A Newton-CG Method for Full-Waveform Inversion in a Coupled Solid-Fluid System

Christian Boehm and Michael Ulbrich

Abstract We present a Newton-CG method for full-waveform seismic inversion. Our method comprises the adjoint-based computation of the gradient and Hessian-vector products of the reduced problem and a preconditioned conjugate gradient method to solve the Newton system in matrix-free fashion. A trust-region globalization strategy and a multi-frequency inversion approach are applied. The governing equations are given by a coupled system of the acoustic and the elastic wave equation for the numerical simulation of wave propagation in solid and fluid media. We show numerical results for the application of our method to marine geophysical exploration.

1 Introduction

Earthquakes excite seismic waves that propagate through the Earth and can be recorded as seismograms at remote receiver locations. Seismic tomography means to infer the Earth's structure based on these observations. An accurate knowledge of the Earth's interior does not only enhance scientific progress in explaining the geodynamics and subsurface processes but can also help to improve reliable Tsunami warning systems and support the search for natural resources. In this article we focus on the application of marine geophysical exploration, where seismic waves are emitted by a research vessel that is equipped with an air gun and cruises on the sea. This requires to model a medium consisting of a solid and a fluid layer as well as the interaction at the interface. The governing equations are given by a coupled system

Christian Boehm
Chair of Mathematical Optimization, Department of Mathematics, Technische Universität München, Boltzmannstr. 3, 85748 Garching b. München, Germany e-mail: boehm@ma.tum.de

Michael Ulbrich
Chair of Mathematical Optimization, Department of Mathematics, Technische Universität München, Boltzmannstr. 3, 85748 Garching b. München, Germany e-mail: mulbrich@ma.tum.de

of the elastic wave equation in the solid domain and the acoustic wave equation in the fluid domain.

Seismic tomography can be stated as an optimization problem with PDE constraints. Here, the spatially heterogeneous, unknown material parameters enter the PDE as coefficients. We refer to [24] for a general overview on seismic tomography. With the availability of high-performance computing clusters, 3D full-waveform inversion has become possible and iterative inversion methods have been applied to datasets on both regional and continental scale [10, 13, 14, 28]. Alternative approaches work in the frequency domain and involve the Helmholtz equation [31].

Here we present a Newton-type method for full-waveform inversion. We apply a trust-region globalization and iteratively solve the resulting subproblems by the Steihaug-CG method [26]. The gradient and Hesse-vector products of the reduced problem are efficiently computed using adjoint-based techniques [16]. A smooth cutoff function ensures that the parameters remain within reasonable bounds without explicitly imposing additional constraints. We invert sequentially for increasing source frequencies [7] and adaptively refine the parameter grid using goal-oriented error estimates [4]. The coupled system is spatially discretized by a high-order continuous Galerkin method and solved with an explicit Newmark time-stepping scheme [17, 22]. The parallel implementation works matrix-free and utilizes MPI communication to tackle large-scale seismic inverse problems.

This article is organized as follows. In section 2 we describe the governing equations and formulate the seismic inverse problem. In section 3 we present our optimization method and conclude with some remarks on the implementation and a numerical example in section 4.

2 The seismic inverse problem

2.1 Wave propagation at a solid-fluid interface

We consider a domain that consists of a solid and a fluid layer. We denote the solid and fluid regions by Ω_S and Ω_F and set $\Omega = \Omega_S \cup \Omega_F \subset \mathbb{R}^d$ with $d = 2,3$. We assume that Ω_S and Ω_F are bounded domains with a C^2-boundary and a smooth interface Γ_{int}. The remaining parts of the boundaries are denoted by $\Gamma_F = \partial\Omega_F \setminus \Gamma_{\text{int}}$ and $\Gamma_S = \partial\Omega_S \setminus \Gamma_{\text{int}}$. The time interval is denoted by $I := (0,T)$ with $T > 0$.

The propagation of waves in the solid medium is governed by the elastic wave equation and, respectively, by the acoustic wave equation in the fluid domain. At the interface continuity of traction and continuity of the normal velocity have to be ensured, cf. [17]. Figure 1 shows the geometry of the domain with both layers.

In the fluid domain, we consider an inviscid fluid medium with a homogeneous density $\rho_F > 0$. p denotes the velocity potential and $c \in \mathbb{R}$ is the speed of compressional waves. In the solid medium, we assume a heterogeneous and positive density $\rho_S(x)$ and a linear elastic rheology. Let u denote the displacement field,

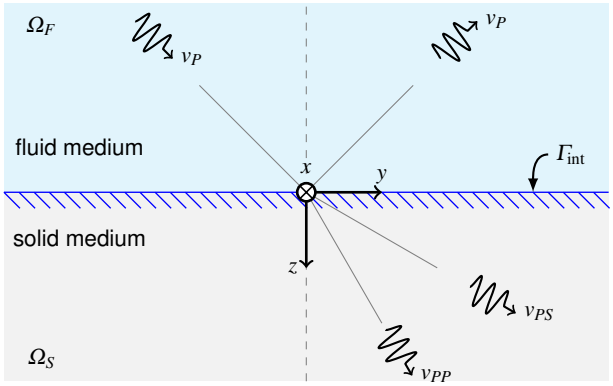

Fig. 1 Sketch of a solid-fluid interface. Parts of the incoming waves are reflected to the acoustic medium while other parts are transmitted to the elastic medium.

$\varepsilon(u) = \frac{1}{2}\left(\nabla u + \nabla u^T\right)$ the stress tensor of u and $\Psi = (\Psi_{ijkl})$ is a fourth order elastic tensor. To cover the general case, we place seismic source functions f_F and f_S into both media. The coupled system is then given by:

$$\begin{cases} c^{-2}p_{tt}(x,t) - \Delta p(x,t) = f_F(x,t) & (x,t) \in \Omega_F \times I, \\ p(x,0) = 0, \quad p_t(x,0) = 0 & x \in \Omega_F, \\ p(x,t) = 0 & (x,t) \in \Gamma_F \times I, \\ \\ \rho_S(x)u_{tt}(x,t) - \nabla \cdot (\Psi(x) : \varepsilon(u)(x,t)) = f_S(x,t) & (x,t) \in \Omega_S \times I, \\ u(x,0) = 0, \quad u_t(x,0) = 0 & x \in \Omega_S, \\ (\Psi(x) : \varepsilon(u)(x,t)) \cdot \mathbf{n}_S = 0 & (x,t) \in \Gamma_S \times I, \\ \\ (\Psi(x) : \varepsilon(u)(x,t)) \cdot \mathbf{n}_S = -\rho_F p_t(x,t)\, \mathbf{n}_F & (x,t) \in \Gamma_{\text{int}} \times I, \\ -\mathbf{n}_F \cdot \nabla p(x,t) = \mathbf{n}_S \cdot u_t(x,t) & (x,t) \in \Gamma_{\text{int}} \times I. \end{cases} \tag{1}$$

Remark 1. The tensor Ψ of elastic moduli has the symmetry properties $\Psi_{ijkl} = \Psi_{jikl} = \Psi_{klij}$ and has to be uniformly coercive. In this general form we can handle anisotropic material. However, in a perfectly elastic, isotropic medium the tensor simplifies to

$$\Psi_{ijkl} = \lambda \delta_{ij}\delta_{kl} + \mu(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{kj}) \tag{2}$$

with the Lamé parameters $\lambda(x)$ and $\mu(x)$. In this case we have the relation

$$v_p = \sqrt{\frac{\lambda + 2\mu}{\rho_S}}, \quad v_s = \sqrt{\frac{\mu}{\rho_S}}, \tag{3}$$

where v_p and v_s denote the speed of compressional and shear waves [21]. In what follows, we will consider an isotropic material but stick to the tensor notation and use the parameterization described in (2).

2.2 Seismic tomography as PDE-constrained optimization problem

Depending on the parameterization of the governing equations the unknown material parameters can be the Lamé parameters, the velocity of compressional and/or shear waves or further elasticity parameters. In either case, the unknown parameter field is heterogeneous in space and does not depend on the time. Due to the interdependencies of ρ_S and Ψ, we keep the density fixed and invert for Ψ only. Furthermore, we assume to have a reference model $\overline{\Psi} \in L^\infty(\Omega_S)^{d^4}$ available and choose a parameterization

$$\Psi(m) = \overline{\Psi} + \Phi(m). \tag{4}$$

Here, $m \in M \subset\subset L^\infty(\Omega_S)^n$ parameterizes deviations from the reference model $\overline{\Psi}$. n denotes the dimension of the unknown parameters and M is a Hilbert space. The reference model should already capture major discontinuities of the material parameters and $\Phi(m)$ characterizes smooth variations from $\overline{\Psi}$ with a linear function $\Phi : \mathbb{R}^n \to \mathbb{R}^{d^4}$.

Remark 2. In all applications that we are concerned with, the existence of a suitable reference model based on a-priori knowledge is guaranteed. Often those reference models only vary in depth and assume a homogeneous parameter field in the horizontal plane. For instance, the most popular choice for global seismic tomography is the Preliminary Reference Earth Model (PREM) [9].

In the next step, we turn to the weak formulation of (1). To shorten the notation, we define the following forms:

$$\begin{aligned}
a_S(m)(v,w) &= (\Psi(m) : \varepsilon(v), \varepsilon(w))_{L^2(\Omega_S)^{d\times d}} && \forall\, v,w \in V_S,\ \forall\, m \in M,\\
a_F(v,w) &= \rho_F\,(\nabla v, \nabla w)_{L^2(\Omega_F)^d} && \forall\, v,w \in V_F,\\
a_{\mathrm{int}}(v,w) &= \rho_F \int_{\Gamma_{\mathrm{int}}} w\, v \cdot \mathbf{n}_s\, dS && \forall\, v \in L^2(\Omega_S)^d,\ \forall\, w \in L^2(\Omega_F).
\end{aligned}$$

Here, $V_S = H^1(\Omega_S)^d$ and $V_F = H_0^1(\Gamma_F, \Omega_F)$ denotes the space of all H^1-functions that vanish on Γ_F. The normal vector $\mathbf{n}_S$ in the definition of a_{int} is pointing outwards of the solid and into the fluid domain. The variational form of (1) is given by:
$\forall\, v \in V_F, \forall\, w \in V_S$ and a.a. $t \in I$:

$$\begin{aligned}
\rho_F c^{-2} \langle p_{tt}(t), v\rangle_{V_F^*,V_F} + a_F(p(t),v) + a_{\mathrm{int}}(u_t(t),v) &= \rho_F \langle f_F(t), v\rangle_{V_F^*,V_F},\\
\langle \rho_S u_{tt}(t), w\rangle_{V_S^*,V_S} + a_S(m)(u(t),w) - a_{\mathrm{int}}(w, p_t(t)) &= \langle f_S(t), w\rangle_{V_S^*,V_S}.
\end{aligned} \tag{5}$$

Here, we multiplied the acoustic wave equation by ρ_F. In [5] we discuss existence and uniqueness of solutions in appropriate function spaces. This analysis, however, is beyond the scope of this article, so we just state the following concepts for fixed $m \in M$:

- For the acoustic wave equation with homogeneous Dirichlet boundary conditions and a source $f_F \in L^2(I, V_F^*)$ there exists a unique very weak solution $p \in P \overset{\mathrm{def}}{=} C(\bar{I}; L^2(\Omega_F)) \cap C^1(\bar{I}; V_F^*)$.

- For the elastic wave equation with homogeneous Neumann boundary conditions and a source $f_S \in L^2(I, V_S^*)$ there exists a unique very weak solution $u \in U \overset{\text{def}}{=} C(\bar{I}; L^2(\Omega_S)^d) \cap C^1(\bar{I}; V_S^*)$.

Differentiability of u with respect to m can be established by exploiting higher time regularity of the seismic source function.

Remark 3. Note that a higher time regularity is a valid assumptions for the seismic source function as wavelets are commonly used. Since we can admit V_F^* or V_S^* in space, Dirac measures require only a slight smoothing.

We set $Y = U \times P$ and let $y = (u, p)$ denote the state that consists of the displacement field in the solid domain and the velocity potential in the fluid domain. With $f = (f_F, f_S)^T$ the weak form of the coupled system (1) can be written as:

$$E(y, m) = f \quad :\Leftrightarrow (y, m) \text{ satisfies (5) a.e. in } I. \tag{6}$$

Additionally, the initial conditions $y(0) = 0$ and $y_t(0) = 0$ have to be satisfied.

In the application of marine geophysical exploration data is taken from several seismic events, which means that we have so solve a wave equation with a different right-hand-side for every event independently. We denote the number of events by n_s and the corresponding state variables by y_i, $i = 1, \ldots, n_s$. For every i we are given a seismic source f_i with support only in the fluid domain. Furthermore, we assume to have measurements y_i^δ of the displacement field in form of seismograms on $\Omega_i \times I$ with $\Omega_i \subset \Omega_S$. Typically, only sparse observations of the displacement field are available which contributes to the ill-posedness of the problem and necessitates proper regularization. The seismic inverse problem is given as follows

$$\min_{\mathbf{y} \in \mathbf{Y},\, m \in M} J(\mathbf{y}, m) \quad \text{s.t. } E(y_i, m) = \begin{pmatrix} f_i \\ 0 \end{pmatrix} \quad 1 \le i \le n_s, \quad \begin{pmatrix} \mathbf{y}(0) \\ \mathbf{y}_t(0) \end{pmatrix} = 0. \tag{7}$$

Here, $\mathbf{y} = (y_i)_{1 \le i \le n_s}$ denotes a vector of states for different seismic events. Note that the state y_i only enters into the i-th component of E, while the parameters m are the same for all components. We consider cost functions $J : \mathbf{Y} \times M \to \mathbb{R}$ of the form:

$$J(\mathbf{y}, m) = \sum_{i=1}^{n_s} J_{\text{fit}}(y_i, y_i^\delta) + \alpha J_{\text{reg}}(m), \tag{8}$$

with a misfit term J_{fit}, a regularization term J_{reg} and the regularization parameter $\alpha > 0$. In what follows, we will always assume that J_{fit} is twice continuously F-differentiable with respect to y_i, J_{reg} is twice continuously F-differentiable with respect to m, convex and lower semicontinuous. A possible choice for J would be, for instance,

$$J_{\text{fit}}(y_i, y_i^\delta) = \frac{1}{2} \left\| y_i - y_i^\delta \right\|^2_{L^2(\Omega_i \times I)}, \quad J_{\text{reg}}(m) = \frac{1}{2} \|m\|_M^2. \tag{9}$$

Furthermore, we introduce the reduced problem

$$\min_{m\in M} \; j(m) \stackrel{\text{def}}{=} J(\mathbf{y}(m),m), \tag{10}$$

where $\mathbf{y}(m)$ denotes solution of the state equation for given $m \in M$.

A major extension to the problem formulation (7) can be made, if additional box constraints on the material parameters are imposed. This can also be found in [5].

Remark 4. In this paper we consider the inverse problem to determine the spatially heterogeneous material parameters and assume that the location and time evolution of the seismic source f is known. However, our optimization method that is outlined in section 3 could easily be extended to invert for the seismic source as well (either simultaneously or in an alternating manner).

2.3 Adjoint equation

As pointed out above, the state variables for different seismic events completely decouple in (7). Hence, we restrict the analysis of the adjoint equation to a single seismic event and drop the index i. Derivatives of the multi-source problem can easily be computed by adding up the individual contributions. In order to derive the adjoint equation, we introduce the Lagrange function $L(y,m,z) : Y \times M \times Y \to \mathbb{R}$ with the adjoint variable $z = (z^1, z^2)$:

$$\begin{aligned} L(y,m,z) \stackrel{\text{def}}{=} J(y,m) &+ \int_0^T \langle \rho_S u_{tt}(t), z^1(t)\rangle_{V_S^*,V_S}\, dt + \int_0^T a_S(m)(u(t),z^1(t))\, dt \\ &+ \int_0^T \rho_F \langle p_{tt}(t), z^2(t)\rangle_{V_F^*,V_F}\, dt + \int_0^T a_F(p(t),z^2(t))\, dt \\ &- \int_0^T a_{\text{int}}(z^1, p_t(t))\, dt + \int_0^T a_{\text{int}}(u_t(t), z^2(t))\, dt \\ &- \int_0^T \langle f_S(t), z^1(t)\rangle_{V_S^*,V_S}\, dt - \int_0^T \rho_F \langle f_F(t), z^2(t)\rangle_{V_F^*,V_F}\, dt \\ &- (\rho_S u(0), z_t^1(0))_{L^2(\Omega_S)^d} + (\rho_S u_t(0), z^1(0))_{L^2(\Omega_S)^d} \\ &- \rho_F (p(0), z_t^2(0))_{L^2(\Omega_F)} + \rho_F (p_t(0), z^2(0))_{L^2(\Omega_F)}. \end{aligned} \tag{11}$$

The adjoint equation is given by

$$L_y(y,m,z) = 0. \tag{12}$$

By considering the variational form of (12) and carefully integrating by parts with respect to time, we obtain the adjoint equation for given m and $y(m)$ as

$$E^{\text{ad}}(z,m) = -J_y(y(m),m), \quad z(T) = 0, \quad z_t(T) = 0. \tag{13}$$

With the assumption of sufficient regularity, the adjoint equation E^{ad} in strong form can be interpreted as a coupled system like (1) backwards in time with final time in-

stead of initial conditions and a different right-hand-side. Further details of deriving the continuous adjoint can be found in [5].

In the next step, we turn to the adjoint-based representation of first and second derivatives. Using the parameterization defined in (4) we obtain

$$\Psi'(m) = \Phi(.) \in \mathscr{L}\left(M, L^\infty(\Omega_S)^{d^4}\right) \quad \forall\, m \in M.$$

We introduce the form $D : Y \times Y \to \mathscr{L}(M,\mathbb{R})$ defined by

$$D(v,w)(m) = \int_0^T \int_{\Omega_S} \left(\varepsilon(v^1)(x,t) \otimes \varepsilon(w^1)(x,t)\right) :: \left(\Phi(m)(x)\right) dx\,dt \quad \forall\, m \in M. \quad (14)$$

Here, $v = (v^1, v^2), w = (w^1, w^2) \in U \times P$ and we use the notations $(a \otimes b)_{ijkl} = a_{ij} b_{kl}$ and $A :: B = \sum_{ijkl} A_{ijkl} B_{ijkl}$. For given $m \in M$ we denote the corresponding state by $y(m)$ and the adjoint state by $z(m)$. The first derivatives can then be expressed as

$$j'(m) = J'_{\text{reg}}(m) + D(y(m), z(m)). \quad (15)$$

Following the derivation in [16] the second derivative of the reduced cost functional can be derived in a similar fashion. This requires the second derivatives of the Lagrange function with respect to the state and the parameters. Since the state equation is linear in both, y and m, a couple of terms vanish. While the operator $j''(m)$ would be prohibitively expensive to compute, operator-vector products $j''(m)s$ for a given perturbation $s \in M$ can be computed at the cost of two additional simulations by performing the following steps:

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

$$E(\delta_s y, m) = -E_m(y(m), m)s, \quad \delta_s y(0) = 0, \quad \delta_s y_t(0) = 0.$$

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

$$E^{\text{ad}}(\delta_s z, m) = -J_{yy}\delta_s y - E_m(z(m), m)s, \quad \delta_s z(T) = 0, \quad \delta_s z_t(T) = 0.$$

Then $j''(m)s$ is given by:

$$j''(m)s = J''_{\text{reg}}(m)s + D(\delta_s y, z(m)) + D(y(m), \delta_s z).$$

With the adjoint-based representation of the first derivatives and operator-vector products representing the second derivatives applied to a search direction, we have everything at hand to apply a Newton-type optimization method.

3 Optimization method

3.1 Trust-region Newton-CG

In order to solve the seismic inverse problem (10), the algorithm successively computes approximate solutions to the trust-region subproblem

$$\min_{s\in M} \quad q_i(s) \stackrel{\text{def}}{=} j(m^i) + \left\langle j'(m^i), s\right\rangle_{M^*,M} + \tfrac{1}{2}\left\langle j''(m^i)s, s\right\rangle_{M^*,M} \qquad \text{s.t. } \|s\|_M \le \Delta_i. \tag{16}$$

Here m^i denotes the current iterate and Δ_i the trust-region radius in iteration i. q is a quadratic model function that approximates $j(m+s)$. The first derivatives $j'(m^i)$ and operator-vector products $j''(m)s$ are computed using adjoint-based techniques as outlined in the previous section. Note that we use the exact second derivatives in (16), however, approximations of the Hessian, e.g. by a Quasi-Newton method, would also be possible. We compute an approximate solution to (16) by the Steihaug conjugate gradient method [26]. Here, the inner product induced by the norm of M is used as preconditioner. The CG iterations are early terminated, if negative curvature is encountered or the trust region radius is exceeded by the current iterate. Additionally, we solve the Newton system inexactly and stop with a relative tolerance of 0.01. A crucial ingredient of the trust-region method is the ratio of actual and predicted reduction,

$$\frac{j(m^i) - j(m^i + s^i)}{q_i(0) - q_i(s^i)},$$

based on which is determined, whether the step is accepted and how the trust-region radius is updated. For further details, we refer to [5], see also [29] for trust-region methods in function space and [8] for a comprehensive study.

The problem formulation (7) does not include any constraints on the material parameters. However, from a physical, as well as from a theoretical point of view, there exist bounds on m, for instance, nonnegative wave velocities or coercivity of the elastic tensor. In [5] we extend the problem formulation by imposing additional pointwise box constraints on the material parameters. The trust-region algorithm presented above is then applied to a series of problems with an additional penalty term in the objective function that is weighted with a monotonically increasing penalty parameter. For the simplified unconstrained case that is discussed in this article, we apply a smooth cutoff function to ensure that the parameters remain within a certain range. Let $m^{l-}, m^l, m^u, m^{u+} \in \mathbb{R}^n$, $m^{l-} < m^l < m^u < m^{u+}$, denote pointwise bounds on the material parameters. We then choose a smooth function $\varphi : M \to M$ and replace m by $\varphi(m)$ in the weak form (5). Hereby, φ has to satisfy the following properties: φ is monotonically increasing, $\varphi \equiv \mathrm{id}$ on $[m^l, m^u]$ and $\varphi(M) = [\varphi(m^{l-}), \varphi(m^{u+})]$. The bounds should be chosen such that $[\varphi(m^{l-}), \varphi(m^{u+})]$ covers the domain of all physically reasonable models and the optimal parameter model should be within $[m^l, m^u]$.

Thus, it is guaranteed that only reasonable values for m enter into the weak form. Moreover, parameter models that exceed the bounds yield the same misfit while the regularization term J_{reg} penalizes large deviations. Hence, the optimization algorithm will eventually favor parameter models within the bounds. Figure 2 shows an example of this cutoff function φ.

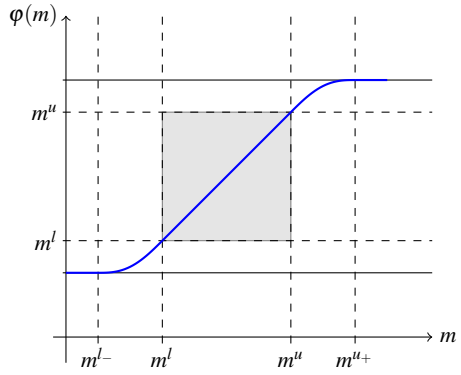

Fig. 2 Smooth cutoff function to ensure that the parameters remain within physically reasonable bounds. A fourth order polynomial ensures the smooth transition in $[m^{l-}, m^l]$ and $[m^u, m^{u+}]$.

3.2 Discretization

We follow a discretize-then-optimize strategy. This involves the temporal and spatial discretization of the state and adjoint equation as well as the spatial discretization of the parameters. We use different meshes for the state and the parameter space. 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. We apply a continuous high-order finite element method for the spatial discretization of the state space and an explicit Newmark time-stepping. We use fourth-order Lagrange polynomials and approximate the integrals by the Legendre-Gauss-Lobatto quadrature rule which results in a diagonal mass matrix. This approach is commonly used in seismic applications, cf. [11, 27, 32]. Moreover, by following an update scheme for the Newmark time-stepping as outlined in [17], we can carry out the time-stepping without solving a linear system.

The seismic inverse problem is computationally very expensive. The costs are vastly dominated by solving the discrete version of the coupled system (1) for several right-hand-sides. $2n_s$ simulations (one forward, one adjoint per seismic event) are required to evaluate the cost function and to compute the first derivatives for a given parameter model m. For a general J, every inner CG iteration requires ad-

ditional $2n_s$ simulations to compute the second derivatives applied to a search direction. However, if J_{fit} is quadratic as in (9) this can be reduced to 2 simulations, because the perturbed forward and the perturbed adjoint wavefield can be computed simultaneously, if the right-hand-sides for all seismic events are added together. This reduces the computational costs of a Newton-step in comparison to a limited memory BFGS approximation considerably. On the other hand, there exist more sophisticated misfit measures that are better suited to capture characteristic differences in the seismograms but are not quadratic in the state variables [30]. In this context, randomized data reduction techniques are an interesting field of research [18, 1].

3.3 Multi-frequency approach

A reconstruction of the material properties with a high resolution requires high frequency information in the observed data. However, the high-frequency data is more prone to errors induced by noisy measurements. Moreover, since only sparse observations are available, there usually exist several models that explain the data equally well. This necessitates the regularization term in the objective function. Additionally, we follow a regularization-by-discretization strategy and combine a multi-frequency inversion approach with an adaptive grid refinement based on goal-oriented error estimates. The multi-frequency approach (sometimes also called multi-scale) means to sequentially invert for increasing source frequencies [7, 12], see also [10]. A bandpass filter can be used to down-sample the observed measurements to lower frequencies. The adaptively refined grid allows to reduce the number of optimization variables without a loss of resolution in the reconstruction. The algorithm works as follows:

0. Choose an initial parameter mesh.

For a sequence of increasing source frequencies $\omega_1 \leq \omega_2 \leq \ldots$

1. Choose the state mesh based on the dominant frequency ω_i of the seismic source and the wave velocities.
2. Solve the discretized problem and obtain a stationary point.
3. Adaptively refine the parameter mesh using goal-oriented adaptivity.

The paradigm for goal-oriented a-posteriori error estimation and dual weighted residuals was established in [4]. We briefly outline the key idea. Let $m^* \in M$ be a stationary point of the Lagrange function L for the continuous problem and $m_h \in M_h \subset M$ a stationary point of L on the discretized subspace M_h. Then the error in the cost function can be represented by

$$j(m^*) - j(m_h) = \frac{1}{2}\langle L_m(y(m_h), m_h, z(m_h)), m^* - v_h\rangle_{M^*,M} + R, \tag{17}$$

with a cubic remainder term R and an arbitrary $v \in M_h$. In order to compute error estimates, the derivative of the Lagrange function with respect to m has to be com-

puted which requires a state and an adjoint equation solve. A remaining difficulty is to estimate the difference $m^* - v_h$, since the continuous solution m^* is unknown. Here, we use higher-order local interpolation in a post-processing step. We refer to [19, 2, 20] for goal-oriented error estimation in general and to [5] for a detailed description of its application in the context of seismic tomography.

4 Numerical Example

4.1 Implementation

Efficient inversion methods rely on a scalable code for the simulation of the elastic-acoustic wave equation. We have implemented the wave propagation code in C++, following a similar approach like SPECFEM3D [22]. 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, our parallel implementation allows to solve a single event on multiple cores using a spatial partitioning of the computational domain and communication with MPI. The algorithm works matrix-free and does not require to solve a linear system during the simulation. We use the Epetra package of Trilinos [15] to handle the distributed data structures. The adaptive refinement of the parameter mesh is implemented using the deal.ii library [3]. The parameter mesh is then interpolated onto the state mesh before every simulation.

4.2 Acoustic-elastic data set

This example shows the application of our method to a synthetic data set in marine geophysical exploration. The geometry and problem setup (Figure 3) is inspired by the Valhall oil field in the North Sea. Prior work on this field can be found in [6, 25, 23]. The geometry is given by a rectangular domain of 8km $\times$ 8km $\times$ 4km.

There is a thin water layer (400m) on top of the solid domain. We use 4 seismic sources that are triggered in the fluid region at 200m water depth. The source time function for all sources is a Ricker wavelet with dominant frequency of 1.25Hz. There are 441 seismic receivers buried into the seafloor at 50m depth that form a dense array of 16m^2 in the center of the domain. In the solid domain, we assume a constant density of 2300 kg/m^3 and a constant Poisson's ratio of 0.25. Thus, we have the relation $v_p = \sqrt{3}v_s$ for the velocity of compressional and shear waves and only one parameter field to invert for. In the fluid domain, we set $\rho_F = 1000$ kg/m^3 and $c = 1500$ m/s. Since the computational domain is artificially truncated, we have to impose absorbing boundary conditions to prevent reflections from non-physical

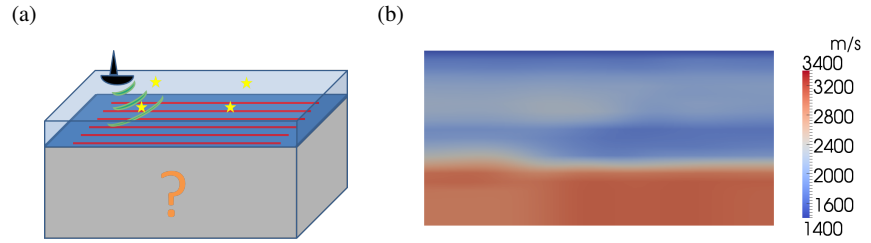

Fig. 3 (a) Data acquisition in marine geophysical exploration. A research vessel equipped with an air gun cruises over the field of interest and emits pressure waves into the sea (green). An array of geophones is placed at the sea floor (red lines) to record seismic waves that are reflected from the subsurface. Based on these measurements, the material structure in the solid domain (gray) shall be reconstructed. (b) shows the P-wave velocity of the target model for a vertical cross section through the domain.

boundaries. Here, we follow ideas from [17] and apply dampers that relate the traction to the velocity.

The synthetic target model has P-wave velocities that range from 1400m/s to 3400m/s. The synthesized data is generated by running a simulation with the target model and adding 0.5% Gaussian noise to the seismograms. The reference model varies only vertically and for every fixed depth we use the average value of the target model in the horizontal plane as the reference value.

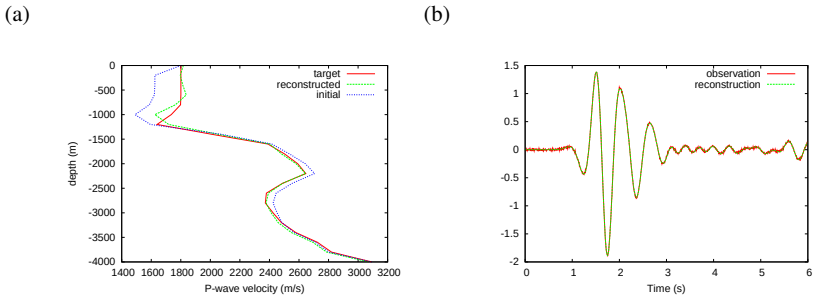

Fig. 4 (a) P-wave velocity for a vertical cross section through the domain at the center of the x-y-plane. (b) Comparison of observed and simulated seismograms (vertical component) at one receiver locations.

We solve the seismic inverse problem by sequentially inverting for source frequencies of 0.75Hz and 1.25Hz. On the coarse level, we use a stronger regularization parameter and the optimization algorithm requires 5 Newton iterations to reduce the norm of the gradient by 7 orders of magnitude. On the fine level we obtain a param-

eter mesh with approx. 35k degrees of freedom. The state mesh has approx. 300k spatial grid points and 1000 time steps.

Figure 4(a) compares the initial, target and reconstructed parameter model at a vertical cross section through the domain. Figure 4(b) shows an example seismogram at a receiver location. There is a good match between observed and reconstructed data. Figure 5 compares horizontal snapshots of the P-wave velocity for the reconstructed and the target model. The inversion output looks reasonable, especially near the surface. As expected the reconstruction becomes less accurate at greater depths.

Table 1 summarizes the optimization process for the fine level of discretization. After 12 iterations the norm of the gradient has been reduced by more than 6 orders of magnitude. From iteration 6 on, the inexact Newton step is always accepted. All computations were carried out on a Linux cluster using 32 processors.

A more sophisticated strategy for steering the interplay of the adaptive grid refinement for multiple frequencies, the update strategy for the regularization parameter and the convergence tolerance is under investigation.

Table 1 Summary of the optimization process.

iteration	$j(m^i)$	$\frac{\left\|\nabla j(m^i)\right\|}{\left\|\nabla j(m^0)\right\|}$	# iter. CG
0	56,1328	1.00	
1	22.0205	5.47e-1	8
2	16.9423	4.52e-1	15
3	14.2688	2.85e-1	22
4	14.0102	2.83e-1	101
5	13.5853	2.77e-1	71
6	8.5083	2.11e-1	93
7	2.0364	9.92e-2	55
8	1.0781	2.39e-2	55
9	0.9318	6.44e-3	65
10	0.9200	7.22e-4	67
11	0.9197	2.03e-5	74
12	0.9197	1.98e-7	74

5 Conclusion

We have presented a Newton-CG method for full-waveform seismic tomography in a coupled solid-fluid system. The optimization framework consists of a trust-region globalization, a smooth cutoff function and a multi-frequency inversion approach. Derivatives are efficiently computed using adjoint techniques. Numerical results show the applicability of our method to large-scale problems in marine geo-

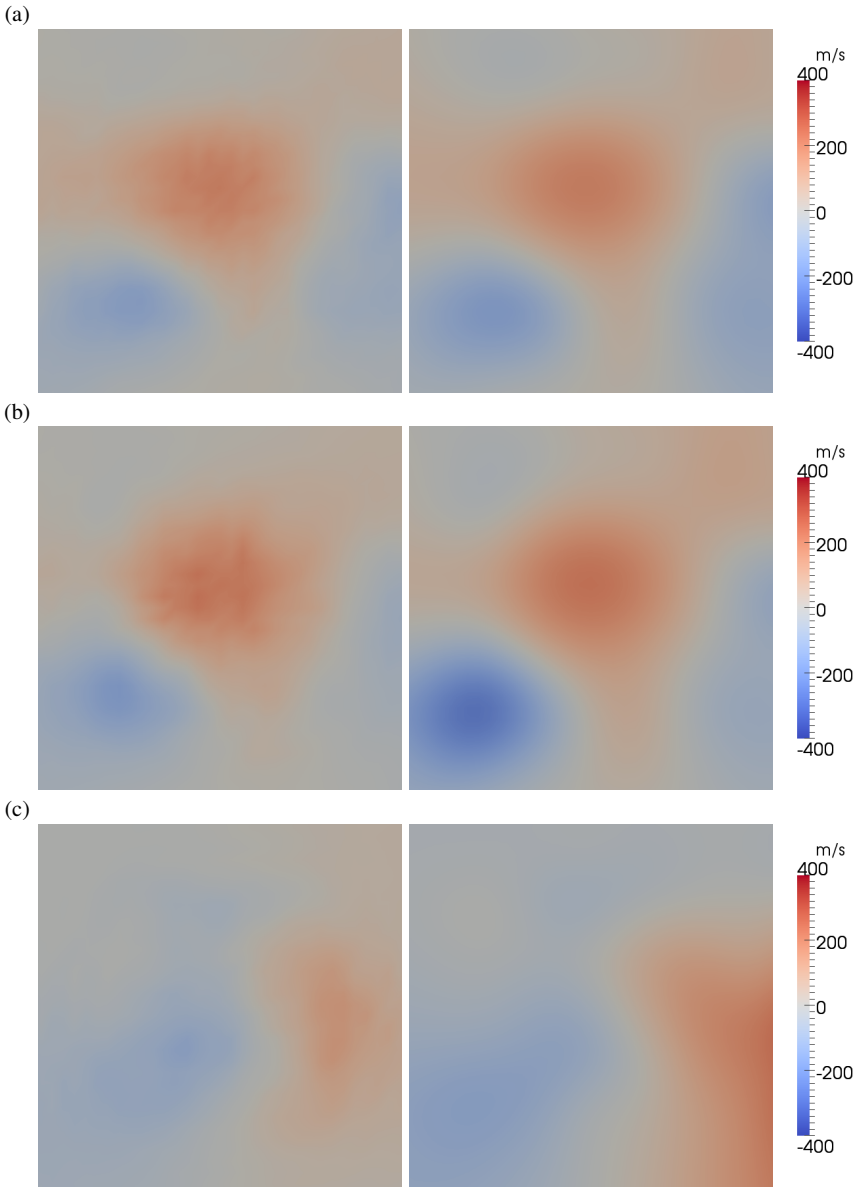

Fig. 5 Horizontal snapshots of the reconstructed (left) and target (right) v_P at (a) 250m, (b) 750m and (c) 2000m depth. The images show the deviation from the reference model that is homogeneous for every fixed depth.

physical exploration. A couple of major extensions in both, theoretical and practical aspects, will appear in [5].

Acknowledgements The project is supported by the Munich Centre of Advanced Computing, Technische Universität München, Germany. The computations were carried out on a Linux cluster that is partially funded by the grant DFG INST 95/919-1 FUGG. In addition, the first author gratefully acknowledges support by the International Graduate School of Science and Engineering at the Technische Universität München, Germany. The authors would like to thank Heiner Igel and Alan Schiemenz for helpful discussions and valuable advices.

References

1. Aravkin, A., Friedlander, M., Herrmann, F., van Leeuwen, T.: Robust inversion, dimensionality reduction, and randomized sampling. Mathematical Programming **134**, 101–125 (2012)
2. Bangerth, W., Geiger, M., Rannacher, R.: Adaptive galerkin finite element methods for the wave equation. Computational Methods in Applied Mathematics **10**(1), 3–48 (2010)
3. Bangerth, W., Kanschat, G.: `deal.II` Differential Equations Analysis Library, Technical Reference. URL http://www.dealii.org. `http://www.dealii.org`
4. Becker, R., Rannacher, R.: An optimal control approach to a posteriori error estimation in finite element methods. Acta Numerica 2001 **10**(1), 1–102 (2001)
5. Boehm, C., Ulbrich, M.: An adaptive semismooth newton-cg method for constrained parameter identification in seismic tomography (2012). In preparation
6. Brossier, R., Operto, S., Virieux, J.: Seismic imaging of complex onshore structures by 2d elastic frequency-domain full-waveform inversion. Geophysics **74**(6), WCC105–WCC118 (2009)
7. Bunks, C., Saleck, F., Zaleski, S., Chavent, G.: Multiscale seismic waveform inversion. Geophysics **60**(5), 1457–1473 (1995)
8. Conn, A.R., Gould, N.I.M., Toint, P.L.: Trust Region Methods. SIAM (2000)
9. Dziewonski, A., Anderson, D.: Preliminary reference earth model. Physics of the earth and planetary interiors **25**(4), 297–356 (1981)
10. Epanomeritakis, I., Akçelik, V., Ghattas, O., Bielak, J.: A newton-cg method for large-scale three-dimensional elastic full-waveform seismic inversion. Inverse Problems **24**(3), 034,015 (2008)
11. Fichtner, A.: Ses3d version 2.1: Programme description and mathematical background. Tech. rep. (2009)
12. Fichtner, A., Bleidinhaus, F., Capdeville, Y.: Full seismic waveform modelling and inversion. Springer Verlag (2010)
13. Fichtner, A., Kennett, B., Igel, H., Bunge, H.: Theoretical background for continental-and global-scale full-waveform inversion in the time–frequency domain. Geophysical Journal International **175**(2), 665–685 (2008)
14. Fichtner, A., Kennett, B., Igel, H., Bunge, H.: Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods. Geophysical Journal International **179**(3), 1703–1725 (2009)
15. Heroux, M., Willenbring, J., Heaphy, R.: Trilinos developers guide. Sandia National Laboratories, SAND2003-1898 (2003)
16. Hinze, M., Pinnau, R., Ulbrich, M., Ulbrich, S.: Optimization with PDE constraints. Springer Verlag (2008)
17. Komatitsch, D., Barnes, C., Tromp, J.: Wave propagation near a fluid-solid interface: A spectral-element approach. Geophysics **65**(2), 623–631 (2000)

18. Krebs, J.R., Anderson, J.E., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A., Lacasse, M.D.: Fast full-wavefield seismic inversion using encoded sources. Geophysics **74**(6), WCC177–WCC188 (2009)
19. Kröner, A.: Adaptive finite element methods for optimal control of second order hyperbolic equations. Computational Methods in Applied Mathematics **11**(2), 214–240 (2011)
20. Meidner, D., Vexler, B.: Adaptive space-time finite element methods for parabolic optimization problems. SIAM Journal on Control and Optimization **46**(1), 116–142 (2007)
21. Nolet, G.: A breviary of seismic tomography. Cambridge University Press (2008)
22. Peter, D., Komatitsch, D., Luo, Y., Martin, R., Le Goff, N., Casarotti, E., Le Loher, P., Magnoni, F., Liu, Q., Blitz, C., Nissen-Meyer, T., Basini, P., Tromp, J.: Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes. Geophysical Journal International **186**(2), 721–739 (2011)
23. Prieux, V., Brossier, R., Gholami, Y., Operto, S., Virieux, J., Barkved, O.I., Kommedal, J.H.: On the footprint of anisotropy on isotropic full waveform inversion: the valhall case study. Geophysical Journal International **187**(3), 1495–1515 (2011)
24. Rawlinson, N., Pozgay, S., Fishwick, S.: Seismic tomography: A window into deep earth. Physics of the Earth and Planetary Interiors **178**(3-4), 101 – 135 (2010)
25. Sirgue, L., Barkved, O., Dellinger, J., Etgen, J., Albertin, U., Kommedal, J.: Full waveform inversion: the next leap forward in imaging at valhall. First Break **28** (2010)
26. Steihaug, T.: The conjugate gradient method and trust regions in large scale optimization. SIAM Journal on Numerical Analysis **20**(3), 626–637 (1983)
27. Tromp, J., Komattisch, D., Liu, Q.: Spectral-element and adjoint methods in seismology. Communications in Computational Physics **3**(1), 1–32 (2008)
28. Tromp, J., Tape, C., Liu, Q.: Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. Geophysical Journal International **160**(1), 195–216 (2005)
29. Ulbrich, M.: Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces. SIAM (2011)
30. Van Leeuwen, T., Mulder, W.A.: A correlation-based misfit criterion for wave-equation traveltime tomography. Geophysical Journal International **182**(3), 1383–1394 (2010)
31. Virieux, J., Operto, S.: An overview of full-waveform inversion in exploration geophysics. Geophysics **74**(6), WCC1–WCC26 (2009)
32. Wilcox, L.C., Stadler, G., Burstedde, C., Ghattas, O.: A high-order discontinuous galerkin method for wave propagation through coupled elastic-acoustic media. Journal of Computational Physics **229**(24), 9373 – 9396 (2010)