Gradient computation for VTI acoustic wavefield tomography

Vladimir Li$^1$, Hui Wang$^1$, Ilya Tsvankin$^1$, Esteban Díaz$^1$ & Tariq Alkhalifah$^2$

$^1$Center for Wave Phenomena, Colorado School of Mines
$^2$King Abdullah University of Science and Technology

ABSTRACT
Wavefield tomography can handle complex subsurface geology better than ray-based techniques and, ultimately, provide a higher resolution. Applied to anisotropic parameter estimation, wavefield methods are commonly implemented under the pseudoo acoustic assumption. Here, we build the foundation for joint wavefield tomography of reflection and vertical seismic profiling (VSP) data from acoustic transversely isotropic (TI) media. The simulation is performed with a wave-equation operator based on a separable approximation of the P-wave dispersion relation. The adjoint-state method is employed to derive the gradients of the objective function for both data-domain and image-domain inversion. The gradients are compared to the ones obtained with a space-time finite-difference (FD) scheme for a system of coupled wave equations. Numerical examples demonstrate the consistency between the kernels computed with different wave-equation operators. The gradients obtained with the pseudospectral method do not contain the imprint of the shear-wave artifact produced by the space-time FD algorithm.

Key words: anisotropy, wave equation, data domain, image domain, tomography, gradient

1 INTRODUCTION

A tomographic approach to building a subsurface model generally leads to an ill-posed nonlinear problem, which means that the wavefield attributes cannot be directly converted into the medium properties. Hence, a subsurface model is commonly updated iteratively with a certain optimization criterion.

Wavefield tomography can be implemented in the data or image domain depending on the way of formulating the objective function. Data-domain methods enforce the similarity between the predicted and observed seismic wavefields. The image-domain approach requires an additional migration step and relies, in accordance with the semblance principle, on the consistency of migrated images for different experiments (Al-Yahya, 1989; Sattlegger, 1975; Perrone and Sava, 2012). There are various modifications of image-domain tomography that employ different migration operators, imaging conditions, and types of image gathers (Sava, 2014). The objective function in either domain is typically minimized using gradient-based techniques, with the gradients obtained by the adjoint-state method (ASM) (Tarantola, 1984; Tromp et al., 2005; Plessix, 2006). Despite the difference in their objective functions, both data- and image-domain methods use the same wave equation and observed wavefields (Sava, 2014).

In this paper, we focus on wavefield extrapolation and gradient derivation, which are common steps for both groups of methods. Our algorithm is designed for transversely isotropic models with a vertical or tilted symmetry axis (VTI or TTI), which are widely used to improve the results of imaging and tomography. Optimally, anisotropic inversion requires elastic wavefield extrapolation and benefits from including shear and mode-converted waves. However, incorporating shear-wave information into wavefield-based inversion remains challenging due to the high cost and complexity of elastic modeling, imaging, and inversion, as well as the limited availability of multicomponent data. Therefore, anisotropic wavefield tomography is typically implemented under the pseudo-acoustic assumption originally proposed by Alkhalifah (1998, 2000).

P-wave kinematics in VTI media is controlled by the vertical velocity $V_{P0}$ and Thomsen parameters $\epsilon$ and $\delta$ (Tsvankin, 2012). Alternative parameter combinations for acoustic VTI media also involve the horizontal velocity $V_{hor} = V_{P0} \sqrt{1 + 2\eta}$, the anellipticity parameter $\eta = (\epsilon - \delta)/(1 + 2\delta)$, and the normal-moveout (NMO) velocity for a horizontal interface $V_{nmo} = V_{P0} \sqrt{1 + 2\delta}$. The main challenge in anisotropic wavefield-based inversion is the trade-off between model parameters, which strongly depends on the chosen parameterization.

Acoustic modeling in TI media is performed using either
explicit space-time finite-difference (FD) schemes or pseudospectral methods. The FD methods operate with coupled second-order partial differential equations (Duveneck et al., 2008; Fletcher et al., 2009; Fowler et al., 2010; Zhang et al., 2011). As discussed by Alkhalifah (1998, 2000) and Greczka et al. (2004), setting the shear-wave symmetry-direction velocity $V_{SH}$ to zero in the acoustic wave equation generates the so-called shear-wave “artifact,” which can contaminate migrated images and hamper the acoustic inversion. The simplest way to suppress the artifact is to place sources and receivers in an elliptic ($\epsilon = \delta, \eta = 0$) or purely isotropic medium (Alkhalifah, 2000; Duveneck et al., 2008). However, this strategy can be legitimately applied only in the case of the data-domain waveform inversion of surface data when the physical sources and receivers, as well as the adjoint sources, reside in the near-surface layer, which can be made elliptic. More elaborate methods for suppressing the artifact involve using a finite $V_{SH}$, wave-mode separation, or introducing a damping term (Fletcher et al., 2009; Le and Levin, 2014; Suh, 2014; Fowler and King, 2011). Another issue with the space-time finite-difference techniques is their numerical instability for models with $\eta < 0$.

The pseudospectral methods are designed to propagate only P-waves by producing decoupled modes in the wavenumber domain (Etgen and Brandsberg-Dahl, 2009; Crawley et al., 2010; Pestana and Stoffa, 2010; Song and Alkhalifah, 2013; Fomel et al., 2013b). A comprehensive review and classification of these methods can be found in Du et al. (2014). Separable P-mode dispersion-relation approximations for TI and orthorhombic media are described in Pestana et al. (2011), Zhan et al. (2012), Du et al. (2014), and Schleicher and Costa (2015).

Anisotropic waveform inversion has drawn considerable attention in the literature, but it is usually implemented only in the data domain (Warner et al., 2013; Gholami et al., 2013; Plessix et al., 2014; Wang and Sava, 2015; Kamath et al., 2015). The most common approach to image-domain tomography involves evaluating the energy focusing in the extended images (Rickett and Sava, 2002; Sava and Fomel, 2006; Sava and Vasconcelos, 2011), which can be done with differential semblance optimization (DSO) (Symes and Carazzzzone, 1991; Shen and Symes, 2008) or image-power estimates (Chavent and Jacewitz, 1995; Soubaras and Gratacos, 2007).

In general, P-wave reflection moveout must be supplemented with borehole (Wang and Tsvankin, 2013a,b) or other information to solve the VTI parameters. Li et al. (2014) and Arntsen (2014) use elastic P-wave extended images to estimate $V_{P0}$, $\epsilon$, and $\delta$. However, their imaging condition is based on a purely isotropic wave-mode separation technique.

Li et al. (2015) analyze the defocusing in the extended domain caused by errors in the VTI parameters and show that the coefficient $\delta$ could be constrained only if it strongly varies laterally. As is the case for conventional moveout analysis, the sensitivity to the anellipticity parameter $\eta$ is higher for dipping interfaces than for horizontal reflectors.

In this paper, we derive the gradients of the data- and image-domain objective functions for VTI media using a wave-equation operator based on the separable P-mode approximation. After reviewing parameterization and wavefield extrapolation for acoustic VTI models, we discuss the objective functions for wavefield tomography, with the main focus on the image-domain approach. For data-domain tomography, the analysis is restricted to the conventional objective function that represents the $\ell_2$-norm of the data-difference. Then we obtain the corresponding gradients of the objective function in both domains using the adjoint-state method. Finally, the gradient computation is tested on VTI models with Gaussian anomalies in the parameter fields.

2 PARAMETERIZATION FOR ACOUSTIC VTI MEDIA

In general, VTI acoustic wavefield tomography in either domain cannot simultaneously constrain all three relevant model parameters due to the parameter trade-offs in surface P-wave data. For data-domain inversion, an optimal parameter choice depends on the directions in which the source and receiver wavefields interact to produce a model update. Alkhalifah and Plessix (2014) analyze the radiation (sensitivity) patterns for horizontal reflectors in acoustic VTI media. They conclude that if the inversion is driven primarily by waves traveling in near-horizontal directions (e.g., diving waves), then the optimal parameter set includes $V_{P0}$, $\eta$, and $\epsilon$. For near-vertical propagation, it is better to operate with $V_{P0}$, $\eta$, and $\delta$.

From the image-domain prospective, parameter trade-offs stem from the properties of P-wave reflection moveout. Alkhalifah and Tsvankin (1995) demonstrate that P-wave reflection moveout for a laterally homogeneous VTI medium above the target horizon (which could be dipping or curved) is controlled by the velocity $V_{P0}$ and parameter $\eta$. For layer-cake VTI media, $\eta$ contributes only to the nonhyperbolic (long-offset) portion of the P-wave moveout. If the reflector is dipping, however, $\eta$ influences the NMO velocity and, therefore, the conventional-spread moveout. P-wave reflection traveltimes are not sensitive to the coefficient $\delta$, unless it varies laterally above the target reflector (Alkhalifah et al., 2001; Tsvankin and Grechka, 2011).

3 WAVEFIELD EXTRAPOLATION METHODS

Pseudo-acoustic modeling operators are widely used in imaging and tomography because of their simplicity and computa-
utional efficiency. Acoustic algorithms, however, cannot accurately predict P-wave amplitudes and must either rely on the phase of recorded arrivals or use a “dummy” model parameter that absorbs unphysical model updates.

3.1 Space-time finite-difference methods

Here, we use the formulation proposed by Fletcher et al. (2009) and Fowler et al. (2010). The 2D VTI version of their equations can be written as:

\[
\frac{\partial^2 u}{\partial t^2} = V_{\text{hor}}^2 \frac{\partial^2 u}{\partial x^2} + V_{P_0}^2 \frac{\partial^2 u}{\partial z^2},
\]

\[
\frac{\partial^2 u}{\partial z^2} = V_{\text{homo}}^2 \frac{\partial^2 u}{\partial x^2} + V_{P_0}^2 \frac{\partial^2 u}{\partial z^2},
\]

(1)

where \( u(x, t) \) and \( u(z, t) \) are the components of the vector wavefield \( \mathbf{u} \). In the matrix-vector notation equation 1 can be expressed as:

\[
\mathbf{L} \begin{bmatrix} u_x \\ u_z \end{bmatrix} + \begin{bmatrix} f_x \\ f_z \end{bmatrix} = 0,
\]

(2)

where the operator \( \mathbf{L} \) is defined as

\[
\mathbf{L} = \begin{bmatrix} V_{\text{hor}}^2 \partial_{xx} - \partial_{tt} & V_{P_0}^2 \partial_{zz} \\ V_{\text{homo}}^2 \partial_{xx} & V_{P_0}^2 \partial_{zz} - \partial_{tt} \end{bmatrix}.
\]

(3)

For gradient computation, we use the system of equations adjacent to equation 1 developed and implemented by Wang and Sava (2015).

3.2 Separable P-mode approximation

The pseudospectral methods evaluate the spatial derivatives of the wavefield in the wavenumber domain. Application to anisotropic wave equations requires developing approximate dispersion relations with separable wavenumber and model-parameter terms. In other words, the contributions of heterogeneity and anisotropy should be decoupled. In the pseudo-acoustic approximation, the 2D P-wave dispersion relation for VTI media can be written as (Alkhalifah, 1998):

\[
\omega^2 = \frac{1}{2} \left( (1 + 2\epsilon) V_{P_0}^2 k_x^2 + V_{P_0}^2 k_z^2 \right)
\]

\[
+ \frac{1}{2} \left( (1 + 2\epsilon) V_{P_0}^2 k_x^2 + V_{P_0}^2 k_z^2 \right) \sqrt{1 - \frac{8(\epsilon - \delta) k_z^2 k_x^2}{(1 + 2\epsilon) k_z^2 + k_x^2}},
\]

(4)

where \( k_x \) and \( k_z \) are the horizontal and vertical wavenumbers. However, equation 4 is not suitable for pseudospectral methods because it contains the radical term. A Taylor series expansion applied to the radical yields:

\[
\omega^2 = (1 + 2\epsilon) V_{P_0}^2 k_x^2 + V_{P_0}^2 k_z^2 - 2(\epsilon - \delta) V_{P_0}^2 \frac{k_z^2 k_x^2}{k_z^2 + F k_x^2},
\]

(5)

where \( F = 1 + 2\epsilon \). Pestana et al. (2011) set \( F \) to a constant to achieve separable formulas suitable for pseudospectral methods. Physically, the Taylor series expansion produces a weak-anisotropy approximation for the dispersion relation.

A more accurate dispersion relation can be obtained from Padé’s expansion. With the first-order Padé expansion, the separable dispersion relation takes the form (Schleicher and Costa, 2015):

\[
\omega^2 = (1 + 2\epsilon) V_{P_0}^2 k_x^2 + V_{P_0}^2 k_z^2 - 2(\epsilon - \delta) V_{P_0}^2 \frac{k_z^2 k_x^2}{k_z^2 + k_x^2} \left[ 1 - 2(\epsilon - \delta) \frac{k_z^2 k_x^2}{(k_z^2 + k_x^2)^2} \right].
\]

(6)

Here, the Padé coefficients \( \alpha \) and \( \beta \) in equation 17 of Schleicher and Costa (2015) are set to 1/2 and 1/4 respectively, and their coefficient \( f \) is set to unity according to the acoustic assumption. Equation 6 is referred to as the “separable strong-anisotropy approximation,” which can be implemented with pseudospectral methods. An extension to TTI media can be obtained by locally applying the appropriate rotation matrix to the wavenumbers because equation 6 remains valid for the wavenumbers \( k_x \) and \( k_z \) in the rotated coordinates.

4 OBJECTIVE FUNCTIONS FOR WAVEFIELD TOMOGRAPHY

4.1 Data domain

Data-domain methods enforce the similarity between the observed and modeled data. The objective function is typically defined as the \( l_2 \)-norm data difference:

\[
J = \frac{1}{2} \left\| \mathbf{K}(r) u - d_{\text{obs}} \right\|^2_2,
\]

(7)

where the operator \( \mathbf{K}(r) \) produces the modeled wavefield at the receiver locations, and \( d_{\text{obs}} \) is the observed data. However, because acoustic wavefield extrapolation cannot adequately predict P-wave amplitudes, application of equation 7 to field data is problematic. Acoustic data-domain tomography requires elaborate inversion algorithms that rely mostly on phase information and, therefore, are less prone to get trapped in local minima (Luo and Schuster, 1991; Alkhalifah, 2015; Choi and Alkhalifah, 2015; Díaz and Sava, 2015).

4.2 Image domain

Image-domain tomography methods use migrated images as input for the inversion. Our treatment is restricted to the residual energy minimization in the so-called extended domain. Extended images are produced by retaining correlation lags between the source and receiver wavefields in the output. The general imaging condition can be formulated as follows (Sava and Vasconcelos, 2011):

\[
I(\mathbf{x}, \lambda, \tau) = \sum_{x,t} W_s(\mathbf{x} - \mathbf{\lambda}, t - \tau)
\]

\[
\times W_r(\mathbf{x} + \mathbf{\lambda}, t + \tau),
\]

(8)

where \( I(\mathbf{x}, \lambda, \tau) \) is the extended image, \( W_s \) and \( W_r \) denote the source and receiver wavefields, respectively, \( \lambda \) is the space.
lag, and $\tau$ is the time lag. To reduce computational cost, one can compute only extended common-image-gathers (CIG), which are space-lag or time-lag extensions at fixed horizontal coordinates (Rickett and Sava, 2002; Sava and Fomel, 2006), or common-image-point (CIP) gathers, which represent multiple extensions computed at sparse points in the image space (Sava and Vasconcelos, 2009). Residual energy at nonzero lags can be used to update the migration velocity model and is most commonly measured with differential semblance optimization (DSO) (Symes and Carazzone, 1991; Shen and Symes, 2008). The DSO objective function for a horizontal space-lag extended image $I$ has the form:

$$J_{DSO} = \frac{1}{2} \| \lambda_s I(x, z, \lambda_s) \|^2_2,$$

where $\lambda_s$ is a penalty operator.

## 5 GRADIENT COMPUTATION USING THE ADJOINT-STATE METHOD

The adjoint-state method (Tarantola, 1984; Tromp et al., 2005; Plessix, 2006) is designed to efficiently evaluate the gradient of the objective function with respect to the model parameters. For seismic wavefield tomography, general gradient expressions for acoustic wavefields written in matrix-vector notation can be found in Sava (2014). Application of the adjoint-state method involves three ingredients: state equations, adjoint equations, and the objective function. Minimization of the objective function $J$ is subject to the constraints $F_s$ and $F_r$:

$$\begin{bmatrix} F_s \\ F_r \end{bmatrix} \begin{bmatrix} 0 & L \\ L^T & 0 \end{bmatrix} \begin{bmatrix} u_s \\ u_r \end{bmatrix} - \begin{bmatrix} d_s \\ d_r \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

where $L$ and $L^T$ are the forward and adjoint wave-equation operators, respectively, $0$ is the zero operator, $d_s$ is the source function, $d_r$ is the observed data, and $u_s$ and $u_r$ are the source and receiver wavefields, respectively. These constraints indicate that the wavefields $u_s$ and $u_r$ used in the minimization problem should be solutions of the wave equation:

$$\begin{bmatrix} L & 0 \\ 0 & L^T \end{bmatrix} \begin{bmatrix} u_s \\ u_r \end{bmatrix} = \begin{bmatrix} d_s \\ d_r \end{bmatrix}.$$  (11)

The method of Lagrange multipliers can be used to formulate the minimization as an unconstrained problem:

$$\mathcal{H} = J - \begin{bmatrix} F_s^T \\ F_r^T \end{bmatrix} \begin{bmatrix} a_s \\ a_r \end{bmatrix};$$

$$T$$ denotes a transpose. The Lagrange multipliers $a_s$ and $a_r$ are referred to as the adjoint-state variables, which are found from the following adjoint equations:

$$\begin{bmatrix} L^T & 0 \\ 0 & L \end{bmatrix} \begin{bmatrix} a_s \\ a_r \end{bmatrix} = \begin{bmatrix} g_s \\ g_r \end{bmatrix}.$$  (13)

The magnitude and spatial distribution of the source terms $g_s$ and $g_r$ are obtained from the objective function:

$$\begin{bmatrix} g_s \\ g_r \end{bmatrix} = \begin{bmatrix} \frac{\partial T}{\partial a_s} \\ \frac{\partial T}{\partial a_r} \end{bmatrix}.$$  (14)

Finally, the gradient of the augmented function $\mathcal{H}$ with respect to the vector $m$ of the model parameters is found as

$$\frac{\partial \mathcal{H}}{\partial m} = \frac{\partial J}{\partial m} + \sum_e \left( \frac{\partial F_s}{\partial m} \right)^T \left( \frac{\partial F_r}{\partial m} \right) \begin{bmatrix} a_s \\ a_r \end{bmatrix},$$

where $e$ indicates summation over all experiments. Additionally, the second term on the right-hand side implies summation over time, which is equivalent to the zero time-lag correlation (Sava, 2014):

$$\frac{\partial \mathcal{H}}{\partial m} = \frac{\partial J}{\partial m} + \sum_{e, \tau} \delta(\tau) \left( \frac{\partial F_s}{\partial m} * a_s + \frac{\partial F_r}{\partial m} * a_r \right),$$  (16)

where ‘$*$’ denotes cross-correlation. Overall, application of the adjoint-state method involves computing the following quantities:

(i) The state variables $u_s$ and $u_r$ by solving the state equations (equation 11).
(ii) The adjoint sources $g_s$ and $g_r$ that depend on the chosen objective function (equation 14).
(iii) The adjoint-state variables $a_s$ and $a_r$ by solving the adjoint equations (equation 13).
(iv) The gradient of the objective function, which depends on the wave-equation operator and parameterization.

Here, we apply the adjoint-state method to the pseudo-acoustic operators $L_{FD}$ and $L_{FS}$ discussed above and obtain gradient expressions for the objective functions in equations 7 and 9.

### 5.1 Coupled second-order equations

For VTI media, the forward (state) wave-equation operator $L$ is defined as (equation 3)

$$L = \begin{bmatrix} V_{hor}^2 \partial_{xx} - \partial_{tt} & V_{hor}^2 \partial_{zz} \\ V_{nmo}^2 \partial_{xx} & V_{nmo}^2 \partial_{zz} - \partial_{tt} \end{bmatrix}.$$  (17)

As shown by Wang and Sava (2015), the corresponding adjoint operator $L^T$ is

$$L^T = \begin{bmatrix} \partial_{xx} V_{hor}^2 - \partial_{tt} & \partial_{zz} V_{hor}^2 \\ \partial_{zz} V_{nmo}^2 & \partial_{xx} V_{nmo}^2 - \partial_{tt} \end{bmatrix}.$$  (18)

#### 5.1.1 Data domain

For the data-domain objective function (equation 7), the gradients can be found in Wang and Sava (2015). For 2D models, they define the data residual as $K \left( (u^d + u^m)^2 - d_{obs} \right)$, the model parameters as $m = \{ V_{hor}^2, V_{nmo}^2, V_{hor}^2 \}$, and obtain the following expressions:
Gradient computation for VTI acoustic wavefield tomography

but not for wavefield extrapolation. For VTI media, the forward (state) wave-equation operator $L$ equation 6, which simplifies the gradient expressions. However, we truncate equation 6 only for deriving the gradient expressions.

In anisotropic inversion, the initial Thomsen anisotropy parameters are often set to zero and their magnitude is typically small.

5.2 Pseudospectral approach

Application of the chain rule yields the gradient expressions for the vector $\mathbf{m} = \{V_{\text{hor}}, \eta, \epsilon\}$:

$$
\frac{\partial J}{\partial \mathbf{m}} = \frac{\partial J}{\partial \epsilon} \frac{\partial \epsilon}{\partial \delta} \frac{\partial \delta}{\partial \mathbf{m}} = \sum_{e, \tau} \delta(\tau) \begin{bmatrix}
\frac{\partial J}{\partial V_{\text{hor}}} & \frac{\partial J}{\partial \eta} & \frac{\partial J}{\partial \epsilon} \\
\frac{\partial J}{\partial \eta} & \frac{\partial J}{\partial \epsilon} & 0 \\
\frac{\partial J}{\partial \epsilon} & 0 & 0 \\
\end{bmatrix}
\begin{bmatrix}
\partial_{xx} u^g \ast (\alpha^p + \alpha^q) \\
\partial_{xx} u^p \ast a^q \\
\partial_{xx} u^p \ast a^p \\
\end{bmatrix}.
$$

5.1.2 Image domain

We define the space-lag common-image gather through the sum of the $p$ and $q$ components of the source and receiver wavefields:

$$
I(x, \lambda) = \sum_{e,t} W_e(x, z) W_r(x, z),
$$

where

$$
W_i(e, x, t) = u_i^p(e, x, t) + u_i^q(e, x, t), \ i = s, r.
$$

Thus, for the objective function in equation 9, equations 14 for the adjoint sources take the following form:

$$
\begin{bmatrix}
g^p_i \\
g^q_i \\
\end{bmatrix} = \sum_{\lambda x} \lambda^2_e \begin{bmatrix}
\frac{I(x + \lambda x, \lambda x) W_e(x + 2\lambda x, t)}{I(x + \lambda x, \lambda x) W_e(x + 2\lambda x, t)} \\
\frac{I(x - \lambda x, \lambda x) W_e(x - 2\lambda x, t)}{I(x - \lambda x, \lambda x) W_e(x - 2\lambda x, t)} \\
\end{bmatrix} ^i,
$$

After the adjoint wavefields are computed, the source- and receiver-side gradients with respect to the vector $\mathbf{m} = \{V_{\text{nmo}}, \eta, \delta\}$ are found as:

$$
\begin{bmatrix}
\frac{\partial J}{\partial V_{\text{nmo}}} & \frac{\partial J}{\partial \eta} & \frac{\partial J}{\partial \delta} \\
\frac{\partial J}{\partial \eta} & \frac{\partial J}{\partial \delta} & 0 \\
\frac{\partial J}{\partial \delta} & 0 & 0 \\
\end{bmatrix}
\begin{bmatrix}
\partial_{xx} u_i^q \ast (\alpha_i^p + \alpha_i^q) \\
(1 + 2\eta)\partial_{xx} u_i^p \ast a_i^q + \partial_{xx} u_i^p \ast a_i^q + \partial_{xx} u_i^q \ast (\alpha_i^p + \alpha_i^q) \\
\partial_{xx} u_i^p \ast a_i^p \\
\end{bmatrix},
$$

where $i$ denotes either the source or receiver side.

5.2 Pseudospectral approach

In anisotropic inversion, the initial Thomsen anisotropy parameters are often set to zero and their magnitude is typically small. Therefore, for most TI models, sufficient accuracy can be provided by the three leading terms of the separable dispersion relation in equation 6, which simplifies the gradient expressions. However, we truncate equation 6 only for deriving the gradient expressions but not for wavefield extrapolation. For VTI media, the forward (state) wave-equation operator $L$ can be defined as

$$
L = \omega^2 - V_{\text{hor}}^2 k_x^2 - \frac{V_{\text{hor}}^2}{1 + 2\eta} k_x^2 + 2\eta \frac{V_{\text{hor}}^2}{1 + 2\eta} k_z^2 k^2 + k_z^2,
$$

\[(25)\]
The image residual can be defined as:

\[ L = \omega^2 - V_{nmo}^2 k_x^2 - \frac{V_{nmo}^2}{1 + 2\delta} k_x^2 - 2\eta V_{nmo}^2 \frac{k_x^4}{k_x^2 + k_z^2}. \]  \hspace{1cm} (26)

The corresponding adjoint operator \( L^\dagger \) is:

\[ L^\dagger = \omega^2 - k_x^2 V_{hor}^2 - k_x^2 \frac{V_{hor}^2}{1 + 2\epsilon} + 2 \frac{k_x^2 k_z^2}{k_x^2 + k_z^2} \eta \frac{V_{hor}^2}{1 + 2\eta}, \]  \hspace{1cm} (27)

or

\[ L^\dagger = \omega^2 - k_x^2 V_{nmo}^2 - k_x^2 \frac{V_{nmo}^2}{1 + 2\delta} - 2 \frac{k_x^4}{k_x^2 + k_z^2} \eta V_{nmo}^2. \]  \hspace{1cm} (28)

### 5.2.1 Data domain

Below, we obtain the gradient expressions for the data-domain objective function in equation 7. The data residual is defined as \( K_r u - d_{obs} \). Therefore, equation 14 becomes:

\[
\begin{bmatrix}
g_s \\
g_r
\end{bmatrix} =
\begin{bmatrix}
K_r^T (K_r u - d_{obs}) \\
-\lambda K_r^T (K_r u - d_{obs})
\end{bmatrix}.
\]  \hspace{1cm} (29)

For data-domain methods, only the adjoint source wavefield is relevant (Sava, 2014), and the gradient with respect to the model parameter \( \mathbf{m} = \{ V_{hor}, \eta, \epsilon \} \) is given by the following expression:

\[
\frac{\partial J}{\partial \mathbf{m}} =
\begin{bmatrix}
\frac{\partial J}{\partial \epsilon} \\
\frac{\partial J}{\partial \eta} \\
\frac{\partial J}{\partial V_{hor}}
\end{bmatrix} =
\begin{bmatrix}
\frac{-2V_{hor}^2}{(1 + 2\epsilon)^2} & 0 & 0 \\
0 & \frac{-2V_{hor}^2}{(1 + 2\eta)^2} & 0 \\
0 & 0 & 2V_{hor}
\end{bmatrix}
\begin{bmatrix}
k_x^2 u \ast a \\
k_x^2 k_z^2 u \ast a \\
k_x^2 u \ast a + \frac{k_x^2 u}{1 + 2\epsilon} \ast a - \frac{2\eta}{1 + 2\eta} k_x^2 k_z^2 u \ast a
\end{bmatrix}.
\]  \hspace{1cm} (30)

### 5.2.2 Image domain

The image residual can be defined as:

\[ \lambda I(x, \lambda) = \lambda \left( \sum_{e, t} u_s(e, x - \lambda, t) u_r(e, x + \lambda, t) \right). \]  \hspace{1cm} (31)

Hence, for the objective function in equation 9, equation 14 for the adjoint sources becomes:

\[
\begin{bmatrix}
g_s \\
g_r
\end{bmatrix} = \sum_{s, r} \lambda_s^2 \begin{bmatrix} I(x + \lambda_s, \lambda_s) u_s(x + 2\lambda_s, t) \\ I(x - \lambda_s, \lambda_s) u_s(x - 2\lambda_s, t) \end{bmatrix}.
\]  \hspace{1cm} (32)

Then the source- and receiver-side gradients with respect to the model vector \( \mathbf{m} = \{ V_{nmo}, \eta, \delta \} \) are given by:

\[
\left. \begin{bmatrix}
\frac{\partial J}{\partial \mathbf{m}}
\end{bmatrix} \right|_i =
\begin{bmatrix}
\frac{\partial J}{\partial \delta} \\
\frac{\partial J}{\partial V_{nmo}} \\
\frac{\partial J}{\partial \eta}
\end{bmatrix} =
\begin{bmatrix}
\frac{-2V_{nmo}^2}{(1 + 2\delta)^2} & 0 & 0 \\
0 & 2V_{nmo} & 0 \\
0 & 0 & 2V_{nmo}^2
\end{bmatrix}
\begin{bmatrix}
k_x^2 u_i \ast a_i \\
k_x^2 k_z^2 u_i \ast a_i \\
k_x^2 k_z^2 u_i \ast a_i + \frac{k_x^2}{1 + 2\epsilon} u_i \ast a_i + 2\eta \frac{k_x^2 k_z^2}{k_x^2 + k_z^2} u_i \ast a_i
\end{bmatrix},
\]  \hspace{1cm} (33)

\[ i = s, r. \]
Gradient computation for VTI acoustic wavefield tomography

6 SYNTHETIC EXAMPLES

Below, we test the gradient expressions derived above on several simple VTI models. The medium parameters are specified on a rectangular grid, and the density is assumed to be constant. For forward and adjoint wavefield extrapolation, we use both the FD explicit scheme (operators $L_{FD}$ and $L^+_{FD}$) and the mixed-domain algorithm (operators $L_{MD}$ and $L^+_{MD}$) described above.

For data-domain examples (models 1 and 2), we use a vertical array of receivers to simulate VSP acquisition geometry, and the model vector is defined as $\mathbf{m} = \{V_{hor}, \eta, \epsilon\}$. In this parameterization, the parameter $\eta$ is poorly constrained by waveform tomography, even when very long offsets are available (Alkhalifah and Plessix, 2014). Hence, we assume $\eta$ to be known and compute the gradients only with respect to $V_{hor}$ and $\epsilon$.

For the image-domain test (model 3), the sources and receivers are located at the surface, and the model vector is defined as $\mathbf{m} = \{V_{\text{nmo}}, \eta, \delta\}$. Here, the velocity $V_{\text{nmo}}$ is assumed to be known, so the gradients are computed only for the anisotropy coefficients.

6.1 Model 1

In the first test, we compute the inversion gradients for the parameter $\epsilon$. The model includes identical Gaussian anomalies in the parameters $\eta$ and $\epsilon$ (reaching 0.15 at the center) embedded in a homogeneous isotropic background (Figure 1). Only transmitted waves are employed to generate parameter updates. The source function is a Ricker wavelet with a central frequency of 2 Hz. Using the actual $\eta$-field, we compute the gradients for understated and overstated peak values of the $\epsilon$-anomaly ($\epsilon = 0$ and 0.3; the background $\epsilon = 0$ is correct).

For the chosen parameterization ($V_{\text{hor}}, \eta, \epsilon$), the coefficient $\epsilon$ should be constrained for near-vertical propagation, if $V_{\text{hor}}$ has been estimated from long-offset data (Alkhalifah and Plessix, 2014). The gradients for one source ($x = 2$ km, $z = 0$ km) and one receiver ($x = 4$ km, $z = 4$ km) (Figure 2), referred to as “sensitivity kernels”, show what areas of the $\epsilon$-field can be updated with this acquisition geometry (Figure 3), the kernels are obtained with the mixed-domain operator. Note that for the peak frequency of the source signal (2 Hz) and the model size, the $\epsilon$-errors cause data residuals that do not exceed half a cycle.

The corresponding kernels computed with the FD algorithm have a similar spatial distribution (Figure 4). We then repeat the computations using the entire vertical receiver array that crosses the center of the anomaly. The gradients generated by both operators are similar and, as expected, change sign depending on the sign of the $\epsilon$-error (Figure 5 and 6).

Because here we use the data-difference objective function, the gradient for the actual $\epsilon$-field goes to zero (it may not be the case in the image domain). However, as mentioned above, the data-difference estimate is questionable for real-data applications because the acoustic approximation cannot provide accurate P-wave amplitudes.

6.2 Model 2

The second VTI model includes a Gaussian anomaly in the horizontal velocity $V_{\text{hor}}$, while the coefficients $\eta$ and $\epsilon$ are constant (Figure 7). This time, we keep the anisotropy coefficients $\eta$ and $\epsilon$ fixed at the actual values and evaluate the gradient for understated and overstated values of $V_{\text{hor}}$. In VTI velocity analysis, P-wave reflections can be combined with zero-offset and walkaway VSP data to constrain the Thomsen parameters (Wang and Tsvankin, 2013a,b). Here, we use walkaway VSP acquisition geometry with a horizontal source array at the surface and a vertical receiver array at $x = 7$ km. Both positive and negative perturbations in the $V_{\text{hor}}$-field introduce data residuals that satisfy the half-cycle requirement.

Similar to the previous example, we first obtain the gradients (sensitivity kernels) for a single source ($x = 1$ km) - receiver ($z = 4$ km) pair (Figure 7); the kernels are shown in Figures 8 and 9. Because the coefficient $\eta$ is positive, the FD algorithm produces the shear-wave artifact, which is visible near the source and receiver locations (Figure 9). Then we use the wavefields from all sources to compute the gradients (Figures 10 and 11). Despite summation over all experiments, the gradients computed with the FD algorithm still feature artifact-related vertical stripes at depth between 0 and 0.5 km (Figure 11b).

6.3 Model 3

Here, we obtain the gradients in the image domain that could help update the $\eta$-field using reflection data. The model includes a horizontal interface beneath a homogeneous VTI layer with $V_{\text{nmo}} = 2$ km/s, $\eta = \delta = 0.15$, and a thickness of 2 km. The near-surface layer 0.2 km-thick is assumed to be elliptic ($\epsilon = \delta$) to suppress the shear-wave artifact produced by the FD extrapolator. For this model, the coefficient $\delta$ cannot be constrained from P-wave reflection moveout, so it is kept at the actual value. We compute horizontal-space-lag extended images (Figure 12) and obtain the $\eta$-gradients for understated (0) and overstated (0.3) values of $\eta$. The $\eta$-errors induce residual energy in extended images (Figure 12) that has a linear (“V”-like) shape, which is typical for near-horizontal interfaces (Sava and Alkhalifah, 2012; Li et al., 2015). For both extrapolators, the extended images computed with the under-stated and even actual $\eta$-fields also contain considerable residual energy that spreads from the image point up to the surface. These kinematic artifacts, caused by illumination problems, may introduce bias in the DSO objective function and lead to false model updates.

First, we compute the sensitivity kernels using a single source-receiver pair and one image point (Figure 13). The kernels for the two operators (Figures 14 and 15) are consistent, but for $\eta = 0.3$ the FD implementation has an imprint of the shear-wave artifact near the image-point location (Figure 15b).

The gradients computed using the full surface acquisition geometry (Figure 16) and the entire extended image are shown in Figures 17 (mixed-domain operator) and 18 (FD operator). With either extrapolation operator, the gradient for the understated $\eta$-field is strongly influenced by the kinematic artifacts.
in the extended image. These artifacts are caused by having the reflector illuminated only from the surface. The contribution of the artifact is even larger than that of the residual induced by the $\eta$-error because the artifact is located closer to the physical sources and receivers. Such artifacts could be suppressed by tapering an extended image before applying a DSO-type penalty operator. However, this strategy is applicable only to simple models with few interfaces. A better approach requires proper accounting for illumination in the imaging or DSO penalty operators (Lameloise et al., 2015; Hou and Symes, 2015; Yang and Sava, 2015).

7 CONCLUSIONS
Wavefield extrapolation and gradient computation are key steps of wave-equation-based inversion algorithms. We implemented forward and adjoint mixed-domain extrapolation operators for acoustic VTI media based on a separable dispersion-relation approximation and derived the corresponding gradient expressions. This work is mostly focused on image-domain wavefield tomography, which is less susceptible to amplitude distortions produced by acoustic algorithms. However, because estimation of all three relevant VTI parameters (e.g., $V_{P0}$, $\epsilon$, and $\delta$) is seldom feasible using only P-wave reflection moveout, we also derived data-domain gradients, which can incorporate borehole information.

The gradients of the image- and data-domain objective functions were computed for several VTI models and different acquisition geometries. The similarity between the gradients obtained with the mixed-domain and FD operators validates our analytic results. However, the gradients computed with these two operators do not have exactly the same spatial distribution, which can be explained by the difference in amplitude variation along the simulated wavefronts. This difference becomes larger as the parameter $\eta$ increases. For sources and receivers placed in a layer with $\eta > 0$, the gradients obtained with FD algorithm contain an imprint of the shear-wave artifact.

Image-domain examples reveal illumination-related issues with the DSO objective function applied to cross-correlation extended images. Kinematic artifacts caused by insufficient illumination substantially contaminate the gradients and should be suppressed prior to updating the model.

Ongoing work involves improving the imaging and inversion steps of anisotropic image-domain tomography. Also, we are extending the proposed methodology to tilted TI media.

8 ACKNOWLEDGMENTS
We thank Paul Fowler, Jörg Schleicher, and the A(nisotropy)-and i(maging)-Teams at CWP for fruitful discussions. This work was supported by the Consortium Project on Seismic Inverse Methods for Complex Structures at CWP and the competitive research funding from King Abdullah University of Science and Technology (KAUST). The reproducible numeric examples in this paper are generated with the Madagascar open-source software package (Fomel et al., 2013a) freely available from http://www.ahay.org.

REFERENCES


Li, Y. E., B. Biondi, D. Nichols, and R. Clapp, 2016, Toward a closed loop from seismic imaging to earth-model building: The Leading Edge, 35, 135–139.


Zhan, G., R. C. Pestana, and P. L. Stoffa, 2012, Decoupled
equations for reverse time migration in tilted transversely
Zhang, Y., H. Zhang, and G. Zhang, 2011, A stable TTI re-
verse time migration and its implementation: Geophysics,
76, WA3–WA11.
Gradient computation for VTI acoustic wavefield tomography

Figure 1. VTI model with identical Gaussian anomalies in the parameters $\eta$ and $\epsilon$: (a) $V_{\text{hor}}$, (b) $\eta$, and (c) $\epsilon$ (model 1).

Figure 2. Source (red dot)-receiver (green dots) geometry for model 1.

Figure 3. Sensitivity kernels for model 1 computed using the mixed-domain extrapolator with different peak values of $\epsilon$: (a) 0 and (b) 0.3. The source is located at $x = 2$ km, $z = 0$ km and the receiver at $x = 2$ km, $z = 4$ km. The actual peak value is $\epsilon = 0.15$.

Figure 4. Sensitivity kernels for model 1 computed using the FD extrapolator with different peak values of $\epsilon$: (a) 0 and (b) 0.3. The actual peak value is $\epsilon = 0.15$. 
Figure 5. Gradients for model 1 computed using the mixed-domain extrapolator with different peak values of $\epsilon$: (a) 0 and (b) 0.3.

Figure 6. Gradients for model 1 computed using the FD extrapolator with different peak values of $\epsilon$: (a) 0 and (b) 0.3.

Figure 7. VTI model with a Gaussian anomaly in the parameter $V_{\text{hor}}$ and constant coefficients $\eta = \epsilon = 0.15$ (model 2). The peak value of the anomaly is 3.1 km/s and the background value is 2.7 km/s. Red and green dots mark the source and receiver locations, respectively.

Figure 8. Sensitivity kernels for model 2 computed using the mixed-domain extrapolator with different peak values of the $V_{\text{hor}}$-anomaly: (a) 2.7 and (b) 3.5 km/s. The actual peak value is 3.1 km/s.
Figure 9. Sensitivity kernels for model 2 computed using the FD extrapolator with different peak values of the $V_{\text{hor}}$-anomaly: (a) 2.7 and (b) 3.5 km/s. The actual peak value is 3.1 km/s.

Figure 10. Gradients for model 2 computed using the mixed-domain extrapolator with different peak values of the $V_{\text{hor}}$-anomaly: (a) 2.7 and (b) 3.5 km/s.

Figure 11. Gradients for model 2 computed using the FD extrapolator with different peak values of the $V_{\text{hor}}$-anomaly: (a) 2.7 and (b) 3.5 km/s.
Figure 12. Space-lag CIGs for model 3 computed at $x = 4$ km using the mixed-domain extrapolator with (a) $\eta = 0$, (b) $\eta = 0.15$ (actual value), and (c) $\eta = 0.3$.

Figure 13. Source (red dot), receiver (green dot), and image (blue dot) locations used in computing the sensitivity kernels for model 3.

Figure 14. Sensitivity kernels for model 3 computed using the mixed-domain extrapolator with (a) $\eta = 0$ and (c) $\eta = 0.3$. The actual $\eta = 0.15$.

Figure 15. Sensitivity kernels for model 3 computed using the FD extrapolator with (a) $\eta = 0$ and (c) $\eta = 0.3$. The actual $\eta = 0.15$. 
Gradient computation for VTI acoustic wavefield tomography

Figure 16. Source (red dots)-receiver (green dots) geometry used in computing the gradients for model 3.

Figure 17. Gradients for model 3 computed using the mixed-domain extrapolator with (a) $\eta = 0$ and (c) $\eta = 0.3$.

Figure 18. Gradients for model 3 computed using the FD extrapolator with (a) $\eta = 0$ and (c) $\eta = 0.3$. 