First-arrival delayed tomography using 1st and 2nd order adjoint-state method

F. Bretaudeau, BRGM (previously at ISTerre, Université Grenoble Alpes), R. Brossier, ISTerre, Université Grenoble Alpes, L. Métivier, LJK-CNRS, Université Grenoble Alpes, J. Virieux, ISTerre, Université Grenoble Alpes*

SUMMARY

We compare the standard procedure of delayed first-arrival time tomography based on two nested loops workflow and ray tracing with two alternative ray-free approaches based on eikonal solver and adjoint formulations. The standard approach requires the explicit computation and storage of Fréchet derivatives. Then the workflow is composed of an outer loop on velocity model updating and an inner loop for the iterative estimation of the model perturbation (usually LSQR). The adjoint approaches avoid computation and storage of Fréchet matrix, and use instead an efficient estimation of the gradient of the data misfit function, which is enough for steepest-descent or *l*-BFGS quasi-Newton methods. In the *l*-BFGS approach, successive gradients are stored and used for an approximate estimation of the influence of the inverse Hessian. Such a workflow is only composed of a single loop optimization. Using a second-order adjoint formulation, we can also obtain adjoint formulas to compute efficiently Hessian-vector product. Those formulas yields a new manner to perform the two-nested-loops workflow (either in the Gauss-Newton approximation or with the full Hessian matrix). The composition of the memory-demanding Fréchet derivative matrix is not required and the computational cost is independent of the number of receivers. This leads to the so-called matrix-free *Truncated Newton* optimization method. Through a synthetic 2D example and a real example of travel-time tomography in the Nankai Trough, we show that the results provided by the l-BFGS quasi-Newton method and the truncated Newton method are similar. This promotes the use of the l-BFGS method for the mono-parameter velocity model reconstruction as it is based on a single loop formalism and is therefore less computationally expensive.

INTRODUCTION

Since the early development in the 1970s (Aki and Lee, 1976) of travel-time tomography based on delayed first-arrival times, the reconstruction of the velocity structure has been performed both for geotechnical (Greenhalgh et al., 1986), exploration (Zhou et al., 1992) and geodynamic purposes (Spakman and Nolet, 1988). Different approaches have been used for computing travel times based on ray tracing equations (Zelt and Smith, 1992) or on eikonal solvers (Le Meur et al., 1997). The latter has been the method of choice because of its robustness in complex structures. The standard minimisation strategy of travel-time residues leads to a two-nested loop optimization problem where the outer loop consists in updating the velocity model, while the inner loop computes the current velocity model update by solving the sparse linear system built up on the current model. Different iterative methods have been proposed as LSQR (Paige and Saunders, 1982), SIRT or DSIRT (Trampert and Lévêque, 1990).

We investigate here the behavior of different methods for this local optimization problem where the Hessian influence should provide the necessary scaling between different physical model parameters, leading to a consistent model updating from the observations. We compare and discuss the standard approach with two matrix-free alternative approaches relying on adjoint-state formulations. In the first proposed approach, only one loop is required. We take benefit of the information deduced from previous steps to approximate the inverse Hessian with an l-BFGS algorithm. In the second one, a two-loop strategy is considered, but in a matrix-free fashion: neither the Hessian or the Fréchet derivatives matrices are computed.

Although these two approaches could be performed whatever is the method selected for computing travel-times, we shall concentrate our attention to those where the forward problem is solved using an eikonal solver. We shall first formulate the optimisation problem with a description of the different solving strategies: (1) the standard two-loops Fréchet based Gauss-Newton (GN) optimization method, (2) the single-loop quasi-Newton (QN) 1st order adjoint-state method and, (3) the two-loops matrix-free truncated Newton method with 2nd order adjoint-state method. Then we illustrate on 2D synthetic and real data examples the effects of the different approaches.

METHOD

Inversion of seismic velocity is often performed in a two-loops optimisation after picking procedure of first-arrival times $t_{obs_{sr}}$ on recorded traces at receivers r for a known source positions s. We consider a simple least-squares misfit function as a sum over sources and receivers given by

$$\mathscr{C}(\mathbf{m}) = \frac{1}{2}\sum_{sr}\left(W_{d_{rs}}(t_{obs_{sr}} - t_{syn_{sr}}(\mathbf{m}))\right)^2 + \varepsilon\, P(\mathbf{m}). \quad (1)$$

The first term is the data misfit in the data domain Ω_d with defined weights $W_{d_{rs}}$ where the synthetic travel times in the current model are denoted by $t_{syn_{sr}}$. The second term $P(\mathbf{m})$ is a regularization term that includes prior model term and/or Tikhonov regularization term in the model domain Ω_m. The hyper-parameter mixing data and model components is denoted by the scalar ε.

The two-loops procedure of the linearized tomography could be resumed as a loop on model updates k with the following equation

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \Delta m_k, \quad (2)$$

where Δm_k is the model update and α_k is a scalar parameter computed through a linesearch or a trust-region procedure (Nocedal and Wright, 2006). Within the framework of Newton algorithms, the increment Δm_k is defined by solving the linear system

$$H(\mathbf{m}_k)\Delta m_k = -\gamma(\mathbf{m}_k), \quad (3)$$

where $\gamma(\mathbf{m}_k)$ and $H(\mathbf{m}_k)$ are respectively the gradient and the Hessian of the misfit function with respect to the model parameter $\mathbf{m}$.

DOI http://dx.doi.org/10.1190/segam2014-0987.1

In the standard method, inside the current velocity model $\mathbf{m}$, travel-times $t_{syn_{sr}}$ are computed as well as rays for each couple (source,receiver). Defining a representation of the model space Ω_m, Fréchet derivatives $J(\mathbf{m})$ can be deduced along these rays assuming that perturbations of the velocity are only on the ray: the so-called *frozen ray* approximation. With the help of the Fréchet derivatives $J(\mathbf{m})$, the gradient is approximated as

$$\gamma(\mathbf{m}_k) = J^t(\mathbf{m})(t_{obs} - t_{syn}) = J^t(\mathbf{m})\Delta\mathbf{t}. \quad (4)$$

The linear system (3) is solved under the Gauss-Newton approximation as

$$J^t(\mathbf{m})J(\mathbf{m})\Delta\mathbf{m} = -J^t(\mathbf{m})\Delta\mathbf{t} \quad (5)$$

with an iterative damped conjugate gradient. The symbol ^{t} denotes the transpose of the matrix. The LSQR algorithm will require products $J(\mathbf{m})\mathbf{x}$ and $J^t(\mathbf{m})\mathbf{y}$ where vectors $\mathbf{x}$ and $\mathbf{y}$ are respectively in the model space and in the data space (Nolet, 1985). One may have a look at the review by Thurber and Ritsema (2007). The standard approach thus relies on the explicit computation of the Fréchet derivatives matrix under the frozen-ray approximation, which assumes that the velocity perturbation are located along the rays.

On the other hand, adjoint formulations (Chavent, 1974; Plessix, 2006) provide a framework in which the explicit computation of the Fréchet derivatives is not required. Adjoint formulations for first-arrival travel-time tomographic problems have been considered previously by Leung and Qian (2006); Taillandier et al. (2009) for the estimation of the gradient. They have used steepest-descent method or quasi-Newton method as the l-BFGS algorithm (Byrd et al., 1995) for updating the velocity structure, leading to a one-loop procedure. Successive gradients are stored to build up an approximation of the inverse Hessian $Q_k \simeq H^{-1}(\mathbf{m_k})$, and the model update Δm_k is computed as

$$\Delta m_k = -Q_k\gamma(\mathbf{m}_k), \quad (6)$$

How this strategy will compete with two-loops procedures is the question we want to address. For such adjoint approach of first-arrival travel-times, they directly solve the eikonal equation for travel-time maps in the whole domain by fast marching or fast sweeping methods (Vidale, 1990; Zhao, 2005). In a medium with continuous velocity $v(\mathbf{x})$, the first-arrival travel-time τ is the solution of the eikonal equation

$$\|\nabla\tau(\mathbf{x})\|^2 = \frac{1}{v^2(\mathbf{x})} = s^2(\mathbf{x}), \quad (7)$$

with boundary condition $\tau(\mathbf{x}_s) = 0$ at the source. We select the slowness $s = 1/v$ as the medium parameter to be reconstructed. Travel-times are extracted at the receiver positions through the operator R_r defined by $t_{syn_{sr}} = R_r(\tau(\mathbf{s},\mathbf{r},s))$ and time delays are defined by $\Delta t_{sr} = t_{obs_{sr}} - t_{syn_{sr}}$ used in the least-square misfit function $\mathscr{C}$ for each data. In the medium, an adjoint field λ is estimated through the following transport equation

$$\nabla\cdot(\lambda(\mathbf{x})\nabla\tau(\mathbf{x})) = -\sum_{rs} R_r^t W_{d_{rs}}^2 \Delta t_{sr} \quad (8)$$

with boundary condition $\lambda(\mathbf{x})(\nabla\tau(\mathbf{x})\cdot\mathbf{n}) = \sum_{rs} W_{d_{rs}}^2 \Delta t_{sr}$ over the surface Γ surrounding the model space Ω_s with a local normal vector defined by $\mathbf{n}$ if receivers are deployed on boundaries. Taillandier et al. (2009) has only considered a continuous receiver surface located at the boundaries of Ω_s, while our formula consider discrete receivers anywhere in the medium leading to the right-hand-side terms of the equation (8).

At the iteration k, the misfit velocity gradient at each point of the medium is readily obtained by the expression

$$\gamma_{s_k}(\mathbf{x}) = -\lambda(\mathbf{x})s_k(\mathbf{x}) + \varepsilon\,\gamma_P(\mathbf{x}). \quad (9)$$

This estimation of the gradient is the only required ingredient for one-loop steepest-descent or l-BFGS approaches.

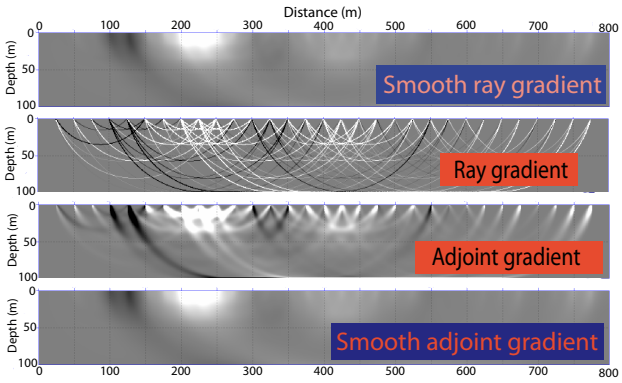

Figure 1: Absolute gradient contribution (same colorscale for all figures) for some source/receiver couples for a 2D toy refraction example: ray gradients are only on the ray while adjoint gradients are diffused around the ray. As we discretize the medium, we smooth gradients and both ray and adjoint gradients look the same.

By computing the gradient through the adjoint formulation, we avoid any ray tracing and the explicit Fréchet derivative J computation and storage. This also means that no assumption regarding where should be located the velocity perturbation have been made. The adjoint approach includes automatically ray perturbation implied by velocity perturbation, while this effect could be estimated with rays in a more difficult manner. The figure 1 shows individual contributions of the gradient estimated by the adjoint formulation and the standard one using the operator J. When these contributions are added for all sources and receivers, the complete gradients are almost identical supporting the hypothesis of the frozen ray approximation when ray density is significant while zones with sparse ray sampling will still show differences by the two estimations. As we prevent significant updating in these poorly sampled zones through proper parameterization and regularisation (equation 3), final results will be almost similar although tuning regularization might give small differences between the frozen-ray and adjoint approaches.

However, one should want to take benefit of second-order local optimization method such as Gauss-Newton (GN) and/or full Newton (FN) methods for speeding up the convergence of the minimisation problem and for correcting partially the impact of the acquisition configuration, still working in the adjoint formulation framework (no explicit computation of the Hessian or the Fréchet derivatives matrix). This can be performed through the solution of (3) with a matrix-free conjugate gra-

DOI http://dx.doi.org/10.1190/segam2014-0987.1

dient algorithm and the computation of Hessian-vector products through second-order adjoint state formulas, leading to truncated Newton optimization methods (Métivier et al., 2013, 2014).

Consider the functional $\mathscr{E}_w(s)$

$$\mathscr{E}_w(s) = < \nabla_s \mathscr{C}(s(\mathbf{x})) | w >, \tag{10}$$

where $< .|. >$ denotes the scalar product on the model space Ω_s. By definition, the gradient of the functional $\mathscr{E}_w(s)$ is the quantity $\nabla_s \mathscr{E}_w(s) = H(s)\ \mathbf{w}(\mathbf{x})$, which is exactly the product of the Hessian by an arbitrary vector $\mathbf{w}(\mathbf{x})$ in the model space. Second-order adjoint formulation yields formula to compute this gradient without explicitly building the Hessian operator. An additional adjoint field μ solution of a similar transport equation as the 1st-order adjoint field λ with different boundary and source conditions should be considered. The Hessian product comes out as

$$H(s)\mathbf{w} = \mathbf{w}(\mathbf{x})\lambda(\mathbf{x}) + s(\mathbf{x})\mu(s) + H_P.\mathbf{w} \tag{11}$$

The field μ is obtained using two sources terms, one corresponding specifically to the 2nd-order term of the Hessian. One can compute the effect of the FN or GN Hessian operators on the vector $\mathbf{w}$ at a similar cost by simply considering the whole expression or neglecting the 2nd-order term when computing the adjoint field μ. We may proceed by solving iteratively the equation (3) using a conjugate gradient method. The number of iterations provides the truncation on this matrix-free approach corresponding to the second loop.

TOY EXAMPLE

We have designed a toy example where a 23 km long and 5 km deep synthetic velocity structure includes a constant velocity gradient and circular positive and negative velocity anomalies of different sizes as shown in the figure 3. 115 receivers are deployed at the free surface and 223 events at a constant depth of 4.2 km in order to mimic a transmission regime. All sensors record each event. The initial model is the constant velocity gradient and we invert for the Vp velocity. The figure 3 provides a nice reconstruction whatever is the optimisation approach. The misfit function decreases approximately the same way for single-loop l-BFGS approach as well as two-loops GN and FN approaches based on the TCN strategy (figure 2). Because we take benefit from previous computations of gradients, l-BFGS method is the fastest and should be selected especially when dealing with large models.

REAL EXAMPLE: NANKAI TROUGH

We consider a travel-time dataset from a wide-aperture seismic experiment using 100 ocean bottom seismometers (OBS) deployed along a line for imaging the eastern Nankai subduction system. Dessa et al. (2004) have performed a travel-time tomography along this line. We have performed similar tomographic reconstruction with this dataset using standard tools (eikonal solver and LSQR algorithm) with two-loops procedure and the new proposed algorithm using one loop over the eikonal solved and l-BFGS optimization. In the first case, we have performed 6 iterations of the outer loop related to the eikonal solver and between 20-30 inner loops for solving the normal equations using the conjugate-gradient LSQR method

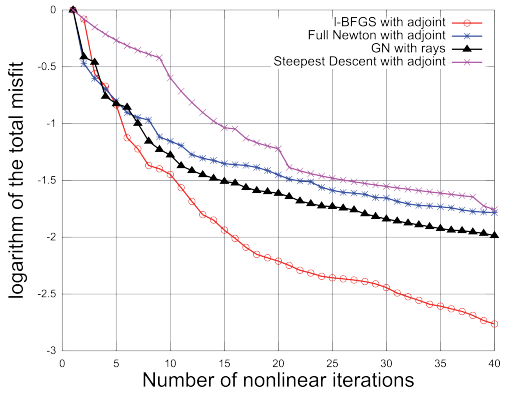

Figure 2: Comparison of the minimization strategies. The *l*-BFGS method exhibits the best convergence properties. The TCN method and the standard LSQR strategies yields comparable results. The steepest decent is based only on the gradient direction and provides the slowest convergence, especially if we consider a 10% decrease often met for real data.

with a reduction of the RMS residuals by more than 80 %. In the second case, we have performed 70 iterations and stored 5 previously computed gradients avoiding the second loop. The RMS residuals is reduced by more than 88 %. The figure 4 shows the initial model we have started with, borrowed from the work by Dessa et al. (2004) as well as the two tomographic results showing similar features with differences coming from the way we perform regularization.

CONCLUSIONS

We have shown that the delayed travel-time tomographic problem could be performed in a single loop as well as in a double loop related to the estimation of travel-times in the updated velocity model using adjoint-state method. By storing previously computed gradients as proposed in the l-BFGS method, we succeed in keeping a good approximation of the Hessian while avoiding the second loop. We have shown on a synthetic example that the second loop based on GN or FN approaches provides the same result as the l-BFGS approach. In a real application, we illustrate that differences are mainly due to regularization procedure when considering two independent computer codes designed for the delayed travel-time tomography. As the adjoint formulation is quite efficient, extensions to 3D geometries can be considered with the l-BFGS method. We shall investigate as well the influence of the noise in observed arrival times.

ACKNOWLEDGEMENTS

This work is partially funded by the SEISCOPE consortium (*http://seiscope2.osug.fr*). This study is granted access to the HPC facilities of CIMENT (Université Joseph Fourier, Grenoble) and GENCI-CINES under grand 046091. Authors would like to thank S. Operto and V. Monteiller for fruitful discussions.

DOI http://dx.doi.org/10.1190/segam2014-0987.1

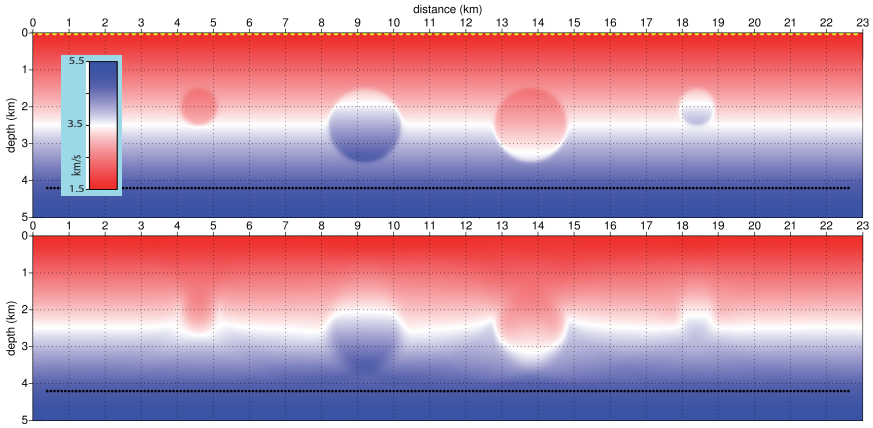

Figure 3: Top panel: synthetic model with both positive and negative anomalies. The initial model is taken without anomalies. Bottom panel: the reconstructed velocity model through the three optimization methods. Sources are the dotted black line at depth of 4.2 km while receivers are yellow dots at the free surface.

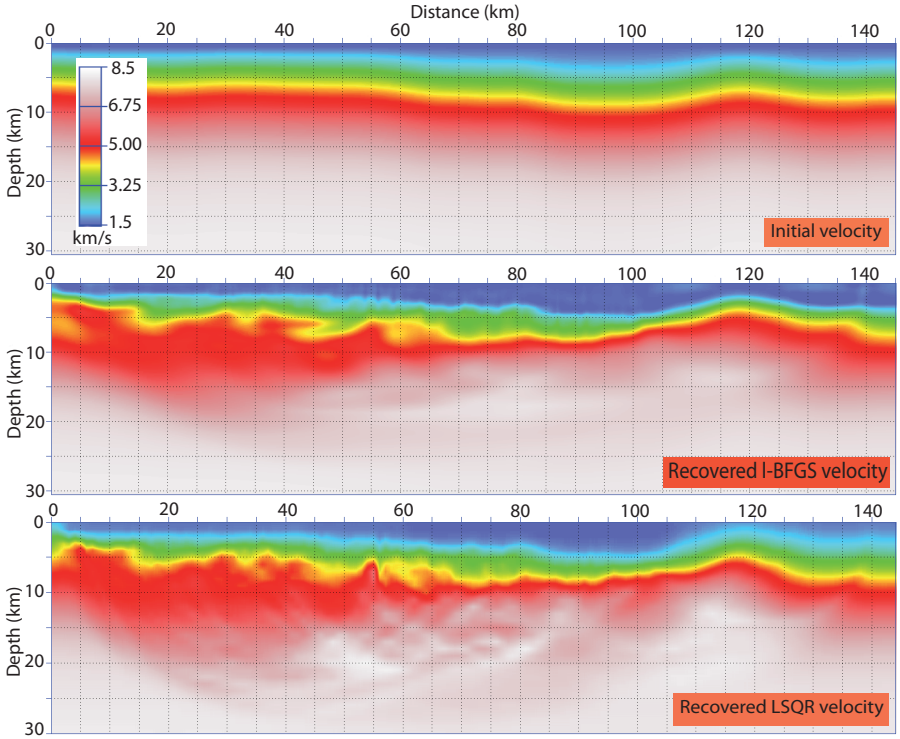

Figure 4: Top panel displays the initial model while the middle panel shows the adjoint reconstructed velocity and the bottom panel the velocity using a standard approach based on LSQR. Similar although not identical features are obtained when comparing the two approaches.

DOI http://dx.doi.org/10.1190/segam2014-0987.1

http://dx.doi.org/10.1190/segam2014-0987.1

EDITED REFERENCES

Note: This reference list is a copy-edited version of the reference list submitted by the author. Reference lists for the 2014 SEG Technical Program Expanded Abstracts have been copy edited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES

Aki, K., and W. Lee, 1976, Determination of three-dimensional velocity anomalies under a seismic array using first P-arrival times from local earthquakes: A homogenous initial model: Journal of Geophysical Research, **81**, no. 23, 4381–4399, http://dx.doi.org/10.1029/JB081i023p04381.

Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu, 1995, A limited memory algorithm for bound constrained optimization: SIAM Journal on Scientific and Statistical Computing, **16**, no. 5, 1190–1208, http://dx.doi.org/10.1137/0916069.

Chavent, G., 1974, Identification of parameter distributed systems, in Identification of function parameters in partial differential equations: American Society of Mechanical Engineers, 31–48.

Dessa, J. X., S. Operto, S. Kodaira, 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**, B2, B02111, http://dx.doi.org/10.1029/2003JB002689.

Greenhalgh, S., D. Burns, and I. Mason, 1986, A crosshole and face-to-borehole in-seam seismic experiment at invincible Colliery, Australia: Geophysical Prospecting, **34**, no. 1, 30–55, http://dx.doi.org/10.1111/j.1365-2478.1986.tb00451.x.

Le Meur, H., J. Virieux, and P. Podvin, 1997, Seismic tomography of the Gulf of Corinth: A comparison of methods: Annali di Geofisica, **XL**, 1–24.

Leung, S., and J. Qian, 2006, An adjoint state method for three-dimensional transmission traveltime tomography using first-arrivals: Communications in Mathematical Sciences, **4**, no. 1, 249–266, http://dx.doi.org/10.4310/CMS.2006.v4.n1.a10.

Métivier, L., F. Bretaudeau, R. Brossier, S. Operto, and J. Virieux, 2014, Full-waveform inversion and the truncated Newton method: Quantitative imaging of complex subsurface structures: Geophysical Prospecting, http://dx.doi.org/10.1111/1365-2478.12136.

Métivier, L., R. Brossier, J. Virieux, and S. Operto, 2013, Full-waveform inversion and the truncated Newton method: SIAM Journal on Scientific Computing, **35**, no. 2, B401–B437, http://dx.doi.org/10.1137/120877854.

Nocedal, J., and S. J. Wright, 2006, Numerical optimization, 2nd ed.: Springer.

Nolet, G., 1985, Solving or resolving inadequate and noisy tomographic systems: Journal of Computational Physics, **61**, no. 3, 463–482, http://dx.doi.org/10.1016/0021-9991(85)90075-0.

Paige, C. C., and M. A. Saunders, 1982, Algorithm 583 LSQR: Sparse linear equations and least squares problems: ACM Transactions on Mathematical Software, **8**, no. 2, 195–209, http://dx.doi.org/10.1145/355993.356000.

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**, no. 2, 495–503, http://dx.doi.org/10.1111/j.1365-246X.2006.02978.x.

DOI http://dx.doi.org/10.1190/segam2014-0987.1

Spakman, W., and G. Nolet, 1988, in Imaging algorithms, accuracy and resolution in delay time tomography: Dordrecht, 155–188.

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, http://dx.doi.org/10.1190/1.3250266.

Thurber, C., and J. Ritsema, 2007, Theory and observations — Seismic tomography and inverse methods: Treatise of geophysics: Elsevier.

Trampert, J., and J.-J. Lévêque, 1990, Simultaneous iterative reconstruction technique: Physical interpretation based on the generalized least squares solution: Journal of Geophysical Research, **95**, B8, 12553–12559, http://dx.doi.org/10.1029/JB095iB08p12553.

Vidale, D., 1990, Finite-difference calculation of traveltimes in three dimensions: Geophysics, **55**, 521–526, http://dx.doi.org/10.1190/1.1442863.

Zelt, C., and R. B. Smith, 1992, Seismic traveltime inversion for 2D crustal velocity structure: Geophysical Journal International, **108**, no. 1, 16–34, http://dx.doi.org/10.1111/j.1365-246X.1992.tb00836.x.

Zhao, H., 2005, A fast sweeping method for eikonal equations: Mathematics of Computation, **74**, no. 250, 603–628, http://dx.doi.org/10.1090/S0025-5718-04-01678-3.

Zhou, B., C. Sinadinovski, and S. Greenhalgh, 1992, Nonlinear inversion traveltime tomography: Imaging high-contrast inhomogeneities: Exploration Geophysics, **23**, no. 2, 459–464, http://dx.doi.org/10.1071/EG992459.

DOI http://dx.doi.org/10.1190/segam2014-0987.1