Karlsruhe Institute of Technology

wave phenomena
analysis and numerics

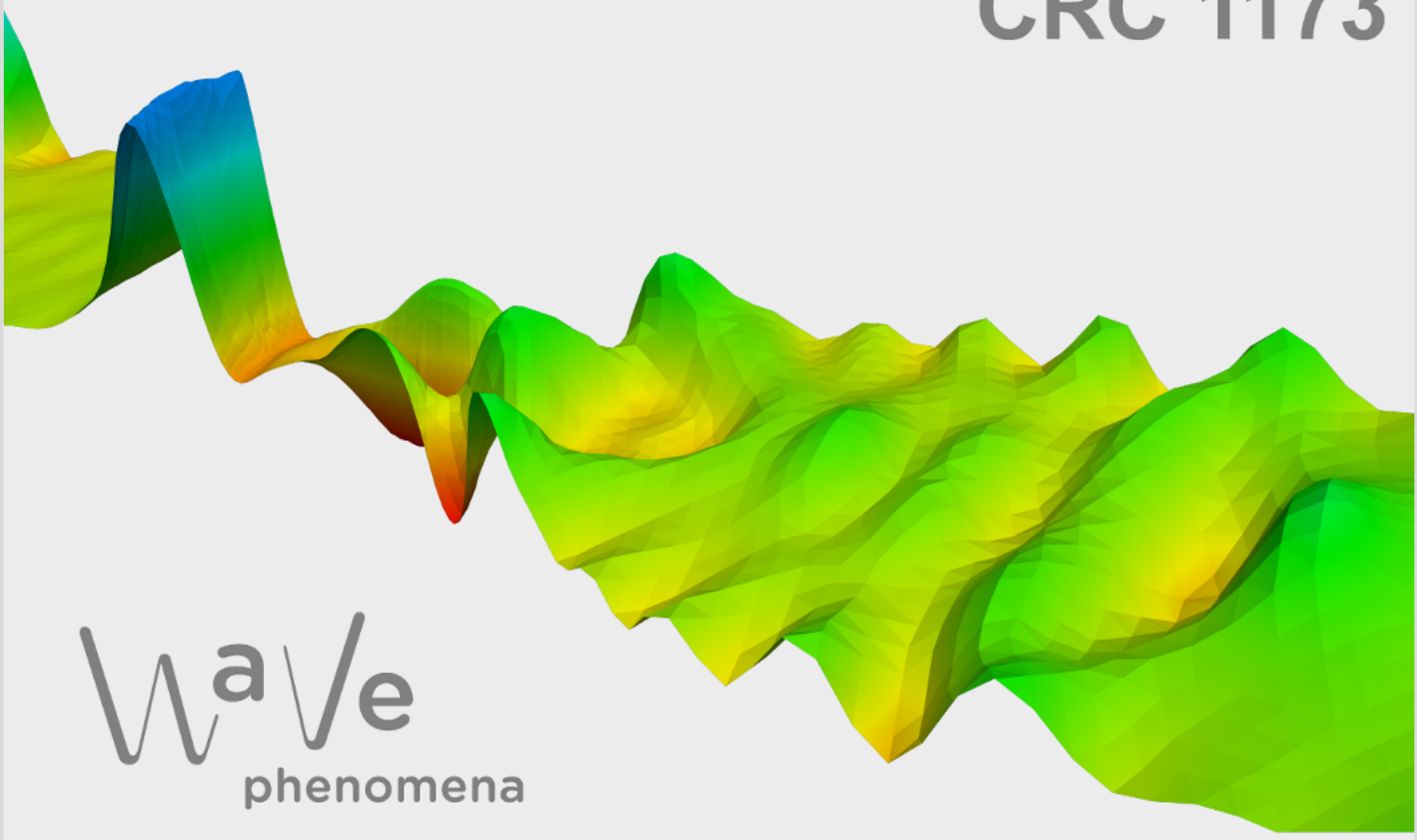
An all-at-once solver for visco-acoustic full waveform inversion in the time-domain

Christian Rheinbay, Andreas Rieder

CRC Preprint 2025/31, July 2025

KARLSRUHE INSTITUTE OF TECHNOLOGY

CRC 1173



Participating universities

Universität Stuttgart

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Funded by

ISSN 2365-662X

AN ALL-AT-ONCE SOLVER FOR VISCO-ACOUSTIC FULL WAVEFORM INVERSION IN THE TIME-DOMAIN

CHRISTIAN RHEINBAY AND ANDREAS RIEDER

Department of Mathematics, Karlsruhe Institute of Technology (KIT),
D-76128 Karlsruhe, Germany

Abstract. Full waveform inversion (FWI) is the cutting-edge seismic imaging technique used to reconstruct subsurface material properties such as wave velocities, mass density, and attenuation from measurements of reflected wavefields. In this paper, we present the first implementation of FWI in the visco-acoustic regime, based on a new time-domain all-at-once (AAO) formulation of the seismic inverse problem [Math. Meth. Appl. Sci. 2021; 44: 6376-6388]. The AAO approach solves for the material properties and the wavefield simultaneously. This offers advantages over classical methods, particularly in mitigating cycle-skipping by reducing dependency on initial models. Our algorithm combines an inexact Newton-type method, a problem-specific line search, and preconditioning to address the ill-posed nature of seismic imaging. Numerical experiments demonstrate the algorithm's robustness against noise and its increased domain of convergence compared to classical methods. Our study suggests combining the AAO algorithm with classical FWI algorithms to leverage the former's larger domain of convergence and the latter's higher accuracy when it converges. In conclusion, our AAO-FWI algorithm is a viable tool for seismic imaging, offering flexibility and robustness, with potential for further advancements and extensions to practical field data applications.

1. Introduction

Full waveform inversion (FWI) is a technique for obtaining information about the Earth's interior material, such as shear or compressional wave velocity, mass density, and attenuation, from surface or ocean seismic wave measurements. The objective is to identify material values that, when plugged into the underlying wave propagation model, yield a simulated wavefield that can account for the full information content of the seismic measurements (called data or seismograms).

FWI has become the state of the art in seismic imaging due to its increased resolution compared to techniques that rely only on amplitude or phase information [40]. The classical formulation, based on ideas of Tarantola [36], can be expressed as a constrained

E-mail address: christian.rheinbay@posteo.de, andreas.rieder@kit.edu.

Date: July 4, 2025.

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project-ID 258734477 - SFB 1173.

optimization problem for the searched material v:

$$\text{find } (u,v) \text{ such that } J(u,v) = C(u) + \mathcal{R}(v) \to \min \text{ subject to } \mathrm{L}(v)u = f, \tag{1.1}$$

where the objective functional J consists of two terms C and R. The former quantifies the discrepancy between the recorded and simulated data, whereas the latter is a regularization term to stabilize the minimization process. Furthermore, the constraint ensures that the simulated wavefield u is a solution to the underlying partial differential equation that models wave propagation based on v, that is, $\mathrm{L}(v)$ is the differential operator and the source f initiates the wavefield.

While C often represents an appropriate L^2 distance (as proposed by Tarantola and which we will call the *classical* formulation), numerous other options have been considered in the literature, such as the L^1 distance [11], total variation regularization [15], or distances based on optimal transport [28, 14]. Tarantola's ansatz (1.1) has been widely used in the past and continues to be improved to more efficiently reconstruct multiple correlated parameters at once [30] or to take advantage of modern GPU architectures to speed up simulations [44].

A common challenge with the classical formulation, especially for the inversion of pressure wave velocity, is the phenomenon called cycle-skipping [40]: If the initial and recorded data are viewed as sums over a finite set of frequencies (due to bandwidth limitations), and the phase differences for these individual frequencies exceed half a period, the optimization procedure will not converge to the global minimum. This is because gradient-based optimization schemes converge to a local minimum that belongs to a seismogram phase-shifted by one period, that is, one cycle is skipped. Therefore, the classical formulation requires a good initial model.

In recent years, new methods have emerged that relax the equality constraint in (1.1), allowing the wavefield not to be an exact solution of the underlying wave equation. In contrast to (1.1), we extend the search space and consider, for a $\lambda > 0$,

$$\text{find } (u,v) \text{ such that } \widetilde{J}(u,v) = C(u) + \lambda \| \mathrm{L}(v)u - f\|_2^2 + \mathcal{R}(u,v) \to \min\,. \tag{1.2}$$

As the penalty factor λ increases, the above formulation places more emphasis on the adhesion of the wavefield to the wave propagation model. In the limit $\lambda \to \infty$ we even regain the classical formulation (1.1). The advantage of (1.2) over (1.1), as reported in the literature, see, e.g., [38, 33], is that the objective functional $\widetilde{J}$ of (1.2) is less nonlinear and is expected to exhibit fewer local minima than J of the classical formulation.

FWI in the form of (1.2) can be interpreted as an all-at-once (AAO) approach [20], since we solve for both the material and the wavefield simultaneously. Some mathematical results for abstract AAO formulations regarding regularization methods, such as Landweber or Tikhonov, have been shown in [22]. The AAO formulation for visco-elastic FWI in the time domain and some analytical properties of the associated operator (ill-posedness, Fréchet differentiability) were established in [33]. We follow the framework of this paper and adopt the name all-at-once when talking about the FWI formulation proposed therein.

The first publications [1, 38, 39] that combined the AAO idea with FWI considered the acoustic wave equation in the frequency domain. The resulting methods are known as

Wavefield Reconstruction Inversion (WRI), for which problem-adapted Newton methods that require the Hessian matrix or gradient-based optimization schemes are employed to minimize $\widetilde{J}$. It has been reported that the domain of convergence of these methods when applied to (1.2) is larger than when applied to (1.1).

More recently, still in the frequency domain, the alternating-direction method of multipliers (ADMM) [10] was utilized and the resulting algorithm was called Iteratively-Refined Wavefield Reconstruction Inversion (IR-WRI) [2, 4]. This algorithm uses the augmented Lagrangian function to solve the optimization problem (1.2) by splitting it into subproblems and updating the primal variable, v, and the dual variable, u, individually and sequentially. Unlike penalty methods, IR-WRI does not require manual tuning of penalty parameters. This ansatz was then extended to time domain [17, 18], which substantially increases the size of the wavefield variable. Therefore the algorithm was adapted accordingly: First, the subproblem involving the augmented wave equation cannot be solved exactly for u as it can in the frequency domain. Thus, strategies were developed to iteratively approximate the solution to the augmented wave equation. Second, to avoid storing full wavefields, related optimization tasks in seismogram space were established, and algorithms adapted to this concept were developed. In [5], a proximal Newton method is used in conjunction with the ADMM method.

Thus far, we have only discussed publications that consider the acoustic wave equation. However, IR-WRI has been extended to the elastic wave equation in [6]. Another ansatz related to the AAO formulation is the Extended Source Inversion (ESI) [21, 35], which was recently transferred to the time domain [19]. This approach adds the right-hand side f of the forward equation to the minimization (1.2), replacing u. The resulting objective function incorporates a regularization term $\|Bf\|^2$, where B is a pseudo-differential operator called annihilator. See [45] for a comprehensive comparison of WRI and ESI. For an extensive list of publications, developed methods, and related techniques for the AAO concept in geophysics we refer to [29].

Note that, although it is often argued that these all-at-once-like algorithms can mitigate the phenomenon of cycle-skipping, it has been shown in [45] that WRI and ESI can be interpreted as weighted L^2 minimization of the reduced problem, meaning that both methods will exhibit local minima. For publications demonstrating the occurrence of cycle-skipping in WRI see [37, 34]. Therefore, it is not expected that any algorithm will solve the cycle-skipping challenge. However, mitigating the problem to a certain degree can enable inversions that previously yielded no results due to an initial model that was outside the domain of convergence.

In this work we propose an algorithm of inexact Newton-type for the time-domain AAO formulation of [33] combined with a problem-specific line search method and a problem-adapted preconditioner. As our implementation relies on a similar minimization formulation as the aforementioned publications, we encounter similar issues, such as high dimensionality. Nevertheless, we consider our operator-based approach to be a strong competitor to the purely optimization-based methods.

The paper is organized as follows: We start the next section by introducing the visco-acoustic wave equation, where we will model attenuation using the theory of General

Standard Linear Solids. Then, we first recall and discuss the corresponding classical (reduced) approach to seismic imaging before deriving and explaining our time-domain AAO formulation. Section 3 is devoted to discretization and implementation issues: We solve the visco-acoustic wave equation using discontinuous finite elements in the spatial domain and the implicit trapezoidal rule for time integration. Based on this discretization, we solve the AAO inverse problem using our inexact Newton-type regularization `PMSD-REGINN` in which a preconditioned multistep deepest descent method serves as the inner iteration. We provide a detailed framework for implementing the AAO solver, emphasizing on computational efficiency and robustness against the challenges of seismic inversion. In Section 4, we present results of numerical experiments to evaluate the performance of `PMSD-REGINN` in two test scenarios for visco-acoustic FWI. One scenario uses the well-known Camembert model, which features a circular inclusion with varying pressure wave velocity embedded in a homogeneous background. It is designed to test the algorithm's ability to handle highly nonlinear problems that are prone to cycle-skipping. The results show that `PMSD-REGINN` can reconstruct a higher velocity contrast than the classical reduced method can, thus demonstrating a larger domain of convergence. While the classical method yields lower errors when it converges, combining `PMSD-REGINN` with the classical method improves the results even more. Starting the classical method with the output of `PMSD-REGINN` significantly reduces the final error. The Marmousi benchmark model, which has a complex, layered geometry, serves as the second scenario. Three initial velocity models are tested, and Gaussian noise is added to the simulated seismograms to mimic realistic conditions. While shallow layers are reconstructed well across all initial models, the reconstruction quality of deeper layers is positively correlated with the quality of the initial model. Notably, the classical formulation does not converge for any of the three initial models. In the concluding Section 5, we discuss the advantages and disadvantages of `PMSD-REGINN`, pointing out areas for improvement.

2. All-at-Once Approach in Time Domain

2.1. **Visco-acoustic wave equation.** Wave propagation in isotropic materials can be modeled by the visco-elastic wave equation. It contains both pressure (longitudinal) and shear (transversal) waves and further takes dissipation into account which results in attenuation of amplitudes and other dispersive effects. A simplification which is sufficient for our study only containing the pressure waves is the visco-acoustic wave equation. We consider the formulation introduced in [46] which uses the General Standard Linear Solid (GSLS) ansatz to model the attenuation. Let $\Omega \subset \mathbb{R}^d$, $d \in \{2,3\}$, be a Lipschitz domain and ℓ be the number of GSLS damping mechanisms [26]. Then, the evolution of the particle velocity field $\boldsymbol{w}\colon [0,T]\times\Omega \to \mathbb{R}^d$, the elastic part of the pressure $p_0\colon [0,T]\times\Omega \to \mathbb{R}$, and the viscous parts of the pressure $p_l\colon [0,T]\times\Omega \to \mathbb{R}$, $l = 1,\dots,\ell$, is described by

$$
\begin{aligned}
\varrho\, \partial_t \boldsymbol{w} &= \nabla(p_0 + \cdots + p_\ell) + f && \text{in } (0,T)\times\Omega\,,\\
\frac{1+\alpha\tau}{\varrho v^2}\, \partial_t p_0 &= \operatorname{div} \boldsymbol{w} + g_0 && \text{in } (0,T)\times\Omega\,,\\
\frac{1+\alpha\tau}{\varrho v^2 \tau}\, \partial_t p_l &= \operatorname{div} \boldsymbol{w} - \omega_l \frac{1+\alpha\tau}{\varrho v^2 \tau}\, p_l + g_l \qquad l = 1,\dots,\ell && \text{in } (0,T)\times\Omega\,.
\end{aligned}
\tag{2.1}
$$

Further, we assume zero initial conditions $\boldsymbol{w}(0) = 0$, $p_l(0) = 0$, $l = 0, \dots, \ell$. The wave is often excited by a pressure source $g_0 \colon [0,T] \times \Omega \to \mathbb{R}$. Furthermore, we assume that the sources have a product structure, for example, $g_0(t,x) = \psi(t)\phi(x)$ with $\psi \colon [0,T] \to \mathbb{R}$ and $\phi \colon \mathbb{R}^d \to \mathbb{R}$. When we refer to the source frequency ω_c, we are referring to the frequency at which the modulus of the Fourier transform of ψ attains its maximum value.

However, it is also possible to consider source terms in $f \colon [0,T] \times \Omega \to \mathbb{R}^d$ and $g_l \colon [0,T] \times \Omega \to \mathbb{R}$. Equation (2.1) contains three material parameters, pressure wave velocity $v \colon \Omega \to \mathbb{R}$, mass density $\varrho \colon \Omega \to \mathbb{R}$, and a scaling parameter related to attenuation $\tau \colon \Omega \to \mathbb{R}$. In this work we only try to reconstruct $v \in P_{\text{adm}} \subset L^\infty(\Omega)$, where P_{adm} contains functions that are bounded from above and below by physically reasonable values. We assume ϱ and τ to be known and also bounded from above and below.

In order to model the attenuation, certain model parameters have to be determined a priori. First, the number of GSLS parameters ℓ must be chosen. While higher values can more accurately model the effects, they also increase the numerical effort. Since the dispersive effects are inherently frequency-dependent, it is necessary to select values $\omega_1, \dots, \omega_\ell$ in the vicinity of the source frequency ω_c to achieve a satisfying result in the frequency band surrounding the source. Finally, the dissipation coefficient τ must be selected to reflect the level of attenuation. The selection of these values is discussed in detail in [7, 8]. With these values chosen, we can calculate the last remaining quantity of (2.1):

$$\alpha = \sum_{l=1}^{\ell} \frac{\omega_c^2 \omega_l^{-2}}{1 + \omega_c^2 \omega_l^{-2}} \,. \tag{2.2}$$

Set

$$H = L^2(\Omega, \mathbb{R}^{d+1+\ell}) \quad \text{and} \quad Y = L^2([0,T], H).$$

For $v \in P_{\text{adm}}$ and $u = (\boldsymbol{w}, p_0, \dots, p_\ell) \in Y$ we define the operators

$$\begin{aligned} \mathrm{M}(v)u &= (\varrho\, \boldsymbol{w}, (1+\alpha\tau)/(\varrho v^2)\, p_0, (1+\alpha\tau)/(\varrho v^2 \tau)\, p_1, \dots, (1+\alpha\tau)/(\varrho v^2 \tau)\, p_\ell)\,, \\ \mathrm{D}u &= (0, \dots, 0, 0, \omega_1 p_1, \dots, \omega_\ell p_\ell)\,, \\ Au &= -(\nabla(p_0 + \cdots + p_\ell), \operatorname{div} \boldsymbol{w}, \dots, \operatorname{div} \boldsymbol{w})\,, \quad u \in \mathcal{D}(A), \end{aligned}$$

where $\mathrm{M}(v) \colon H \to H$ is a symmetric positive definite operator, $\mathrm{D} \colon H \to H$ is a symmetric semi-positive definite operator and $A \colon \mathcal{D}(A) \subset H \to H$ is a skew-symmetric differential operator (boundary values are incorporated into the domain of definition $\mathcal{D}(A)$). With this notation in place we can formulate the initial value problem corresponding to equation (2.1) as

$$\mathrm{L}(v)u := \mathrm{M}(v)\partial_t u(t) + (A + \mathrm{M}(v)\mathrm{D})u(t) = f(t), \quad t \in [0,T], \quad u(0) = 0\,. \tag{2.3}$$

This evolution equation is uniquely solvable for reasonable requirements on the right-hand side and yields a classical solution $u \in C^1([0,T], H)$, see, e.g., [9] for all details.

2.2. **Classical formulation of FWI.** Although an overview of the various formulations for FWI was provided in the introduction, this section will offer a more detailed explanation in the context of our case study. While the end goal of seismic imaging is to recover a

good fit to the material distribution, the additional objective of FWI is to find a wavefield that not only complies with the underlying physics but also aligns with the data at the receiver locations. Thus, for a given right-hand side $f \in Y$ and observed seismograms $s_{\text{obs}} \in S$, the goal is to find u and v such that the *model error* $\|\,\mathrm{L}(v)u - f\|^2_{[0,T]\times\Omega}$ is small for a given norm. Moreover, the same wavefield should yield a good data fit, that is, the *data error* $\|\Psi u - s_{\text{obs}}\|^2_S$ should be small. Here $\Psi\colon Y \to S \cong \mathbb{R}^N$ is the linear operator that models the measurement procedure, where N is the dimension of the measurements, that is, number of receivers multiplied by the number of data points in each time series. The classical formulation of FWI only reduces the problem to the minimization of the data error by forcing $\mathrm{L}(v)u = f$. Consequently, the data error provides the sole criterion for reconstructing the material in the classical formulation. Typically, the functional $J_{\mathbf{R}}(v) = \|\Psi\,\mathrm{L}(v)^{-1}f - s_{\text{obs}}\|_S$ is riddled with local minima, which can cause gradient-like methods to become trapped at a local minima. Furthermore, although less critical, the differential operator L will never include all physical effects influencing wave propagation in the real world, so not forcing $J_{\mathbf{R}}$ to become zero acknowledges this limitation. Then, we can define the forward map as

$$\Phi_{\mathbf{R}}\colon P_{\text{adm}} \subset L^\infty(\Omega) \mapsto S, \quad v \mapsto \Psi\,\mathrm{L}(v)^{-1}f.$$

The inverse problem is then given by

$$\text{find } v \in P_{\text{adm}} \text{ such that } \Phi_{\mathbf{R}}(v) = s_{\text{obs}}. \tag{RED}$$

The map $\Phi_{\mathbf{R}}$ is Fréchet differentiable at any interior point v of P_{adm} with derivative

$$\Phi'_{\mathbf{R}}(v)\colon L^\infty(\Omega) \to S\,, \quad h \mapsto -\Psi\,\mathrm{L}(v)^{-1}(\mathrm{L}'(v)[h]\Phi_{\mathbf{R}}(v)), \tag{2.4}$$

where

$$\mathrm{L}'(v)[\cdot]u\colon L^\infty(\Omega) \to Y, \quad h \mapsto \mathrm{M}'(v)[h](\partial_t + \mathrm{D})u,$$

and

$$\mathrm{M}'(v)[\cdot]u\colon L^\infty(\Omega) \mapsto Y, \quad \mathrm{M}'(v)[h]u = -2h\,\frac{1+\alpha\tau}{\varrho v^3}\left(0, p_0, \tau^{-1}p_1, \dots, \tau^{-1}p_\ell\right) \tag{2.5}$$

are the Fréchet derivatives of $\mathrm{L}(\cdot)u\colon P_{\text{adm}} \subset L^\infty(\Omega) \to Y$ and $\mathrm{M}\colon P_{\text{adm}} \subset L^\infty(\Omega) \to \mathcal{L}(H)$, respectively. The adjoint of $\Phi'_{\mathbf{R}}(v)$ is given by

$$\Phi'^*_{\mathbf{R}}(v)\colon S \to L^\infty(\Omega)', \quad y \mapsto \mathrm{L}'(v)[\mathrm{L}^\star(v)^{-1}\Psi^* y]^*\Phi_{\mathbf{R}}(v)$$

with

$$\mathrm{L}^\star(v) = -\mathrm{M}(v)\partial_t u(t) + (-A + \mathrm{M}(v)\mathrm{D})u(t), \quad t \in [0,T], \quad u(T) = 0,$$

and with the adjoint of $\mathrm{L}'(v)[\cdot]u$, which is

$$\mathrm{L}'(v)[\cdot]^*u\colon Y \to L^\infty(\Omega)', \tag{2.6}$$

$$(\boldsymbol{v}, q_0, \dots, q_\ell) \mapsto \frac{-2(1+\alpha\tau)}{\varrho v^3}\left(\int_0^T q_0\partial_t p_0\,\mathrm{d}t + \sum_{l=1}^{\ell}\frac{\omega_l}{\tau}\,q_l p_l\,\mathrm{d}t\right) \in L^1(\Omega) \subset L^\infty(\Omega)',$$

where p_i's are the pressure components of the wavefield $\Phi_{\mathbf{R}}(v)$. The above formal representations of $\Phi'_{\mathbf{R}}$ and $\Phi'^*_{\mathbf{R}}$ can be rigorously derived in a functional analytic framework, see [23] and [46].

The inverse problem (RED) is locally ill-posed at any parameter point in the interior of P_{adm} in the following sense [23]:

$$\text{In any neighborhood of } v^+ \in \operatorname{int}(P_{\text{adm}}) \text{ there exists a sequence } \{v_k\} \text{ with} \quad \lim_{k\to\infty} \|\Phi_{\mathbf{R}}(v_k) - \Phi_{\mathbf{R}}(v^+)\|_S = 0 \quad \text{but} \quad v_k \not\to v^+ \text{ in } L^\infty(\Omega). \tag{2.7}$$

Solving a locally ill-posed problem is challenging since reducing the residual as much as possible does not necessarily improve the outcome. The reduced formulation (RED) of the seismic inverse problem is typically solved by gradient, Newton-type methods, and Quasi-Newton methods.

2.3. **All-at-once formulation of FWI.** To formulate the time domain AAO version of the seismic inverse problem we rely on the framework of [33] which requires a weaker solution concept for the wave equation (2.3). To this end we extend Y to a larger Hilbert space $\widetilde{Y}$ and also extend the differential operator A to $\widetilde{A}$ with a larger domain of definition in $\widetilde{Y}$. This extension is necessary for a sound mathematical framework [33, Remark 3.2], but nothing is lost in terms of understanding the algorithm if one considers the strong formulation introduced at the beginning of this section. The weaker version of (2.3) reads

$$\mathrm{M}(v)u(t) + (\widetilde{A} + \mathrm{DM}(v))Ju(t) = Jf(t), \qquad t \in [0,T], \tag{2.8}$$

where $Jz(t) := \int_0^t z(s)\,\mathrm{d}s$ (note that $u(0) = 0$ is implicitly imposed). With the differential operator

$$\mathfrak{L}(v) := \mathrm{M}(v) + (\widetilde{A} + \mathrm{DM}(v))J,$$

we define the AAO forward operator

$$\Phi_{\mathbf{A}} \colon P_{\text{adm}} \times \widetilde{Y} \subset L^\infty(\Omega) \times \widetilde{Y} \to S \times \widetilde{Y}, \quad (v,u) \mapsto \big(\Psi u, \lambda\,(\mathfrak{L}(v)u - Jf)\big),$$

where $\lambda > 0$ is a factor with which we can force, during inversion, how accurately the recovered wavefield will solve (2.8), see Remark 3.1 below.

The time domain AAO version of FWI can be formulated as the following inverse problem

$$\text{find } (v,u) \in P_{\text{adm}} \times \widetilde{Y} \text{ such that } \Phi_{\mathbf{A}}(v,u) = (s_{\text{obs}}, 0), \tag{AAO}$$

which is locally ill-posed as well, that is, statement (2.7) holds accordingly for $\Phi_{\mathbf{A}}$, see [33, Proposition 4.2].

The next assertions follow from more general results in the visco-acoustic and visco-elastic regimes presented in [31, 33]: The forward operator $\Phi_{\mathbf{A}}$ is Fréchet differentiable at any interior point (v,u) of $P_{\text{adm}} \times \widetilde{Y}$ with derivative

$$\Phi'_{\mathbf{A}}(v,u) \colon L^\infty(\Omega) \times \widetilde{Y} \to S \times \widetilde{Y}, \quad \begin{pmatrix} h_v \\ h_u \end{pmatrix} \mapsto \begin{pmatrix} \Psi h_u \\ \lambda\big(\mathfrak{L}'(v)[h_v]u + \mathfrak{L}(v)h_u\big) \end{pmatrix},$$

where

$$\mathfrak{L}'(v)[\cdot]u \colon L^\infty(\Omega) \to \widetilde{Y}, \quad h \mapsto \mathrm{M}'(v)[h](I + \mathrm{D}J)u,$$

see (2.5) for the derivative of M. Its adjoint is given by

$$\Phi'^{*}_{\mathbf{A}}(v,u)\colon S\times\widetilde{Y}\to L^{\infty}(\Omega)',\quad \begin{pmatrix}s\\ y\end{pmatrix}\mapsto\begin{pmatrix}\lambda\mathfrak{L}'(v)[y]^{*}u\\ \Psi^{*}s+\lambda\mathfrak{L}^{\star}(v)y\end{pmatrix},$$

where

$$\mathfrak{L}^{\star}(v)=\mathrm{M}(v)+(-\widetilde{A}+\mathrm{DM}(v))J^{*},\qquad J^{*}z(t)=\int_{t}^{T}z(s)\,\mathrm{d}s,$$

and where $\mathfrak{L}'(v)[\cdot]^{*}u$ can be calculated analogously to (2.6) when $\Phi_{\mathbf{R}}(v)$ is replaced by u.

3. Discretization and Implementation

Before discussing the inversion algorithm, we briefly address some numerical implementation issues. We focus primarily on the performance of the algorithm for FWI tasks, and do not provide a mathematical analysis. Therefore, for all purposes, we will consider objects to be discrete without always labeling them with h, which subsequently denotes the spatial discretization step size.

3.1. Discrete spaces and time stepping.

For the space discretization of (2.8) we use the discontinuous Galerkin method as detailed in [9]. We decompose Ω into K disjoint open convex polyhedral sets $\mathcal{K}_j, j=1,\dots,K$, which we call cells. They fulfill $\Omega=\cup_{j=1}^{K}\overline{\mathcal{K}}_j$ and the have maximal diameter h. Further we define the discrete space

$$H_h=\{x_h\in L^2(\Omega,\mathbb{R}^{d+1+\ell})\colon x_h\in\mathbb{P}_p^{d+1+\ell}\text{ piecewise on each cell}\}$$

where $\mathbb{P}_p$ denotes space of d-variate polynomials with total degree $p\in\mathbb{N}_0$. The corresponding discrete operators $A_h\colon H_h\to H_h$, $\mathrm{M}_h\colon H_h\to H_h$, and $\mathrm{D}_h\colon H_h\to H_h$ are defined in [9, Section 3.1]. Finally, we introduce the discrete material space P_h which consists of functions being constant on each cell. Then, $P_{\mathrm{adm},h}:=P_h\cap P_{\mathrm{adm}}$.

For the discretization of the time interval $[0,T]$ we choose NT time points $t_n=n\mathrm{dt}$ for $n=0,\dots,\mathrm{NT}$ with $\mathrm{dt}:=T/\mathrm{NT}$, so that the finite dimensional version of $\widetilde{Y}$ is $Y_h=H_h^{\mathrm{NT}}=H_h\times\cdots\times H_h$.

Now, the discrete version of the wave equation (2.8) at the N-th time instance reads

$$\begin{aligned}\mathrm{M}_h u_N+(A_h+\mathrm{M}_h\mathrm{D}_h)\mathrm{dt}\left(0.5(u_0+u_N)+\sum_{n=1}^{N-1}u_n\right)&\\ =\mathrm{dt}\left(0.5(f_0+f_N)+\sum_{n=1}^{N-1}f_n\right),\quad N=0,\dots,\mathrm{NT},&\end{aligned}$$

where the integration operator J is approximated by the trapezoidal rule with respect to the time steps. Setting

$$T_h:=\mathrm{dt}(A_h+\mathrm{M}_h\mathrm{D}_h)$$

we get the square matrix representation of the wave operator $\mathfrak{L}$:

$$\mathfrak{L} = \begin{pmatrix} \mathrm{M}_h & 0 & \cdots & \cdots & \cdots & 0 \\ 0.5T_h & \mathrm{M}_h + 0.5T_h & \ddots & \ddots & \ddots & \vdots \\ 0.5T_h & T_h & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0.5T_h & T_h & \cdots & \cdots & T_h & \mathrm{M}_h + 0.5T_h \end{pmatrix}$$

which has the order $\dim(H_h)(\mathrm{NT}+1)$. This lower triangular matrix can be inverted by forward substitution. It involves solving one linear system with the non-diagonal operator $\mathrm{M}_h + 0.5T_h$. However, we modified the linear space-time system by Gaussian elimination such that we are able to solve the wave equation with the implicit trapezoidal Runge-Kutta time stepping scheme and subsequently integrate the solution; see [31, Section 5.2] for details.

The adjoint $\mathfrak{L}^\star$ is then approximated by

$$\mathfrak{L}^\star = \begin{pmatrix} \mathrm{M}_h & 0.5T_h^\star & \cdots & \cdots & \cdots & 0.5T_h^\star \\ 0 & \mathrm{M}_h + 0.5T_h^\star & T_h^\star & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & T_h^\star \\ 0 & \cdots & \cdots & \cdots & 0 & \mathrm{M}_h + 0.5T_h^\star \end{pmatrix}$$

where $T_h^\star = \mathrm{dt}(-A_h + \mathrm{M}_h\mathrm{D}_h)$.

3.2. `PmSD-REGINN`. The ill-posedness of (AAO) demands a regularization to ensure a stable approximation of its solution. Inexact Newton-type methods such as `REGINN` [32] are an efficient tool for regularizing this kind of nonlinear problem, as demonstrated in [9]. `REGINN`[1] methods comprise two components: an outer iteration and an inner iteration, whose iteration indices we denote by k and l, respectively. The outer iteration is basically a Newton scheme terminated by the discrepancy principle, while the inner iteration approximately solves the locally linearized problem

$$\Phi'_{\mathbf{A}}(v^k, u^k)\nu = (s_{\mathrm{obs}}, 0) - \Phi_{\mathbf{A}}(v^k, u^k)$$

at the actual outer iterate (v^k, u^k) to generate the Newton update ν^l. For ease of notation, we use the abbreviations below

$$B^k = \Phi'_{\mathbf{A}}(v^k, u^k) \quad \text{and} \quad r^k = (s_{\mathrm{obs}}, 0) - \Phi_{\mathbf{A}}(v^k, u^k).$$

In order to clarify the dependence of the operators $\mathfrak{L}, \mathfrak{L}^\star, \mathfrak{L}', \mathfrak{L}^\star$ on the material v^k in the following, we will denote the dependence with the superindex k. We avoid solving the linearized equation in a Banach space setting (which would theoretically be necessary

[1] `REG`ularization by `IN`exact `N`ewton method

due to the parameters being in L^∞) and instead iteratively decrease the corresponding least-squares functional $\mathrm{J}\colon P_h \times Y_h \to [0,\infty)$,

$$\mathrm{J}(\nu) = \frac{1}{2}\|B^k\nu - r^k\|^2_{S\times Y_h}, \tag{3.1}$$

which is possible since we are in a finite-dimensional setting. So we benefit from the simpler Hilbert space structure.

Remark 3.1. *When written out, the functional* J *is of the form* (1.2)*. In fact, we just included the penalty factor* λ *in the operator* $\Phi_{\boldsymbol{A}}$*.*

For this approach to work, we need to choose an efficient method for reducing J. Typical minimization schemes, such as conjugate gradient or standard gradient descent, rely on one-dimensional optimization problems to determine an appropriate step size. Based on our experience with the AAO formulation, these algorithms fail because useful updates require a careful balance between the material and the wavefield components. Consequently, we propose a method that is able to do this.

3.3. **Line search.** Let $\nu^l \in P_h \times Y_h$ be the current iterate in the process of decreasing the value of J from (3.1). Furthermore, let $d \in P_h \times Y_h$ be a descent direction of J at ν^l. Then, the standard optimal (one-dimensional) step size ς is

$$\varsigma = \operatorname*{argmin}_{\bar\varsigma\in\mathbb{R}_+} \mathrm{J}(\nu^l + \bar\varsigma d) = \frac{-(B^{k,*}r^k, d)_{P_h\times Y_h}}{\|B^k d\|^2_{S\times Y_h}}.$$

We switch to a multidimensional version of this ansatz. To this end, split $P_h \times Y_h$ into M orthogonal subspaces V_m, $m = 1,\dots,M$,

$$P_h \times Y_h = \bigoplus_{m=1}^{M} V_m,$$

and denote the orthogonal projection of d onto the m-th subspace by d_m, so that $d = \sum_{m=1}^M d_m$. The objective is to find a vector $a \in \mathbb{R}^M$ such that

$$a = \operatorname*{argmin}_{(\widetilde a_1,\dots,\widetilde a_M)\in\mathbb{R}^M} \mathrm{J}\Big(\nu^l + \sum_{m=1}^{M} \widetilde a_m d_m\Big).$$

Since $a = (a_1,\dots,a_M)$ satisfies $(\partial_{\widetilde a_n})\mathrm{J}(\nu^l + \sum_{m=1}^M a_m d_m) = 0$, $n = 1,\dots,M$, we find that

$$0 = \mathrm{J}'\Big(\nu^l + \sum_{m=1}^{M} a_m d_m\Big)[d_n] = \big(B^{k,*}(B^k\nu^l - r^k), d_n\big)_{P_h\times Y_h} + \sum_{m=1}^{M} a_m \big(B^{k,*}B^k d_m, d_n\big)_{P_h\times Y_h}.$$

Define the matrix $E \in \mathbb{R}^{M\times M}$ and the vector $c \in \mathbb{R}^M$, according to

$$E_{m,n} = \big(B^{k,*}B^k d_m, d_n\big)_{P_h\times Y_h} = \big(B^k d_m, B^k d_n\big)_{S\times Y_h}, \quad m,n = 1,\dots,M,$$
$$c_n = \big(B^{k,*}(B^k\nu^l - r^k), d_n\big)_{P_h\times Y_h}, \quad n = 1,\dots,M,$$

then $a \in \mathbb{R}^M$ is the solution of

$$Ea = c. \tag{3.2}$$

Assuming injectivity of B^k and $d_m \neq 0$, $m = 1, \dots, M$, E is a symmetric positive definite matrix and thus invertible

For our implementation of this line search, we decompose the parameter space P_h by splitting it geometrically according to the decomposition of Ω into individual cells, that is, we define the subspaces V_m as

$$V_m = \operatorname{span}\{\varphi_m\} \times \{0\}, \quad m = 1, \dots, K,$$

where φ_m is the indicator function of the cell $\mathcal{K}_m$. We do not split Y_h and define

$$V_{K+1} = \{0\} \times Y_h.$$

Hence, the V_m's, $m = 1, \dots, K+1$, decompose $P_h \times Y_h$ into $M = K+1$ orthogonal subspaces. Accordingly, the matrix E of (3.2) of order $K+1$ has block form:

$$E = \begin{pmatrix} D & R \\ R^\top & e \end{pmatrix}, \quad D \in \mathbb{R}^{K\times K}, \quad R \in \mathbb{R}^{K\times 1}, \quad e \in \mathbb{R}\,.$$

With $d = (d_v, d_u) \in P_h \times Y_h$, the entries of the diagonal matrix D can be calculated by

$$D_{mn} = \big(B^k[d_{v,n}, 0], B^k[d_{v,m}, 0]\big)_{P_h\times Y_h} = \delta_{mn}\lambda^2 \|\mathfrak{L}'^{,k} d_{v,m}\|^2_{\mathcal{K}_m\times[0,T]}, \quad m, n = 1, \dots K\,,$$

where the last equality holds due to these arguments: The cells $\{\mathcal{K}_m\}_m$ are pairwise disjoint, so that the supports of the $d_{v,m}$'s are pairwise disjoint as well. Additionally, $\mathfrak{L}'^{,k} d_{v,m}$ and $d_{v,m}$ share the same support, since $\mathrm{M}'^{,k}$ is a multiplication operator, see (2.5).

Further, $R \in \mathbb{R}^{K\times 1}$, which is the $(K+1)$-th column of E without the $(K+1)$-th entry, has the following components

$$R_m = \big(B^k[0, d_u], B^k[d_{v,m}, 0]\big)_{S\times Y_h} = \lambda^2 \big(\mathfrak{L}^k d_u, \mathfrak{L}'^{,k} d_{v,m}\big)_{\mathcal{K}_m\times[0,T]}, \quad m = 1, \dots, K\,.$$

Finally, the bottom diagonal entry of E can be calculated via

$$e = E_{K+1,K+1} = \|B^k[0, d_u]\|^2_{S\times Y_h} = \lambda^2 \left\|\mathfrak{L}^k d_u\right\|^2_{Y_h} + \|\Psi d_u\|^2_S.$$

To simplify the components of the left-hand side c of (3.2), we denote the residual of the current iterate ν^l of the inner iteration by

$$\widetilde{r}^l = B^k \nu^l - r^k = (\widetilde{r}^l_S, \widetilde{r}^l_u) \in S \times Y_h.$$

We obtain

$$c_m = \big((\widetilde{r}^l_S, \widetilde{r}^l_u), B^k[d_{v,m}, 0]\big)_{S\times Y_h} = \lambda^2 \big(\widetilde{r}^l_u, \mathfrak{L}'^{,k} d_{v,m}\big)_{\mathcal{K}_m\times[0,T]}, \quad m = 1, \dots, K,$$

and

$$c_{K+1} = \big((\widetilde{r}^l_S, \widetilde{r}^l_u), B^k[0, d_u]\big)_{S\times Y_h} = \big(\widetilde{r}^l_S, \Psi d_u\big)_S + \lambda^2 \big(\widetilde{r}^l_u, \mathfrak{L}^k d_u\big)_{Y_h}.$$

Introducing the Schur complement $s = e - e^{-1} R^\top R \in \mathbb{R}\backslash\{0\}$ yields first

$$a_{K+1} = s^{-1}\big(c_K - R^\top D^{-1}(c_1, \dots, c_K)^\top\big) \tag{3.3a}$$

and then

$$(a_1, \dots, a_K)^\top = D^{-1}\big((c_1, \dots, c_K)^\top - a_{K+1} R\big). \tag{3.3b}$$

The main computational cost of setting up the system (3.2) lies in applying the linear operators $\mathfrak{L}^k$ and $\mathfrak{L}'^{,k}$, as well as in computing the norm of the full waveform required

for the matrix entry e. From the above expressions it is clear that an evaluation of $\mathfrak{L}^k$ and $\mathfrak{L}'^{,k}$ is required. Due to the locality and linearity of both operators, it makes no computational difference whether we perform the total evaluation once or for each cell individually. Each total operator evaluation consists of a matrix-vector multiplication per time step, which is relatively inexpensive compared to solving the full wave system. We stress that the locality and linearity of the operators allow an efficient computation of the inner products for all R_m, $m = 1, \dots, K$, and the matrix entry e, as well as the norms needed to build the diagonal matrix D. In addition, solving the linear system (3.2) is computationally efficient due to its structurally simple and well-conditioned form.

Remark 3.2. *The framework presented above easily extends to multiple parameters and other decompositions of Ω; see* [31].

3.4. **Preconditioning.** The line search introduced in the previous subsection represents only one aspect of our algorithm as it does not yield convergence in general. So we add a preconditioner to ensure an update of the reconstructed wavefield and hence of the material in each outer iteration. To this end our preconditioner is inspired by the diagonal entry of the operator corresponding to the normal equation. This is the part that maps wavefields to wavefields:

$$(B^{k,*}B^k)_{u,u} = \lambda^2 \mathfrak{L}'^{*,k}\mathfrak{L}^k + \Psi^*\Psi.$$

The efficient inversion of this block, known as the augmented wave equation, remains an unsolved problem in time domain. Various approaches to master this challenge can be found, e.g., in [3, 17, 41]. On the other hand, we invert only the block $\mathfrak{L}'^{*,k}\mathfrak{L}^k$, which we realize by forward substitution as described above, at the additional cost of solving two wave equations. The preconditioned wavefield component is then given by

$$\triangle u^l{}_{\mathrm{PC}} = \left(\lambda^2 \mathfrak{L}'^{*,k}\mathfrak{L}^k\right)^{-1} \triangle u^l\,. \tag{3.4}$$

A similar preconditioning ansatz was used in [3], called backward-forward recursion, however, for solving the augmented wave equation.

As all operators in this section are restricted to finite dimensional spaces, that is, they can be represented by matrices, we can apply the Woodbury formula to get

$$\begin{aligned}(\lambda^2 \mathfrak{L}'^{*,k}\mathfrak{L}^k + \Psi^*\Psi)^{-1} = {} & (\lambda^2 \mathfrak{L}'^{*,k}\mathfrak{L}^k)^{-1} \\ & - (\lambda^2 \mathfrak{L}'^{*,k}\mathfrak{L}^k)^{-1}\Psi^*(I + \Psi(\lambda^2 \mathfrak{L}'^{*,k}\mathfrak{L}^k)^{-1}\Psi^*)^{-1}\Psi(\lambda^2 \mathfrak{L}'^{*,k}\mathfrak{L}^k)^{-1}\,.\end{aligned}$$

So, our preconditioner neglects the second term. Our experiments have shown that the algorithm is generally not very sensitive to the choice of λ. However, in [31] it is argued that

$$\widetilde{\lambda} = \frac{\|\Psi(\mathfrak{L}^{k,\star}\mathfrak{L}^k)^{-1}\Psi^*\Delta s\|_S}{\|(\mathfrak{L}^{k,\star})^{-1}\Psi^*\Delta s\|_S}$$

is a critical value to stay below. Here Δs is the second component of the current residual $\widetilde{r}^l = (\Delta u, \Delta s)$. The threshold $\widetilde{\lambda}$ can be calculated on the fly while running the inversion. In our experiments, we always choose either $\lambda = 1$ or $\lambda = 0.1\widetilde{\lambda}$.

Algorithm 1 Preconditioned Multistep Steepest Descent (`PMSD`) algorithm

Input: $\nu^0 \in S \times Y$ % starting guess for the inner loop
$r^k \in Y \times P_{\text{adm}}$ % residual of the outer loop
$l_{\max} \in \mathbb{N}, \vartheta > 0$ % stopping criteria

1: $l \leftarrow 0$
2: $\widetilde{r}^l \leftarrow r^k - B^k \nu^0$
3: **repeat**
4: $\quad d^l \leftarrow B^{k,*} \widetilde{r}^l$
5: $\quad d_{\text{PC}} \leftarrow \text{Precon}(d^l)$ % see (3.4)
6: $\quad a \leftarrow \text{CalculateA}(B^k, d_{\text{PC}})$ % see (3.3)
7: $\quad \nu^{l+1} \leftarrow \nu^l + \sum_{m=1}^{M} a_m d_{m,\text{PC}}$
8: $\quad \widetilde{r}^{l+1} \leftarrow \widetilde{r}^l - \sum_{m=1}^{M} a_m B^k d_{m,\text{PC}}$ % $B^k d_{m,\text{PC}}$ can be reused from line 6
9: $\quad l \leftarrow l + 1$
10: **until** $(\|\widetilde{r}^l\|_{[0,T]\times\Omega} \leq \vartheta \|b\|_{[0,T]\times\Omega}$ **or** $l \geq l_{\max})$ **and** $l \geq l_{\min}$
11: **return** $\nu^{l+1} \in P \times Y$

The combination of the line search and the preconditioner yields the `PMSD`-method (Preconditioned Multistep Steepest Descent), see Algorithm 1 for a rough algorithmic outline.

3.5. **Full algorithm.** We proceed with the description of the outer loop/iteration of our AAO algorithm. In typical applications we have more than one measurement from different, say, $\Xi \in \mathbb{N}$ sources. To adapt to this situation, we realize a Kaczmarz version of the algorithm, which computes an update of the wavefield and the material for each shot. So we additionally iterate over the sources. However, we only pass the material update over to the next shot, since storing the wavefield for each shot is not feasible due to the large memory footprint. Therefore, we discard the wavefield results from the inner loop after each shot. For the next iteration, we do not want to start with a wavefield that equals zero. Instead, we use the solution of equation (2.8), computed with the current material model, as the initial wavefield for the next shot. To terminate the outer loop we use a fixed number of $k_{\max}$ instead of a discrepancy principle. As previously mentioned, `REGINN` allows an adaptive stopping criteria for the inner loop. We use the proven rule from [9, 31, 43]: We choose ϑ^k and $l^k_{\max}$ according to

$$\vartheta^k = \min\{0.999, \widetilde{\vartheta}^k\}$$

with

$$\widetilde{\vartheta}^k = \begin{cases} 1 & \text{for } k = 0, \\ \|r_0\|_{S\times Y_h} / \|r_1\|_{S\times Y_h} & \text{for } k = 1, \\ 1 - \frac{l^{k-2}}{l^{k-1}}(1 - \vartheta^{k-1}) & \text{for } l^{k-1} > l^{k-2} \wedge k \geq 2, \\ 0.9\,\vartheta^{k-1} & \text{otherwise,} \end{cases} \tag{3.5}$$

where l^k denotes the number of passes through the repeat loop in Algorithm 1 for the k-th outer iteration. The maximum amount of repeat loop iterations is determined by

$$l^k_{\max} = \begin{cases} 1 & \text{for } k = 0, \\ 2 & \text{for } k = 1, \\ l^{k-1} + l^{k-2} & \text{otherwise.} \end{cases} \tag{3.6}$$

Additionally, a minimal amount $l_{\min}$ of inner loop passes has proven to be necessary for the convergence and can even positively influence the convergence of the algorithm. The full algorithm is shown in Algorithm 2 where the dependency of the variables on the different shots is indicated by the subscript ξ.

The minimal computational cost of a single outer loop for one shot consists of solving the wave equation once (for line 6 in Algorithm 2) and solving it twice for each inner iteration, with at least $l_{\min}$ inner iterations required. This results in a minimal number of $1 + 2l_{\min}$ solutions of the wave equation per outer loop.

Algorithm 2 `PMSD-REGINN`

Input: $v^0 \in P_{\text{adm}}$ % starting guess
$s_{\text{obs}} \in S$ % seismograms

1: $k \leftarrow 1$
2: **while** $k < k_{\max}$ **do**
3: $\quad \xi \leftarrow 1$
4: $\quad v^k_\xi \leftarrow v^k$
5: $\quad$ **repeat**
6: $\quad\quad u^k_\xi \leftarrow \mathfrak{L}^{-1} J f_\xi$ % one solution of the integrated forward wave equation (2.8).
7: $\quad\quad r^k_\xi \leftarrow (0, \Psi u^k_\xi - s_{\text{obs},\xi})$ % wave residual is zero, since $\mathfrak{L} u^k_\xi - J f_\xi = 0$.
8: $\quad\quad$ determine $l^k_{\max,\xi}$ and ϑ^k_ξ according to (3.5) and (3.6) to terminate the inner loop (line 10 of Algorithm 1)
9: $\quad\quad \triangle v^k_\xi, - \leftarrow$ `PMSD`$(0, r^k_\xi, l^k_{\max,\xi}, \vartheta^k_\xi)$
10: $\quad\quad v^k_{\xi+1} \leftarrow v^k_\xi + \triangle v^k_\xi$
11: $\quad\quad \xi \leftarrow \xi + 1$
12: $\quad$ **until** ξ equals Ξ
13: $\quad v^{k+1} \leftarrow v^k_\Xi$
14: $\quad k \leftarrow k + 1$
15: $v \leftarrow v^k$
16: **return** v

4. Numerical examples

In this section, we investigate two test scenarios to demonstrate the performance of `PMSD-REGINN` in comparison to a `REGINN` solver for the classical formulation. The code in the C++ programming language for all experiments and the scripts to generate the figures are available for download from the repository: https://gitlab.kit.edu/andreas.rieder/aao-paper.

4.1. **The Camembert model.** First, we present a model scenario in which the reduced method is unsuccessful, whereas the AAO method succeeds. This fact is underpinned by a modified version of the Camembert model, initially introduced in [16] as a case study of a highly nonlinear problem resulting in the phenomenon of cycle-skipping which requires a starting model (initial guess) close to the sought-after ground truth, as detailed in Section 1.

Here, we consider acoustic wave propagation without attenuation, that is, $\ell = 0$ and $\alpha = 0$ in (2.1). For the geometry we choose the rectangle $\Omega = (0, 4800)\text{m} \times (0, 6000)\text{m}$ with an absorbing layer around the domain to prevent reflections at the boundary from propagating back into Ω during the measurement period. The full calculation domain is $\Omega_{\text{calc}} = (-900, 6300)\text{m} \times (-9000, 7500)\text{m}$ with the homogeneous background velocity $v_{\text{bg}} = 4000\,\text{m/s}$. A circular inclusion, the Camembert, with midpoint $x_{\text{mid}} = (2400, 3000)^\top\text{m}$ and radius $r = 1600\text{m}$, is embedded into Ω equipped with a velocity that differs from the background by a constant factor $1 + 0.01p$, $p > 0$, i.e. within the Camembert the velocity is

$$v_\Delta = (1 + 0.01p)v_{\text{bg}}. \tag{4.1}$$

On the right side of the domain, we placed 160 receivers, equally spaced on the line segment from $(4706.25, 18.75)^\top$ to $(4706.5, 5981.25)^\top$. This geometric layout can be seen in Figure 1. Similar configurations for the Camembert model are used in [14] and [17]. According to [16], the strong nonlinearity arises from the large diameter of the inclusion. This makes a good initial model more important due to the local nature of gradient-based inversion schemes such as Newton-type methods.

For the discretization, as explained in Section 3.1, we decompose Ω into squares with a side length of 37.5m ($h = \sqrt{2} \cdot 37.5\text{m}$) and set the final time $T = 2.4$. Further, we use the implicit trapezoid rule as time integrator with step size $\text{dt} = 0.003$. On the left side of Ω, there are $\Xi = 14$ equally spaced sources: $x_{\text{src},i} = (100, 200 + 5600i/(\Xi - 1))^\top, i = 0, \ldots, \Xi - 1$. The applied source signal is the integrated Ricker wavelet,

$$Jf(t) = \int_0^t (1 - 2\sigma^2(s)) \exp(-\sigma^2(s))\text{d}s \quad \text{where } \sigma(s) = \pi\,\omega_c\,(s - t_{\text{shift}}),\ \omega_c,\ t_{\text{shift}} \in \mathbb{R}, \tag{4.2}$$

with the central frequency $\omega_c = 10\,\text{Hz}, t_{\text{shift}} = 1/(2\omega_c) = 0.05\,\text{s}$. Due to the AAO version (2.8) of the wave equation we need the source in integrated from.

In this study, we compare the inversion results obtained through the reduced, classical algorithm `CG-REGINN`, as presented in [9], with those obtained by `PMSD-REGINN`, our AAO implementation. For a quantitative comparison, we define the relative L^2-error as

$$e_p = \frac{\|v_{\text{true},p} - v_{\text{final},p}\|_\Omega}{\|v_{\text{true},p}\|_\Omega}$$

with $v_{\text{true},p}$ as the true value of the material distribution and $v_{\text{final},p}$ as the result of the inversion process. The goal is to find the maximum value of p beyond which the inversion leads to relative L^2-errors that exceed the initial error. To this end, we conducted inversions using the reduced method for $p \in [4.5, 8.75]$ and the AAO method for $p \in [4.5, 10.5]$.

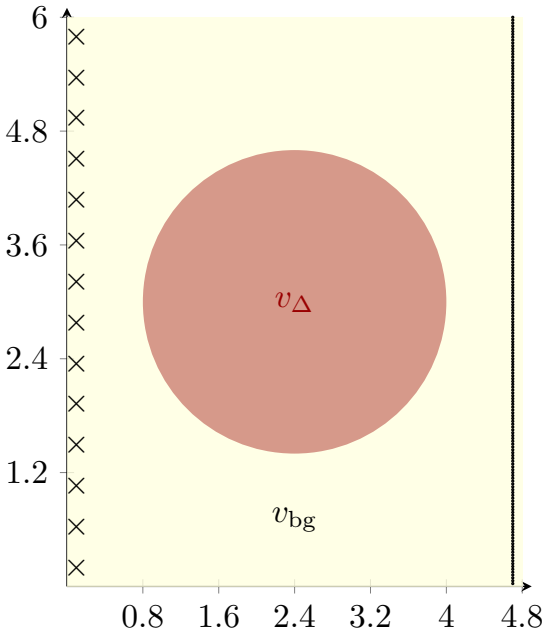


FIGURE 1. Geometry of the Camembert model: sources (marked by $\times$), inclusion (colored red), and the receivers on the line at the right.

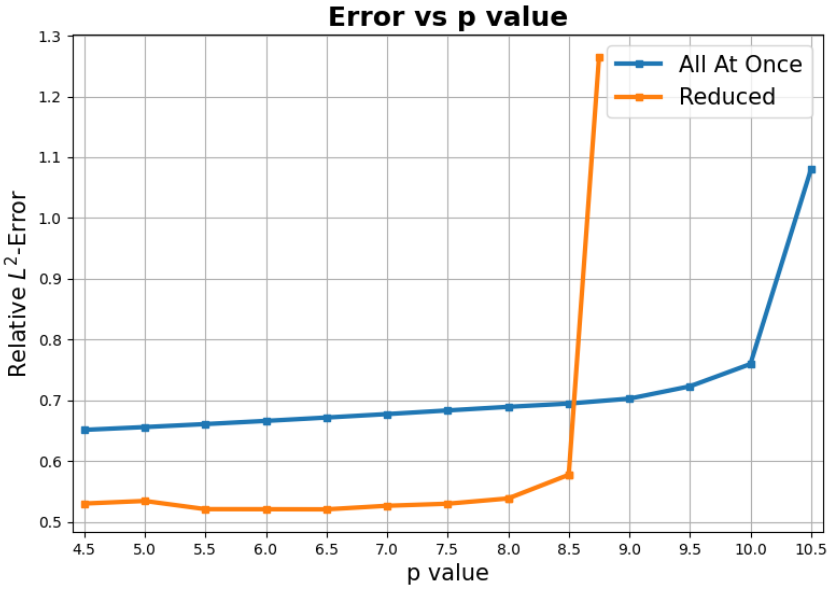


FIGURE 2. Relative L^2-error e_p after 15 iterations as a function of p, see (4.1), for both inversion algorithms.

In Figure 2, we plot the error of the final iterate of the algorithms versus the value of p (both algorithms were run for a fixed number of 15 iterations). It is evident that both algorithms are effective for values below $p = 8.5$, which is the threshold at which the

classical formulation algorithm fails. However, the AAO method is capable of reconstructing values up to $p = 10.0$, indicating a larger domain of convergence for this particular problem. Final reconstructions for a "moderate" value of $p = 7.5$ for both algorithms, as well as values just before each of the methods breaks down are shown in Figure 3.

Figure 4 displays a detailed convergence/divergence history of the two algorithms. There, we show how the relative L^2-error evolves relative to the outer iteration index for different p values. In the plot on the right for `PMSD-REGINN`, we observe that the error for $p = 10.5$ exhibits a strictly monotone decrease after reaching its peak. So it appears that convergence may still occur, and in fact, the relative L^2-error falls below 100% after 40 iterations as additional computations show. We stopped our computations with the 115th iterate having a relative L^2-error of 95.1%. Although the error decrease was strictly monotone, we consider `PMSD-REGINN` to have failed for $p = 10.5$, as a significant amount of numerical effort would be required to reach an acceptable error level, if at all.

In Figure 2 we observe that the classical formulation yields lower errors when successful. It is important to note that each iteration of the classical version is computationally cheaper. Therefore, if the solution is sufficiently close to the initial value, this formulation should be preferred. To further improve the reconstruction, we combine both algorithms: We start the classical algorithm with the final value of the AAO method to further reduce the error for values outside of the original domain of convergence for the classical formulation. The result of this combination for $p = 10$ and 20 iterations of the classical method is shown in Figure 5. The final relative error in the AAO reconstruction is 0.78, after the continuation the error dropped to 0.72, supporting the hypothesis that starting an additional reduced inversion after the AAO inversion can further improve the results.

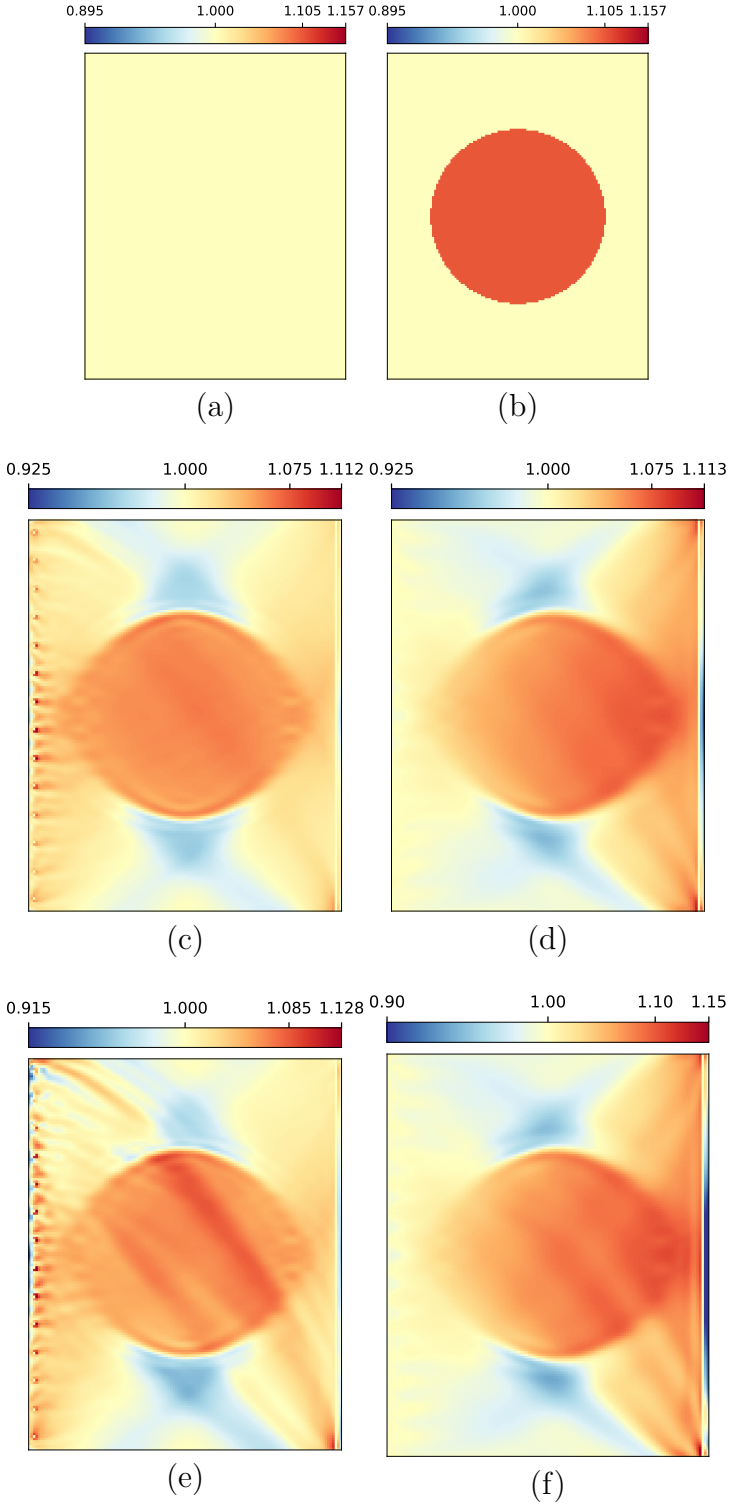
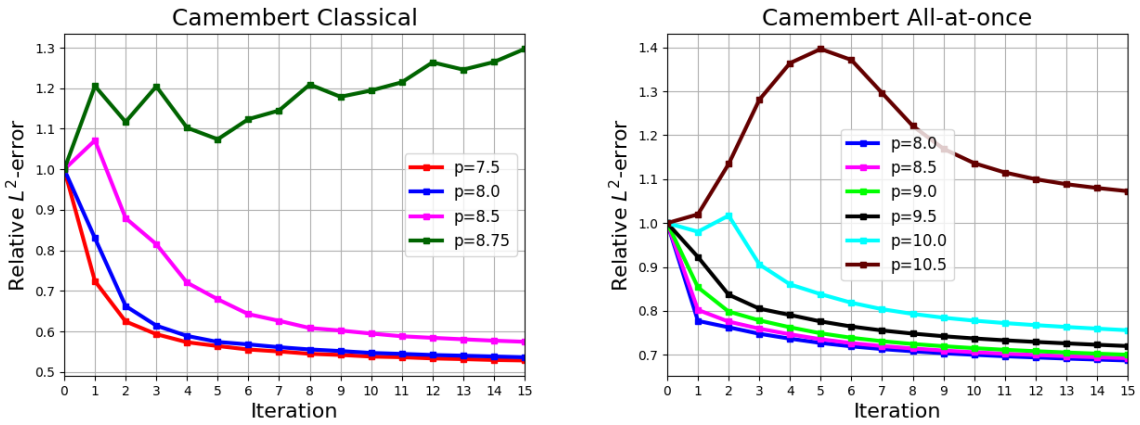
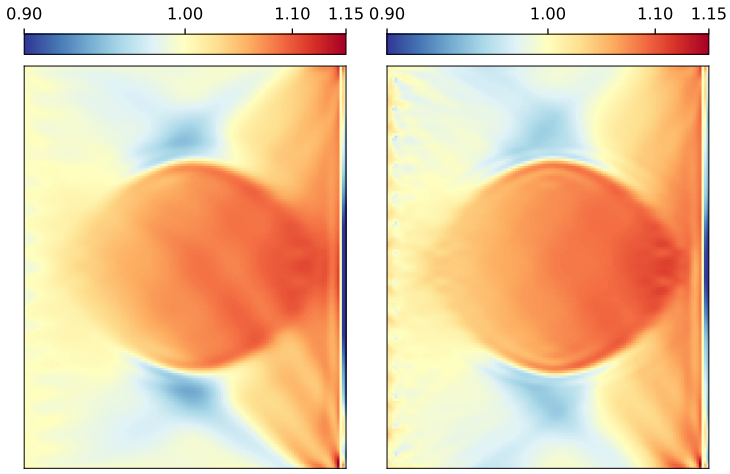


FIGURE 3. Camembert velocity reconstructions. (a) initial value, (b) ground truth. Final reconstructions for the classical method in (c) $p = 7.5$ and (e) $p = 8.5$; for the AAO method in (d) $p = 7.5$ and (f) $p = 10.0$. To improve visibility of deviations from the background velocity, the velocity values in all plots are divided by $4000\,{}^{\mathrm{m}}/_{\mathrm{s}}$.



(a) (b)

FIGURE 4. Relative L^2-error for different p plotted versus the outer iteration index. (a) reduced method; (b) AAO method `PMSD-REGINN`. Note the different scales at the vertical axes.



(a) (b)

FIGURE 5. Continued Camembert velocity reconstructions. (a) initial value (final value of the AAO reconstruction), (b) final reconstruction of the classical method after 20 iterations. The normalization of the plots is as in Figure 3.

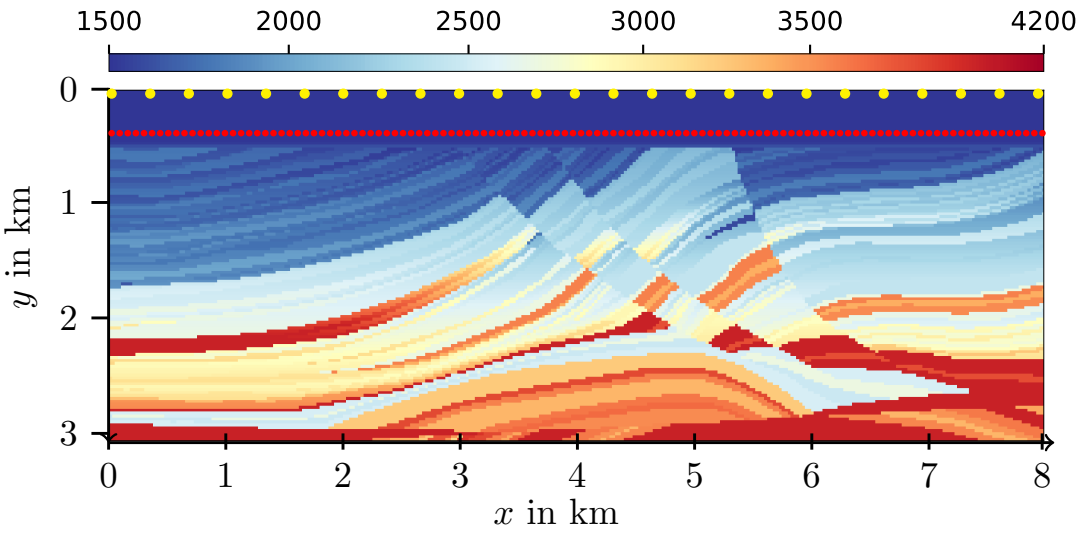


FIGURE 6. The velocity distribution v of the Marmousi model. Source positions are marked with yellow dots, receiver positions with red dots.

4.2. **Marmousi.** We evaluate the algorithm using a more complex, layered geometry derived from the pressure wave velocity distribution known as Marmousi2 (hereafter referred to as *Marmousi*) [27], a well-established model for testing inversion algorithms in geophysics, see Figure 6: At the top, a layer of water extends to a depth of 450 m. Below this layer are sediments with different wave velocities. In the water layer, the wave speed v is set to $1500\,\mathrm{m/s}$, the density ϱ is set to $1000\,\mathrm{kg/m^3}$ with the dissipation coefficient τ set to 0.0001. In the ground, the density ϱ is constant at $2000\,\mathrm{kg/m^3}$ with $\tau = 0.0274$. Moreover we choose the central frequency $\omega_c = 9\,\mathrm{Hz}$ and the time shift $t_{\mathrm{shift}} = 0.5\,\mathrm{s}$ for the Ricker wavelet (4.2). We work with $\ell = 3$ damping terms in (2.1) and with frequencies $\omega_1^{-1} = 1.0540\,\mathrm{s}$, $\omega_2^{-1} = 0.0825\,\mathrm{s}$, and $\omega_3^{-1} = 0.0084\,\mathrm{s}$ to define α in (2.2). This configuration yields a mean quality factor of approximately 62 over the relevant frequency band in the ground region, indicating that 1/62 of the wave energy dissipates per cycle.[2] In the water layer, we have a substantial Q value in the order of 10000, which approximates nearly dissipation-free propagation.

The true velocity model is scaled such that the wave speed varies between the extrema $v_{\mathrm{p,min}} = 1500\,\mathrm{m/s}$ and $v_{\mathrm{p,max}} = 4000\,\mathrm{m/s}$. Our parameter settings above for the Marmousi model match the configuration proposed in [24].

The computational domain is $\Omega = (0, 8000)\mathrm{m} \times (0, 3000)\mathrm{m}$. Along its upper boundary, we impose Dirichlet boundary conditions such that the sum of all pressure components is zero, a condition commonly known in geophysics as a *free surface boundary*. An absorbing layer is added to the remaining three boundaries, where we increase the attenuation parameter to effectively dampen the wave.

[2]The frequency dependence of wave propagation in real media is modeled by a constant Q, which is the ratio of full energy versus dissipated energy, see, e.g., [7, 13].

The computational mesh is uniformly Cartesian with a sidelength of 25m, and we use affine-linear ansatz functions. The wave equations are solved using the implicit trapezoidal rule with time step $\mathrm{dt} = 0.0056\,\mathrm{s}$ and final time $T = 5.88\,\mathrm{s}$.

We position 25 equally spaced sources along a line near the water surface, from $(100.5\,\mathrm{m}, 20.0\,\mathrm{m})^\top$ to $(7899.5\,\mathrm{m}, 20.0\,\mathrm{m})^\top$. In addition, 350 equally spaced receivers are placed along a line near the seafloor from $(0.5\,\mathrm{m}, 435.0\,\mathrm{m})^\top$ to $(7999.5\,\mathrm{m}, 435.0\,\mathrm{m})^\top$.

In the sequel, we illustrate that `PMSD-REGINN` has an extended domain of convergence, even for material distributions that are more complex than the Camembert model. To this end, we consider three initial values that fall outside the domain of convergence of the classical approach when no additional preconditioning techniques, such as frequency filtering [12], are implemented[3]: a 'blurred' model, an 'averaged' model (constructed by averaging the blurred model over each row), and a 'linear' model (which assumes a linearly increasing initial model based on depth from $1500\,\mathrm{m/s}$ to $4000\,\mathrm{m/s}$), see Figure 7. To weaken the committed inverse crime[4], we add Gaussian noise: For each trace (that is, the recorded time series of one receiver) in the simulated seismograms s^r_{obs} (r is the number of the receiver), we generate s^r_δ with entries $s^r_{\delta,i} \sim \mathcal{N}(0,1)$, $i = 0, \dots, \mathrm{NT}$. Then, we normalize it and add it to the simulated seismograms:

$$s^{r,\delta}_{\mathrm{obs}} := s^r_{\mathrm{obs}} + \frac{\|s^r_{\mathrm{obs}}\|_S}{\|s^r_\delta\|_S}\, s^r_\delta,$$

which yields a relative noise level of 100% in each trace.

We ran `PMSD-REGINN` with a minimum number $l_{\min} = 3$ of inner iterations and with $k_{\max} = 20$ outer iterations. The results presented in Figure 8 indicate that our algorithm produces reconstructions of competitive quality. However, the quality of the reconstructions deteriorates as the initial value is further away from the ground truth. Reconstruction from the blurred model gives the best quality, while the linear model gives the worst. The shallow sections have a decent quality for all initial values, but the deeper layers are not accurately reconstructed, especially for the averaged and linear initial values. Nevertheless, Figure 9 shows a consistent and strictly monotone reduction in error across all initial values.[5] The error values are normalized to the initial error of the affine linear initial guess. This normalization allows for a quantitative comparison: the final error of the reconstruction that starts with the linear model remains larger than the error of the averaged initial model, while the final error of the iteration with this starting model is still larger than the error of the blurred model. This observation suggest that the L^2-error metric alone may be misleading because all the final reconstructions in Figure 8 ostensibly capture more subsurface information than any of three initial models.

Over the past decade, the structural similarity index (SSIM) has received considerable attention from the image processing community as a measure for evaluating perceptual

[3]The code corresponding to the experiments showing divergence for the three initial models can be found in the Git repository provided at the beginning of this section.

[4]An inverse crime is committed when synthetic measurements are generated with the same forward solver that will later be used for inversion.

[5]The observed, strictly monotone L^2-error decrease complies with the theoretical prediction in [25, Theorem 3.1].

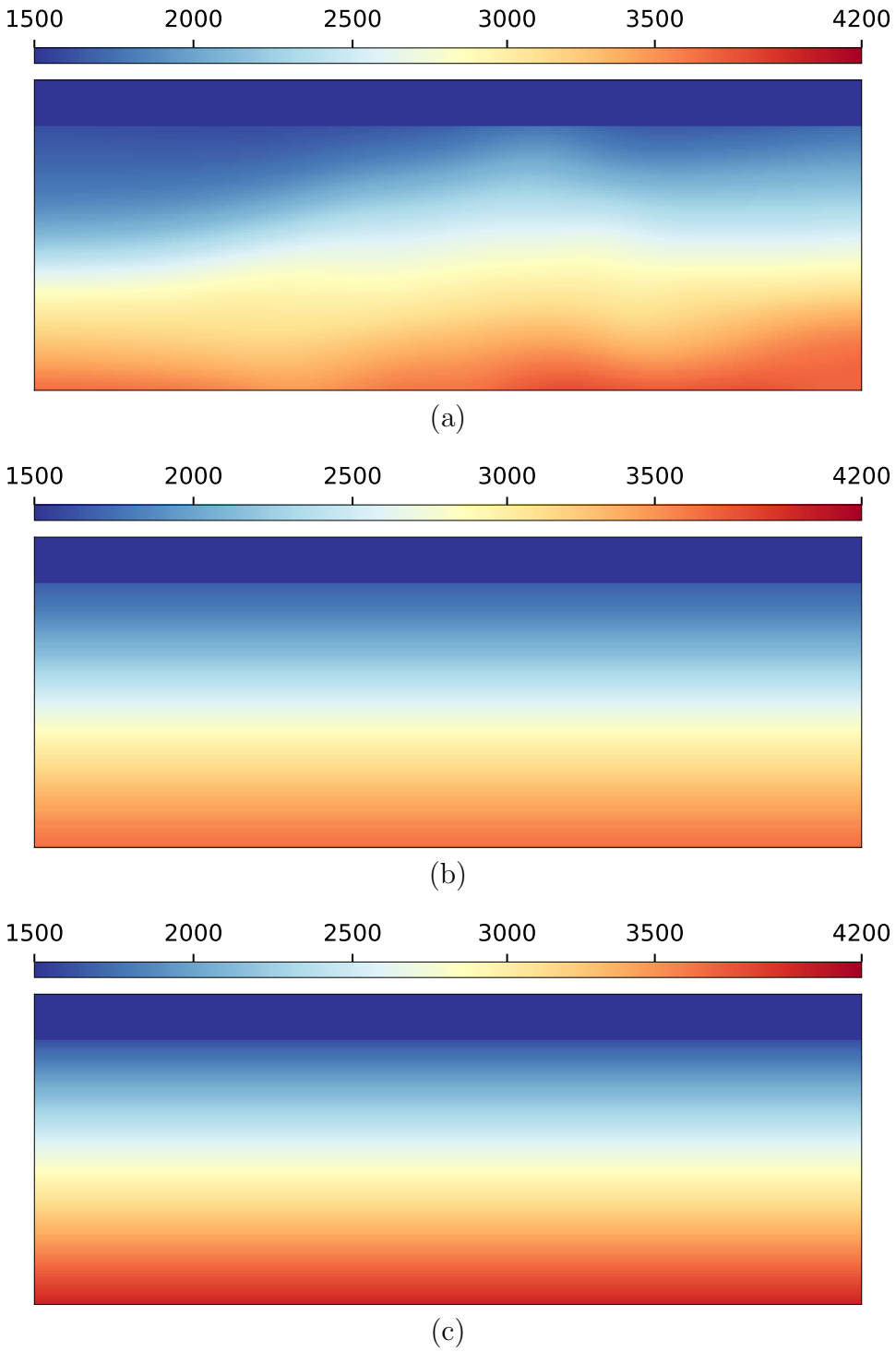
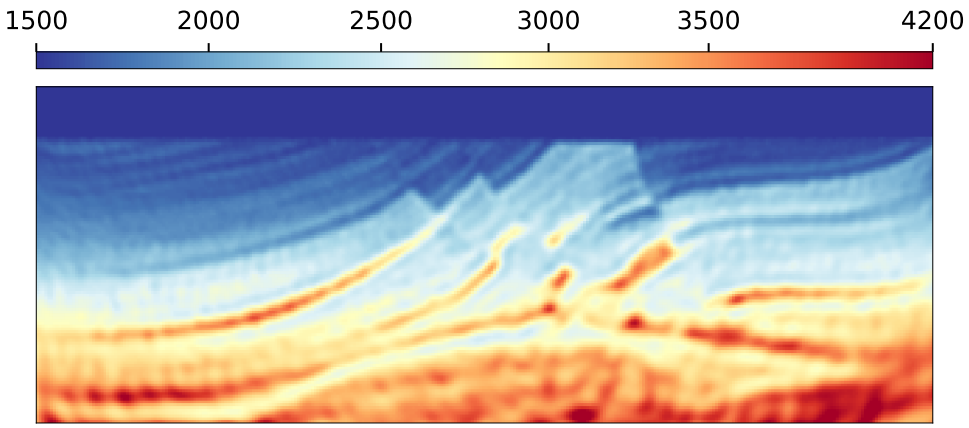
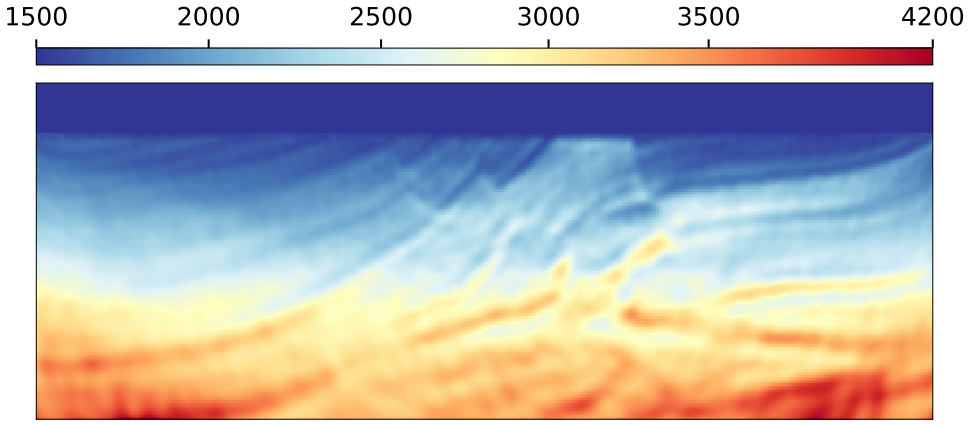


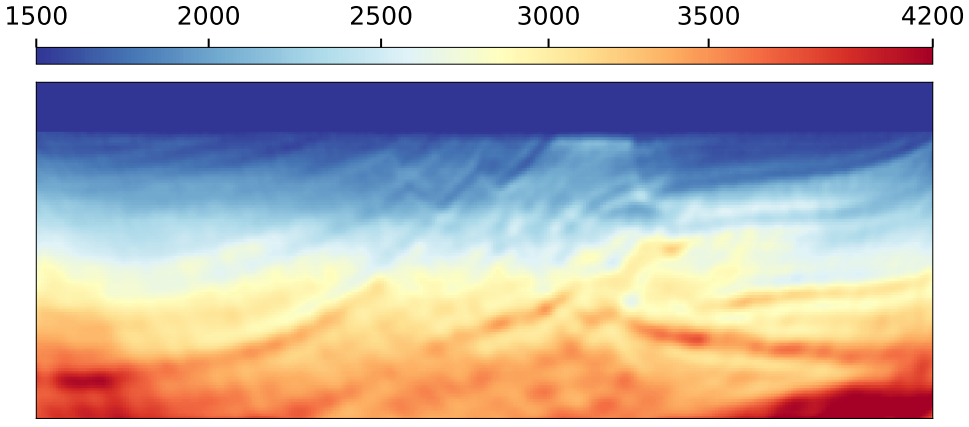
FIGURE 7. Initial values for the reconstructions: (a) blurred true model, (b) row-averaged version of (a), (c) depth linear model (1500 m/s to 4000 m/s)



(a)



(b)



(c)

FIGURE 8. Results of the Marmousi experiment with initial values: (a) blurred, (b) averaged , (c) depth linear

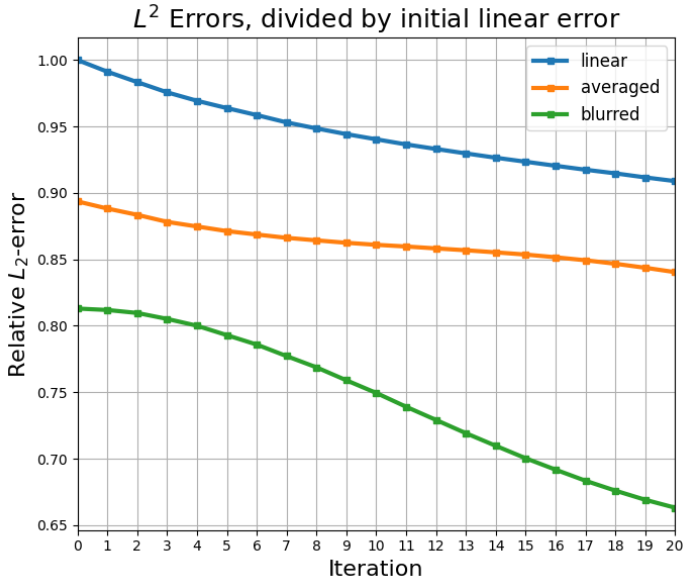


FIGURE 9. Relative L^2-errors of the Marmousi experiment with the three different initial values shown in Figure 7. The error values are normalized to the initial error of the affine linear initial guess.

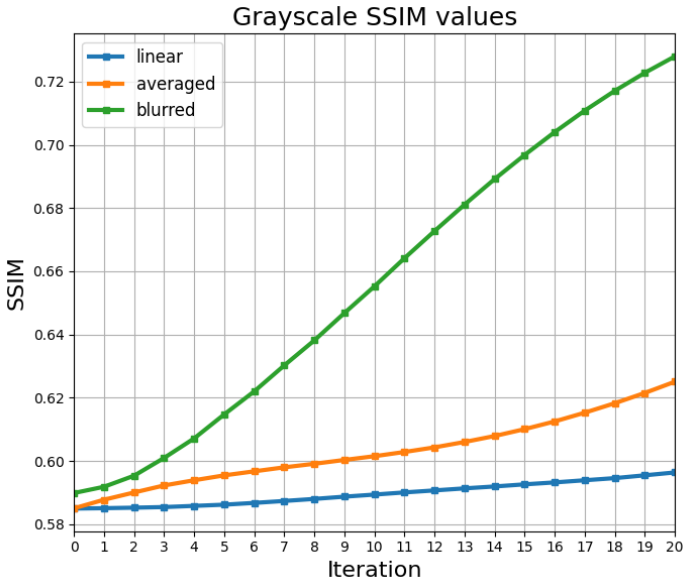


FIGURE 10. Structural similarity index of the Marmousi experiment with the three different initial values shown in Figure 7.

differences between digital images [42]. The SSIM takes values between -1 and 1, with a value of 1 only if the compared images are identical. We have computed the SSIM index

for the Marmousi experiment[6], see Figure 10. Although the initial models all exhibit comparable SSIM quality, we observe a convergence pattern similar to that of the L^2-error: The SSIM index with respect to the blurred initial model increases quickly, clearly outperforming the other initial models. Starting with the linear model, the quality of the reconstructions improves only marginally over 20 iterations, and the index evolution with respect to the averaged initial model behaves only slightly better. Nonetheless, the SSIM index indeed corresponds to our visual perception: the final iterations, considering both the averaged and linear initial models, exhibit better quality compared to the blurred initial model.

It is worth noting that changes to the parameter $l_{\min}$ could potentially enhance reconstruction outcomes; however, such optimizations are beyond the scope of this study.

5. Conclusion and outlook

We derived the inversion scheme `PMSD-REGINN` as an implementation of the time-domain all-at-once formulation of FWI in the visco-acoustic regime. Through numerical experiments, we demonstrated its effectiveness in inverting the pressure wave speed and its robustness against noise. Additionally, we demonstrated that `PMSD-REGINN` mitigates cycle-skipping to some extent: In cases where the classical formulation failed, we observed convergence with `PMSD-REGINN`. This illustrates that the algorithm is less dependent on the initial value chosen.

We also found that a combination of `PMSD-REGINN` with the classical formulation can be advantageous, that is, starting an inversion algorithm based on the classical formulation with the output of `PMSD-REGINN`. With this combination we get the benefits of both algorithms: `PMSD-REGINN` has a larger domain of convergence and the algorithm for classical formulation yields better reconstructions when it converges.

The ultimate objective of most seismic inversion techniques is to make them practical and high-performing for real-world applications, particularly for inverting field data in three dimensions. With additional features, such as source inversion, `PMSD-REGINN` can, in principle, cope with actual measured data. However, conducting three-dimensional inversions presents added challenges since the time-domain all-at-once approach requires storing at least four wavefields simultaneously: one for the current iterate, one for the update, one for the residual, and one temporary field while applying the preconditioner. Given the typical application scale of approximately 10^9 grid points, such storage demands significant memory resources, available only on large, modern computer clusters. For a 3D simulation with three damping terms (i.e., three velocity components and four pressure components), one time step in double precision consumes around 1 GB of RAM. Thus, a full wavefield, which typically incorporates at least 1000 time steps, requires over 1 TB of memory.

Therefore, to adapt the algorithm for three-dimensional problems, it is essential to reduce memory requirements. In conventional reduced methods, a common approach is to save snapshots and recalculate segments of the wavefield as needed. However, this

[6]We first converted our raw data to a gray scale images with range $[0, 255]$ and then we applied the Python routine `structural_similarity` from the `scikit-image` package.

technique is infeasible here because the wavefields to be stored are not just solutions of the wave equation. Additionally, we cannot store the right-hand sides that generate the wavefields because they are outputs of prior computations rather than known inputs. A first step to reducing memory consumption is exploring alternative preconditioners that do not depend on storing an extra wavefield.

Our proof of concept justifies further exploration and the `REGINN` framework is flexible enough to incorporate the resulting further improvements.

References

[1] A. Abubakar, W. Hu, T. M. Habashy, and P. M. van den Berg, *Application of the finite-difference contrast-source inversion algorithm to seismic full-waveform data*, Geophysics, 74 (2009), pp. WCC47–WCC58, https://doi.org/10.1190/1.3250203.

[2] H. S. Aghamiry, A. Gholami, and S. Operto, *Improving full-waveform inversion by wavefield reconstruction with the alternating direction method of multipliers*, Geophysics, 84 (2019), pp. R125–R148, https://doi.org/10.1190/geo2018-0093.1.

[3] H. S. Aghamiry, A. Gholami, and S. Operto, *Accurate and efficient data-assimilated wavefield reconstruction in the time domain*, Geophysics, 85 (2020), pp. A7–A12, https://doi.org/10.1190/geo2019-0535.1.

[4] H. S. Aghamiry, A. Gholami, and S. Operto, *Compound regularization of full-waveform inversion for imaging piecewise media*, IEEE Transactions on Geoscience and Remote Sensing, 58 (2020), pp. 1192–1204, https://doi.org/10.1109/TGRS.2019.2944464.

[5] H. S. Aghamiry, A. Gholami, and S. Operto, *Full waveform inversion by proximal Newton method using adaptive regularization*, Geophysical Journal International, 224 (2020), pp. 169–180, https://doi.org/10.1093/gji/ggaa434.

[6] K. Aghazade, A. Gholami, H. S. Aghamiry, and H. R. Siahkoohi, *Robust elastic full-waveform inversion using an alternating direction method of multipliers with reconstructed wavefields*, Geophysics, 89 (2024), pp. R287–R302, https://doi.org/10.1190/geo2023-0411.1.

[7] J. O. Blanch, J. O. A. Robertsson, and W. W. Symes, *Modeling of a constant Q: Methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique*, Geophysics, 60 (1995), pp. 176–184, https://doi.org/10.1190/1.1443744.

[8] T. Bohlen, *Viskoelastische FD-Modellierung seismischer Wellen zur Interpretation gemessener Seismogramme.*, PhD thesis, Christian-Albrechts-Universität zu Kiel, 1998, https://bit.ly/2LMOSWr.

[9] T. Bohlen, M. R. Fernandez, J. Ernesti, C. Rheinbay, A. Rieder, and C. Wieners, *Visco-acoustic full waveform inversion: From a DG forward solver to a Newton-CG inverse solver*, Computers and Mathematics with Applications, 100 (2021), pp. 126–140, https://doi.org/10.1016/j.camwa.2021.09.001.

[10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2011), p. 1–122, https://doi.org/10.1561/2200000016.

[11] R. Brossier, S. Operto, and J. Virieux, *Robust elastic frequency-domain full-waveform inversion using the L1 norm*, Geophysical Research Letters, 36 (2009), https://doi.org/10.1029/2009GL039458.

[12] C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent, *Multiscale seismic waveform inversion*, Geophysics, 60 (1995), pp. 1457–1473, https://doi.org/10.1190/1.1443880.

[13] J. M. Carcione, *Wave Fields in Real Media: Wave Propagation in Anisotropic, Anelastic, Porous and Electromagnetic Media*, ISSN, Elsevier Science, 2014, https://doi.org/10.1016/C2021-0-00938-X.

[14] B. Engquist and Y. Yang, *Optimal transport based seismic inversion: Beyond cycle skipping*, Communications on Pure and Applied Mathematics, 75 (2022), pp. 2201–2244, https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.21990.

[15] E. Esser, L. Guasch, T. van Leeuwen, A. Y. Aravkin, and F. J. Herrmann, *Total variation regularization strategies in full-waveform inversion*, SIAM J. Imaging Sci., 11 (2018), pp. 376–406, https://doi.org/10.1137/17M111328X.

[16] O. Gauthier, J. Virieux, and A. Tarantola, *Two-dimensional nonlinear inversion of seismic waveforms: Numerical results*, Geophysics, 51 (1986), pp. 1387–1403, https://doi.org/10.1190/1.1442188.

[17] A. Gholami, H. S. Aghamiry, and S. Operto, *Extended-space full-waveform inversion in the time domain with the augmented Lagrangian method*, Geophysics, 87 (2022), pp. R63–R77, https://doi.org/10.1190/geo2021-0186.1.

[18] G. Guo, S. Operto, A. Gholami, and H. S. Aghamiry, *A practical implementation of data-space Hessian in the time-domain formulation of source extended full-waveform inversion*, Society of Exploration Geophysicists, 2022, pp. 757–761, https://doi.org/10.1190/image2022-3749861.1.

[19] G. Guo, S. Operto, A. Gholami, and H. S. Aghamiry, *Time-domain extended-source full-waveform inversion: Algorithm and practical workflow*, Geophysics, 89 (2024), pp. R73–R94, https://doi.org/10.1190/geo2023-0055.1.

[20] E. Haber and U. M. Ascher, *Preconditioned all-at-once methods for large, sparse parameter estimation problems*, Inverse Problems, 17 (2001), pp. 1847–1864, https://doi.org/10.1088/0266-5611/17/6/319.

[21] G. Huang, R. Nammour, and W. W. Symes, *Source-independent extended waveform inversion based on space-time source extension: Frequency-domain implementation*, Geophysics, 83 (2018), pp. R449–R461, https://doi.org/10.1190/geo2017-0333.1.

[22] B. Kaltenbacher, *Regularization based on all-at-once formulations for inverse problems*, SIAM J. Numer. Anal., 54 (2016), pp. 2594–2618, https://doi.org/10.1137/16M1060984.

[23] A. Kirsch and A. Rieder, *Inverse problems for abstract evolution equations with applications in electrodynamics and elasticity*, Inverse Problems, 32 (2016), pp. 085001, 24, https://doi.org/10.1088/0266-5611/32/8/085001.

[24] A. Kurzmann, *Applications of 2D and 3D full waveform tomography in acoustic and viscoacoustic complex media*, PhD thesis, Karlsruher Institut für Technologie (KIT), 2012, https://doi.org/10.5445/IR/1000034421.

[25] A. Lechleiter and A. Rieder, *Towards a general convergence theory for inexact Newton regularizations*, Numer. Math., 114 (2010), pp. 521–548, https://doi.org/10.1007/s00211-009-0256-0.

[26] H.-P. Liu, D. L. Anderson, and H. Kanamori, *Velocity dispersion due to anelasticity; implications for seismology and mantle composition*, Geophysical Journal International, 47 (1976), pp. 41–58, https://doi.org/10.1111/j.1365-246X.1976.tb01261.x.

[27] G. Martin, R. Wiley, and K. Marfurt, *Marmousi2: An elastic upgrade for Marmousi*, The Leading Edge, 25 (2006), pp. 156–166, https://doi.org/10.1190/1.2172306.

[28] L. Métivier, A. Allain, R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, *Optimal transport for mitigating cycle skipping in full-waveform inversion: A graph-space transform approach*, Geophysics, 83 (2018), pp. R515–R540, https://doi.org/10.1190/geo2017-0807.1.

[29] S. Operto, A. Gholami, H. Aghamiry, G. Guo, S. Beller, K. Aghazade, F. Mamfoumbi, L. Combe, and A. Ribodetti, *Extending the search space of full-waveform inversion beyond the single-scattering born approximation: A tutorial review*, Geophysics, 88 (2023), pp. R671–R702, https://doi.org/10.1190/geo2022-0758.1.

[30] S. Operto, Y. Gholami, V. Prieux, A. Ribodetti, R. Brossier, L. Métivier, and J. Virieux, *A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice*, The Leading Edge, 32 (2013), pp. 1040–1054, https://doi.org/10.1190/tle32091040.1.

[31] C. Rheinbay, *All-At-Once and Reduced Solvers for Visco-Acoustic Full Waveform Inversion*, PhD thesis, Karlsruher Institut für Technologie (KIT), 2024, https://doi.org/10.5445/IR/1000167623.

[32] A. Rieder, *On the regularization of nonlinear ill-posed problems via inexact Newton iterations*, Inverse Problems, 15 (1999), pp. 309–327, https://doi.org/10.1088/0266-5611/15/1/028.

[33] A. Rieder, *An all-at-once approach to full waveform inversion in the viscoelastic regime*, Math. Methods Appl. Sci., 44 (2021), pp. 6376–6388, https://doi.org/10.1002/mma.7190.

[34] W. W. Symes, *Wavefield reconstruction inversion: an example*, Inverse Problems, 36 (2020), p. 105010, https://dx.doi.org/10.1088/1361-6420/abaf66.

[35] W. W. Symes, H. Chen, and S. E.Minkoff, *Full-waveform inversion by source extension: Why it works*, 2020, pp. 765–769, https://library.seg.org/doi/abs/10.1190/segam2020-3424509.1.

[36] A. Tarantola, *Inversion of seismic reflection data in the acoustic approximation*, Geophysics, 49 (1984), pp. 1259–1266, https://doi.org/10.1190/1.1441754.

[37] T. van Leeuwen, *A note on extended full waveform inversion*, 2019, https://arxiv.org/abs/1904.00363.

[38] T. van Leeuwen and F. J. Herrmann, *Mitigating local minima in full-waveform inversion by expanding the search space*, Geophysical Journal International, 195 (2013), pp. 661–667, https://doi.org/10.1093/gji/ggt258.

[39] T. van Leeuwen and F. J. Herrmann, *A penalty method for PDE-constrained optimization in inverse problems*, Inverse Problems, 32 (2015), p. 015007, https://doi.org/10.1088/0266-5611/32/1/015007.

[40] J. Virieux and S. Operto, *An overview of full-waveform inversion in exploration geophysics*, Geophysics, 74 (2009), pp. WCC1–WCC26, https://doi.org/10.1190/1.3238367.

[41] C. Wang, D. Yingst, P. Farmer, and J. Leveille, *Full-waveform inversion with the reconstructed wavefield method*, Society of Exploration Geophysicists, 2016, pp. 1237–1241, https://doi.org/10.1190/segam2016-13870317.1.

[42] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, *Image quality assessment: from error visibility to structural similarity*, IEEE Transactions on Image Processing, 13 (2004), pp. 600–612, https://doi.org/10.1109/TIP.2003.819861.

[43] R. Winkler and A. Rieder, *Model-aware Newton-type inversion scheme for electrical impedance tomography*, Inverse Problems, 31 (2015), pp. 045009, 27, https://doi.org/10.1088/0266-5611/31/4/045009.

[44] P. Yang, J. Gao, and B. Wang, *A graphics processing unit implementation of time-domain full-waveform inversion*, Geophysics, 80 (2015), pp. F31–F39, https://doi.org/10.1190/geo2014-0283.1.

[45] P. Yang and W. Zhou, *Algorithmic analysis towards time-domain extended source waveform inversion*, 2023, https://arxiv.org/abs/2211.06300.

[46] U. C. Zeltmann, *The Viscoelastic Seismic Model: Existence, Uniqueness and Differentiability with Respect to Parameters*, PhD thesis, Karlsruher Institut für Technologie (KIT), 2019, https://doi.org/10.5445/IR/1000093989.