Gradient computation for elastic full-waveform inversion in 2D VTI media

Nishant Kamath & Ilya Tsvankin
Center for Wave Phenomena, Colorado School of Mines

ABSTRACT

Full-waveform inversion (FWI) has proved effective in significantly improving the spatial resolution of seismic models. However, it is implemented mostly for isotropic media and the applications to anisotropic models are often limited to acoustic approximations. In a previous publication, we introduced a methodology for performing FWI of multicomponent (PP and PS) reflection data from a stack of horizontal, homogeneous VTI (transversely isotropic with a vertical symmetry axis) layers. Taking into account both anisotropy and elasticity in the description of the reflected wavefield makes it possible to resolve the interval Thomsen parameters and constrain the depth scale of the model. Here, we develop a foundation for extending elastic FWI to laterally heterogeneous VTI media using the adjoint state method. The main focus is on deriving the gradients of the objective function with respect to the model parameters. To test the algorithm, we introduce Gaussian anomalies in the Thomsen parameters of a homogeneous VTI medium and study the influence of the size of the anomaly, the source-receiver configuration, and the source polarization on the gradients. The model update for FWI can be produced by an appropriate scaling of the gradient using the steepest descent or similar methods.

1 INTRODUCTION

Full-waveform inversion (FWI) is a technique for estimating subsurface properties by using entire seismic waveforms recorded at the surface or in a borehole. Depending on the problem and availability of forward-modeling algorithms, FWI can be performed in the time domain (Kolb et al., 1986; Gauthier, 1986; Mora, 1987; Bunks et al., 1995) or frequency domain (Song and Williamson, 1995; Song et al., 1995; Pratt, 1999; Pratt and Shipp, 1999). Evaluation of the gradient of the objective function is often based on the adjoint state method, as described in Tarantola (1984a), Fichtner et al. (2006), and Liu and Tromp (2006).

Köhn et al. (2012) discuss the influence of parameterization on elastic isotropic FWI and conclude that describing the model in terms of P- and S-wave velocities and density gives better results than using impedances. Liu and Tromp (2006) derive the gradients of the objective function for earthquake data with respect to the parameters (bulk and shear moduli, density, source location, and source function) of an elastic earth model. FWI has been extended to anisotropic media, but typically in the acoustic approximation (Plessix and Rynja, 2010; Gholami et al., 2011; Plessix and Cao, 2011; Shen, 2012). Such "anisotropic acoustic" algorithms, however, suffer from trade-offs between the model parameters. Elastic FWI of synthetic multicomponent surface seismic data (consisting of both diving waves and reflections) for VTI media is performed by Lee et al. (2010), but suboptimal parameterization in terms of the stiffness coefficients causes ambiguity in their results.

In our previous work (Kamath and Tsvankin, 2012; hereafter, referred to as Paper I), we invert multicomponent reflection data from a horizontally layered VTI model for the the interval Thomsen parameters − the P- and S-wave vertical velocities (V_{P0} and V_{S0}) and anisotropy parameters ϵ and δ. Including density as an unknown parameter makes the objective function highly nonlinear, thereby causing the algorithm to get trapped in local minima. Therefore, the densities in all layers are fixed at the correct values. Although PP-waves alone may be sufficient to resolve V_{P0}, V_{S0}, ϵ, and δ, stable parameter estimation for layers at depth requires employing long-offset data (with the spreadlength-to-depth ratio reaching at least two) or the addition of PS-waves. Inversion of multicomponent data benefits from using a multiscale approach, which helps reduce the sensitivity to the choice of the initial model (Bunks et al., 1995).

Here we build the foundation for extending FWI to laterally heterogeneous VTI media. The model is parameterized in terms of V_{P0}, V_{S0}, and the P-wave NMO ($V_{\mathrm{nmo},P}$) and horizontal ($V_{\mathrm{hor},P}$) velocities. To compute the gradient of the objective function, we adapt the results of Liu and Tromp (2006) obtained with the adjoint-state method. The algorithm is applied to VTI models with Gaussian anomalies in the Thomsen parameters inserted into a homogeneous background.

2 METHODOLOGY

2.1 Full-waveform inversion for VTI media

FWI in the time domain is designed to minimize the following objective function:

$$\mathcal{F} = \frac{1}{2}\sum_{r=1}^{\mathrm{N}}\int_0^T \|\mathbf{u}(\mathbf{x}_r,t)-\mathbf{d}(\mathbf{x}_r,t)\|^2 \mathrm{d}t\ , \qquad (1)$$

where N is the number of receivers, T is the trace length, $\mathbf{u}(\mathbf{x}_r,t)$ is the modeled displacement, and $\mathbf{d}(\mathbf{x}_r,t)$ is the observed displacement at receiver location $\mathbf{x}_r$. Since the relationship between the data and the model is nonlinear, the inversion is performed iteratively, and the model update at each iteration can be found as:

$$\mathbf{\Delta m} = [\mathbf{J^T J}]^{-1}\,\mathbf{J^T \Delta d}, \qquad (2)$$

where $\mathbf{J}$ is the Fréchet derivative matrix obtained by perturbing each model parameter, $\mathbf{J^T J}$ is the approximate Hessian matrix, $\mathbf{T}$ denotes transposition, and $\mathbf{\Delta d}$ is the difference between the observed data and those computed for a trial model.

In Paper I, we use PP reflections or a combination of PP and PS events generated for a horizontally layered VTI model to invert for the interval parameters V_{P0}, V_{S0}, ϵ, and δ. For the purpose of inversion, it is convenient to use quantities that can be constrained by the data and have the same units. Hence, instead of ϵ and δ we invert for the P-wave NMO velocity ($V_{\mathrm{nmo},P} = V_{P0}\sqrt{1+2\delta}$) and the horizontal velocity ($V_{\mathrm{hor},P} = V_{P0}\sqrt{1+2\epsilon}$). If the number of layers (which is fixed during the inversion) is not large, it is possible to compute the Fréchet matrix and the approximate Hessian explicitly by perturbing each model parameter.

2.2 Inverse problem in 2D

In the case of laterally heterogeneous media, computation of the Fréchet derivatives becomes prohibitively expensive since it involves calculating as many forward models at each iteration as the number of parameters (typically defined on a grid). It is more practical to calculate the gradient ($\mathbf{J^T \Delta d}$ in equation 2) of the objective function with the adoint state method, which has been widely used in FWI (Tarantola, 1984b; Plessix, 2006; Liu and Tromp, 2006; Fichtner et al., 2006). The model update, which is a scaled version of the gradient, is then calculated using steepest-descent or conjugate-gradient algorithms. Alternatively, either the so-called BFGS (Broyden-Fletcher-Goldfarb-Shanno) method or its limited-memory equivalent, the L-BFGS method (both are quasi-Newton techniques), can be employed to scale the gradient by the inverse of an approximate Hessian (Virieux and Operto, 2009).

The adjoint state method involves computing the so-called adjoint wavefield to obtain the gradient. Because the wave equation is self-adjoint, it can be solved for the adjoint wavefield with the data residuals treated as sources. The residuals at each time step are injected "backward in time" (i.e., starting from the last sample), which is commonly described as back-propagation of data residuals. For 2D multicomponent data, the vertical and horizontal displacement components of the data residuals should be injected into the medium simultaneously. The gradient is obtained by applying the imaging condition to the spatial derivatives of the forward and adjoint wavefields.

Here, we assume that the properties of the VTI medium vary in 2D and consider only in-plane polarized waves. Hence the model is described by four stiffness coefficients (written in the Voigt notation): C_{11}, C_{33}, C_{13}, and C_{55}. However, it is certain combinations of the stiffnesses that control traveltimes and amplitudes of seismic waves (Tsvankin, 2012). In particular, description of wave propagation and inversion of seismic data can be facilitated by employing Thomsen parameters and their simple combinations (e.g., the anellipticity coefficient η). Lee et al. (2010), who parameterize the VTI model in terms of the stiffnesses, are unable to resolve the coefficient C_{13}, likely because of the tradeoff between C_{13} and C_{55} in P-wave kinematic signatures. In Paper I we were able to constrain the relevant Thomsen parameters (V_{P0}, V_{S0}, ϵ, and δ), although the algorithm operated with the vertical, horizontal and NMO velocities. Here, we parameterize the model in terms of the squared velocities V_{P0}^2, V_{S0}^2, $V_{\mathrm{nmo},P}^2$, and $V_{\mathrm{hor},P}^2$. Squaring the velocities is a matter of convenience because it simplifies the gradient expressions.

The gradients of the objective function (equation 1) with respect to the elements of the stiffness tensor are derived in Appendix A using the results of Liu and Tromp (2006):

$$\frac{\partial \mathcal{F}}{\partial c_{ijkl}} = -\int_0^T \frac{\partial u_i}{\partial x_j}\frac{\partial \psi_k}{\partial x_l}\,\mathrm{d}t\ , \qquad (3)$$

where $\mathbf{u}$ and $\boldsymbol{\psi}$ are the forward and adjoint wavefields, respectively. Using the chain rule, we can find the gradient for the squared velocities V_n^2 ($V_{P0}^2,\ V_{S0}^2,\ V_{\mathrm{nmo},P}^2,$ and $V_{\mathrm{hor},P}^2$):

$$\frac{\partial \mathcal{F}}{\partial V_n^2} = \sum_{ijkl}\frac{\partial \mathcal{F}}{\partial c_{ijkl}}\,\frac{\partial c_{ijkl}}{\partial V_n^2}. \qquad (4)$$

The stiffness coefficients are expressed in terms of the velocities in equations A18$-$A21. Combining equations 3, 4, and A18$-$A21 yields the gradients with respect to the squared velocities (equations A22$-$ A25).

Here, FWI is implemented in the time domain, primarily because the finite-difference modeling software available to us performs time-domain computations. The multiscale approach described in Bunks et al. (1995) and Paper I helps reduce the nonlinearity of the objective function and ensure that each updated solution lies in the basin of convergence. At first, diving waves are employed to update the low-wavenumber component of the model and obtain a close approximation for traveltimes of the recorded arrivals. Reflection events can then be used to increase the vertical and lateral resolution of the inverted model. Application of this methodology to 2D synthetic data from heterogeneous VTI models will be discussed in a sequel paper.

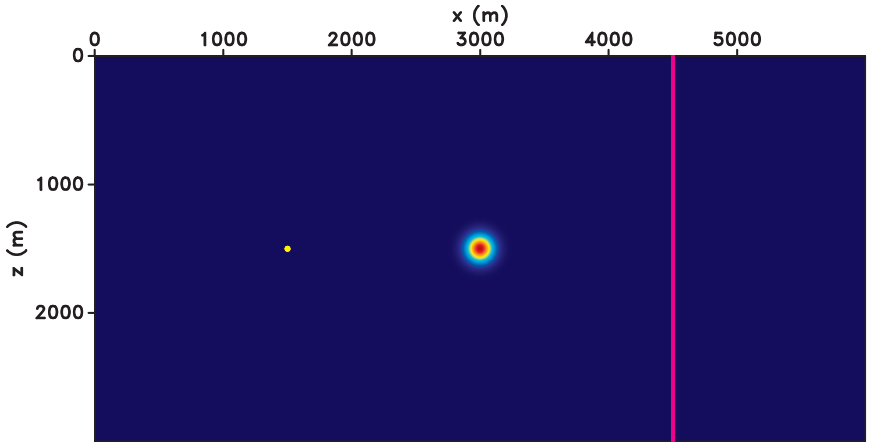
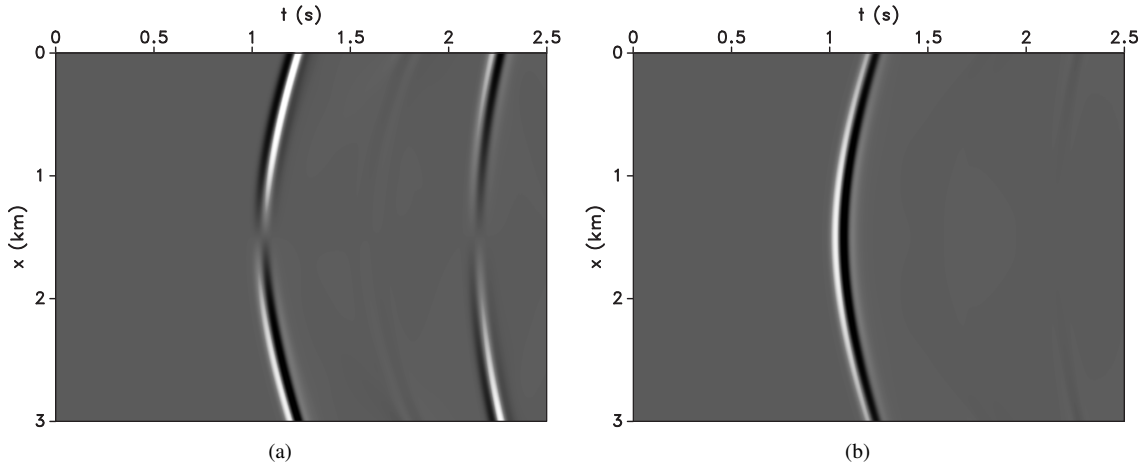


Figure 1. VTI model with a Gaussian anomaly (standard deviation $\sigma = 75\,\mathrm{m}$) in the anisotropy parameter ϵ. The background and maximum values of ϵ are 0.1 and 0.128, respectively. The other Thomsen parameters are spatially invariant: $V_{P0} = 3000\,\mathrm{m/s}$, $V_{S0} = 1500\,\mathrm{m/s}$, and $\delta = -0.1$. The white dot marks the source location and the white vertical line represents an array of receivers placed at each grid point (6.6 m apart).



(a) (b)

Figure 2. (a) Vertical and (b) horizontal displacements generated for the model in Figure 1.

3 GRADIENTS FOR GAUSSIAN ANOMALIES

Next, we perform tests for a simple synthetic model to verify the accuracy of the gradient computation. Because the initial stage of FWI involves diving waves, we compute the gradient for transmission experiments. The model includes Gaussian anomalies in the parameters V_{P0}, V_{S0}, ϵ, and δ inserted into a homogeneous VTI medium between a buried source and a vertical array of receivers (Figure 1). A point-displacement source with a peak frequency of 10 Hz polarized in the horizontal direction generates the data. The horizontal and vertical displacement components ("observed data") recorded by the receivers in the presence of the anomaly in ϵ are shown in Figure 2. Then the "modeled" data are generated in the background medium, and the adjoint source is obtained as the difference between the two wavefields (Figure 3).

The gradients with respect to the model parameters (squared velocities) for an anomaly in ϵ are calculated from equations A22−A25 (Figure 4). For the source-receiver geometry in Figure 1, waves travel close to the isotropy plane, and are influenced primarily by $V_{\mathrm{hor},P}$, which is a function of ϵ. Hence, the largest gradient is that for the velocity $V_{\mathrm{hor},P}$, which correctly identifies ϵ as the parameter that needs updating. However, the gradient with respect to V_{S0} is also significant, most likely because of the P-to-S conversion that takes place at the anomaly. The observed data include such conversions, while the wavefield modeled in the background medium does not. As a result, the obtained gradient resembles an "image" of the scatterer responsible for the conversion.

For a Gaussian anomaly with a larger radius (Figure 5), the gradient with respect to V_{S0} decreases (Figure 6), and there are fewer artifacts. This is likely due to the absence of the anomaly-induced PS conversions. In addition, the gradients for V_{P0} and $V_{\mathrm{nmo},P}$ are also small. In the presence of such a smooth anomaly in ϵ, we expect to see a change only

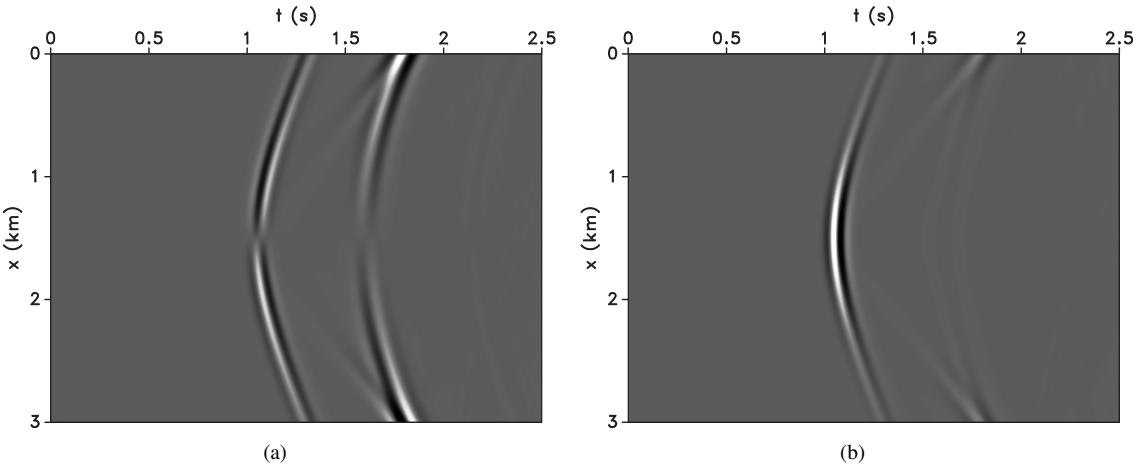


Figure 3. (a) Vertical and (b) horizontal components of the adjoint source for the model in Figure 1 obtained as the difference between the observed data from Figure 2 and the modeled data for the background medium (in the absence of the anomaly).

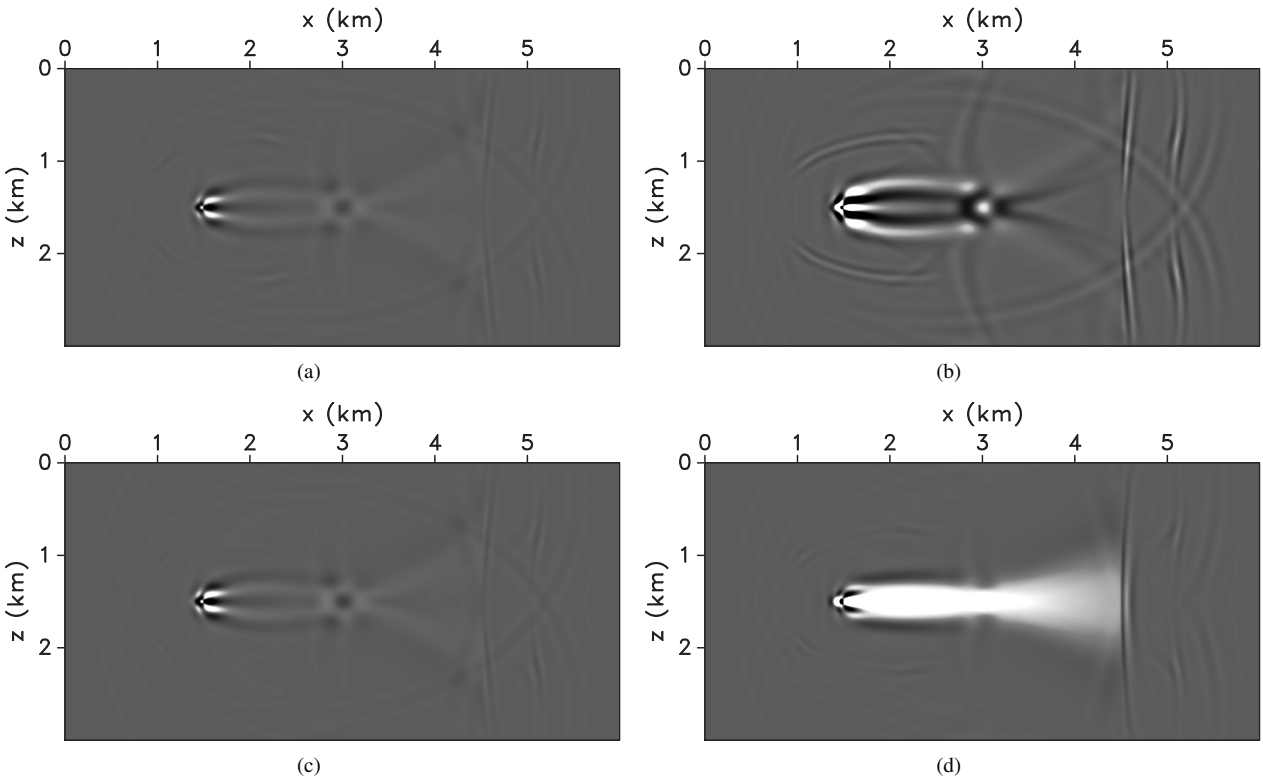


Figure 4. Gradients of the objective function (equation 1) with respect to (a) V_{P0}^2, (b) V_{S0}^2, (c) $V_{\mathrm{nmo},P}^2$, and (d) $V_{\mathrm{hor},P}^2$ for the model with an anomaly in ϵ (Figure 1). All four gradients here and on the plots below are plotted on the same scale.

in $V_{\mathrm{hor},P}$, which is indeed the case (Figure 6). As mentioned above, this is the desired result because $V_{\mathrm{hor},P}$ is the only parameter influenced by ϵ. All gradients in Figures 4 and 6 reach extremely large values at the source where the wavefield has the highest amplitude. This problem can be mitigated by applying an appropriate preconditioner (e.g., a Gaussian taper at the source). The large gradients along the horizontal raypath from the source indicate the nonuniqueness of the inverse problem for the configuration with a single source. Increasing the number of shots is likely to help focus the largest gradient values at the anomaly.

Similar tests performed for Gaussian anomalies in the pa-

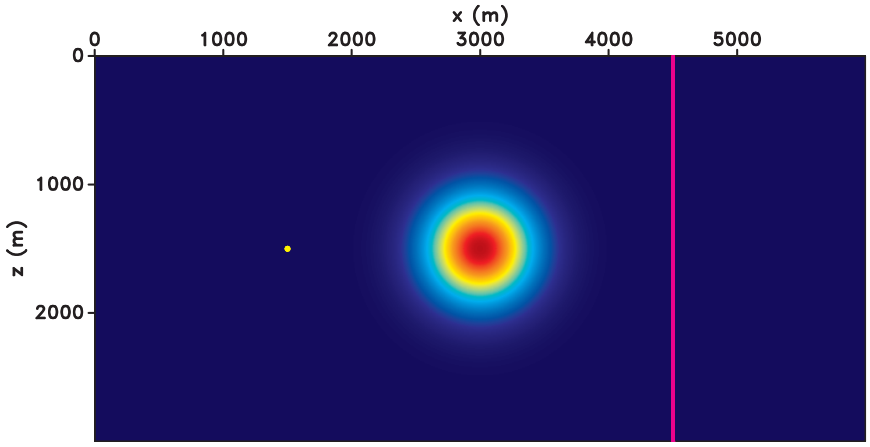


Figure 5. VTI model with a larger (in spatial extent) Gaussian anomaly in the anisotropy parameter ϵ. The standard deviation of the anomaly $\sigma = 300$ m. The background and maximum values of ϵ are 0.1 and 0.128, respectively. The other parameters are $V_{P0} = 3000\,\mathrm{m/s}$, $V_{S0} = 1500\,\mathrm{m/s}$, and $\delta = -0.1$.

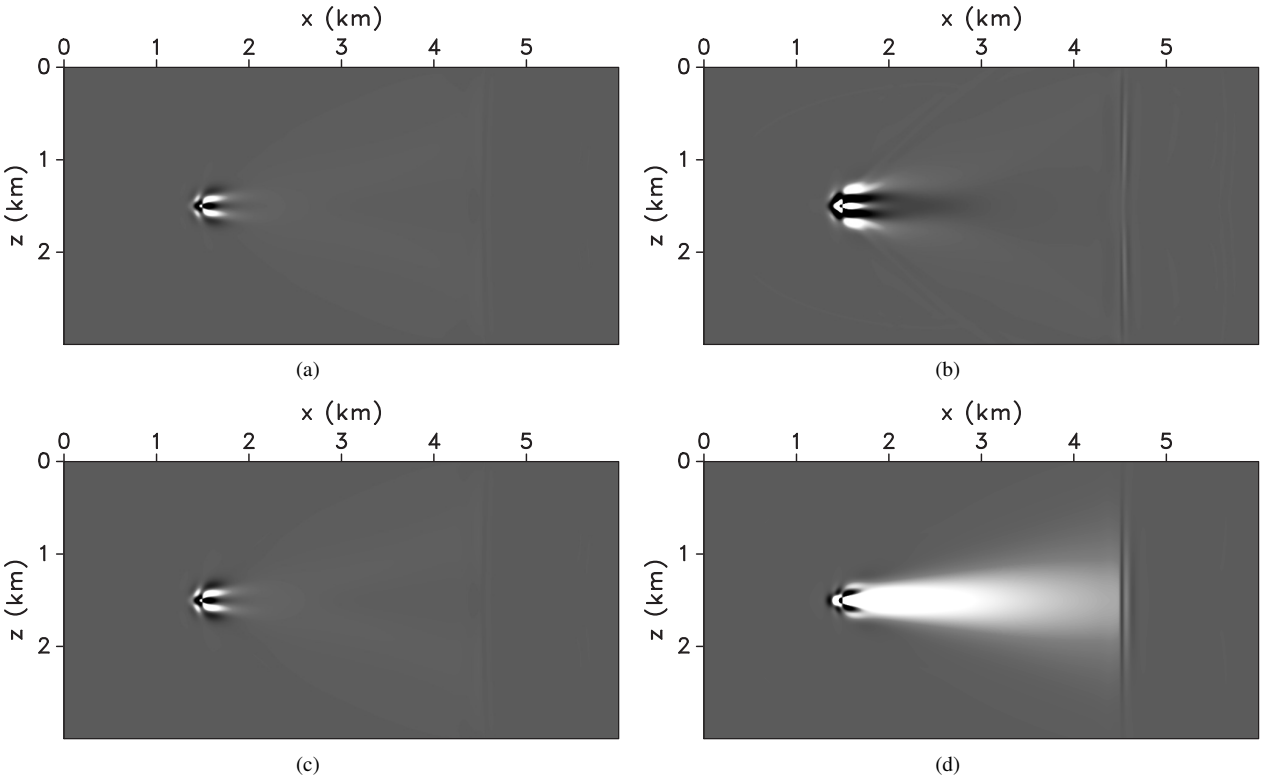


Figure 6. Gradients of the objective function with respect to (a) V_{P0}^2, (b) V_{S0}^2, (c) $V_{\mathrm{nmo},P}^2$, and (d) $V_{\mathrm{hor},P}^2$ for the model with an anomaly in ϵ from Figure 5.

rameter δ show that the gradients with respect to all four velocities are smaller compared to those obtained for the anomaly in ϵ.

Next we introduce an anomaly in the P-wave vertical velocity V_{P0} (Figure 7). As was the case for the model in Figure 5, there are no PS conversions because of the increased spatial extent of the anomaly. The largest gradient is the one with respect to the horizontal velocity (Figure 8), whereas the gradients with respect to V_{P0} and $V_{\mathrm{nmo},P}$ (which also depend on V_{P0}) are negligible. Although $V_{\mathrm{hor},P}$ depends on V_{P0}, updating the horizontal velocity without changing V_{P0} will result in an update in ϵ and steer the inversion away from recovering

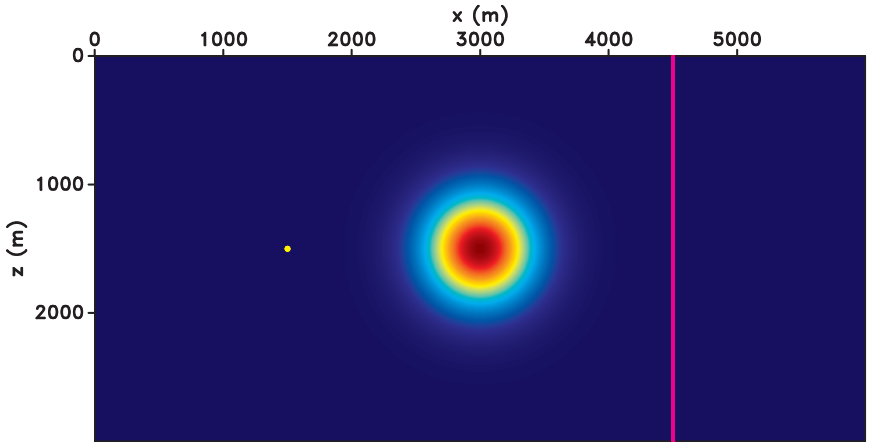


Figure 7. VTI model with a Gaussian anomaly in the P-wave vertical velocity V_{P0}. The background and maximum values of V_{P0} are 3000 m/s and 3140 m/s, respectively. The other parameters are $V_{S0} = 1500\,\mathrm{m/s}$, $\delta = -0.1$, and $\epsilon = 0.1$.

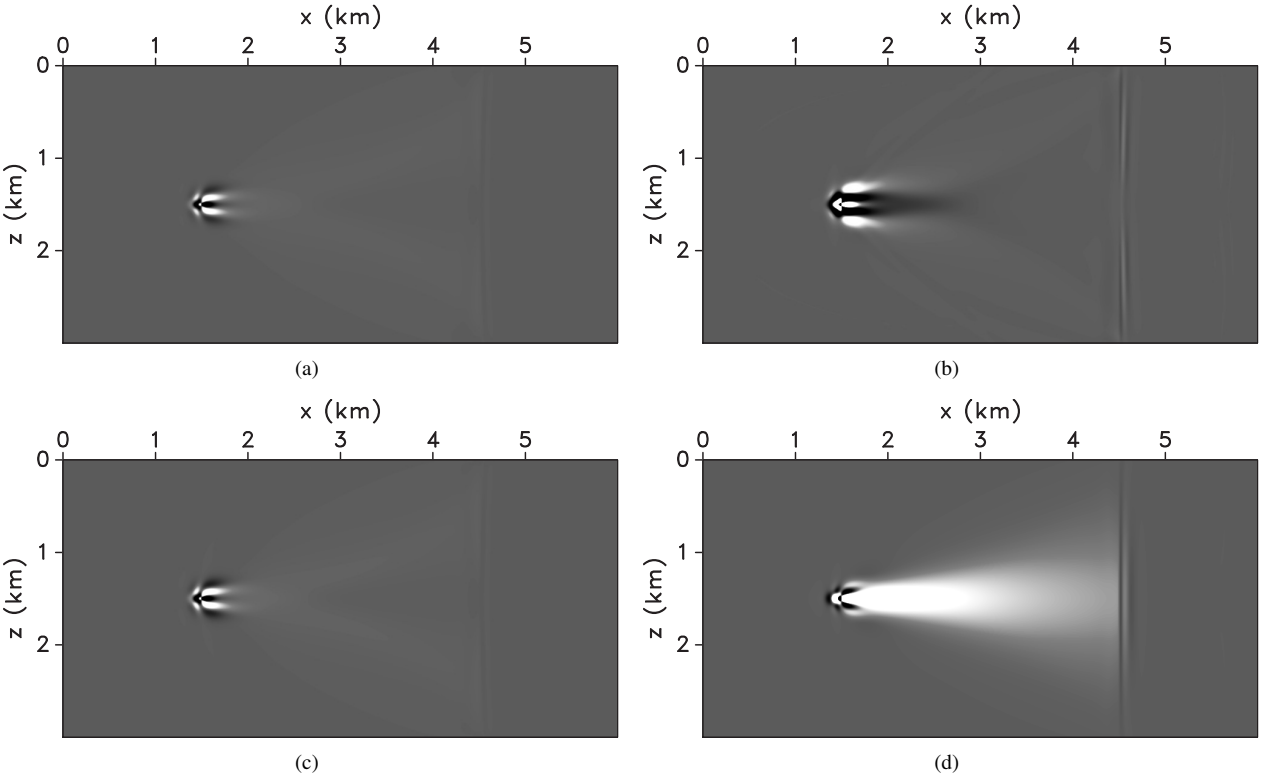


Figure 8. Gradients of the objective function with respect to (a) V_{P0}^2, (b) V_{S0}^2, (c) $V_{\mathrm{nmo},P}^2$, and (d) $V_{\mathrm{hor},P}^2$ for the model with an anomaly in V_{P0} from Figure 7.

the anomaly in V_{P0}. Increasing the number of sources, with a larger polar angle coverage, is likely to elevate the gradients with respect to all three velocities, and, potentially improve the convergence of the inversion algorithm.

When the source-receiver configuration is rotated by 90° (Figure 9) and the wave propagation is predominantly vertical, the largest gradient is the one with respect to V_{P0} (Figure 10). In principle, the gradients for V_{nmo} and $V_{\mathrm{hor},P}$ should not vanish, which would ensure that the parameters ϵ and δ do not get updated. As in the previous case, multiple shots with improved spatial coverage should increase the gradients for the NMO and horizontal velocities.

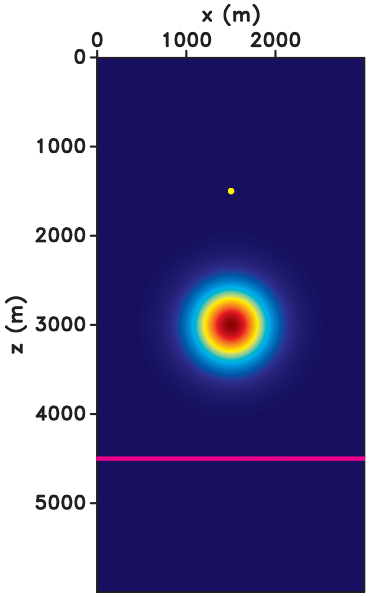


Figure 9. VTI model with a Gaussian anomaly ($\sigma = 300\,m$) in the P-wave vertical velocity V_{P0} and different source-receiver geometry. The source-receiver configuration from Figure 7 is rotated by 90°. The background and maximum values of V_{P0} are 3000 m/s and 3140 m/s, respectively. The other parameters are $V_{S0} = 1500\,\mathrm{m/s}$, $\delta = -0.1$, and $\epsilon = 0.1$.

The gradients for a Gaussian anomaly in V_{S0} and a horizontal receiver array (Figure 11) are displayed in Figure 12. As expected, the gradient with respect to V_{S0} is much larger than the other three. However, because the source is polarized in the vertical direction, the SV-wave amplitude vanishes along the symmetry axis and increases away from it. As a result, the gradient for V_{S0} reaches its maximum at oblique incidence angles near the symmetry axis.

4 CONCLUSIONS

One of the most important steps in performing FWI is calculation of the gradient of the objective function with respect to the model parameters. We discussed gradient computation for elastic multicomponent wavefields from 2D VTI media. The in-plane polarized waves (P and SV) are controlled by combinations of four stiffness coefficients: C_{11}, C_{13}, C_{33}, and C_{55}. The adjoint state method was employed to derive analytic expressions for the gradients of the least-squares objective function with respect to the stiffnesses. A more convenient parameterization, however, includes the squares of the P-wave vertical (V_{P0}), NMO ($V_{\mathrm{nmo},P}$), and horizontal ($V_{\mathrm{hor},P}$) velocities, and of the S-wave vertical velocity (V_{S0}).

After obtaining the gradients for V_{P0}^2, V_{S0}^2, $V_{\mathrm{nmo},P}^2$, and V_{hor}^2, we conducted numerical tests for Gaussian anomalies in the Thomsen parmeters V_{P0}, V_{S0}, ϵ, and δ embedded in a homogeneous VTI model. The gradients were computed for a single buried source and an array of receivers placed at each grid point along a straight line. If the spatial extent of the anomaly is small, it acts as a secondary source that generates P-to-S conversions. The obtained gradients in this case essentially represent images of the anomaly and their meaning requires further analysis.

For anomalies with a larger spatial extent, the converted waves are weak and there are fewer artifacts. The magnitude of the gradient with respect to each model parameter is governed by the source position and polarization, as well as the location and configuration of the receiver array. For example, an anomaly in V_{P0} produces a large gradient with respect to $V_{\mathrm{hor},P}$ when the wavefield propagates predominantly in the horizontal direction. The gradient for V_{P0}, however, is negligible, which would result in an undesired change in the parameter ϵ. However, for the same V_{P0} anomaly, but a different source-receiver geometry, when waves travel close to the vertical, the dominant gradient in the one with respect to V_{P0}. For the same source-receiver configuration with near-vertical wave propagation, an anomaly in the S-wave vertical velocity produces a large gradient with respect to V_{S0}. The maximum gradient, however, is observed at oblique incidence because the source polarization is vertical. Increasing the number of shots should improve illumination and result in gradients that are more consistent with the anomaly. Multiple shots are also expected to reduce the artifacts in the gradient fields.

The obtained gradients should be combined with an appropriate line-search technique to make meaningful model updates. In addition, it is necessary to investigate if the same step length can be used for all model parameters.

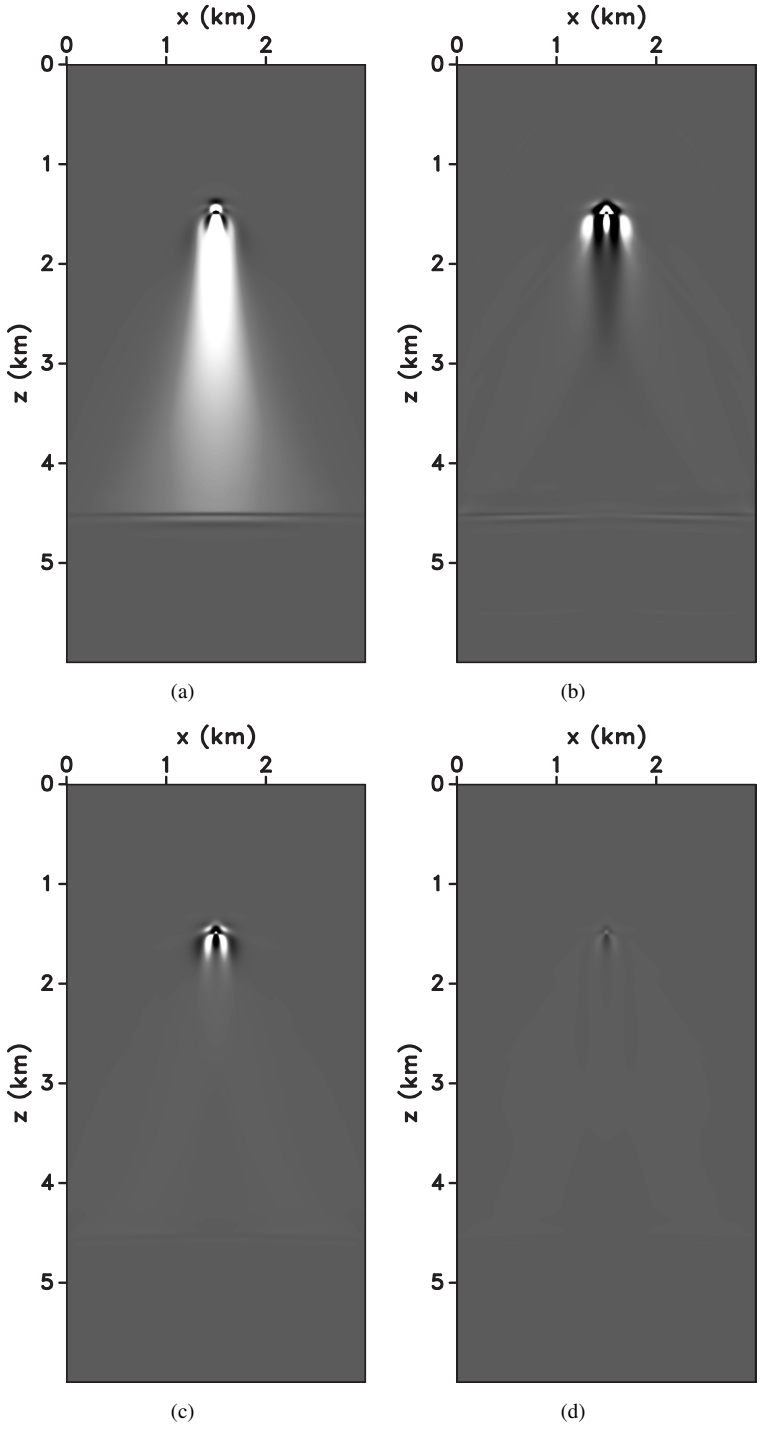


Figure 10. Gradients of the objective function with respect to (a) V_{P0}^2, (b) V_{S0}^2, (c) $V_{\mathrm{nmo},P}^2$, and (d) $V_{\mathrm{hor},P}^2$ for the model with an anomaly in V_{P0} from Figure 9.

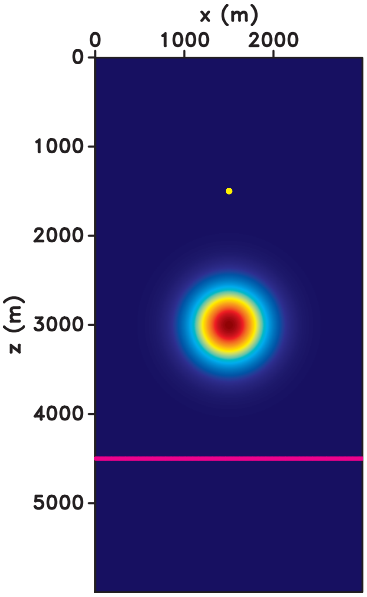


Figure 11. VTI model with a Gaussian anomaly ($\sigma = 300\,m$) in the S-wave vertical velocity V_{S0}. The background and maximum values of V_{S0} are 1500 m/s and 1640 m/s, respectively. The other parameters are $V_{P0} = 3000\,\mathrm{m/s}$, $\delta = -0.1$, and $\epsilon = 0.1$.

5 ACKNOWLEDGMENTS

We are grateful to the members of the A(nisotropy) and i(maging) teams at CWP and to Francesco Perrone, Esteban Diaz and Simon Luo (CWP), Daniel Köhn (University of Kiel), Ken Matson (Shell), and Jon Sheiman (contractor, Shell), for fruitful discussions. We would also like to thank John Stockwell (CWP) and Paul Martin (Dept. of Mathematics, CSM) for help with mathematical issues. This work was supported by the Consortium Project on Seismic Inverse Methods for Complex Structures at CWP and by the CIMMM Project of the Unconventional Natural Gas Institute at CSM. The reproducible numeric examples in this paper are generated with the Madagascar open-source software package freely available from http://www.ahay.org.

REFERENCES

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Fichtner, A., H.-P. Bunge, and H. Igel, 2006, The adjoint method in seismology: I. theory: Physics of the Earth and Planetary Interiors, **157**, 86 – 104.

Gauthier, O., 1986, Two-dimensional nonlinear inversion of seismic waveforms: Numerical results: Geophysics, **51**, 1387–1403.

Gholami, Y., R. Brossier, S. Operto, V. Prieux, A. Ribodetti, and J. Virieux, 2011, Two-dimensional acoustic anisotropic (VTI) full waveform inversion: The Valhall case study: SEG Technical Program Expanded Abstracts, **30**, 2543–2548.

Kamath, N., and I. Tsvankin, 2012, Full-waveform inversion of multicomponent data for layered VTI media: CWP Report, **706**, 11–19.

Köhn, D., D. De Nil, A. Kurzmann, A. Przebindowska, and T. Bohlen, 2012, On the influence of model parametrization in elastic full waveform tomography: Geophysical Journal International, **191**, 325–345.

Kolb, P., F. Collino, and P. Lailly, 1986, Pre-stack inversion of a 1-D medium: Proceedings of the IEEE, **74**, 498–508.

Lee, H., J. M. Koo, D. Min, B. Kwon, and H. S. Yoo, 2010, Frequency-domain elastic full waveform inversion for VTI media: Geophysical Journal International, **183**, 884–904.

Liu, Q., and J. Tromp, 2006, Finite-frequency kernels based on adjoint methods: Bulletin of the Seismological Society of America, **96**, 2383–2397.

Mora, P., 1987, Nonlinear two-dimensional elastic inversion of multioffset seismic data: Geophysics, **52**, 1211–1228.

Plessix, R., and Q. Cao, 2011, A parametrization study for surface seismic full waveform inversion in an acoustic vertical transversely isotropic medium: Geophysical Journal International, **185**, 539–556.

Plessix, R.-E., and H. Rynja, 2010, VTI full waveform inversion: a parameterization study with a narrow azimuth streamer data example: SEG Technical Program Expanded Abstracts, **29**, 962–966.

Plessix, R. ., 2006, A review of the adjointstate method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503.

Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain; part 1, theory and verification in a physical scale model: Geophysics, **64**, 888–901.

Pratt, R. G., and R. M. Shipp, 1999, Seismic waveform in-

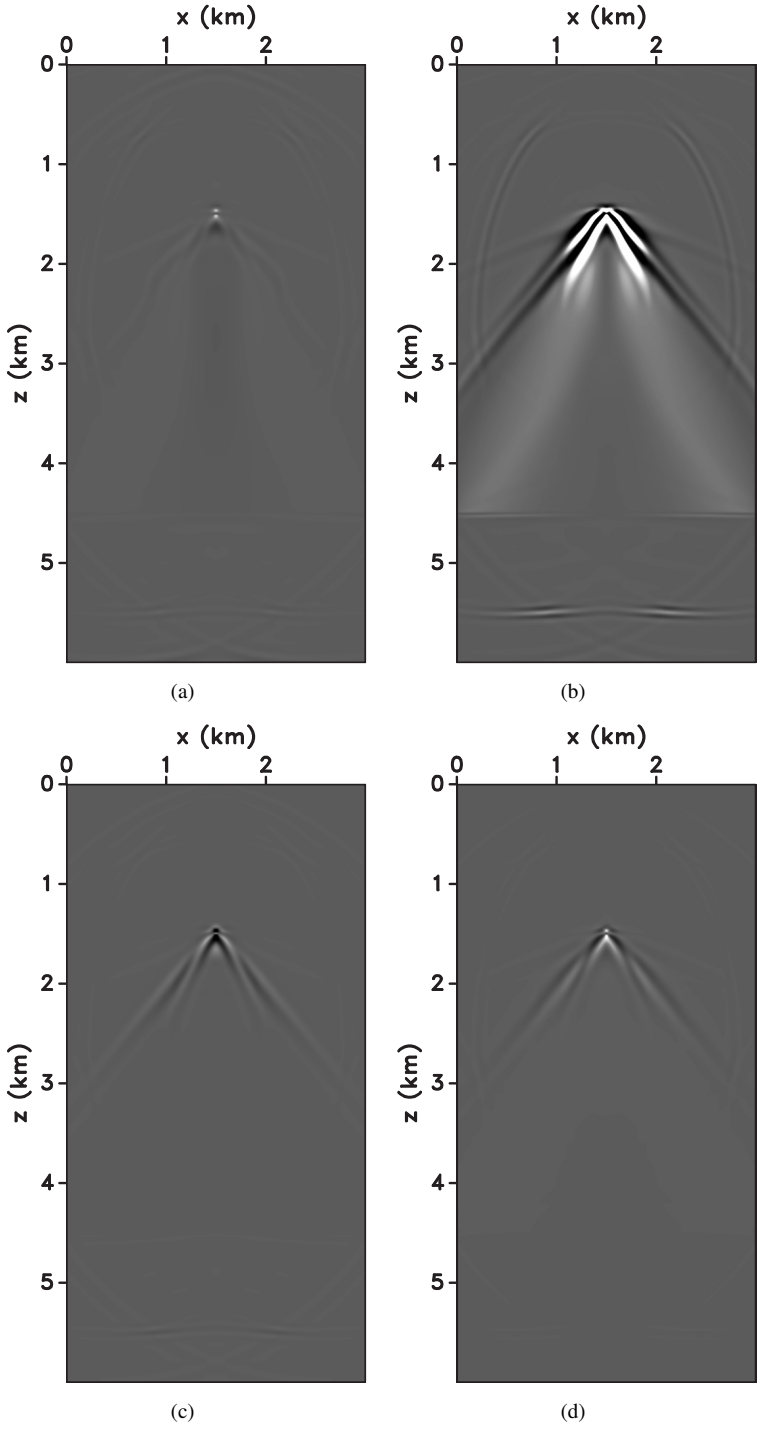


Figure 12. Gradients of the objective function with respect to (a) V_{P0}^2, (b) V_{S0}^2, (c) $V_{\mathrm{nmo},P}^2$, and (d) $V_{\mathrm{hor},P}^2$ for the model with an anomaly in V_{S0} from Figure 11.

version in the frequency domain; part 2; fault delineation in sediments using crosshole data: Geophysics, **64**, 902–914.

Shen, X., 2012, Early-arrival waveform inversion for near-surface velocity and anisotropic parameter: Parametrization study: SEG Technical Program Expanded Abstracts 2012, SEG, 1–5.

Song, Z., and P. R. Williamson, 1995, Frequency-domain acoustic-wave modeling and inversion of crosshole data; part 1, 2.5-D modeling method: Geophysics, **60**, 784–795.

Song, Z., P. R. Williamson, and R. G. Pratt, 1995, Frequency-domain acoustic-wave modeling and inversion of crosshole data; Part II — Inversion method, synthetic experiments and real-data results: Geophysics, **60**, 796–809.

Strang, G., 1991, Calculus: Wellesley-Cambridge Press.

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

———, 1984b, Linearized inversion of seismic reflection data: Geophysical Prospecting, **32**, 998–1015.

Tsvankin, I., 2012, Seismic signatures and analysis of reflection data in anisotropic media, Third ed.: Society of Exploration Geophysicists.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–26.

APPENDIX A: APPENDIX A: GRADIENT COMPUTATION FOR VTI MEDIA USING THE ADJOINT STATE METHOD

As discussed in Plessix (2006) and Liu and Tromp (2006), the objective function in equation 1 is minimized under the constraint that the modeled displacement $\mathbf{u}(\mathbf{x}_r, t)$ satisfies the elastic wave equation,

$$\rho \frac{\partial^2 u_i}{\partial t^2} - \frac{\partial}{\partial x_j}\left(c_{ijkl}\frac{\partial u_k}{\partial x_l}\right) = f_i \ , \tag{A1}$$

where ρ is the density, c_{ijkl} are components of the stiffness tensor and $\mathbf{f}$ is the body force per unit volume. All indices change from 1 to 3 and summation over repeated indices is implied. The displacement wavefield is subject to the initial conditions,

$$\mathbf{u}(\mathbf{x}, 0) = 0, \quad \frac{\partial \mathbf{u}(\mathbf{x}, 0)}{\partial t} = 0, \tag{A2}$$

and the radiation boundary condition,

$$\mathbf{u}(\mathbf{x}, t)|_{\mathbf{x}\to\infty} \to 0. \tag{A3}$$

The method of Lagrange multipliers (Strang, 1991) is used to define the Lagrangian Λ:

$$\Lambda = \frac{1}{2}\sum_r \int_0^T \|\mathbf{u}(\mathbf{x}_r, t) - \mathbf{d}(\mathbf{x}_r, t)\|^2 \mathrm{d}t - \int_0^T \int_\Omega \lambda_i \left[\rho \frac{\partial^2 u_i}{\partial t^2} - \frac{\partial}{\partial x_j}\left(c_{ijkl}\frac{\partial u_k}{\partial x_l}\right) - f_i\right] \mathrm{d}V\mathrm{d}t, \tag{A4}$$

where Ω is the domain of integration, $\partial\Omega$ is the surface of Ω, and $\boldsymbol{\lambda}(\mathbf{x}, t)$ is the vector Lagrange multiplier that needs to be determined. Note that when $\mathbf{u}(\mathbf{x}, t)$ satisfies the wave equation, the second term in the above equation disappears and Λ reduces to the objective function in equation 1. Each of the model parameters (c_{ijkl} in this case) and the state variable $\mathbf{u}$ are perturbed. After integration by parts and application of the Gauss divergence theorem, we obtain the change in the Lagrangian,

$$\begin{aligned}
\delta\Lambda = & \int_0^T \int_\Omega \sum_{r=1}^{\mathrm{N}} \Big(u_i(\mathbf{x}, t) - d_i(\mathbf{x}, t)\Big)\delta(\mathbf{x} - \mathbf{x}_r)\delta u_i \ \mathrm{d}V\mathrm{d}t \\
& - \int_0^T \int_\Omega \delta c_{ijkl} \ \frac{\partial u_k}{\partial x_l} \ \frac{\partial \lambda_i}{\partial x_j} \ \mathrm{d}V\mathrm{d}t - \int_0^T \int_\Omega \left[\rho \frac{\partial^2 \lambda_i}{\partial t^2} - \frac{\partial}{\partial x_j}\left(c_{ijkl} \ \frac{\partial \lambda_k}{\partial x_l}\right)\right] \delta u_i \ \mathrm{d}V\mathrm{d}t \\
& - \int_\Omega \left[\rho\lambda_i \frac{\partial(\delta u_i)}{\partial t} - \rho \ \delta u_i \ \frac{\partial \lambda_i}{\partial t}\right]\Bigg|_0^T \ \mathrm{d}V \\
& + \int_0^T \int_{\partial\Omega} \lambda_i \left[\delta c_{ijkl} \ \frac{\partial u_k}{\partial x_l} + c_{ijkl} \ \frac{\partial(\delta u_k)}{\partial x_l}\right] n_j \ \mathrm{d}S \ \mathrm{d}t - \int_0^T \int_{\partial\Omega} \delta u_i \ c_{ijkl} \ \frac{\partial \lambda_k}{\partial x_l} \ n_j \ \mathrm{d}S \ \mathrm{d}t \ ,
\end{aligned} \tag{A5}$$

where $\mathbf{n}$ is the vector normal to the surface $\partial\Omega$ and $r = 1, 2 \ldots \mathrm{N}$ denotes the receivers. Perturbing $\mathbf{u}(\mathbf{x}, t)$ in equations A2 and A3 yields the initial and boundary conditions for $\delta\mathbf{u}(\mathbf{x}, t)$:

$$\delta\mathbf{u}(\mathbf{x}, 0) = 0, \quad \frac{\partial \delta\mathbf{u}(\mathbf{x}, t)}{\partial t} = 0, \quad \delta\mathbf{u}(\mathbf{x}, t)|_{\mathbf{x}\to\infty} \to 0. \tag{A6}$$

The wavefield $\boldsymbol{\lambda}$ is constrained by the 'final' conditions (i.e., those at time T),

$$\boldsymbol{\lambda}(\mathbf{x}, T) = \mathbf{0}, \quad \frac{\partial \boldsymbol{\lambda}(\mathbf{x}, T)}{\partial t} = \mathbf{0}, \tag{A7}$$

and the boundary condition,

$$\boldsymbol{\lambda}(\mathbf{x}, t) \ |_{\mathbf{x}\to\infty} \to 0. \tag{A8}$$

Equation A5 then reduces to

$$\delta\Lambda = \int_0^T \int_\Omega \sum_{r=1}^{\mathrm{N}} \Big(u_i(\mathbf{x}, t) - d_i(\mathbf{x}, t)\Big)\delta(\mathbf{x} - \mathbf{x}_r)\delta u_i \ \mathrm{d}V\mathrm{d}t$$

$$-\int_0^T \int_\Omega \delta c_{ijkl}\, \frac{\partial u_k}{\partial x_l}\, \frac{\partial \lambda_i}{\partial x_j}\, \mathrm{d}V\mathrm{d}t - \int_0^T \int_\Omega \left[\rho \frac{\partial^2 \lambda_i}{\partial t^2} - \frac{\partial}{\partial x_j}\left(c_{ijkl}\, \frac{\partial \lambda_k}{\partial x_l}\right)\right] \delta u_i\, \mathrm{d}V\mathrm{d}t\,. \tag{A9}$$

According to Plessix (2006), the condition $\partial\Lambda/\partial\boldsymbol{\lambda} = 0$ leads to the "state equations" (in this case, the elastic wave equation); this can be verified from equation A4. The adjoint state equations are obtained by setting $\partial\Lambda/\partial\mathbf{u} = 0$. Taking the derivative in equation A9 gives the adjoint state equation,

$$\rho \frac{\partial^2 \lambda_i}{\partial t^2} - \frac{\partial}{\partial x_j}\left(c_{ijkl} \frac{\partial \lambda_k}{\partial x_l}\right) = \sum_{r=1}^{\mathrm{N}} \Big[u_i(\mathbf{x}_r, t) - d_i(\mathbf{x}_r, t)\Big], \tag{A10}$$

subject to the conditions at time T (equation A7) and boundary conditions (equation A8). In addition, $\partial\Lambda/\partial\mathbf{m} = \partial\mathcal{F}/\partial\mathbf{m}$, where $\mathbf{m}$ are the model parameters (Plessix, 2006). Hence, the change in the objective function $\delta\mathcal{F}$ caused by perturbations of the stiffness coefficients is given by:

$$\delta\mathcal{F} = -\int_0^T \int_\Omega \delta c_{ijkl}\, \frac{\partial u_i}{\partial x_j}\, \frac{\partial \lambda_k}{\partial u_l}\, \mathrm{d}V\mathrm{d}t\,. \tag{A11}$$

This is a general result for an anisotropic medium described by the complete stiffness tensor c_{ijkl}. Expressions for models with specific symmetries can be derived from equation A11 by substituting the appropriate stiffness tensors or matrices.

Note that the boundary conditions for $\mathbf{u}(\mathbf{x}, t)$ and $\boldsymbol{\lambda}(\mathbf{x}, t)$ can be modified to include a free surface where the tractions due to $\mathbf{u}$ and $\boldsymbol{\lambda}$ go to zero. However, the addition of the free surface causes complications in finite-difference modeling and produces surface multiples. Instead, we impose the radiation condition to create absoring boundaries on all sides of the model.

To simulate the Lagrange multiplier wavefield, it is convenient to define an "adjoint wavefield" $\boldsymbol{\psi}$ (Liu and Tromp, 2006):

$$\boldsymbol{\psi}(\mathbf{x}, t) \equiv \boldsymbol{\lambda}(\mathbf{x}, T - t). \tag{A12}$$

The wavefield $\boldsymbol{\psi}$ satisfies the wave equation A10 but with the source function reversed in time:

$$\rho \frac{\partial^2 \psi_i}{\partial t^2} - \frac{\partial}{\partial x_j}\left(c_{ijkl} \frac{\partial \psi_k}{\partial x_l}\right) = \sum_{r=1}^{\mathrm{N}} \Big[u_i(\mathbf{x}_r, T - t) - d_i(\mathbf{x}_r, T - t)\Big]\,. \tag{A13}$$

The initial conditions for $\boldsymbol{\psi}$ (using equations A7 and A12) are as follows:

$$\boldsymbol{\psi}(\mathbf{x}, 0) = 0, \quad \frac{\partial \boldsymbol{\psi}(\mathbf{x}, 0)}{\partial t} = 0. \tag{A14}$$

The wavefield $\boldsymbol{\psi}$ also satisfies the radiation boundary condition:

$$\boldsymbol{\psi}(\mathbf{x}, t)|_{\mathbf{x}\to\infty} \to 0. \tag{A15}$$

Note that defining the adjoint wavefield makes it possible to simulate the Lagrange multiplier wavefield (which starts at time T and propagates backward in time).

From equations A11 and A12, we can find the gradient of the objective function with respect to the stiffness coefficients:

$$\frac{\partial \mathcal{F}}{\partial c_{ijkl}} = -\int_0^T \frac{\partial u_i}{\partial x_j} \frac{\partial \psi_k}{\partial x_l}\, \mathrm{d}t. \tag{A16}$$

If, instead of c_{ijkl}, the model is described by parameters m_n, the gradient of $\mathcal{F}$ can be found from the chain rule:

$$\frac{\partial \mathcal{F}}{\partial m_n} = \sum_{ijkl} \frac{\partial \mathcal{F}}{\partial c_{ijkl}}\, \frac{\partial c_{ijkl}}{\partial m_n}. \tag{A17}$$

Here, we parameterize the model in terms of the squared velocities V_{P0}^2, V_{S0}^2, $V_{\mathrm{nmo},P}^2$, and $V_{\mathrm{hor},P}^2$. The stiffness coefficients (written in the two-index notation) represent the following functions of the velocities (Tsvankin, 2012):

$$C_{11} = \rho\, V_{\mathrm{hor},P}^2, \tag{A18}$$

$$C_{33} = \rho\, V_{P0}^2, \tag{A19}$$

$$C_{13} = \rho \sqrt{(V_{P0}^2 - V_{S0}^2)(V_{\mathrm{nmo},P}^2 - V_{S0}^2)} - \rho\, V_{S0}^2, \tag{A20}$$

$$C_{55} = \rho\, V_{S0}^2. \tag{A21}$$

Using equations A16, A17, and A18 − A21, we obtain the following gradients of the objective function with respect to V_n^2:

$$\frac{\partial \mathcal{F}}{\partial \left(V_{P0}^2\right)} = -\int_0^T \rho \left[\frac{\partial \psi_3}{\partial x_3}\frac{\partial u_3}{\partial x_3} + \frac{1}{2}\sqrt{\frac{V_{\mathrm{nmo},P}^2 - V_{S0}^2}{V_{P0}^2 - V_{S0}^2}} \left(\frac{\partial \psi_1}{\partial x_1}\frac{\partial u_3}{\partial x_3} + \frac{\partial \psi_3}{\partial x_3}\frac{\partial u_1}{\partial x_1} \right) \right] \mathrm{d}t, \tag{A22}$$

$$\begin{aligned} \frac{\partial \mathcal{F}}{\partial \left(V_{S0}^2\right)} = -\int_0^T \rho \Bigg\{ & \left[\frac{2V_{S0}^2 - V_{P0}^2 - V_{\mathrm{nmo},P}^2}{2\sqrt{(V_{\mathrm{nmo},P}^2 - V_{S0}^2)(V_{P0}^2 - V_{S0}^2)}} - 1 \right] \left(\frac{\partial \psi_1}{\partial x_1}\frac{\partial u_3}{\partial x_3} + \frac{\partial \psi_3}{\partial x_3}\frac{\partial u_1}{\partial x_1} \right) \\ & + \left(\frac{\partial \psi_1}{\partial x_3} + \frac{\partial \psi_3}{\partial x_1} \right) \left(\frac{\partial u_1}{\partial x_3} + \frac{\partial u_3}{\partial x_1} \right) \Bigg\} \mathrm{d}t, \end{aligned} \tag{A23}$$

$$\frac{\partial \mathcal{F}}{\partial \left(V_{\mathrm{nmo},P}^2\right)} = -\int_0^T \frac{\rho}{2}\sqrt{\frac{V_{P0}^2 - V_{S0}^2}{V_{\mathrm{nmo},P}^2 - V_{S0}^2}} \left(\frac{\partial \psi_1}{\partial x_1}\frac{\partial u_3}{\partial x_3} + \frac{\partial \psi_3}{\partial x_3}\frac{\partial u_1}{\partial x_1} \right) \mathrm{d}t, \tag{A24}$$

$$\frac{\partial \mathcal{F}}{\partial \left(V_{\mathrm{hor},P}^2\right)} = -\int_0^T \rho \frac{\partial \psi_1}{\partial x_1}\frac{\partial u_1}{\partial x_1}\, \mathrm{d}t, \tag{A25}$$