Pseudo-acoustic wavefield tomography with model constraints

Hui Wang and Paul Sava
Center for Wave Phenomena, Colorado School of Mines

ABSTRACT
Anisotropic wavefield tomography seeks to recover anisotropic parameters based on seismic wavefields. In terms of computational efficiency and degrees of freedom of the model parameters, the acoustic approximation to the anisotropic elastodynamics is more suitable than the full anisotropic elastic equations to the anisotropic tomography problem. We implement the wavefield tomography in VTI media based on the pseudo-acoustic coupled wave equations. Although three independent parameters are enough to describe acoustic VTI media, the choices of model parameterization remain case-dependent and ambiguous. To circumvent subjective choices and trade-offs in model parameterizations, we propose relating the model parameters using inequality constraints. Moreover, we show the potentials of using such inequality constraints to alleviate the acquisition signatures in anisotropic tomography. For example, one can mitigate the problem where, in VTI media, the data acquired from the cross-well acquisition geometry can hardly resolve phase velocities along the symmetry axis.

Key words: anisotropy, wavefield tomography, pseudo-acoustic, inequality constraints

1 INTRODUCTION

Seismic wavefield tomography is a general framework for model building using wavefield based techniques, and can be formulated both in the data and in the image domains (Sava, 2014). Full waveform inversion (FWI) directly manipulates the waveform data to obtain the physical properties of the Earth's subsurface (e.g., velocity). The method belongs to the data domain family of seismic wavefield tomography, and has seen a rapid usage increase due to expansion of computing resources (Pratt, 1999; Sirgue and Pratt, 2004; Plessix, 2009; Virieux and Operto, 2009; Symes et al., 2010). Seismic wavefield tomography can also be formulated in the image domain based on migrated images which serve as proxies for seismic wavefields simulated in the interior of the Earth. Typical migration velocity analysis (MVA) linearizes the seismic inversion process to some extent, and therefore provides a more stable way to attack the seismic inverse problem than waveform inversion.

Despite recent advances, seismic wavefield tomography is still challenging when applied to multi-parameter estimations. The huge number of unknown model parameters increases the size of the null space. The modeling engines with multi-parameter inversions, i.e., the wave equations in complex media, are typically more costly to implement than acoustic wave equations. Although computationally intensive, anisotropic seismic wavefield tomography has been discussed in the literature, in both the data and image domain. Ideally, anisotropic wavefield tomography ought to be implemented using elastic fields; in this paper we discuss an approximation to the approach based on the so-called pseudo-acoustic approximation (Alkhalifah, 1998, 2000). Li et al. (2011, 2014) extend acoustic wave equation migration velocity analysis (WEMVA) originally developed for acoustic waves (Sava and Biondi, 2004a,b) to the pseudo-acoustic case. Weibull and Arntsen (2013) discuss the application of WEMVA to field data. In the data domain, Warner et al. (2013) discuss the details of 3D full waveform inversion in anisotropic media. In this paper, we focus on the wavefield tomography formulated in the data domain and restrict out attention to the pseudo-acoustic wave equation for modeling because of its computational efficiency over the elastic wavefield extrapolators.

The coupled VTI system of partial differential equations (PDE) can be used to describe the pseudo-acoustic wave propagations (Fletcher et al., 2009). The family of various coupled systems of TI equations (Zhou et al., 2006a,b; Fletcher et al., 2008, 2009; Fowler et al., 2010; Duveneck and Bakker, 2011; Zhang et al., 2011) share the same characteristic, namely the same P-wave dispersion relation in TI media, which is referred to as the pseudo-acoustic approximation to the compressional waves in elastic media (Alkhalifah, 1998, 2000). These wave equations can accurately describe the kinematics of P-wave propagations in elastic media, with reduced computational costs compared to the original anisotropic elastic wave equations.

We implement our wavefield tomography method based

on forward and adjoint wavefield modeling using the coupled VTI system. The gradients of our selected objective functionals are then computed using the adjoint-state method (Tromp et al., 2005; Plessix, 2006). However, gradients calculated from conventional adjoint-state methods suffer from illumination issues, i.e., stronger energy at source and receiver positions and weaker energy at places away from them. Thus, we apply illumination compensation to the gradients by normalizing using source and receiver wavefields. In anisotropic wavefield tomography, different parameters have different sensitivity to the data residual given certain acquisition geometries. To avoid unphysical updates on certain model parameters, we introduce a set of inequality constraints to physically relate the model parameters such that the updates on one parameter boost updates on other parameters. With synthetic examples, we demonstrate that the gradients behave in a more stable and reliable way when preconditioned by illumination compensations and penalized by inequality model parameter constraints.

2 PSEUDO-ACOUSTIC WAVE PROPAGATORS

In this section, we discuss the forward and adjoint pseudo-acoustic wave propagators in discretized forms. Following Fletcher et al. (2009), the pseudo-acoustic wave propagation can be described by

$$\begin{cases} \partial_{tt} p = v_\perp^2 H_2[p] + v_\parallel^2 H_1[q] \\ \partial_{tt} q = v_{\rm n}^2 H_2[p] + v_\parallel^2 H_1[q], \end{cases} \tag{1}$$

where $p(\mathbf{x},t)$ and $q(\mathbf{x},t)$ are scalar wavefield variables; the TI medium is parameterized using $v_\parallel$ (P-wave velocity in the direction of the symmetry axis), $v_{\rm n}$ (P-wave NMO velocity), and $v_\perp$ (P-wave velocity in the direction normal to the symmetry axis); the spatial differential operators $H_1[\cdot]$ and $H_2[\cdot]$ are given by

$$\begin{aligned} H_1[\cdot] &= \sin^2\theta\cos^2\phi\,\partial_{xx} + \sin^2\theta\sin^2\phi\,\partial_{yy} + \cos^2\theta\,\partial_{zz} \\ &\quad + \sin^2\theta\sin 2\phi\,\partial_{xy} + \sin 2\theta\sin\phi\,\partial_{yz} \\ &\quad + \sin 2\theta\cos\phi\,\partial_{xz}, \\ H_2[\cdot] &= \partial_{xx} + \partial_{yy} + \partial_{zz} - H_1[\cdot], \end{aligned} \tag{2}$$

where $\theta(\mathbf{x})$ is the dip angle of the symmetry axis measured away from the vertical direction and $\phi(\mathbf{x})$ is the azimuth angle of the symmetry axis. The three P-wave velocities are related by the Thomsen parameters ϵ, δ, and the anellipticity parameter η:

$$\begin{aligned} v_{\rm n}^2 &= v_\parallel^2(1+2\delta), \\ v_\perp^2 &= v_\parallel^2(1+2\epsilon), \\ v_\perp^2 &= v_{\rm n}^2(1+2\eta), \end{aligned} \tag{3}$$

where η can be constructed from δ and ϵ:

$$\eta = \frac{\epsilon-\delta}{1+2\delta}. \tag{4}$$

Equations (1) are suitable for modeling 3D pseudo-acoustic TTI waves. However inverting for the tilt angles of the symmetry axis introduces complexities to the wavefield tomography problem since the equation is nonlinear with respect to the angles. For simplicity, in this paper we focus on the VTI wavefield tomography problem. In this case the spatial differential operators H_1 and H_2 reduce to

$$\begin{aligned} H_1[\cdot] &= \partial_{zz}, \\ H_2[\cdot] &= \partial_{xx} + \partial_{yy}. \end{aligned} \tag{5}$$

By adding the forcing terms f_p and f_q, we obtain the system

$$\begin{aligned} \partial_{tt} p - v_\perp^2\left(\partial_{xx}+\partial_{yy}\right)p - v_\parallel^2\partial_{zz} q &= f_p, \\ \partial_{tt} q - v_{\rm n}^2\left(\partial_{xx}+\partial_{yy}\right)p - v_\parallel^2\partial_{zz} q &= f_q, \end{aligned} \tag{6}$$

where $f_p(\mathbf{x},t)$ and $f_q(\mathbf{x},t)$ are the forcing terms acting on wavefield variables p and q. The physical meaning of p and q is still an open research question. One possible physical meaning is the one consistent with the linearized elastodynamic equations, i.e., p and q represent the normal stress along transverse and vertical directions (Duveneck and Bakker, 2011). In this case f_p and f_q are interpreted as acceleration sources. Notice that if we only have the knowledge of single component pressure wavfield data, the source injection of this scalar wave equation system becomes ambiguous. One can inject the source function either on p or q, or both on p and q. Following the normal stress interpretation of p and q, we decide to output the wavefield as $p+q$ for 2D ($2p+q$ for 3D), which resembles the pressure field (the trace of the stress tensor), and then simultaneously inject the source function on the p and q variable, i.e., $f_p = f_q = f$ for 2D ($f_p = 2f$, $f_q = f$ for 3D). However, the source injection and wavefield output choices are not arbitrary in the adjoint modeling. The adjoint system has to be numerically adjoint to the forward system, and therefore the source injection and wavefield output choices in the adjoint system have to be adjoint to those in the forward system.

In order to perform wavefield tomography, one need to derive the adjoint system associated with the forward modeling system (6). We choose the standard second-order in time, even order in space finite-difference discretization:

$$\begin{aligned} p^+ + p^- - 2p^o - v_\perp^2\left(\mathbf{D}_{xx}+\mathbf{D}_{yy}\right)p^o - v_\parallel^2\mathbf{D}_{zz}q^o &= f_p^o, \\ q^+ + q^- - 2q^o - v_{\rm n}^2\left(\mathbf{D}_{xx}+\mathbf{D}_{yy}\right)p^o - v_\parallel^2\mathbf{D}_{zz}q^o &= f_q, \end{aligned} \tag{7}$$

where p, q, and f denote vectors of dimension $N\times 1$, and $N = nx\times ny\times nz$ is the model grid size; sign $^+$ indicates the future time step, $^-$ indicates the previous time step, and ^{o} indicates current time step of the wavefield variables. The bold symbols are matrices; $\mathbf{D}_{xx}$, $\mathbf{D}_{yy}$ and $\mathbf{D}_{zz}$ are spatial finite-difference matrices along the x, y and z directions, respectively, of dimension $N\times N$. For convenience, the time sampling Δt is normalized to 1.

We define the unknown wavefield vector $P = [p^-,\ p^o,\ p^+,\ q^-,\ q^o,\ q^+]^T$, and the source function vector $F = [f_p^-,\ f_p^o,\ f_p^+,\ f_q^-,\ f_q^o,\ f_q^+]^T$. The discretization (7)

yields the linear system

$$\left[\begin{array}{ccc:ccc} \mathbf{I} & 0 & 0 & 0 & 0 & 0 \\ \mathbf{S}_p & \mathbf{I} & 0 & \mathbf{S}_q & 0 & 0 \\ \mathbf{I} & \mathbf{S}_p & \mathbf{I} & 0 & \mathbf{S}_q & 0 \\ \hdashline 0 & 0 & 0 & \mathbf{I} & 0 & 0 \\ \mathbf{T}_p & 0 & 0 & \mathbf{T}_q & \mathbf{I} & 0 \\ 0 & \mathbf{T}_p & 0 & \mathbf{I} & \mathbf{T}_q & \mathbf{I} \end{array}\right] \left[\begin{array}{c} p^- \\ p^o \\ p^+ \\ q^- \\ q^o \\ q^+ \end{array}\right] = \left[\begin{array}{c} f_p^- \\ f_p^o \\ f_p^+ \\ f_q^- \\ f_q^o \\ f_q^+ \end{array}\right], \tag{8}$$

where $\mathbf{I}$ is the identity matrix, and the elements S_p, S_q, T_p, T_q are block matrices in the form of

$$\begin{aligned} \mathbf{S}_p &= -2\mathbf{I} - v_\perp^2(\mathbf{D}_{xx} + \mathbf{D}_{yy}), \\ \mathbf{S}_q &= -v_\parallel^2 \mathbf{D}_{zz}, \\ \mathbf{T}_p &= -v_{\mathrm{n}}^2(\mathbf{D}_{xx} + \mathbf{D}_{yy}), \\ \mathbf{T}_q &= -2\mathbf{I} - v_\parallel^2 \mathbf{D}_{zz}. \end{aligned} \tag{9}$$

Rows three and six of equations (8) are simply the compact representations of equation (7). Notice that these block matrices are not symmetric in the case of variable velocities. Their transpose of the matrices are given by

$$\begin{aligned} \mathbf{S}_p^T &= -2\mathbf{I} - (\mathbf{D}_{xx} + \mathbf{D}_{yy})v_\perp^2, \\ \mathbf{S}_q^T &= -\mathbf{D}_{zz} v_\parallel^2, \\ \mathbf{T}_p^T &= -(\mathbf{D}_{xx} + \mathbf{D}_{yy})v_{\mathrm{n}}^2, \\ \mathbf{T}_q^T &= -2\mathbf{I} - \mathbf{D}_{zz} v_\parallel^2, \end{aligned} \tag{10}$$

assuming that the finite-difference matrices are symmetric. Notice that the transposed matrices first apply the velocity scaling operators, while the original matrices first apply the differential operators.

We transpose system (8) to get the discretized adjoint modeling system

$$\left[\begin{array}{ccc:ccc} \mathbf{I} & \mathbf{S}_p^T & \mathbf{I} & 0 & \mathbf{T}_p^T & 0 \\ 0 & \mathbf{I} & \mathbf{S}_p^T & 0 & 0 & \mathbf{T}_p^T \\ 0 & 0 & \mathbf{I} & 0 & 0 & 0 \\ \hdashline 0 & \mathbf{S}_q^T & 0 & \mathbf{I} & \mathbf{T}_q^T & \mathbf{I} \\ 0 & 0 & \mathbf{S}_q^T & 0 & \mathbf{I} & \mathbf{T}_q^T \\ 0 & 0 & 0 & 0 & 0 & \mathbf{I} \end{array}\right] \left[\begin{array}{c} f_p^- \\ f_p^o \\ f_p^+ \\ f_q^- \\ f_q^o \\ f_q^+ \end{array}\right] = \left[\begin{array}{c} p^- \\ p^o \\ p^+ \\ q^- \\ q^o \\ q^+ \end{array}\right]. \tag{11}$$

For equations (11), we need to solve the system bottom-up by updating the wavefield variables from $^+$ and ^{o} time steps to $^-$ time step. Thus the adjoint systems describe the time-reversal propagation of waves. By extracting row one and row four, and substituting the formulae S_p^T, S_q^T, T_p^T, and T_q^T, we arrive at a time-stepping update for the adjoint system:

$$\begin{aligned} f_p^- + f_p^+ - 2f_p^o - (\mathbf{D}_{xx} + \mathbf{D}_{yy})\left(v_\perp^2 f_p^o + v_{\mathrm{n}}^2 f_q^o\right) &= p^-, \\ f_q^- + f_q^+ - 2f_q^o - \mathbf{D}_{zz} v_\parallel^2 \left(f_q^o + f_p^o\right) &= q^-. \end{aligned} \tag{12}$$

We use sponge and one-way wave equations absorbing boundaries for the forward and adjoint system. The forward and adjoint pseudo-acoustic VTI wavefields are required for the adjoint-state method discussed next.

3 THE ADJOINT-STATE METHOD APPLIED TO PSEUDO-ACOUSTIC WAVE PROPAGATORS

The adjoint-state method summarized in Appendix A (Tromp et al., 2005; Plessix, 2006), can be applied to different wave equations. Here, we apply it to the pseudo-acoustic wave propagators. Three ingredients are required to perform the adjoint-state method: state equations, adjoint equations, and the objective functional.

The state VTI coupled system (6) can be written in a compact form:

$$\mathcal{F}(u^p, u^q) = \mathbf{L}\begin{bmatrix} u^p \\ u^q \end{bmatrix} + \begin{bmatrix} f^p \\ f^q \end{bmatrix} = 0, \tag{13}$$

where $u^p(\mathbf{x}, t)$ and $u^q(\mathbf{x}, t)$ are the state wavefield variables and the two-by-two operator $\mathbf{L}$ is defined as

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_{11} & \mathbf{L}_{12} \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{bmatrix} = \begin{bmatrix} v_\perp^2\left(\partial_{xx} + \partial_{yy}\right) - \partial_{tt} & v_\parallel^2 \partial_{zz} \\ v_{\mathrm{n}}^2\left(\partial_{xx} + \partial_{yy}\right) & v_\parallel^2 \partial_{zz} - \partial_{tt} \end{bmatrix}. \tag{14}$$

Similarly, the adjoint VTI coupled system yields

$$\mathbf{L}^T \begin{bmatrix} a^p \\ a^q \end{bmatrix} + \begin{bmatrix} g^p \\ g^q \end{bmatrix} = 0, \tag{15}$$

where $a^p(\mathbf{x}, t)$ and $a^q(\mathbf{x}, t)$ denote the adjoint wavefield variables; $g^p(\mathbf{x}, t)$ and $g^q(\mathbf{x}, t)$ are the forcing terms of the adjoint system; and the $\mathbf{L}^T$ is defined as

$$\mathbf{L}^T = \begin{bmatrix} \mathbf{L}_{11}^T & \mathbf{L}_{21}^T \\ \mathbf{L}_{12}^T & \mathbf{L}_{22}^T \end{bmatrix} = \begin{bmatrix} \left(\partial_{xx} + \partial_{yy}\right) v_\perp^2 - \partial_{tt} & \left(\partial_{xx} + \partial_{yy}\right) v_{\mathrm{n}}^2 \\ \partial_{zz} v_\parallel^2 & \partial_{zz} v_\parallel^2 - \partial_{tt} \end{bmatrix}. \tag{16}$$

Equations (13) and (15) are referred to as the state equations and the adjoint state equations, respectively, in the framework of the adjoint-state method.

The third ingredient of the adjoint-state method is the objective functional for the wavefield tomography problem. For a transmission-type full waveform inversion problem, a typical and widely used choice is the ℓ_2-norm measure of the difference between the modeled data and observed data, i.e.,

$$\mathcal{J} = \frac{1}{2}\left\| d - d^{\mathrm{obs}} \right\|_2^2. \tag{17}$$

Since we choose $u^p + u^q$ for 2D ($2u^p + u^q$ for 3D) as the output wavefield from the modeling, the 2D objective functional is

$$\mathcal{J} = \frac{1}{2}\left\| \mathbf{K}_r(u^p + u^q) - d^{\mathrm{obs}} \right\|_2^2, \tag{18}$$

and in 3D

$$\mathcal{J} = \frac{1}{2}\left\| \mathbf{K}_r(2u^p + u^q) - d^{\mathrm{obs}} \right\|_2^2, \tag{19}$$

where the operator $\mathbf{K}_r$ restricts the modeled wavefields to the known receiver position. According to the adjoint-state

method, the source functions in the adjoint equations (15) are

$$\begin{bmatrix} g^p \\ g^q \end{bmatrix} = \begin{bmatrix} \nabla_{u^p}\mathcal{J} \\ \nabla_{u^q}\mathcal{J} \end{bmatrix} = \begin{bmatrix} \mathbf{K}_r^T \\ \mathbf{K}_r^T \end{bmatrix} \left[\mathbf{K}_r(u^p + u^q) - d^{\text{obs}}\right] \quad (20)$$

in 2D, and

$$\begin{bmatrix} g^p \\ g^q \end{bmatrix} = \begin{bmatrix} \nabla_{u^p}\mathcal{J} \\ \nabla_{u^q}\mathcal{J} \end{bmatrix} = \begin{bmatrix} 2\mathbf{K}_r^T \\ \mathbf{K}_r^T \end{bmatrix} \left[\mathbf{K}_r(2u^p + u^q) - d^{\text{obs}}\right] \quad (21)$$

in 3D.

Wavefield tomography is essentially a PDE constrained optimization problem which can be posed as

$$\begin{aligned} &\underset{\mathbf{m}}{\text{minimize}} \;\; \mathcal{J}(\mathbf{m}, u^p, u^q), \\ &\text{subject to} \;\; \mathcal{F}(\mathbf{m}, u^p, u^q) = 0. \end{aligned} \quad (22)$$

For large scale seismic problems, one has to use the gradient based local optimization which requires knowledge of the gradients of the objective functional, i.e., $\nabla_{\mathbf{m}}\mathcal{J}$ at each iteration.

We define the model parameters as $\mathbf{m} = \{v_{\parallel}^2, v_{\text{n}}^2, v_{\perp}^2\}$. After performing the adjoint-state method, we get the following gradients:

$$\begin{aligned} \frac{\partial \mathcal{J}}{\partial v_{\parallel}^2} &= \sum_{e,\tau} \delta(\tau)\Big(\partial_{zz} u^q \star (a^p + a^q)\Big), \\ \frac{\partial \mathcal{J}}{\partial v_{\text{n}}^2} &= \sum_{e,\tau} \delta(\tau)\Big((\partial_{xx} + \partial_{yy}) u^p \star a^q\Big), \\ \frac{\partial \mathcal{J}}{\partial v_{\perp}^2} &= \sum_{e,\tau} \delta(\tau)\Big((\partial_{xx} + \partial_{yy}) u^p \star a^p\Big), \end{aligned} \quad (23)$$

where $\star$ represents temporal cross-correlation of the wavefields; the summation over τ indicates an accumulation of cross-correlated wavefields over time which extracts the zero-lag of the cross-correlation; and the summation over e indicates the need to sum over all shot experiments.

A typical algorithm of the adjoint-state method applied to pseudo-acoustic wave propagators consists of the following steps:

(i) Compute the state variables u^p and u^q by solving the state equations (13), i.e., forward modeling the seismic wavefields with known source functions f^p and f^q.

(ii) Compute the adjoint source functions g^p and g^q using equation (20) for 2D case (equation (21) for 3D case).

(iii) Compute the adjoint variables a^p and a^q by solving the adjoint equations (15), i.e., adjoint modeling the seismic wavefields with the computed adjoint source functions.

(iv) Compute the gradients of the objective function using equations (23).

We illustrate the above algorithm with a simple 2D model depicted in Figure 1. The velocity models have an isotropic water layer at the top and linearly increase below the water layer. The source is at $(0.32, 2.56)$ km and the receiver is at $(0.32, 17.92)$ km, which are shown as dots on the velocity models. The state variable is the linear combination of wavefields ($u^p + u^q$ for 2D, $2u^p + u^q$ for 3D) modeled from the forward system, as shown in figure 3 from top to bottom by increasing time. The trace shown in figure 2(a) is the modeled data with true veloc-

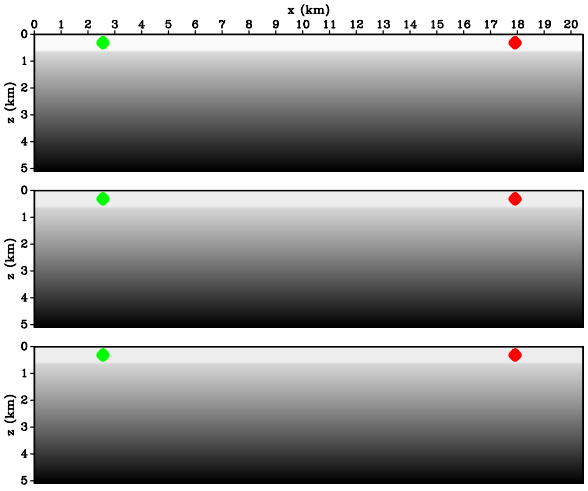
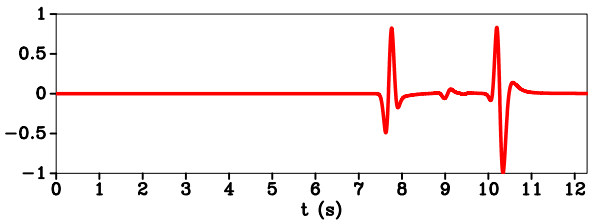
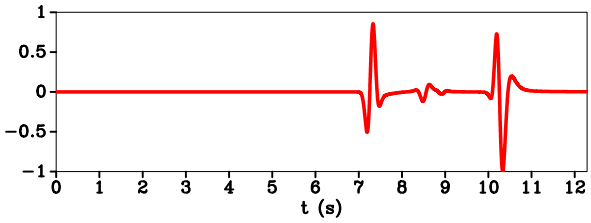


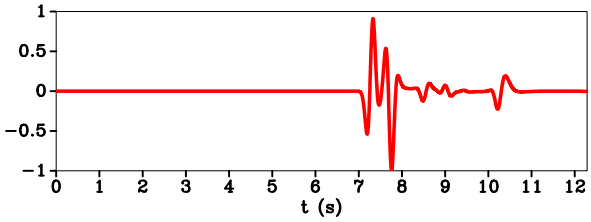
Figure 1. Top to bottom: $v_{\parallel}$, v_{n}, and $v_{\perp}$ velocity models. All three models have isotropic water layers on the top, and TI layers below the water layer. The velocities in TI layers increase linearly with depth.



(a)



(b)



(c)

Figure 2. Traces at the receiver position modeled with (a) true velocities and (b) perturbed velocities. Panel (c) is the difference of (a) and (b), and represents the adjoint source function used in the adjoint system to generate adjoint wavefield variables shown in figure 4.

ities at the receiver position. The recorded data contain three arrivals. The first arrival with time around $t = 7.8$ s is the diving wave; the second one is the head wave which arrives at $t = 9$ s; and the event arriving at about $t = 10.2$ s is the direct wave. As a comparison, figure 2(b) shows the modeled data with background velocities (3% positive perturbations on the three velocities), which is the state variable restricted to the receiver position. Because of the positive perturbations in velocities, the diving waves arrive earlier at around $t = 7.2$ s, and the head wave arrives at $t = 8.5$ s. The adjoint source is shown in figure 2(c) which is the difference of figure 2(a) and 2(b) in the 2D case of difference misfit functional. After subtraction, the residuals from the diving and head waves show that these arrivals do not fall within half cycles even with 3% small velocity perturbations because of the long offset geometry.

As indicated earlier, the adjoint variable ($a^p + a^q$ in 2D, and $2a^p + a^q$ in 3D) is obtained by extrapolating wavefields using the adjoint system. All pulses in the adjoint source are back-propagated in the adjoint wavefield variable and they interfere with the forward simulated state variable along several paths. These interferences are represented in the gradients shown in figure 5. First, wavefield coincidence occurs along the elliptical-shaped diving wavepaths. Second, the small portion of residual between two direct waves contributes to the wavefield coincidence along the path connecting directly the source with the receiver. Third, the residual of head waves coincidence along the water bottom interface. Note that the gradients in figure 5 are asymmetric, although in principle they should be for the example shown here. This is due to the imperfect absorbing boundaries used in this example. The waves enter the left-corner of models at pretty large gazing angles which causes the weak absorbing effects. These waves reflected back from the boundary interfere with the adjoint variables and distort the shapes of the gradients.

Another point to notice is that the gradients shown in figure 5 have stronger energy at the source and receiver positions and weaker energy at the deeper part of the models, which is often referred to as the illumination issue in the context of migrations. For migrations, illumination compensations through normalizations by source wavefields are often applied to address this issue. In the next section we show a similar technique to compensate for the imbalance of energy distribution of the gradients.

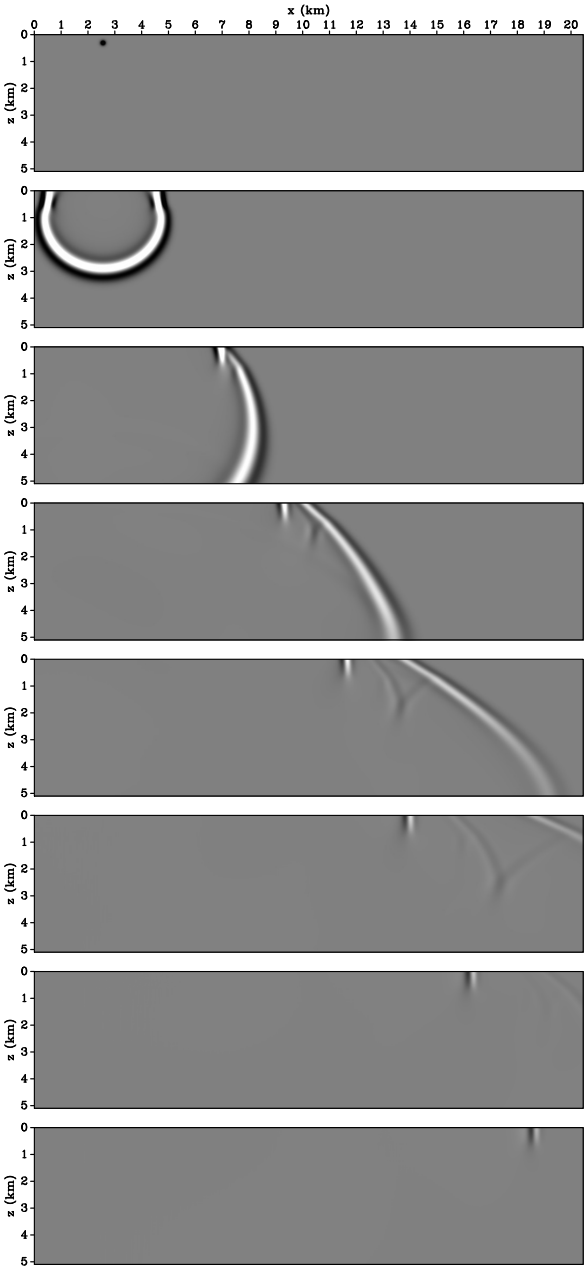


Figure 3. Top to bottom: the state wavefield variable $u^p + u^q$ ordered by increasing time. The waves emanate from the source, part of which travels in water layer as direct waves, and part of which propagate through the TI media as diving waves. Head waves start to appear after critical offset and travel at velocities between the direct and the diving waves.

4 ILLUMINATION COMPENSATION

Gradients represent the amount of model update; for a given data residual, we wish to update the models as evenly as possible rather than only update in certain areas (e.g., at source and receiver positions). This requires that we compensate for the source and receiver illuminations. In other words, one needs to suppress the energy (or amplitudes) near the source and receiver positions while enhancing the update away from them.

Intuitively, one can precompute illumination maps which reflect the energy distribution of source and receiver wavefields and then divide them from the gradients. One possible way to obtain the illumination map is to model the wavefield for a given source and take the envelope of the wavefield. However, we choose to directly divide the "filtered" state and adjoint wavefields from the gradients. By "filtered", we mean the state wavefields are applied with spatial derivative operators. More specifically, we apply three different scaling opera-

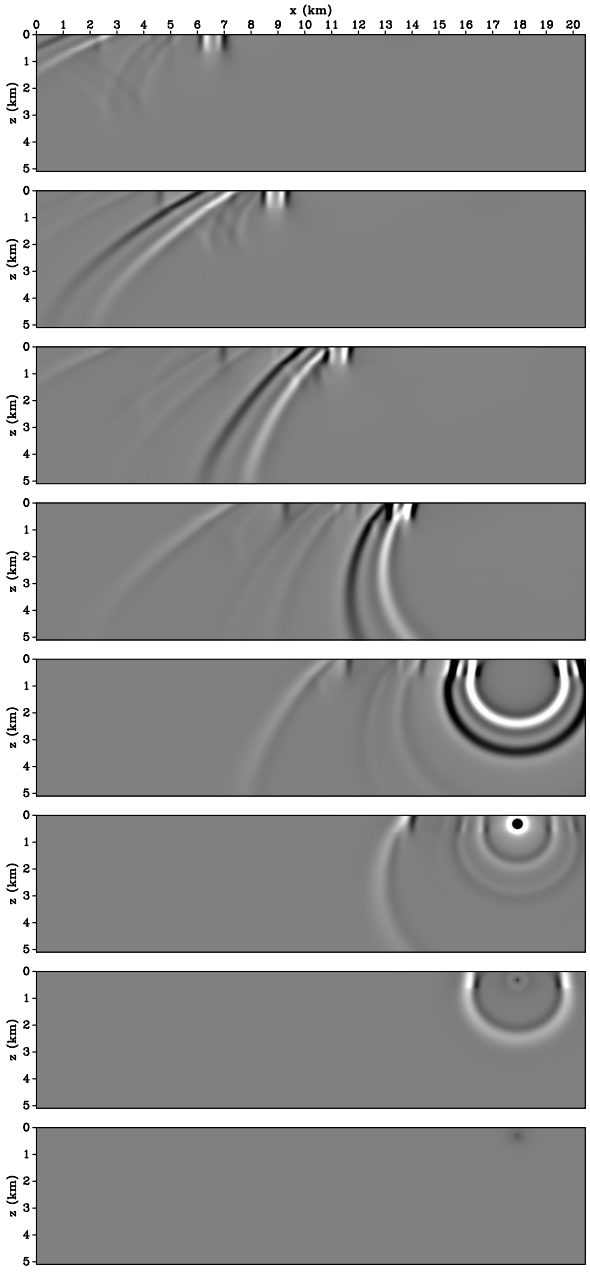


Figure 4. Top to bottom: adjoint wavefield variable $a^p + a^q$ ordered by increasing time. Wavefields are generated by time-reversal propagating the adjoint source function shown in figure 2(c). The waves start from the left part of the models and shrink to the receiver position on the right.

tors to equation (23), namely

$$\frac{\partial \mathcal{J}}{\partial v_{\parallel}^2} = \sum_e \frac{\sum_\tau \delta(\tau)\big(\partial_{zz} u^q \star (a^p + a^q)\big)}{\sqrt{\sum_t (\partial_{zz} u^q)^2 + \kappa}\,\sqrt{\sum_t (a^p + a^q)^2 + \kappa}},$$

$$\frac{\partial \mathcal{J}}{\partial v_{\mathrm{n}}^2} = \sum_e \frac{\sum_\tau \delta(\tau)\big((\partial_{xx} + \partial_{yy}) u^p \star a^q\big)}{\sqrt{\sum_t \big((\partial_{xx} + \partial_{yy}) u^p\big)^2 + \kappa}\,\sqrt{\sum_t \big(a^q\big)^2 + \kappa}},$$

$$\frac{\partial \mathcal{J}}{\partial v_{\perp}^2} = \sum_e \frac{\sum_\tau \delta(\tau)\big((\partial_{xx} + \partial_{yy}) u^p \star a^p\big)}{\sqrt{\sum_t \big((\partial_{xx} + \partial_{yy}) u^p\big)^2 + \kappa}\,\sqrt{\sum_t \big(a^p\big)^2 + \kappa}}, \tag{24}$$

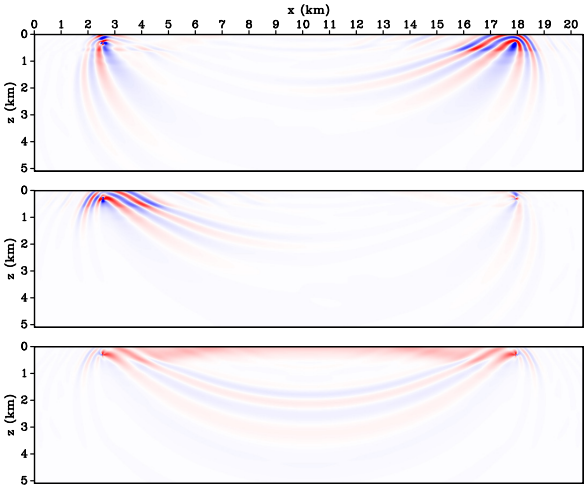


Figure 5. Top to bottom: gradients of difference-measured objective functional $\mathcal{J}$ (18) with respect to $v_{\parallel}^2$, v_{n}^2, and $v_{\perp}^2$. All gradients are generated using equation (23) given the true model shown in figure 1 and the background model with 3% positive perturbation relative to the true model. Energies are stronger at source and receiver positions than those at the positions away from sources and receivers.

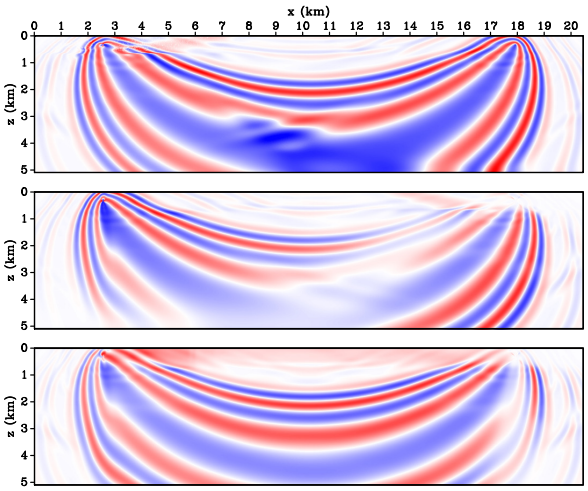


Figure 6. Top to bottom: gradients of difference-measured objective functional $\mathcal{J}$ with respect to $v_{\parallel}^2$, v_{n}^2, and $v_{\perp}^2$ after applying illumination compensation equations (24). All gradients are generated using the same model used for figure 5. The relative strength at the deeper parts are enhanced compared to figure 5.

where κ is a regularization factor to avoid zero division. Illumination compensation applied to the gradients in figure 5, leads to more balanced energy distribution, as shown in figure 6, where the deeper parts of the gradients brighten up. We also notice that the gradients become dimensionless quantities after applying the scaling operators. This is useful if we have gradients from model constraints which may have different units prior to normalizations.

5 INEQUALITY MODEL CONSTRAINTS

Aside from the illumination issue associated with the data misfit gradient, multi-parameter estimations have another challenge in model parameterizations. In the case of pseudo-acoustic wavefield tomography, three independent model parameters are needed to describe the media. However, different choices of three model parameters are possible. For example, one can parameterize the media using $v_\parallel$, δ, and ϵ, or $v_\parallel$, δ, and η, or $v_\parallel$, ϵ, and η, or $v_\parallel$, $v_\perp$, and v_n. Choices of parameterizations are based the formulations of anisotropic wave equations and the source and/or receiver acquisition geometry. Typical selection criteria are that the chosen model parameters should be sensitive to the perturbation in data misfit and the resultant gradients associated with the parameters are easy to compute. Following the second criterion, we choose $v_\parallel^2$, v_n^2, and $v_\perp^2$ as the model parameterizations for pseudo-acoustic media because the coupled VTI system is linear in these parameters. However, these parameters are sensitive to source and receiver acquisition geometries. For certain geometry, some of the parameters may not get updated through gradient-based iterations.

We illustrate this phenomena with a constant velocities example shown in figure 7. The sources are at $z = 0.16$ km and receivers are located at $z = 4.96$ km. Both of them are located in regions with elliptical anisotropic areas to avoid unwanted shear wave artifacts (associated with the nature of pseudo-acoustic approximation (Alkhalifah, 1998, 2000)). We use a plane source, i.e., all sources excite at the same time, to generate the vertically propagating plane waves. In this scenario, waves propagate using the phase velocity $v_\parallel^2$ and therefore, there is no chance for the waves to carry information about the other model parameters v_n^2 and $v_\perp^2$. We then perturb all three model parameters relative to the true model parameters (10% for $v_\parallel^2$ relative to the true $v_\parallel^2$; 46.7% for $v_\perp^2$ relative to the true $v_\perp^2$; and 32% for v_n^2 relative to the true v_n^2). The perturbations are chosen such that the background models fall close to the boundaries of inequality constraints, which will be explained later in this section. We then perform the adjoint state method to obtain the gradients for these parameters, shown in figure 8. As expected, only the gradient for $v_\parallel^2$ is non-zero since the recorded data are not impacted by the velocities v_n^2, and $v_\perp^2$. The inversion is thus ill-posed because the velocities v_n^2 and $v_\perp^2$ all lie in the null space of this particular objective function, under this chosen experiment geometry.

To address this problem, we add model constraints to constrain the anisotropic wavefield tomography process. The goal is a technique that connects the models updates as prescribed by the laws of linear elasticity. For example, in the above plane wave case, we wish to update v_n^2 and $v_\perp^2$ using the information from the updates on $v_\parallel^2$ through iterations. One possible solution to the problem is to incorporate the physical constraints between these three parameters.

According to equation (3), $v_\perp^2$, v_n^2 and $v_\parallel^2$ are related to the anisotropic parameters ϵ, δ. As a priori information, ϵ, δ can fall in the range of $(\epsilon_{\min}, \epsilon_{\max})$ and $(\delta_{\min}, \delta_{\max})$. Correspondingly, the three velocities can be contained using four

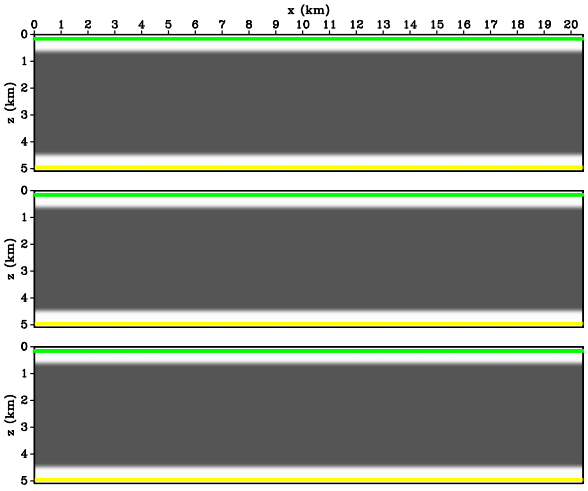


Figure 7. Top to bottom: $v_\parallel$, v_{pn}, and $v_\perp$ velocity models. All three models have isotropic water layers on the top and bottom, and constant velocity TI layers in the middle. Source and receiver positions are overlaid on the velocity models represented as dots (sources are at the top and receivers are at the bottom).

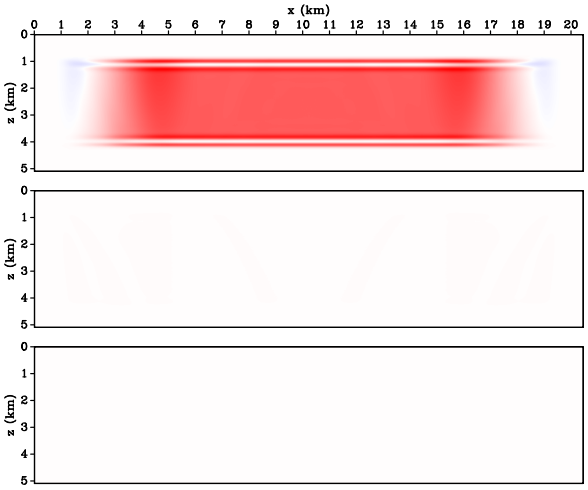


Figure 8. Top to bottom: gradients of difference objective functional $\mathcal{J}$ (18) with respect to $v_\parallel^2$, v_n^2, and $v_\perp^2$ without inequality model constraints. The gradients are generated using equation (23) given the true model shown in figure 7 and the background model with positive perturbations relative to the true model. Only the gradient for $v_\parallel^2$ is non-zero due to the vertically plane wave propagation.

inequality relations:

$$\begin{aligned} v_\mathrm{n}^2 - v_\parallel^2(1+2\delta_{\max}) &\leq 0, \\ v_\mathrm{n}^2 - v_\parallel^2(1+2\delta_{\min}) &\geq 0, \\ v_\perp^2 - v_\parallel^2(1+2\epsilon_{\max}) &\leq 0, \\ v_\perp^2 - v_\parallel^2(1+2\epsilon_{\min}) &\geq 0. \end{aligned} \tag{25}$$

Notice that the inequality conditions have to be satisfied at ev-

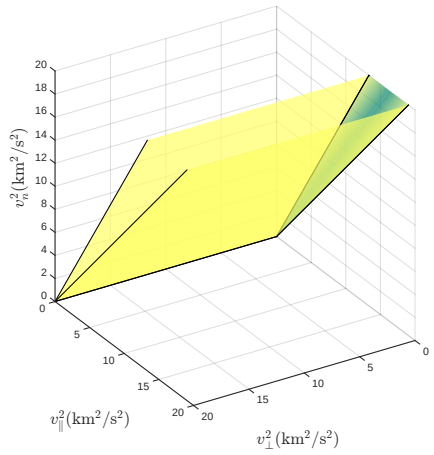


Figure 9. Inequality constraints of $\mathcal{C}_1$ and $\mathcal{C}_2$ shown in the model space $\{v_\parallel^2, v_\perp^2, v_\mathrm{n}^2\}$ based on the formula (26). Each of $\mathcal{C}_1$ and $\mathcal{C}_2$ represents a half-space, the intersection is a convex set formed by the two planes. The model parameters are feasible only if they are in the convex set. The shaded triangular area is a projection of the convex set onto the $v_\perp^2 = 0$ plane. The dark colors in the center of the triangular indicates small logarithmic barrier weights in that area. Closer to the boundaries of the triangular region, the weights increase as shown with brighter colors.

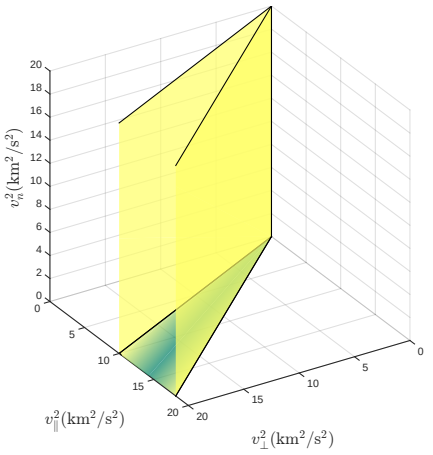


Figure 10. Inequality constraints of $\mathcal{C}_3$ and $\mathcal{C}_4$ shown in the model space $\{v_\parallel^2, v_\perp^2, v_\mathrm{n}^2\}$ based on the formula (26). Each of $\mathcal{C}_3$ and $\mathcal{C}_4$ represents a half-space, the intersection is a convex set formed by the two planes. The model parameters are feasible only if they are in the convex set. The shaded triangular area is a projection of the convex set onto the $v_\mathrm{n}^2 = 0$ plane. The dark colors in the center of the triangular indicates small logarithmic barrier weights in that area. Closer to the boundaries of the triangular region, the weights increase as shown with brighter colors.

ery grid point and therefore each inequality relation in equation (25) can be written in matrix form. Incorporating these inequality constraints, we can reformulate the pseudo-acoustic wavefield tomography as

$$\begin{aligned}
\underset{\mathbf{v}_\parallel^2,\mathbf{v}_\mathrm{n}^2,\mathbf{v}_\perp^2}{\text{minimize}}\ & \mathcal{J}(\mathbf{v}_\parallel^2, \mathbf{v}_\mathrm{n}^2, \mathbf{v}_\perp^2, u^p, u^q), \\
\text{subject to}\ & \mathcal{F}(\mathbf{v}_\parallel^2, \mathbf{v}_\mathrm{n}^2, \mathbf{v}_\perp^2, u^p, u^q) = 0, \\
& \mathcal{C}_1 = \mathbf{v}_\mathrm{n}^2 - (\mathbf{I} + 2\boldsymbol{\delta}_{\text{max}})\mathbf{v}_\parallel^2 \preceq 0, \\
& \mathcal{C}_2 = (\mathbf{I} + 2\boldsymbol{\delta}_{\text{min}})\mathbf{v}_\parallel^2 - \mathbf{v}_\mathrm{n}^2 \preceq 0, \\
& \mathcal{C}_3 = \mathbf{v}_\perp^2 - (\mathbf{I} + 2\boldsymbol{\epsilon}_{\text{max}})\mathbf{v}_\parallel^2 \preceq 0, \\
& \mathcal{C}_4 = (\mathbf{I} + 2\boldsymbol{\epsilon}_{\text{min}})\mathbf{v}_\parallel^2 - \mathbf{v}_\perp^2 \preceq 0,
\end{aligned} \tag{26}$$

where each velocity term is a vector and the symbol $\preceq$ represents element-wise less equal; $\mathbf{I}$ is the identity matrix; $\boldsymbol{\delta}_{\text{min}}$, $\boldsymbol{\delta}_{\text{max}}$, $\boldsymbol{\epsilon}_{\text{min}}$ and $\boldsymbol{\epsilon}_{\text{max}}$ are predefined diagonal matrices containing a priori information. Each inequality relation of (26) describes a half-space as shown in figures 9 and 10. In figure 9, the shaded triangular area in the $v_\perp^2 = 0$ plane is a projection of the convex set formed by the first two inequalities $\mathcal{C}_1$ and $\mathcal{C}_2$. The minimum and maximum slopes of the triangles are controlled by δ_{min} and δ_{max}. The triangle in figure 10 has a similar meaning but the minimum and maximum slopes are controlled by ϵ_{min} and ϵ_{max}. Notice that the shaded triangles in both figures have dark colors at the center, which indicates that the constraints are satisfied and the penalty associated with these constraints is small. In these areas, the inequalities are already satisfied and play little roles in constraining the model parameters. By combining the four inequalities, one can define a convex set in the model space where the model parameters are feasible (i.e., they satisfy the inequality constraints). The convex set is a pyramid with four faces shown in figure 11, which is just the intersection of four half-spaces from figure 9 and 10. As indicated by the colors of the triangles, the inequalities constrains the model parameters more when they fall closer to the boundaries of the triangles. The inequalities essentially push the model parameters towards the center of the convex set which is denoted by the red line in figure 11. The inequalities do not force the model to be on the red line, but rather in the vicinity.

The inequality constrained optimization problem (26) can be solved by transforming it into an equivalent unconstrained optimization problem:

$$\begin{aligned}
\underset{\mathbf{m}}{\text{minimize}}\ & \mathcal{J}(\mathbf{m}, u^p, u^q) + \sum_{i=1}^{4} I(\mathcal{C}_i(\mathbf{m})), \\
\text{subject to}\ & \mathcal{F}(\mathbf{m}, u^p, u^q) = 0,
\end{aligned} \tag{27}$$

where $I(u)$ is the indicator function

$$I(u) = \begin{cases} 0 & u \le 0 \\ \infty & u > 0 \end{cases}. \tag{28}$$

In general, the interior-point methods (Nocedal and Wright, 2006) can be used to solve such problem. We choose the logarithmic barrier method to proceed, which approximates the indicator function with a continuously differentiable function

$$I(u) \approx \begin{cases} -\frac{1}{t}\log(-u), & u < 0 \\ \infty & u > 0 \end{cases}, \tag{29}$$

where $t > 0$ is a parameter that controls the accuracy of ap-

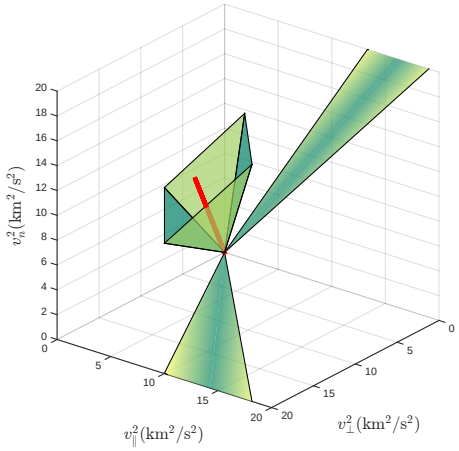


Figure 11. Inequality constraints of $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ and $\mathcal{C}_4$ shown in the model space $\{v_\parallel^2, v_\perp^2, v_\mathrm{n}^2\}$ based on the formula (26). The pyramid with four faces is the intersection of four half-spaces which are defined by the inequality constraints (26). The shaded triangular areas have the same explanations as figure 9 and 10. The red line represents the center of the pyramid, where the inequality constraints have the minimum weights.

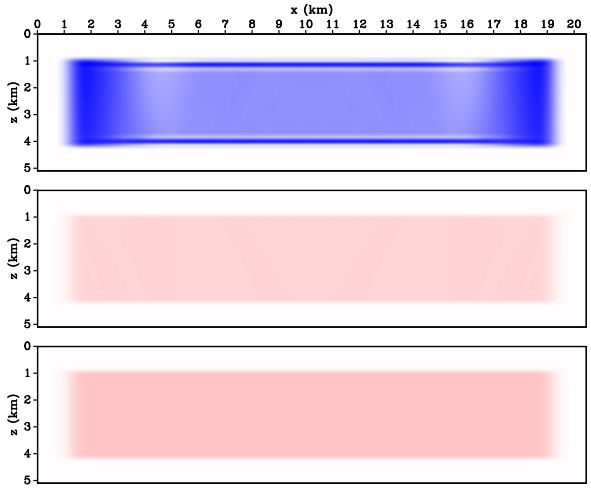


Figure 12. Top to bottom: gradients of difference objective functional $\mathcal{J}$ (18) with respect to $v_\parallel^2$, v_n^2, and $v_\perp^2$ after applying inequality model constraints with parameter $t = 0.8$. Red and blue color indicates positive and negative values, respectively. All panels are generated using the same model that is used for generating figure 8. The gradient for $v_\perp^2$ and v_n^2 are nonzero due to the logarithmic barrier in the objective functional which enforces physical constraints between the model parameters. The $v_\perp^2$ v_n^2 gradients are uniform is because the background model at current iteration is constant.

proximation. Notice that for each t, we get a solution $\mathbf{m}_t$, and the solution converges to the original solution $\mathbf{m}^\star$ as the parameter t increases. Hence barrier method requires an outer loop over t, and for each t one needs to solve an equality constrained optimization problem. In the context of seismic inversion, solving one seismic related constrained optimization problem is expensive and therefore we heuristically choose the parameter t and keep it fixed. We can justify this choice by the fact that our inequalities do not have to be satisfied strictly, i.e, we can relax the inequality constraints due to the uncertainty in our apriori. In this case we hope that the chosen logarithmic function we choose approximates the indicator function to provide enough accuracy. Substituting the logarithmic approximation (29) into (27), we arrive at

$$\begin{aligned} &\underset{\mathbf{m}}{\text{minimize}}\ \ \mathcal{J}(\mathbf{m}, u^p, u^q) - \frac{1}{t}\sum_{i=1}^{4}\log\left(-\mathcal{C}_i(\mathbf{m})\right), \\ &\text{subject to}\ \ \mathcal{F}(\mathbf{m}, u^p, u^q) = 0. \end{aligned} \tag{30}$$

We denote the logarithmic barrier term as $\mathcal{J}^B = -\frac{1}{t}\sum_{i=1}^{4}\log\left(-\mathcal{C}_i(\mathbf{m})\right)$, and its gradients with respect to model parameters as

$$\nabla_\mathbf{m}\mathcal{J}^B = -\frac{1}{t}\sum_{i=1}^{4}\frac{1}{\mathcal{C}_i(\mathbf{m})}\nabla_\mathbf{m}\mathcal{C}_i(\mathbf{m}), \tag{31}$$

Substituting the expression of $\mathcal{C}_i$ shown in (26) into (31), we obtain the gradients with respect to $v_\parallel^2$, v_n^2 and $v_\perp^2$:

$$\begin{bmatrix} \frac{\partial\mathcal{J}^B}{\partial v_\parallel^2} \\ \frac{\partial\mathcal{J}^B}{\partial v_\mathrm{n}^2} \\ \frac{\partial\mathcal{J}^B}{\partial v_\perp^2} \end{bmatrix} = -\frac{1}{t}\mathbf{A}^T\boldsymbol{\mathcal{C}}, \tag{32}$$

where the symbol ^{T} indicates a matrix transpose and the block matrix $\mathbf{A}^T$ is defined as

$$\mathbf{A}^T = \begin{bmatrix} -\mathbf{I} - 2\boldsymbol{\delta}_{\max} & \mathbf{I} + 2\boldsymbol{\delta}_{\min} & -\mathbf{I} - 2\boldsymbol{\epsilon}_{\max} & \mathbf{I} + 2\boldsymbol{\epsilon}_{\min} \\ \mathbf{I} & -\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & -\mathbf{I} \end{bmatrix}; \tag{33}$$

the vector $\boldsymbol{\mathcal{C}}$ is defined as

$$\boldsymbol{\mathcal{C}} = \begin{bmatrix} 1/\mathcal{C}_1 & 1/\mathcal{C}_2 & 1/\mathcal{C}_3 & 1/\mathcal{C}_4 \end{bmatrix}^T, \tag{34}$$

with $\mathcal{C}_i$ shown in (26).

We demonstrate the functionality of the inequality model constraints using the same plane-wave example shown in figure 7. By choosing the parameter $t = 0.8$ in the logarithmic barrier term of the objective functional, we obtain updates on the the model parameters $v_\perp^2$ and v_n^2, as indicated by the nonzero gradients in middle and bottom panels of figure 12. In this example, the perturbed velocities are higher than the true model velocities, thus the gradients from the data misfit should have positive signs, as shown in figure 8 for the $v_\parallel^2$ gradient. However, incorporating the inequality constraints, the $v_\parallel^2$ gradient become less positive or even negative as shown in top panel of figure 12. This is because the starting models fall close to the boundaries of $\mathcal{C}_1$ and $\mathcal{C}_3$, then model is strongly compelled away from these boundaries. That means the ratios $v_\mathrm{n}^2/v_\parallel^2$ and $v_\perp^2/v_\parallel^2$ should decrease if we assign enough weight to the logarithmic barrier misfit. Therefore $v_\parallel^2$ should increase meanwhile v_n^2 and $v_\perp^2$ should decrease according to the barrier. This explains why the gradients of v_n^2 and $v_\perp^2$ are positive

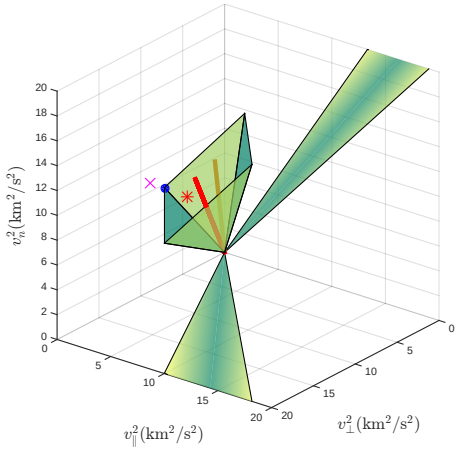


Figure 13. Model distributions overlaid on the inequality constraints for the plane wave example. The dot symbol represents background model position, and it lies close to the edges but inside the pyramid defined by the inequality constraints. The cross symbol is the model after first iteration only using the gradient from data misfit part, and is located outside of the pyramid meaning that it violates the physics controlled by the constraints. The star symbol is the model after first iteration using both gradients from data and logarithmic barrier misfit, and it lies inside the pyramid, thus indicating that the update satisfies the physical constraints.

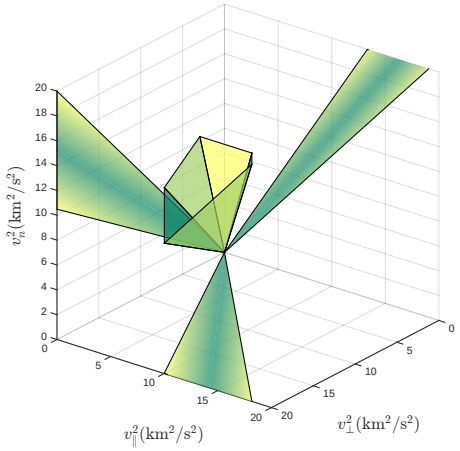


Figure 14. Additional inequality $v_\mathrm{n}^2 - v_\perp^2 \leq 0$ intersected with the pyramid shown in figure 11. This inequality further confines the solutions in a five-faces pyramid.

in figure 12 while the gradient of $v_\parallel^2$ is less positive or even negative. Notice that the gradients show uniform updates because the inequality constraints essentially use the models at current iteration and force the gradient to follow the same pattern, and the initial background models are constant positive perturbations relative to the constant true models.

To see whether the inequalities constrain the model parameters as expected, we run the inversion for one iteration. We take the grid points ranging from $x = 4$ to $x = 16$ km and from $z = 2$ to $z = 3$ km, and check the three velocities before and after the application of the inequality constraints. The three velocities are averaged over the chosen grid points and the result is then plotted in figure 13, overlaid with the inequality constraints. The original model parameters fall close to the edges of the pyramid, as denoted by the blue dot in figure 13, but are still inside the feasible convex set. The model parameters fall outside the convex set before applying inequality constraints which is represented as cross symbol. After applying the logarithmic barrier, the models are confined within the convex set represented as red star symbol, and they tend toward the center of the pyramid, as indicated by the red line.We thus conclude that the inequality constraints should force the model parameters to be in the feasible region at each iterations.

6 DISCUSSION

During the implementation of logarithmic barrier method, we fix the parameter t, which accounts for how accurately the logarithmic barrier functions approximate the indicator functions. In theory, this parameter t should evolve from small to large which results in a sequence of equality constrained optimizations to be solved. In order to minimize the computational cost, in this paper we use an equality constrained optimization for fixed t. However, we conjecture that the parameter t could vary within the equality constrained optimization such that the solution better approximates the original solution of the inequality constrained optimization. This requires theoretical proof which is out of the scope of this paper.

In this paper, we define the inequality relations for the anisotropic parameters using the phase velocities in VTI media. We only use the Thomsen parameters ϵ and δ to enforce the inequalities. However, we could in principle add an additional constraint on parameter η, which is related to ϵ and δ and thus not independent. Moreover, the pseudo-acoustic coupled wave equations require $\epsilon \geq \delta$ for stability. This requirement can cast as an inequality constraint, limiting the velocities of the half-space defined by $v_\mathrm{n}^2 - v_\perp^2 \leq 0$, as shown in figure 14. This inequality rules out the unstable model updates making the pseudo-acoustic wavefield modelings unstable during tomography.

7 CONCLUSIONS

We implement anisotropic wavefield tomography in VTI media based on the pseudo-acoustic coupled systems. The gradients of the objective functionals derived using the adjoint-state method are normalized by state and adjoint wavefield variables. This normalization accounts for both source and receiver illuminations, and thus results in even model updates. We use inequality constraints on the anisotropic phase velocities, to enforce known physical relations between the invented parameters. Such inequality constraints reduce the tomography nullspace when some model parameters are insensitive to the data misfit perturbations under particular acquisition geometries. The inequality constraints also shrink the model

space to a subspace describing physically reasonable parameters.

8 ACKNOWLEDGMENTS

We thank sponsor companies of the Consortium Project on Seismic Inverse Methods for Complex Structures. The reproducible numeric examples in this paper use the Madagascar open-source software package (Fomel et al., 2013) freely available from http://www.ahay.org. Thanks to Esteban Díaz for implementing the software interfaces to wavefield tomography solvers used in this paper, and to Yuting Duan for fruitful discussions on wavefield tomography. We are grateful to Ken Larner and Diane Witters for reviewing this paper.

REFERENCES

Alkhalifah, T., 1998, Acoustic approximations for processing in transversely isotropic media: Geophysics, **63**, 623–631.

———, 2000, An acoustic wave equation for anisotropic media: Geophysics, **65**, 1239–1250.

Duveneck, E., and P. M. Bakker, 2011, Stable P-wave modeling for reverse-time migration in tilted TI media: Geophysics, **76**, S65–S75.

Fletcher, R., X. Du, P. J. Fowler, et al., 2008, A new pseudo-acoustic wave equation for TI media: Presented at the 78th SEG Annual Meeting, Society of Exploration Geophysicists.

Fletcher, R. P., X. Du, and P. J. Fowler, 2009, Reverse time migration in tilted transversely isotropic (TTI) media: Geophysics, **74**, WCA179–WCA187.

Fomel, S., P. Sava, I. Vlad, Y. Liu, and V. Bashkardin, 2013, Madagascar: open-source software project for multidimensional data analysis and reproducible computational experiments: Journal of Open Research Software, **1**, e8.

Fowler, P. J., X. Du, and R. P. Fletcher, 2010, Coupled equations for reverse time migration in transversely isotropic media: Geophysics, **75**, S11–S22.

Li, Y., B. Biondi, R. Clapp, and D. Nichols, 2014, Wave-equation migration velocity analysis for VTI models: Geophysics, **79**, WA59–WA68.

Li, Y., B. Biondi, et al., 2011, Migration velocity analysis for anisotropic models: SEG Technical Program Expanded Abstracts, 201–206.

Lions, J. L., 1971, Optimal control of systems governed by partial differential equations problèmes aux limites: Springer.

Nocedal, J., and S. J. Wright, 2006, Least-squares problems: Springer.

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

———, 2009, Three-dimensional frequency-domain full-waveform inversion with an iterative solver: Geophysics, **74**, WCC149–WCC157.

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.

Sava, P., 2014, A comparative review of wavefield tomography methods: CWP Annual Report, 119–144.

Sava, P., and B. Biondi, 2004a, Wave-equation migration velocity analysis. I. theory: Geophysical Prospecting, **52**, 593–606.

———, 2004b, Wave-equation migration velocity analysis. II. Subsalt imaging examples: Geophysical Prospecting, **52**, 607–623.

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

Symes, W., et al., 2010, Waveform Inversion-progress and prospects: Presented at the 2010 SEG Annual Meeting, Society of Exploration Geophysicists.

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

———, 1987, Inversion of travel times and seismic waveforms, *in* Seismic tomography: Springer, 135–157.

———, 1988, Theoretical background for the inversion of seismic waveforms including elasticity and attenuation: Pure and Applied Geophysics, **128**, 365–399.

Tromp, J., C. Tape, and Q. Liu, 2005, Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels: Geophysical Journal International, **160**, 195–216.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.

Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Štekl, L. Guasch, C. Win, et al., 2013, Anisotropic 3D full-waveform inversion: Geophysics, **78**, R59–R80.

Weibull, W. W., and B. Arntsen, 2013, Anisotropic migration velocity analysis using reverse-time migration: Geophysics, **79**, R13–R25.

Zhang, Y., H. Zhang, and G. Zhang, 2011, A stable TTI reverse time migration and its implementation: Geophysics, **76**, WA3–WA11.

Zhou, H., G. Zhang, and R. Bloor, 2006a, An anisotropic acoustic wave equation for modeling and migration in 2D TTI media: Presented at the 76th Annual International Meeting, Society of Exploration Geophysicists.

———, 2006b, An anisotropic acoustic wave equation for VTI media: Presented at the 68th EAGE Conference & Exhibition.

9 APPENDIX A: THE ADJOINT-STATE METHOD

The adjoint-state method, first seen in the optimal control theory (Lions, 1971) and further introduced to geophysical community by Tarantola (1984, 1987, 1988), is a technique to explicitly evaluate the gradient of an objective functional (misfit functional). The main feature is that only two numerical simulations are required to calculate the gradient of the objective functional: one for state variables and the other for adjoint variables. Compared to the brute-force calculation of the Fréchet derivatives, which requires an equal number of numerical simulations and model parameters, the adjoint-state method makes the seismic inversion practical. The adjoint-state method follows the method of Lagrange multipliers, which is the necessary condition for an equality constrained optimization problem.

Consider the generic problem of minimizing a functional $\mathcal{J}(m,u)$ with the equality constraints $\mathcal{F}(u,m)=0$:

$$\begin{aligned} &\underset{m}{\text{minimize}} \;\; \mathcal{J}(m,u), \\ &\text{subject to} \;\; \mathcal{F}(m,u)=0. \end{aligned} \tag{A-1}$$

The method of Lagrange multipliers says the original problem can be equivalently posed as an unconstrained problem:

$$\underset{m}{\text{minimize}} \;\; \mathcal{H}(m,u,a)=\mathcal{J}(m,u)-a\mathcal{F}(u,m), \tag{A-2}$$

where the objective functional $\mathcal{J}(m,u)$ is augmented into a functional $\mathcal{H}(m,u,a)$ with higher dimensional independent variables. The variables a are the Lagrange multipliers associated with the equality constraints. Intuitively, there exists a Lagrange multiplier $\hat{a}$ that pulls the minimizer $\{\hat{m},\hat{u}\}$ of $\mathcal{J}(m,u)$ into a saddle point $\{\hat{m},\hat{u},\hat{a}\}$ of $\mathcal{H}(m,u,a)$. Following the definition of saddle point, one arrives at

$$\nabla_m \mathcal{H}(m,u,a)=0. \tag{A-3}$$

Therefore the gradient of $\mathcal{J}(m,u)$, i.e., $\nabla_m\mathcal{J}$, should be linearly proportional to the gradient of the equality constraints. More explicitly,

$$\nabla_m \mathcal{J}=a\nabla_m\mathcal{F}(u,m), \tag{A-4}$$

where a, u, and m are assumed independent. The variable a is feasible if it satisfies certain equations related to the definition of the saddle point of $\mathcal{H}(m,u,a)$:

$$\nabla_u \mathcal{H}(m,u,a)=0. \tag{A-5}$$

Substituting $\mathcal{H}(m,u,a)$ with its definition, one gets

$$a\nabla_u\mathcal{F}(m,u)=\nabla_u\mathcal{J}(m,u). \tag{A-6}$$

For linear elasticity, $\nabla_u\mathcal{F}(m,u)$ gives us an equation which relates a to m, and the equation happens to be the adjoint of the operator $\mathcal{F}$ applied on u:

$$\mathcal{F}^T(m,a)=\nabla_u\mathcal{J}(m,u). \tag{A-7}$$

Notice that in general the computation of a requires knowledge of u. One can obtain u by applying the last saddle point condition of $\mathcal{H}(m,u,a)$:

$$\nabla_a \mathcal{H}(m,u,a)=0, \tag{A-8}$$

which gives the original equality constraints, also known as the state equations:

$$\mathcal{F}(m,u)=0. \tag{A-9}$$

To sum up, the adjoint-state method can be explained as follows:

(i) Augment the original objective functional $\mathcal{J}$ into $\mathcal{H}$, shown in (A-2) by introducing the adjoint variables a. Notice that a has the same dimension as u.

(ii) Apply the saddle point condition of $\mathcal{H}$, i.e., $\nabla_{(a,u,m)}\mathcal{H}=0$, which gives the state equations (A-8), adjoint equations (A-7), and the gradient evaluation formulae (A-4).

(iii) Sequentially solve the state and adjoint equations.

(iv) Extract the gradient of the original objective functional by accessing the state and adjoint variables.