GEOPHYSICS, VOL. 74, NO. 6 (NOVEMBER-DECEMBER 2009); P. WCB1–WCB10, 11 FIGS.
10.1190/1.3250266

First-arrival traveltime tomography based on the adjoint-state method

Cédric Taillandier[1], Mark Noble[2], Hervé Chauris[2], and Henri Calandra[3]

ABSTRACT

Classical algorithms used for traveltime tomography are not necessarily well suited for handling very large seismic data sets or for taking advantage of current supercomputers. The classical approach of first-arrival traveltime tomography was revisited with the proposal of a simple gradient-based approach that avoids ray tracing and estimation of the Fréchet derivative matrix. The key point becomes the derivation of the gradient of the misfit function obtained by the adjoint-state technique. The adjoint-state method is very attractive from a numerical point of view because the associated cost is equivalent to the solution of the forward-modeling problem, whatever the size of the input data and the number of unknown velocity parameters. An application on a 2D synthetic data set demonstrated the ability of the algorithm to image near-surface velocities with strong vertical and lateral variations and revealed the potential of the method.

INTRODUCTION

First-arrival traveltime tomography can be used to image the earth's interior at various scales, from near-surface to global scale using passive sources. For seismic imaging, determination of the near-surface velocity structure is a key step when trying to image deeper structures. In most cases, refraction tomography — or more generally, first-arrival traveltime tomography based on direct, diffracted, refracted, or diving waves — is used to assess a velocity model of the subsurface that best explains the data (Zhu et al., 1992; Zelt et al., 2006). For many years, such algorithms have proved to be valuable tools for obtaining the shallow structure needed for static corrections in seismic reflection-data processing. Nowadays, first-arrival traveltime tomography can be used to obtain the initial near-surface velocity structure for prestack depth migration and even waveform inversion of wide-angle data (Dessa et al., 2004; Brenders and Pratt, 2007).

Like all traveltime-inversion algorithms, first-arrival tomography needs an efficient and accurate method to calculate traveltimes, and many inversion schemes require computation of Fréchet derivatives. Among a wide variety of algorithms published in the literature, a well-suited forward-modeling algorithm for first-arrival traveltime tomography is the first-order finite-difference eikonal solver (for example, Podvin and Lecomte, 1991, or Hole and Zelt, 1995) that is very fast and supports strong velocity contrasts. The eikonal solver is probably the most commonly used algorithm, but one could use alternative methods such as the wavefront construction method of Vinje et al. (1993). Many refraction-tomography methods combine an eikonal solver with a posteriori ray tracing to compute the Fréchet derivative matrix, and an algorithm such as LSQR (Paige and Saunders, 1982) to solve the linearized tomographic system (Zelt and Barton, 1998; Le Meur, 1994; Baina, 1998; Dessa et al., 2004) iteratively.

Commonly, current seismic acquisition surveys now deploy thousands of sources combined with thousands of receivers, leading to millions or even more acquired traces (Vesnaver, 2008). Moreover, the area under investigation could be very large, leading to a velocity model containing millions of parameters, whatever the type of parametrization. In addition, the properties of the near surface can exhibit strong variations vertically, laterally, and with very short wavelengths, requiring a velocity model with very small grids. For 2D geometries, classical first-arrival tomography algorithms might not suffer from computational limitations. On the other hand, for 3D geometries, the Fréchet derivative matrix could turn out to be very difficult to handle in terms of memory requirements, even if sparsity is accounted for. To overcome this limitation, one could reduce the size of the data to be inverted or the number of parameters of the model, both leading to either a loss of information or to a poorer resolution.

The classical method that includes computation of Fréchet derivatives formulates the minimization of the misfit function as a sequence of linear problems. Here, we formulate the tomographic inversion directly as a nonlinear optimization problem. The key step is

Manuscript received by the Editor 30 December 2008; revised manuscript received 16 June 2009; published online 8 December 2009; corrected version published online 17 December 2009.
[1]Formerly Mines ParisTech, Fontainebleau, France; presently CGGVeritas, Massy, France. E-mail: cedric.taillandier@cggveritas.com.
[2]Mines ParisTech, Centre de Géosciences, Fontainebleau, France. E-mail: mark.noble@mines-paristech.fr; herve.chauris@mines-paristech.fr.
[3]Total E & P, Pau, France. E-mail: henri.calandra@total.com.

DOI:10.1190/1.3250266

determination of the gradient of the misfit function, obtained here by the adjoint-state method. Once the gradient is derived, it can be coupled, for example, to the steepest descent method, the conjugate-gradient method, or the quasi-Newton method.

Computation of the gradient of the misfit function with the adjoint-state method has one main advantage: its cost is equivalent to twice the solution of the forward-modeling problem, no matter the size of the input data. To compute the gradient, we need forward plus backward modeling, both with a similar forward-modeling computational burden. This method was developed originally in the control-theory framework (Lions, 1971). Chavent (1974) introduces it to resolve inverse problems. It has been used for a series of geophysical applications such as seismic migration and waveform inversion (Lailly, 1983; Tarantola, 1984), velocity model building (Sei and Symes, 1994; Chavent and Jacewitz, 1995; Plessix et al., 2000; Mulder and ten Kroode, 2002; Shen et al., 2003), including traveltime tomography (Delprat-Jannaud et al., 1992; Shen et al., 2003; Leung and Qian, 2006; Mulder, 2006). For a more detailed review, especially for geophysical applications, we refer to Plessix (2006). To our knowledge, Sei and Symes (1994) are the first to derive the adjoint-state equations for the eikonal solver, and Leung and Qian (2006) apply the method to traveltime transmission tomography data.

Here we describe implementation of a new nonlinear first-arrival traveltime tomography algorithm based on the original work of Sei and Symes (1994) and Leung and Qian (2006). The combination of this approach with an efficient parallel implementation with a large number of traveltime picks is demonstrated on a realistic-sized model in the context of seismic traveltime tomography.

First we will present the governing equations of the adjoint-state method that enable us to calculate the gradient of the misfit function. Then the practical implementation of our first-arrival tomography algorithm is shown on a very simple model. All mathematical developments and numerical examples are developed in two dimensions and constitute a relevant feasibility study for the application of the method to 3D geometries. Finally, validation results obtained on a 2D synthetic velocity model reveal the potential of the method.

THE MISFIT FUNCTION AND ITS GRADIENT

We define the misfit function used to estimate the quality of a given velocity model and then derive its gradient with the adjoint-state technique. All equations are written for the continuous case. For a single shot, the classical misfit function is

$$J(\mathbf{c}) = \frac{1}{2}\int_{\partial\Omega} \mathrm{d}r |T(\mathbf{c},r) - T_{\mathrm{obs}}(r)|^2, \tag{1}$$

for each receiver r at the acquisition surface $\partial\Omega$, Ω being the subsurface. Observed first-arrival traveltimes T_{obs} picked in the shot gathers are compared to calculated traveltimes T obtained by solving the eikonal equation in velocity model $\mathbf{c}$. This eikonal solution indicates that the norm of the spatial gradient of the traveltime is equal to the square of the slowness and is written for any point x in subsurface Ω as

$$|\nabla T(x)|^2 = \frac{1}{\mathbf{c}^2(x)}, \tag{2}$$

and

$$T(s) = 0. \tag{3}$$

At source position s, traveltime is equal to zero. Equation 2 is valid under the high-frequency approximation for an acoustic, isotropic, heterogeneous velocity model (Cerveny, 2001; Chapman, 2004). From a physical point of view, it means that spatial wavelength of the signals is supposed to be much shorter than the characteristic wavelength of the velocity model. In the first-arrival traveltime tomography approach, the amplitude term that would be provided by the transport equation is not used. The misfit function for a full survey consists simply of summing contributions from all shots. Note that traveltime T is not quadratic in $1/\mathbf{c}$ because the raypaths change with the velocity model.

The classical approach consists of linearizing T with respect to $\mathbf{c}$ around a given velocity model $\mathbf{c}_n$. The optimal velocity $\mathbf{c}_{n+1}$ satisfies

$$T_{\mathrm{obs}} = T(\mathbf{c}_n) + \mathbf{G}(\mathbf{c}_n)(\mathbf{c}_{n+1} - \mathbf{c}_n), \tag{4}$$

where $\mathbf{G}$ is the matrix of the Fréchet derivatives. The size of $\mathbf{G}$ is the number of observed traveltimes multiplied by the number of model parameters. The gradient of J is $\nabla J = -\mathbf{G}^t(T_{\mathrm{obs}} - T(\mathbf{c}_n))$ and the associated Hessian $\mathbf{H} = \mathbf{G}^t\mathbf{G}$. In the Newton approach, the minimization problem is solved as a sequence of linear problems where formally, $\mathbf{c}_{n+1} = \mathbf{c}_n - \mathbf{H}^{-1}\nabla J$.

If we formulate the problem as a nonlinear problem using a simple gradient method, we have

$$\mathbf{c}_{n+1} = \mathbf{c}_n - \alpha_n \nabla J(\mathbf{c}_n), \tag{5}$$

where α is a positive scalar. This approach does not require solving a linear system involving $\mathbf{G}$ (equation 4). The adjoint-state method is a general technique for computing the gradient of a misfit function, without the need for introducing Fréchet derivatives. The two tomography implementations a priori have the same properties. In our case, the state variable is traveltime T, the solution of the eikonal equation providing the link between T and the unknown velocity model $\mathbf{c}$. The adjoint method is introduced here through the Lagrangian formulation. In this case, we define three independent variables: t, $\mathbf{c}$, and the adjoint state λ. The extended misfit function becomes

$$L(\mathbf{c},t,\lambda) = \frac{1}{2}\int_{\partial\Omega} \mathrm{d}r |t(r) - T_{\mathrm{obs}}(r)|^2 - \frac{1}{2}\int_{\Omega} \mathrm{d}x \lambda(x)\left(|\nabla t(x)|^2 - \frac{1}{\mathbf{c}(x)^2}\right). \tag{6}$$

Variable λ can be viewed as a penalty term to ensure that at the optimum, t is the solution to the eikonal equation and thus equal to T. During the minimization process, t and λ are independent of $\mathbf{c}$ until the optimum is reached. The gradient of the original misfit function $\partial J/\partial\mathbf{c}$ is obtained by solving the following system:

$$\frac{\partial L}{\partial \mathbf{c}} = -\int_{\Omega} \mathrm{d}x \frac{\lambda(x)}{\mathbf{c}^3(x)} = \frac{\partial J}{\partial \mathbf{c}}, \tag{7}$$

$$\frac{\partial L}{\partial \lambda} = 0, \tag{8}$$

and

$$\frac{\partial L}{\partial t} = 0. \tag{9}$$

When all equations are satisfied, J and L have the same expression (equation 7). With equation 8, we obtain the eikonal equation. Finally, equation 9 provides the way to compute λ. By definition of L, we have

$$\frac{\partial L}{\partial t} = \int_{\partial\Omega} \mathrm{d}r(t(r) - T_{\mathrm{obs}}(r)) - \int_{\Omega} \mathrm{d}x\lambda(x)\nabla t(x)\cdot\frac{\partial \nabla t}{\partial t}. \tag{10}$$

The second part of the equation is integrated by parts, yielding

$$\frac{\partial L}{\partial t} = \int_{\partial\Omega} \mathrm{d}r(t(r) - T_{\mathrm{obs}}(r) - \lambda\mathbf{n}\cdot\nabla t) + \int_{\Omega} \mathrm{d}x\nabla\cdot\lambda(x)\nabla T(x), \tag{11}$$

where $\mathbf{n}$ is the unit vector normal to surface $\partial\Omega$. Among several possibilities, we impose that the two terms in equation 11 are null. On the surface, λ is the solution of

$$\Lambda(r)\mathbf{n}(r)\cdot\nabla T(r) = T(r) - T_{\mathrm{obs}}(r), \tag{12}$$

and within the subsurface, λ is the solution of

$$\nabla\cdot\Lambda(x)\nabla T(x) = 0. \tag{13}$$

Variables t and λ are equal to T and Λ when equations 8 and 9 are satisfied. In practice, one needs to solve the direct problem to obtain residuals $t - T_{\mathrm{obs}}$. The adjoint variable is initialized at the surface as being the residuals, divided by the normal component of the ray vector. Then, λ is back-propagated in the current velocity model to the source. The gradient of the misfit function finally is obtained according to equation 7 where λ is normalized by velocity values $\mathbf{c}^3$. More precisely, if we consider a ray tube limited in 2D by two rays, time T, velocity $\mathbf{c}$, diameter S of the ray tube, and adjoint-state variable Λ can be parametrized by s, the arc length along the ray (Figure 1). We apply the divergence theorem within the ray tube between s and $s + ds$, changing from a volume integral to a surface integral. The contribution perpendicular to the rays is null in isotropic models. Knowing that the norm of the gradient of T is given by the eikonal equation, equation 13 simply means that quantity $Q = \Lambda(s)/\mathbf{c}(s)dS(s)$ is constant along the ray tube. With the initial conditions provided by equation 12, the adjoint-state system can be read as back-propagation of residuals along the ray tube to the source position. Next, we present the main practical aspects for the estimation of the misfit function and its gradient.

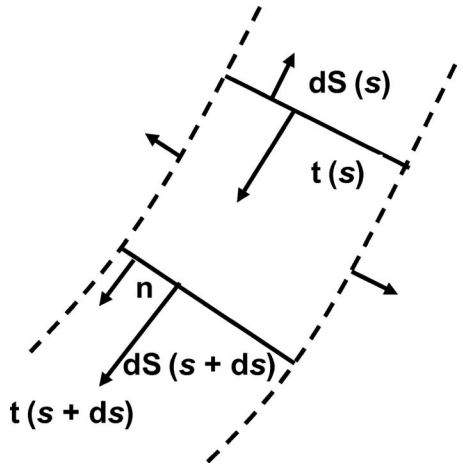

Figure 1. Representation of a ray tube defined by two rays (dotted lines). The variable S is the arc along the ray.

PRACTICAL IMPLEMENTATION

We use a classical steepest-descent gradient-based approach for minimization of the misfit function. It consists mainly of iteratively computing the value of the misfit function and its associated gradient. The gradient provides the correct direction in the model space, and the length of the update is solved for by determining a multiplicative scalar or step length such that the misfit function is minimized. To compute the step length, we used the method proposed by Pica et al. (1990). It should be mentioned that a steepest-descent method has a linear convergence rate: One could expect better convergence with an optimizer such as a nonlinear conjugate gradient or a quasi-Newton method.

Local and global scheme for eikonal and adjoint-state system

The same strategy is used to solve the eikonal system (equations 2 and 3) and the adjoint-state system (equations 12 and 13). It consists of a combination of two schemes: first, for a given point in the medium, traveltimes and λ are predicted locally from neighboring points (local scheme) and then a global scheme is applied to propagate traveltimes or λ across the whole medium.

For the local scheme applied to the eikonal equation, we refer to Podvin and Lecomte (1991) and Zhang et al. (2005). This approach is based on Huygens' principle (Figure 2). Knowing the traveltimes at points A_1 and A_2, three different traveltimes are predicted for point B by considering a diffraction source in A_1, a plane wave propagating between A_1 and A_2, or a head wave coming from A_2. In complex velocity models, waves can propagate from any direction to point B.

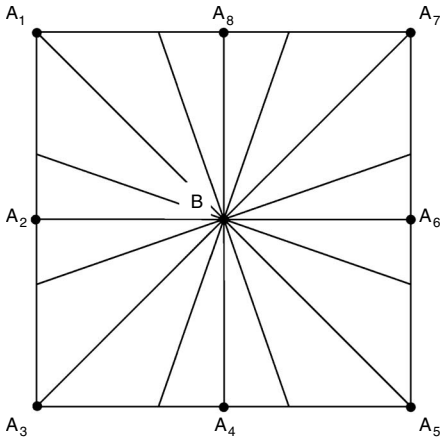

Figure 2. Local scheme for the determination of the first-arrival traveltime, after Zhang et al. (2005). This approach (Podvin and Lecomte, 1991) is based on Huygens' principle. For example, knowing the traveltimes at points A_1 and A_2, three different traveltimes are predicted for point B by considering a diffraction source in A_1, a plane wave propagating between A_1 and A_2 or a head wave coming from A_2. This operation is performed for all other A points, leading to the computation of 16 possibly different traveltimes in B. Only the minimum time is retained.

As we can see in Figure 2, there are 16 possible wave propagations to the middle point of interest. Only the minimum time is retained. The local scheme used to solve the adjoint-state equation is based on the same principle as the eikonal scheme. We used the same approach as Leung and Qian (2006). For completeness, the scheme is described in Appendix A.

For the global scheme, we adopt the fast-sweeping method as proposed by Zhao (2005) and Kuster (2006). The global scheme implementation is exactly the same for the eikonal and the adjoint-state system. For the eikonal system of equations, initially we assign traveltime equal to zero at the source location and very large values at all other grid points. Equivalently, the adjoint state λ is initialized according to equation 12 on the receiver locations and with very large values at all other grid points. Then, the fast-sweeping method proceeds by updating values of t or λ for all the grid points in a certain order: it conducts sweeps in the positive and negative x and positive and negative z directions, each time using the local scheme described above.

Gradient calculation for a canonical case

We illustrate the key elements to calculate the gradient of the misfit function for one shot gather in a very simple model, consisting of a constant vertical-gradient velocity model where a local positive velocity anomaly of 200 m/s is added (Figure 3). Topography is modeled by a sinusoidal function and the velocity above is equal to 330 m/s. The model is 2 km in depth and 7.5 km laterally and is discretized on a grid of 12.5 by 12.5 m. The acquisition consists of 600 sources, and there are 600 receivers for each source. Sources and receivers are spread uniformly along the topography of the model, with a 12.5-m distance between sources or receivers.

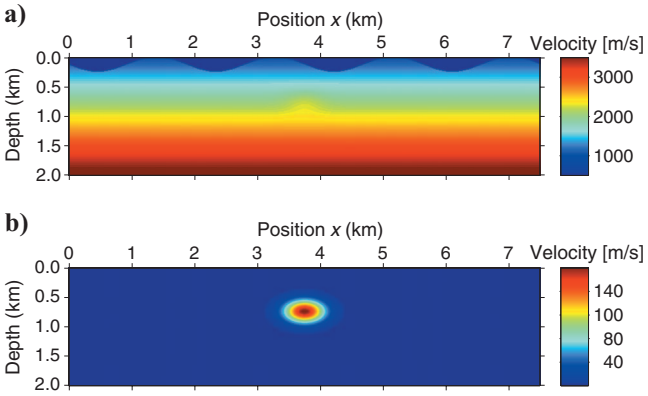

Figure 3. Velocity model used to illustrate computation of the gradient of the misfit function. (a) Exact velocity model, and (b) perturbation to be retrieved.

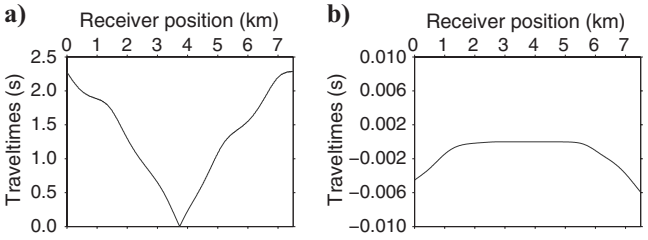

Figure 4. (a) Observed traveltimes calculated in exact velocity model (Figure 3a) for a source placed on the topography at $x = 3.75$ km and (b) traveltime residuals for the same source.

The initial model is the same as the exact model, except that it does not contain the local velocity anomaly. For a source in the middle of the model placed on the topography at $x = 3.75$ km, first-arrival traveltimes are obtained by solving equations 2 and 3 in the true and initial models to get observed traveltimes and residuals (Figure 4). To compute the gradient of the misfit function, first we need to initialize λ at the surface (Figure 5a and proceed with sweeping in four directions (Figure 5). For the first sweep (Figure 5b), the sweep is from left to right and top to bottom. It means we can only retrieve information when the ray also goes from left to right and top to bottom. At the point where the ray is turning back to the surface, λ is not updated at that time. A single iteration (four sweeps) is enough to converge in this case. For more complicated models where rays oscillate, more iterations are needed until λ does not change. The gradient of the misfit function for a single source (Figure 6a) is obtained by dividing λ by the velocity at the third power according to equation 7. After summing the contribution of all sources, we start to retrieve the negative velocity anomaly (Figure 6b). The asymmetry is be-

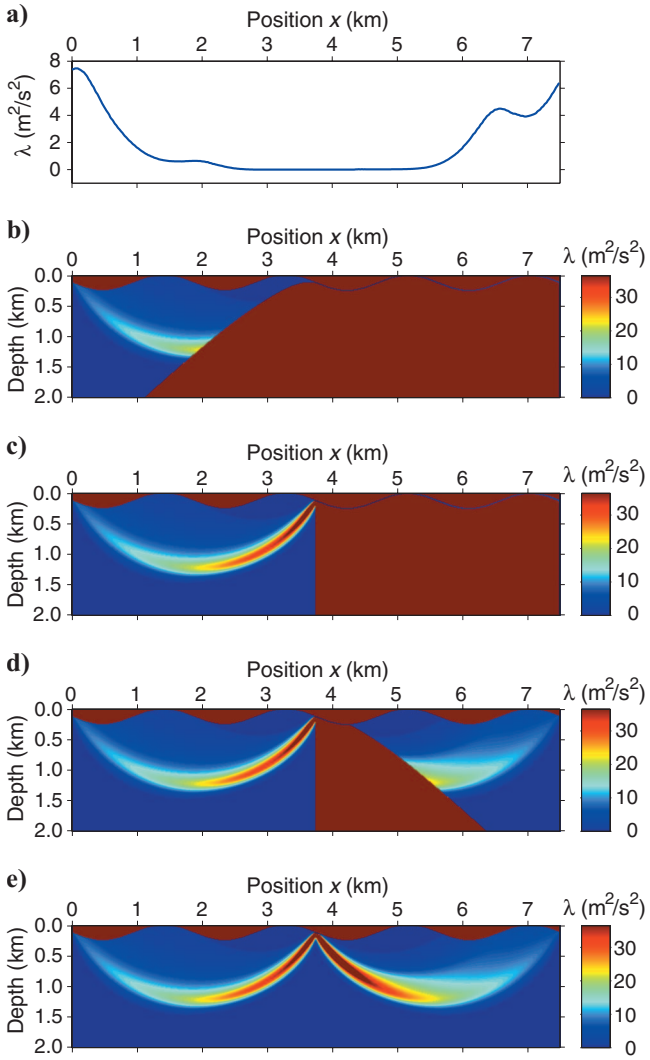

Figure 5. (a) Initial values for λ at the surface, and (b-e) after the four successive sweeps.

cause of the topography of the model. We present only computation of the gradient for the first iteration in the minimization process.

Finally, to validate our algorithm, the gradient calculated with the adjoint-state method is compared to a gradient obtained by a finite-difference approach. According to Figure 6c, the two results are almost identical.

Regularization and choice of starting model

The only regularization we use during the inversion is a low-pass filter applied on the gradient of the misfit function at each iteration. The fact that our algorithm does not use rays explicitly does not exempt us from any low-pass filtering. Even though the gradient for one single source-receiver pair is slightly fatter than a ray (Figure 7), the adjoint-state method has the same behavior as any classical tomography algorithm based on ray tracing, and thus requires regularization or, more precisely, filtering. Correct choice of filter length is crucial to prevent the algorithm from getting trapped in a local minimum. In practice, we do the inversion in two steps. We start by applying a very strong Gaussian filter on the gradient at each iteration. The length of the filter is approximately equal to the maximum source-receiver offset horizontally, and half this distance vertically. After convergence, we do another inversion, this time with a very small filter whose length is equal to one tenth of the maximum offset. This empirical rule concerning the length of our regularization filter seems to be fairly robust for seismic exploration data with a dense acquisition.

Another important and well-known issue is the choice of the initial velocity model. If the initial model is faster than the expected velocities we are inverting for and it contains a vertical velocity gradient that increases with depth, we can start our inversion with almost any velocity model and converge to the correct answer. This is because we are using strong regularization (a very large filter) at the beginning of the inversion. With an initial velocity model that is too slow, our algorithm might get trapped in a local minimum of the misfit function and give an unrealistic model. In the very first iterations, the inversion scheme will create a strong refractor at a depth that is too shallow. Once this refractor is in place, and despite a strong regularization, no rays are able to pass through the refractor and thus to update the velocity model correctly.

Our conclusions concerning the choice of the initial velocity model are probably true only for the case of seismic refraction tomography. Certainly our algorithm will behave differently for crosshole data for example, where the initial velocity model is not required to contain any velocity gradient.

Memory requirements and parallelism

The first benefit of implementing the gradient method is a low memory requirement. The memory required is independent of the amount of input traveltimes we want to invert for, and it depends only on the size of the velocity model. The second benefit of such an implementation is intrinsically a parallel algorithm. For each shot, one needs to compute the value of the misfit function and its gradient. These computations are totally independent from one shot to another, and they can be distributed easily across the different processors of a cluster. We have parallelized our code using the message-passing-interface (MPI) library. In practice, each processor handles one shot at a time and only needs to hold the part of the current velocity model and λ field that is covered by the acquisition geometry for the particular shot in memory. The total gradient is simply the sum of all source contributions; thus there is a limited exchange of information between nodes. Gain in CPU time will vary depending on the architecture of the cluster, the number of processors, and the amount of data to process. For the example described in the next section, using a 100-processor cluster, the inversion runs 80 times faster than on a single processor. The main communications between processors occur at the end of the computation of the gradient, where all partial gradients are summed together on the master processor and sent back to the slaves for the step-length calculation.

2D SYNTHETIC-DATA EXAMPLE

The first-arrival tomographic inversion based on the adjoint-state technique is tested on a realistic-size acquisition using a synthetic model. This model was created at the Amoco Tulsa Research Lab in 1994 by Mike O'Brien as part of a project to study methods for attacking statics in land data. It contains many different types of near-

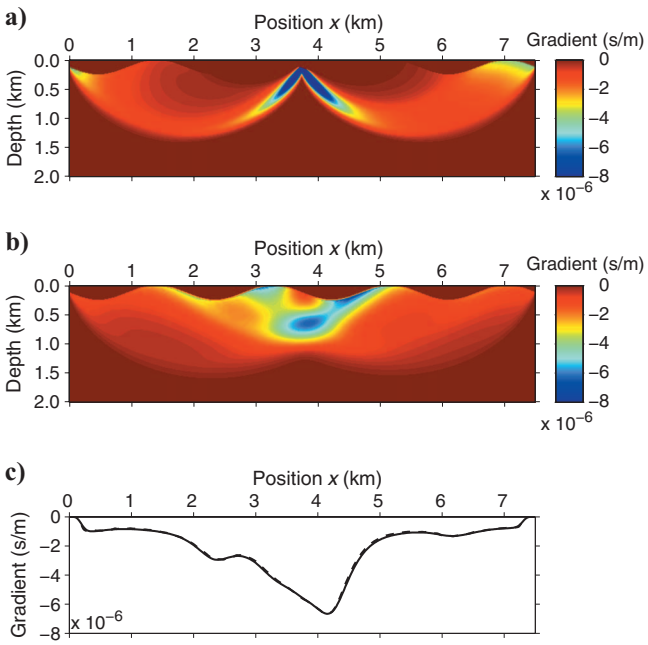

Figure 6. (a) Gradient of the misfit function for a single source, (b) for all sources, (c) comparison of the gradient for all shots obtained by the adjoint-state method with a gradient computed by finite differences, at a depth of $z = 500$ m. Adjoint state in solid line and finite difference in dashed line.

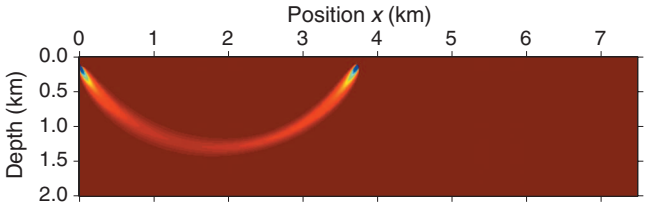

Figure 7. Gradient for only one source-receiver couple.

DOI:10.1190/1.3250266

surface geology, generally representing geology thought to be responsible for statics. Figure 8d shows the first 3.5-km depth section of this model. The entire profile has a length of 50 km and a depth of 6 km and is characterized by large variations in both velocity and topography. Observed first-arrival traveltimes from this model were simulated with an eikonal solver with 2000 sources located on the topography. Each recorded up to 1500 receivers also placed on the topography on either side of the source. Source and receiver intervals are 25 and 5 m, respectively. The nearest offset is 5 m and maximum offset is 7500 m. This geometry leads to approximately 5.5 million first arrivals for the near-surface velocity estimation.

As an initial model, we took a 1D velocity field with a constant vertical gradient, on which we set all velocity values to 330 m/s above the topography. This model is displayed in Figure 8a. Above the topography, values are held constant during the inversion. The velocity model defined as a grid consists of $1200 \times 10{,}000$ cells with a grid spacing of 5 m in both the vertical and horizontal directions. The first inversion ran with a strong filter length: 7500 m in the horizontal direction and 3000 m in the vertical direction (shown in Figure 8b). After a second run with a filter length of 600 m in both directions, the final velocity model is shown in Figure 8c. Having limited offset and only inverting for first-arrival traveltimes, we retrieve only the near-surface velocity structure and not the entire model.

Comparison of Figure 8c and d shows close agreement between the inverted and true velocity models, at least for the near surface. Even some small details in the recovered model, such as the high-velocity anomaly close to the topography at $x = 10.5$ km, indicates that our method is capable of recovering small velocity structures in complex areas. Figure 9 shows several velocity logs for the initial (medium gray), inverted (light gray) and true (black) velocity models at different locations: 10.5, 18, 24, 30, 36, and 42 km (Figure 9a-f, respectively). There is a good agreement between the true model and inverted model down to approximately 500 to 700 m below the acquisition surface, depending on location and ray coverage. On the first log at 10.5 km, the small high-velocity anomaly close to the topography (also visible on Figure 8) is almost recovered. Figure 10 shows velocity profiles for the initial (medium gray), final inverted (light gray) and true (black) velocity models at constant depths below the topography: 40, 60, 100, 200, and 400 m for Figure 10a-e, respectively. There is good agreement between the true and inverted model through the whole profile except on the edges where ray coverage is poor and becomes worse with increasing depth. This figure

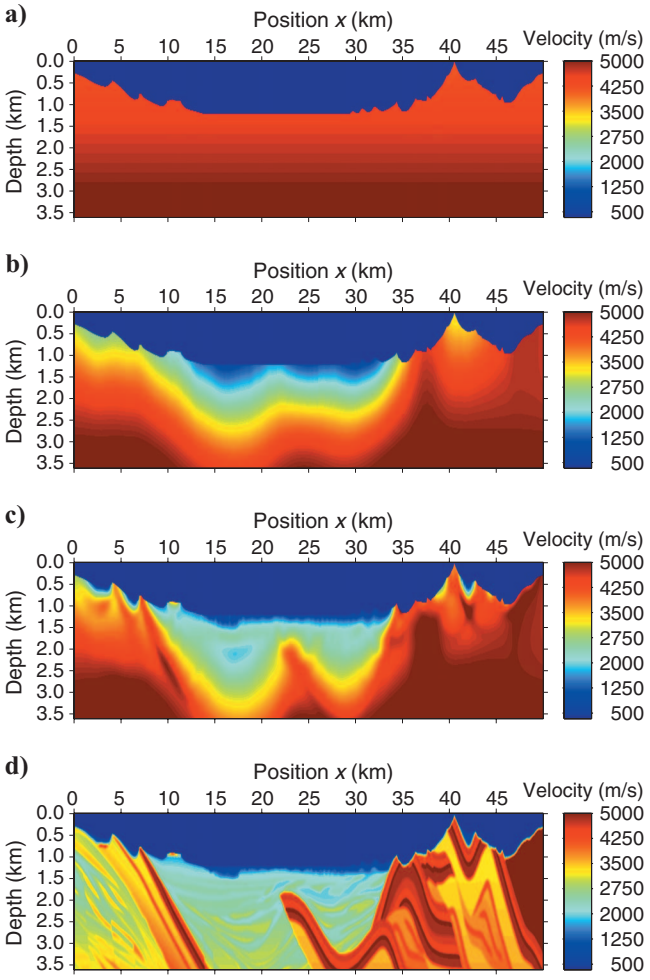

Figure 8. (a) 2D synthetic example, initial, (b) first inversion result with strong filter, (c) final inverted, and (d) true velocity models.

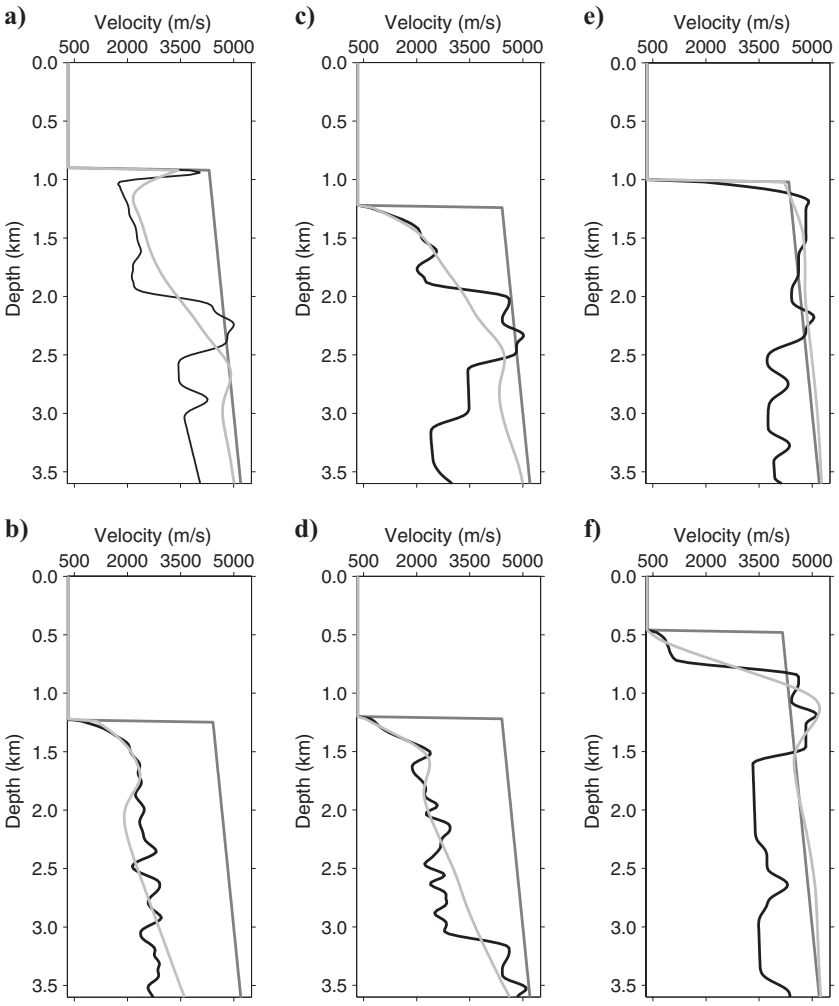

Figure 9. Velocity logs for the initial (medium gray), inverted (light gray) and true (black) velocity models at different locations: (a) 10.5, (b) 18, (c) 24, (d) 30, (e) 36, and (f) 42 km, respectively. The first profile at 10.5 km shows a small strong positive velocity anomaly close to the topography that our inversion has recovered partially.

indicates that our tomographic method is quite flexible in accommodating strong lateral variations. Finally, the initial and final residuals for one shot are plotted in Figure 11. This inversion required 60 iterations to converge, and it took approximately 20 minutes using a cluster with 100 processors. The same inversion using just one processor would take more than 24 hours.

DISCUSSION

We have developed a first-arrival traveltime tomography algorithm based on a steepest-descent gradient approach using the adjoint-state technique for computation of the gradient of the misfit function. The main goal in developing this algorithm is to handle very large data sets. This inversion scheme presented is promising from a computation point of view because the memory needed is only a few times the memory size of the velocity model. In addition, the algorithm is parallel by nature, adapted to supercomputers, and it confirms the great potential for 3D applications.

Here we have not addressed the difficult issue of picking the traveltimes. Our algorithm behaves the same way as any other existing method and is dependent on the quality of the picks.

Like all optimization problems that need to handle large data volumes, the logical solution was to choose a gradient approach (c.f., Tarantola, 1987, and others). Compared to classical tomography methods, our method probably requires more iterations. This overhead is compensated largely by the fact that our algorithm is fully parallel. Probably we could gain quite a few iterations by using a limited-memory quasi-Newton method, which provides an estimate of the Hessian using a few gradients from previous iterations.

One of the most interesting features of our algorithm is its CPU time efficiency that enables us to do many runs and try different regularization (filtering) parameters to ensure optimal convergence. Like all standard tomography algorithms, we have made the assumption that the misfit function is quadratic, but it is not. To overcome the nonlinearity of the problem, we start the inversion by applying a large filter in both directions (vertical and horizontal) first to retrieve the long wavelength of the velocity model. Then we reduce the filter length to get the shorter wavelengths. Although we have found a few empirical rules (mentioned earlier), defining the appropriate length of the filters requires a few trial-and-error runs that we can afford with our algorithm.

This algorithm is not limited to refraction tomography; it could be used with no modification for walkaway or crosshole acquisitions. At present, using such an algorithm for seismic reflection tomography is not straightforward.

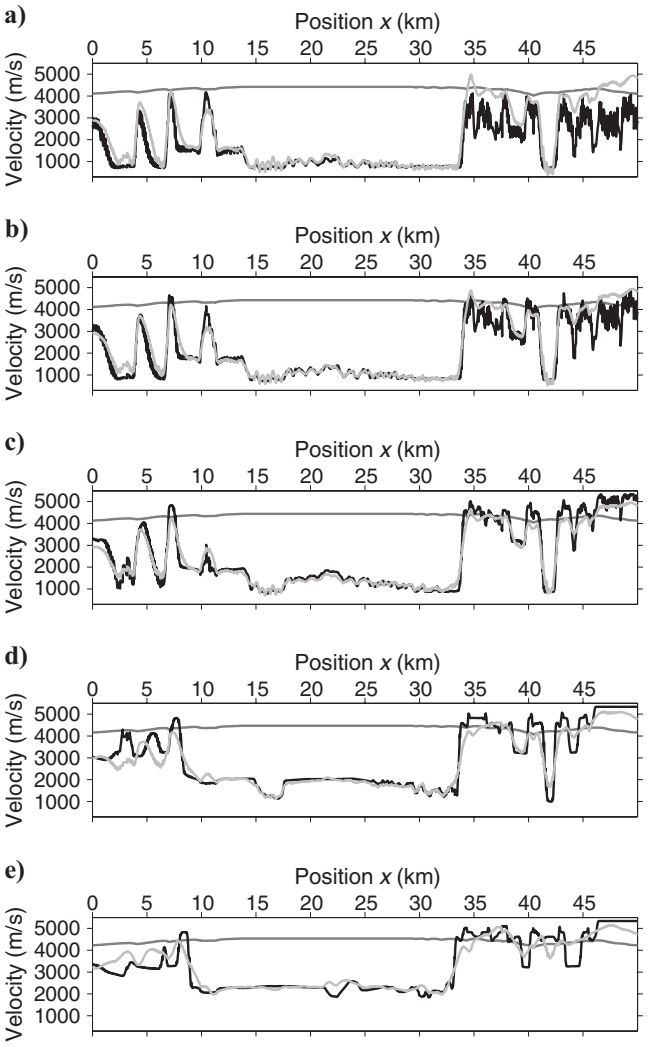

Figure 10. Velocity profiles for the initial (medium gray), inverted (light gray) and true (black) velocity models at constant depths bellow the topography: (a) 40, (b) 60, (c) 100, (d) 200, and (e) 400 m, respectively.

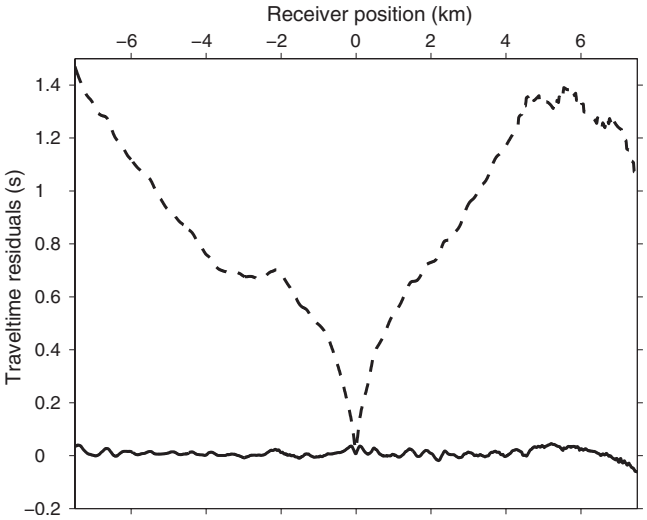

Figure 11. Initial residuals for one shot gather (dashed line). The source is at $x = 25$ km. Final residuals (solid line).

CONCLUSIONS

We have shown how the combination of a gradient-based approach with the adjoint-state technique for computation of the gradient of the misfit function provides the key elements for an efficient first-arrival traveltime tomography algorithm. The main computational advantage of this method is the memory requirement, which depends only on the size of the velocity model, not the amount of input traveltimes we want to invert for. The algorithm can handle a data set of any size easily. The steepest-descent method that we implemented has a linear convergence rate and requires many iterations to minimize the misfit function. We could improve the convergence rate by using an optimizer such as a nonlinear conjugate gradient or a quasi-Newton method. This feasibility test confirms the potential of the method for 3D first-arrival tomography.

ACKNOWLEDGMENTS

The authors would like to thank Total E&P for partially funding the project. The synthetic velocity model used in the application section was created by Mike O'Brien and Carl Regone and is provided courtesy of Amoco and BP. It can be found at http://software.seg.org.

APPENDIX A

LOCAL AND GLOBAL SCHEME FOR THE ADJOINT STATE

We provide the main equations of the fast-sweeping method (Zhao, 2005) to solve the adjoint-state equations 12 and 13 (Leung and Qian, 2006).

Local scheme

In 2D, equation 13 can be rewritten as

$$\frac{\partial}{\partial x}(a\lambda) + \frac{\partial}{\partial z}(b\lambda) = 0, \tag{A-1}$$

where $a = \partial t(x,z)/\partial x$ and $b = \partial t(x,z)/\partial z$ are the two components of the gradient of t. For a rectangular gridding of the subsurface with cells of size Δx by Δz, $\lambda_{j,i}$ denotes the adjoint-state variable at the position (z_j, x_i) in the subsurface. Values of a and b are specified at intermediate positions $(z_{j\pm 1/2}, x_i)$ and $(z_j, x_{i\pm 1/2})$, leading to

$$a_{j,i-\frac{1}{2}} = \frac{t_{j,i} - t_{j,i-1}}{\Delta x}, \quad a_{j,i+\frac{1}{2}} = \frac{t_{j,i+1} - t_{j,i}}{\Delta x} \tag{A-2}$$

and

$$b_{j-\frac{1}{2},i} = \frac{t_{j,i} - t_{j-1,i}}{\Delta z}, \quad b_{j+\frac{1}{2},i} = \frac{t_{j+1,i} - t_{j,i}}{\Delta z}. \tag{A-3}$$

We introduce the following notations:

$$a^{\pm}_{j,i+\frac{1}{2}} = \frac{a_{j,i+\frac{1}{2}} \pm \left| a_{j,i+\frac{1}{2}} \right|}{2}, \tag{A-4}$$

$$a^{\pm}_{j,i-\frac{1}{2}} = \frac{a_{j,i-\frac{1}{2}} \pm \left| a_{j,i-\frac{1}{2}} \right|}{2}, \tag{A-5}$$

$$b^{\pm}_{j+\frac{1}{2},i} = \frac{b_{j+\frac{1}{2},i} \pm \left| b_{j+\frac{1}{2},i} \right|}{2}, \tag{A-6}$$

and

$$b^{\pm}_{j-\frac{1}{2},i} = \frac{b_{j-\frac{1}{2},i} \pm \left| b_{j-\frac{1}{2},i} \right|}{2}. \tag{A-7}$$

Equation A-1 becomes

$$\frac{1}{\Delta x}\left[\left(a^{-}_{j,i+\frac{1}{2}}\lambda_{j,i} + a^{+}_{j,i+\frac{1}{2}}\lambda_{j,i+1}\right) - \left(a^{-}_{j,i-\frac{1}{2}}\lambda_{j,i-1} + a^{+}_{j,i-\frac{1}{2}}\lambda_{j,i}\right)\right] + \frac{1}{\Delta z}\left[\left(b^{-}_{j+\frac{1}{2},i}\lambda_{j,i} + b^{+}_{j+\frac{1}{2},i}\lambda_{j+1,i}\right) - \left(b^{-}_{j-\frac{1}{2},i}\lambda_{j-1,i} + b^{+}_{j-\frac{1}{2},i}\lambda_{j,i}\right)\right] = 0, \tag{A-8}$$

or equivalently,

$$\left(\frac{a^{-}_{j,i+\frac{1}{2}} - a^{+}_{j,i-\frac{1}{2}}}{\Delta x} + \frac{b^{-}_{j+\frac{1}{2},i} - b^{+}_{j-\frac{1}{2},i}}{\Delta z}\right)\lambda_{j,i} = \frac{a^{-}_{j,i-\frac{1}{2}}\lambda_{j,i-1} - a^{+}_{j,i+\frac{1}{2}}\lambda_{j,i+1}}{\Delta x} + \frac{b^{-}_{j-\frac{1}{2},i}\lambda_{j-1,i} - b^{+}_{j+\frac{1}{2},i}\lambda_{j+1,i}}{\Delta z}. \tag{A-9}$$

Equation A-9 is the way to express $\lambda_{j,i}$ as a function of its neighboring values $\lambda_{j,i\pm1}$ and $\lambda_{j\pm1,i}$. It is used to build the iterative scheme of the fast-sweeping method.

Global scheme: Fast-sweeping method

The fast-sweeping method to compute λ over the whole model is an iterative process. It is stopped only when convergence is reached. Upper index n of $\lambda^{(n)}$ represents the iteration index.

First of all, after initializing $\lambda^{(0)}$ at the surface according to equation 12, $\lambda^{(0)}$ is set to any value larger than the maximum expected traveltime residual. At each iteration and for each grid point, a new $\lambda^{e}_{j,i}$ is computed according to equation A-9 and the new $\lambda^{(n)}$ is the minimum between $\lambda^{(n-1)}$ and $\lambda^{e}_{j,i}$. The algorithm can be summarized as:

1) Initialization of $\lambda^{(0)}$
2) $n = 1$
3) $\tilde{\lambda} = \lambda^{(n-1)}$
4) Sweep from $i = 1{:}I$, then from $j = 1{:}J$, computation of $\lambda^{e}_{j,i}$, $\tilde{\lambda}_{j,i} = \max(\tilde{\lambda}_{j,i}, \lambda^{e}_{j,i})$
5) Sweep from $i = 1{:}I$, then from $j = J{:}1$, computation of $\lambda^{e}_{j,i}$, $\tilde{\lambda}_{j,i} = \max(\tilde{\lambda}_{j,i}, \lambda^{e}_{j,i})$
6) Sweep from $i = I{:}1$, then from $j = 1{:}J$, computation of $\lambda^{e}_{j,i}$, $\tilde{\lambda}_{j,i} = \max(\tilde{\lambda}_{j,i}, \lambda^{e}_{j,i})$
7) Sweep from $i = I{:}1$, then from $j = J{:}1$, computation of $\lambda^{e}_{j,i}$, $\tilde{\lambda}_{j,i} = \max(\tilde{\lambda}_{j,i}, \lambda^{e}_{j,i})$
8) $\lambda^{(n)} = \tilde{\lambda}_{j,i}$
9) If convergence ($\lambda^{(n)} = \lambda^{(n-1)}$) stop; otherwise $n = n + 1$ and go to 3

Generally, a single iteration is enough to converge. However, for complex media where, for example, rays are oscillating as they do with tube waves, more iterations are needed.

REFERENCES

Baina, R., 1998, Tomographie sismique entre puits: Mise en œuvre et rôle de l'analyse a posteriori; vers une prise en compte de la bande passante: Ph.D. thesis, Université de Rennes I (in French).

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.

Cerveny, V., 2001, Seismic ray theory: Cambridge University Press.

Chapman, C., 2004, Fundamentals of seismic wave propagation: Cambridge University Press.

Chavent, G., 1974, Identification of functional parameters in partial differential equations, *in* R. E. Goodson and M. Polis, eds., Identification of parameters in distributed systems: ASME, 31–48.

Chavent, G., and C. Jacewitz, 1995, Determination of background velocities by multiple migration fitting: Geophysics, **60**, 476–490.

Delprat-Jannaud, F., P. Lailly, E. Becache, and E. Chovet, 1992, Reflection tomography: How to cope with multiple arrivals?: 62nd Annual International Meeting, SEG, Expanded Abstracts, 741–744.

Dessa, J.-X., S. K. 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.

Hole, J., and B. Zelt, 1995, 3-D finite-difference reflection traveltimes: Geophysical Journal International, **121**, 427–434.

Kuster, C. M., 2006, Fast numerical methods for evolving interfaces: Ph.D. thesis, North Carolina State University.

Lailly, P., 1983, The seismic inverse problem as a sequence of before-stack migrations, *in* J. B. Bednar, R. Redner, E. Robinson, and A. Weglein, eds., Conference on inverse scattering: Theory and application: SIAM, 206–220.

Le Meur, H., 1994, Tomographie tridimensionnelle à partir des temps des premières arrivées des ondes P et S: Ph.D. thesis, Université Paris VII.

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

Lions, J. L., 1971, Optimal control systems governed by partial differential equations: Springer Verlag.

Mulder, W. A., 2006, The perturbed traveltime equation and the adjoint-states gradient of the traveltime error: Geophysical Journal International, **167**, 679–683.

Mulder, W. A., and A. P. E. ten Kroode, 2002, Automatic velocity analysis by differential semblance optimization: Geophysics, **67**, 1184–1191.

Paige, C., and M. A. Saunders, 1982, LSQR: Sparse linear equations and least squares problems, part I and part II: ACM Transactions on Math Software, **8**, 43–71.

Pica, A., J. P. Diet, and A. Tarantola, 1990, Nonlinear inversion of seismic reflection data in a laterally invariant medium: Geophysics, **55**, 284–292.

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.

Plessix, R.-E., W. A. Mulder, and A. P. E. ten Kroode, 2000, Automatic cross-well tomography by semblance and differential semblance optimization: Theory and gradient computation: Geophysical Prospecting, **48**, 913–935.

Podvin, P., and I. Lecomte, 1991, Finite difference computation of traveltimes in very contrasted velocity model: A massively parallel approach and its associated tools: Geophysical Journal International, **105**, 271–284.

Sei, A., and W. W. Symes, 1994, Gradient calculation of the traveltime cost function without ray tracing: 64th Annual International Meeting, SEG, Expanded Abstracts, 1351–1354.

Shen, P., W. W. Symes, and C. Stolk, 2003, Differential semblance velocity analysis by wave-equation migration: 73rd Annual International Meeting, SEG, Expanded Abstracts, 2132–2135.

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

———, 1987, Inverse problem theory: methods for data fitting and model parameter estimation: Elsevier.

Vesnaver, A., 2008, Yardsticks for industrial tomography: Geophysical Prospecting, **56**, 457–465.

Vinje, V., E. Iversen, and H. Gjystdal, 1993, Traveltime and amplitude esti-

mation using wavefront construction: Geophysics, **58**, 1157–1166.
Zelt, C. A., A. Azaria, and A. Levander, 2006, 3D seismic refraction traveltime tomography at a groundwater contamination site: Geophysics, **71**, no. 5, H67–H78.
Zelt, C. A., and P. J. Barton, 1998, Three-dimensional seismic refraction tomography: A comparison of two methods applied to data from the Faeroe Basin: Journal of Geophysical Research, **103**, 7187–7210.
Zhang, L., J. W. Rector, and G. M. Hoversten, 2005, Eikonal solver in the celerity domain: Geophysical Journal International, **162**, 1–8.
Zhao, H., 2005, A fast sweeping method for eikonal equations: Mathematics of Computation, **74**, 603–627.
Zhu, X., D. P. Sixta, and B. G. Angstman, 1992, Tomostatics: Turning-ray tomography + static corrections: The Leading Edge, **11**, 15–23.

DOI:10.1190/1.3250266