, we use 531,441 grid points and 4,000 time steps. The parameter mesh has approximately 68,921 degrees of freedom and is discretized by $41 \times 41 \times 41$ grid points. Figure 5.4 shows the reconstruction on the right-hand side. Only 6 Newton iterations with a maximum of 40 CG steps are required to solve the problem to a relative tolerance of 10^{-6}. Here, the misfit has been reduced by more than 99%. Table 5.4 shows the iteration tableau. Note that this inverse problem is considerably easier to solve than the previous example. 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.

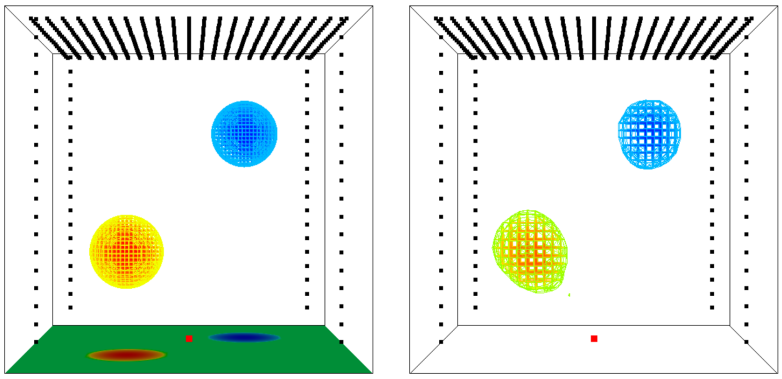

FIG. 5.4. *Borehole Tomography. The test setup is shown on the left with black dots representing the receivers' locations and the red dot indicating the position of the seismic source. Furthermore, the two perturbations of the true material are visualized and projected to the bottom for a better visibility. The right image shows the reconstruction which captures the two perturbations very well.*

TABLE 5.4
Iteration tableau for the borehole example. The second column shows the relative decrease of the objective function and the third column indicates the relative reduction of the optimality criterion. All iterates are feasible.

it	rel. obj.	rel. optim.	cg it
0	1.00	1.00	
1	3.95e-01	7.09e-01	9
2	4.54e-02	6.04e-02	19
3	2.25e-02	2.57e-02	40
4	2.21e-02	5.66e-04	34
5	2.21e-02	9.34e-06	40
6	2.21e-02	3.83e-07	40

5.4. 2d elastic inversion with the Marmousi model. The last example is based on the Marmousi data set provided by the Institut Français du Pétrole Énergies Nouvelles [47]. 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. 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 parameter 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 the top row of Figure 5.5. Here, we only consider the upper part of the domain with up to 1km depth, as the reconstruction becomes less accurate for deeper structures.

In our test setup, we place 191 seismic sources at 36m depth and use a Ricker wavelet with a dominant frequency of 5Hz as source time function. 384 receivers that are placed equidistantly on a horizontal line at 100m depth record the signal (with 1% Gaussian noise added). This setup mimics a marine seismic exploration with sources located in a water layer and geophones measuring data at the seafloor. Note, however, that we do not explicitly model the fluid layer. For the discretized problem we use a uniformly refined parameter mesh with 49,665 degrees of freedom. The state equation is discretized with 197,633 degrees of freedom and 6,000 time-steps.

The large number of seismic events makes it computationally very expensive to consider every seismic source independently. A promising approach is to exploit the linearity of the elastic wave operator with respect to the displacement field and to trigger the sources simultaneously by building the weighted sum of the individual right-hand sides [16]. In particular, we choose weights $w^k \in \mathbb{R}^{n_s}$, $k = 1, \dots, K$, and compute $\mathbf{u}^k = \mathbf{u}(m, w^k)$ by solving the following elastic wave equation for every k:

$$E(\mathbf{u}^k, m) = \sum_{i=1}^{n_s} w_i^k f_i, \quad \mathbf{u}^k(0) = 0, \quad \mathbf{u}_t^k(0) = 0. \tag{5.2}$$

The misfit term in the cost functional compares now the seismograms generated by $\mathbf{u}^k$ with the weighted sum of the observed data, i.e.,

$$J_{\text{fit}}\left(\mathbf{u}(m, w^k), \sum_{i=1}^{n_s} w_i^k u_i^\delta\right), \quad k = 1, \dots, K. \tag{5.3}$$

Hence, we only have to consider K (super)-sources instead of n_s. In this example, we choose $K = 8$ and weights w_i^k as i.i.d. samples of Rademacher's distribution, i.e., $w^k \in \{-1, 1\}^{n_s}$ with $P(w_i^k = 1) = P(w_i^k = -1) = 0.5$, as has been suggested in [2].

We start the inversion using a reference model that varies only in depth and uses the average of the true material model in the horizontal plane, see middle row of Figure 5.5. The reconstructed material is shown in the bottom row of Figure 5.5. Here, the misfit has been reduced by 91%. 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. Table 5.5 summarizes the optimization process for the different tolerances.

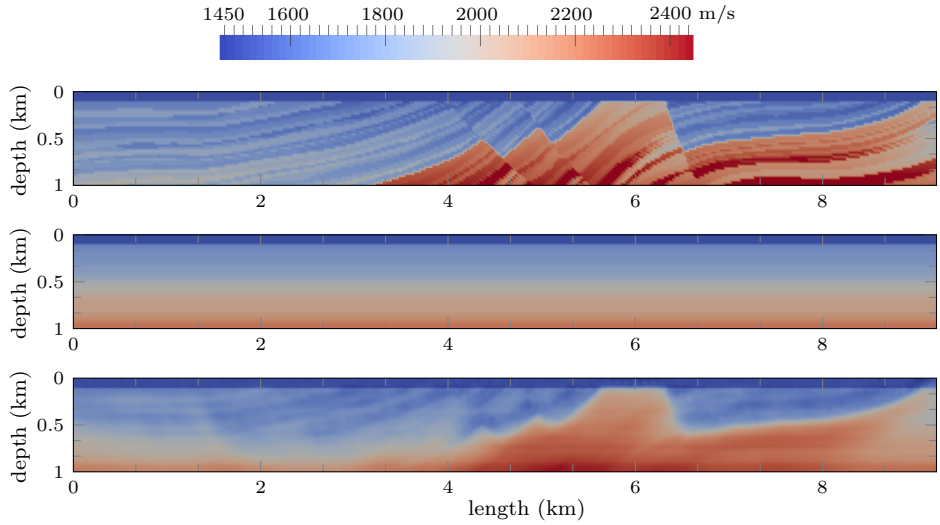

FIG. 5.5. *P-wave velocity profile of the Marmousi model. Top row: true material, middle row: initial model, bottom row: reconstruction.*

In a second step, we additionally enforce 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) to $v_p \geq 1450[m/s]$. This

TABLE 5.5
Computational effort to solve the Marmousi test problem with and without additional constraints. The first column indicates the relative tolerance for the stopping criterion.

tol	unconstrained			constrained		
	it Newton	avg. cg it	# PDEs	it Newton	avg. cg it	# PDEs
10^{-3}	24	17.0	7120	25	17.9	7800
10^{-6}	30	21.6	11104	33	23.3	13112

TABLE 5.6
Iteration tableau for the Marmousi test case with constraints for increasing penalty parameter γ.

γ	it Newton	avg. cg it	# PDEs
1	7	2.4	456
10	3	7.0	416
10^2	2	8.0	304
10^3	6	19.2	1992
10^4	15	39.9	9944

bound is only relevant near the surface as the velocities increase with depth and the true Marmousi model is feasible with respect to this bound. There is a slight improvement of the reconstruction in shallow depths up to 100m compared to the unconstrained case. Here, the maximum relative error is reduced from 0.22094 to 0.22068. The solutions coincide for deeper structures. The number of iterations for an increasing penalty parameter is indicated in Table 5.6. γ is updated as outlined at the end of section 3.3. Summing up, this test shows a nice application of additional 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 as Table 5.5 indicates.

6. Conclusion. In this paper, we proposed a semismooth Newton method for solving problems in full-waveform seismic tomography with additional constraints on the material parameters. We established the continuity and Fréchet differentiability of the parameter-to-state operator for the elastic wave equation in a suitable function space setting. Constraints on the parameters are treated by the Moreau-Yosida regularization, which results in a penalized problem, where the first order optimality conditions are given by a semismooth operator equation. The numerical results underline the applicability of the proposed semismooth Newton-PCG method to solve seismic inverse problems in 2d and 3d. A superlinear rate of convergence can be observed with a suitable stopping tolerance on the CG iterations.

Since the computational costs are vastly dominated by repeatedly solving the elastic wave equation, strategies to reduce the required number of simulations are highly desirable and an interesting field for future research. For instance, only a subset of the seismic sources can be used to compute Hessian-vector products. Alternatively, an initial phase using L-BFGS [5] approximations could be incorporated requiring only $2n_s$ simulations per iterations. Moreover, a preconditioner for the misfit term can be constructed using the L-BFGS approximation, cf. [10].

Acknowledgements. The authors would like to thank two anonymous referees and the Associate Editor Tim Kelley for their valuable suggestions that helped to improve the paper. The authors gratefully acknowledge support by the International

Graduate School of Science and Engineering at the Technische Universität München, Germany, as well as from the International Research Training Group IGDK1754, funded by the German Science Foundation (DFG). Some of the computations were carried out on a Linux cluster that was partially funded by the grant DFG INST 95/919-1 FUGG.

REFERENCES

[1] R.A. Adams and J.J.F. Fournier. *Sobolev Spaces*, volume 140 of *Pure and Applied Mathematics*. Elsevier/Academic Press, Amsterdam, 2003.
[2] A. Aravkin, M. Friedlander, F. Herrmann, and T. van Leeuwen. Robust inversion, dimensionality reduction, and randomized sampling. *Math. Program.*, 134:101–125, 2012.
[3] K.D. Blazek, C. Stolk, and W.W. Symes. A mathematical framework for inverse wave problems in heterogeneous media. *Inverse Problems*, 29(6):065001, 2013.
[4] C. Boehm and M. Ulbrich. A newton-cg method for full-waveform inversion in a coupled solid-fluid system. In Michael Bader, Hans-Joachim Bungartz, and Tobias Weinzierl, editors, *Advanced Computing*, volume 93 of *Lect. Notes Comput. Sci. Eng.*, pages 99–117. Springer, Berlin, 2013.
[5] R. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.*, 16(5):1190–1208, 1995.
[6] E. Chaljub, D. Komatitsch, J.-P. Vilotte, Y. Capdeville, B. Valette, and G. Festa. Spectral-element analysis in seismology. In Valerie Maupin Ru-Shan Wu and Renata Dmowska, editors, *Advances in Wave Propagation in Heterogenous Earth*, volume 48 of *Advances in Geophysics*, pages 365 – 419. Elsevier, 2007.
[7] S. Dain. Generalized korn's inequality and conformal killing vectors. *Calc. Var. Partial Differential Equations*, 25(4):535–540, 2006.
[8] J. Dennis, Jr. and J. Mor. Quasi-newton methods, motivation and theory. *SIAM Rev.*, 19(1):46–89, 1977.
[9] A.M. Dziewonski and D.L. Anderson. Preliminary reference earth model. *Physics of the earth and planetary interiors*, 25(4):297–356, 1981.
[10] I. Epanomeritakis, V. Akçelik, O. Ghattas, and J. Bielak. A newton-cg method for large-scale three-dimensional elastic full-waveform seismic inversion. *Inverse Problems*, 24(3):034015, 2008.
[11] L.C. Evans. *Partial Differential Equations*. Graduate Studies in Mathematics. AMS, 2010.
[12] A. Fichtner. Ses3d version 2.1: Programme description and mathematical background. Technical report, Ludwig-Maximilians-Universität München, 2009.
[13] A. Fichtner. *Full seismic waveform modelling and inversion*. Springer, Heidelberg, 2011.
[14] A. Fichtner, B.L.N. Kennett, H. Igel, and H.P. Bunge. Theoretical background for continental- and global-scale full-waveform inversion in the time–frequency domain. *Geophys. J. Int.*, 175(2):665–685, 2008.
[15] A. Fichtner, B.L.N. Kennett, H. Igel, and H.P. Bunge. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods. *Geophys. J. Int.*, 179(3):1703–1725, 2009.
[16] E. Haber, M. Chung, and F. Herrmann. An effective method for parameter estimation with pde constraints with multiple right-hand sides. *SIAM J. Optim.*, 22(3):739–757, 2012.
[17] M.A. Heroux, J.M. Willenbring, and R. Heaphy. Trilinos developers guide. *Sandia National Laboratories, SAND2003-1898*, 2003.
[18] M. Hintermüller and M. Hinze. Moreau–yosida regularization in state constrained elliptic control problems: Error estimates and parameter adjustment. *SIAM J. Numer. Anal.*, 47(3):1666–1683, 2009.
[19] M. Hintermüller, K. Ito, and K. Kunisch. The primal-dual active set strategy as a semismooth newton method. *SIAM J. Optim.*, 13(3):865–888, 2002.
[20] M. Hintermüller and K. Kunisch. Feasible and noninterior path-following in constrained minimization with low multiplier regularity. *SIAM J. Control Optim.*, 45(4):1198–1221, 2006.
[21] M. Hintermüller and K. Kunisch. Path-following methods for a class of constrained minimization problems in function space. *SIAM J. Optim.*, 17(1):159–187, 2006.
[22] M. Hintermüller and K. Kunisch. Pde-constrained optimization subject to pointwise constraints on the control, the state, and its derivative. *SIAM J. Optim.*, 20(3):1133–1156, 2010.
[23] 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. *SIAM J. Op-*

tim., 24(1):108–126, 2014.
[24] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE constraints*. Springer, Berlin, 2008.
[25] C. Horgan. Korn's inequalities and their applications in continuum mechanics. *SIAM Rev.*, 37(4):491–511, 1995.
[26] T.J.R. Hughes. *The finite element method*. Dover Publications, Mineola, NY, 2000.
[27] A. Kirsch and A. Rieder. On the linearization of operators related to the full waveform inversion in seismology. *Math. Methods Appl. Sci.*, 2013. first published online, doi: 10.1002/mma.3037.
[28] D. Komatitsch, C. Barnes, and J. Tromp. Wave propagation near a fluid-solid interface: A spectral-element approach. *Geophysics*, 65(2):623–631, 2000.
[29] M. Kristekova, J. Kristek, and P. Moczo. Time-frequency misfit and goodness-of-fit criteria for quantitative comparison of time signals. *Geophys. J. Int.*, 178(2):813–825, 2009.
[30] A. Kröner, K. Kunisch, and B. Vexler. Semismooth newton methods for optimal control of the wave equation with control constraints. *SIAM J. Control Optim.*, 49(2):830–858, 2011.
[31] Axel Kröner. Adaptive finite element methods for optimal control of second order hyperbolic equations. *Comput. Methods Appl. Math.*, 11(2):214–240, 2011.
[32] J.L. Lions and E. Magenes. *Non-homogeneous boundary value problems and applications*, volume 1. Springer, Berlin, 1972.
[33] R. Madariaga. Seismic source theory. In G. Schubert, editor, *Treatise on Geophysics*, pages 59 – 82. Elsevier, Amsterdam, 2007.
[34] L. Métivier, R. Brossier, J. Virieux, and S. Operto. Full waveform inversion and the truncated newton method. *SIAM J. Sci. Comput.*, 35(2):B401–B437, 2013.
[35] G. Nolet. *A breviary of seismic tomography*. Cambridge University Press, Cambridge, 2008.
[36] 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. *Geophys. J. Int.*, 186(2):721–739, 2011.
[37] R.G. Pratt, C. Shin, and G.J. Hick. Gauss-newton and full newton methods in frequency-space seismic waveform inversion. *Geophys. J. Int.*, 133(2):341–362, 1998.
[38] 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. *Geophys. J. Int.*, 2013.
[39] N. Rawlinson, S. Pozgay, and S. Fishwick. Seismic tomography: A window into deep earth. *Phys. Earth Planet. Int.*, 178(3-4):101 – 135, 2010.
[40] T. Steihaug. The conjugate gradient method and trust regions in large scale optimization. *SIAM J. Numer. Anal.*, 20(3):626–637, 1983.
[41] C. Tape, Q. Liu, A. Maggi, and J. Tromp. Seismic tomography of the southern california crust based on spectral-element and adjoint methods. *Geophys. J. Int.*, 180(1):433–462, 2010.
[42] J. Tromp, D. Komatitsch, and Q. Liu. Spectral-element and adjoint methods in seismology. *Commun. Comput. Phys.*, 3(1):1–32, 2008.
[43] J. Tromp, C. Tape, and Q. Liu. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophys. J. Int.*, 160(1):195–216, 2005.
[44] M. Ulbrich. Semismooth newton methods for operator equations in function spaces. *SIAM J. Optim.*, 13(3):805–841, 2002.
[45] M. Ulbrich. *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*. SIAM, Philadephia, PA, 2011.
[46] T. Van Leeuwen and W. A. Mulder. A correlation-based misfit criterion for wave-equation traveltime tomography. *Geophys. J. Int.*, 182(3):1383–1394, 2010.
[47] R. Versteeg. The marmousi experience: Velocity model determination on a synthetic complex data set. *The Leading Edge*, 13(9):927–936, 1994.
[48] J. Virieux and S. Operto. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26, 2009.
[49] L.C. Wilcox, G. Stadler, C. Burstedde, and O. Ghattas. A high-order discontinuous galerkin method for wave propagation through coupled elastic-acoustic media. *J. Comput. Phys.*, 229(24):9373 – 9396, 2010.