

Ray-illumination compensation for adjoint-state first-arrival traveltime tomography

Jiangtao Hu¹, Jianliang Qian², Junxing Cao¹, Xingjian Wang¹, Huazhong Wang³, and Shingyu Leung⁴

ABSTRACT

First-arrival traveltime tomography is an essential method for obtaining near-surface velocity models. The adjoint-state first-arrival traveltime tomography is appealing due to its straightforward implementation, low computational cost, and low memory consumption. Because solving the point-source isotropic eikonal equation by either ray tracers or eikonal solvers intrinsically corresponds to emanating discrete rays from the source point, the resulting traveltime gradient is singular at the source point, and we denote such a singular pattern the imprint of ray-illumination. Because the adjoint-state equation propagates traveltime residuals back to the source point according to the negative traveltime gradient, the resulting adjoint state will inherit such an imprint of ray-illumination, leading to singular gradient-descent directions when updating the velocity model in the adjoint-state traveltime tomography. To mitigate this imprint, we solve the

adjoint-state equation twice but with different boundary conditions: one being taken to be regular data residuals and the other taken to be ones uniformly, so that we are able to use the latter adjoint state to normalize the regular adjoint state and we further use the normalized quantity to serve as the gradient direction to update the velocity model; we call this process ray-illumination compensation. To overcome the issue of limited aperture, we have developed a spatially varying regularization method to stabilize the new gradient direction. A synthetic example demonstrates that our method is able to mitigate the imprint of ray-illumination, remove the footprint effect near source points, and provide uniform velocity updates along raypaths. A complex example extracted from the Marmousi2 model and a migration example illustrate that the new method accurately recovers the velocity model and that an offset-dependent inversion strategy can further improve the quality of recovered velocity models.

INTRODUCTION

Reliable velocity models are crucial for seismic imaging and are usually inferred from kinetic properties of seismic waves, such as first-arrival or multiarrival traveltimes. First-arrival traveltime usually captures the minimum traveltime of direct, refracted, or diving waves. Because these seismic waves normally propagate in the near-surface region, first-arrival traveltime tomography is widely used to build velocity models for such regions (Zhu et al., 1992; Zelt and Barton, 1998). Because reliable near-surface velocity

models are useful in at least three situations — providing an initial model for the full-waveform inversion to obtain a further improved velocity model, providing a velocity structure for near-surface migration velocity analysis, and providing a velocity structure for near-surface static correction, which greatly affects the quality of imaging in areas with complex surface conditions (Zhu et al., 1992) — we are motivated to develop effective first-arrival traveltime tomography methods.

First-arrival traveltime tomography is a classic data-fitting nonlinear inverse problem that needs a forward-modeling method to

Manuscript received by the Editor 4 March 2020; revised manuscript received 15 April 2021; published ahead of production 17 June 2021; published online 18 August 2021.

¹State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation (Chengdu University of Technology), Key Laboratory of Earth Exploration and Information Techniques (Chengdu University of Technology), Ministry of Education, No. 1, Dongsanlu, Erxianqiao, Chengdu 610059, China. E-mail: jiangtao_hu@126.com (corresponding author); caojx@cdut.edu.cn; wangxj@cdut.edu.cn.

²Michigan State University, Department of Mathematics and Department of CMSE, East Lansing, Michigan 48824, USA. E-mail: qian@math.msu.edu.

³Tongji University, WPI, School of Ocean and Earth Science, 1239 Siping Road, Shanghai 200092, China. E-mail: herbhuak@vip.163.com.

⁴Hong Kong University of Science and Technology, Department of Mathematics, Hong Kong SAR 999077, China. E-mail: masyleung@ust.hk.

© 2021 Society of Exploration Geophysicists. All rights reserved.

generate first-arrival traveltimes from the current model and a method to compute the gradient of the data-fitting functional with respect to the current model. Various methods can be used to obtain the first-arrival traveltimes, such as ray-tracing techniques (Vinje et al., 1993; Lambaré et al., 1996; Červený, 2001), finite-difference eikonal solvers (Vidale, 1988; Podvin and Lecomte, 1991; van Trier and Symes, 1991; Qin et al., 1992; Gray and May, 1994; Sethian and Popovici, 1999; Qian and Symes, 2002; Kao et al., 2004; Zhao, 2005; Qian et al., 2007a, 2007b), and Fermat's principle-based methods (Schneider et al., 1992; Wang et al., 1999; Hu et al., 2017, 2018). Ray tracing computes traveltimes along raypaths so that the resulting traveltimes are on irregular grids. When a velocity model has complex structures, ray-tracing techniques may suffer from shadow zones. The wavefront construction method (Vinje et al., 1993; Lambaré et al., 1996) alleviates this problem by carrying out ray interpolation. However, finite-difference eikonal solvers and Fermat's principle-based methods produce traveltimes at regular grid points including all regularly spaced receivers so that it is convenient to use these methods to update traveltime misfits at receivers.

Because the traveltime tomography problem is in essence an eikonal-constrained optimization problem, naturally there are two possible approaches for computing the gradient of the traveltime-misfit functional with respect to the (current) slowness function: One is the primal formulation and the other is the dual (adjoint-state) formulation. The primal approach is based on linearization (first-order perturbation) of the traveltime with respect to slowness along raypaths, which states that the first-order traveltime perturbation is the integral of the slowness perturbation along the reference (current) ray; this is in fact a consequence of Fermat's least-traveltime principle (Snieder and Sambridge, 1992). Therefore, because the resulting gradient of the traveltime with respect to slowness is defined through a well-behaved integral of slowness perturbation along the ray, after discretization the Jacobian of traveltime with respect to slowness is a sparse matrix, in which nonzero entries are lengths of ray segments and may appear in arbitrary locations due to the underlying possibly twisted raypaths (Peterson et al., 1985; Berryman, 1989; Bregman et al., 1989; Nolet, 1993; Wang and Braile, 1995; Hu and Marcinkovich, 2012). Because of this irregular sparsity of the Jacobian matrix, it is challenging to manage storage of this matrix at large scale, giving rise to implementational difficulties for large-scale seismic applications; see Nolet (2012) for an overview of seismic tomography and many applications.

Conversely, the dual or adjoint-state formulation introduces the adjoint state (or the Lagrange multiplier) as a dual variable to the eikonal so that the eikonal equation as a constraint comes into play directly when minimizing the traveltime-misfit functional (Leung and Qian, 2006); consequently, the gradient of the traveltime-misfit functional with respect to the slowness is obtained from the adjoint-state variable defined in the entire space rather than on individual rays. Because of this dual property, the dual variable satisfies a linear conservative equation, the so-called adjoint-state equation (Sei and Symes, 1994, 1995; Leung and Qian, 2006). To develop an efficient algorithm to implement this dual formulation, Leung and Qian (2006) propose a fast sweeping method to solve this adjoint-state equation; the resulting framework is easy to implement, and it is memory efficient and computationally efficient. Since then, the adjoint-state tomography method has undergone ex-

tensive development (Taillandier et al., 2009; Noble et al., 2010; Li and Leung, 2013; Li et al., 2014; Waheed et al., 2016).

Because solving the point-source isotropic eikonal equation either by ray tracers or by eikonal solvers intrinsically corresponding to emanating discrete rays from the source point, the resulting traveltime gradient is singular at the source point, and we term such a singular pattern the imprint of ray-illumination. Because traveltime tomography is based on either ray tracers or eikonal solvers, primal and dual approaches to traveltime tomography face the issue of imprint of ray-illumination. For example, to overcome the nonconvergence difficulty in applying the algebraic reconstruction technique to straight-ray traveltime tomography, Dines and Lytle (1979) develop a minimax-correction-based simultaneous iterative-reconstruction technique (SIRT) to account for the ray density map by updating the velocity model only after accumulating the updates from all available rays, which mitigates the effect of imprint of ray-illumination. Many works on ray-tracing traveltime tomography have developed various strategies to deal with the ray-coverage issue in traveltime tomography (Peterson et al., 1985; Berryman, 1989; Bregman et al., 1989; Nolet, 1993; Wang and Braile, 1995; Hu and Marcinkovich, 2012).

Because the adjoint-state equation propagates traveltime residuals back to the source point according to the reversed ray direction, namely, the negative traveltime gradient from the eikonal equation, the resulting adjoint state will inherit a similar imprint of ray-illumination as the primal approach does. This imprint of ray-illumination leads to singular gradient-descent directions in adjoint-state traveltime tomography. To mitigate this imprint, we propose to solve the adjoint-state equation twice but with different boundary conditions: one being taken to be regular data residuals and the other taken to be ones uniformly, so that we are able to use the latter adjoint state to normalize the regular adjoint state and we further use the normalized quantity to serve as the gradient direction to update the slowness model. We call this process the ray-illumination compensation. We mention that a similar idea has been used in Benaïchouche et al. (2015) for estimating the Hessian of the misfit functional with respect to the slowness. Although our formulation is analogous to that in Benaïchouche et al. (2015), our motivation is different from that in Benaïchouche et al. (2015); moreover, we emphasize that we also provide rigorous mathematical derivation to justify our motivation, and such mathematical justification is missing in Benaïchouche et al. (2015). We will show the mathematical meaning of this normalized quantity. In addition, a spatially varying regularization method is proposed to stabilize the ray-illumination compensation. As we can see from the analysis of our illustrative example, our normalization strategy just distributes the boundary traveltime residual uniformly along a raypath without considering the ray length. Therefore, our strategy is different from the minimax-correction-based SIRT (Dines and Lytle, 1979) in that the latter uses the ray length to balance the velocity updating process so that longer rays tend to dominate, smearing any zones of anomalous velocity (Peterson et al., 1985). Hence, to some extent, our normalization method provides a mathematically justified strategy to compensate for the imprint of ray-illumination in adjoint-state traveltime tomography.

This paper is organized as follows. First, the methodology of adjoint-state first-arrival traveltime tomography is analyzed to illustrate the effect of ray-illumination, and we further propose to solve the adjoint-state equation with a different boundary condition

so that we obtain a normalized gradient direction of the misfit functional. We then discuss the regularization method with spatially varying regularization factors. Finally, numerical experiments on two synthetic data sets demonstrate the feasibility of the proposed method.

ADJOINT-STATE FIRST-ARRIVAL TRAVELTIME TOMOGRAPHY

Adjoint-state equation for gradient

We start from the point-source eikonal equation to consider the first-arrival traveltime tomography in terms of an adjoint-state (dual) formulation. The traveltime misfit functional for a single shot takes the following form (Sei and Symes, 1994, 1995; Leung and Qian, 2006):

$$J[v] = \frac{1}{2} \int_{\Gamma} |T(\mathbf{x}; \mathbf{x}_s; v) - T_{\text{obs}}(\mathbf{x}; \mathbf{x}_s)|^2 d\Gamma + \frac{1}{2} \int_{\Omega} \lambda(\mathbf{x}) \left(|\nabla T(\mathbf{x}; \mathbf{x}_s)|^2 - \frac{1}{v(\mathbf{x})^2} \right) d\mathbf{x}, \quad (1)$$

where Ω is the subsurface domain with boundary $\partial\Omega$, $\Gamma \subset \partial\Omega$ is the acquisition surface, $d\Gamma$ is the surface differential element on Γ , T_{obs} is the observed first-arrival traveltime data on Γ due to the source $\mathbf{x}_s$, T is the first-arrival traveltime due to the velocity v , $\mathbf{x} = (x, z)$ is the spatial coordinate, λ is the adjoint-state variable (the Lagrange multiplier), and $\mathbf{x}_s$ is the shot location.

To minimize the misfit function $J[v]$ with respect to v , we need to find its gradient, which can be obtained from the following first-order variation of J in terms of the first-order variation δv (Leung and Qian, 2006):

$$\delta J = \left\langle \frac{\partial J}{\partial v}, \delta v \right\rangle_{L^2} = \int_{\Omega} \frac{\delta v}{v^3(\mathbf{x})} \lambda(\mathbf{x}) d\mathbf{x}, \quad (2)$$

so that

$$\frac{\partial J}{\partial v} = \frac{\lambda(\mathbf{x})}{v^3(\mathbf{x})}, \quad (3)$$

where $\langle \cdot, \cdot \rangle_{L^2}$ indicates the inner product in the L^2 space. The adjoint-state variable λ satisfies the following linear conservative advection equation (Leung and Qian, 2006):

$$-\nabla \cdot (\lambda(\mathbf{x}) \nabla T(\mathbf{x})) = 0 \quad \text{in } \Omega, \quad (4)$$

$$\mathbf{n}(\mathbf{x}) \cdot \nabla T(\mathbf{x}) \lambda(\mathbf{x})|_{\Gamma} = T_{\text{obs}} - T, \quad (5)$$

where $\mathbf{n}$ is the outward normal to Γ . Equations 4 and 5 can be solved by the fast sweeping method designed in Leung and Qian (2006).

Consequently, we take the negative gradient direction to update the velocity model v so as to minimize $J[v]$,

$$v^{(k+1)}(\mathbf{x}) = v^{(k)}(\mathbf{x}) - \gamma \frac{\partial J}{\partial v}, \quad (6)$$

where k is the iteration number with $k = 0, 1, 2, \dots$ and γ is the step length. We use a three-point parabolic interpolation method (Vigh et al., 2009) to determine the step length.

To carry out this adjoint-state tomography strategy, we need an efficient method to compute the traveltime and its gradient because the traveltime gradient is used as known coefficients to back propagate the traveltime residual into the entire computational domain. In practice, we may choose any efficient eikonal solver to solve the eikonal equations, but we choose a Fermat's principle-based fast marching method (Hu et al., 2017), which is a variant of the popular fast marching method developed by Sethian and Popovici (1999) method. This method yields highly accurate traveltimes at each grid point by solving a nonlinear optimization problem, so that the traveltime gradient is also highly accurate.

Adjoint-state equation for ray-illumination

As we allude in the preceding section, similar to the primal approach, the adjoint-state defined gradient direction has singular behavior at the source due to the imprint of ray-illumination. The discrete acquisition geometry further amplifies this singular behavior, causing the so-called footprint phenomenon in the usual eikonal-based first-arrival traveltime tomography (Leung and Qian, 2006). Such footprint effects can be depressed by heavy regularization at the cost of degraded overall resolution.

We use a simple velocity model to demonstrate the behavior of this gradient. The exact velocity model is shown in Figure 1a. The

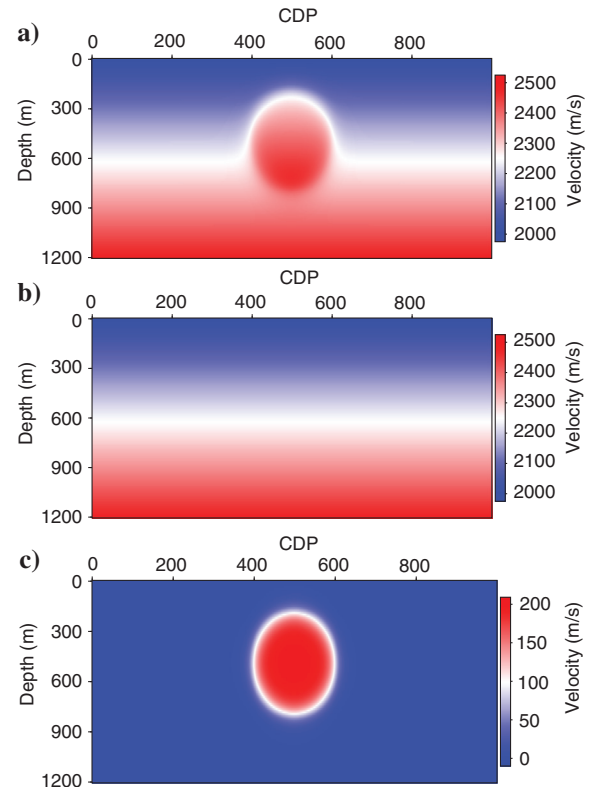


Figure 1. Velocity model with elliptical velocity perturbation. (a) The exact velocity model, (b) the velocity model defined by a linear function in the spatial coordinates, and (c) the elliptical velocity perturbation.

model is 1.2 km in depth and 10 km laterally. The grid spacing in the x - and z -axes is 10 m. This model is a perturbed velocity model, in which the background consists of a velocity model defined by a linear function in the spatial coordinates (Figure 1b) and the added perturbation is defined by an elliptical shape (Figure 1c). The source and receivers are located on the surface. We simulate 80 shots using the exact velocity model (Figure 1a) at a 100 m shot interval. The lateral position of the first shot is located at $x = 1000$ m (i.e., common depth point [CDP] 100). The maximum offset for each shot is 7 km, and the receiver interval is 10 m. The velocity model with a constant gradient (Figure 1b) is used as the initial model in tomography. Therefore, the inversion target is to recover the elliptical velocity perturbation in Figure 1c.

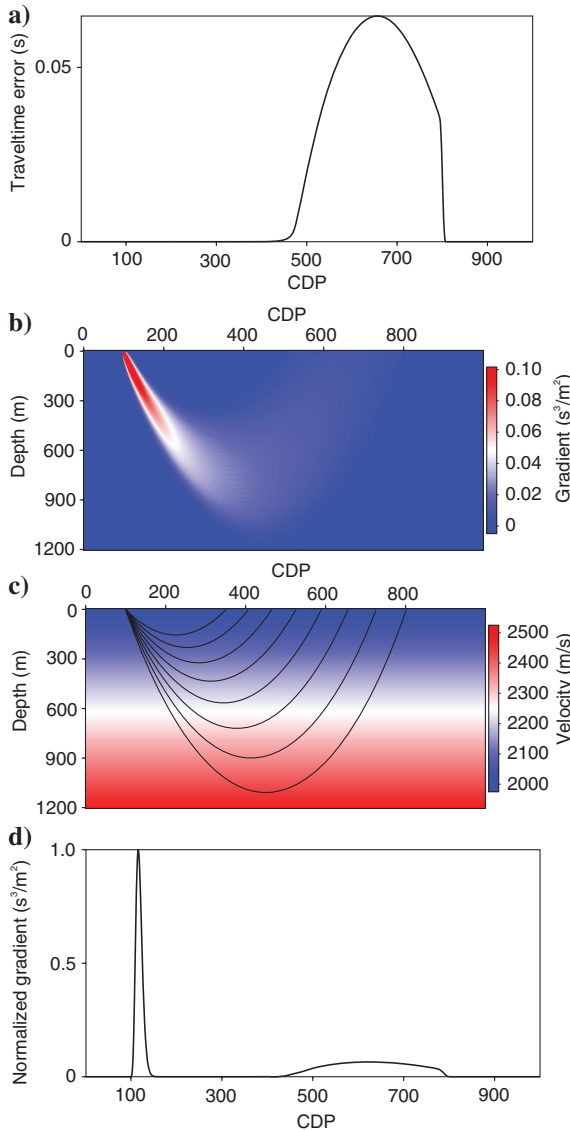


Figure 2. Illustration of gradient for a single source. The (a) traveltime error, (b) normalized gradient, (c) raypath for a single source located at $x = 1000$ m, and the (d) normalized gradient's common depth curve at a depth of 100 m. Because the gradient near the source is too strong, we clip the maximum value in (b) for the purpose of display.

We show the gradient of the misfit functional due to the first shot. We first calculate the first-arrival traveltimes from the current model, which are compared with the observed first-arrival traveltimes to obtain traveltime residuals, as shown in Figure 2a. These traveltime residuals are used to solve equation 4 to obtain the adjoint-state variable, which is then divided by v^3 to obtain the gradient of the misfit functional for this shot, as shown in Figure 2b. Figure 2c also shows a part of raypaths for this shot, and the primal approach back propagates the traveltime residuals along these raypaths. The gradient in Figure 2b shows that the adjoint-state method back propagates the traveltime residuals in a similar way using the adjoint-state equation. Although this is a valid way of back propagating traveltime residuals in traveltime tomography, the amplitude distribution of this gradient is not appealing. The gradient near the source is quite strong, and this may cause footprints in the overall gradient in the setting of discrete acquisition geometry. Moreover, the gradient is nonuniformly distributed along the raypath, and it gets stronger as it moves from the receivers to the source point, as shown in Figure 2b and 2d. However, the strength of the desired gradient should reveal the magnitude of the velocity errors. The functional gradient in Figure 2b shows that the velocity of the neighborhood near the source will be updated in a distorted or biased manner that is not suitable. It also can be observed that the strength of the functional gradient in Figure 2b is similar to the ray-illumination in Figure 2c. The strength of the functional gradient is weak near the neighborhood of a receiver because the ray coverage at each receiver is coarse, and the strength of the gradient is strong near the source neighborhood because the rays converge. To understand the behavior of the adjoint-state gradient, the adjoint-state equation should be analyzed.

Singular behavior of the adjoint state: An example

We consider the eikonal equation on a unit sphere $\Omega = B(\mathbf{0}, 1)$ for $d = 2, 3$, where Ω is a unit disk when $d = 2$ and a unit sphere when $d = 3$:

$$|\nabla T| = s_0 \quad \text{on } \Omega \setminus \{\mathbf{0}\}, \quad T|_{\mathbf{x}=\mathbf{0}} = 0, \quad (7)$$

where the slowness s_0 is constant, and the source is set at the origin.

To carry out traveltime tomography, we need to solve the following adjoint-state equation after solving the eikonal equation:

$$-\nabla \cdot (\lambda \nabla T) = 0, \quad \frac{\partial T}{\partial \mathbf{n}} \lambda \Big|_{\partial \Omega} = f, \quad (8)$$

where f is the traveltime residual and $\partial \Omega = \partial B(\mathbf{0}, 1)$. We denote the solution of this equation by λ^f .

To normalize the singular behavior of the adjoint state, we propose to solve the adjoint-state equation 8 once more by setting $f \equiv 1$, and we denote the resulting solution by λ^1 . It is easy to check that

$$\lambda^1 = 1/(s_0 |\mathbf{x}|^{d-1}), \quad (9)$$

which is singular at the source (the origin). Such a singular behavior holds in general for the solution of the adjoint-state equation.

Then, we propose to normalize λ^f by λ^1 so that we have a new function

$$\beta(\mathbf{x}) = \lambda^f(\mathbf{x}) / \lambda^1(\mathbf{x}), \quad (10)$$

which is nonsingular. To see this, in fact we can show that

$$\beta(\mathbf{x}) = f\left(\frac{\mathbf{x}}{|\mathbf{x}|}\right). \quad (11)$$

First, we construct a tube of rays ending at $\mathbf{x}$ and $\mathbf{x}/|\mathbf{x}|$. Because, in an isotropic medium, the point-source first-arrival ray direction is defined by the traveltime gradient, in our constant case here, $\nabla T(\mathbf{x})$ is parallel to the vector $\mathbf{x}$ which is the ray from the source at the origin to the point $\mathbf{x}$. Thus, the tube of rays is a cylinder whose longitudinal direction is parallel to $\nabla T(\mathbf{x})$ and transversal direction is perpendicular to $\nabla T(\mathbf{x})$. We denote the ray tube by G .

Second, we integrate the adjoint-state equation 8 around the ray tube G , yielding

$$\begin{aligned} 0 &= \int_G \nabla \cdot (\lambda \nabla T) dx = \int_{\partial G} \mathbf{n} \cdot \lambda \nabla T d\Gamma \\ &= \left(\frac{\partial T}{\partial \mathbf{n}} \lambda \right) \Big|_{\frac{\mathbf{x}}{|\mathbf{x}|}} ds_1 - s_0 \lambda(\mathbf{x}) ds_2, \end{aligned} \quad (12)$$

where ds_1 and ds_2 indicate the area of the top and bottom transversal sections of the ray tube, respectively. Applying the two boundary conditions $f \equiv 1$ and generic f in the above integration result, we end up with

$$s_0 \lambda^f(\mathbf{x}) ds_2 = f\left(\frac{\mathbf{x}}{|\mathbf{x}|}\right) ds_1, \quad (13)$$

$$s_0 \lambda^1(\mathbf{x}) ds_2 = 1 ds_1, \quad (14)$$

and these two equalities yield the desired claim in equation 11.

Modified gradient direction via normalized adjoint-state

Consequently, we propose to solve the adjoint-state equation twice with two different boundary conditions to normalize the adjoint-state variable so as to modify the gradient, yielding a well-behaved gradient descent direction. The new boundary condition for the adjoint-state equation 4 is

$$\mathbf{n}(\mathbf{x}) \cdot \nabla T(\mathbf{x}) \lambda(\mathbf{x})|_{\Gamma} = 1, \quad (15)$$

which yields the solution λ^R .

Denoting the solution of equations 4 and 5 by λ , we normalize the adjoint state λ by λ^R to get

$$\lambda_c(\mathbf{x}) = \lambda(\mathbf{x}) / \lambda^R(\mathbf{x}), \quad (16)$$

so that the gradient of the misfit functional is modified to be $\lambda_c(\mathbf{x}) / v^3(\mathbf{x})$.

Figure 3a shows the computed ray-illumination using equations 4 and 15 for the first shot. The strength of this ray-illumination gets stronger as it approaches the source, and this is quite similar to the ray distribution in Figure 2c. Therefore, we can compensate for the gradient using this ray-illumination. Equation 11 shows that the normalized adjoint state from one shot is nonsingular. Therefore, in this work, we perform the ray-illumination compensation shot by shot, and then stack all of the compensated gradients to form the final gradient for all the shots.

Regularization method with spatially varying regularization factors

We will use the normalized adjoint state to update the velocity model. However, when the aperture is incomplete, λ^R may be zero or very small, which may cause instability issues. Figure 3b shows a common depth curve of Figure 3a at a depth of 100 m, and it clearly shows that the illumination factor is zero or very small in a certain area. The regularization method is a common way of avoiding this instability issue. The conventional regularization method normally uses a constant regularization factor, yielding

$$\lambda_c(\mathbf{x}) = \frac{\lambda(\mathbf{x})}{\lambda^R(\mathbf{x}) + \alpha}, \quad (17)$$

where α is the regularization factor, which is a positive value related to the smallest ray-illumination factor among all the receiver points $\lambda_{\min}^R$. The regularization factor is an important parameter to balance between the compensation quality and stability.

Figure 4 shows compensation results using different regularization factors. After the ray-illumination compensation, the strength of the gradient is balanced along the raypath. The strong gradient near the source also disappears, and this helps remove the footprints near the source point. However, this regularization method is not

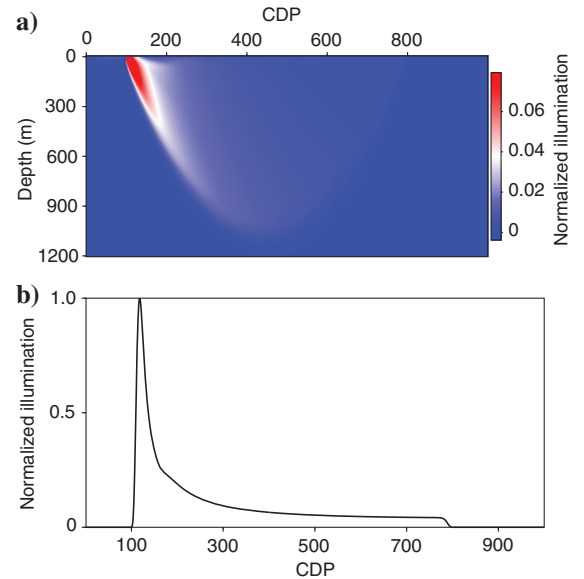


Figure 3. Ray illumination for a single source located at $x = 1000$ m. (a) Ray illumination and (b) its common depth curve at a depth of 100 m. Because the ray-illumination near the source is too strong, we clip the maximum value in (a) for display purposes.

optimal. When the regularization factor is large, the compensation is stable, but the compensation result is not optimal because the receiver side is undercompensated, as shown in Figure 4a and 4b. When the regularization factor is small, the compensation

quality is better in terms of amplitude balancing along the raypath, but it also amplifies small values leading to instability issues, as shown in Figure 4c–4f.

For the effective raypath, the adjoint-state equation 4 with the boundary condition in equation 15 shows that the value of the ray-illumination factor is big enough at each receiver point. Its value gradually increases as the observation point moves toward the source point. Therefore, it should have no instability issues. As the observation point gets away from the effective raypath, the corresponding value of the ray-illumination factor becomes smaller, and the instability issue arises. Therefore, we can apply spatially varying regularization factors according to the spatial distribution of the ray-illumination factor. We set the regularization factor to be small in areas with large illumination factors so that the regularization does not compromise the quality of compensation. We set the regularization factor to be large in areas with small illumination factors so that the instability issue is alleviated. This strategy can be implemented by replacing the constant regularization factor in equation 17 with a spatially varying regularization factor,

$$\alpha_r = \begin{cases} \alpha_{\min} & \text{if } \lambda^R(\mathbf{x}) \geq \lambda_{\max}, \\ \alpha_{\min} + \frac{(\lambda^R(\mathbf{x}) - \lambda_{\max})(\alpha_{\max} - \alpha_{\min})}{\lambda_{\min} - \lambda_{\max}} & \text{if } \lambda_{\max} > \lambda^R(\mathbf{x}) > \lambda_{\min}, \\ \alpha_{\max} & \text{if } \lambda^R(\mathbf{x}) \leq \lambda_{\min}, \end{cases} \quad (18)$$

Figure 4. Ray-illumination compensated gradients using the conventional regularization method. (a) Normalized gradient using $\alpha = \lambda_{\min}^B$ and (b) its common depth curve at a depth of 100 m; (c) normalized gradient using $\alpha = 0.1 * \lambda_{\min}^B$ and (d) its common depth curve at a depth of 100 m; and (e) normalized gradient using $\alpha = 0.01 * \lambda_{\min}^B$ and (f) its common depth curve at a depth of 100 m. The term $\lambda_{\min}^B$ is the smallest ray-illumination factor among all of the receiver points. Because there are instability issues in (c and e), we clip their maximum values for display purposes.

where $\alpha_{\min}$ and $\alpha_{\max}$ are the minimum and maximum values of the regularization factor, respectively, and $\lambda_{\min}$ and $\lambda_{\max}$ are the minimum and maximum values of the ray-illumination factor that we wish for compensation, respectively.

The term $\alpha_{\min}$ is a very small positive value for large ray-illumination factors, and it also can be set to zero. It is used to guarantee the quality of compensation in areas with an illumination factor larger than $\lambda_{\max}$. The term $\alpha_{\max}$ is a sufficiently large positive value for small ray-illumination factors. It is used to guarantee the stability in areas with an illumination factor smaller than $\lambda_{\min}$. The terms $\lambda_{\max}$ and $\lambda_{\min}$ should be set according to the smallest ray-illumination factor $\lambda_{\min}^B$ among all of the receiver points. The term $\lambda_{\max}$ should be equal to or a little bit smaller than $\lambda_{\min}^B$, and $\lambda_{\min}$ should be smaller than $\lambda_{\min}^B$. When the ray-illumination factor ranges from $\lambda_{\min}$ to $\lambda_{\max}$, its regularization factor is linearly interpolated to enable a smooth variation of regularization factors. Therefore, this spatially varying regularization factor can guarantee the stability of the ray-illumination compensation without compromising its quality.

To test the feasibility of this regularization method, the values of $\lambda_{\min}$ and $\lambda_{\max}$ are set to $0.01 * \lambda_{\min}^B$ and $\lambda_{\min}^B$, respectively; and the values of $\alpha_{\min}$ and $\alpha_{\max}$ are set to $0.01 * \lambda_{\min}^B$ and $\lambda_{\min}^B$, respectively. Figure 5 shows the ray-illumination compensation result. The original gradient in Figure 2b is unbalanced along the raypath due to the ray-illumination. However, the gradient along the raypath is well balanced, as shown in Figure 5. Therefore, the compensated gradient along the raypath is not smeared (or biased) due to the

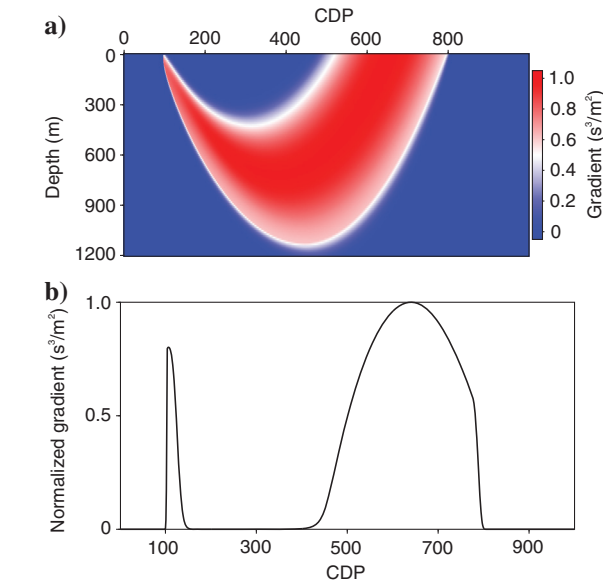


Figure 5. (a) Ray-illumination compensated gradient using the regularization method with spatially varying regularization factors and (b) its common depth curve at a depth of 100 m.

ray-illumination compensation. In addition, this compensated gradient does not suffer from instability issues.

Inversion strategy

The adjoint-state traveltime tomography can update the raypath (i.e., the traveltime field) and velocity field simultaneously. Updating the raypath is quite important if the initial velocity model is far away from the true solution. This makes the whole inversion process a typical nonlinear inverse problem. To stabilize the nonlinear inversion process, the inversion strategy is important.

In the early stage when the velocity mismatch is large, the raypath provided by the traveltime field is far away from the true raypath. To ensure convergence of the inversion process, we intentionally choose smaller step lengths to gradually update the raypath and velocity field simultaneously. This step helps to obtain a reasonable background velocity model.

After the early stage, the traveltime residuals become small, and the step length is not restricted. The velocity updating from far offsets usually contributes to the deep part, and that from near offsets usually contributes to the shallow part. However, because the raypath from far offsets overlays with that from near offsets, the velocity updating from near offsets is affected by that from far offsets. This may degrade the inverted velocity in the shallow part. To avoid this issue, we adopt an offset-dependent inversion strategy. The maximum offset used in inversion is gradually reduced during inversion. In this way, the later iteration will focus on updating the shallow area.

NUMERICAL EXAMPLES

In this section, the proposed method is tested on synthetic data sets to demonstrate its feasibility.

In the first example, the synthetic data set in Figure 1 is further tested. The previous section shows the gradient from one single shot. Figure 6 shows the functional gradient for all shots. Because the conventional adjoint-state equation does not incorporate the ray-illumination (Figure 2a), its cumulative gradient suffers from strong footprints near sources, as shown in Figure 6a. This is not an ideal gradient for updating the velocity function because it brings strong footprints into the inversion result. After the ray-illumination compensation, the strength of the gradient is balanced along the whole raypath. Therefore, its cumulative gradient does not suffer from footprints near sources, and the overall gradient is smooth (as shown in Figure 6b), which is useful for tomographic velocity updating. In addition, due to the ray-illumination compensation, the effective updating depth of gradient with ray-illumination compensation is deeper than that of the gradient without ray-illumination compensation. Figure 6c shows the common depth curve comparison between these two different gradients at a depth of 30 m. It shows that the gradient without ray-illumination compensation changes rapidly in the shallow area due to the strong footprints near sources, and the gradient with ray-illumination is much smoother. Figure 6d shows the CDP curve comparison at CDP500. It shows that the maximum values for gradients without and with ray-illumination compensation are located at depths of approximately 400 and 500 m, respectively. Because the center of this elliptical velocity perturbation is located at a depth of 500 m, the gradient with ray-illumination compensation performs better. The comparison between gradients shows that the proposed method can effectively

remove the imprint of ray-illumination and produce a more reliable gradient. Because the conventional gradient contains footprints near sources, it results in incorrect inverted models near sources if they are not processed. Therefore, to obtain inverted results from conventional gradients, we remove these large values near sources for the conventional gradient and apply a smooth filter to it. We also apply the same smoothing filter to the gradient with the ray-illumination compensation for comparison purposes. Figure 7 shows the inverted velocity perturbations from the conventional gradient and the proposed gradient. Because the ray-illumination compensation balances the gradient in the deeper area, the inverted velocity perturbation with ray-illumination compensation is better in the deeper

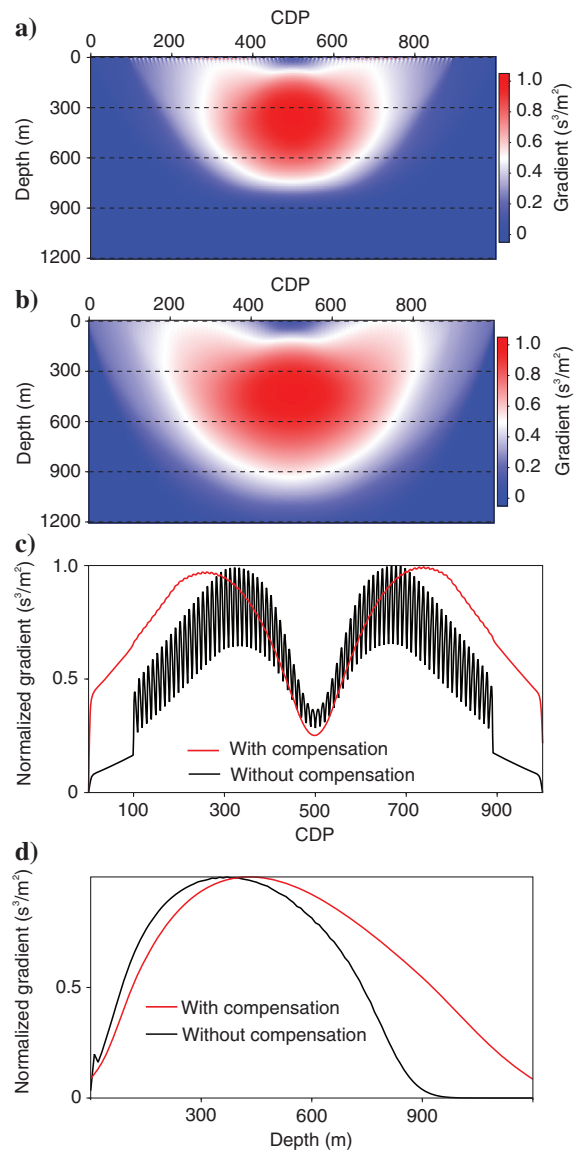


Figure 6. Gradient for all sources. (a) The normalized gradient without ray-illumination compensation, (b) the normalized gradient with ray-illumination compensation, (c) the common depth curve comparison at a depth of 30 m, and (d) the CDP curve comparison at CDP500. The black curve is the gradient without ray-illumination compensation, and the red curve is the gradient with ray-illumination compensation.

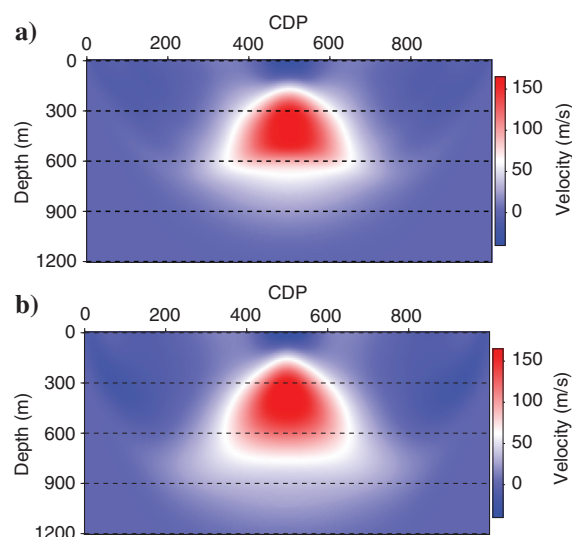


Figure 7. Inverted velocity perturbation comparison. (a) The inverted velocity perturbation without ray-illumination compensation and (b) the inverted velocity perturbation with ray-illumination compensation.

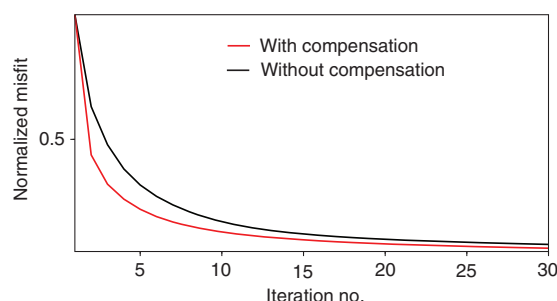


Figure 8. Misfit curve. The black and red curves are misfit curves without and with ray-illumination compensation, respectively.

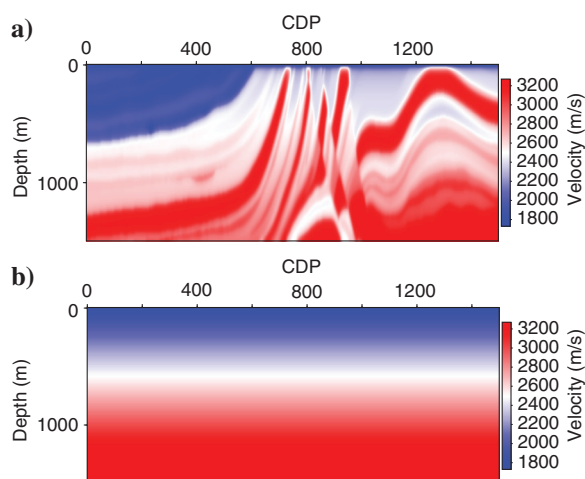


Figure 9. Modified Marmousi2 velocity model extracted from the Marmousi2 model. (a) The exact velocity model and (b) the initial velocity model taken to be a linear function in the spatial coordinates.

area. Moreover, the ray-illumination compensation also enhances convergence, as shown in Figure 8.

In the second example, a synthetic data set from the modified Marmousi2 velocity model (Figure 9a) is tested. This velocity model is extracted from the original Marmousi2 model (Martin et al., 2006). The model dimension is 1.5 km in depth and 15 km laterally. The grid spacing in the x - and z -axes is 10 m. The source and receivers are located on the surface. The maximum offset for each shot is 6 km, and the receivers are located on both sides of the source. The receiver interval is 10 m. We acquire 150 shots using the exact velocity model (Figure 9a), and the shot interval is 100 m. The lateral position of the first shot is located at $x = 10$ m (i.e., CDP = 1).

After simulating the first-arrival traveltimes for each shot, we use a velocity model defined to be a linear function in the spatial coordinates (Figure 9b) as the initial model. Figure 10 shows the functional gradients obtained by the conventional adjoint-state method and the proposed method. The regularization parameters are the same as the first example. The conventional functional gradient is affected by the imprint of the ray-illumination. There is a strong footprint in the shallow region, where the velocity needs updating, but the footprint causes a velocity distortion in the shallow region. The proposed method compensates for the ray-illumination, and there is no footprint near the source. The functional gradient better reveals the trend of velocity updating.

Because the footprints in the conventional gradient lead to incorrect velocity models in the shallow area, we remove these big values near sources and apply a smoothing filter to obtain the inverted results for the conventional method. Figure 11a shows the result from the conventional method. Figure 11b shows the result using ray-illumination compensation. Because the raypaths from large-offset data cover deep and shallow areas, they interfere with the velocity updating contributed by the near-offset data. Then, we use the offset-dependent inversion strategy to improve the result by using the velocity model in Figure 11b as an initial model. During the inversion, the maximum offset is gradually reduced from 6 to 2 km with a 2 km interval. Figure 11c shows the result from the offset-dependent inversion strategy. These three tomographic results successfully

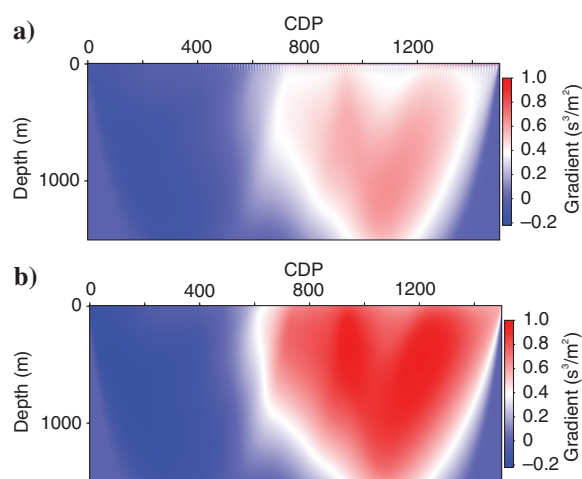


Figure 10. Gradient for all sources for the modified Marmousi2 velocity model. (a) The normalized gradient without ray-illumination compensation and (b) the normalized gradient with ray-illumination compensation.

recover the main velocity structure. To compare these results in detail, Figure 12 shows the inverted velocity perturbations, and Figure 13 shows velocity logs laterally and vertically. Compared with the true velocity perturbation in Figure 12a, these three results recover reliable background velocity models. Comparing Figure 12b with Figure 12c, the conventional method and the ray-illumination compensation method perform similarly. However, the velocity log comparison in Figure 13 shows that the ray-illumination compensation method obtains a better result in the shallow area (Figure 13a), and it also gets more velocity updates in the deeper area due to the balanced gradient in the deeper area (Figure 13b). Moreover, Figure 14 also shows that the ray-illumination compensation method converges faster than the conventional method. Comparing Figure 12d with the true velocity perturbation, the offset-dependent inversion strategy enhances details of the velocity structure in the shallow area, which also can be observed in Figure 13.

Figure 15 shows migration results using different velocity models. Because the initial velocity model deviates far from the true velocity model, its migration result in Figure 15b is of low quality, especially in the two circled areas. However, because the reflection signals for these shallow reflectors are recorded by near-offset data with quite limited offset range, the migration result in Figure 15b shows mostly the mispositioned reflectors, and it does not exhibit obvious focusing or defocusing effect. Because the tomographic velocity model improves the accuracy of velocity, the migration result from the inverted velocity model without ray-illumination compensation in Figure 15c greatly corrects the positions of these reflectors. Compared with Figure 15c, the migration result from the inverted velocity model with ray-illumination compensation

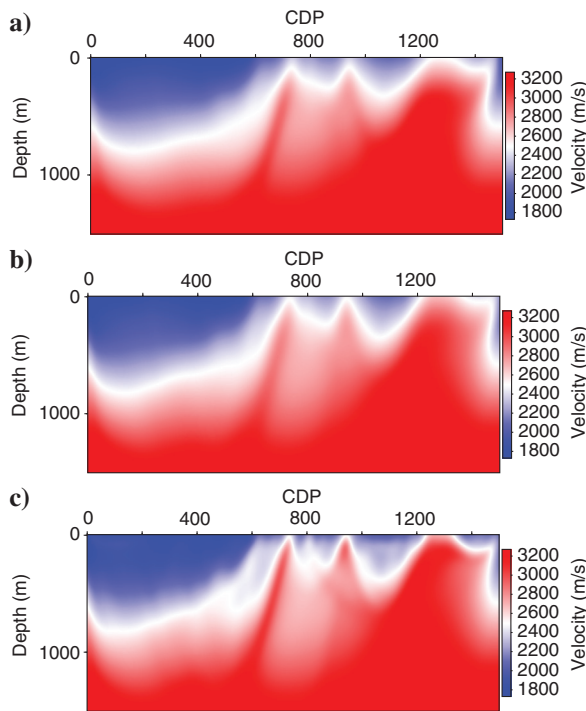


Figure 11. Tomographic result comparison for the modified Marmousi2 velocity model. (a) The inverted velocity model without ray-illumination compensation, (b) the inverted velocity model with ray-illumination compensation, and (c) the inverted velocity model using the offset-dependent inversion strategy.

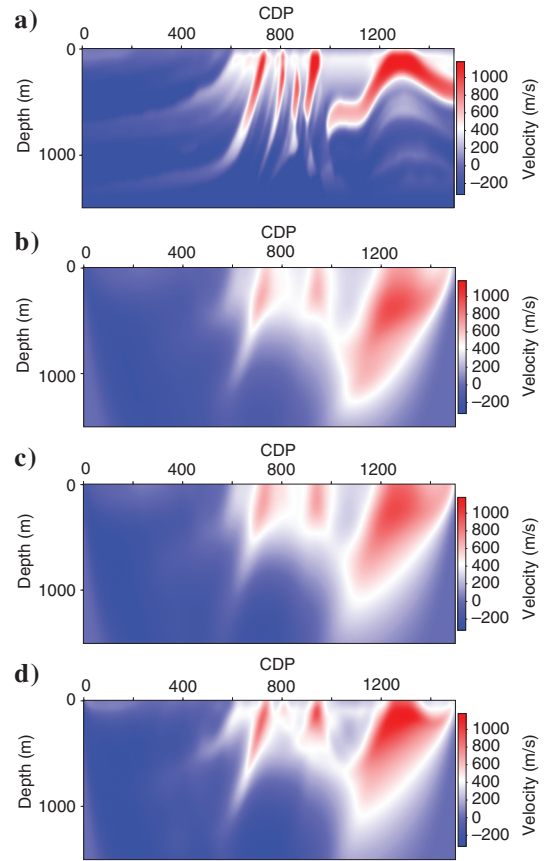


Figure 12. Velocity perturbation comparison. (a) The true velocity perturbation, (b) the inverted velocity perturbation without ray-illumination compensation, (c) the inverted velocity perturbation with ray-illumination compensation, and (d) the inverted velocity perturbation using the offset-dependent inversion strategy.

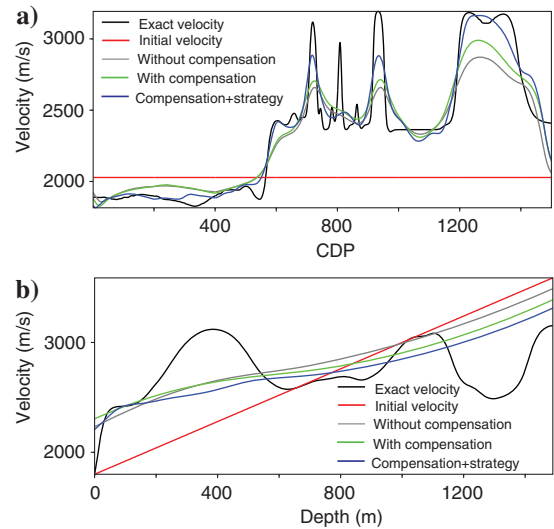


Figure 13. Velocity log comparison: (a) 200 m depth and (b) CDP800. Velocity logs from the exact velocity model, the initial velocity model, the inverted velocity model without ray-illumination compensation, the inverted velocity model with ray-illumination compensation, and the inverted velocity model using the offset-dependent inversion strategy are indicated by the black, red, gray, green, and blue curves, respectively.

in Figure 15d further improves the positions of these reflectors, especially in the marked lower left and upper right regions. Compared with Figure 15c, the migration result using the velocity model with offset-dependent inversion strategy further corrects the positions of these reflectors. The reflector positions in the left and upper right regions in Figure 15e are comparable with those in the migration result using the true velocity model shown in Figure 15a. Figure 15f compares traces from different migration results at CDP1285. As can be seen, there are large errors in depth in the migration result from the initial velocity model. The migration result from the inverted velocity model without ray-illumination compensation greatly reduces errors in depth, but there are

still some residual errors. The result from the inverted velocity model with ray-illumination compensation further reduces the errors in depth. Finally, the migration result from the inverted velocity model with offset-dependent inversion strategy significantly reduces the errors in depth.

CONCLUSION

We propose a method to eliminate the imprint of ray-illumination in the adjoint-state first-arrival traveltome tomography by analyzing the mechanism of the imprint of ray-illumination. The proposed adjoint-state equation with a modified boundary condition is effective in obtaining the ray-illumination. A regularization method with spatially varying regularization factors stabilizes the ray-illumination compensation. Numerical experiments on two synthetic examples show that the proposed method can effectively remove footprints caused by the imprint of ray-illumination. The inversion strategy succeeds in recovering the velocity structure and minimizing the traveltome residuals. The proposed method also can be used in other seismic acquisition geometries, such as vertical seismic profiling and cross-hole geometry, and it can be extended to adjoint-state reflection traveltome tomography.

ACKNOWLEDGMENTS

Authors Hu, Cao, and Wang appreciate the financial support from the National Natural Science Foundation of China (grant nos. 42074166, 42030812, 42074163, 41974160, and U1562219). Qian is partially supported by the National Science Foundation (grant nos. 1614566 and 2012046). Leung is partially supported by the Hong Kong RGC (grant no. 16302819). We thank J. Shragge, J. Blanch, F. Q. Liu, M. Noble, and two anonymous reviewers for helpful comments and suggestions.

DATA AND MATERIALS AVAILABILITY

Data associated with this research are available and can be obtained by contacting the corresponding author.

REFERENCES

- Benaichouche, A., M. Noble, and A. Gesret, 2015, First arrival traveltome tomography using the fast marching method and the adjoint state technique: 77th Annual International Conference and Exhibition, EAGE, Extended Abstracts, Tu N102 09, doi: [10.3997/2214-4609.201412568](https://doi.org/10.3997/2214-4609.201412568).
- Berryman, J. G., 1989, Weighted least-squares criteria for seismic traveltome tomography: IEEE Transactions on Geoscience and Remote Sensing, **27**, 302–309, doi: [10.1109/36.17671](https://doi.org/10.1109/36.17671).
- Bregman, N., R. Langan, and J. Resnick, 1989, Cross-hole seismic tomography: Geophysics, **54**, 200–215, doi: [10.1190/1.1442644](https://doi.org/10.1190/1.1442644).
- Červený, V., 2001, Seismic ray theory: Cambridge University Press.
- Dines, K. A., and R. J. Lytle, 1979, Computerized geophysical tomography: Proceedings of IEEE, **67**, 1065–1073, doi: [10.1109/PROC.1979.11390](https://doi.org/10.1109/PROC.1979.11390).
- Gray, S., and W. May, 1994, Kirchhoff migration using eikonal equation traveltomes: Geophysics, **59**, 810–817, doi: [10.1190/1.1443639](https://doi.org/10.1190/1.1443639).

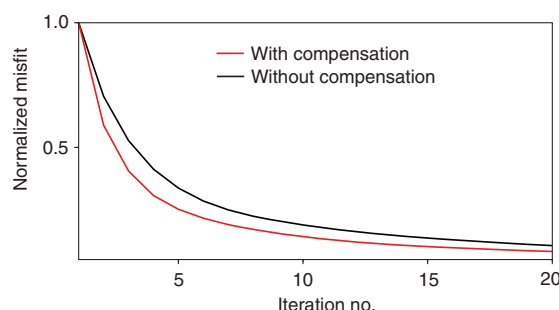


Figure 14. Misfit curve for the modified Marmousi2 velocity model. The black and red curves are misfit curves without and with ray-illumination compensation, respectively.

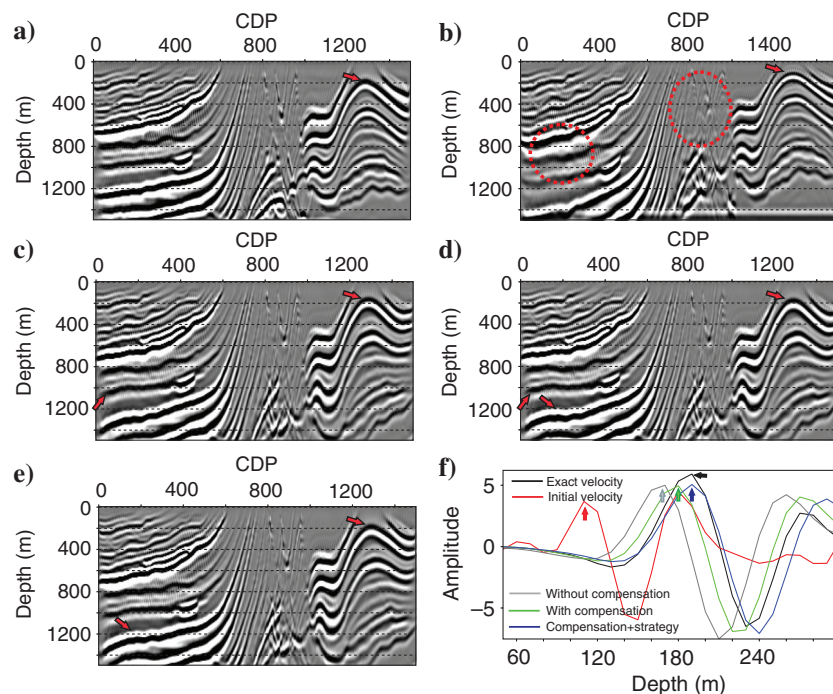


Figure 15. Migration results using different velocity models. (a) The true velocity model, (b) the initial velocity model, (c) the inverted velocity model without ray-illumination compensation, (d) the inverted velocity model with ray-illumination compensation, and (e) the inverted velocity model using the offset-dependent inversion strategy. (f) The comparison of traces between different migration results at CDP1285. Migration results from (a-e) are indicated by the black, red, gray, green, and blue curves, respectively.

- Hu, J. T., J. X. Cao, H. Z. Wang, S. Y. Liu, and X. J. Wang, 2018, 3D traveltimes computation for quasi-P wave in orthorhombic media using dynamic programming: *Geophysics*, **83**, no. 1, C27–C35, doi: [10.1190/geo2016-0558.1](https://doi.org/10.1190/geo2016-0558.1).
- Hu, J. T., J. X. Cao, H. Z. Wang, X. J. Wang, and X. D. Jiang, 2017, First-arrival traveltimes computation in transversely isotropic media using Fermat's principle-based fast marching: *Journal of Geophysics and Engineering*, **14**, 1484–1491, doi: [10.1088/1742-2140/aa8159](https://doi.org/10.1088/1742-2140/aa8159).
- Hu, W., and C. Marcinkovich, 2012, A sensitivity controllable target-oriented tomography algorithm: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: [10.1190/segam2012-1336.1](https://doi.org/10.1190/segam2012-1336.1).
- Kao, C. Y., S. J. Osher, and J. Qian, 2004, Lax-Friedrichs sweeping schemes for static Hamilton-Jacobi equations: *Journal of Computational Physics*, **196**, 367–391, doi: [10.1016/j.jcp.2003.11.007](https://doi.org/10.1016/j.jcp.2003.11.007).
- Lambaré, G., P. S. Lucio, and A. Hanyga, 1996, Two-dimensional multivalued traveltimes and amplitude maps by uniform sampling of a ray field: *Geophysical Journal International*, **125**, 584–598, doi: [10.1111/j.1365-246X.1996.tb00021.x](https://doi.org/10.1111/j.1365-246X.1996.tb00021.x).
- Leung, S., and J. Qian, 2006, An adjoint state method for three-dimensional transmission traveltimes tomography using first-arrivals: *Communications in Mathematical Sciences*, **4**, 249–266, doi: [10.4310/CMS.2006.v4.n1.a10](https://doi.org/10.4310/CMS.2006.v4.n1.a10).
- Li, W., and S. Leung, 2013, A fast local level set based adjoint state method for first arrival transmission traveltimes tomography with discontinuous slowness: *Geophysical Journal International*, **195**, 582–596, doi: [10.1093/gji/ggt244](https://doi.org/10.1093/gji/ggt244).
- Li, W., S. Leung, and J. Qian, 2014, A level-set adjoint-state method for crosswell transmission-reflection traveltimes tomography: *Geophysical Journal International*, **199**, 348–367, doi: [10.1093/gji/ggu262](https://doi.org/10.1093/gji/ggu262).
- Martin, G., R. Wiley, and K. Marfurt, 2006, Marmousi2: An elastic upgrade for Marmousi: *The Leading Edge*, **25**, 156–166, doi: [10.1190/1.2172306](https://doi.org/10.1190/1.2172306).
- Noble, M., P. Thierry, C. Taillandier, and H. Calandra, 2010, High-performance 3D first-arrival traveltimes tomography: *The Leading Edge*, **29**, 86–93, doi: [10.1190/1.3284057](https://doi.org/10.1190/1.3284057).
- Nolet, G., 1993, Solving large linearized tomographic problems, in H. M. Iyer and K. Hirahara, eds., *Seismic tomography: Theory and practice*: Chapman and Hall, 227–247.
- Nolet, G., 2012, *Seismic tomography: With applications in global seismology and exploration geophysics*: Springer Science & Business Media.
- Peterson, J. E., B. Paulsson, and T. V. McEvilly, 1985, Applications of algebraic reconstruction techniques to crosshole seismic data: *Geophysics*, **50**, 1566–1580, doi: [10.1190/1.1441847](https://doi.org/10.1190/1.1441847).
- Podvin, P., and I. Lecomte, 1991, Finite difference computation of traveltimes in very contrasted velocity models: A massively parallel approach and its associated tools: *Geophysical Journal International*, **105**, 271–284, doi: [10.1111/j.1365-246X.1991.tb03461.x](https://doi.org/10.1111/j.1365-246X.1991.tb03461.x).
- Qian, J., and W. W. Symes, 2002, An adaptive finite difference method for traveltimes and amplitude: *Geophysics*, **67**, 167–176, doi: [10.1190/1.1451472](https://doi.org/10.1190/1.1451472).
- Qian, J., Y. T. Zhang, and H. K. Zhao, 2007a, Fast sweeping methods for eikonal equations on triangulated meshes: *SIAM Journal on Numerical Analysis*, **45**, 83–107, doi: [10.1137/050627083](https://doi.org/10.1137/050627083).
- Qian, J., Y. T. Zhang, and H. K. Zhao, 2007b, Fast sweeping methods for static convex Hamilton-Jacobi equations: *Journal of Scientific Computing*, **31**, 237–271, doi: [10.1007/s10915-006-9124-6](https://doi.org/10.1007/s10915-006-9124-6).
- Qin, F., Y. Luo, K. B. Olsen, W. Cai, and G. T. Schuster, 1992, Finite difference solution of the eikonal equation along expanding wavefronts: *Geophysics*, **57**, 478–487, doi: [10.1190/1.1443263](https://doi.org/10.1190/1.1443263).
- Schneider, W. A. J., K. Ranzinger, A. Balch, and C. Kruse, 1992, A dynamic programming approach to first arrival traveltimes computation in media with arbitrarily distributed velocities: *Geophysics*, **57**, 39–50, doi: [10.1190/1.1443187](https://doi.org/10.1190/1.1443187).
- Sei, A., and W. W. Symes, 1994, Gradient calculation of the traveltimes cost function without ray tracing: 64th Annual International Meeting, SEG, Expanded Abstracts, 1351–1354, doi: [10.1190/1.1822780](https://doi.org/10.1190/1.1822780).
- Sei, A., and W. W. Symes, 1995, Convergent finite-difference traveltimes gradient for tomography: 65th Annual International Meeting, SEG, Expanded Abstracts, 1258–1261, doi: [10.1190/1.1887581](https://doi.org/10.1190/1.1887581).
- Sethian, J. A., and A. M. Popovici, 1999, 3-D traveltimes computation using the fast marching method: *Geophysics*, **64**, 516–523, doi: [10.1190/1.1444558](https://doi.org/10.1190/1.1444558).
- Snieder, R., and M. Sambridge, 1992, Ray perturbation theory for traveltimes and ray paths in 3-D heterogeneous media: *Geophysical Journal International*, **109**, 294–322, doi: [10.1111/j.1365-246X.1992.tb00099.x](https://doi.org/10.1111/j.1365-246X.1992.tb00099.x).
- Taillandier, C., M. Noble, H. Chauris, and H. Calandra, 2009, First-arrival traveltimes tomography based on the adjoint-state method: *Geophysics*, **74**, no. 6, WCB1–WCB10, doi: [10.1190/1.3250266](https://doi.org/10.1190/1.3250266).
- van Trier, J., and W. W. Symes, 1991, Upwind finite-difference calculation of traveltimes: *Geophysics*, **56**, 812–821, doi: [10.1190/1.1443099](https://doi.org/10.1190/1.1443099).
- Vidale, J., 1988, Finite-difference calculation of travel times: *Bulletin of the Seismological Society of America*, **78**, 2062–2076.
- Vigh, D., E. Starr, and J. Kapoor, 2009, Developing earth model with full waveform inversion: *The Leading Edge*, **28**, 432–435, doi: [10.1190/1.3112760](https://doi.org/10.1190/1.3112760).
- Vinje, V., E. Iversen, and H. Gjoystdal, 1993, Traveltimes and amplitude estimation using wavefront construction: *Geophysics*, **58**, 1157–1166, doi: [10.1190/1.1443499](https://doi.org/10.1190/1.1443499).
- Waheed, U. B., G. Flagg, and C. E. Yarman, 2016, First-arrival traveltimes tomography for anisotropic media using the adjoint-state method: *Geophysics*, **81**, no. 4, R147–R155, doi: [10.1190/geo2015-0463.1](https://doi.org/10.1190/geo2015-0463.1).
- Wang, B., and L. W. Braille, 1995, Effective approaches to handling non-uniform data coverage problem for wide-aperture refraction/reflection profiling: 65th Annual International Meeting, SEG, Expanded Abstracts, 659–662, doi: [10.1190/1.1887398](https://doi.org/10.1190/1.1887398).
- Wang, H. Z., Z. T. Ma, and P. Hubral, 1999, Traveltimes calculation in 3D media with arbitrary velocity distribution: 69th Annual International Meeting, SEG, Expanded Abstracts, 1263–1266, doi: [10.1190/1.1820738](https://doi.org/10.1190/1.1820738).
- Zelt, C. A., and P. J. Barton, 1998, Three-dimensional seismic refraction tomography: A comparison of two methods applied to data from the Fae-ro Basin: *Journal of Geophysical Research*, **103**, 7187–7210, doi: [10.1029/97JB03536](https://doi.org/10.1029/97JB03536).
- Zhao, H. K., 2005, Fast sweeping method for eikonal equations: *Mathematics of Computation*, **74**, 603–628, doi: [10.1090/S0025-5718-04-01678-3](https://doi.org/10.1090/S0025-5718-04-01678-3).
- Zhu, X., D. Sixta, and B. Angstman, 1992, Tomostatics: Turning-ray tomography + static corrections: *The Leading Edge*, **11**, 15–23, doi: [10.1190/1.1436864](https://doi.org/10.1190/1.1436864).

Biographies and photographs of the authors are not available.