

*A numerical study of multi-parameter
full waveform inversion with iterative
regularization using multi-frequency
vibroseis data*

**Jia Shi, Elena Beretta, Maarten V. de
Hoop, Elisa Francini & Sergio Vessella**

Computational Geosciences
Modeling, Simulation and Data Analysis

ISSN 1420-0597
Volume 24
Number 1

Comput Geosci (2020) 24:89–107
DOI 10.1007/s10596-019-09897-6

Your article is protected by copyright and all rights are held exclusively by Springer Nature Switzerland AG. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



A numerical study of multi-parameter full waveform inversion with iterative regularization using multi-frequency vibroseis data

Jia Shi¹ · Elena Beretta² · Maarten V. de Hoop³ · Elisa Francini⁴ · Sergio Vessella⁴

Received: 6 October 2018 / Accepted: 3 September 2019 / Published online: 23 November 2019
 © Springer Nature Switzerland AG 2019

Abstract

We study the inverse boundary value problem for time-harmonic elastic waves, for the recovery of P - and S -wave speeds from vibroseis data or the Neumann-to-Dirichlet map. Our study is based on our recent result pertaining to the uniqueness and a conditional Lipschitz stability estimate for parametrizations on unstructured tetrahedral meshes of this inverse boundary value problem. With the conditional Lipschitz stability estimate, we design a procedure for full waveform inversion (FWI) with iterative regularization. The iterative regularization is implemented by projecting gradients, after scaling, onto subspaces associated with the mentioned parametrizations yielding Lipschitz stability. The procedure is illustrated in computational experiments using the continuous Galerkin finite element method of recovering the rough shapes and wave speeds of geological bodies from simple starting models, near and far from the boundary, that is, the free surface.

Keywords Full waveform inversion · Finite element method · Stability and convergence

1 Introduction

Seismic data from land acquisition can mathematically be represented by the Neumann-to-Dirichlet map [3] since the normal traction (Neumann boundary value) is applied to the boundary and the displacement (Dirichlet boundary value) is measured. This map forms the data for the inverse boundary value problem for time-harmonic elastic waves corresponding with vibroseis data. We present FWI with iterative regularization, which aids in avoiding over-parameterization of the original problem. This approach is based on our recent result [6] pertaining to uniqueness and a conditional Lipschitz stability estimate, that is, well-posedness for parametrizations on unstructured

tetrahedral meshes of this nonlinear inverse boundary value problem. The unstructured tetrahedral meshes form domain partitions, while the wave speeds on these are chosen to be piecewise constant. The conditional Lipschitz stability linking the model differences and the data residuals provides theoretical control of the reconstruction on the stable subspace.

Following the mentioned parameterizations and a natural tetrahedral mesh refinement procedure (*while elements may simply, randomly change as well*), we form a hierarchy of subspaces generating sequences of increasingly accurate approximations of “true” models. One may view these subspaces as *setup for compression* of true models [2]. The piecewise constant parameter representations are reminiscent of expansions of parameters in terms of Haar wavelets [34]. We note that the stability constant will grow exponentially in the number of elements of the mesh. This reflects the ill-posedness of the problem. The study of the interplay between growing stability constants and compression rates in multi-level iterative reconstruction was studied in generality by De Hoop et al. [18].

The contribution of this paper is a convergence study of the above mentioned, multi-level approach to FWI by computational experiments. We formulate the inverse boundary value problem in terms of a constrained minimization of a suitable misfit functional justifiably derived from a Hilbert-Schmidt norm. The adjoint state method yields an

✉ Jia Shi
 jia.shi@rice.edu

¹ Department of Earth, Environmental and Planetary Sciences, Rice University, Houston, TX, USA

² Dipartimento di Matematica “F. Brioschi”, Politecnico di Milano, Milan, Italy

³ Simons Chair in Computational and Applied Mathematics and Earth Science, Rice University, Houston, TX, USA

⁴ Dipartimento di Matematica e Informatica “U. Dini”, Università di Firenze, Milan, Italy

adjoint boundary value problem. Elastic-wave boundary value problems can be naturally discretized by the finite element method. To mimic the target half-space problem with a Neumann boundary condition on the top, we introduce a constraint for the damping function in constructing the perfectly matched layers (PMLs) [4]. The local matrices are constructed elementwise after discretizing the relevant weak formulation. We construct the global matrix pattern on unstructured meshes and then apply a parallel strategy for generating a distributed matrix. We refer to [35] for the use of the finite element method and associated parametrizations in FWI. Unstructured tetrahedral meshes aligned with finite element discretizations [53] were considered using purely imaginary frequencies.

Following earlier works [25, 29, 58] for the acoustic wave equation, the reconstruction of subsurface elastic parameters using iterative minimization was originally introduced in the time domain [37, 60]. The time-harmonic or frequency-domain formulation of the seismic inverse problem was later considered for the acoustic case [45] and then for the elastic case [43]. Multi-scale FWI [11, 55] was designed to mitigate the occurrence of local minima without proof. However, many case studies [1, 7, 24, 27] have confirmed its computational efficiency. Here, frequency progression comes into play. *In our formulation, the optimal frequency for the next level minimizes the Lipschitz stability constant for the next level yielding the largest possible radius of convergence to the next approximation.* We illustrate this in our computational experiments.

As far as optimization is concerned, the application of Newton-type methods in FWI dates back to the 1980s [37] and 1990s [44]. Typically, one adopts a matrix-free approach through the adjoint state method. We mention a few results describing various strategies to mitigate the computational cost. The diagonal of the Gauss-Newton Hessian [15] was used to scale the gradients of the misfit function for P - and S -wave speeds, presumably to speed up convergence. Other standard methods, such as a limited-memory variant of the quasi-Newton BFGS method known as the L-BFGS algorithm [10] and the truncated Newton method [36], have also been adopted in FWI. We limit ourselves to scaling the gradient with the diagonal of the Gauss-Newton Hessian primarily due to the computational cost in three-dimensional multi-parameter reconstruction while the efficiency of a Gauss-Newton method remains questionable anyway.

We give a brief overview of recent work concerning multi-parameter inversion. Multi-parameter FWI was applied to marine and land data examples [41] and studied for multi-component ocean-bottom-cable data over the Valhall field, where P - and S -wave speeds were jointly updated [46]. Here, hydrophone data were utilized to update the long

and intermediate wavelengths of the S -wave speeds from the amplitude-versus-offset variations of the PP reflections. To reduce the computational costs, a multi-scale FWI scheme, which promotes the construction of full waveform tomographic models that describe the geological structures at multiple scales, was used in [24, 64]. Multi-parameter FWI was also applied to a large wide-azimuth long-offset land data set in Oman [56], where large wave speed variations occur between shale and carbonate layers. The data contained low frequencies down to 1.5 Hz with long-offsets and wide azimuths. *We feel that this case study justifies the use of low-frequency data in our computational experiments.* Time-lapse FWI [47] as a monitoring tool for directly resolving changes was applied to elastic parameter models to detect a carbon dioxide gas cloud. Incorporation of surface topography is important for successful elastic FWI of land seismic data [39]. We note that surface topography can naturally be incorporated into unstructured tetrahedral meshes. Multi-component three-dimensional elastic FWI with both surface and body waves has been recently applied to detect near-surface anomalies [8]. *In our formulation, surface and body waves co-exist.* For the recovery of a high-wave-speed variations, total variation regularization was used for blocky updates [19]. Such a regularization is inherent in our approach. An iterative solver of the Helmholtz problem was implemented using a complex-shifted incomplete LU-based preconditioner [40] and applied to elastic FWI recently.

As conventional elastic multi-parameter FWI [8, 24, 40, 41, 46, 56, 64] is commonly initiated with some smooth tomographic model, we initiate our iterations with a very coarse mesh with piecewise constant parameters. At low frequencies, this yields a relatively large radius of convergence to a proper coarse approximation. We note that *the computational and parameter meshes typically are not the same* in our approach.

The outline of the paper is as follows. In Section 2, we introduce the direct problem for modeling land vibroseis data and the corresponding inverse boundary value problem. In Section 3, we present the adjoint state equation for the inverse boundary value problem and an idealized example to verify the theory. In Section 4, we describe our multi-level, multi-frequency, and multi-parameter iterative scheme and illustrate its properties with a computational experiment recovering geobodies from simple initial models. In Section 5, we show two computational experiments where the true models are piecewise smooth with high contrasts and do not belong to the hierarchy of stable subspaces, to illustrate the recovery of best approximation in general applications. In Section 6, we discuss our approach and the reasoning behind it. To ensure reproducibility of our experiments, we present our use of the continuous Galerkin formula for Neumann boundary value problems with PMLs

in Appendix A, and the first- and second-order adjoint state method for inverse boundary value problems in Appendices B and C, respectively.

2 Direct and inverse problem

We consider seismic land acquisition where the forward modeling can be viewed as solving an elastic boundary value problem. Vibroseis data (omitting correlation in time) are modeled by the Neumann-to-Dirichlet map: the boundary values are given by the normal traction underneath the base plate of the vibroseis and are zero (free surface) elsewhere. The applied signal is essentially time-harmonic (suppressing the sweep) [3, (2.52)–(2.53)]. The particle velocities—from which the displacements can be obtained—are measured by the geophones.

Time-harmonic elastic waves are described by the operator,

$$P_{il} = -\rho(x)\delta_{il}\omega^2 + A_{il}, \quad A_{il} = -\partial_{x_j}c_{ijkl}(x)\partial_{x_k}, \quad x \in X,$$

where $i, j, k, l \in \{1, 2, 3\}$, $X \subset \mathbb{R}^3$ and X is the open bounded domain of interest and ω denotes the frequency. The relevant boundary value problem is given by

$$P_{il}u_l = 0, \quad (1)$$

$$(c_{inkl}\partial_{x_k}u_l)|_{\Sigma} = g, \quad (2)$$

where u is the displacement vector, $\Sigma \subseteq \partial X$ signifies the part of the surface on which the acquisition geometry is defined, and g denotes the time-harmonic boundary normal traction, or the Neumann boundary condition. The other computational boundaries that are not the surface, that is, $\partial X/\Sigma$, are treated as the PMLs. Details are discussed in Appendix A. Let m represent the model coefficients, $m = (c, \rho)$. Proceeding as in the analysis [6], we find a suitable range of frequencies, that is away from the eigenfrequency of the problem (1)–(2), such that the problem has a solution for any model m satisfying suitable prior assumptions. The displacement fields are measured at the surface, which are also the Dirichlet data. Thus, we define the so-called local Neumann-to-Dirichlet map

$$\Lambda_m^{\Sigma} : g \mapsto u|_{\Sigma},$$

The vibroseis data probe the Neumann-to-Dirichlet map via applying the boundary normal tractions at Σ and collecting the displacement information at Σ . The properties of the data operator Λ_m^{Σ} depend on the model m and the acquisition set Σ . The forward map is given by

$$F : m \rightarrow \Lambda_m^{\Sigma}. \quad (3)$$

Here, we assume that the parameters are real-valued and known in a neighborhood of Σ and otherwise piecewise constant on a tetrahedral partition. In the case of isotropic

media, $c_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})$. If the parameters are piecewise constant on a domain partition, $\bar{X} = \bigcup_{j=1}^N D_j$, where D_j , $j = 1, 2, \dots, N$ are connected and pairwise nonoverlapping open subdomains, here, tetrahedra, we arrive at the parametrizations,

$$\begin{aligned} \lambda &= \sum_{j=1}^N \lambda_j \chi_{D_j}(x), \quad \mu = \sum_{j=1}^N \mu_j \chi_{D_j}(x), \\ \rho &= \sum_{j=1}^N \rho_j \chi_{D_j}(x), \end{aligned} \quad (4)$$

where χ indicates the characteristic function. If the partition is known with reasonable assumptions [6, Assumptions 2.4–2.6], we can show by choosing suitable normal traction functions g that F is injective and that F^{-1} is Lipschitz continuous [6, Theorem 2.7], that is, there exists a constant C such that

$$\|m_1 - m_2\|_{L^2} \leq C \|\Lambda_{m_1}^{\Sigma} - \Lambda_{m_2}^{\Sigma}\|_{*}, \quad (5)$$

where m_1 and m_2 are two different real-valued coefficients, that is, the collections of $\{\lambda_j\}_{j=1}^N$, $\{\mu_j\}_{j=1}^N$ and $\{\rho_j\}_{j=1}^N$ in Eq. 4; $*$ denotes the operator norm. The constant C grows essentially exponentially with the number of subdomains. This number is directly related to spatial scale. The idea is to pair scale and frequency through the stability constant, which controls the radius of convergence within a subspace associated with this scale. We do not need to know the domain partition if it consists of subanalytic sets; at the same time, the domain partition can be viewed as yielding an approximation and its precise subdomains are not so relevant.

Frequency progression is carried out as a multi-level nonlinear projected steepest descent iteration, reminiscent of a multi-grid approach, which was introduced and analyzed in [18]. We emphasize that the (scale-dependent) meshes defining our domain partitions are chosen independently from the (frequency-dependent) computational meshes as illustrated in Fig. 1. The tetrahedral domain partitions are generated using Tetgen [54]. The linear system (1) and (2) from a realistically sized problem can be solved by a massively parallel algorithm [62] with randomized numerical linear algebra [33]. To solve a large-scale problem with above 10 million model parameters, we may need to utilize an iterative method [32] with highly parallel matrix-vector products [51] to construct a numerical solution at the target frequency.

3 Adjoint state method for vibroseis data

In this section, we discuss the adjoint problem as a boundary value problem and formulate the adjoint state method for vibroseis data. We construct a constrained

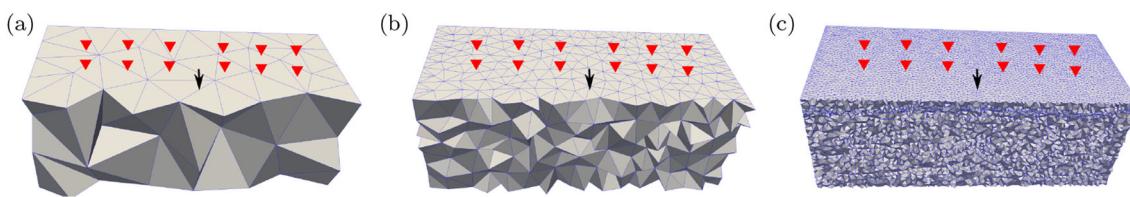


Fig. 1 Illustrations of different meshes. The arrow and red triangles represent the source and receivers. (a) A coarse parameter representation. (b) A finer parameter representation. (c) The computational mesh

optimization problem with the data residual norm as our misfit functional. We derive the gradient in an abstract setting for general parametrization and then consider the isotropic case for piecewise constant parameters on an unstructured tetrahedral mesh. For the application of Newton's method and a broader understanding, we give the first- and second-order adjoint state method for the Neumann-to-Dirichlet map as the data in Appendices B and C.

3.1 Misfit functional of the inverse boundary value problem

To implement a reconstruction procedure for the solution of the inverse problem, we can reformulate the problem as a constrained optimization problem. More precisely, we consider

$$\tilde{\Psi} = \frac{1}{2} \|\Lambda_m^\Sigma - \Lambda_{m^*}^\Sigma\|_*^2,$$

where m varies in the class of piecewise constant coefficients and m^* denotes the true model. From the theory developed by [6], one should use the data operator norm $*.$ Assuming our models are known in the subdomain of the partition containing Σ on its boundary, $\Lambda_m^\Sigma - \Lambda_{m^*}^\Sigma$ will be a Hilbert-Schmidt (HS) operator. We then introduce the misfit functional using the HS norm [17],

$$\begin{aligned} \tilde{\Psi} &= \frac{1}{2} \|\Lambda_m^\Sigma - \Lambda_{m^*}^\Sigma\|_{HS}^2 \\ &= \frac{1}{2} \sum_{j=1}^{\infty} \|(\Lambda_m^\Sigma - \Lambda_{m^*}^\Sigma) \psi_j\|_{H^{1/2}(\Sigma)}^2, \end{aligned} \quad (6)$$

where $\{\psi_j\}_{j=1}^\infty$ is an orthonormal basis of the space of the boundary sources and $\{\Lambda_{m^*}^\Sigma \psi_j\}_{j=1}^\infty$ are measured. In practice, we replace $\{\psi_j\}_{j=1}^\infty$ by the finite set $\{\hat{g}_s\}_{s=1}^{N_s}$ and omit $(1 - \Delta)^{1/2}$ in our computational experiments while dealing with $H^{1/2}$ norm. This operator $(1 - \Delta)^{1/2}$ can also be implemented as a filter [57].

To estimate such an operator norm that is controlled by the HS norm, the linear combination of sources and receivers needs to be enough to probe the data operator.

Within one Love wavelength, a few sources and receivers are needed. Hence, choosing suitable boundary sources $\{g^s\}_{s=1}^{N_s}$ with N_s sufficiently large, the misfit functional

$$\Psi^{HS} = \frac{1}{2} \sum_{s=1}^{N_s} \int_{\Sigma} \chi_{\Sigma} (R u^s - \Lambda_{m^*}^\Sigma g^s) \cdot (R u^s - \Lambda_{m^*}^\Sigma g^s) \, dx, \quad (7)$$

gives a good approximation of $\tilde{\Psi}$. We note that $\{g^s\}_{s=1}^{N_s}$ is selected as a set of compact support point sources and naturally approximates an orthonormal basis. The Neumann-to-Dirichlet map generates measurements $\Lambda_m^\Sigma g^s = R u^s$, for $s = 1, 2, \dots, N_s$, where s is the source index, N_s is the total number of sources, $\Lambda_{m^*}^\Sigma g^s$ represents the data generated from the true model, R restricts u^s for each boundary source to the surface, and χ_{Σ} represents a smooth cutoff function over Σ . In practice, we can use fewer sources and receivers for reconstruction at low frequency.

3.2 Adjoint state equation and gradient for the inverse boundary value problem

The adjoint state method was introduced in optimal control theory [31]. This method [13] was designed to efficiently calculate the gradient of a functional without computing Fréchet derivatives of the forward operator F in Eq. 3. The standard formulation uses the elastic wave equation and point-source data [59, 60].

Our adjoint state equation can be viewed as an extension of the classic adjoint state method [42]. To deal with the boundary data, we note that the adjoint sources are essentially boundary values instead of body forces, which is often overlooked. Since our model contains sharp jumps, the use of the weak formula is necessary. It is also beneficial for describing major geological discontinuities. The adjoint equation for inverse boundary value problem should be

$$-\omega^2 \rho \gamma_l \delta_{il} - \partial_{x_j} c_{ijkl} \partial_{x_k} \gamma_l = 0, \quad (8)$$

with the adjoint boundary value,

$$(c_{ijkl} \partial_{x_k} \gamma_l) \nu_j|_{\partial X} = -\chi_{\Sigma} R(\tilde{u}_i - u_i^*), \quad (9)$$

where γ denotes the adjoint wave field; $-\chi_{\Sigma} R(\tilde{u}_i - u_i^*)$ denotes the adjoint source; \tilde{u} is the solution of the forward

problem (1) and (2). Since the objective function $\Psi^{HS}(\tilde{u})$ depends on the model m , we then let

$$\mathcal{E}(m) = \Psi^{HS}(\tilde{u}).$$

Combining contributions from all the available sources, we obtain,

$$D_m \mathcal{E}[m] \delta m = \sum_s \left\{ \int_X -\omega^2 \tilde{u}_i^s \tilde{\gamma}_i^s \frac{\partial \rho}{\partial m} \delta m \, dx \right. \\ \left. + \int_X (\partial_{x_j} \tilde{u}_i^s) (\partial_{x_k} \tilde{\gamma}_i^s) \frac{\partial c_{ijkl}}{\partial m} \delta m \, dx \right\} = (\nabla \mathcal{E}, \delta m), \quad (10)$$

where $\tilde{\gamma}$ denotes the solution of the adjoint problem (8) and (9); $D_m \mathcal{E}[m]$ denotes the derivative; and $\nabla \mathcal{E}[m]$ denotes the gradient. In this work, we will update V_P and V_S alternatively from the updated Lamé parameters. More details of the derivation can be found in Appendix B. It has been pointed out that the difference between the derivative $D_m \mathcal{E}[m]$ and gradient $\nabla \mathcal{E}[m]$ is sometimes overlooked and the corresponding scaling of the gradient is essential for the convergence of the gradient-based optimization approaches [30]. The scaling factors rely on the choice of inner products. The discretized inner product, $(m_1, m_2) = m_1^T \mathbf{W} m_2$, contains a symmetric positive definite weight \mathbf{W} . We note that the weight \mathbf{W} can naturally be constructed via the Galerkin approximation from the predefined inner product.

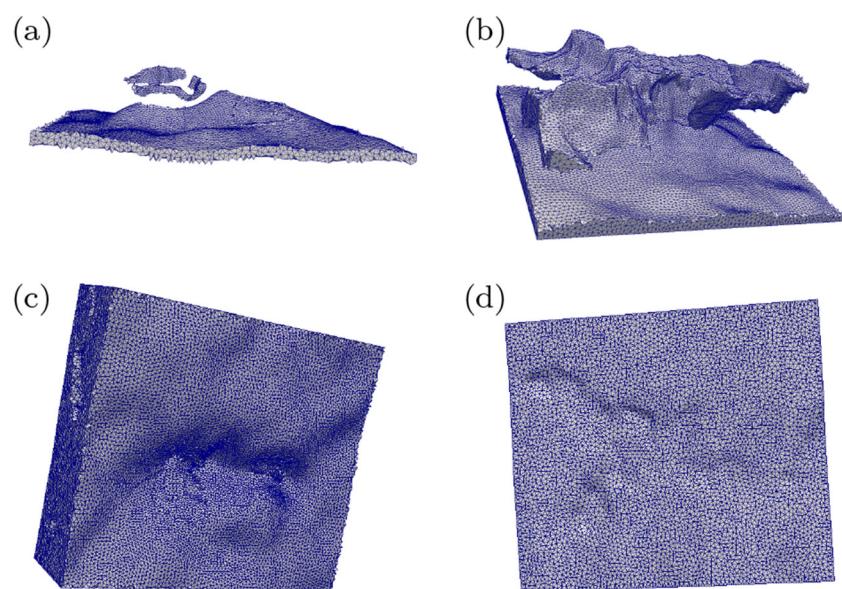
3.3 An idealized computational experiment using single-frequency data

To follow the theory [6] closely, here, we use an example to show the convergence using single-frequency data, if the exact partition is known. Some preliminary results can be found in [50]. We use a complex salt problem using the

extended SEG Advanced Modeling Program (SEAM) phase I model [22]. To describe the model using fully unstructured tetrahedral meshes, we need to triangulate the exterior and internal discontinuities. To illustrate the procedure, which is similar to the work in medical imaging [21], we utilize image segmentation for generating the surface meshes for discontinuities using the Computation Geometry Algorithms Library (CGAL) [20] and then construct the entire volume meshes. We separate the model into four domains to capture the major geological features (see Fig. 2, similar to the early work [63]). Once these surface meshes are generated, we use Tetgen to generate the entire unstructured tetrahedral mesh. We note that this procedure is flexible and allows us to control the quality of our desired mesh, including the smoothness of the surface mesh, the number of triangles and tetrahedra.

The true model, which is shown in Fig. 3 (at) and (bt) for V_P and V_S , respectively, is piecewise constant. The model size is $7 \text{ km} \times 8 \text{ km} \times 3 \text{ km}$. Each model contains 14 subdomains that form the four main subdomains in Fig. 2. Each main subdomain is equally divided into three to four subdomains. We use 3.0 Hz data to perform the inversion with 40 iterations. Fifty-six ($n_x = 7, n_y = 8$) sources are regularly spaced on the top boundary and represent three directional tractions. Fifty-six ($n_x = 7, n_y = 8$) receivers are also regularly spaced in the top boundary. The sources and receivers are not spatially coincident. The smallest offset is around 50 m. In Fig. 3 (a0) and (b0), we show the starting V_P and V_S models. Since the partitioning of the true models is known and the number of subdomains is small, we match the assumption in the theory [6] and expect that the stability constant is small. Hence, the reconstruction shown in Fig. 3 (af) and (bf) is good. The values of the bottom two layers are affected due to the PMLs. We

Fig. 2 Composition of the mesh of the extended SEAM phase I model. (a) Mesh underneath the salt body. (b) Mesh of the salt body. (c) Mesh around the salt body. (d) Mesh of the original water bottom



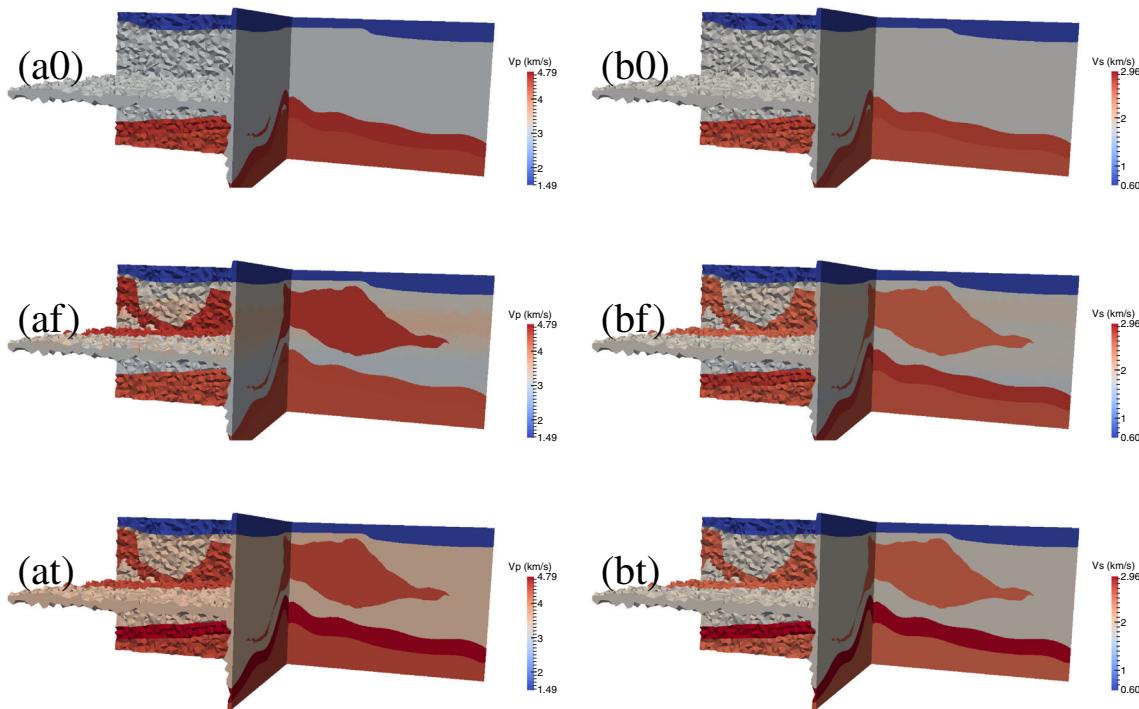


Fig. 3 Left column: V_P models, (a0) starting V_P model, (af) reconstructed V_P model, (at) true V_P model; right column: V_S models, (b0) starting V_S model, (bf) reconstructed V_S model, (bt) true V_S model. The model size is $7 \text{ km} \times 8 \text{ km} \times 3 \text{ km}$. Middle slices, that is, $x = 3.5 \text{ km}$, $y = 4.0 \text{ km}$, $z = 1.5 \text{ km}$, are shown in all figures. The x axis points outwards and the y axis points from left to right

model. The model size is $7 \text{ km} \times 8 \text{ km} \times 3 \text{ km}$. Middle slices, that is, $x = 3.5 \text{ km}$, $y = 4.0 \text{ km}$, $z = 1.5 \text{ km}$, are shown in all figures. The x axis points outwards and the y axis points from left to right

verify the theoretical analysis [6] using this example with single-frequency partial boundary data and known domain partition. The salt body that was completely missing in the initial model is successfully recovered. In fact, since the number of subdomains is known, we can start with relatively higher frequency.

4 Computational study: progressive refinement

In this section, we develop a computational, multi-level approach consistent with the theory accounting for the fact that the stability constant will grow exponentially with the number of subdomains in the partition. This multi-level technique allows the radius of convergence to be enlarged. We scale the gradient with the diagonal of the Gauss–Newton Hessian, H^{GN} say; this scaling appears as a weight matrix on the left-hand side of (10). We note that the use of projections (corresponding to iterative regularization) allows us to avoid over-parameterization. We demonstrate the convergence with piecewise constant models containing shallow and deep geological structures. The true model used in this section contains complex geological features using mentioned tetrahedral representations.

In the iteration, we make a simplification by choosing a fixed step size following rules explained below. We update V_P and V_S alternatively from the updated Lamé parameters. A level-wise stopping criterion, as well as rules for *gradual* refinement of the domain partition together with frequency progression, is discussed in the following subsections. We demonstrate that a piecewise constant 1D layered initial model suffices to obtain convergence. This initial model has very few parameters and will be an approximation (in L^2 norm) to the true model. We typically start the multi-level scheme with 1.0- or 1.5-Hz data necessitated by the general complexity of the true models.

4.1 Iterative regularization and strategy

In this subsection, we discuss our iterative regularization strategy and rules associated with the aforementioned multi-level projected steepest descent method. Our initial subdomains have roughly cubical shapes since we have limited knowledge about the subsurface. We choose the frequency roughly proportional to the cubic root of the number of subdomains, N . This choice is motivated by minimizing the upper bound for the stability constant (maximizing the radius of convergence) in frequency for a given number of subdomains [5, (41)]. In the acoustic case,

the stability constant is studied quantitatively as well and the choice of frequency can be evaluated from a quantitative estimate [5]. Given \bar{r} as the average radius of the subdomain and c_m as the shortest (shear-)wave speed, we have

$$N = \frac{\text{Vol}(X)}{\bar{r}^3} = \frac{\text{Vol}(X)}{[\alpha(c_m/\omega)]^3} = \alpha^{-3}\omega^3 c_m^{-3} \text{Vol}(X), \quad (11)$$

where $\alpha = \bar{r}/(c_m/\omega)$ is a scaling factor. In Table 1, we summarize our observations of different choices of (ω, N) pairs. A small N may cause an issue of poor resolution while a large N may result in divergence. If N is increased too rapidly the current model might no longer be within the radius of convergence of the next level approximation of the “true” model. In Fig. 4, we show the choices of (ω, N) pairs in the later sections. In practice, we relate the diameter of a subdomain, \bar{r} , to the wavelength and determine the normalization of the diameter of a subdomain by the shortest (shear-)wavelength, c_m/ω . This is not dissimilar from homogenization [12].

The (fixed) step length is chosen in accordance with the following rule. First, we scale the gradient of the misfit functional by the diagonal of the Gauss-Newton Hessian. Then, we determine the energy norm of the gradient and multiply it with the maximum value of V_P or V_S at the current iteration. Since the computational cost for linear search is quite high, we typically take 10% of this value as the step size. At the lowest levels, we can enlarge the step size for computational efficiency.

It will also help us to check if we obtain linear convergence, since the stability constant C in Eq. 5 is related to the slope of the residual curve, which is shown in Section 5. We monitor the decrease in residual as well as the norm of the gradient [38, Chapter 3]. We stop the level-wise iteration when the relative change in residual when updating V_S is less than 1%; however, if the norm of the gradient determining the update of V_S becomes less than 1% of the norm of the initial gradient, we stop the iteration as well.

We also monitor the initial convergence rate: if it is linear, we are within the radius of convergence to obtain the best approximation at the next level. The convergence is valid only on the projected space. Hence, a range of models can be good approximations to the true one as long as they stay in the convergence radius of the next level. Here we exploit that due to Lipschitz stability, the convergence is necessarily linear as proved in [18]. We adjust the refinement of the

Table 1 Observations of different choices of (ω, N) pairs

Choices	Observations
$\omega \ll \alpha c_m [N/\text{Vol}(X)]^{1/3}$	Possible to diverge
$\omega \approx \alpha c_m [N/\text{Vol}(X)]^{1/3}$	Relatively optimal
$\omega \gg \alpha c_m [N/\text{Vol}(X)]^{1/3}$	Convergence with poor resolution

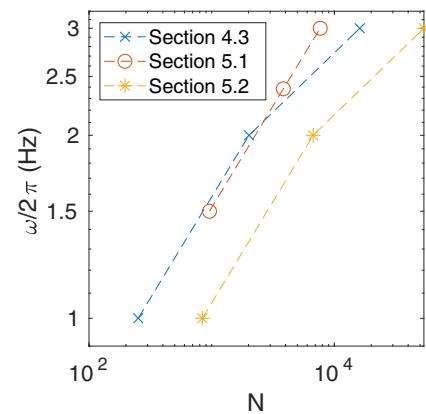


Fig. 4 Choices of different (ω, N) pairs in different sections. The numerical slopes of the curves indicate that $\omega \sim N^{0.264}$, $N^{0.333}$ and $N^{0.264}$ for the experiments in Section 4.3, Section 5.1 and Section 5.2, respectively

domain partitioning accordingly. The choice of N in Eq. 4 and ω pair is important.

The initial convergence rate can illustrate the success of the reconstruction and let us know if the starting model is in the convergence radius.

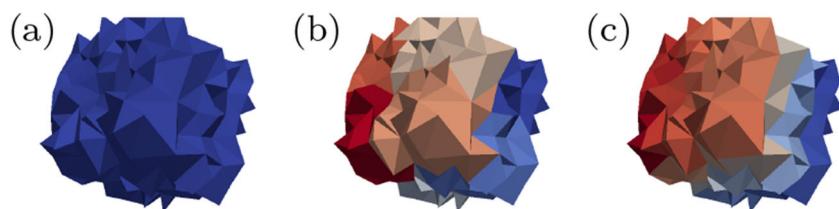
4.2 Domain partition

The computational subdomain refinement procedure is important for the reconstruction. In principle, the fully unstructured mesh allows us to design arbitrary domain partition. Figure 5 illustrates the local refinement of the parametric representations from level to level. At each level, we partition the computational mesh into the subdomains with a size of approximately half of the shear wavelength. During the inversion, these subdomains are utilized as the projections onto a hierarchy of stable subspaces. In the later Section 4.3 and 5, we utilize box-like partitions at different levels as shown in Fig. 5. Ideally, the local partition can be performed based on the gradient and model information. Other techniques, such as the random mesh projectors [28] used in machine learning community and shape optimization [52], can further be applied to improve the modeling of complex geological features. We note that the classical Tikhonov weights can also be space-dependent and can help to adjust the local velocities. The meshing technique extends the generality and can help to model known discontinuities, such as topography, arid terrains, and karst fields [48].

4.3 An idealized computational experiment illustrating convergence radii

To verify that our proposed iterative regularization using projections onto a hierarchy of stable subspaces leads to

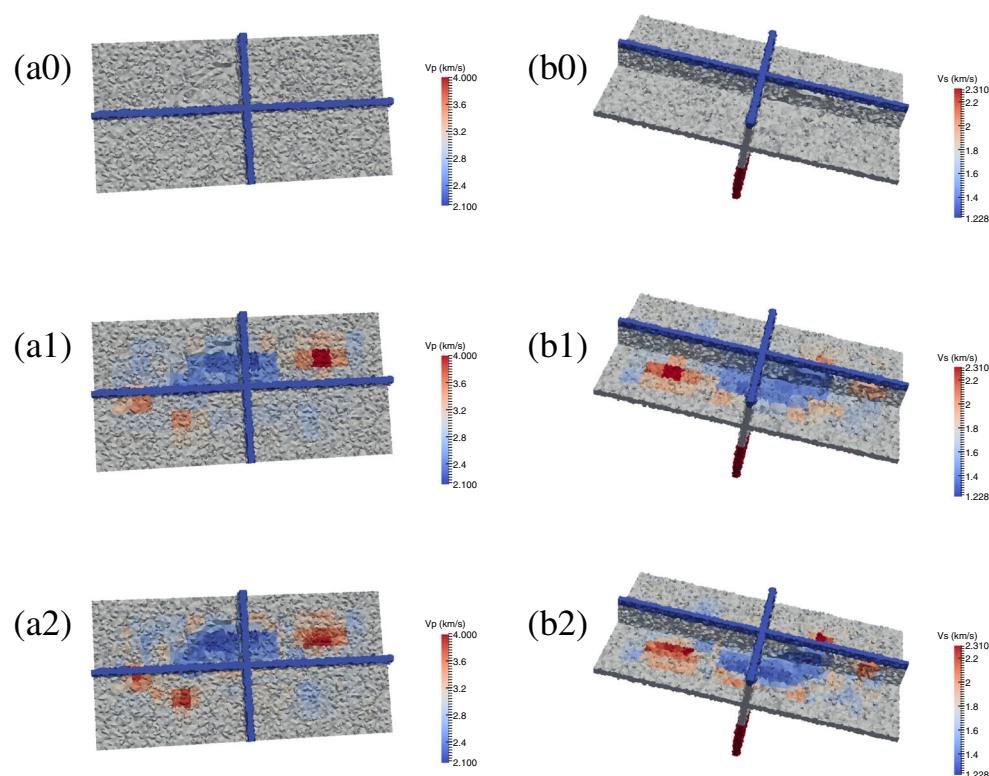
Fig. 5 Refinement (from (a) to (b) to (c)) on the parametric representations. (a) A single piece at the first level. (b) Eight pieces at the second level. (c) Sixty-four pieces at the third level



convergence, we present a computational experiment in which the true model is piecewise constant, and it is recovered on a tetrahedral mesh. We consider geological bodies with rough shapes and high-contrast variations in P - and S -wave speeds. In the experiment, the wave speeds at the boundary are assumed to be known. The model contains a background structure of three (plane) layers with constant wave speeds while the heterogeneous bodies are contained in the middle layer. The background structure is used as the starting model.

First, we study the reconstruction of shallow heterogeneities. Figure 7 (at) and (bt) show the true V_P and V_S models. Figure 7 (bt) shows the mesh. These shallow bodies have different shapes and wave speed values higher or lower than the background wave speed values. The sizes of the bodies vary from 200 m to 3 km. The P -wave speed contrast varies from -0.9 to 0.8 km/s and the S -wave speed contrast varies from -0.52 to 0.46 km/s. At 1.0 Hz, the diameter of the subdomain is 0.5 km and the shortest wavelength is around 1 km.

Fig. 6 Left column: V_P models, (a0) starting V_P model, (a1) reconstructed V_P model at the first level, (a2) reconstructed V_P model at the first level; right column: V_S models, (b0) starting V_S model, (b1) reconstructed V_S model at the first level, (b2) reconstructed V_S model at the second level. The model size is $8 \text{ km} \times 4 \text{ km} \times 3 \text{ km}$. In the bottom layer, $V_P = 4.0$ km/s and $V_S = 2.31$ km/s



The depths of the bodies range from 200 m to 1.5 km. The starting model has three constant P - and S -wave speeds in layers shown in Fig. 6 (a0) and (b0). The model size is $8 \text{ km} \times 4 \text{ km} \times 3 \text{ km}$ and contains 1.3 million elements.

We design three levels for the recovery, with each level providing a rough box-like domain partition of the model. The first level contains 252 ($n_x = 14, n_y = 6, n_z = 3$) subdomains, the second level contains 2016 ($n_x = 28, n_y = 12, n_z = 6$) subdomains, and the third level contains 16128 ($n_x = 56, n_y = 24, n_z = 12$) subdomains. For each level, we perform a maximum of 150 iterations for the reconstruction at a fixed frequency. Eighty ($n_x = 16, n_y = 5$) sources are regularly spaced in the top boundary and represent three directional tractions. Eighty ($n_x = 16, n_y = 5$) receivers are also regularly spaced in the top boundary. The sources and receivers are not spatially coincident. The smallest offset is around 50 m.

Following Section 4.1, at the first level, we select 1.0 -Hz data, and the results are shown in Fig. 6 (a1) and (b1). At the

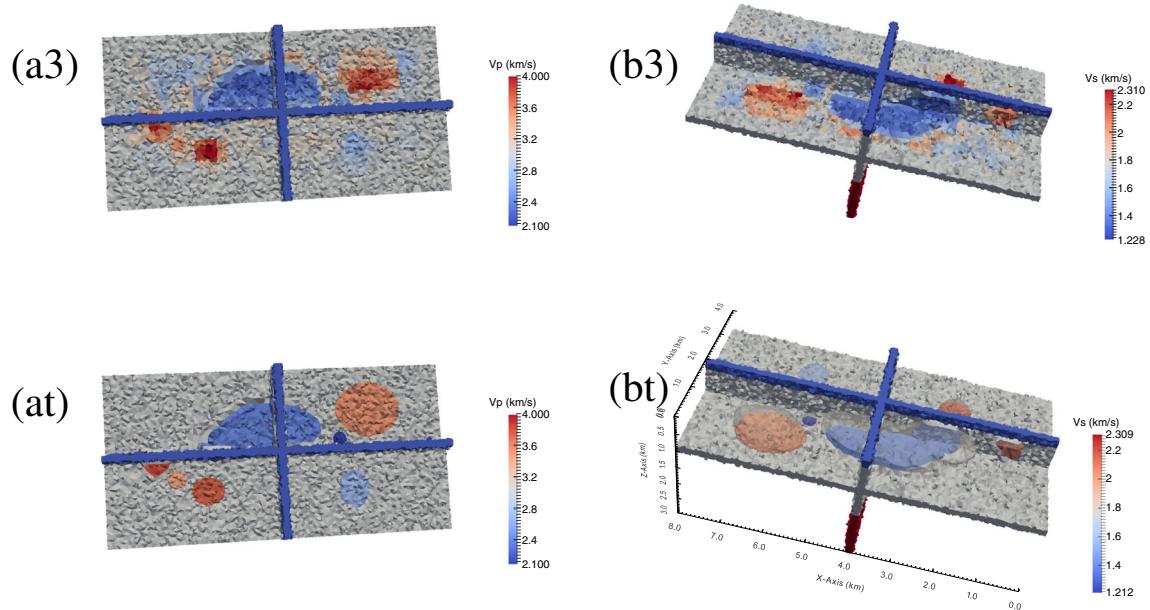


Fig. 7 Final reconstructions: (a3) inverted V_P model at the third level, (b3) inverted V_S model at the third level; true model: (at) V_P , (bt) V_S . The model size is $8 \text{ km} \times 4 \text{ km} \times 3 \text{ km}$. In the bottom layer, $V_P = 4.0 \text{ km/s}$ and $V_S = 2.31 \text{ km/s}$

second level, we select 2.0-Hz data; the results are shown in Fig. 6 (a2) and (b2). At the third level, we select 3.0-Hz data, and the results are shown in Fig. 7 (a3) and (b3). Figure 8 shows the true u_y data at different frequencies. Figure 8 (a1–a2, b1–b2, c1–c2) shows a u_y data residual (associated with a centrally located source) from level to level.

For the piecewise constant case, the model representation follows Eq. 4 exactly. Since we begin with a constant layered model, low-frequency data are important to find the rough anomalies. The later local refinement allows us to capture details of the anomalies. The multi-level strategy with sub-domain refinement is suitable to capture the main anomalies.

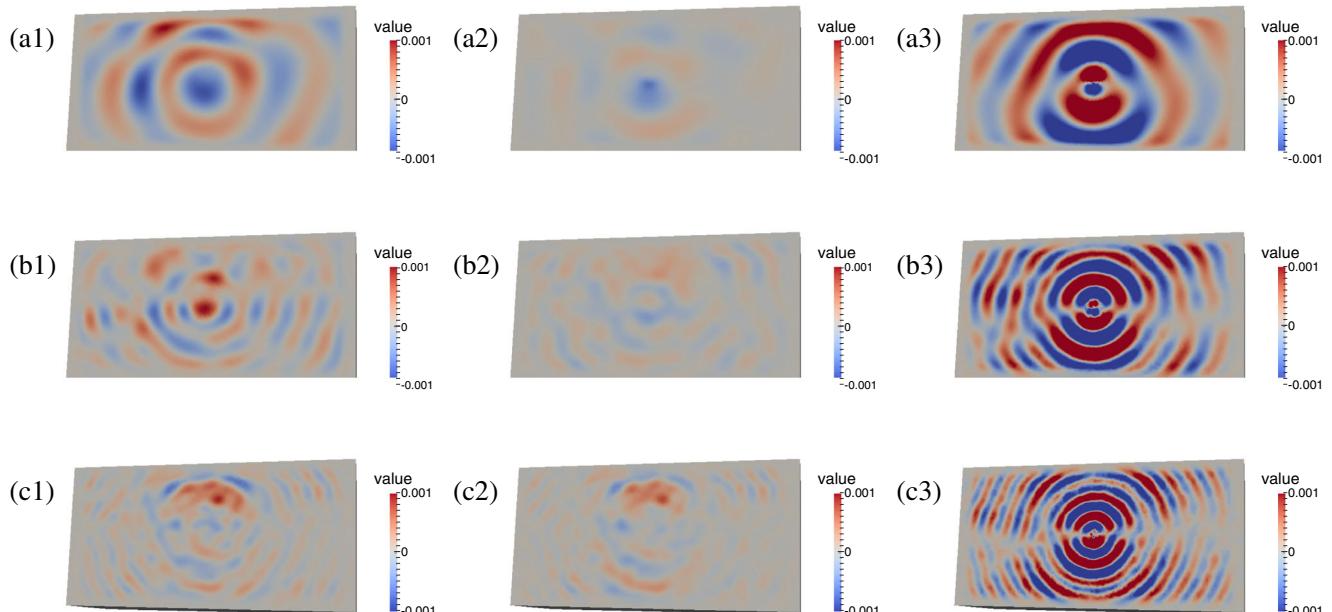
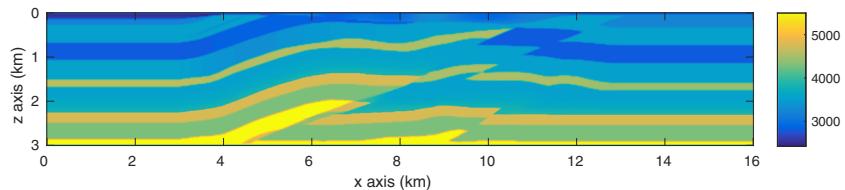


Fig. 8 Data differences from a given boundary traction at the center of the upper surface: at 1.0 Hz, (a1) u_y difference between the starting (cf. Fig. 6(a0,b0)) and true (cf. Fig. 7(at,bt)) models, (a2) u_y difference between the level 1 (cf. Fig. 6(a1,b1)) and true models, (a3) u_y data from the true models; at 2.0 Hz, (b1) u_y difference between the level 1 and true models, (b2) u_y difference between the level 2 (cf. Fig. 6(a2,b2)) and true models, (b3) u_y data from the true models; at 3.0 Hz, (c1) u_y difference between the level 2 and true models, (c2) u_y difference between the level 3 (cf. Fig. 7(a3,b3)) and true models, (c3) u_y data from the true models

the level 1 and true models, (b2) u_y difference between the level 2 (cf. Fig. 6(a2,b2)) and true models, (b3) u_y data from the true models; at 3.0 Hz, (c1) u_y difference between the level 2 and true models, (c2) u_y difference between the level 3 (cf. Fig. 7(a3,b3)) and true models, (c3) u_y data from the true models

Fig. 9 A vertical slice of the SEG thrust model from a three-dimensional regular file



5 Computational experiments: multi-parameter elastic FWI

In this section, we study the performance of our FWI algorithm in two cases as the best approximate models are computed in a hierarchy of stable subspaces defined by tetrahedral mesh refinement. We note that both true models are piecewise smooth. The true data are generated from the true model with the true mesh with complicated geometries. The simulated data are generated from the model with the computational mesh. In current experiments, no noise is included in the simulated data. However, for the low-frequency updates, since we use fewer subdomains, the stability constant is small, the reconstruction can tolerate noise.

5.1 SEG thrust model

To verify that our proposed iterative regularization (using projections onto a hierarchy of stable subspaces) leads to convergence, we present one computational experiment, in which the true model is represented and recovered on a tetrahedral mesh. In this example, we use the SEG thrust model as the true one.

To illustrate the mesh generation, we use a two-dimensional slice, see Fig. 9, as an example to perform image segmentation. The top row of Fig. 10 shows several individual features in Fig. 9. We illustrate the three-dimensional surface mesh in the bottom row of Fig. 10. This procedure is helpful for multi-resolution analysis.

Figure 11 (a0) and (b0) show the starting V_P and V_S models, which are far away from the true models in Fig. 12 (a0) and (b0). The starting models can be obtained from tomographic results. The size of the SEG thrust model is 16 km \times 4.41 km \times 3 km. The left columns in Figs. 11 and

12 show the vertical slice of the V_P models at $y = 1.5$ km and both right columns show the horizontal slice of the V_S models at $z = 0.8$ km. As mentioned in Section 1, we start from 1.5 Hz, which can be obtained in the field [56].

We design three levels for the recovery, with each level providing a rough box-like domain partition of the model. The first level contains 960 ($n_x = 24, n_y = 8, n_z = 5$) subdomains, the second level contains 3840 ($n_x = 48, n_y = 8, n_z = 10$) subdomains, and the third level contains 7680 ($n_x = 48, n_y = 16, n_z = 10$) subdomains. For each level, we perform a maximum of 120 iterations for the reconstruction at a fixed frequency. A total of 240 ($n_x = 30, n_y = 8$) sources are regularly spaced in the top boundary and represent three directional tractions. These three components are treated separately. Two hundred forty ($n_x = 30, n_y = 8$) receivers are also regularly spaced in the top boundary. The sources and receivers are not spatially coincident. The smallest offset is around 50 m.

Following Section 4.1, at the first level, we select 1.5-Hz data, and the results are shown in Fig. 11 (a1) and (b1). At the second level, we select 2.385-Hz data; the results are shown in Fig. 11 (a2) and (b2). At the third level, we select 3.0-Hz data, and the results are shown in Fig. 12 (a3) and (b3). The width of the PMLs is 0.7 km and the models (16 km \times 4.41 km \times 3 km) that are shown contain the PMLs. Hence, the boundary box of the V_S shown in Fig. 12 (b3) is not well updated.

We show that the proposed strategy can work on the classic layered model problem. While the conventional FWI approaches usually begin with a smooth tomographic model as the starting model, based on the analysis, we start with a low-dimensional piecewise constant model. We check if the start model is in the convergence radius by monitoring convergence rates. Fig. 13 shows the residuals of the misfit at different levels. The residuals drop to 10%,

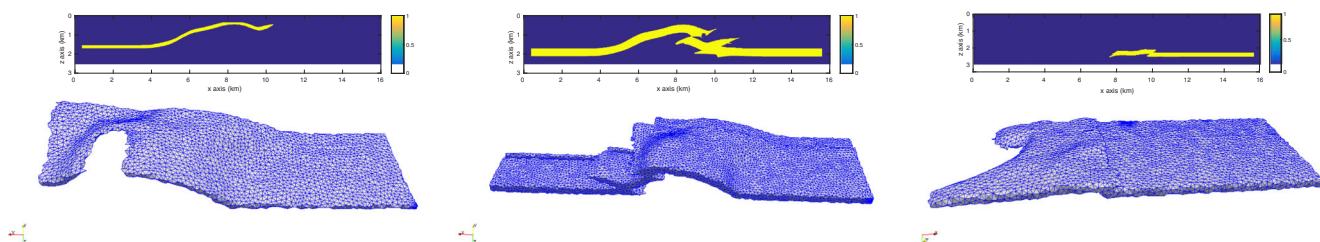


Fig. 10 Surface mesh generation. Top row: image segmentations of each individual geological features; Bottom row: three-dimensional surface mesh for the corresponding features

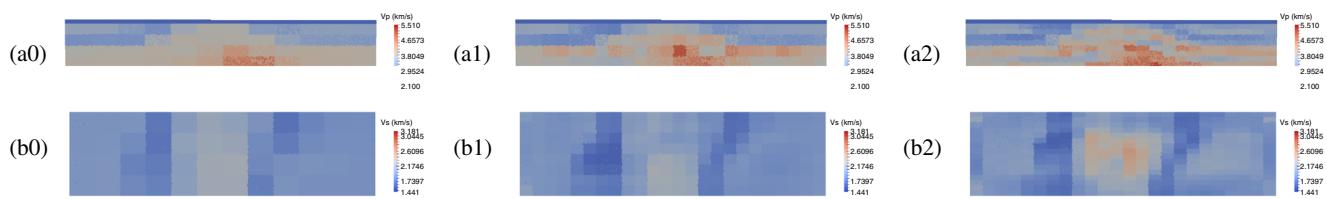


Fig. 11 Top row: V_P models; (a0) starting V_P model, (a1) reconstructed V_P model at the first level, (a2) reconstructed V_P model at the second level; Bottom row: V_S models; (b0) starting V_S model, (b1) reconstructed V_S model at the first level, (b2) reconstructed V_S model

at the second level. The model size is $16 \text{ km} \times 4.41 \text{ km} \times 3 \text{ km}$. In (a0), (a1), and (a2), $y = 1.5 \text{ km}$ and the slice size is $16 \text{ km} \times 3 \text{ km}$. In (b0), (b1), and (b2), $z = 0.8 \text{ km}$ and the slice size is $16 \text{ km} \times 4.41 \text{ km}$

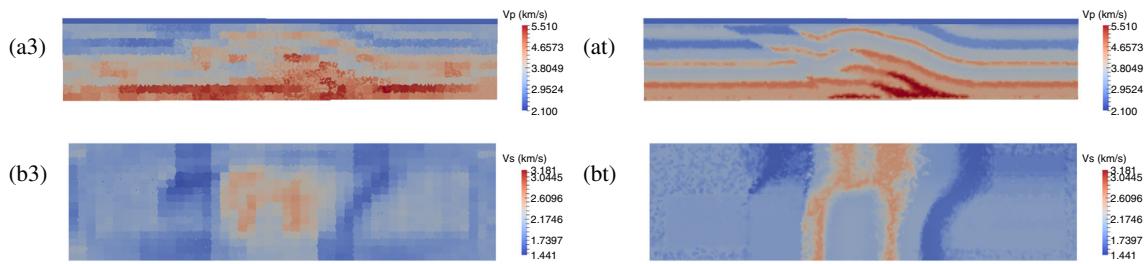


Fig. 12 Final reconstructions: (a3) inverted V_P model at the third level, (b3) inverted V_S model at the third level; true model: (at) V_P , (bt) V_S . The model size is $16 \text{ km} \times 4.41 \text{ km} \times 3 \text{ km}$. In (a3) and (at),

$y = 1.5 \text{ km}$ and the slice size is $16 \text{ km} \times 3 \text{ km}$. In (b3) and (bt), $z = 0.8 \text{ km}$ and the slice size is $16 \text{ km} \times 4.41 \text{ km}$

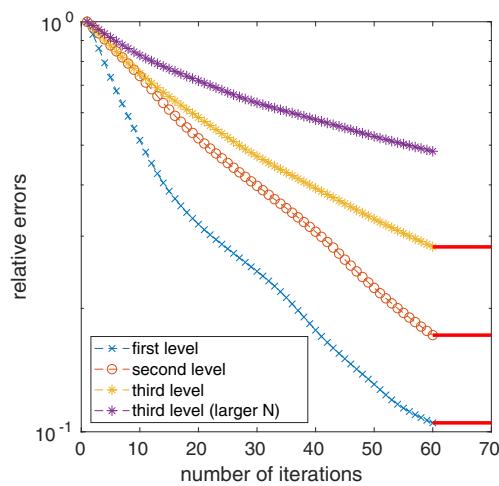


Fig. 13 The residuals of the misfit at different levels. The residuals drop to 10%, 17%, and 30% of the starting value at the first, second, and third levels, respectively. The purple line shows the residual decay with a much larger $N = 61448$ at the third level. The y axis is on the log scale. The red lines show that the residuals reach the stopping criteria

17%, and 30% of the starting value at the first, second, and third levels, respectively. It also shows the residual decay with a much larger $N = 61448$ at the third level. The residual decay is slow and the updated model does not have much improvements. It confirms our observations in Table 1. Since the slope of the residuals stays approximately constant, it is consistent with the analysis that the slopes of the residual curves are proportional to the stability constant C in Eq. 5. We monitor the residual curves and stop the iteration when the residual decay is roughly no longer linear. We expect that the results can further be improved using higher frequency data with higher computational costs.

5.2 SEAM

Here, we return to the SEAM as our example. In our test case, the starting model shown in Fig. 14 (a0–b0) has six planar layers; on each layer, the P - and S -wave speeds are constant. The true model shown in Fig. 15 (at–bt) is adapted from SEAM Phase I; the original water layer and model size have been modified. The top layer contains a Poisson solid with constant P -wave speed 2.1 km/s. Note that the true model is piecewise smooth, but not piecewise constant, and therefore lies outside the stable subspace hierarchy. The model size is 7 km \times 8 km \times 3 km and each model contains 1.1 million elements. Figure 16 shows the true

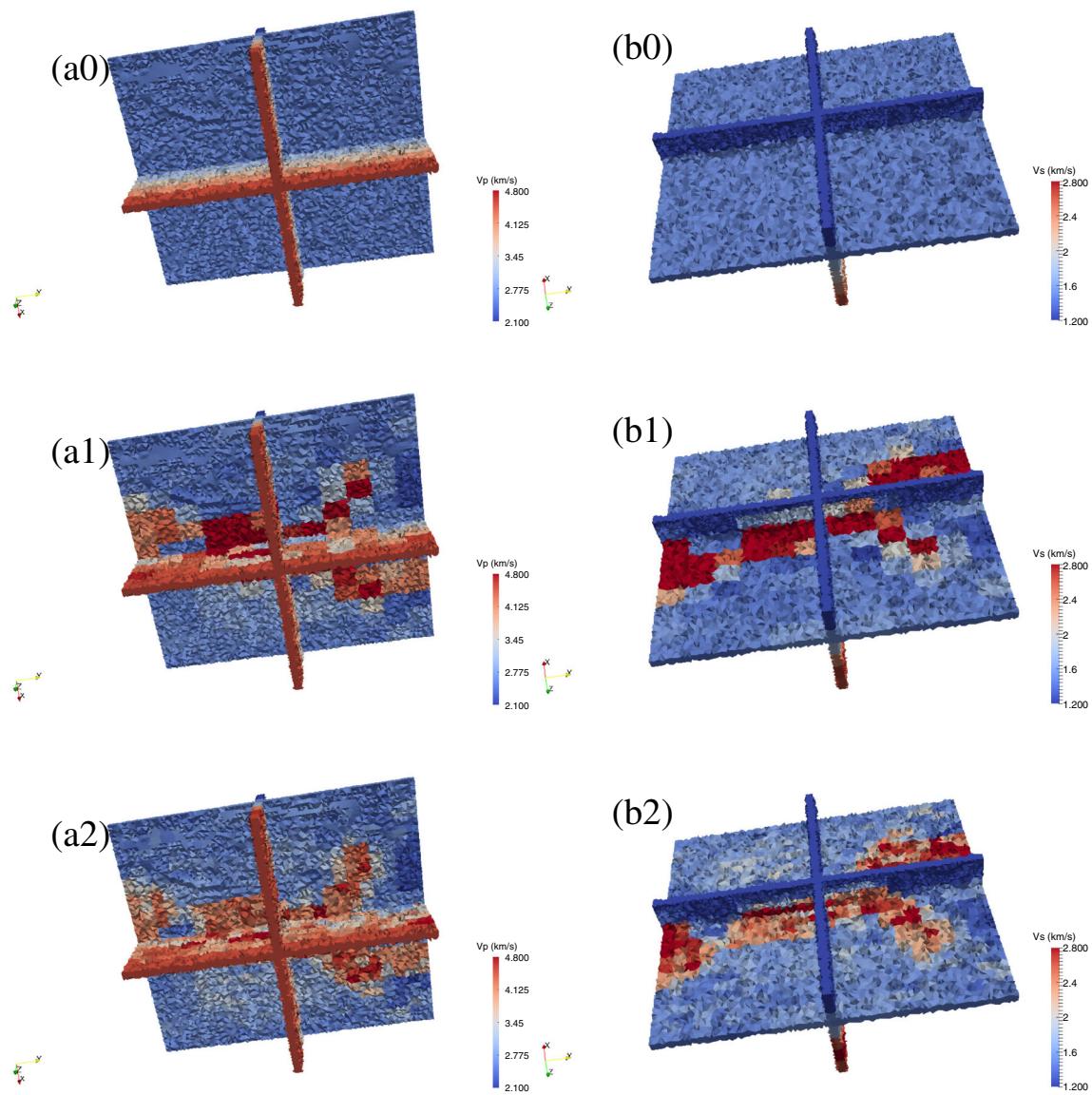


Fig. 14 Left column: V_P models; (a0) starting V_P model, (a1) reconstructed V_P model at the first level, (a2) reconstructed V_P model at the first level; Right column: V_S models; (b0) starting V_S model, (b1)

reconstructed V_S model at the first level, (b2) reconstructed V_S model at the second level. The model size is 7 km \times 8 km \times 3 km. The x axis points inwards and the y axis points from left to right

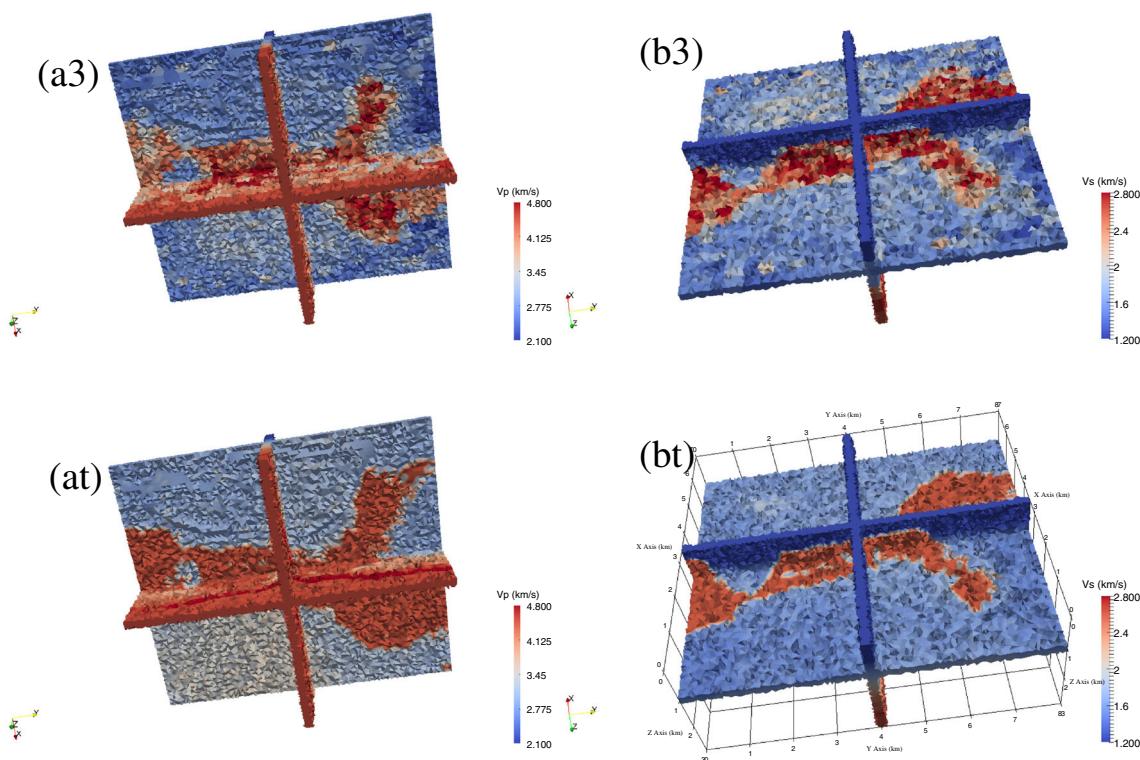


Fig. 15 Final reconstructions: (a3) inverted V_P model at the third level, (b3) inverted V_S model at the third level; true model: (at) V_P , (bt) V_S . The model size is 7 km \times 8 km \times 3 km

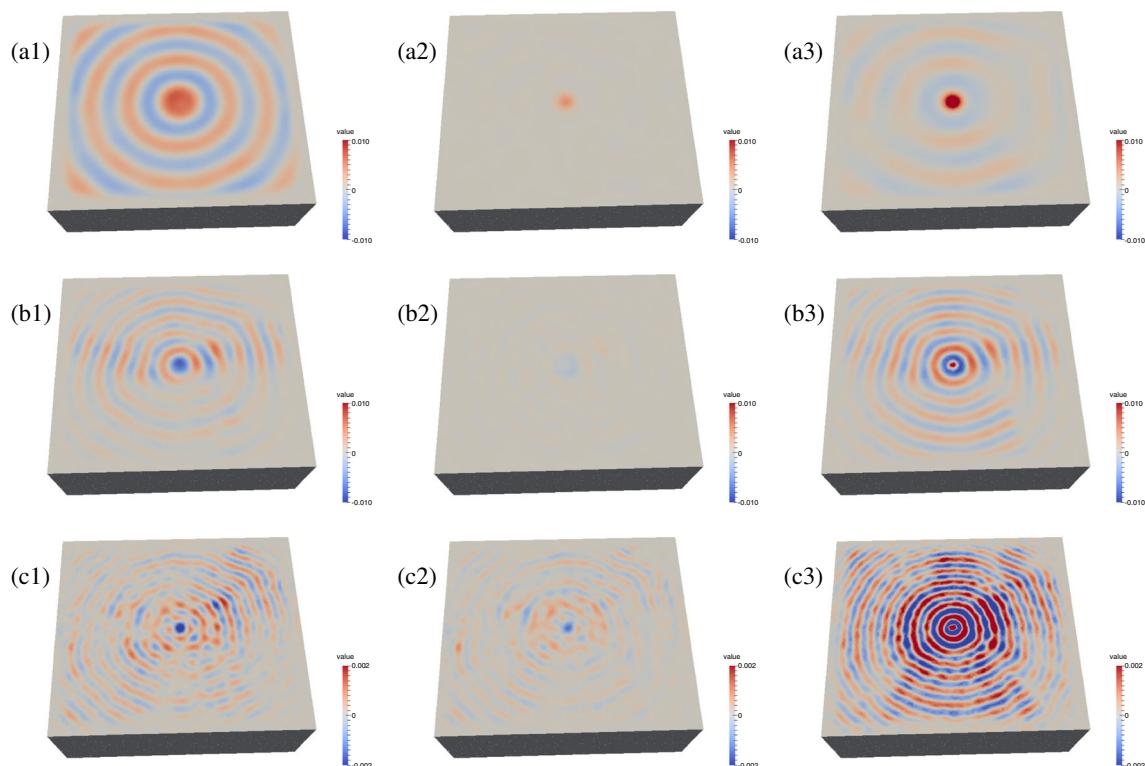
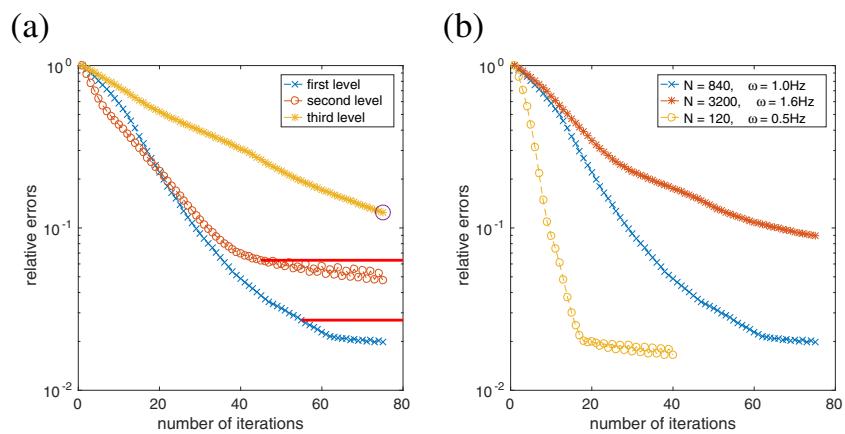


Fig. 16 Data differences from a given boundary traction at the center of the upper surface: at 1.0 Hz, (a1) u_z difference between the starting (cf. Fig. 14(a0,b0)) and true (cf. Fig. 15(c,d)) models, (a2) u_z difference between the level 1 and true models, (a3) u_z data from the true models. At 2.0 Hz, (b1) u_z difference between the level 1 and true

models, (b2) u_z difference between the level 2 and true models, (b3) u_z data from the true models. At 3.0 Hz, (c1) u_z difference between the level 1 and true models, (c2) u_z difference between the level 3 and true models, (c3) u_z data from the true models

Fig. 17 Error log plots for data. (a) The data residuals at different levels; the red lines show that the residuals reach the stopping criteria; the circle indicates that we stop at the maximum iteration. (b) The data residual at the first level with different number of subdomains coupled with different frequencies. The blue curve is the same blue curve in (a)



vertical displacements at different frequencies. At 1.0 Hz, the diameter of the subdomain is again about 0.5 km and the shortest wavelength is around 1 km.

As before, we design three levels for the recovery, each level providing a rough box-like domain partition of the model. In this example, we also include lateral PML regions for the update. The first level contains 840 ($n_x = 14, n_y = 12, n_z = 5$) subdomains, the second 6720 ($n_x = 28, n_y = 12, n_z = 10$) subdomains, and the third 53760 ($n_x = 56, n_y = 24, n_z = 20$) subdomains. At each level, we perform 75 iterations for the reconstruction at a fixed frequency. A total of 56 ($n_x = 7, n_y = 8$) sources are regularly spaced on the top boundary and represent three directional tractions. Fifty-six ($n_x = 7, n_y = 8$) receivers are also regularly spaced in the top boundary. The sources and receivers are not spatially coincident. The smallest offset is around 50 m.

We also select 1.0-Hz, 2.0-Hz, and 3.0-Hz data respectively for the first level (Fig. 14 (a1–b1)), second (Fig. 14 (a2–b2)), and third levels (Fig. 15 (a3–b3)). Figure 16 (a1–a2, b1–b2, c1–c2) shows the decay in u_z data residual (for a centrally located source) from level to level.

In Figs. 17 and 18, we show data and model errors during our iterations. Figure 17(a) plots the data residual over the course of the algorithm at different levels. As the

plot shows, the behavior of the residual change is consistent with a projected gradient-based method. In the first several iterations, the errors decay linearly, which indicates the procedure is converging to the next best approximation that we can obtain. In Fig. 17(b), we plot the data residual over different number of subdomains ($N = 120, 840, 3200$, plus a fixed near surface layer) at the first level. It shows that when N is too small, that is, the diameter of the subdomain is too large, the residual may diverge in the later inversion because the inverted model may not stay in the convergence radius of the next level; when N is too large, that is, the diameter of the subdomain is too small, the residual may also diverge even using a smaller step size. It confirms our observations in Table 1. Since we start without prior information in this experiment, it is challenging to capture large-scale anomalies. It is important to set up a proper N and ω pair. As described in Section 4.1, we check whether the coupled choice of N and ω brings us within the radius of convergence. We point out that the convergence is solid only on the projected space. We check this by monitoring convergence rates as illustrated in Fig. 17(a) and (b). The projected model errors in Fig. 18 show linear convergence rates at different levels as they should. The slopes of the residual curves are related to the stability constant C in Eq. 5. These linear convergence rates apply directly to the

Fig. 18 Error log plots for Figs. 14 and 15. (a) The projected λ model errors at different levels. (b) The projected μ model errors at different levels. Here we compute the errors between the current model and the best projected model at the current level

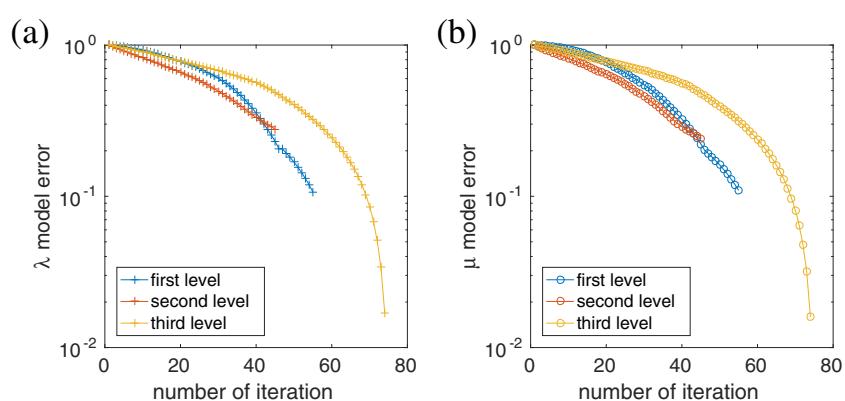


Table 2 Final model relative errors between models at different level and the projected true model and final relative residuals at different levels and frequencies

Relative errors	Start	Level 1	Level 2	Level 3
$V_P (L^2)$	1.0	0.447	0.272	0.192
$V_S (L^2)$	1.0	0.361	0.213	0.168
L^2 residuals at 1.0 Hz	1.0	0.027		
L^2 residuals at 2.0 Hz		1.0	0.0632	
L^2 residuals at 3.0 Hz			1.0	0.125

reconstructions level by level, as illustrated in Fig. 18. The convergence of the best approximations in each level to the true model and relative residuals is illustrated in Table 2. We note that the projected true model may not always be the best approximation at each level.

6 Discussion

We presented a scheme for multi-parameter elastic FWI with iterative regularization, motivated by a new result [6] on uniqueness and a conditional Lipschitz stability estimate for model representations that are piecewise constant on unstructured tetrahedral meshes in the inverse boundary value problem associated with vibroseis data. We developed a procedure to generate a hierarchy of such representations or parametrizations via adaptive mesh refinement. Such a hierarchy enabled an implementation of the multi-level scheme—with frequency-scale progression—introduced and analyzed in [18] which comes with conditions for convergence. The iterative regularization is numerically implemented to avoid over-parameterization of the original problem. Note that we do not need strict mesh refinement as we can change coarser elements from level to level. We could even run Poisson-Voronoi with multiple realizations at each level. In a second pass, for each level, you can search for the best linear combination of the realizations. The hierarchy of parametrizations allows robust estimation of salt bodies with rough shapes and complex geological structures from simple starting models. Our starting models are typically constructed with a few blocks. Recent progress in the development of massively parallel structured direct solvers [62] makes it possible to apply our scheme to realistically sized problems. From the underlying mathematical analysis, we expect that the estimation of highly complex geological structures far from, say, models with piecewise smooth P - and S -wave speeds requires low-frequency data as confirmed by our computational experiments. This is in agreement with recent work [49].

Acknowledgments J.S. would like to thank Jean Virieux, Peter Caday, Ruichao Ye, and Florian Faucher for useful discussions. The authors would also like to thank two anonymous referees for their helpful comments.

Funding information MVdH and JS were financially supported by the Simons Foundation under the MATH + X program, the National Science Foundation under grant DMS-1815143, and by the members of the Geo-Mathematical Group at Rice University.

Appendix A: Formulation of perfectly matched layers with the Neumann boundary conditions

The PML technique [4] was first implemented and used in a finite-difference time-domain method for the computation of electromagnetic waves. A more straightforward and convenient approach [14] was introduced by using complex coordinate stretching to build the same PMLs. Here, we will follow the work for the isotropic case [65] with the caveat that the half-space problem with a Neumann boundary condition on the top requires some adaptations. That is, we will need a constraint for the damping function in constructing the PMLs.

We let $S_i(x_i)$ be a complex-valued damping function. We note that each $S_i(x_i)$ is only a function of x_i and is independent of other coordinates. We adjust the partial derivatives, $\partial_{x_i} \rightarrow \frac{1}{S_i} \partial_{x_i}$, with S_i being identically one in the domain of interest and complex-valued inside the PML region.

Numerically, we expect $u_l|_{\partial X/\Sigma} \rightarrow 0$. The boundary value problem (1)–(2) takes the form

$$\begin{cases} (-\rho(x)\omega^2\delta_{il} - \frac{1}{S_j}\partial_{x_j}c_{ijkl}(x)\frac{1}{S_k}\partial_{x_k})u_l = 0, \\ (c_{ijkl}\frac{1}{S_k}\partial_{x_k}u_l)v_j|_{\Sigma} = g_i. \end{cases} \quad (12)$$

To arrive at the weak formulation, we need to carry out the following steps. We multiply both sides of (12) by $S_1S_2S_3$,

$$(-S_1S_2S_3\rho(x)\omega^2\delta_{il} - \partial_{x_j}c_{ijkl}(x)\frac{S_1S_2S_3}{S_jS_k}\partial_{x_k})u_l = 0,$$

noting that $S_1S_2S_3/S_j$ is not a function of x_j . We now introduce coefficients,

$$\tilde{\rho}(x) = S_1S_2S_3\rho(x), \quad \tilde{c}_{ijkl}(x) = c_{ijkl}(x)\frac{S_1S_2S_3}{S_jS_k}, \quad x \in X \cup \partial X, \quad (13)$$

where $X \cup \partial X$ is the computational box with PML inside X . When we apply the classical PML coefficient S_j , we observe the reflected surface waves from the corners of the upper surface. This is because we have a mismatch between PML and the Neumann boundary condition. Here

we modify the PML coefficient so that we can deal with the boundary conditions properly. We let

$$S_j|_{\partial X} = 1, \quad \text{for } j = 1, 2, 3. \quad (14)$$

We multiply $S_1 S_2 S_3$ to both sides of the Neumann boundary condition,

$$\begin{aligned} (c_{ijkl} \frac{S_1 S_2 S_3}{S_k} \partial_{x_k} u_l) v_j|_{\Sigma} &= (c_{ijkl} \frac{S_1 S_2 S_3}{S_j S_k} \partial_{x_k} u_l) v_j|_{\Sigma} \\ &= (\tilde{c}_{ijkl} \partial_{x_k} u_l) v_j|_{\Sigma} = S_1 S_2 S_3 u_l. \end{aligned} \quad (15)$$

We note that we replace the original coefficients c_{ijkl} with the new coefficients \tilde{c}_{ijkl} at the boundary.

Considering that $S_1 S_2 S_3 g_i|_{\Sigma} = g_i|_{\Sigma}$, we obtain the modified strong formulation,

$$(-\tilde{\rho}(x) \omega^2 \delta_{il} - \partial_{x_j} \tilde{c}_{ijkl}(x) \partial_{x_k}) u_l = 0, \quad (16)$$

$$(\tilde{c}_{ijkl} \partial_{x_k} u_l) v_j|_{\Sigma} = g_i. \quad (17)$$

Since we now have standard derivatives without any complex functions, we are now able to apply the continuous Galerkin finite element approximation to the system with PMLs. We then construct the local matrices on each element and assemble these local matrices into the global matrix. The strategy is similar to the standard work [26].

Appendix B: First-order adjoint state method: the gradient

Elastic FWI can be formulated as an optimization problem with equality constraints.

Since we deal with inverse boundary value problems, to extract the adjoint boundary values for misfit functional, we revisit the classical first-order adjoint state method. We consider a single source g here and sum over all the available sources later. The optimization problem minimizing $\Psi^{HS}(u)$ in Eq. 7 takes the form,

$$\begin{aligned} \operatorname{argmin}_m \Psi^{HS}(u) \text{ subject to} \\ \int_X \left(-\omega^2 \rho u_i v_i + (\partial_{x_j} v_i) c_{ijkl} \partial_{x_k} u_l \right) dx \\ = \int_{\Sigma} g_i v_i dx, \quad \forall v \in H^1(X), \end{aligned} \quad (18)$$

where the constraint in Eq. 18 represents the weak form of the entire boundary value problem (1)–(2), u denotes the weak solution and v denotes the test function. H^1 denotes the Sobolev space of square-integrable functions with square-integrable weak first-order derivatives. We point out that the boundary value problems with discontinuities in the media can naturally be solved in the weak sense. Additionally, to obtain the adjoint boundary value, one needs to derive the adjoint formula in the weak sense.

To compute the gradient of the functional involved, we use a Lagrangian approach, the constrained optimization

problem is cast into a formulation with Lagrange multipliers γ ,

$$\begin{aligned} \mathcal{L}(m, u, \gamma) &= \frac{1}{2} \int_{\partial X} \chi_{\Sigma} R(u_i - u_i^*) \cdot R(u_i - u_i^*) dx \\ &\quad + \int_X \left(-\omega^2 \rho u_i \gamma_i + (\partial_{x_j} u_i) c_{ijkl} \partial_{x_k} \gamma_l \right) dx \\ &\quad - \int_{\Sigma} g_i \gamma_i dx, \end{aligned} \quad (19)$$

where u^* denotes the solution in the true model m^* . Given some m , we let $\tilde{u} = \tilde{u}(m)$ be the solution to the forward boundary value problem and write

$$\mathcal{L}(m, \tilde{u}, \gamma) = \Psi^{HS}(\tilde{u}) = \mathcal{E}(m). \quad (20)$$

Since we consider piecewise constant models as described in Eq. 4, \mathcal{E} is a Fréchet differentiable function $\mathcal{E} : V \rightarrow \mathbb{R}$, where V is a finite-dimensional vector space, the derivative $D_m \mathcal{E}[m]$ exists. Since the Fréchet derivative is continuous, the Riesz representation theorem can be applied, here, using the L^2 inner product in the model space [9]:

$$D_m \mathcal{E}[m] \delta m = (\nabla \mathcal{E}, \delta m), \quad \forall m \in V,$$

where $\nabla \mathcal{E}$ denotes the gradient and $D_m \mathcal{E}$ is defined as the linear operator

$$D_m \mathcal{E}[m] : \delta m \mapsto \frac{d}{dt} \Big|_{t=0} \mathcal{E}(m + t \delta m), \quad \delta m \in V.$$

Since the Fréchet derivative of $\tilde{u}(m)$ exists, the Fréchet derivative of $\mathcal{E}(m)$ with respect to m in the direction δm attains the form

$$\begin{aligned} D_m \mathcal{E}[m] \delta m &= D_m \mathcal{L}(m, \tilde{u}, \gamma) \delta m \\ &= \int_X -\omega^2 \tilde{u}_i \gamma_i \frac{\partial \rho}{\partial m} \delta m dx \\ &\quad + \int_X (\partial_{x_j} \tilde{u}_i) (\partial_{x_k} \gamma_l) \frac{\partial c_{ijkl}}{\partial m} \delta m dx \\ &\quad + \int_X \left(-\omega^2 \rho \gamma_i (D_m \tilde{u}_i[m] \delta m) \right. \\ &\quad \left. + \partial_{x_j} (D_m \tilde{u}_i[m] \delta m) c_{ijkl} \partial_{x_k} \gamma_l \right) dx \\ &\quad + \int_{\partial X} \left(R(D_m \tilde{u}_i[m] \delta m) \chi_{\Sigma} R(\tilde{u}_i - u_i^*) \right) dx. \end{aligned} \quad (21)$$

We choose the adjoint state, $\tilde{\gamma} = \tilde{\gamma}(m)$, so that $(m, \tilde{u}, \tilde{\gamma})$ is a stationary point of the Lagrangian [16, 42, 61].

Thus, applying the calculus of variations, we let $\tilde{\gamma}$ solve

$$\begin{aligned} \int_X \left(-\omega^2 \rho \gamma_i v_i + (\partial_{x_j} v_i) c_{ijkl} (\partial_{x_k} \gamma_l) \right) dx \\ + \int_{\partial X} v_i \chi_{\Sigma} R(\tilde{u}_i - u_i^*) dx = 0, \quad \forall v \in H^1(X). \end{aligned} \quad (22)$$

From Eq. 22, it follows that the first-order adjoint state equation for the boundary value problem takes the form of Eqs. 8 and 9. Clearly, the adjoint boundary value problem

(8)–(9) is well-posed in the weak sense. Substituting $v = D_m \tilde{u}[m] \delta m$ in Eq. 21, we avoid computing $D_m \tilde{u}[m]$ explicitly and obtain,

$$D_m \mathcal{E}[m] \delta m = (\nabla \mathcal{E}, \delta m) = \int_X -\omega^2 \tilde{u}_i \tilde{\gamma}_i \frac{\partial \rho}{\partial m} \delta m \, dx + \int_X (\partial_{x_j} \tilde{u}_i) (\partial_{x_k} \tilde{\gamma}_i) \frac{\partial c_{ijkl}}{\partial m} \delta m \, dx. \quad (23)$$

Summing over all available sources, we arrive at Eq. 10.

Appendix C: Second-order adjoint state method for the inverse boundary value problem

Since the vibroseis data lead to the inverse boundary value problems, we present the evaluation of (full and Gauss-Newton) Hessian-vector multiplication. For the analogous evaluation in the case of traditional FWI, several previous works [23, 36, 44] have been performed.

C.1 Full Hessian-vector product computation

To begin with, we consider the optimization problem with equality constraints with a single source,

$$\begin{aligned} \min_m \Psi_1(u, u_1) \quad & \text{subject to} \\ \int_X \left(-\omega^2 \rho u_i v_{1i} + (\partial_{x_j} v_{1i}) c_{ijkl} \partial_{x_k} u_l \right) dx \\ = \int_{\Sigma} g_i v_{1i} \, dx, \quad & \forall v_1 \in H^1(X), \\ \int_X \left(-\omega^2 \rho u_{1i} v_i + (\partial_{x_j} v_i) c_{ijkl} \partial_{x_k} u_{1l} \right) dx \\ = - \int_X \left(-\omega^2 (\delta l^{\rho}) u_i v_i + (\partial_{x_j} v_i) (\delta l^c)_{ijkl} \partial_{x_k} u_l \right) dx \\ + \int_{\partial X} -[(\delta l^c)_{ijkl} \partial_{x_k} u_l] v_j v_i \, dx, \quad & \forall v \in H^1(X), \end{aligned}$$

in which

$$\Psi_1(u, u_1) = D_m \Psi(u) \delta l = \int_{\partial X} \chi_{\Sigma} R(u_i - u_i^*) R u_{1i} \, dx, \quad (24)$$

where Ψ was introduced in Eq. 18, δl is the parameter perturbation, $m + \delta l$, δl^c is the stiffness tensor part of parameter perturbation δl , δl^{ρ} is the density part of parameter perturbation δl , and u_1 is the first-order perturbed field with respect to m along δl .

We derive the full Hessian-vector product for the inverse boundary value problem. We have two forward problems: \tilde{u} is the weak solution to the direct problem (1)–(2) and the other generates \tilde{u}_1 , which is the solution to

$$P_{il} u_{1l} = \omega^2 (\delta l^{\rho}) \tilde{u}_l \delta_{il} + \partial_{x_j} (\delta l^c)_{ijkl} \partial_{x_k} \tilde{u}_l,$$

supplemented with the boundary condition,

$$(c_{ijkl} \partial_{x_k} u_{1l}) v_j |_{\partial X} = -[(\delta l^c)_{ijkl} \partial_{x_k} \tilde{u}_l] v_j |_{\partial X},$$

We introduce two Lagrangian multi-parameters γ and γ_1 to replace v and v_1 . Following a similar argument in Appendix B, we choose $\tilde{\gamma}$ to be the weak solution to the first adjoint boundary value problem (8)–(9), and $\tilde{\gamma}_1$ to be the weak solution to the the second adjoint boundary value problem, which is given by

$$\begin{aligned} P_{il} \gamma_{1l} &= \delta l^{\rho} \omega^2 \tilde{\gamma}_l \delta_{il} + \partial_{x_j} [(\delta l^c)_{ijkl} \partial_{x_k} \tilde{\gamma}_l], \\ (c_{ijkl} \partial_{x_k} \gamma_{1l}) v_j |_{\partial X} &= -[(\delta l^c)_{ijkl} \partial_{x_k} \tilde{\gamma}_l] v_j |_{\partial X} - \chi_{\Sigma} R \tilde{u}_{1i}. \end{aligned}$$

When summing over available boundary sources, g^s , we obtain the Hessian-vector product,

$$\begin{aligned} H \delta l(\cdot) = & \sum_s \int \left[-\omega^2 \tilde{u}_i^s \tilde{\gamma}_i^s \frac{\partial \rho}{\partial m} (\cdot) + (\partial_{x_j} \tilde{u}_i^s) (\partial_{x_k} \tilde{\gamma}_i^s) \frac{\partial c_{ijkl}}{\partial m} (\cdot) \right] dx, \\ & + \int \left[-\omega^2 \tilde{u}_i^s \tilde{\gamma}_i^s \frac{\partial \rho}{\partial m} (\cdot) + (\partial_{x_j} \tilde{u}_i^s) (\partial_{x_k} \tilde{\gamma}_i^s) \frac{\partial c_{ijkl}}{\partial m} (\cdot) \right] dx \\ & + \int (\partial_m^2 P \delta l(\cdot) \tilde{u}^s) \cdot \tilde{\gamma}^s \, dx, \end{aligned} \quad (25)$$

where the data residual information is hidden in the adjoint wavefield, $\tilde{\gamma}^s$ and $\tilde{\gamma}_1^s$; $P \delta l$ is a short-hand representation of P_{il} acting on δl .

C.2 Gauss-Newton Hessian-vector product computation

For the Gauss-Newton method, we consider the least-squares misfit and aim to compute the Gauss-Newton Hessian-vector product via the constrained minimization problem [36]. We consider a new objective function Ψ^{GN} ,

$$\begin{aligned} \min_m \Psi^{GN}(u) \quad & \text{subject to} \\ \int_X \left(-\omega^2 \rho u_i v_i + (\partial_{x_j} v_i) c_{ijkl} \partial_{x_k} u_l \right) dx \\ = \int_{\Sigma} g_i v_i \, dx, \quad & \forall v \in H^1(X), \end{aligned}$$

in which

$$\Psi^{GN}(u) = \int_{\partial X} \chi_{\Sigma} R u_i R \tilde{u}_{1i} \, dx.$$

With analogous derivation as the second-order adjoint state method, we introduce a Lagragian multiplier η and let $\tilde{\eta}$ to be the weak solution to the Gauss-Newton adjoint equation

$$P_{il} \eta_l = 0, \quad (26)$$

$$v_j (c_{ijkl} \partial_{x_k} \eta_l) |_{\partial X} = -\chi_{\Sigma} R \tilde{u}_{1i}. \quad (27)$$

We have a new adjoint equation for Gauss-Newton Hessian-vector product, which means we need to solve one more equation to retrieve a Gauss-Newton Hessian-vector multiplication.

Then, for any choice of the parameters, we have

$$H^{GN} \delta l(\cdot) = \sum_s \left\{ \int -\omega^2 \tilde{u}_i^s \tilde{\eta}_i^s \frac{\partial \rho}{\partial m}(\cdot) dx \right. \\ \left. + \int (\partial_{x_j} \tilde{u}_i^s) (\partial_{x_k} \tilde{\eta}_i^s) \frac{\partial c_{ijkl}}{\partial m}(\cdot) dx \right\}, \quad (28)$$

Note that δl is hidden in the Gauss-Newton adjoint wavefield $\tilde{\eta}$.

References

1. Akcelik, V., Biros, G., Ghattas, O.: Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation. In: ACM/IEEE 2002 Conference Supercomputing, pp. 41:1–41:15. IEEE (2002)
2. Avena, L., Castell, F., Gaudilli  re, A., M  lot, C.: Intertwining wavelets or multiresolution analysis on graphs through random forests. *Applied and Computational Harmonic Analysis* (2018)
3. Baeten, G.J.M.: Theoretical and Practical Aspects of the Vibroseis Method. PhD thesis, Technische Universiteit Delft (1989)
4. Berenger, J.-P.: A perfectly matched layer for the absorption of electromagnetic waves. *J. Comput. Phys.* **114**(2), 185–200 (1994)
5. Beretta, E., de Hoop, M.V., Faucher, F., Scherzer, O.: Inverse boundary value problem for the Helmholtz equation: quantitative conditional Lipschitz stability estimates. *SIAM J. Math. Anal.* **48**(6), 3962–3983 (2016)
6. Beretta, E., de Hoop, M.V., Francini, E., Vessella, S., Zhai, J.: Uniqueness and Lipschitz stability of an inverse boundary value problem for time-harmonic elastic waves. *Inverse Probl.* **33**(3), 035013 (2017)
7. Boonyasiriwat, C., Valasek, P., Routh, P., Cao, W., Schuster, G.T., Macy, B.: An efficient multiscale method for time-domain waveform tomography. *Geophysics* **74**(6), WCC59–WCC68 (2009)
8. Borisov, D., Smith, J., Tromp, J., Miller, R., Peterie, S., Cudney, H., Sloan, S., Moran, M.: Multi-component 3D elastic full waveform inversion using surface and body waves for detecting near surface anomalies. In: 79th EAGE Conference and Exhibition 2017 (2017)
9. Borzi, A., Schulz, V.: Computational Optimization of Systems Governed by Partial Differential Equations. SIAM (2011)
10. 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)
11. Bunks, C., Saleck, F.M., Zaleski, S., Chavent, G.: Multiscale seismic waveform inversion. *Geophysics* **60**, 1457–1473 (1995)
12. Capdeville, Y., Marigo, J.-J.: Second order homogenization of the elastic wave equation for non-periodic layered media. *Geophys. J. Int.* **170**(2), 823–838 (2007)
13. Chavent, G.: Identification of functional parameters in partial differential equations. *Joint Autom. Control Conf.* **12**, 155–156 (1974)
14. Chew, W.C., Weedon, W.H.: A 3D perfectly matched medium from modified Maxwell's equations with stretched coordinates. *Microwave Opt. Technol. Lett.* **7**(13), 599–604 (1994)
15. Choi, Y., Min, D.-J., Shin, C.: Frequency-domain elastic full waveform inversion using the new pseudo-Hessian matrix: experience of elastic Marmousi-2 synthetic data. *Bull. Seismol. Soc. Am.* **98**, 2402–2415 (2008)
16. Ciarlet, P.G., Miara, B., Thomas, J.-M.: Introduction to Numerical Linear Algebra and Optimisation. Cambridge University Press (1989)
17. Datchev, K., de Hoop, M.V.: Iterative reconstruction of the wave speed for the wave equation with bounded frequency boundary data. *Inverse Probl.* **32**(2), 025008 (2016)
18. de Hoop, M.V., Qiu, L., Scherzer, O.: An analysis of a multi-level projected steepest descent iteration for nonlinear inverse problems in Banach spaces subject to stability constraints. *Numer. Math.* **129**(1), 127–148 (2015)
19. Esser, E., Guasch, L., van Leeuwen, T., Aravkin, A.Y., Herrmann, F.J.: Total variation regularization strategies in full-waveform inversion. *SIAM J. Imag. Sci.* **11**(1), 376–406 (2018)
20. Fabri, A., Pion, S.: CGAL: The computational geometry algorithms library. In: Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 538–539. ACM (2009)
21. Fang, Q., Boas, D.A.: Tetrahedral mesh generation from volumetric binary and grayscale images. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009. ISBI'09, pp. 1142–1145. IEEE (2009)
22. Fehler, M., Kelih  r, P.J.: SEAM Phase 1: Challenges of subsalt imaging in tertiary basins, with emphasis on deepwater Gulf of Mexico. Society of Exploration Geophysicists (2011)
23. Fichtner, A., Trampert, J.: Hessian kernels of seismic data functionals based upon adjoint techniques. *Geophys. J. Int.* **185**(2), 775–798 (2011)
24. Fichtner, A., Trampert, J., Cupillard, P., Saygin, E., Taymaz, T., Capdeville, Y., Villase  r, A.: Multiscale full waveform inversion. *Geophys. J. Int.* **194**(1), 534–556 (2013)
25. Gauthier, O., Vieux, J., Tarantola, A.: Two-dimensional nonlinear inversion of seismic waveforms: Numerical results. *Geophysics* **51**, 1387–1403 (1986)
26. Hughes, T.J.: The finite element method: linear static and dynamic finite element analysis. Courier Corporation (2012)
27. Kormann, J., Rodr  guez, J.E., Ferrer, M., Farr  s, A., Guti  rrez, N., de la Puente, J., Hanzich, M., Cela, J.M.: Acceleration strategies for elastic full waveform inversion workflows in 2D and 3D. *Comput. Geosci.* **21**(1), 31–45 (2017)
28. Kothari, K., Gupta, S., de Hoop, M.V., Dokmanic, I.: Random mesh projectors for inverse problems. ICLR (2019)
29. Lailly, P.: The Seismic Inverse Problem as a Sequence of Before Stack Migrations. Society for Industrial and Applied Mathematics (1983)
30. Li, D., Gurnis, M., Stadler, G.: Towards adjoint-based inversion of time-dependent mantle convection with nonlinear viscosity. *Geophys. J. Int.* **209**(1), 86–105 (2017)
31. Lions, J.L., Mitter, S.K.: Optimal Control of Systems Governed by Partial Differential Equations, vol. 1200. Springer, Berlin (1971)
32. Liu, X., Xi, Y., Saad, Y., de Hoop, M.V.: Solving the 3D High-Frequency Helmholtz Equation Using Contour Integration and Polynomial Preconditioning. arXiv:1811.12378 (2018)
33. Liu, X., Xia, J., de Hoop, M.V.: Parallel randomized and matrix-free direct solvers for large structured dense linear systems. *SIAM J. Sci. Comput.* **38**(5), S508–S538 (2016)
34. Loris, I., Nolet, G., Daubechies, I., Dahlen, F.: Tomographic inversion using l_1 -norm regularization of wavelet coefficients. *Geophys. J. Int.* **170**(1), 359–370 (2007)
35. Marfurt, K.J.: Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations. *Geophysics* **49**(5), 533–549 (1984)
36. M  tivier, L., Brossier, R., Virieux, J., Operto, S.: Full waveform inversion and the truncated Newton method. *SIAM J. Sci. Comput.* **35**(2), B401–B437 (2013)
37. Mora, P.: Nonlinear two-dimensional elastic inversion of multioffset seismic data. *Geophysics* **52**(9), 1211–1228 (1987)
38. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer (2006)

39. Nuber, A., Manukyan, E., Maurer, H.: Ground topography effects on near-surface elastic full waveform inversion. *Geophys. J. Int.* **207**(1), 67–71 (2016)

40. Pan, G., Liang, L., Habashy, T.M.: A numerical study of 3D frequency-domain elastic full-waveform inversion. *Geophysics* **84**(1), R113–R122 (2018)

41. Plessix, R., Milcik, P., Rynja, H., Stopin, A., Matson, K., Abri, S.: Multiparameter full-waveform inversion: marine and land examples. *Lead. Edge* **32**(9), 1030–1038 (2013)

42. Plessix, R.-E.: A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophys. J. Int.* **167**(2), 495–503 (2006)

43. Pratt, R.G.: Inverse theory applied to multi-source cross-hole tomography. Part 2: Elastic wave-equation method. *Geophys. Prospect.* **38**(3), 311–329 (1990)

44. Pratt, R.G., Shin, C., Hicks, G.J.: Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophys. J. Int.* **133**, 341–362 (1998)

45. Pratt, R.G., Worthington, M.H.: Inverse theory applied to multi-source cross-hole tomography. Part 1: Acoustic wave-equation method. *Geophys. Prospect.* **38**, 287–310 (1990)

46. Prieux, V., Brossier, R., Operto, S., Virieux, J.: Multiparameter full waveform inversion of multicomponent OBC data from valhall. Part 2: Imaging compressional and shearwave velocities. *Geophys. J. Int.* **194**, 3 (2013)

47. Raknes, E.B., Weibull, W., Arntsen, B.: Seismic imaging of the carbon dioxide gas cloud at Sleipner using 3D elastic time-lapse full waveform inversion. *Int. J. Greenhouse Gas Control* **42**, 26–45 (2015)

48. Regone, C., Stefani, J., Wang, P., Gerea, C., Gonzalez, G., Oristaglio, M.: Geologic model building in SEAM Phase II—Land seismic challenges. *Lead. Edge* **36**(9), 738–749 (2017)

49. Shen, X., Ahmed, I., Brenders, A., Dellinger, J., Etgen, J., Michell, S.: Full-waveform inversion: the next leap forward in subsalt imaging. *Leading Edge* **37**(1), 67b1–67b6 (2018)

50. Shi, J., de Hoop, M., Faucher, F., Calandra, H.: Elastic full-waveform inversion with surface and body waves. In: SEG Technical Program Expanded Abstracts 2016, pp. 1120–1124. Society of Exploration Geophysicists (2016)

51. Shi, J., Li, R., Xi, Y., Saad, Y., de Hoop, M.V.: Computing planetary interior normal modes with a highly parallel polynomial filtering eigensolver. In: Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis, SC'18, Dallas, TX, November 11–16, 2018, pp. 71:1–71:13. ACM/IEEE (2018)

52. Shi, J., Ye, R., de Hoop, M.: Full-waveform inversion with 3D-shape optimization on unstructured meshes. In: SEG Technical Program Expanded Abstracts 2017, pp. 1528–1532. Society of Exploration Geophysicists (2017)

53. Shin, J., Kim, Y., Shin, C., Calandra, H.: Laplace-domain full waveform inversion using irregular finite elements for complex foothill environments. *J. Appl. Geophys.* **96**, 67–76 (2013)

54. Si, H.: Tetgen: A Quality Tetrahedral Mesh Generator and Three-Dimensional Delaunay Triangulator. Weierstrass Institute for Applied Analysis and Stochastic, Berlin (2006)

55. Sirgue, L., Pratt, R.G.: Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics* **69**, 231–248 (2004)

56. Stopin, A., Plessix, R.-É., Al Abri, S.: Multiparameter waveform inversion of a large wide-azimuth low-frequency land data set in Oman. *Geophysics* **79**(3), WA69–WA77 (2014)

57. Symes, W.W.: Approximate linearized inversion by optimal scaling of prestack depth migration. *Geophysics* **73**(2), R23–R35 (2008)

58. Tarantola, A.: Inversion of seismic reflection data in the acoustic approximation. *Geophysics* **49**, 1259–1266 (1984)

59. Tarantola, A.: Inversion of travel times and seismic waveforms. In: *Seismic Tomography*, pp. 135–157. Springer (1987)

60. Tarantola, A.: Theoretical background for the inversion of seismic waveforms including elasticity and attenuation. *Pure Appl. Geophys.* **128**(1–2), 365–399 (1988)

61. Tromp, J., Tape, C., Liu, Q.: Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophys. J. Int.* **160**, 195–216 (2005)

62. Xin, Z., Xia, J., de Hoop, M.V., Cauley, S., Balakrishnan, V.: A distributed-memory randomized structured multifrontal method for sparse direct solutions. *SIAM J. Sci. Comput.* **39**(4), C292–C318 (2017)

63. Ye, R., de Hoop, M.V., Petrovitch, C.L., Pyrak-Nolte, L.J., Wilcox, L.C.: A discontinuous Galerkin method with a modified penalty flux for the propagation and scattering of acousto-elastic waves. *Geophys. J. Int.* **205**(2), 1267–1289 (2016)

64. Yuan, Y.O., Simons, F.J., Bozdağ, E.: Multiscale adjoint waveform tomography for surface and body waves. *Geophysics* **80**(5), R281–R302 (2015)

65. Zheng, Y., Huang, X.: Anisotropic perfectly matched layers for elastic waves in cartesian and curvilinear coordinates. Technical report, Massachusetts Institute of Technology Earth Resources Laboratory (2002)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.