

ITERATIVE RECONSTRUCTION OF PIECEWISE SMOOTH WAVESPEEDS USING A CRITERION DERIVED FROM THE SCATTERING RELATION

RUICHAO YE* AND MAARTEN V DE HOOP†

Abstract. We study the inverse boundary value problem for the wave equation and recovery of the wavespeed in the high-frequency limit. In the low-frequency regime we have studied the convergence of iterative methods via the Fourier transform in time, using the Helmholtz equation. Here, we focus on high-resolution reconstruction, again, using iterative methods. We restrict our analysis to piecewise smooth wavespeed models containing interfaces. We assume that the presence (but not the location) of interfaces is known. We introduce directional boundary sources (such as wave packets) and a mismatch functional given by the cross correlation power constrained by a structure tensor residual.

1. Introduction. We study the inverse boundary value problem for the wave equation and recovery of the wavespeed in the high-frequency limit. Specifically, we focus on the formulation of iterative methods for the inverse boundary value problem for the wave equation referred to as full waveform inversion (FWI) at relatively high frequencies on the one hand and address the integration of these methods with wave equation reflection tomography on the other hand. (For work related to this aspect, see Symes [45].) Our approach is motivated by the geodesic X -ray transform using the (multiply) broken scattering relation as the data, which can be obtained from the Dirichlet-to-Neumann map. We develop a strategy derived from a uniqueness theorem (partial reconstruction) for the associated inverse problem [24]. We constrain our models to *models with structure* (piecewise smooth with discontinuities, sometimes referred to as blocky), represented on an unstructured (tetrahedral) mesh. We synthesize directional (Gaussian) wavepackets (belonging to a frame [20]) from ‘point’ sources, and we may require certain preprocessing of the data containing multiple scattered wave constituents involving microlocal cutoffs, including time windowing.

The approach we propose, here, is part of a comprehensive procedure we have been developing for the mentioned inverse boundary problem. This procedure is initiated by taking a time-Fourier transform of the data followed by a hierarchical compressed reconstruction following a multi-level scheme for the inverse boundary value problem for the Helmholtz equation using the Dirichlet-to-Neumann map as the data [15, 4, 5]. The effectiveness of this reconstruction has been analyzed and is essentially limited to a minimum size of the subdomains underlying a blocky model representations and to lower frequencies. We then segment (without interpretation) the result on an unstructured mesh with internal boundaries [22] and optimize their shapes in particular and varying wavespeeds in general via a gradient flow using the data. The result is used in the approach developed in this paper. We emphasize that it yields in principle models with limited resolution, though through re-segmentation progressively the resolution can be enhanced while interfaces can vanish. We use a subset of (multiple) reflections in the data which is controlled by window functions in our misfit functional and data simulated in the current model. We use the adjoint-state method and an augmented Lagrangian to arrive at our iterative scheme. (For a review of this method, see Plessix & Mulder [34].)

It has been recognized that upon considering a misfit based on least-squares, the separation in time between measured and modeled data needs to be within a half period to avoid convergence toward a local minimum (Beydoun & Tarantola [6]; one refers to this as cycle skipping). Pratt *et al.* [37] analyzed this limitation more carefully and related the cycle skipping to the relative time error to the reciprocal of the length of the propagation path in number of wavelengths. This was our motivation to develop a proxy to detecting a distance between scattering relations (Stefanov & Uhlmann [43]) from finite-frequency measured and modeled data. The scattering relation contains the surface source and receiver locations, the source and receiver ‘horizontal’ slownesses (‘slopes’)

*Department of Earth and Atmospheric Sciences, Purdue University, 150 N. University Street, West Lafayette, IN 47907, USA.

†Department of Mathematics, Purdue University, 150 N. University Street, West Lafayette, IN 47907, USA.

and travel times corresponding with (broken) ray paths. (We note that the scattering relation is also used in stereotomography as the data (Billette & Lambaré [7] and Lambaré *et al.* [30].) In principle, the scattering relation can be obtained from the Dirichlet-to-Neumann map. To obtain a criterion, we introduce directional boundary sources (Gaussian wave packets) and a mismatch functional given by the cross correlation power regularized by a structure tensor residual with Gaussian window functions.

The use of cross correlation power, for the purpose of velocity analysis via local optimization, was proposed by Van Leeuwen & Mulder [48, 49], and may be compared with the splitting intensity introduced for the purpose of estimating anisotropy from shear-wave data (Long *et al.* [32]). The cross correlation power is sensitive to the travel residual and intuitively more linearly related to large-scale wavespeed variations [11]. The incorporation of a cross correlation in the misfit functional to arrive at an ad hoc integrated approach to FWI and wave-equation tomography was considered by Wang, Singh & Jian [52].

Double beamforming was used by Brossier & Roux [10] in time-harmonic FWI. Under the assumption that the wavespeed is known near the surface where the sources and receivers are located, they exploited beamforming to select arrivals based on slopes and arrived at a hierarchical level with the purpose to mitigate the nonlinearity of the inverse problem. Downey, Routh & Cha [17] introduced a method of grouping shots into encoded multi-shot gathers. The encoding is convolutional in time and the process results in the generation of random beams. The motivation was a significant speedup in the computation of the gradient. We also mention the work of Vigh & Starr [50] in which plane-wave gathers are used as the data.

An energy norm in the misfit can be theoretically justified in the frequency-domain formulation. However, the effectiveness of FWI at higher frequencies has been identified as challenging (Virieux & Operto [51]), which has been understood by analyzing the stability of the time-harmonic seismic inverse problem. The time-harmonic formulation was initially promoted by Pratt and collaborators [39, 36] and specifically for wide-angle reflection data in [38]. Pratt, Shin & Hicks [35] developed the Newton method using the discretized Helmholtz equation as a point of departure.

Lailly [28] and Tarantola [46, 47] introduced the formulation of the seismic inverse problem as a local optimization problem encompassing a least-squares minimization of a misfit functional. They noticed that the gradient of this functional has the form of an imaging procedure with a cross correlation imaging condition. We also mention the original work of Bamberg, Chavent & Lailly [3, 2]. Chavent [12] introduced a method based on the adjoint state in inverse problems to compute the gradient of a functional without the Fréchet derivatives. This originated from control theory [31]. Kolb, Collino & Lailly [27] carried out initial computational experiments with such a procedure restricting the models to the planar layered class.

The apparent ability of wide aperture data to resolve the long wavelengths of the model has prompted studies to consider long-offset acquisition as a way to design an improved FWI [38, 42, 40]. Indeed, the least-squares misfit, for short-offset acquisition, is quite insensitive to intermediate wavelength variations in the model. We also assume the availability of long-offset data.

A strategy to mitigate the nonlinearity in waveform inversion is layer stripping. Here, the wavespeed is only recovered in a pre-determined layer, possibly conforming geology using interpreted horizons, hierarchically from shallow to deep. (One can more generally ‘mask’ subdomains in the inversion process.) This strategy might be coupled to source-receiver offset weighting, that is, progressively introducing longer offsets, shorter apertures and longer recording times, with the goal to optimize sensitivity and resolution in given regions in the subsurface [42, 53].

Jurado, Sinoquet & Ehinger [25] considered reflection tomography for piecewise smooth models (in reflection seismology also referred to as ‘blocky’ models). They parametrized the smooth parts of the models by B -splines, and the interfaces by a level set function, that is, depth as a function of surface coordinates also using B -splines. The interfaces can intersect one another and thus form a pinch out by allowing a zero jump across portions of interfaces. The data were reflection times associated with single scattered waves. These could be multi-valued while they parameterized the data by source angle. They incorporated inequality constraints and used a Gauss-Newton method.

In this context, we also mention the work of Lailly & Sinoquet [29] and Delprat-Jannaud & Lailly [16], who analyzed ‘ill-posed and well-posed’ formulations of the reflection tomography problem. Later, Clapp, Biondi & Claerbout [14] incorporated geologic information into reflection tomography. Malcolm & Nicholls [33] studied (constrained quadratic) optimization for interface reconstruction for impenetrable period acoustic media using time-harmonic data.

The use of unstructured polyhedral meshes to represent (piecewise constant) wavespeed models in inverse problems has been proposed in global optimization (genetic algorithms) by Schoenauer, Ehinger and Braunschweig [41]; they based their evolving representations on Voronoï diagrams.

The process of progressively increasing the complexity of the models has been proposed both in local (iterative) and global optimization. In the latter category we mention the pseudo-subspace method of Boschetti [8] where the progression follows a user-defined evolution. We follow a strategy or re-segmentation. For an alternative method for model reparameterization via preconditioning and explicitly estimating the dips in the gradient, see Guitton, Ayeni & Díaz [21].

One typically applies a ‘nonlinear’ conjugate gradient method, a Gauss-Newton method, or a quasi-Newton method (L-BFGS; for a review, see Brossier, Operto & Virieux [9]). For the application of multi-scale Newton methods, see Akcelik [1]. We carry out numerical experiments with the nonlinear conjugate gradient and L-BFGS methods. In the gradient method, the step length is typically estimated by a simple line search for which a linearization of the direct problem is used (Gauthier, Virieux & Tarantola [18]). This estimation is challenging in practice and may lead to a failure of convergence. In various approaches one accounts for the Hessian, at least its diagonal, in the Gauss-Newton scheme [26]. In certain earthquake seismology applications, one builds the Fréchet derivative or Jacobian (sensitivity, for example, [13]) explicitly and then applies LSQR. The second term in the Hessian (Newton method) can be interpreted in terms of multiple scattering (see, for example, Pratt, Shin & Hicks [35]).

The inverse boundary value problem: Modelling the data. We focus on the acoustic wave modelling with constant density on the spatial domain \mathbb{R}^n with $n = 1, 2, 3$, and the data are modelled as solutions to the inhomogeneous Cauchy initial value problem,

$$(1.1) \quad [c^{-2}(x)\partial_t^2 - \Delta] u_i(t, x) = f_i(t, x),$$

$$(1.2) \quad u_i(0, x) = u_{0;i}(x) = 0, \quad \partial_t u_i(0, x) = u_{1;i}(x) = 0.$$

Here, $c = c(x)$ is a piecewise smooth wavespeed and contains conormal singularities, and $f_i(t, x) = \bar{f}_i(t, x')\delta(x_n)$, is a directional boundary source. To simplify the notation, we will sometimes write $\bar{x} = (t, x')$. We choose the directional source to be a curvelet-like Gaussian wave packet: in dimension two,

$$(1.3) \quad \varphi_{j,k,\lambda}(\bar{x}) := 2^{3/2j}\varphi(A_{j,k}\bar{x} - \lambda), \quad (j \geq 1, 0 \leq k \leq 2^j - 1, \lambda \in \Lambda),$$

where

$$A_{j,k} := \begin{bmatrix} 4^j & 0 \\ 0 & 2^j \end{bmatrix} \begin{bmatrix} \cos(\frac{2k\pi}{2^j}) & -\sin(\frac{2k\pi}{2^j}) \\ \sin(\frac{2k\pi}{2^j}) & \cos(\frac{2k\pi}{2^j}) \end{bmatrix},$$

and $\Lambda \subseteq \mathbb{R}^2$ is a lattice. If φ is fast decaying, the packet $\varphi_{j,k,\lambda}$ is localized near $A_{j,k}^{-1}\lambda$ and aligned at an angle of $\frac{-2\pi k}{2^j}$ radians. We assign to the multi-index (j, k, λ) .

We observe the solutions u_i restricted to $[0, T] \times \Sigma_r \ni (t, x_r) = \bar{x}_r$, where Σ_r is an open subset of the surface, $x_n = 0$.

2. Iterative scheme.

2.1. Correlation power criterion. To conduct the inverse scattering, we require a comparison between pairs of seismograms generated by each source in the acquisition geometry. We consider the functional

$$(2.1) \quad J = \sum_i \frac{1}{S_{t,i}} \int_{\Sigma_r} \left(\int_{-T_{\max}}^{T_{\max}} W_\sigma(\Delta t) |C_t[d_i, u_i](\Delta t, x_r)|^2 d\Delta t \right) dx_r,$$

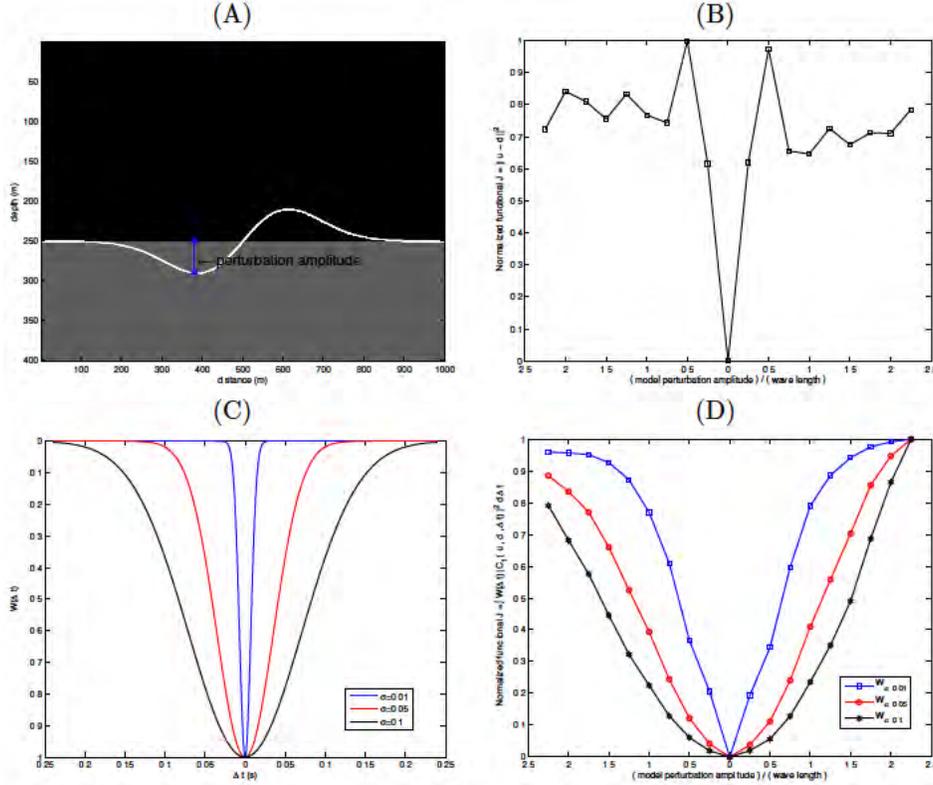


FIG. 1. Comparison between the L^2 and cross-correlation power criteria. (A) Illustration of the interface perturbation. (B) L^2 misfit as a function of the perturbation. (C) Weight functions used with different values of σ . (D) Cross-correlation functionals (normalized by its L^∞ norm and shifted to zero) as a function of the perturbation.

where C_t denotes the cross-correlation between observed data d_i and modelled data u_i ,

$$(2.2) \quad C_t[d_i, u_i](\Delta t, x_r) = \int_0^T d_i(t + \Delta t, x_r, x_{s;i}) u_i(t, x_r) dt;$$

implicitly, here, u_i has been restricted to Σ_r . W_σ is a Gaussian weighting function with variance σ ,

$$W_\sigma(\Delta t) = -\exp\left[-\frac{\Delta t^2}{\sigma^2}\right],$$

which penalizes correlation energy away from zero time shift. This weighting function is dependent on the frequency content of the data. We choose σ such that W_σ has a width that is of the same order as the first peak of the frequency-dependent correlation, by which means, a complete seismic phase that contains one or several wave period is picked out and separated from other phases that come from different seismic events. Furthermore,

$$(2.3) \quad S_{t;i} = \int_{\Sigma_r} \left(\int_{-T_{\max}}^{T_{\max}} |C_t[d_i, u_i](\Delta t, x_r)|^2 d\Delta t \right) dx_r,$$

represents a normalization which ensures that the functional is not minimized simply by correlating low-amplitude background and shifting events in the data outside the correlation windows.

Upon describing the scattered waves in the vicinity of the surface asymptotically, we note that J is sensitive to the travel time residual. Figure 1 shows an illustration of J as a function of a certain

model perturbation. Here, the interface is represented by an amplitude times a given smooth function $h = h(x)$, see Figure 1 A. We generate synthetic data and compute the functional J with data generated by a flat (zero amplitude) interface. The traditional L^2 functional exhibits local minima and gives a narrow range of convergence as expected; see Figure 1 B. The cross-correlation power criterion gives a much larger range of convergence. It gives a proper descent direction even if the data misfit caused by a model perturbation is as large as 2.5 wavelengths, see Figure 1 C. We note that this is sensitive to the choice of σ ; see Figures 1 C and D. We can choose to vary σ in our iterative scheme.

2.2. Augmented Lagrangian. We develop an adjoint method for determining the gradient of the error functional with respect to the model parameters. To this end, we introduce the augmented functional (cf. (1.1), (1.2), (2.1) and (2.2))

$$\begin{aligned}
 \mathfrak{L} = & J + \sum_i \int_{\Sigma_r} \int_{-T_{\max}}^{T_{\max}} \lambda_{C;i}(\Delta t, x_r) \left(C_{t;i}(\Delta t, x_r) - \int_0^T d_i(t + \Delta t, x_r) u_i(t, x_r) dt \right) d\Delta t dx_r \\
 & + \sum_i \int_0^T \int_{\Omega} \lambda_{u;i}(t, x) \left([c^{-2}(x)\partial_t^2 - \Delta] u_i(t, x) - f_i(x, t) \right) dx dt \\
 (2.4) \quad & + \sum_i \int_{\Omega} \lambda_{u_0;i}(x) u_i|_{t=0}(x) dx + \sum_i \int_{\Omega} \lambda_{u_1;i}(x) \partial_t u_i|_{t=0}(x) dx \\
 & + \int_{\Omega} \lambda_c(x) \left(c(x) - \sum_{k=1}^{N_c} \gamma_k \psi_k(x) \right) dx,
 \end{aligned}$$

in which c has been expanded into basis functions:

$$(2.5) \quad c(x) = \sum_{k=1}^{N_c} \gamma_k \psi_k(x).$$

The basis functions will vary depending on the structure of the perturbation under consideration.

Gradient. We first carry out an integration by parts:

$$\begin{aligned}
 \int_0^T \int_{\Omega} \lambda_{u;i}(t, x) c^{-2}(x) \partial_t^2 u_i(t, x) dx dt &= \int_0^T \int_{\Omega} c^{-2}(x) (\partial_t^2 \lambda_{u;i})(t, x) u_i(t, x) dx dt \\
 (2.6) \quad &+ \int_{\Omega} \lambda_{u;i}(T, x) c^{-2}(x) \partial_t u_i(T, x) dx - \int_{\Omega} \lambda_{u;i}(0, x) c^{-2}(x) \partial_t u_i(0, x) dx \\
 &- \int_{\Omega} c^{-2}(x) (\partial_t \lambda_{u;i})(T, x) u_i(T, x) dx + \int_{\Omega} c^{-2}(x) (\partial_t \lambda_{u;i})(0, x) u_i(0, x) dx,
 \end{aligned}$$

and replace the third term of Lagrangian. We take the gradient of the Lagrangian with respect to the variables $C_t, u_i, u_{0;i}, u_{1;i}$ and c yielding

$$(2.7) \quad \frac{\partial \mathcal{L}}{\partial C_{t;i}} = \frac{2}{S_{t;i}} (W_\sigma(\Delta t) - J) C_{t;i}(\Delta t, x_r) + \lambda_{C;i}(\Delta t, x_r),$$

$$(2.8) \quad \begin{aligned} \frac{\partial \mathcal{L}}{\partial u_i} &= [c^{-2}(x)\partial_t^2 - \Delta] \lambda_{u;i}(t, x) \\ &\quad - \int_{\Sigma_r} \delta(x - x_r) \int_{-T_{\max}}^{T_{\max}} \lambda_{C;i}(\Delta t, x_r) d_i(t + \Delta t, x_r) d\Delta t dx_r, \end{aligned}$$

$$(2.9) \quad \frac{\partial \mathcal{L}}{\partial u_{0;i}} = c^{-2}(x)(\partial_t \lambda_{u;i})(0, x) + \lambda_{u_{0};i}(x),$$

$$(2.10) \quad \frac{\partial \mathcal{L}}{\partial u_{1;i}} = -c^{-2}(x)\lambda_{u;i}(0, x) + \lambda_{u_{1};i}(x),$$

$$(2.11) \quad \frac{\partial \mathcal{L}}{\partial c} = \sum_i \int_0^T \lambda_{u;i}(t, x) (-2)c^{-3}(x)\partial_t^2 u_i(t, x) dt + \lambda_c(x),$$

$$(2.12) \quad \frac{\partial \mathcal{L}}{\partial \gamma_k} = \frac{\partial J}{\partial \gamma_k} - \int_\Omega \lambda_c(x)\psi_k(x) dx,$$

whereas

$$(2.13) \quad \lambda_{u;i}(T, x) = 0, \quad \partial_t \lambda_{u;i}(T, x) = 0.$$

The adjoint states, $\lambda_{u_{0};i}, \lambda_{u_{1};i}$ do not play a role in the gradient of J . To obtain the gradient, we have to solve the final value problem

$$(2.14) \quad [c^{-2}(x)\partial_t^2 - \Delta] \lambda_{u;i}(t, x) = \int_{\Sigma_r} \delta(x - x_r) \int_{-T_{\max}}^{T_{\max}} \lambda_{C;i}(\Delta t, x_r) d_i(t + \Delta t, x_r) d\Delta t dx_r,$$

$$(2.15) \quad \lambda_{u;i}(T, x) = 0, \quad \partial_t \lambda_{u;i}(T, x) = 0,$$

where

$$(2.16) \quad \lambda_{C;i}(\Delta t, x_r) = -\frac{2}{S_{t;i}} (W_\sigma(\Delta t) - J) C_t[d_i, u_i](\Delta t, x_r).$$

We note the appearance of cross correlation power appears in the source on the right-hand side. Then

$$(2.17) \quad \frac{\partial J}{\partial \gamma_k} = - \int_\Omega \left(\sum_i \int_0^T \lambda_{u;i}(t, x) (-2)c^{-3}(x)\partial_t^2 u_i(t, x) dt \right) \psi_k(x) dx.$$

Interfaces: Shape optimization. Here, we adapt (2.12) to the gradient with respect to vertices, x_κ , defining interfaces, Γ say. The interface is triangulated. The (smooth) wavespeed functions on either side of Γ are c_1 and c_2 . We replace λ_c by

$$(2.18) \quad \tilde{\lambda}_c(x) = - \sum_i \int_0^T \lambda_{u;i}(t, x) \partial_t^2 u_i(t, x) dt,$$

and obtain

$$DJ(\Omega_\alpha, c_1, c_2)V = \int_\Gamma (c_1^{-2}(x) - c_2^{-2}(x)) \tilde{\lambda}_c(x)\nu(x) \cdot V(x) d\sigma.$$

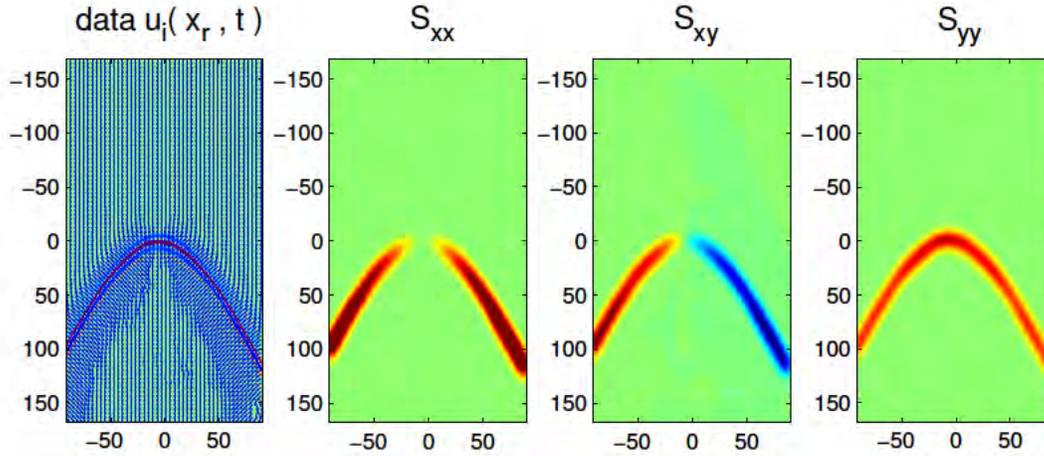


FIG. 2. An example of a structure tensor computation for a two-dimensional data set. The left figure shows the original data; the blue arrows indicate predominant gradient direction. The right three figures illustrate the independent components of the structure tensor.

Let ϕ_κ denote piecewise linear interpolating basis functions, such that $\phi_\kappa(x_{\kappa'}) = \delta_{\kappa\kappa'}$, while $\sum_\kappa \phi_\kappa(x) = 1$. Then

$$(2.19) \quad \frac{\partial J}{\partial x_\kappa} = \sum_{j \in J_\kappa} \int_{\mathcal{T}_j} \left(\sum_i \int_0^T \lambda_{u_i}(t, x) (c_1^{-2}(x) - c_2^{-2}(x)) \partial_t^2 u_i(t, x) dt \right) \nu_j(x) \phi_\kappa(x) d\sigma,$$

where J_κ represents the set of triangles containing vertex x_κ , and ν_j is the normal of triangle \mathcal{T}_j . Then

$$DJ(\Omega_\alpha, c_1, c_2)V = \sum_\kappa V_\kappa \cdot \frac{\partial J}{\partial x_\kappa}.$$

In shape optimization, we set $V = v\nu$.

2.3. Directional constraints: Structure tensor. In the following discussion we use the shorthand notation, $\bar{x}_r = (t, x_r)$. The multi-scale structure tensor is defined as

$$(2.20) \quad (Sd_{i;\lambda})(\bar{x}_r) := \int (\nabla d_{i;\lambda})(\bar{x}_r - \bar{y}_r) (\nabla d_{i;\lambda})^T(\bar{x}_r - \bar{y}_r) W_\mu(\bar{y}_r) d\bar{y}_r,$$

where

$$(2.21) \quad d_{i;\lambda}(\bar{x}_r) = W_\lambda(\cdot) \overset{(\bar{x}_r)}{*} d_i(\cdot),$$

in which W_λ and W_μ are chosen to be Gaussians in \mathbb{R}^n . The structure tensor detects the predominant directions of the gradient in a specified neighborhood of a point, and the degree to which those directions are coherent. The structure tensor is symmetric and, generically, positive definite. An example of a structure tensor in dimension two is given in Figure 2.

In the development of the adjoint state method we will suppress W_λ . We denote the right eigenvector associated with the largest eigenvalue of $S_i(\bar{x}_r) = (Sd_{i;\lambda})(\bar{x}_r)$ by $ES_i(\bar{x}_r)$, which coincides with the predominant direction of the gradient. The remaining eigenvectors are denoted by e_i . We assume that $\{E_i, (e_i)\}$ are orthonormal. Furthermore, Σ_i is the largest eigenvalue and the remaining eigenvalues are denoted by σ_i . See also Hale [23]. The contribution, J_s , to J reflecting the constraint is defined as

$$J_s = \frac{\alpha}{2} \sum_i \int_{\Sigma_r} \int_0^T \|E_i(t, x_r) - (E(Sd_{i;\lambda}))(t, x_r)\|^2 dt dx_r,$$

where $\|\cdot\|$ denotes the Euclidean norm in \mathbb{R}^n .

We obtain additional contributions to the augmented Lagrangian:

$$\begin{aligned}
(2.22) \quad \mathfrak{L}_s &= J_s + \sum_i \int_{\Sigma_r} \int_0^T \lambda_{E;i}(\bar{x}_r) \cdot (E_i(\bar{x}_r) - (ES_i)(\bar{x}_r)) \, d\bar{x}_r \\
&+ \sum_i \int_{\Sigma_r} \int_0^T \lambda_{S;i}(\bar{x}_r) : \left(S_i(\bar{x}_r) - \int p_i(\bar{x}_r - \bar{y}_r) p_i^T(\bar{x}_r - \bar{y}_r) W_\mu(\bar{y}_r) \, d\bar{y}_r \right) \, d\bar{x}_r \\
&+ \sum_i \int_{\Sigma_r} \int_0^T \lambda_{p;i}(\bar{x}_r) \cdot (p_i(\bar{x}_r) - \nabla u_{i;\lambda})(\bar{x}_r)) \, d\bar{x}_r.
\end{aligned}$$

We apply an integration by parts,

$$\int_{\Sigma_r} \int_0^T \lambda_{p;i}(\bar{x}_r) \cdot (p_i(\bar{x}_r) - \nabla u_{i;\lambda})(\bar{x}_r)) \, d\bar{x}_r = \int_{\Sigma_r} \int_0^T (\lambda_{p;i}(\bar{x}_r) \cdot p_i(\bar{x}_r) + (\nabla \cdot \lambda_{p;i})(\bar{x}_r) u_{i;\lambda}(\bar{x}_r)) \, d\bar{x}_r.$$

Taking the gradient of \mathfrak{L}_s with respect to the variables E_i , S_i , p_i and u_i yields

$$(2.23) \quad \frac{\partial \mathfrak{L}_s}{\partial E_i} = \alpha E_i(t, x_r) + \lambda_{E;i}(t, x_r),$$

$$(2.24) \quad \frac{\partial \mathfrak{L}_s}{\partial (S_i)_{kl}} = - \sum_{m \neq n} \frac{(e_{i,m})_k (e_{i,n})_l}{\sigma_{i,n} - \sigma_{i,m}} (2 - \delta_{kl}) (\lambda_{E;i} \cdot e_{i,m})(t, x_r) + (\lambda_{S;i})_{kl}(t, x_r),$$

where $e_{i,n} = E_i$, $\sigma_{i,n} = \Sigma_i$,

$$(2.25) \quad \frac{\partial \mathfrak{L}_s}{\partial (p_i)_l} = \sum_{m,n} (\lambda_{S;i})_{mn}(\bar{x}_r) (\delta_{lm} (p_i)_n + (p_i)_m \delta_{ln})(\bar{x}_r) W_\mu(\bar{0}) + (\lambda_{p;i})_l(\bar{x}_r),$$

$$(2.26) \quad \frac{\partial \mathfrak{L}_s}{\partial u_i} = \int_{\Sigma_r} \delta(x - x_r) \nabla_{x_r} \cdot \lambda_{p;i}(t, x_r);$$

the right-hand side of the final equation needs to be added to (2.8).

Checkpointing. The final value problem is solved reverse in time. However, in our computations we solve for $u_i^b(t, x) = \lambda_{u;i}(T - t, x)$, which satisfies

$$(2.27) \quad [c^{-2}(x) \partial_t^2 - \Delta] u_i^b(t, x) = \int_{\Sigma_r} \delta(x - x_r) \int_{-T_{\max}}^{T_{\max}} \lambda_{C;i}(\Delta t, x_r) d_i(T - t + \Delta t, x_r) \, d\Delta t \, dx_r,$$

$$(2.28) \quad u_i^b(0, x) = 0, \quad \partial_t u_i^b(0, x) = 0,$$

which is an initial value problem, forward in time. For the computation of the gradient in (2.17) and (2.19) we use checkpointing [19, 44].

3. Computational experiments. We carry out three basic experiments to test our functional: (i) the reconstruction of the shape of a curved interface, (ii) the recovery of interval wavespeed, and (iii) the joint reconstruction of interface shape and interval wavespeed. We restrict our experiments to dimension two. We build a model which is 10 km wide and 4 km deep (including a 1 km padding while incorporating a perfect matching layer), consisting of two layers. The top layer has smoothly varying wavespeed ranging from 1.5 km/s to 3.0 km/s. The bottom layer has a wavespeed of 3.3 km/s and will be kept fixed during iterations. The data consist of 45 shots with a 200 m spacing; the acquisition surface is positioned at 1 km depth in the computational domain. As the source signature we use a Ricker wavelet with a peak frequency of 10 Hz. We observe the wavefield at 182 receivers with a 50 m spacing. We consider different models for the different experiments. Throughout, we assume the source to be known.

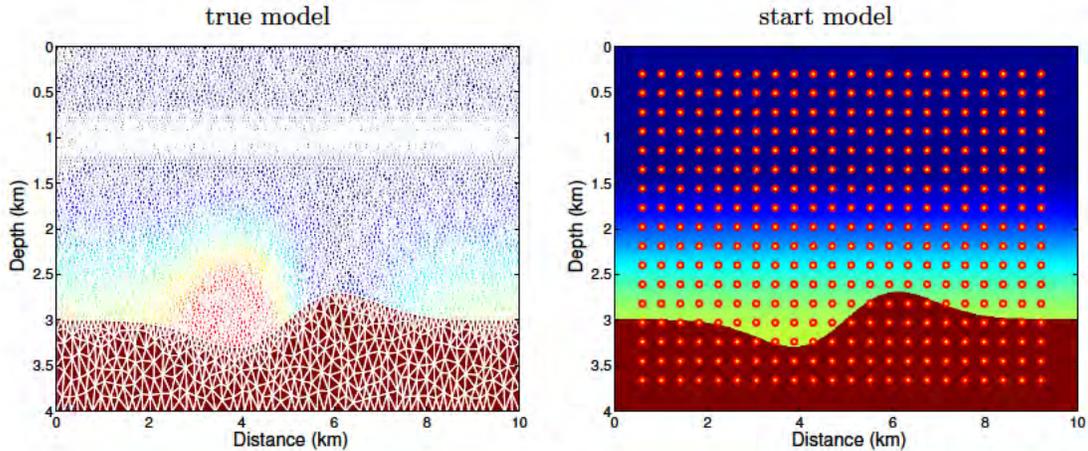


FIG. 3. Wavespeed model designed for testing the recovery of interval wavespeed. Left: the true model with a body-fitting mesh that will remain the same during the iterations. Right: the initial model; the dots indicate the centers of spline basis functions used in the parametrization.

3.1. Interval wavespeed. Here, the shape of the interface is fixed. The relative difference in wavespeed between the initial (one-dimensional profile) model and the true model is 75 % (see Figure 3). The reflections in the data generated in these two models are more than a wavelength apart. The shape of the interface and the heterogeneity with high and low wavespeeds yield the formation of caustics. The initial and the true models differ significantly in this respect.

We expand the interval wavespeed into 22×17 B-spline basis functions; see Figure 3. We use the cross-correlation power functional and compute the gradient according to (2.17). We use the BFGS method for computing the updates. The iterations gradually recover the heterogeneity while the modelled wavefield approaches the synthetic data both in phase and amplitude; see Figure 4.

3.2. Strong reflector: shape. Here the function describing the wavespeed in the top layer depends on depth only, and remains unchanged during the iterations. The wavespeed function in the bottom layer is a constant. The iteration updates the interface between these two layers, and, implicitly, which wavespeed function is to be used at any given point. The initial interface is flat. The reflection in the data generated in the initial model does not contain a caustic while the reflection in the data generated in the true model does. Thus the initial and the true models differ significantly.

We use 21 cubic spline polynomials to (sparsely) represent the shape of the strong reflector, which is assumed to be smooth. The central points of the polynomials are marked as red dots in Figure 5-A and Figure 7-A. We use the cross-correlation power functional and compute the gradient according to (2.19). We use the BFGS method for computing the updates. Each update of the interface shape is accompanied with generating a new body-fitted triangular mesh. The iterations gradually deform the interface approaching its true shape. The corresponding modelled data gradually show the formation of a caustic; see Figure 5.

3.3. Piecewise smooth wavespeed. Here, we test the joint inversion for interval wavespeed and interface shape. This is more challenging, as a perturbation of the smooth interval wavespeed will distort the gradient of shape optimization and thus the update of the interface. Also, a perturbation of interface shape can distort the gradient of wavespeed optimization and the corresponding update. We include the term (2.22) containing the structure tensor residual in the functional. An example of the effect of this term on the shape optimization is shown in Figure 6. We note a slight improvement in the update.

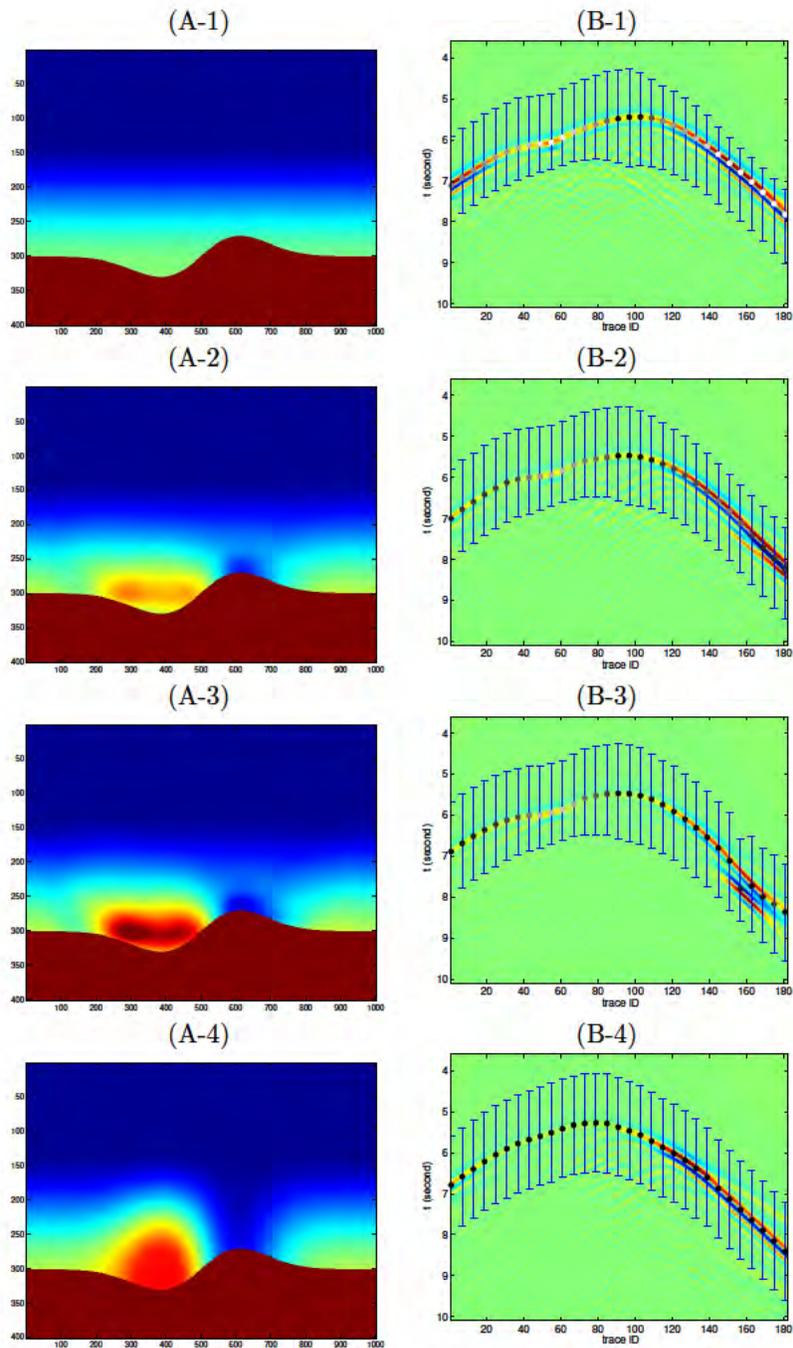


FIG. 4. Wavespeed model and data (single shot). We indicate the intervals, $[-T_{\max}, T_{\max}]$, used in the cross correlation by blue bars. The darkness of gray points at center of each window corresponds to the cross-correlation amplitude between synthetic data and observation data at this trace, i.e., the measurement of data misfit. The dots, for example, in (B-4) are totally dark because the observation data matches itself. (A-1) Initial model; (B-1) data generated in the initial model. (A-2) Updated model after the first iteration; (B-2) data generated in the updated model after the first iteration. (A-3) Updated model after the second iteration; (B-3) data generated in the updated model after the second iteration. (A-4) True model; (B-4) data generated in the true model.

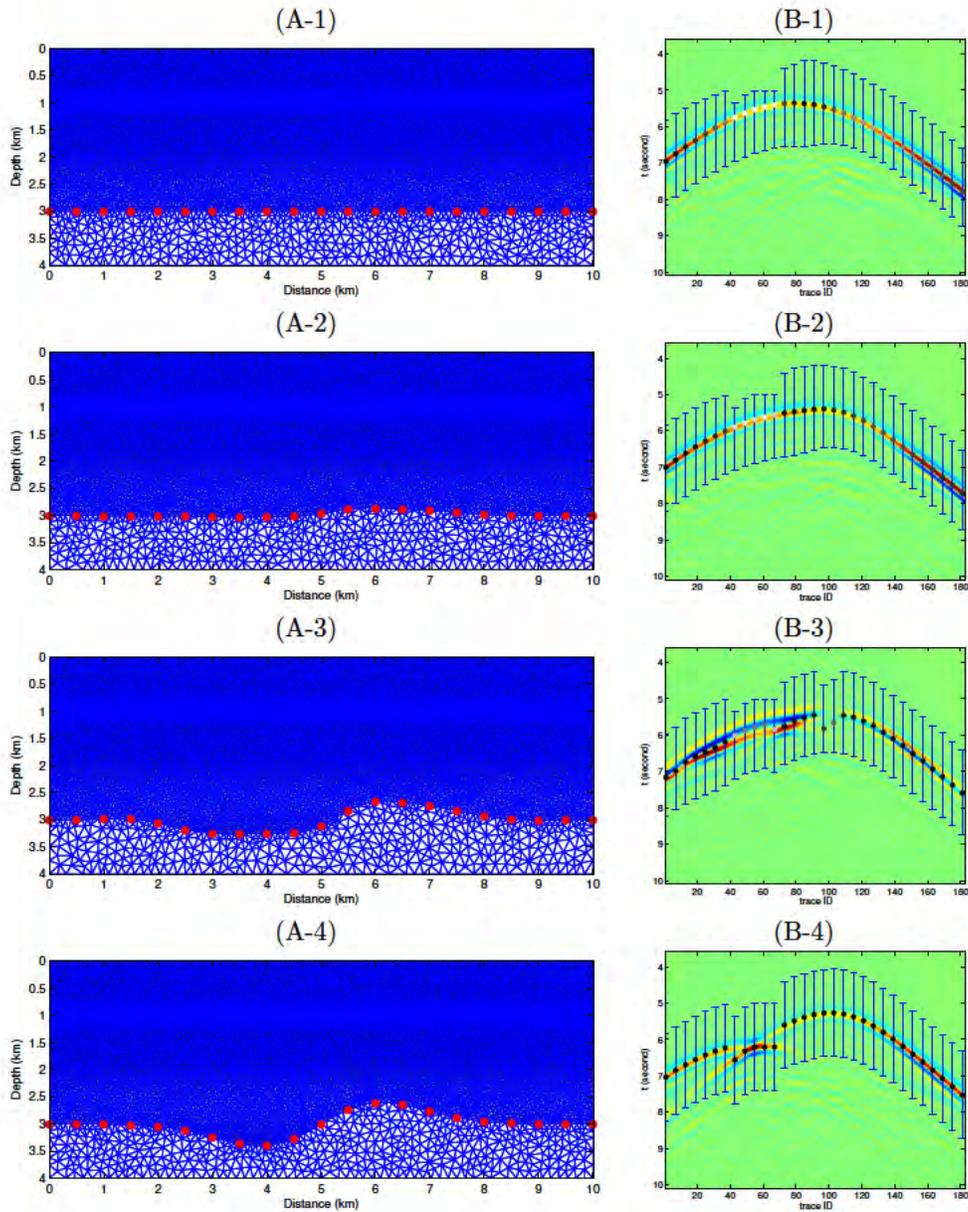


FIG. 5. *Wavespeed model and data (single shot).* The red dots indicate the locations of centers of cubic polynomial basis functions used to represent the interface. Blue bars in the right figures indicates the cross-correlation window $[-T_{\max}, T_{\max}]$; the darkness of gray points at center of each window corresponds to the cross-correlation amplitude between synthetic data and observation data at this trace, i.e., the measurement of data misfit. The dots, for example, in (B-4) are totally dark because the observation data matches itself. (A-1) Initial model; (B-1) data generated in the initial model. (A-2) Updated model after the first iteration; (B-2) data generated in the updated model after the first iteration. (A-3) Updated model after the third iteration; (B-3) data generated in the updated model after the third iteration. (A-4) True model; (B-4) data generated in the true model.

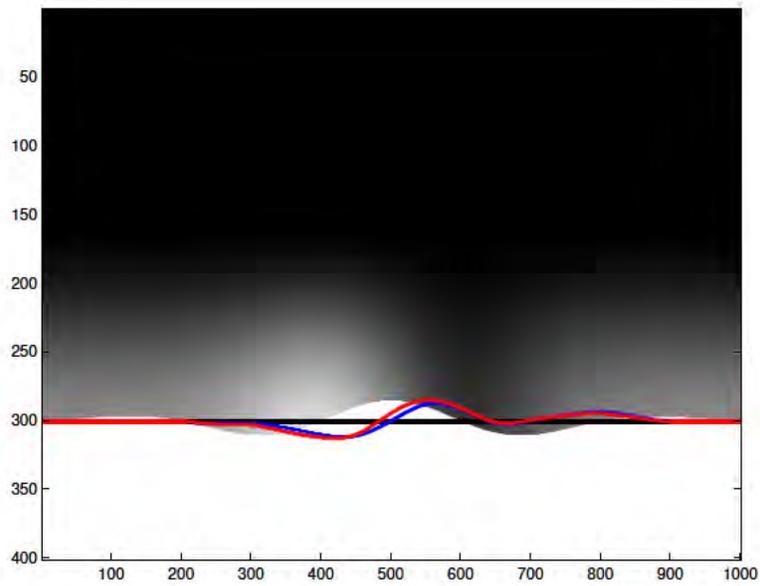


FIG. 6. *The first iteration of the shape optimization using the cross-correlation power criterion with and without (directional) structure tensor regularization. The background illustrates true model. The black line indicates the initial shape. The blue line shows the updated interface shape from the cross-correlation power gradient. The red line shows the updated interface shape from the cross-correlation power and structure tensor residual combined.*

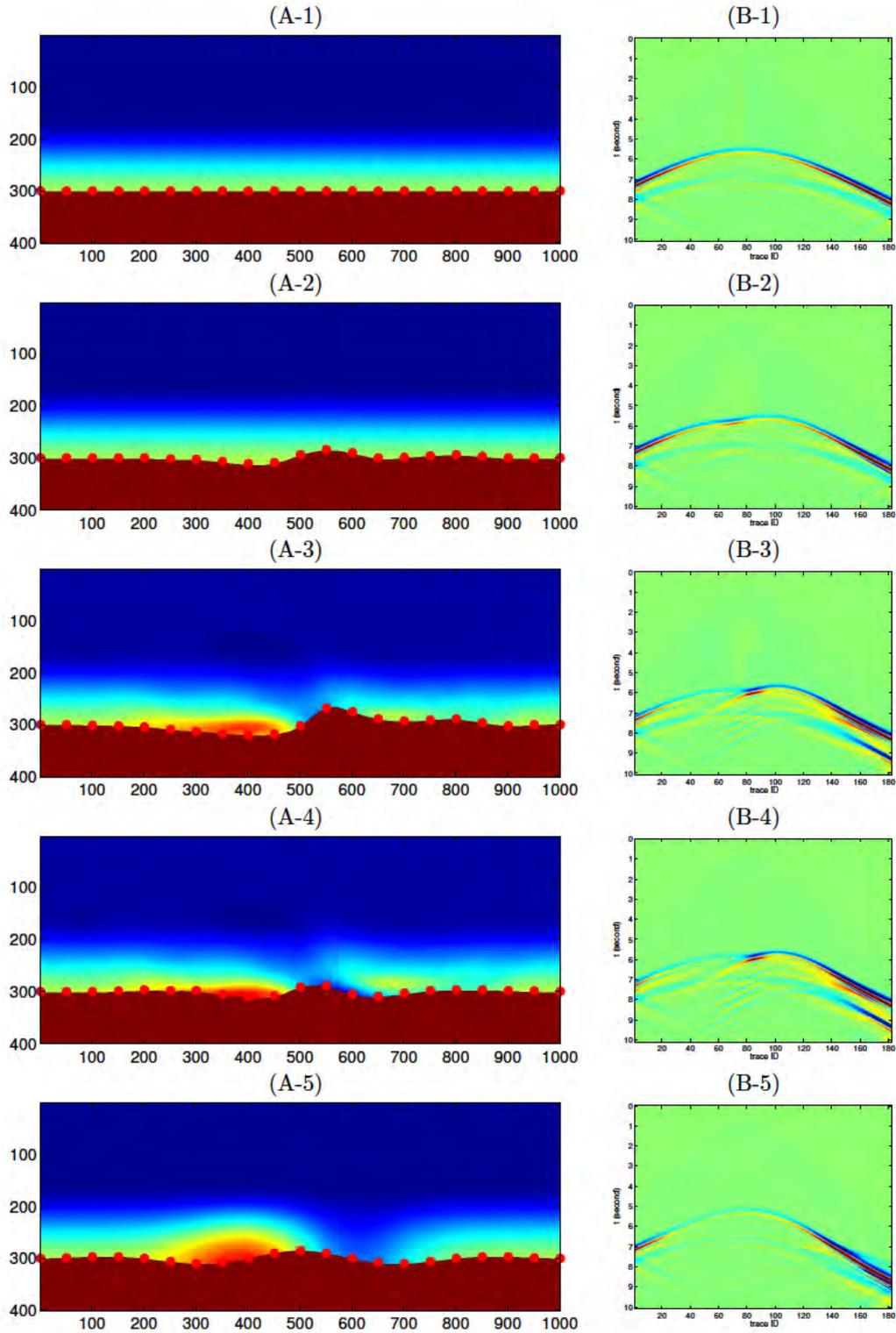


FIG. 7. Wavespeed model and data (single shot). The red dots indicate the locations of the center of cubic polynomial basis functions that are used to represent the interface. (A-1) Initial model; (B-1) data generated in the initial model. (A-2) Updated model after the first iteration; (B-2) data generated in the updated model after the first iteration. (A-3) Updated model after the third iteration; (B-3) data generated in the updated model after the third iteration. (A-4) Updated model after the fifth iteration; (B-4) data generated in the updated model after the fifth iteration. (A-5) True model; (B-5) data generated in the true model.

The test model designed for the joint inverse problem combines the features of both previous test models. The interface in the true model is described by a smooth function, $150 \sin(\pi x/200) \exp(-(\pi(x-500)/1000)^2)$. There are two smooth perturbation up to 60 % in the interval wavespeed. The phase misfit between synthetic and modelled data is more than one period. Here, the sensitivity of the data to the interface shape appears to exceed the sensitivity to the wavespeed. We alternate the updating for interface shape and wavespeed. Model updates are illustrated in Figure 7, as well as the corresponding modelled data.

4. Discussion. We presented the formulation of an iterative method for the inverse boundary value problem for the wave equation (full waveform inversion) at high frequencies based on a novel misfit functional. We synthesize directional (Gaussian) wavepackets from ‘point’ sources, and generate a functional given by the cross correlation (in time) power regularized by a structure tensor residual with Gaussian window functions. We use a subset of (multiple) reflections in the data which is controlled by window functions in our misfit functional and data simulated in the initial/current model, as well as slopes in the data. We constrain our models to models with structure, namely, piecewise smooth with discontinuities (sometimes referred to as blocky).

We carried out computational experiments, revealing the basic properties of the functional. The method we developed is in some way derived from wave equation tomography with multiple reflections and as such has natural limitations. Refining the structure in the iterative method beyond the class of models considered here is a topic of ongoing research.

5. Acknowledgements. This research was supported in part by the members of GMIG at Purdue University, BGP, ExxonMobil, PGS, Statoil and Total.

REFERENCES

- [1] V. Akcelik, G. Biros, and O. Ghattas. Parallel multiscale gauss-newton-krylov methods for inverse wave propagation. In *Supercomputing, ACM/IEEE 2002 Conference*, pages 41–41, 2002.
- [2] A. Bamberger, G. Chavent, C. Hemon, and P. Lailly. Inversion of normal incidence seismograms. *Geophysics*, 47(5):757–770, 1982.
- [3] A. Bamberger, G. Chavent, and P. Lailly. Une application de la théorie du contrôle à un problème inverse de sismique. *Annales de Géophysique*, 33:689 – 690, 1977.
- [4] E. Beretta, M. V. de Hoop, and L. Qiu. Lipschitz stability of an inverse boundary value problem for time harmonic elastic waves, part i: recovery of the density. GMIG2013.
- [5] E. Beretta, M. V. de Hoop, L. Qiu, and O. Scherzer. Inverse boundary value problem for the helmholtz equation with multi-frequency data. GMIG2013.
- [6] W. B. Beydoun and A. Tarantola. First born and rytov approximations: Modeling and inversion conditions in a canonical example. *The Journal of the Acoustical Society of America*, 83(3):1045–1055, 1988.
- [7] F. Billette and G. Lambaré. Velocity macro-model estimation from seismic reflection data by stereotomography. *Geophysical Journal International*, 135(2):671–690, 1998.
- [8] F. Boschetti. *Application of Genetic Algorithms to the Inversion of Geophysical Data*. PhD thesis, University of Western Australia, 1995.
- [9] R. Brossier, S. Operto, and J. Virieux. Seismic imaging of complex onshore structures by 2d elastic frequency-domain full-waveform inversion. *GEOPHYSICS*, 74(6):WCC105–WCC118, 2009.
- [10] R. Brossier and P. Roux. Hierarchical waveform inversion with double beamforming. In *SEG Technical Program Expanded Abstracts*, pages 2389–2394, 2011.
- [11] M. Cara and J. J. Lévêque. Waveform inversion using secondary observables. *Geophysical Research Letters*, 14(10):1046–1049, 1987.
- [12] G. Chavent. *Identification of function parameters in partial differential equations, in Identification of parameter distributed systems*. eds Goodson, R.E. & Polis, New-York, ASME, 1974.
- [13] P. Chen, L. Zhao, and T. H. Jordan. Full 3d tomography for the crustal structure of the los angeles region. *Bulletin of the Seismological Society of America*, 97(4):1094–1120, 2007.
- [14] R. G. Clapp, B. Biondi, and J. F. Claerbout. Incorporating geologic information into reflection tomography. *Geophysics*, 69(2):533–546, 2004.
- [15] M. V. de Hoop, L. Qiu, and O. Scherzer. A convergence analysis of a multi-level projected steepest descent iteration for nonlinear inverse problems in Banach spaces subject to stability constraints. *ArXiv e-prints*, June 2012.
- [16] F. Delprat-Jannaud and P. Lailly. What information on the earth model do reflection travel times provide? *Journal of Geophysical Research: Solid Earth*, 97(B13):19827–19844, 1992.

- [17] N. Downey, P. Routh, and Y. H. Cha. Random-beam full-wavefield inversion. In *SEG Technical Program Expanded Abstracts*, pages 2423–2427, 2011.
- [18] O. Gauthier, J. Virieux, and A. Tarantola. Two-dimensional nonlinear inversion of seismic waveforms; numerical results. *Geophysics*, 51(7):1387–1403, 1986.
- [19] A. Griewank and A. Walther. Algorithm 799: revolve: an implementation of checkpointing for the reverse or adjoint mode of computational differentiation. *ACM Trans. Math. Softw.*, 26(1):19–45, Mar. 2000.
- [20] K. Gröchenig and M. Ruzhansky. New frames of gaussian wave packets following parabolic scaling of phase space. preprint (2013).
- [21] A. Guitton, G. Ayeni, and E. DĀaz. Constrained full-waveform inversion by model reparameterization. *Geophysics*, 77(2):R117–R127, 2012.
- [22] Z. Guo and M. V. de Hoop. Adaptive domain partitioning and deformation of representations for the wavespeed using unstructured meshes in full waveform inversion. preprint (2013).
- [23] D. Hale. Image-guided blended neighbor interpolation of scattered data. In *SEG Technical Program Expanded Abstracts*, pages 1127–1131, 2009.
- [24] S. Holman and M. V. de Hoop. Linearization and microlocal analysis of reflection tomography. preprint (2013).
- [25] F. Jurado, D. Sinoquet, and A. Ehinger. 3d reflection tomography designed for complex structures. In *SEG Technical Program Expanded Abstracts*, pages 711–714, 1996.
- [26] J.-M. Kang, W. Shin, C.-G. Park, and D.-I. Lee. Extended blp security model based on process reliability for secure linux kernel. In *Dependable Computing, 2001. Proceedings. 2001 Pacific Rim International Symposium on*, pages 299–303, 2001.
- [27] P. Kolb, F. Collino, and P. Lailly. Prestack inversion of a 1d medium. *Proceedings of IEEE*, 74:498 – 506, 1986.
- [28] P. Lailly. The seismic inverse problem as a sequence of before stack migrations: Conference on inverse scattering, theory and application,. In *Society of Industrial and Applied Mathematics, Expanded Abstracts, 206–220.*, 1983.
- [29] P. Lailly and D. Sinoquet. Smooth velocity models in reflection tomography for imaging complex geological structures. *Geophysical Journal International*, 124(2):349–362, 1996.
- [30] G. Lambaré, M. Alerini, R. Baina, and P. Podvin. Stereotomography: a semi-automatic approach for velocity macromodel estimation. *Geophysical Prospecting*, 52(6):671–681, 2004.
- [31] J. L. Lions. *Some aspects of the optimal control of distributed parameter systems*. Society for Industrial and Applied Mathematics (Philadelphia), 1972.
- [32] M. D. Long, M. V. De Hoop, and R. D. Van Der Hilst. Wave-equation shear wave splitting tomography. *Geophysical Journal International*, 172(1):311–330, 2008.
- [33] A. Malcolm and D. P. Nicholls. Operator expansions and constrained quadratic optimization for interface reconstruction: Impenetrable periodic acoustic media. 2012.
- [34] W. Mulder and R. Plessix. Full waveform tomography and nonlinear migration. In *68th EAGE Conference & Exhibition, Expanded Abstracts, A038.*, 2006.
- [35] G. Pratt, C. Shin, and Hicks. Gauss-newton and full newton methods in frequency-space seismic waveform inversion. *Geophysical Journal International*, 133(2):341–362, 1998.
- [36] R. G. Pratt and N. R. Goult. Combining wave-equation imaging with travelttime tomography to form high-resolution images from crosshole data. *Geophysics*, 56(2):208–224, 1991.
- [37] R. G. Pratt, L. Sirgue, B. Hornby, and J. Wolfe. Cross-well waveform tomography in fine-layered sediments meeting the challenges of anisotropy. In *70th Conference & Technical Exhibition, EAGE, Extended Abstracts.*, 2008.
- [38] R. G. Pratt, Z.-M. Song, P. Williamson, and M. Warner. Two-dimensional velocity models from wide-angle seismic data by wavefield inversion. *Geophysical Journal International*, 124(2):323–340, 1996.
- [39] R. G. PRATT and M. H. WORTHINGTON. Inverse theory applied to multi-source cross-hole tomography. *Geophysical Prospecting*, 38(3):287–310, 1990.
- [40] C. Ravaut, S. Operto, L. Imbrota, J. Virieux, A. Herrero, and P. Dell’Aversana. Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: application to a thrust belt. *Geophysical Journal International*, 159(3):1032–1056, 2004.
- [41] M. Schoenauer, A. Ehinger, and B. Braunschweig. Non-parametric identification of geological models. In *Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on*, pages 136–141, 1998.
- [42] R. M. Shipp and S. C. Singh. Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data. *Geophysical Journal International*, 151(2):325–344, 2002.
- [43] P. Stefanov and G. Uhlmann. Linearizing non-linear inverse problems and its applications to inverse backscattering. *J. Funct. Anal*, pages 2842–2866, 2009.
- [44] W. Symes. Reverse time migration with optimal checkpointing. *GEOPHYSICS*, 72(5):SM213–SM221, 2007.
- [45] W. W. Symes. Migration velocity analysis and waveform inversion. *Geophysical Prospecting*, 56(6):765–790, 2008.
- [46] A. Tarantola. Inversion of seismic reflection data in the acoustic approximation. *GEOPHYSICS*, 49(8):1259–1266, 1984.
- [47] A. Tarantola. *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Amsterdam: Elsevier, 1987.
- [48] T. Van Leeuwen and W. Mulder. Velocity analysis based on data correlation. *Geophysical Prospecting*, 56(6):791–

- 803, 2008.
- [49] T. Van Leeuwen and W. A. Mulder. A correlation-based misfit criterion for wave-equation traveltime tomography. *Geophysical Journal International*, 182(3):1383–1394, 2010.
 - [50] D. Vigh and E. Starr. 3d prestack plane-wave, full-waveform inversion. *GEOPHYSICS*, 73(5):VE135–VE144, 2008.
 - [51] J. Virieux and S. Operto. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26, 2009.
 - [52] H. Wang, S. Singh, H. Jian, and H. Calandra. Integrated inversion of long and short wavelength velocity structures using combined wave equation tomography and fwi. In *74th EAGE Conference & Exhibition*, 2012.
 - [53] Y. Wang and Y. Rao. Reflection seismic waveform tomography. *Journal of Geophysical Research: Solid Earth*, 114(B3):n/a–n/a, 2009.