GEOPHYSICS, VOL. 78, NO. 6 (NOVEMBER-DECEMBER 2013); P. U89–U101, 19 FIGS.
10.1190/GEO2013-0058.1

First-break traveltime tomography with the double-square-root eikonal equation

Siwei Li[1], Alexander Vladimirsky[2], and Sergey Fomel[1]

ABSTRACT

First-break traveltime tomography is based on the eikonal equation. Because the eikonal equation is solved at fixed-shot positions and only receiver positions can move along the ray-path, the adjoint-state tomography relies on inversion to resolve possible contradicting information between independent shots. The double-square-root (DSR) eikonal equation allows not only the receivers but also the shots to change position, and thus describes the prestack survey as a whole. Consequently, its linearized tomographic operator naturally handles all shots together, in contrast with the shotwise approach in the traditional eikonal-based framework. The DSR eikonal equation is singular for the horizontal waves, which require special handling. Although it is possible to recover all branches of the solution through postprocessing, our current forward modeling and tomography focuses on the diving wave branch only. We consider two upwind discretizations of the DSR eikonal equation and show that the explicit scheme is only conditionally convergent and relies on nonphysical stability conditions. We then prove that an implicit upwind discretization is unconditionally convergent and monotonically causal. The latter property makes it possible to introduce a modified fast matching method thus obtaining first-break traveltimes efficiently and accurately. To compare the new DSR eikonal-based tomography and traditional eikonal-based tomography, we perform linearizations and apply the same adjoint-state formulation and upwind finite-differences implementation to both approaches. Synthetic model examples justify that the proposed approach converges faster and is more robust than the traditional one.

INTRODUCTION

The first-break traveltime tomography (Zhu et al., 1992; Osypov, 2000; Leung and Qian, 2006; Taillandier et al., 2009; Noble et al., 2010) is an established tool for estimating near-surface macrofeature seismic velocities. Starting from a prior model, tomographic inversion gradually modifies the velocities such that the misfits between predicted and observed first breaks decrease. Because the problem is nonlinear, several linearization iterations may be required until convergence. Moreover, inversion must be carried out with careful choice of regularization to avoid local minima (Stefani, 1993; Simmons and Bernitsas, 1994; Engl et al., 1996). The estimated model has a direct influence on subsequent applications; for example, static corrections (Marsden, 1993; Cox, 1999; Bergman et al., 2004) where it provides a medium-to-long wavelength near-surface model, and waveform tomography (Sheng et al., 2006; Brenders and Pratt, 2007; Virieux and Operto, 2009), where it serves as a low-frequency prior.

The traditional first-break traveltime tomography is based on the eikonal equation that arises from high-frequency approximation of the wave equation (Chapman, 2002). During forward modeling, the first breaks computed through the eikonal equation are naturally shot-indexed because only receiver coordinates move while the source is fixed. At the tomography stage, one may formulate the minimization of cost function as a sequence of explicitly linearized problems or directly as a nonlinear optimization problem. The first choice (Zelt and Barton, 1998; Zhu et al., 2000; Dessa et al., 2004; Pei, 2009) requires computation of Fréchet derivatives, which is usually carried out by combining an eikonal solver with posterior ray tracing. Then, an algorithm such as LSQR (Paige and Saunders,

Manuscript received by the Editor 14 February 2013; revised manuscript received 1 June 2013; published online 9 October 2013.
[1]The University of Texas at Austin, Bureau of Economic Geology, John A. and Katherine G. Jackson School of Geosciences, Austin, Texas, USA. E-mail: siwei.li@utexas.edu; sergey.fomel@beg.utexas.edu.
[2]Cornell University, Department of Mathematics, Ithaca, New York, USA. E-mail: vlad@math.cornell.edu.

1982) is applied to solve the linearized tomographic system iteratively. Although this approach accounts for information from source and receiver dimensions, it faces computational limitations when the Fréchet derivative matrix becomes difficult to handle because of a large number of model parameters. The nonlinear optimization approach, on the other hand, can be combined with the adjoint-state method (Plessix, 2006) and avoids an explicit computation of Fréchet derivatives (Taillandier et al., 2009). The cost of computing the gradient is equivalent to twice the solution of the forward modeling problem, regardless of the size of input data. However, one major drawback of this approach, as we will show later, is that the resulting gradient disregards information available along the shot dimension.

The drawback of eikonal-based adjoint-state tomographies is that they always face conflicting information that propagates across different shots. Such conflicts must be resolved during inversion, or else an erroneous model update may appear. In practice, the inversion may be less robust and may take more iterations to converge, compared to the situation where we replace the eikonal equation with another governing equation that allows both source and receiver positions to change along raypaths. The double-square-root (DSR) eikonal equation is a promising candidate in this regard because it describes the prestack data as a whole by linking the evolution of traveltimes to subsurface source and receiver positions. In this paper, we investigate the feasibility of using the DSR eikonal equation for first-break traveltime tomography with the adjoint-state method.

DSR eikonal was analyzed previously by Belonosova and Alekseev (1974), Duchkov and de Hoop (2010), and Alkhalifah (2011). Ray-tracing methods applied to DSR are capable of providing multiarrivals by extrapolating isochron rays (Iversen, 2004) or using perturbation theory, but their extra costs in computing non-first-breaks are not necessary for first-break tomography purpose. We first prove that an implicit discretization of the DSR eikonal equation is causal and thus can be solved by a Dijkstra-like noniterative method (Dijkstra, 1959). The DSR singularity and two DSR branches that are noncausal need special treatment. Our current implementation employs a modified fast-marching (Sethian, 1999) DSR eikonal solver. We first test its accuracy by DSR forward modeling. Next, we linearize the DSR eikonal equation and use the resulting operators in adjoint-state tomography. For comparison, we apply an analogous linearization and adjoint-state formulation to the traditional tomography based on shot-indexed eikonal equation. Then, we demonstrate the differences between the proposed and traditional approaches and justify advantages of the new method using several synthetic model examples. We conclude by discussing possible further improvements and extensions of our method.

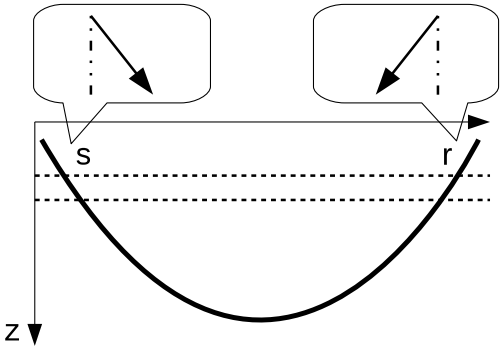

Figure 1. A diving ray and zoom-in of the ray segments between two depth levels.

THEORY

DSR eikonal equation

The DSR eikonal equation can be derived by considering a raypath and its segments between two depth levels. Figure 1 illustrates a diving ray (Zhu et al., 1992) in 2D with velocity $v = v(z, x)$. We denote $T(z, r, s)$ as the total traveltime of the raypath beneath depth z, where r and s are *subsurface* receiver and source lateral positions, respectively.

At source and receiver sides the traveltime satisfies the eikonal equation, therefore

$$\frac{\partial T}{\partial z} = -\sqrt{\frac{1}{v^2(z,s)} - \left(\frac{\partial T}{\partial s}\right)^2} - \sqrt{\frac{1}{v^2(z,r)} - \left(\frac{\partial T}{\partial r}\right)^2}. \quad (1)$$

The negative signs before the two square-roots in equation 1 correspond to a decrease of traveltime with increasing depth, or geometrically a *downward* pointing of slowness vectors on s and r sides. Because the slowness vectors could also be pointing *upward* and the directions may be different at r and s, the DSR eikonal equation (Belonosova and Alekseev, 1974) should account for all the possibilities (Figure 2)

$$\frac{\partial T}{\partial z} = \pm\sqrt{\frac{1}{v^2(z,s)} - \left(\frac{\partial T}{\partial s}\right)^2} \pm \sqrt{\frac{1}{v^2(z,r)} - \left(\frac{\partial T}{\partial r}\right)^2}. \quad (2)$$

The boundary condition for DSR eikonal equation is that traveltimes at the subsurface zero-offset plane, i.e., $r = s$, are zero: $T(z, r = s) = 0$.

Equation 2 has a *singularity* when $\partial T/\partial z = 0$, in which case the slowness vectors at s and r sides are horizontal and equation 2 reduces to

$$\left(\frac{\partial T}{\partial s}\right)^2 = \frac{1}{v^2(z,s)}; \quad \left(\frac{\partial T}{\partial r}\right)^2 = \frac{1}{v^2(z,r)}. \quad (3)$$

The two independent equations in 3 are not in conflict according to the source-receiver reciprocity because they are the same with an exchange of s and r.

Note that equations 2 and 3 describe T in full prestack domain (z, r, s) by allowing not only receivers but also sources to change positions. In contrast, the eikonal equation

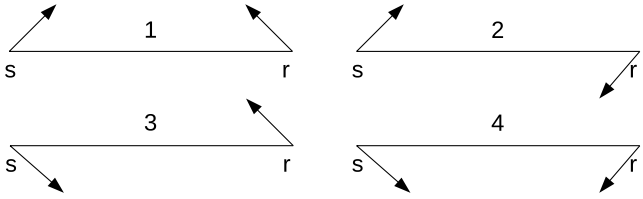

Figure 2. All four branches of DSR eikonal equation from different combination of upward or downward pointing of slowness vectors. Whether the slowness vector is pointing leftward or rightward does not matter, because the partial derivatives with respect to s and r in equation 2 are squared. Figure 1 and equation 1 belong to the last situation.

$$\left(\frac{\partial T}{\partial z}\right)^2+\left(\frac{\partial T}{\partial x}\right)^2=\frac{1}{v^2(z,x)} \tag{4}$$

with boundary condition $T(z=0,x=s)=0$ can be used only for one fixed source position at a time and thus traveltimes of different shots are independent of each other. In equation 4, s is *surface* source lateral position. In the 3D case, the scalars s, r, and x in equations 2, 3, and 4 become 2D vectors **s**, **r**, and **x** that contain inline and crossline positions. The prestack traveltime is then in a 5D space. Our current work is restricted to 2D and we consider the 3D extension in the "Discussion" section.

Similarly to the eikonal equation, the DSR eikonal equation is a nonlinear first-order partial differential equation. Its solutions include, in general, not only first breaks but all arrivals, and can be computed by solving separate eikonal equations for each subsurface source-receiver pair followed by extracting the traveltime and putting the value into prestack volume. However, such an implementation is impractical due to the large amount of computations. Meanwhile, for first-break tomography purposes, we are only interested in the first-arrival solutions but require an efficient and accurate algorithm. In this regard, a finite-difference DSR eikonal solver analogous to the fast-marching (Sethian, 1999) or fast-sweeping (Zhao, 2005) eikonal solvers is preferable.

In upwind discretizations of the DSR eikonal equation on the grid in (z,r,s) domain, one has to make a decision about the z-slice, in which the finite differences are taken to approximate $\partial T/\partial s$ and $\partial T/\partial r$. In Figure 1, it appears natural to approximate these partial derivatives in the z-slice below $T(z,r,s)$. We refer to the corresponding scheme as *explicit*, because it allows to directly compute the grid value $T(z,r,s)$ based on the already known T values from the next-lower z. An alternative *implicit* scheme is obtained by approximating $\partial T/\partial s$ and $\partial T/\partial r$ in the same z-slice as $T(z,r,s)$, which results in a coupled system of nonlinear discretized equations. In Appendix A, we prove the following:

- The explicit scheme is very efficient to use on a fixed grid, but only conditionally convergent. This property is also confirmed numerically in "Synthetic model examples" section.
- The implicit scheme is *monotone causal*, meaning $T(z,r,s)$ depends on the smaller neighboring grid values only. This enables us to apply a Dijkstra-like method (Dijkstra, 1959) to solve the discretized system efficiently. Importantly, the DSR singularity requires a special ordering in the selection of upwind neighbors, which switches between equations 1 and 3 when necessary. We provide a modified fast-marching (Sethian, 1999) DSR eikonal solver along with such an ordering strategy in the "Numerical implementation" section.
- The causality analysis in Appendix A applies only to the first and last *causal branches* out of all four shown in Figure 2. Additional postprocessings, albeit expensive, can be used to recover the rest two noncausal branches as they may be decomposed into summations of the causal ones.

In practice, we find that, for moderate velocity variations, the first breaks correspond only to causal branches. An example in the "Synthetic model examples" section serves to illustrate this observation. Therefore, for efficiency, we turn off the noncausal branch postprocessings in forward modeling and base the tomography solely on equations 1 and 3.

DSR tomography

The first-break traveltime tomography with DSR eikonal equation (DSR tomography) can be established by following a procedure analogous to the traditional one with the shot-indexed eikonal equation (standard tomography). To further reveal their differences, in this section we will derive both approaches.

For convenience, we use slowness-squared $w\equiv 1/v^2$ instead of velocity v in equations 1, 3, and 4. Based on analysis in Appendix A, the velocity model $w(z,x)$ and prestack cube $T(z,r,s)$ are Eulerian discretized and arranged as column vectors **w** of size $nz\times nx$ and **t** of size $nz\times nx\times nx$. We denote the observed first breaks by $\mathbf{t}^{\text{obs}}$, and use $\mathbf{t}^{\text{std}}$ and $\mathbf{t}^{\text{dsr}}$ whenever necessary to discriminate between **t** computed from shot-indexed eikonal equation and DSR eikonal equation.

The tomographic inversion seeks to minimize the l_2 (least-squares) norm of the data residuals. We define an objective function as follows

$$E(\mathbf{w})=\frac{1}{2}(\mathbf{t}-\mathbf{t}^{\text{obs}})^T(\mathbf{t}-\mathbf{t}^{\text{obs}}), \tag{5}$$

where the superscript T represents transpose. A Newton method of inversion can be established by considering an expansion of the misfit function 5 in a Taylor series and retaining terms up to the quadratic order (Bertsekas, 1982)

$$E(\mathbf{w}+\delta\mathbf{w})=E(\mathbf{w})+\delta\mathbf{w}^T\nabla_w E(\mathbf{w})+\frac{1}{2}\delta\mathbf{w}^T\mathbf{H}(\mathbf{w})\delta\mathbf{w}+O(|\delta\mathbf{w}|^3). \tag{6}$$

Here, $\nabla_w E$ and **H** are gradient vector and Hessian matrix, respectively. We may evaluate the gradient by taking partial derivatives of equation 5 with respect to **w**, yielding

$$\nabla_w E\equiv\frac{\partial E}{\partial\mathbf{w}}=\mathbf{J}^T(\mathbf{t}-\mathbf{t}^{\text{obs}}), \tag{7}$$

where **J** is the Frechét derivative matrix and can be found by further differentiating **t** with respect to **w**.

We start by deriving the Frechét derivative matrix of standard tomography. Denoting

$$D_m\equiv\frac{\partial}{\partial m};\quad m=z,x,r,s \tag{8}$$

as the partial derivative operator in the m direction, equation 4 can be rewritten as

$$D_z t_k^{\text{std}}\cdot D_z t_k^{\text{std}}+D_x t_k^{\text{std}}\cdot D_x t_k^{\text{std}}=w;\quad k=1,2,3,\ldots,nx. \tag{9}$$

Here, we assume that there are, in total, nx shots and use t_k^{std} for first breaks of the kth shot. Applying $\partial/\partial w$ to both sides of equation 9, we find

$$J_k^{\text{std}}\equiv\frac{\partial t_k^{\text{std}}}{\partial w}=\frac{1}{2}(D_z t_k^{\text{std}}\cdot D_z+D_x t_k^{\text{std}}\cdot D_x)^{-1}. \tag{10}$$

Kinematically, each J_k^{std} contains characteristics of the kth shot. Because shots are independent of each other, the full Frechét derivative is a concatenation of individual J_k^{std}, as follows

$$J^{\text{std}} = [J_1^{\text{std}} J_2^{\text{std}} \cdots J_{nx}^{\text{std}}]^T. \tag{11}$$

Inserting equation 11 into equation 7, we obtain

$$\nabla_w E = \sum_{k=1}^{nx} (\mathbf{J}_k^{\text{std}})^T (\mathbf{t}_k^{\text{std}} - \mathbf{t}_k^{\text{obs}}), \tag{12}$$

where, similar to t_k^{std}, t_k^{obs} stands for the observed first breaks of the kth shot.

Figure 3 illustrates equation 12 schematically, i.e., the gradient produced by standard tomography. The first step on the left depicts the transpose of the kth Frechét derivative acting on the corresponding kth data residual. It implies a back-projection that takes place in the $z - r$ plane of a fixed s position. The second step on the right is the summation operation in equation 12.

To derive the Frechét derivative matrix associated with DSR tomography, we first rewrite equation 1 with definition 8

$$D_z t^{\text{dsr}} = -\sqrt{w_s - D_s t^{\text{dsr}} \cdot D_s t^{\text{dsr}}} - \sqrt{w_r - D_r t^{\text{dsr}} \cdot D_r t^{\text{dsr}}}, \tag{13}$$

where w_s and w_r are w at subsurface source and receiver locations, respectively. Note that, in equation 13, w appears twice. Thus, a differentiation of t^{dsr} with respect to w must be carried out through the chain-rule

$$J^{\text{dsr}} \equiv \frac{\partial t^{\text{dsr}}}{\partial w} = \frac{\partial t^{\text{dsr}}}{\partial w_s}\bigg|_{w_r} \frac{\partial w_s}{\partial w} + \frac{\partial t^{\text{dsr}}}{\partial w_r}\bigg|_{w_s} \frac{\partial w_r}{\partial w}. \tag{14}$$

We recall that w and t^{dsr} are of different lengths. Meanwhile, in equation 13, w_s and w_r have the size of t^{dsr}. Clearly, in equation 14, $\partial w_s/\partial w$ and $\partial w_r/\partial w$ must achieve dimensionality enlargement. In fact, according to Figure 1, w_s and w_r can be obtained by spraying w such that $w_s(z, r, s) = w(z, s)$ and $w_r(z, r, s) = w(z, r)$. Therefore, $\partial w_s/\partial w$ and $\partial w_r/\partial w$ are essentially spraying operators and their adjoints perform stackings along s and r dimensions, respectively.

In Appendix B, we prove that J^{dsr} has the following form:

$$J^{\text{dsr}} = B^{-1}(C_s + C_r). \tag{15}$$

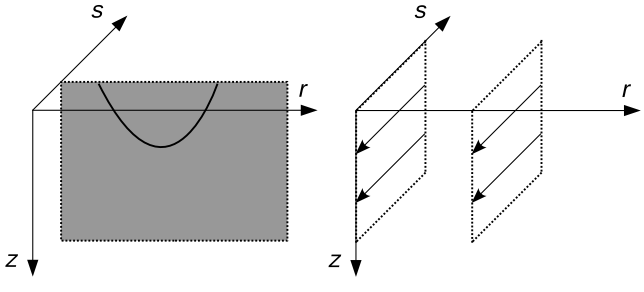

Figure 3. The gradient produced by standard tomography. The solid curve indicates a shot-indexed characteristic.

Combining equations 7 and 15 results in

$$\nabla_w E = (\mathbf{C}_s^T + \mathbf{C}_r^T)\mathbf{B}^{-T}(\mathbf{t}^{\text{dsr}} - \mathbf{t}^{\text{obs}}). \tag{16}$$

Note that, unlike equation 12, equation 16 cannot be expressed as an explicit summation over shots.

Figure 4 shows the gradient of DSR tomography. Similar to the standard tomography, the gradient produced by equation 16 is a result of two steps. The first step on the left is a back-projection of prestack data residuals according to the adjoint of operator B^{-1}. Because B contains DSR characteristics that travel in prestack domain, this back-projection takes place in (z, r, s) and is different from that in standard tomography, although the data residuals are the same for both cases. The second step on the right follows the adjoint of operators C_s and C_r and reduces the dimensionality from (z, r, s) to (z, x). However, compared to standard tomography this step involves summations in not only s but also r.

NUMERICAL IMPLEMENTATION

Following the analysis in Appendix A, we consider an implicit Eulerian discretization. For forward modeling, we solve the DSR eikonal equation by a version of the fast-marching method (FMM) (Sethian, 1999). First, a plane wave with $T = 0$ at subsurface zero-offset $r = s$ is initialized. Next, in the update stage, the traveltime at a grid point is computed from its upwind neighbors. A priority queue keeps track of the first-break wavefront, and the computation is nonrecursive.

To properly handle the DSR singularity, we design an ordering of the combination of upwind neighbors during the update stage. Assuming that T^i is the upwind neighbor of T in the i's direction for $i = z, r, s$, we summarize the ordering as follows:

1) First, try a three-sided update:
   Solve equation A-9, return T if $T \geq \max(T^z, T^r, T^s)$;
2) Next, try a two-sided update: solve equations A-10, A-12, and A-13 and keep the results as T_{rs}, T_{zr}, and T_{zs}, respectively.
   If $T_{zr} \geq \max(T^z, T^r)$ and $T_{zs} \geq \max(T^z, T^s)$,
   return $\min(T_{zr}, T_{zs}, T_{rs})$;
   If $T_{zr} < \max(T^z, T^r)$ and $T_{zs} \geq \max(T^z, T^s)$,
   return $\min(T_{zs}, T_{rs})$;
   If $T_{zr} \geq \max(T^z, T^r)$ and $T_{zs} < \max(T^z, T^s)$,
   return $\min(T_{zr}, T_{rs})$;
3) Finally, try a one-sided update:
   Solve equation A-14, return $\min(T, T_{rs})$.

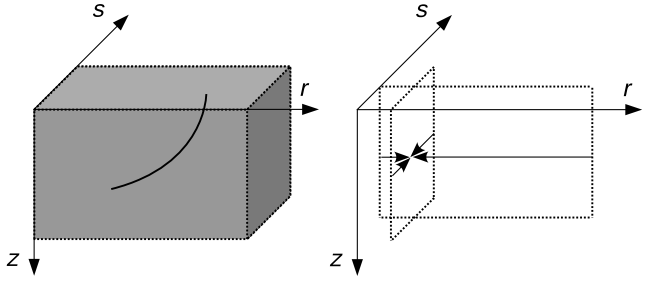

Figure 4. The gradient produced by DSR tomography. The solid curve indicates a DSR characteristic, which has one end in plane $z = 0$ and the other in plane $s = r$. Compare with Figure 3.

An optional search routine A-17 may be added after the update to recover all branches of the DSR eikonal equation. The overall cost can be reduced roughly by half by acknowledging the source-receiver reciprocity and thus computing only the positive (or negative) subsurface offset region.

For an implementation of linearized tomographic operators 12 and 16, we choose upwind approximations (Franklin and Harris, 2001; Li et al., 2011; Lelièvre et al., 2011) for the difference operators in equation 8. In Appendix C, we show that the upwind finite-differences result in triangulation of matrices 11 and 15. Therefore, the costs of applying $\mathbf{J}^{\text{std}}$ and $\mathbf{J}^{\text{dsr}}$ and their transposes are inexpensive. Moreover, although our implementation belongs to the family of adjoint-state tomographies, we do not need to compute the adjoint-state variable as an intermediate product for the gradient.

Additionally, the Gauss-Newton approach approximates the Hessian in equation 6 by $\mathbf{H} \approx \mathbf{J}^T\mathbf{J}$. An update $\delta\mathbf{w}$ at current $\mathbf{w}$ is found by taking derivative of equation 6 with respect to $\delta\mathbf{w}$, which results in the following normal equation:

$$\delta\mathbf{w} = [\mathbf{J}^T\mathbf{J}]^{-1}\mathbf{J}^T(\mathbf{t}^{\text{obs}} - \mathbf{t}). \tag{17}$$

To add model constraints, we combine equation 17 with Tikhonov regularization (Tikhonov, 1963) with the gradient operator and use the method of conjugate gradients (Hestenes and Stiefel, 1952) to solve for the model update $\delta\mathbf{w}$.

SYNTHETIC MODEL EXAMPLES

The numerical examples in this section serve several different purposes. The first example will test the accuracy of the modified FMM DSR eikonal equation solver (DSR FMM) and shows the drawbacks of the alternative explicit discretization. The second example will demonstrate the effect of considering noncausal branches of the DSR eikonal equation in forward modeling. The third example will compare the sensitivity kernels of DSR tomography and standard tomography in a simple model. The last example will present a tomographic inversion and demonstrate advantages of the DSR method over the traditional method.

Figure 5 shows a 2D velocity model with a constant-velocity-gradient background plus a Gaussian anomaly in the middle. We use Δ to denote the grid spacing in z and δ in x. The traveltimes on the surface $z = 0$ km of a shot at $(0, 0)$ km are computed by DSR FMM at a gradually refined Δ or δ while fixing the other one. For reference, we also calculate first breaks by a second-order FMM (Rickett and Fomel, 1999; Popovici and Sethian, 2002) for the same shot at a very fine grid spacing of $\Delta = \delta = 1$ m. In Figure 6, a grid refinement in Δ and δ helps reducing errors of the implicit discretization, although improvements in the Δ refinement case are less significant because most of the raypaths are nonhorizontal. The results are consistent with the analysis in Appendix A, which shows that the implicit discretization is unconditionally convergent. On the other hand, as shown in Figure 7, the explicit discretization is only conditionally convergent when $\Delta/\delta \to 0$ under grid refinement, in order to resolve the flatter parts of the raypaths. This explains why its accuracy deteriorates when refining δ and fixing Δ. A more detailed error analysis remains open for future research.

Next, we use a smoothed Marmousi model (Figure 8) and run two DSR FMMs, one with the search process for noncausal DSR branches turned on and the other turned off. In Figure 9, again we compute reference values by a second-order FMM. The three groups of curves are traveltimes of shots at (0,0) km, (0.75,0) km, and (1.5,0) km, respectively. The maximum absolute differences between the two DSR FMMs, for all three shots, are approximately 5 ms at the largest offset. This shows that, if the near-surface model is moderately complex, then the first breaks are of causal types described by equations 1 and 3, and we therefore can use their linearizations (equation 15) for tomography.

According to equations 11 and 15, the sensitivity kernels (a row of Frechét derivative matrix) of standard tomography and DSR tomography are different. Figure 10 compares sensitivity kernels for the same source-receiver pair in a constant-velocity-gradient model. We use a fine model sampling of $\Delta = \delta = 2.5$ m. The standard tomography kernel appears to be asymmetric. Its amplitude

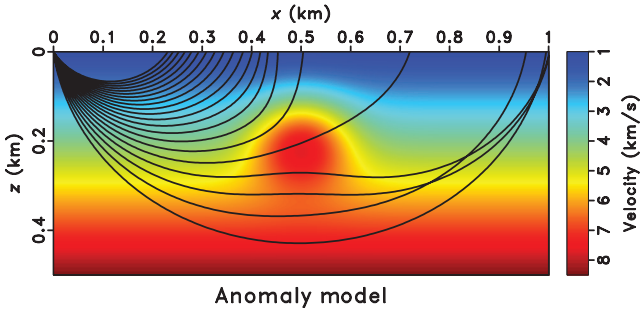

Figure 5. The synthetic model used for DSR FMM accuracy test. The overlaid curves are rays traced from a shot at (0,0) km.

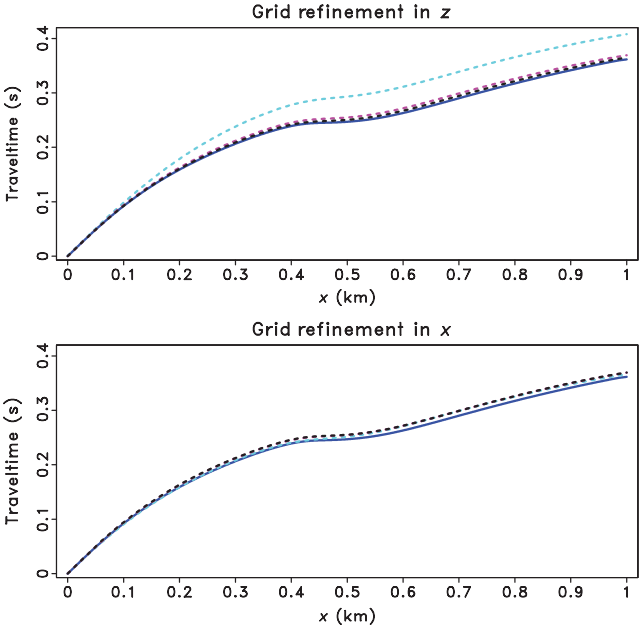

Figure 6. Grid refinement experiment (implicit discretization). In both figures, the solid blue curve is the reference values and the dashed curves are computed by DSR FMM. Top: fixed $\delta = 10$ m and $\Delta = 50$ m (cyan), 10 m (magenta), 5 m (black). Bottom: fixed $\Delta = 10$ m and $\delta = 50$ m (cyan), 10 m (magenta), 5 m (black).

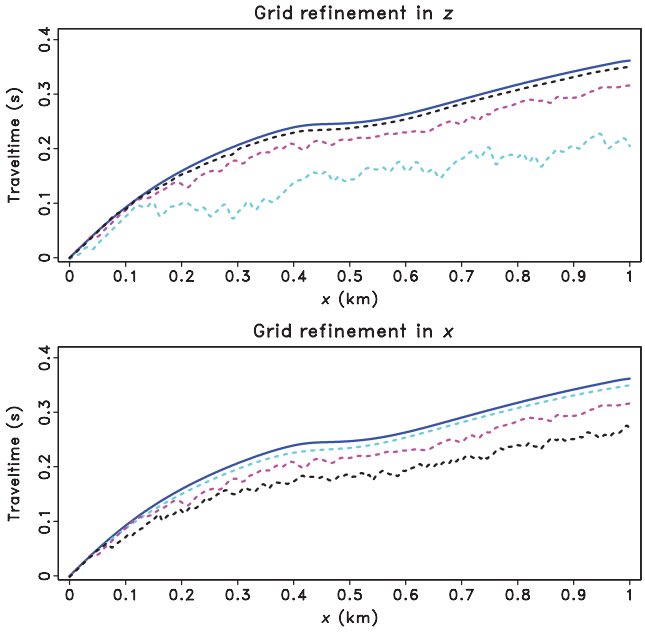

Figure 7. Grid refinement experiment (explicit discretization). The experiment setups are the same as in Figure 6.

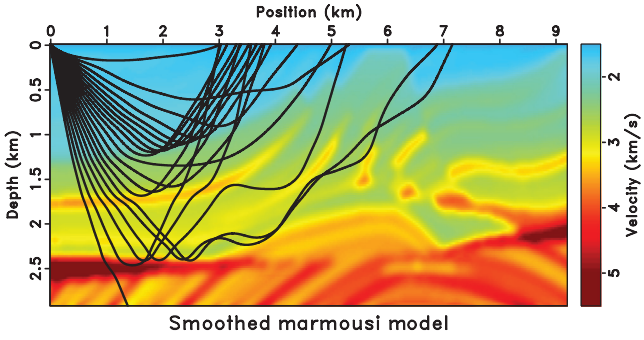

Figure 8. A smoothed Marmousi model overlaid with rays traced from a shot at (0,0) km. Because of velocity variations, multipathing is common in this model, especially at large offsets.

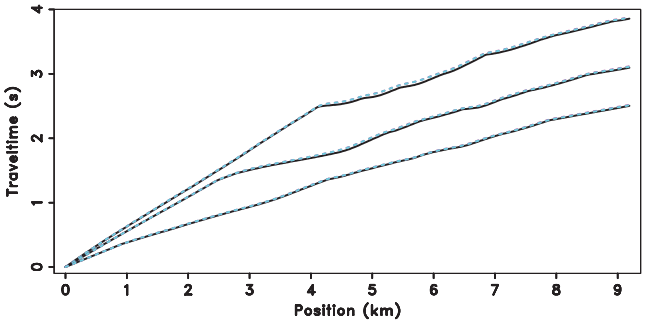

Figure 9. DSR FMM with noncausal branches. The solid black lines are reference values. There are two groups of dashed lines, from DSR FMM but one with the optional search process turned-on and the other without. The differences between them are negligible and hardly visible.

has a bias toward the source side, while the width is broader on the receiver side. These phenomena are related to our implementation, as described in Appendix C. Note in the top plot of Figure 10, the curvature of first-break wavefront changes during propagation. Upwind finite-differences take the curvature variation into consideration and, as a result, backproject data misfit with different weights along the raypath. Meanwhile, the DSR tomography kernel is symmetric in amplitude and width, even though it uses the same discretization and upwind approximation as in standard tomography. The source-receiver reciprocity may suggest averaging the standard tomography kernel with its own mirroring around $x = 1$ km; however, the result will still be different from the DSR tomography kernel as the latter takes into consideration all sources at the same time.

Finally, Figure 11 illustrates a prestack first-break traveltime modeling of the Marmousi model by DSR FMM. We use a constant-velocity-gradient model as the prior for inversion. There are 287 shots evenly distributed on the surface; each shot has a maximum absolute receiver offset of 6 km. Figure 12 shows a zoom-in of the exact

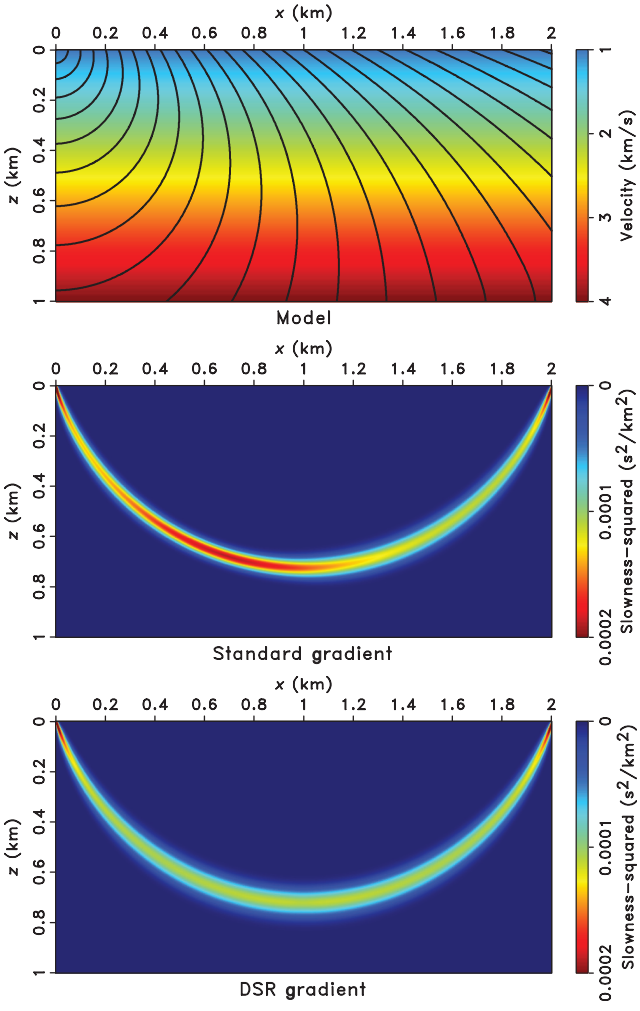

Figure 10. (Top) model overlaid with traveltime contours of a source at (0,0) km and sensitivity kernels of (middle) the standard tomography and (bottom) the DSR tomography.

model that is within the tomographic aperture. The DSR tomography and standard tomography are performed with the same parameters: Ten conjugate gradient iterations per linearization update and four linearization updates in total. Figure 13 shows the convergence histories. While both inversions converge, the relative l_2 data misfit of DSR tomography decreases faster than that of standard tomography. Figure 14 compares the recovered models. Although both results resemble the exact model in Figure 12 at the large scale, the standard tomography model exhibits several undesired structures. For example, a near-horizontal structure with a velocity of around 2.75 km/s at location (0.85,4.8) km is false. It indicates the presence of a local minimum that has trapped the standard tomography. In practice, it is helpful to tune the inversion parameters so that the standard tomography takes more iterations with a gradually reducing regularization. The inversion parameters are usually empirical and hard to control. Our analysis in preceding sections suggests that part of the role of regularization is to deal with conflicting information between shots. In contrast, we find DSR tomography less dependent on regularization and hence more robust.

The advantage of DSR tomography becomes more significant in the presence of noise in the input data. We first generate random noise with normal distribution and zero mean that ranges between ± 600 ms, then threshold the result with a minimum absolute value of 250 ms. This is to mimic the spiky errors in first breaks estimated from an automatic picker. After adding noise to the data, we run

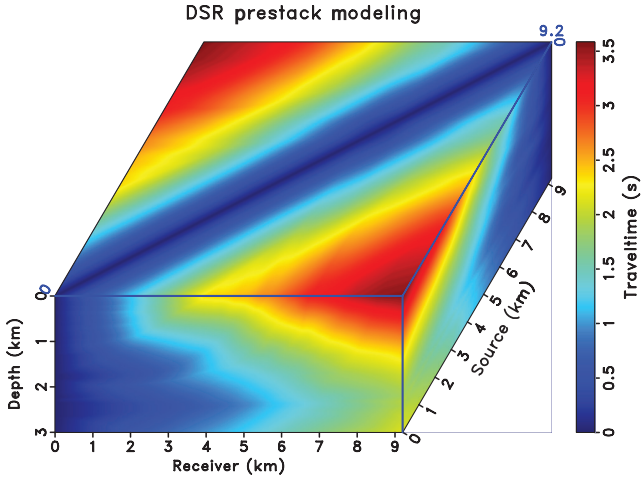

Figure 11. DSR first-break traveltimes in the Marmousi model. The original model is divided by two in vertical and lateral directions, such that $nz = 376$, $nx = 1151$ and $\Delta = \delta = 8$ m.

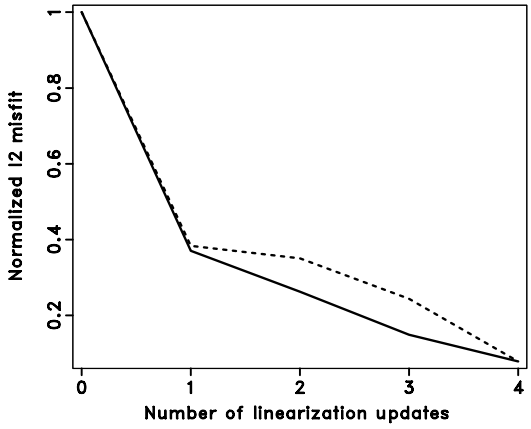

Figure 13. Convergency history of DSR tomography (solid) and standard tomography (dashed). There is no noticeable improvement on misfit after the fourth update.

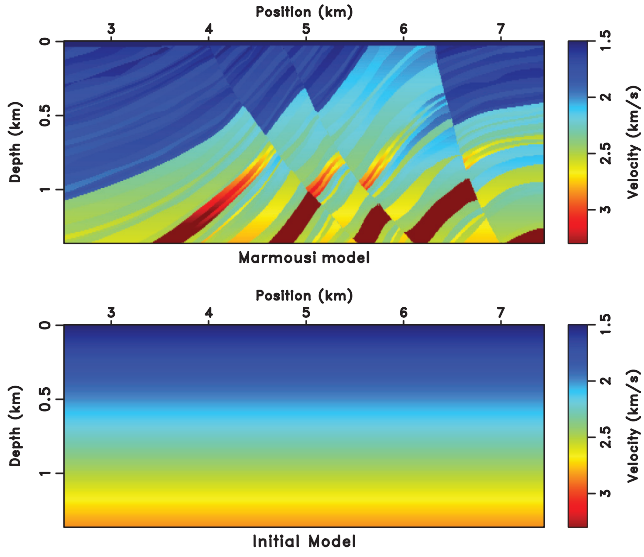

Figure 12. (Top) a zoom-in of Marmousi model and (bottom) the initial model for tomography.

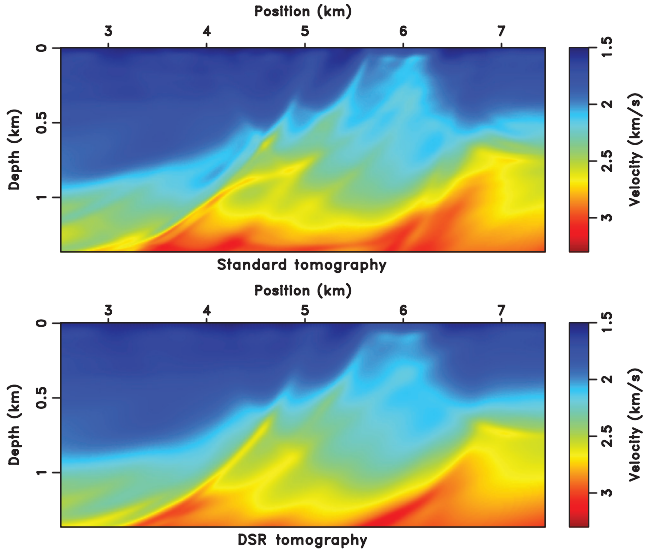

Figure 14. Inverted model of (top) standard tomography and (bottom) DSR tomography. Compare with Figure 12.

inversions with the same parameters as in Figures 13 and 14. Figures 15 and 16 show the convergence history and inverted models. Again, the standard tomography seems to provide a model with higher resolution, but a close examination reveals that many small scale details are in fact nonphysical. On the other hand, DSR tomography suffers much less from the added noise. Adopting a $l1$ norm in objective function 5 can improve the inversion, especially for standard tomography. However, it also raises the difficulty in selecting appropriate inversion parameters.

DISCUSSION

There are three main issues in the DSR tomography. The first issue comes from a large dimensionality of the prestack space, which results in a considerable computational domain size after discretization. The memory consumption becomes an immediate problem for 3D models, where the prestack traveltime belongs to a 5D space and may require distributed storage.

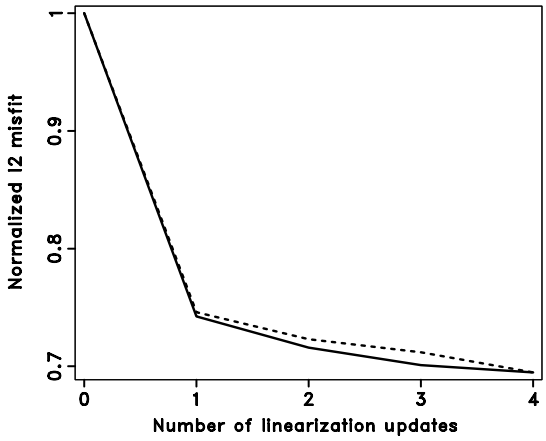

Figure 15. Inversion with noisy data. Convergency history of DSR tomography (solid) and standard tomography (dashed). No significant decrease in misfit appears after the fourth update.

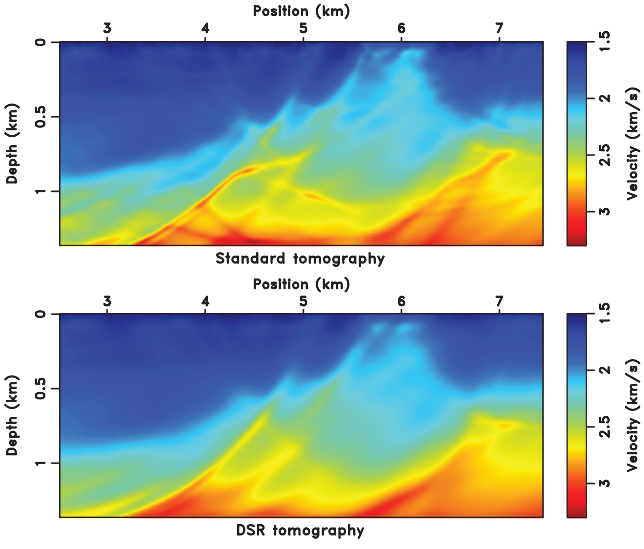

Figure 16. Inversion with noisy data. Inverted model of (top) standard tomography and (bottom) DSR tomography. Compare with Figure 14.

The second issue is related to the computational cost. The FMM DSR we have introduced in this paper has a computational complexity of $O(N \log N)$, where N is the total number of grid points after discretization, $N = nz \times nx^2$. The $\log N$ factor arises in the priority queue used in FMM for keeping track of expanding wavefronts. Some existing works could accelerate FMM to an $O(N)$ complexity and may be applicable to the DSR eikonal equation (Kim, 2001; Yatziv et al., 2006). Other fast methods developed for the eikonal equation might be similarly applicable to the DSR eikonal equation. These include fast-sweeping (Zhao, 2005), hybrid two-scale marching-sweeping methods and various label-correcting algorithms (see Chacon and Vladimirsky [2012] and references therein).

The last issue is possible parallelization of the proposed method. Our current implementation of the FMM DSR tomography algorithm is sequential, whereas the traditional tomography could be parallelized among different shots. However, we notice that the DSR eikonal equation has a plane wave source; therefore, a distributed wavefront propagating at the beginning followed by a subdomain merging is possible. A number of parallelizable algorithms for the eikonal equation have been developed (Zhao, 2007; Jeong and Whitaker, 2008; Weber et al., 2008; Detrixhe et al., 2013; A. Chacon, personal communication, 2012). Extending these methods to the DSR eikonal equation would be the first step in parallelizing DSR tomography.

CONCLUSIONS

We propose to use the DSR eikonal equation for the first-break traveltime tomography. The proposed method relies on an efficient DSR solver, which is realized by a version of the fast-marching method based on an implicitly causal discretization. Because the DSR eikonal equation allows changing of source position along the raypath, its linearization results in a tomographic inversion that naturally handles possible conflicting information between different shots. Our numerical tests show that, compared to the traditional tomography with a shot-indexed eikonal equation, the DSR tomography converges faster and is more robust. Its benefits may be particularly significant in the presence of noise in the data.

ACKNOWLEDGMENTS

This research was supported in part by Saudi Aramco. A. Vladimirsky's work was also partially supported by the NSF grant DMS-1016150. We thank Anton Duchkov, Mauricio Sacchi, Jörg Schleicher, Aldo Vesnaver, and four anonymous reviewers for valuable comments and suggestions. We thank Tariq Alkhalifah and Timothy Keho for useful discussions. This publication is authorized by the Director of the Bureau of Economic Geology, The University of Texas at Austin.

APPENDIX A

CAUSAL DISCRETIZATION OF DSR EIKONAL EQUATION

To simplify the analysis, we consider first the DSR branch as shown in Figure 1 and described by equation 1. We assume a rectangular 2D

velocity model $v(z,x)$ and thus a cubic 3D prestack volume $T(z,r,s)$ with r and s axes having the same dimension as x. After an Eulerian discretization of v and T, we denote the grid spacing in z as Δ, and in x, r, and s as δ.

In Figure A-1, we study the traveltime at grid point $\mathbf{y}=(z,r,s)$ and its relationship with neighboring grid points $\mathbf{y}^z=(z+\Delta,r,s)$, $\mathbf{y}^r=(z,r-\delta,s)$, and $\mathbf{y}^s=(z,r,s+\delta)$ with a semi-Lagrangian scheme. According to the geometry in Figure 1, in the (z,r,s) space, the DSR characteristic (Duchkov and de Hoop, 2010) is straddled by $\mathbf{y}^z\mathbf{y}^r\mathbf{y}^s$.

To compute T, we could continue along this characteristic up until its intersection with the simplex $\mathbf{y}^z\mathbf{y}^r\mathbf{y}^s$. Suppose the intersection point is $\tilde{\mathbf{y}}=(\tilde{z},\tilde{r},\tilde{s})$ and α_i are its barycentric coordinates, i.e.,

$$\alpha_z,\alpha_r,\alpha_s\in[0,1];\quad \alpha_z+\alpha_r+\alpha_s=1;\quad \tilde{\mathbf{y}}=\alpha_z\mathbf{y}^z+\alpha_r\mathbf{y}^r+\alpha_s\mathbf{y}^s. \tag{A-1}$$

This leads to the following discretization

$$T\equiv T(\mathbf{y}) = \min_{\tilde{\mathbf{y}}\in\mathbf{y}^z\mathbf{y}^r\mathbf{y}^s}\left\{T(\tilde{\mathbf{y}})+\frac{\sqrt{(z-\tilde{z})^2+(r-\tilde{r})^2}}{v(z,r)}+\frac{\sqrt{(z-\tilde{z})^2+(s-\tilde{s})^2}}{v(z,s)}\right\}. \tag{A-2}$$

Here, we further assume that $v(z,r)$ and $v(z,s)$ are locally constant and that ray segments between z and $z+\Delta$ are well-approximated by straight lines. This also means that a linear interpolation in T within the simplex $\mathbf{y}^z\mathbf{y}^r\mathbf{y}^s$ is exact, i.e., $T(\tilde{\mathbf{y}})=\alpha_zT^z+\alpha_rT^r+\alpha_sT^s$, where $T^i=T(\mathbf{y}^i)$ for $i=z,r,s$. The minimization over all possible intersection points in equation A-2 guarantees a first-arrival traveltime.

Defining the ratio in grid spacing as $\mu\equiv\Delta/\delta$ and denoting $v_r=v(z,r)$ and $v_s=v(z,s)$, equation A-2 can be rewritten with the barycentric coordinates in A-1 as

$$T=\min_{(\alpha_z,\alpha_r,\alpha_s)}\left\{(\alpha_zT^z+\alpha_rT^r+\alpha_sT^s)+\frac{\delta}{v_r}\sqrt{\alpha_r^2+\mu^2\alpha_z^2}+\frac{\delta}{v_s}\sqrt{\alpha_s^2+\mu^2\alpha_z^2}\right\}. \tag{A-3}$$

Figure A-1 is based on a particular direction of the diving wave: rightward from the source and leftward from the receiver, as in Figure 1. This yields the above positions of $\mathbf{y}^s$ and $\mathbf{y}^r$, and the formula A-3 becomes an update from the $\mathbf{y}^z\mathbf{y}^r\mathbf{y}^s$ quadrant. Because, in general, the direction of a diving wave is not known a priori, we compute one such update from each of the lower quadrants and take the smallest amongst them as a value of T.

To explore the causal properties of equation A-3, we first assume that the minimum is attained at some $\tilde{\mathbf{y}}^*=\xi_z\mathbf{y}^z+\xi_r\mathbf{y}^r+\xi_s\mathbf{y}^s$ such that $\xi_i>0$ for $i=z,r,s$. From the Kuhn-Tucker optimality conditions (Kuhn and Tucker, 1951), there exists a Lagrange multiplier λ such that

$$\lambda=T^z+\delta\left(\frac{\mu^2\xi_z}{v_r\sqrt{\xi_r^2+\mu^2\xi_z^2}}+\frac{\mu^2\xi_z}{v_s\sqrt{\xi_s^2+\mu^2\xi_z^2}}\right); \tag{A-4}$$

$$\lambda=T^i+\delta\left(\frac{\xi_i}{v_i\sqrt{\xi_i^2+\mu^2\xi_z^2}}\right);\quad i=r,s. \tag{A-5}$$

Taking a linear combination of A-4 and A-5 to match the right side of A-3, we find that $\lambda=T$ and thus

$$T-T^z=\delta\left(\frac{\mu^2\xi_z}{v_r\sqrt{\xi_r^2+\mu^2\xi_z^2}}+\frac{\mu^2\xi_z}{v_s\sqrt{\xi_s^2+\mu^2\xi_z^2}}\right)>0; \tag{A-6}$$

$$T-T^i=\delta\left(\frac{\xi_i}{v_i\sqrt{\xi_i^2+\mu^2\xi_z^2}}\right)>0;\ i=r,s. \tag{A-7}$$

This means that if T defined by equation A-3 depends on T^i then $T>T^i$ for $i=z,r,s$, or

$$T>\max(T^z,T^r,T^s) \tag{A-8}$$

and a Dijkstra-like method (Dijkstra, 1959) is applicable to solve the discretized system.

A direct substitution from equations A-6 and A-7 results in

$$\frac{T-T^z}{\Delta}=\sqrt{\frac{1}{v_r^2}-\left(\frac{T-T^r}{\delta}\right)^2}+\sqrt{\frac{1}{v_s^2}-\left(\frac{T-T^s}{\delta}\right)^2}. \tag{A-9}$$

If T^z, T^r, and T^s are known, then T can be recovered by solving the fourth order polynomial equation A-9 and choosing the smallest real root that satisfies condition A-8. This gives a three-sided update at T. The discretization is *implicitly causal* and provides unconditional consistency and convergence.

If there is no real root or none of the real roots satisfy A-8, the minimizer $\tilde{\mathbf{y}}$ cannot lie in the interior of simplex $\mathbf{y}^z\mathbf{y}^r\mathbf{y}^s$ and at least one of the ξ_is must be zero. If $\xi_z=0$, it is easy to show that one of the other barycentric coordinates is also zero and equation A-3 simplifies to

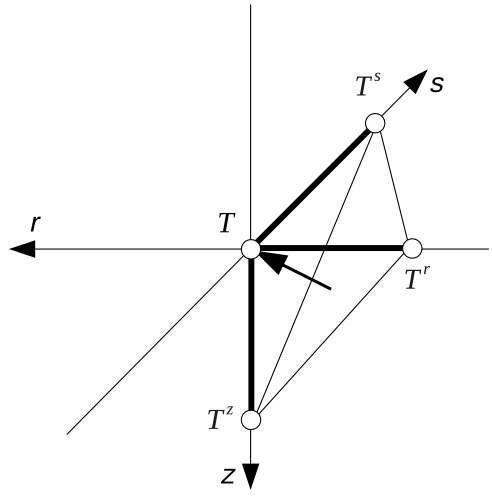

Figure A-1. An implicit discretization scheme. The arrow indicates a DSR characteristic. Its root is located in the simplex $T^zT^rT^s$.

$$T = \min\left(T^r + \frac{\delta}{v_r}, T^s + \frac{\delta}{v_s}\right), \tag{A-10}$$

which is a causal discretization of the DSR singularity in equation 3. On the other hand, if $\xi_z \neq 0$ and $\xi_r \neq 0$, but $\xi_s = 0$, i.e., the slowness vector at s is vertical, then

$$T = (\xi_z T^z + \xi_r T^r) + \frac{\delta}{v_r}\sqrt{\xi_r^2 + \mu^2 \xi_z^2} + \frac{\Delta}{v_s}\xi_z. \tag{A-11}$$

A similar Kuhn-Tucker-type argument shows that equation A-11 is also causal: If $\xi_z, \xi_r > 0$, then $T > \max(T^z, T^r)$. In this case, T can be computed by solving

$$\frac{T - T^z}{\Delta} = \sqrt{\frac{1}{v_r^2} - \left(\frac{T - T^r}{\delta}\right)^2} + \frac{1}{v_s}. \tag{A-12}$$

Equation A-12 is equivalent to setting $\partial T/\partial s = 0$ in equation 1. Analogously, when $\xi_z \neq 0$ and $\xi_s \neq 0$ but $\xi_r = 0$, we have $\partial T/\partial r = 0$ and

$$\frac{T - T^z}{\Delta} = \frac{1}{v_r} + \sqrt{\frac{1}{v_s^2} - \left(\frac{T - T^s}{\delta}\right)^2}, \tag{A-13}$$

with the causal solution satisfying $T > \max(T^z, T^s)$. Equations A-12 and A-13 provide a two-sided update at T. Finally, if $\xi_z \neq 0$ but $\xi_s = 0$ and $\xi_r = 0$, i.e., $\partial T/\partial s = 0$ and $\partial T/\partial r = 0$, we obtain the simplest one-sided update

$$\frac{T - T^z}{\Delta} = \frac{1}{v_r} + \frac{1}{v_s}. \tag{A-14}$$

We note that the one-sided update A-14 could be considered a special case of two-sided updates: If $T = T^r$ (or T^s), then equation A-14 becomes equivalent to A-12 (or, respectively, A-13). Similarly, the two-sided updates can be viewed as special versions of the three-sided one: e.g., if $T = T^r = \max(T^z, T^r, T^s)$, then equation A-9 becomes equivalent to A-13. This means that the causal criteria for formulas A-9, A-12, and A-13 can be relaxed (the inequalities do not have to be strict). This relaxation is used to streamline the update strategy in the "Numerical implementation" section.

In Figure A-1 and the corresponding semi-Lagrangian discretization A-2, the raypath is linearly approximated up to its intersection with the simplex $\mathbf{y}^z\mathbf{y}^r\mathbf{y}^s$ at a priori unknown depth $z + \xi_z\Delta$. An alternative *explicit* semi-Lagrangian discretization can be obtained in the spirit of Figure 1 by tracing the ray up to the prespecified depth $z + \Delta$. In Figure A-2, we consider the DSR characteristic being straddled by $\mathbf{y}_*^z\mathbf{y}_*^r\mathbf{y}_*^s$, where $\mathbf{y}_*^z = (z+\Delta, r, s)$, $\mathbf{y}_*^r = (z+\Delta, r-\delta, s)$, and $\mathbf{y}_*^s = (z + \Delta, r, s + \delta)$. Denoting $\tilde{\mathbf{y}}_* = (z + \Delta, \tilde{r}_*, \tilde{s}_*)$ for the intersection point between DSR characteristic and simplex $\mathbf{y}_*^z\mathbf{y}_*^r\mathbf{y}_*^s$, we obtain the following discretization

$$T = \min_{\tilde{\mathbf{y}}_* \in \mathbf{y}_*^z\mathbf{y}_*^r\mathbf{y}_*^s} \left\{ T(\tilde{\mathbf{y}}_*) + \frac{\sqrt{\Delta^2 + (r - \tilde{r}_*)^2}}{v(z,r)} + \frac{\sqrt{\Delta^2 + (s - \tilde{s}_*)^2}}{v(z,s)} \right\}. \tag{A-15}$$

One could perform the same analysis of equations A-3–A-9 to equation A-15. For the sake of brevity, we omit the derivation and show the resulting explicit discretization scheme

$$\frac{T - T_*^z}{\Delta} = \sqrt{\frac{1}{v_r^2} - \left(\frac{T_*^z - T_*^r}{\delta}\right)^2} + \sqrt{\frac{1}{v_s^2} - \left(\frac{T_*^z - T_*^s}{\delta}\right)^2}, \tag{A-16}$$

where $T_*^i = T(\mathbf{y}_*^i)$ for $i = z, r, s$. More generally, to account for various possible directions of the diving wave, we can set $T_*^r = \min(T(z+\Delta, r-\delta, s), T(z+\Delta, r+\delta, s))$ and $T_*^s = \min(T(z + \Delta, r, s - \delta), T(z + \Delta, r, s + \delta))$.

Compared with equation A-9, A-16 does not require solving a polynomial equation. Moreover, T depends only on values in lower z-slices, which means that the system of equations can be solved in a single sweep in the $-z$-direction. Unfortunately, despite this efficiency on a fixed grid, the explicit discretization has a major disadvantage stemming from the requirement that the characteristic should be straddled by $y_*^z y_*^r y_*^s$. This imposes an upper bound on μ based on the slope of the diving wave. Moreover, because every diving ray is horizontal at its lowest point, the convergence is possible only if $\mu \to 0$ under mesh refinement. In practice, this means that the results are meaningful only if Δ is significantly smaller than δ. We note that restrictive stability conditions also arise for time-dependent Hamilton-Jacobi equations of optimal control, where sufficiently strong inhomogenieties can make nonlinear/implicit schemes preferable to the usual linear/explicit approach (A. Vladimirsky, personal communication, 2013).

The above analysis also applies to the first branch of the DSR eikonal equation in Figure 2. However, in the discretized (z, r, s) domain, the slowness vectors at s and r are always aligned in the z-direction, either upward or downward. For this reason, there is no DSR characteristic that accounts for the second and third scenarios. We will refer to the first and last branches in Figure 2 as *causal branches* of DSR eikonal equation, and the leftover two as *noncausal branches*.

Note that when the slowness vectors at s and r are pointing in opposite directions, there must be at least one intersection of the raypath with the z depth level in-between. As shown in Figure A-3, ray segments between these intersections fall into the category of causal branches. Thus, a search process for the intersections is sufficient in recovering the noncausal branches during forward modeling. Moreover, because we are interested in first breaks only, the minimum traveltime requirement allows us to search for only one intersection, such as q denoted in Figure A-3

$$T(z, r, s) = \min_{q \in (s,r)} \{T(z, q, s) + T(z, r, q)\}. \tag{A-17}$$

Other possible intersections in intervals (s, q) and (q, r) have already been recovered when computing $T(z, q, s)$ and $T(z, r, q)$, as long as we enable the intersection searching from the beginning of forward modeling. The traveltime of noncausal branches from equation A-17 is then compared with that from causal branches, and the smaller one should be kept.

Unfortunately, this search routine induces considerable computational cost. Moreover, we note that, under a dominant diving waves assumption, the first DSR branch, despite being causal, becomes useless if the search routine is turned off.

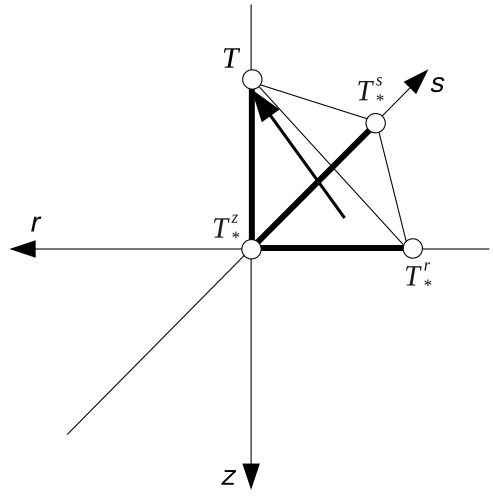

Figure A-2. An explicit discretization scheme. Compare with Figure A-1. The arrow again depicts a DSR characteristic with its root confined in the simplex $T^z_* T^r_* T^s_*$.

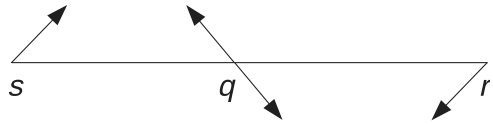

Figure A-3. When slowness vectors at s and r are pointing in the opposite directions, the raypath must intersect with line $s-r$ at certain point q.

APPENDIX B

FRECHÉT DERIVATIVE OF DSR TOMOGRAPHY

To derive the Frechét derivative, we start from equations 13 and 14. Applying $\partial/\partial w_s$ to both sides of equation 13 results in

$$D_z \frac{\partial t^{\mathrm{dsr}}}{\partial w_s} = -\frac{1}{2\sqrt{w_s - D_s t^{\mathrm{dsr}} \cdot D_s t^{\mathrm{dsr}}}} + \left(\frac{D_s t^{\mathrm{dsr}} \cdot D_s}{\sqrt{w_s - D_s t^{\mathrm{dsr}} \cdot D_s t^{\mathrm{dsr}}}} + \frac{D_r t^{\mathrm{dsr}} \cdot D_r}{\sqrt{w_r - D_r t^{\mathrm{dsr}} \cdot D_r t^{\mathrm{dsr}}}} \right) \frac{\partial t^{\mathrm{dsr}}}{\partial w_s}. \tag{B-1}$$

Analogously,

$$D_z \frac{\partial t^{\mathrm{dsr}}}{\partial w_r} = -\frac{1}{2\sqrt{w_r - D_r t^{\mathrm{dsr}} \cdot D_r t^{\mathrm{dsr}}}} + \left(\frac{D_s t^{\mathrm{dsr}} \cdot D_s}{\sqrt{w_s - D_s t^{\mathrm{dsr}} \cdot D_s t^{\mathrm{dsr}}}} + \frac{D_r t^{\mathrm{dsr}} \cdot D_r}{\sqrt{w_r - D_r t^{\mathrm{dsr}} \cdot D_r t^{\mathrm{dsr}}}} \right) \frac{\partial t^{\mathrm{dsr}}}{\partial w_r}. \tag{B-2}$$

Inserting equations B-1 and B-2 into 14 and regrouping the terms, we prove equation 15

$$J^{\mathrm{dsr}} = B^{-1}(C_s + C_r), \tag{B-3}$$

where

$$B = D_z - \left(\frac{D_s t^{\mathrm{dsr}} \cdot D_s}{\sqrt{w_s - D_s t^{\mathrm{dsr}} \cdot D_s t^{\mathrm{dsr}}}} \right) - \left(\frac{D_r t^{\mathrm{dsr}} \cdot D_r}{\sqrt{w_r - D_r t^{\mathrm{dsr}} \cdot D_r t^{\mathrm{dsr}}}} \right), \tag{B-4}$$

and

$$C_s = -\frac{1}{2\sqrt{w_s - D_s t^{\mathrm{dsr}} \cdot D_s t^{\mathrm{dsr}}}} \frac{\partial w_s}{\partial w}; \tag{B-5}$$

$$C_r = -\frac{1}{2\sqrt{w_r - D_r t^{\mathrm{dsr}} \cdot D_r t^{\mathrm{dsr}}}} \frac{\partial w_r}{\partial w}. \tag{B-6}$$

At the singularity of DSR eikonal equation, the operators B, C_s, and C_r take simpler forms and can be derived directly from equation 3.

APPENDIX C

ADJOINT-STATE TOMOGRAPHY WITH UPWIND FINITE-DIFFERENCES

Following Appendix A, we let $T_i^{j,k}$ in the DSR case be the traveltime at vertex (z_i, r_j, s_k) and approximate D_z in equation 8 by a one-sided finite-difference

$$D_z^{\pm} T_i^{j,k} = \pm \frac{T_{i\pm1}^{j,k} - T_i^{j,k}}{\Delta}, \tag{C-1}$$

where the $\pm$ sign corresponds to the two neighbors of $T_i^{j,k}$ in z-direction. An upwind scheme (Franklin and Harris, 2001) picks the sign by

$$D_z T_i^{j,k} = \max\left(D_z^- T_i^{j,k}, -D_z^+ T_i^{j,k}, 0\right). \tag{C-2}$$

The above strategy can be applied to D_r and D_s straightforwardly. For the shot-indexed eikonal equation 9, we approximate D_x with the same upwind method while T in this case is indexed for z and x.

For a Cartesian ordering of the discretized T, i.e., vector $\mathbf{t}$, the discretized operators $D_m T \cdot D_m$ with $m = z, x, r, s$ are matrices. Thanks to upwind finite-differences, these matrices are sparse and contain only two nonzero entries per row. For instance, suppose $T_i^{j,k}$ has its upwind neighbor in z at $T_{i-1}^{j,k}$, then

$$D_z T \cdot D_z = \begin{bmatrix} \ddots & & & & \\ & \ddots & & & \\ & & \ddots & & \\ & -\kappa_z & & \kappa_z & \\ & & & & \ddots \end{bmatrix}, \tag{C-3}$$

where

$$\kappa_z \equiv \frac{D_z T_i^{j,k}}{\Delta} = \frac{T_i^{j,k} - T_{i-1}^{j,k}}{\Delta^2}. \quad \text{(C-4)}$$

Definitions of κ_r, κ_s, and κ_x follow their upwind approximations, respectively. In matrix C-3, $\pm\kappa_z$ are located in the same row as that of $T_i^{j,k}$ in $\mathbf{t}$. While κ_z sits on the diagonal, $-\kappa_z$ has a column index equals to the row of $T_{i-1}^{j,k}$ in $\mathbf{t}$. At $T = 0$, there is no upwind neighbor and the corresponding row contains all zeros.

We can sort entries of $\mathbf{t}$ by their values in an increasing order, which equivalently performs columnwise permutations to $D_m T \cdot D_m$. The results are lower triangular matrices. In fact, during FMM forward modeling, such an *upwind ordering* is maintained and updated by the priority queue and thus can be conveniently imported for usage here.

Note that the summation and subtraction of two (or more) $D_m T \cdot D_m$ matrices are still lower triangular. These matrices are also invertible, except for a singularity at $T = 0$ where we may set the entries to be zero. Naturally, the inverted matrices are also lower triangular. One example is the linearized eikonal equation that gives rise to equation 10. Following notation C-4 and assuming the upwind neighbors of T_i^j are T_{i-1}^j and T_i^{j-1}, the linearized equation 9 reads

$$2\kappa_z(\delta T_i^j - \delta T_{i-1}^j) + 2\kappa_x(\delta T_i^j - \delta T_i^{j-1}) = \delta w_i^j. \quad \text{(C-5)}$$

After regrouping the terms, we get

$$\delta T_i^j = \frac{2\kappa_z \delta T_{i-1}^j + 2\kappa_x \delta T_i^{j-1} + \delta w_i^j}{2(\kappa_z + \kappa_x)}. \quad \text{(C-6)}$$

Equation C-6 means the inverse of the operator $D_z T \cdot D_z + D_x T \cdot D_x$ does not need to be computed by an explicit matrix inversion. Instead, we can perform its application to a vector by a single sweep based on causal upwind ordering. The same conclusion can be drawn for operator B-4.

Lastly, the adjoint-state calculations implied by equations 12 and 16 multiply the transpose of these inverse matrices with the data residual. The matrix transposition leads to upper triangular matrices. Accordingly, we solve the linear system with anticausal *downwind ordering* that follows a decrease in values of $\mathbf{t}$.

REFERENCES

Alkhalifah, T., 2011, Prestack traveltime approximations: 81st Annual International Meeting, SEG, Expanded Abstracts, 3017–3021.

Belonosova, A. V., and A. S. Alekseev, 1974, About one formulation of the inverse kinematic problem of seismics for a two-dimensional continuously heterogeneous medium: Some methods and algorithms for interpretation of geophysical data (in Russian): Nauka, 137–154.

Bergman, B., A. Tryggvason, and C. Juhlin, 2004, High-resolution seismic traveltime tomography incorporating static corrections applied to a till-covered bedrock environment: Geophysics, **69**, 1082–1090, doi: 10.1190/1.1778250.

Bertsekas, D., 1982, Enlarging the region of convergence of Newton's method for constrained optimization: Journal of Optimization Theory and Applications, **36**, 221–252, doi: 10.1007/BF00933831.

Brenders, A. J., and R. G. Pratt, 2007, Efficient waveform tomography for lithospheric imaging: Implications for realistic, two-dimensional acquisition geometries and low-frequency data: Geophysical Journal International, **168**, 152–170, doi: 10.1111/j.1365-246X.2006.03096.x.

Chacon, A., and A. Vladimirsky, 2012, Fast two-scale methods for eikonal equations: SIAM Journal on Scientific Computing, **34**, no. 2, A547–A578, doi: 10.1137/10080909X.

Chapman, C., 2002, Fundamentals of seismic wave propagation: Cambridge University Press.

Cox, M., 1999, Static corrections for seismic reflection surveys: SEG.

Dessa, J. X., S. Operto, A. Nakanishi, G. Pascal, K. Uhira, and Y. Kaneda, 2004, Deep seismic imaging of the eastern Nankai Trough, Japan, from multifold ocean bottom seismometer data by combined traveltime tomography and prestack depth migration: Journal of Geophysical Research, **109**, B02111, doi: 10.1029/2003JB002689.

Detrixhe, M., C. Min, and F. Gibou, 2013, A parallel fast sweeping method for the eikonal equation: Journal of Computational Physics, **237**, 46–55, doi: 10.1016/j.jcp.2012.11.042.

Dijkstra, E. W., 1959, A note on two problems in connection with graphs: Numerische Mathematik, **1**, 269–271, doi: 10.1007/BF01386390.

Duchkov, A., and M. V. de Hoop, 2010, Extended isochron rays in prestack depth (map) migration: Geophysics, **75**, no. 4, S139–S150, doi: 10.1190/1.3459955.

Engl, H. W., M. Hanke, and A. Neubauer, 1996, Regularization of inverse problems: Kluwer Academic Publishers.

Franklin, J. B., and J. M. Harris, 2001, A high-order fast marching scheme for the linearized eikonal equation: Journal of Computational Acoustics, **9**, 1095–1109.

Hestenes, M. R., and E. Stiefel, 1952, Method of conjugate gradients for solving linear systems: Journal of Research of the National Bureau of Standards (United States), **49**, 409–436, doi: 10.6028/jres.049.044.

Iversen, E., 2004, The isochron ray in seismic modeling and imaging: Geophysics, **69**, 1053–1070, doi: 10.1190/1.1778248.

Jeong, W., and R. T. Whitaker, 2008, A fast iterative method for eikonal equations: SIAM Journal on Scientific Computing, **30**, 2512–2534, doi: 10.1137/060670298.

Kim, S., 2001, An $O(N)$ level set method for eikonal equations: SIAM Journal on Scientific Computing, **22**, 2178–2193, doi: 10.1137/S1064827500367130.

Kuhn, H. W., and A. W. Tucker, 1951, Nonlinear programming: Proceedings of the 2nd Berkeley Symposium, 481–492.

Lelièvre, P. G., C. G. Farquharson, and C. A. Hurich, 2011, Inversion of first-arrival seismic traveltimes without rays, implemented on unstructured grids: Geophysical Journal International, **185**, 749–763, doi: 10.1111/j.1365-246X.2011.04964.x.

Leung, S., and J. Qian, 2006, An adjoint state method for three-dimensional transmission traveltime tomography using first arrivals: Communications in Mathematical Sciences, **4**, 249–266.

Li, S., S. Fomel, and A. Vladimirsky, 2011, Improving wave-equation fidelity of Gaussian beams by solving the complex eikonal equation: 71st Annual International Meeting, SEG, Expanded Abstracts, 3829–3834.

Marsden, D., 1993, Static corrections — A review: The Leading Edge, **12**, 43–49, doi: 10.1190/1.1436912.

Noble, M., P. Thierry, C. Taillandier, and H. Calandra, 2010, High-performance 3D first-arrival traveltime tomography: The Leading Edge, **29**, 86–93, doi: 10.1190/1.3284057.

Osypov, K., 2000, Robust refraction tomography: 70th Annual International Meeting, SEG, Expanded Abstracts, 2032–2035.

Paige, C. C., and M. A. Saunders, 1982, LSQR: An algorithm for sparse linear equations and sparse least squares: ACM Transactions on Mathematical Software, **8**, 43–71, doi: 10.1145/355984.355989.

Pei, D., 2009, Three-dimensional traveltime tomography via LSQR with regularization: 79th Annual International Meeting, SEG, Expanded Abstracts, 4004–4008.

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, doi: 10.1111/j.1365-246X.2006.02978.x.

Popovici, A. M., and J. Sethian, 2002, 3-D imaging using higher order fast marching traveltimes: Geophysics, **67**, 604–609, doi: 10.1190/1.1468621.

Rickett, J., and S. Fomel, 1999, A second-order fast marching eikonal solver: Stanford Exploration Project Report, **100**, 287–292.

Sethian, J. A., 1999, Level set methods and fast marching methods: Evolving interfaces in computational geometry, fluid mechanics, computer vision and material sciences: Cambridge University Press.

Sheng, J., A. Leeds, M. Buddensiek, and G. T. Schuster, 2006, Early arrival waveform tomography on near-surface refraction data: Geophysics, **71**, no. 4, U47–U57, doi: 10.1190/1.2210969.

Simmons, J. L., and N. Bernitsas, 1994, Nonlinear inversion of first-arrival times: 64th Annual International Meeting, SEG, Expanded Abstracts, 992–995.

Stefani, J. P., 1993, Possibilities and limitations of turning ray tomography — A synthetics study: 63rd Annual International Meeting, SEG, Expanded Abstracts, 610–612.

Taillandier, C., M. Noble, H. Chauris, and H. Calandra, 2009, First-arrival traveltime tomography based on the adjoint-state method: Geophysics, **74**, no. 6, WCB1–WCB10, doi: 10.1190/1.3250266.

Tikhonov, A. N., 1963, Solution of incorrectly formulated problems and the regularization method: Soviet Mathematics Doklady, 1035–1038.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26, doi: 10.1190/1.3238367.

Weber, O., Y. Devir, A. Bronstein, M. Bronstein, and R. Kimmel, 2008, Parallel algorithms for the approximation of distance maps on parametric surfaces: ACM Transactions on Graphics, **27**, no. 4, 1–16, doi: 10.1145/1409625.1409626.

Yatziv, L., A. Bartesaghi, and G. Sapiro, 2006, A fast $O(N)$ implementation of the fast marching algorithm: Journal of Computational Physics, **212**, 393–399, doi: 10.1016/j.jcp.2005.08.005.

Zelt, C. A., and P. J. Barton, 1998, 3D seismic refraction tomography: A comparison of two methods applied to data from the Faeroe Basin: Journal of Geophysical Research, **103**, 7187–7210, doi: 10.1029/97JB03536.

Zhao, H., 2005, A fast sweeping method for eikonal equations: Mathematics of Computation, **74**, 603–628, doi: 10.1090/S0025-5718-04-01678-3.

Zhao, H., 2007, Parallel implementations of the fast sweeping method: Journal of Computational Mathematics, **25**, 421–429.

Zhu, T., S. Cheadle, A. Petrella, and S. Gray, 2000, First-arrival tomography: Method and application: 70th Annual International Meeting, SEG, Expanded Abstracts, 2028–2031.

Zhu, X., D. P. Sixta, and B. G. Angstman, 1992, Tomostatics: Turning-ray tomography + static corrections: The Leading Edge, **11**, 15–23, doi: 10.1190/1.1436864.