GEOPHYSICS, VOL. 79, NO. 3 (MAY-JUNE 2014); P. WA129–WA140, 21 FIGS., 1 TABLE.
10.1190/GEO2013-0340.1

DOI:10.1190/geo2013-0340.1

Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion

Biondo Biondi[1] and Ali Almomin[1]

ABSTRACT

The convergence of full-waveform inversion can be improved by extending the velocity model along either the subsurface-offset axis or the time-lag axis. The extension of the velocity model along the time-lag axis enables us to linearly model large time shifts caused by velocity perturbations. This linear modeling was based on a new linearization of the scalar wave equation in which perturbation of the extended slowness squared was convolved in time with the second time derivative of the background wavefield. The linearization was accurate for reflected events and transmitted events. We determined that it can effectively model conventional reflection data as well as modern long-offset data containing diving waves. It also enabled the simultaneous inversion of reflections and diving waves, even when the starting velocity model was far from being accurate. We solved the optimization problem related to the inversion with a nested algorithm. The inner iterations were based on the proposed linearization and on a mixing of scales between the short- and long-wavelength components of the velocity model. Numerical tests performed on synthetic data modeled on the Marmousi model and on the "Caspian Sea" portion of the well-known BP model demonstrated the global-convergence properties as well as the high-resolution potential of the proposed method.

INTRODUCTION

Conventional seismic imaging relies on a separation of scales between the migration velocity model (long-wavelength components) and the reflectivity (short-wavelength components). Figure 1 shows a simplified 1D graphical representation of the separation of scales concept. The black line represents the two disjointed wavelength ranges (mapped into corresponding temporal-frequency bands) and the consequent gap in information between long and short wavelengths. This conceptual understanding naturally leads to a sequential approach for seismic imaging; the velocity model is estimated first, and then it is used as input to migration for imaging reflectivity. In current velocity-estimation practice, reflectivity is used only indirectly to measure the focusing power of the velocity model. The only important exceptions occur when migrated volumes are used to interpret boundaries of geobodies (e.g., salt bodies) and to estimate predominant dips in the geologic layering that are then used to constraint a tomographic velocity update.

The sequential imaging process is slowly being undermined by three long-standing trends in the industry: (1) acquisition of lower-frequency data, (2) imaging under a complex overburden that requires higher resolution velocity models to focus and correctly position reflectors, and (3) acquisition of longer offset data. As the industry strives to widen the data frequency band at the low and high ends, the reflectivity band is extended at the low end, as graphically represented by the green line in Figure 1. The high end of the velocity band is also pushed upward (blue line in Figure 1) by the application of sophisticated tomographic methods that enable the estimation of the high-resolution velocity models needed to focus reflectors located under complex overburden. However, tomography (either ray- or wave-equation-based) is a more challenging task than migration, and thus it often falls short of providing the accuracy and resolution necessary to satisfactorily image the high frequencies in the data. The acquisition of longer offset data enables the recording of diving waves and refracted arrivals that provide a complementary illumination of the velocity components in the crucial scale gap and blur the distinction between migration and tomography because they contain forward-scattering perturbations to transmission events.

As the information gap narrows, imaging methods that simultaneously estimate the velocity and reflectivity model by taking advantage of all the information in the data are becoming more attractive. The renewed interest in full-waveform inversion (FWI)

Manuscript received by the Editor 9 September 2013; revised manuscript received 16 January 2014; published online 22 May 2014.
[1]Stanford University, Stanford Exploration Project, Geophysics Department, Stanford, California, USA. E-mail: biondo@sep.stanford.edu; ali@sep.stanford.edu.

DOI:10.1190/geo2013-0340.1

(Bamberger et al., 1982; Tarantola, 1984) could be mostly explained as an attempt to overcome the limitations imposed by the sequential imaging approach, as well as the availability of the computational power sufficient for practical applications of FWI. FWI has been the most successful when applied to the low frequencies in the data (green line in Figure 1) to improve the velocity-model estimation needed for imaging the high frequencies in the data under complex overburdens. FWI has been less successful in using the high-frequencies in the data to tomographically estimate the long wavelengths in the model.

Since the 1980s, it has been recognized that FWI has a migration component and a tomographic component (Mora, 1989). However, to ensure convergence of the tomographic component, the recorded and modeled data must be almost in phase with each other; the rule of thumb being that the residual time shifts must be shorter than the half-period of the dominant frequency in the data. Bootstrapping the inversion by starting from the low frequencies may ameliorate the convergence problems, but it still depends on conventional velocity estimation methods to deliver starting models sufficiently accurate to satisfy the convergence criterion for the FWI tomographic component. It also undermines the goal of simultaneous estimation because the high frequencies in the data contain the high-resolution tomographic information that facilitates the estimation of the velocity components represented by the blue line in Figure 1.

To perform true simultaneous and synergistic inversion for all the model scales, we must address the convergence problems of the tomographic term in FWI. These convergence problems are related to the nonlinearity of the solution of the wave equation with respect to perturbations in the long wavelengths of the velocity model. Long-wavelength perturbations cause substantial time shifts of the propagating wavefields that are poorly approximated by the linearization of the wave equation based on the first-order Born approximation. When recorded events are easily identified and separated, such as in borehole seismology (Woodward, 1990; Luo and Schuster, 1991) and global seismology (Dahlen et al., 2000), the limitation of the Born linearization can be overcome by minimizing the time lag of the maximum of the crosscorrelation between modeled and recorded data. However, modern reflection data do not fulfill this assumption, and new methods are needed to ensure convergence of the tomographic term in FWI.

In this paper, we introduce a linearization of the wave equation based on an extension of the velocity model along the time lag axis (τ). This extension enables the linear modeling of large time shifts in the propagating wavefields, and consequently in the data. Based on this extension, we define an objective function that has a model-

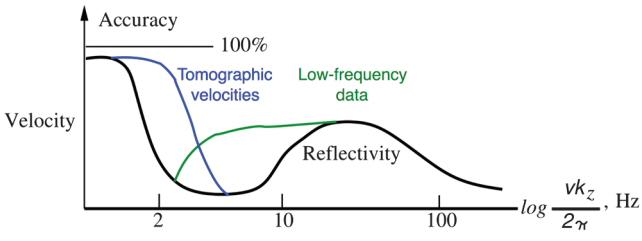

Figure 1. Simplified 1D graphical representation of the separation of scales in seismic imaging (black line) and how current industry trends are narrowing the gap between the estimation of long wavelengths and short wavelengths (blue and green lines). (Adapted from J Claerbout's Imaging the Earth Interior.)

focusing term in addition to the conventional FWI data-fitting term. Numerical examples with realistically complex velocity models demonstrate that this objective function has excellent convergence behavior.

The usefulness of extending the reflectivity model (as prestack images in the angle or offset domain) to manage the nonlinearities in wave-equation velocity analysis was demonstrated in the context of differential semblance optimization (Symes and Carazzone, 1991; Shen and Symes, 2008) and wave-equation migration velocity analysis (WEMVA) (Biondi and Sava, 1999; Sava and Biondi, 2004a, 2004b). The generalization of this idea to the extension of the propagation component of the velocity model (long wavelength) is first introduced by Symes (2008) and more recently is successfully applied by Almomin and Biondi (2012), Biondi and Almomin (2012), and Sun and Symes (2012). These methods are based on a velocity extension along the subsurface-offset or plane-wave ray-parameter axes.

We use an extension along the time lag axis (τ) because it is better suited to describe the large time shifts in wave propagation that are at the root of FWI convergence problems. Furthermore, extending the velocity along the time lag axis can easily handle forward-scattered events recorded at long offsets as well as the reflections recorded at near and intermediate offsets. In contrast, a subsurface-offset extension would require the vertical offset as well as the horizontal offsets to properly model large time shifts for long-offset data. Three subsurface offsets would be required in 3D, and thus the computational burden would increase by several order of magnitude. Yang and Sava (2009, 2010) discuss the use and the computational advantages of time-lag gathers to perform WEMVA for reflected events.

TOMOGRAPHIC FULL-WAVEFORM INVERSION (TFWI)

Conventional FWI is performed by solving the following optimization problem:

$$\min_{\mathbf{s}} J_{\mathrm{FWI}}(\mathbf{s}), \tag{1}$$

where

$$J_{\mathrm{FWI}}(\mathbf{s}) = \frac{1}{2}\|\mathcal{L}(\mathbf{s}) - \mathbf{d}\|_2^2, \tag{2}$$

where $\mathbf{s} = s(\vec{x})$ is the vector of gridded slowness values and $\mathcal{L}$ is a wave-equation operator, whose solutions are nonlinear with respect to slowness perturbations. The data vector $\mathbf{d}$ is a subset of the pressure-field vector $\mathbf{w} = w(t, \vec{x})$ extracted at the surface through a linear sampling operator $\mathbf{K}$, such as $\mathbf{d} = \mathbf{K}\mathbf{w}$.

The wave-equation operator is usually evaluated by recursively solving the following finite difference equation:

$$[\mathbf{S}^2\mathbf{D_2} - \nabla^2]\mathbf{w} = \mathbf{f}, \tag{3}$$

where $\mathbf{D_2}$ is a finite-difference representation of the second derivative in time, ∇^2 is a finite-difference representation of the Laplacian, and $\mathbf{f}$ is the source function. The diagonal matrix $\mathbf{S}^2$ is constructed from the vector $\mathbf{s}^2$; its diagonal elements are the elements of $\mathbf{s}^2$ after replication along the time axis.

DOI:10.1190/geo2013-0340.1

The most efficient solution of the optimization problem expressed in equation 1 is performed by gradient-based methods, and thus it requires the evaluation of the linear operator $\mathbf{L}$, which is the linearization of $\mathcal{L}$ with respect to slowness perturbations $\delta\mathbf{s}^2$. This linear operator can be derived by perturbing equation 3 as follows:

$$[(\mathbf{S_o}^2 + \delta\mathbf{S}^2)\mathbf{D_2} - \nabla^2](\mathbf{w_o} + \delta\mathbf{w}) = \mathbf{f}, \tag{4}$$

where $\mathbf{w_o}$ and $\mathbf{S_o}$ are the background wavefield and slowness, respectively, and $\delta\mathbf{w}$ is the scattered wavefield. The diagonal matrices $\mathbf{S_o}^2$ and $\delta\mathbf{S}^2$ are formed from the corresponding vectors $\mathbf{s_o}^2$ and $\delta\mathbf{s}^2$ in the same way as described above for matrix $\mathbf{S}^2$.

Equation 4 can be rewritten as the following coupled equations:

$$[\mathbf{S_o}^2\mathbf{D_2} - \nabla^2]\mathbf{w_o} = \mathbf{f}, \tag{5}$$

$$[\mathbf{S_o}^2\mathbf{D_2} - \nabla^2]\delta\mathbf{w} = \delta\mathbf{S}^2\mathbf{D_2}(\mathbf{w_o} + \delta\mathbf{w}), \tag{6}$$

which represent a nonlinear relationship between the slowness perturbations and the scattered wavefield. In conventional FWI, to linearize this relationship, we drop the term multiplying the perturbations with each other; that is, we drop the scattered wavefield from the right side of equation 6 and obtain the following coupled equations:

$$[\mathbf{S_o}^2\mathbf{D_2} - \nabla^2]\mathbf{w_o} = \mathbf{f}, \tag{7}$$

$$[\mathbf{S_o}^2\mathbf{D_2} - \nabla^2]\delta\mathbf{w} = \delta\mathbf{S}^2\mathbf{D_2}\mathbf{w_o}. \tag{8}$$

The linear operator $\mathbf{L}$ is evaluated by recursively propagating the background wavefield $\mathbf{w_o}$ and the scattered wavefield $\boldsymbol{\delta}\mathbf{w}$ by solving equations 7 and 8. Its adjoint operator $\mathbf{L}'$, which is needed to compute the gradient of the FWI objective function 1, is evaluated by backward propagating the scattered wavefield solving equation 8 using the data residuals as boundary conditions and evaluating the zero time lag of the crosscorrelation between the background and scattered wavefields.

Equations 7 and 8 define a linear relationship between $\delta\mathbf{w}$ and $\delta\mathbf{s}^2$; however, they cannot model large time shifts between the background wavefield and the scattered wavefield. These large time shifts are correctly modeled by equations 5 and 6 through multiple scattering, that is, by the accumulation of time shifts into the scattered wavefield that enter in the expression of the virtual sources injected by the right side of equation 6. When we drop the scattered wavefield from equation 6, we prevent this accumulation of large time shifts into the scattered wavefield.

These observations suggest that a simple method to improve the capability of the linearization to model large time shifts is to introduce time shifts directly into the slowness perturbations term in the right side of equation 8. We extend the slowness model along the time lag axis τ and convolve its perturbations $\delta\tilde{\mathbf{s}}^2(\tau)$ with the second time derivative of the background wavefield; we rewrite equation 8 as

$$[\tilde{\mathbf{S}}_o^2(\tau = 0)\mathbf{D_2} - \nabla^2]\delta\mathbf{w} = \delta\tilde{\mathbf{S}}^2(\tau)\overset{\tau}{*}\mathbf{D_2}\mathbf{w_o}, \tag{9}$$

where $\overset{\tau}{*}$ denotes convolution in τ. Onward, we use the tilde sign above operators and model vectors to denote their extension along the time-lag axis. The diagonal matrices $\tilde{\mathbf{S}}_o^2$ and $\delta\tilde{\mathbf{S}}^2$ are formed from the corresponding vectors $\tilde{\mathbf{s}}_o^2$ and $\delta\tilde{\mathbf{s}}^2$ in the same way as described above for the matrix $\mathbf{S}^2$.

Equations 7 and 9 define the operator $\tilde{\mathbf{L}}(\tilde{\mathbf{s}}(\tau = 0))$, which is nonlinear with respect to $\tilde{\mathbf{s}}(\tau = 0)$, but defines a linear relationship between $\delta\tilde{\mathbf{s}}^2$ and $\delta\mathbf{w}$. The combination of the wave equation operator $\mathcal{L}(\mathbf{s})$ and of $\tilde{\mathbf{L}}$ defines the extended nonlinear operator

$$\tilde{\mathcal{L}}(\tilde{\mathbf{s}}) = \mathcal{L}(\tilde{\mathbf{s}}_o(\tau = 0)) + \tilde{\mathbf{L}}(\tilde{\mathbf{s}}_o(\tau = 0))\delta\tilde{\mathbf{s}}^2. \tag{10}$$

The modeling equation 10 is used to define the TFWI objective function as

$$J_{\mathrm{TFWI}}(\tilde{\mathbf{s}}) = \frac{1}{2}\|\tilde{\mathcal{L}}(\tilde{\mathbf{s}}) - \mathbf{d}\|_2^2 + \epsilon\||\tau|\tilde{\mathbf{s}}^2\|_2^2. \tag{11}$$

The second term in equation 11 rewards focusing of the data around zero time lag. It introduces a strong tomographic component, which is necessary to constrain the optimization problem because the slowness extension relaxes the constraints on the modeled data kinematics imposed by the data-fitting term (first term) in equation 11. This objective function can be minimized using a nested optimization algorithm with scale mixing, as discussed in the next section.

1D modeling example

We will use a simple 1D numerical example to analyze some of the characteristics of the TFWI method we introduced above. Figure 2 shows the difference between the background wavefield propagated with $\mathbf{v} = 1.2$ km/s and the wavefield propagated with the true velocity of $\mathbf{v} = 1.13$ km/s. The source function is a zero-phase wavelet band limited between 5 and 20 Hz. The wavefield difference is displayed as a function of propagation distance and traveltime. The velocity error is sufficiently high that the wavefields are completely out of phase after propagating for a couple of kilometers. This is therefore a situation like the ones described where the first-order Born linearization (equation 8) would fail to model

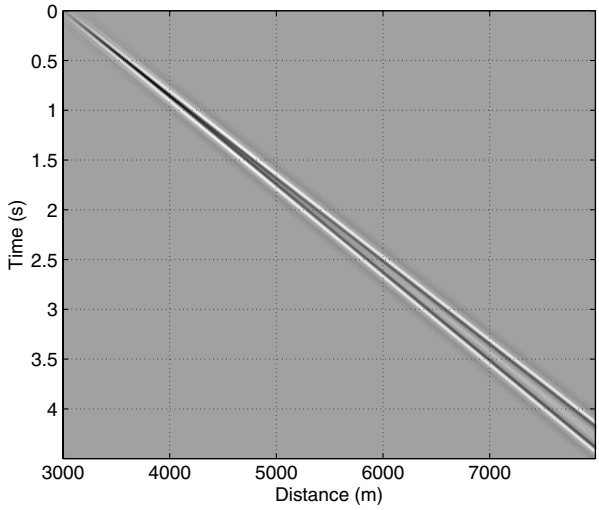

Figure 2. Difference between the background wavefield computed with the starting velocity (1.2 km/s) and the wavefield propagated with the true velocity (1.13 km/s).

the data residuals and conventional FWI would have difficulties to converge, even though the problem is extremely simple.

Figure 3 shows the conventional FWI objective functions when the data are recorded with a single receiver located at 7 km for a total of 4 km offset from the source. The plot shows the value of the initial value of the objective function for several 1D transmission problems sharing the same starting velocity (1.2 km/s) and with different true velocities. If the true velocity is lower than ≈1.18 km/s or larger than ≈1.22 km/s, a gradient based method starting from a velocity of 1.2 km/s will not converge to the right solution. On the contrary, the linearized modeling equation defined in equation 9 would have no problem with modeling the data residual. For example, we can easily reproduce the wavefield difference shown in Figure 2 by setting the extended-velocity perturbation to be a delta function along the τ-axis, where the shift of the delta function linearly increases with the distance from the origin.

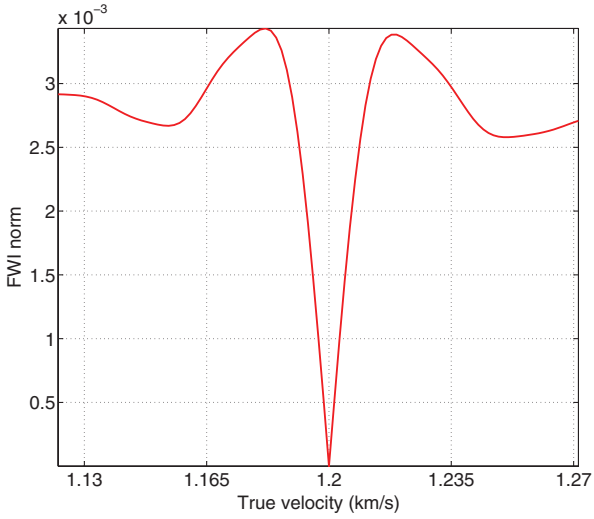

Figure 3. FWI norm as a function of the true velocity, when the starting velocity is equal to 1.2 km/s.

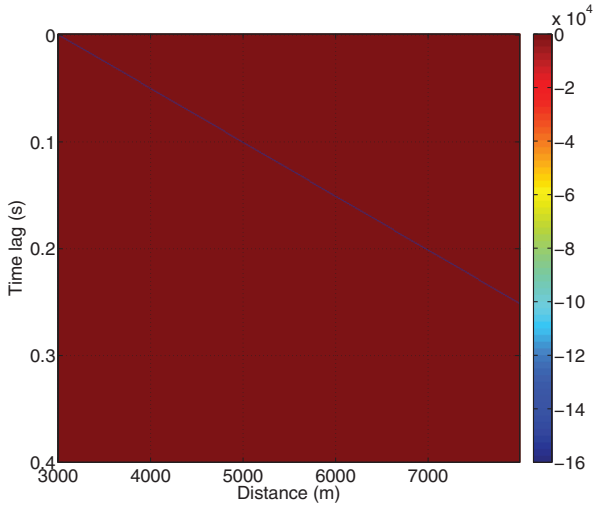

Figure 4. Extended velocity perturbation chosen to approximately model the wavefield difference shown in Figure 2.

This linear shift is computed by integrating the difference in slowness between the background model and the true model. The extended-velocity perturbation is shown in Figure 4. Figure 5 shows the result of solving equation 9 with the model shown in Figure 4. The approximation of the scattered wavefield $\delta\mathbf{w}$ is almost identical to the wavefield difference shown in Figure 2.

Diving-wave modeling example

One of the advantages of extending the velocity model along the time-lag axis τ is the capability to model with a linear operator large time shifts in the diving waves recorded by modern long-offset acquisition geometries. The capability of modeling time shifts in these events is the main reason why the proposed inversion method can robustly converge even when the starting velocity model is far from the correct one, as demonstrated by the second synthetic example presented in this paper.

To show these modeling capabilities, we use one long-offset shot profile recorded over a half-space with a vertical velocity gradient. The starting velocity model is assumed to be uniform and equal to the velocity at the surface. Figure 6a shows the data residual; the recorded diving wave and the data modeled with the starting velocity are clearly visible.

The back projection of the data residuals shown in Figure 6a, by the application of $\tilde{\mathbf{L}}'$, generates the velocity perturbation cube shown in Figure 7. The front panel of the cube shown in Figure 7 displays the zero time lag of the velocity perturbations. A substantial amount of the energy in the residual has been back projected away from the zero time-lag panel.

Figure 6c displays the result of forward modeling the data residuals by the application of $\tilde{\mathbf{L}}$ to the extended velocity perturbation shown in Figure 7. Although a "squaring" of the wavelet is evident in Figure 6c, the kinematics of these modeled residuals are very close to the kinematics of the true residuals shown Figure 6a. In contrast, when we use only the zero time lag of the velocity perturbation to model the data residuals; that is, the conventional first-order Born linearization (front panel of the cube shown in Figure 7), we obtain the seismograms displayed in Figure 6b.

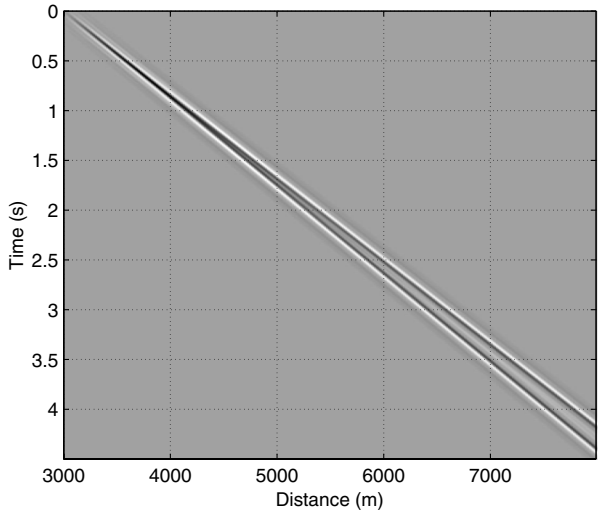

Figure 5. Perturbed wavefield computed by solving equation 9 with the model shown in Figure 4.

The diving wave is totally missing from these modeled residuals because the background wavefield propagates with constant velocity along the horizontal direction.

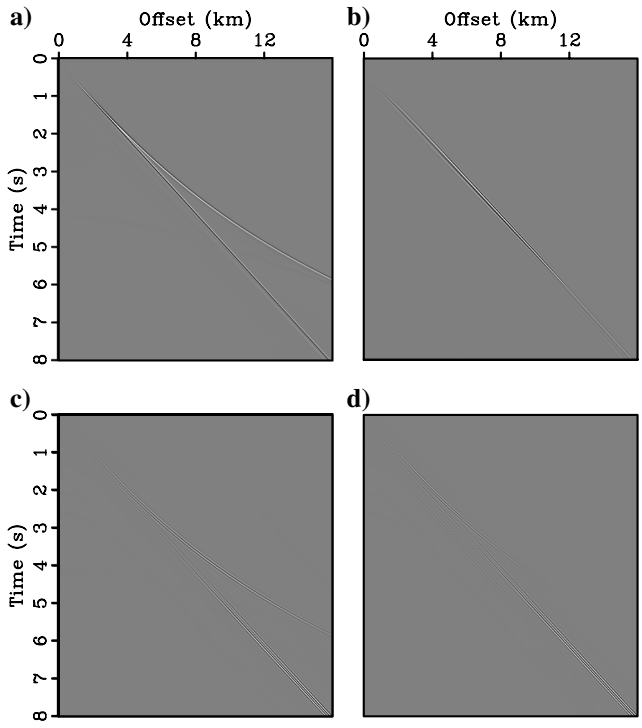

Figure 6. (a) Difference between the background wavefield computed with the starting velocity and the wavefield propagated with the true velocity, (b) data residual modeled from zero lag of the velocity perturbation cubes (front panel in cubes shown in Figures 7 and 8), (c) data residual modeled from the velocity perturbations extended along the time-lag axis (Figure 7), and (d) data residual modeled from the velocity perturbations extended along the horizontal subsurface offset axis (Figure 8).

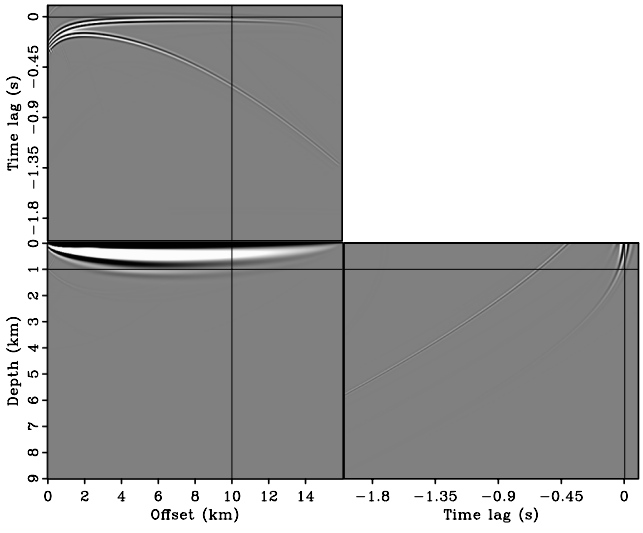

Figure 7. Velocity-perturbation cube extended along the time-lag axis and computed by back projecting the data residuals shown in Figure 6a.

Figure 8 shows the velocity perturbation cube when the velocity is extended along the horizontal subsurface offset axis. The front panel of the cube displays the zero subsurface offset, and thus it is identical to the front panel of the cube shown in Figure 7. Figure 6d displays the result of forward modeling the data residuals starting from the extended velocity perturbation shown in Figure 8. The diving wave event is present in these modeled residuals. However, it dies out at larger offsets, starting at about 8 km. A subsurface offset extension of the velocity has difficulties in modeling large time shifts in transmitted events propagating in directions orthogonal to the subsurface offset axes. Consequently, as the propagation paths of the diving waves have longer vertical components, the less accurate the modeled residuals are. To address this limitation, we could use the vertical subsurface offset in addition to the horizontal ones (Biondi and Symes, 2004), but the computational cost would increase accordingly. In 3D, we would need to extend the velocity along three subsurface offsets, increasing further the dimensionality of the problem, and consequently its computational cost.

OPTIMIZATION METHOD

The minimization of the objective function in equation 11 is a challenging optimization problem because of the nonlinearities in the modeling operator $\tilde{\mathcal{L}}$ and the occasional contradictory search directions suggested by the gradients of the data-fitting term and the focusing term. To overcome these challenges, we devised and tested the nested optimization scheme with scale mixing described below and summarized in Table 1.

Nested inversion

The proposed nested optimization scheme consists of an outer and an inner loop. In the outer loop, we first compute the nonlinear data residual $\mathbf{\Delta d} = \mathbf{d} - \tilde{\mathcal{L}}(\tilde{\mathbf{s}}_o)$, where $\tilde{\mathbf{s}}_o$ is the current slowness model. The nonlinear residual is used as the "observed" data for the inner loop. The output of the inner loop is a velocity update $\Delta\tilde{\mathbf{s}}$.

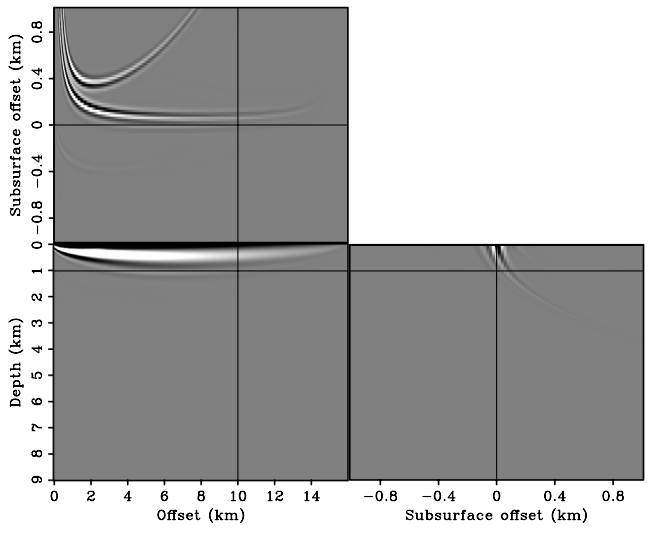

Figure 8. Velocity-perturbation cube extended along the horizontal subsurface-offset axis and computed by back projecting the data residuals shown in Figure 6a.

DOI:10.1190/geo2013-0340.1

In the inner loop, we formally separate the slowness model into a background model $\mathbf{b}$ on which the operator $\tilde{\mathbf{L}}$ depends nonlinearly and a perturbation model $\mathbf{p}^2(\tau)$ on which the output of the operator $\tilde{\mathbf{L}}$ depends linearly. The objective function minimized in the inner loop is

$$J_{\text{ITFWI}}(\mathbf{b},\mathbf{p}) = \frac{1}{2}\|\tilde{\mathbf{L}}(\mathbf{b})\mathbf{p}^2 - \Delta\mathbf{d}\|_2^2 + \epsilon\||\tau|\mathbf{p}^2\|_2^2. \tag{12}$$

The starting models for the inner iterations are $\mathbf{b}_o = \tilde{\mathbf{s}}_o(\tau=0)$ and $\mathbf{p}_o = 0$. The output of the inner loop after N L-BFGS iterations (Nocedal, 1980) is $\Delta\tilde{\mathbf{s}} = (\mathbf{b}_N - \mathbf{b}_o) + \mathbf{p}_N(\tau=0)$.

Unless the starting background slowness $\mathbf{b}_o$ is very close to the true slowness, no choice of $\mathbf{b}$ and $\mathbf{p}$ will simultaneously zero the two terms in the objective function (equation 12). This happens because of the particular choice of the data residual Δd, that is, the difference between the recorded data and the data modeled with $\mathbf{b}_o$. However, the models $\mathbf{b}$ and $\mathbf{p}$ that minimize this objective function provide an effective update $\Delta\tilde{\mathbf{s}}$ for the outer loop of the nested optimization problem.

The output of the modeling operator $\tilde{\mathbf{L}}$ is linear with respect to perturbation $\mathbf{p}^2$ but nonlinear with respect to the background component $\mathbf{b}$. Therefore, another linearization around the "background" background is required to compute the gradient. The Born approximation is used (again) to linearize the $\tilde{\mathbf{L}}$ operator with respect to the background resulting in a data-space tomographic operator, $\mathbf{T}$. Appendix A describes the derivation of this new data-space tomographic operator and how to numerically evaluate it and its adjoint. The expression of the two gradients at the inner iteration i is the following:

$$\nabla_{\mathbf{b}^2} = \mathbf{T}'(\mathbf{b}_i,\mathbf{p}_i)[\tilde{\mathbf{L}}(\mathbf{b}_i)\mathbf{p}_i^2 - \Delta\mathbf{d}], \tag{13}$$

Table 1. Nested optimization algorithm.

iterate { $j = 0, \ldots, M$
- $\Delta\mathbf{d}_i \leftarrow \mathbf{d} - \tilde{\mathcal{L}}(\tilde{\mathbf{s}}_j)$
- $\mathbf{b}_o \leftarrow \tilde{\mathbf{s}}_j(\tau=0)$
- $\mathbf{p}_o \leftarrow 0$
- iterate {$i = 0, \ldots, N$
  - $\mathbf{r}_\mathbf{d} \leftarrow \tilde{\mathbf{L}}(\mathbf{b}_i)\mathbf{p}_i^2 - \Delta\mathbf{d}_j$
  - $\mathbf{r}_\mathbf{m} \leftarrow |\tau|\mathbf{p}_i^2$
  - $\mathbf{r}_{\text{norm}} \leftarrow 0.5\|\mathbf{r}_\mathbf{d}\|_2^2 + 0.5\epsilon\|\mathbf{r}_\mathbf{m}\|_2^2$
  - $\nabla_{\mathbf{b}^2} \leftarrow \mathbf{T}'(\mathbf{b}_i,\mathbf{p}_i)\mathbf{r}_\mathbf{d}$ (equation 13)
  - $\nabla_{\mathbf{p}^2} \longleftarrow \tilde{\mathbf{L}}'(\mathbf{b}_i)r_\mathbf{d} + \epsilon|\tau|r_\mathbf{m}$ (equation 14)
  - $(\Delta\mathbf{p}^2, \Delta\mathbf{b}^2) \leftarrow$ scale mixing $(\nabla_{\mathbf{p}^2}, \nabla_{\mathbf{b}^2})$ (equations 15 and 16)
  - $(\mathbf{p}_{i+1}, \mathbf{b}_{i+1}) \leftarrow$ L-BFGS stepper $(\mathbf{p}_i, \mathbf{b}_i, \Delta\mathbf{p}^2, \Delta\mathbf{b}^2, \mathbf{r}_{\text{norm}})$
  - }
- $\Delta\tilde{\mathbf{s}}^2 \leftarrow$ low-pass filter $(\mathbf{p}_N^2(\tau=0) + \mathbf{b}_N^2 - \tilde{\mathbf{s}}_i^2(\tau=0))$
- $\tilde{\mathbf{s}}_{i+1}^2 \leftarrow \tilde{\mathbf{s}}_i^2 + \Delta\tilde{\mathbf{s}}^2$
- }

$$\nabla_{\mathbf{p}^2} = \tilde{\mathbf{L}}'(\mathbf{b}_i)[\tilde{\mathbf{L}}(\mathbf{b}_i)\mathbf{p}_i^2 - \Delta\mathbf{d}] + \epsilon\tau^2\mathbf{p}_i^2. \tag{14}$$

Scale mixing

In the inner loop, a straightforward use of the gradients is used to update their corresponding models directly. However, this would hinder the simultaneous inversion of different wavelengths of the model. This problem becomes apparent when we examine the result of the two operators in the inner loop. At the first iteration, the application of $\tilde{\mathbf{L}}'$ to the data residual $\Delta\mathbf{d}$ is equivalent to "migrating" the data, and it could give a tomographic update that manifests as a low-wavenumber update. In a conventional migration, this low-wavenumber component is considered noise and filtered out. However, it is actually a tomographic component that should feed into the background model. The opposite argument is also true for the tomographic operator creating short-wavelength perturbations. Therefore, to improve our inversion results, we first mix the two gradients $(\nabla_{\mathbf{b}^2}, \nabla_{\mathbf{p}^2})$ and then separate them in the Fourier domain to get the update of each model as follows:

$$\Delta\mathbf{b}^2(\vec{x}) = -\mathbf{C}_{\mathbf{b}^2}(\nabla_{\mathbf{b}^2}(\vec{x}) + \nabla_{\mathbf{p}^2}(\vec{x},\tau=0)), \tag{15}$$

where $\Delta\mathbf{b}^2(\vec{x})$ is the search direction of the background model and $\mathbf{C}_\mathbf{b}$ is a low-pass filter along the space coordinates $\vec{x}$. Similarly, we can compute the update of the perturbation model as

$$\Delta\mathbf{p}^2(\vec{x},\tau) = -\mathbf{C}_{\mathbf{p}^2}(\nabla_{\mathbf{b}^2}(\vec{x}) + \nabla_{\mathbf{p}^2}(\vec{x},\tau)), \tag{16}$$

where $\Delta\mathbf{p}^2(\vec{x},\tau)$ is the search direction of the perturbation model and $\mathbf{C}_\mathbf{p}$ is a high-pass filter along the space coordinates $\vec{x}$. To sum the two gradients properly, both of them need to have the same units as well as the same scale. This requires careful implementation of each operator at each linearization.

In the examples of this paper, we used a radial cut-off in the Fourier domain with a cosine squared taper. The wavelength cut-off is inversely proportional to the dominant frequency in the data as well as the average slowness of the initial model. The two filters were designed such that they always sum to one at all wavelengths to maintain the energy of the gradients. It is possible to design a more accurate filter that varies with frequency and slowness, but it is not necessary because both models will eventually be added to the slowness. This relative insensitivity is another benefit of applying the nested scheme we presented, in comparison with inverting the models separately and combining them only at the end. If the latter approach were implemented, the final results would be more sensitive to the choice of the scale-separation parameters.

Low-pass filtering of velocity update

We also found that to speed up convergence of the inversion algorithm, it is also helpful to apply a low-pass filter to the velocity update resulting from the last iteration of the inner loop, before adding it to the background model for the next iteration of the outer loop. This low-pass filtering is denoted "low-pass filter" in Table 1.

The passband of the low-pass filter becomes wider with iterations, and at later iterations, the filter is not applied at all. This filtering saves iterations because it spares the inversion process from having to move sharp boundaries in the velocity model from their

DOI:10.1190/geo2013-0340.1

positions determined by the initial velocity to the final positions, which are consistent with the more accurate final velocity.

Brief discussion of optimization algorithm

This nested scheme has many benefits compared to a more conventional way of solving a tomographic inversion followed by an imaging inversion or even doing them simultaneously. The first benefit of this scheme is that the limitation of the linearized, first-order Born operator will not prevent the optimization from inverting higher order scatterings, such as multiple reflections and prismatic events. In this setup, the Born operator will attempt to match the first-order scattering from the background, which is not necessarily primary data because the background itself can generate many orders of scatterings if it contains sharp boundaries (and it will in later iterations). In other words, the linearized operator will account for one additional order of scattering at a time. Higher and higher scattering orders will be introduced with outer-loop iterations, until we eventually invert all the multiple-scattered events that are present in the data. This also means that higher order scattered energy will initially be wrongly positioned in the model, but later iterations will correct for this mispositioning. The second benefit of this scheme is that it produces only one slowness model because we keep pushing background and perturbation into the slowness model, in contrast to the method presented in Almomin and Biondi (2012). Therefore, our goal is to drive the perturbation to a minimum and have the background explain the data.

TFWI OF MARMOUSI DATA

The Marmousi 2 model (Martin et al., 2002) is used for the first synthetic. Figure 9 shows the true velocity model. The water layer is thicker than in the original Marmousi; this thicker layer reduces the amount of refracted energy being recorded in the data. Therefore, this data set tests the capabilities of the inversion method to converge in presence of almost exclusively reflected events. In contrast, the example presented in the next section tests the capabilities of the method to deal with mixed reflected/refracted arrivals.

There are 851 fixed receivers with spacing of 20 m and 171 sources with spacing of 100 m. We use a band-passed wavelet with a frequency range between 5 and 25 Hz and a small taper on both ends. The purpose of using this wavelet is to completely eliminate unrealistically low frequencies in the data. The data were modeled using a constant-density finite-difference solution of the acoustic wave equation. The initial model is shown in Figure 10, which is obtained by laterally averaging the true model, after clipping out the anomalous high-velocity salt bodies.

Figure 11 shows the inversion results after 35 outer-loop iterations, each of them with 10 inner iterations. The relative weight between the data-fitting and the model-focusing terms in the objective function (ϵ in equation 11) was determined heuristically. It was set to the value that balanced the magnitude of the two terms at the second inner iteration. The τ-axis spanned the −0.5–0.5 s range, with 5 ms sampling. The inversion shows remarkable results in reconstructing the velocity model. The results degrade only near the side edges of the model because of limited illumination by the data. To facilitate the comparison, the three velocity models shown in Figures 9–11 are displayed using the same clip values; consequently, the same color corresponds to the same velocity value for all three figures.

Figure 12 shows the norm of the data-residual vector of TFWI (the first term in equation 11) as a function of the number of outer-loop iterations. The data residuals are almost zero after 30 outer

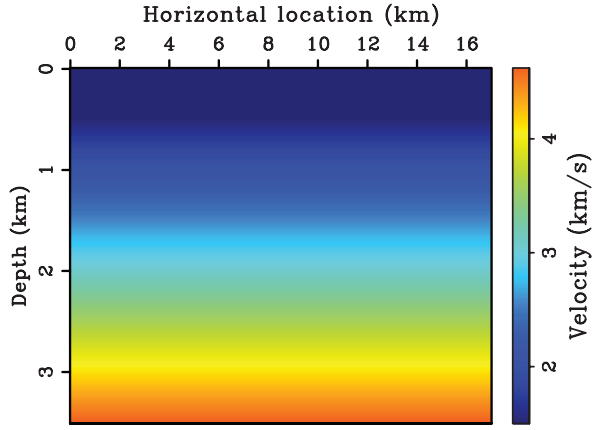

Figure 10. The initial velocity of the Marmousi example.

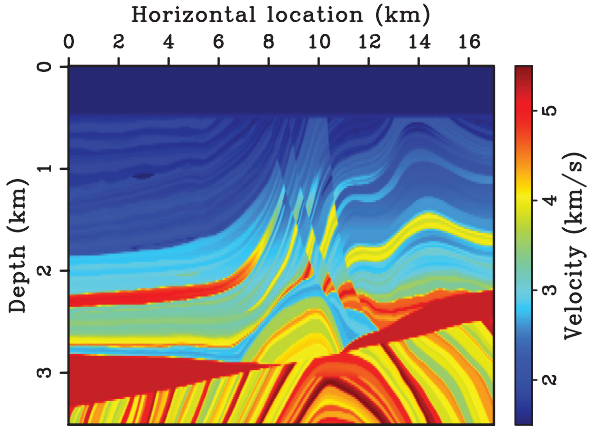

Figure 9. The true velocity of the Marmousi example.

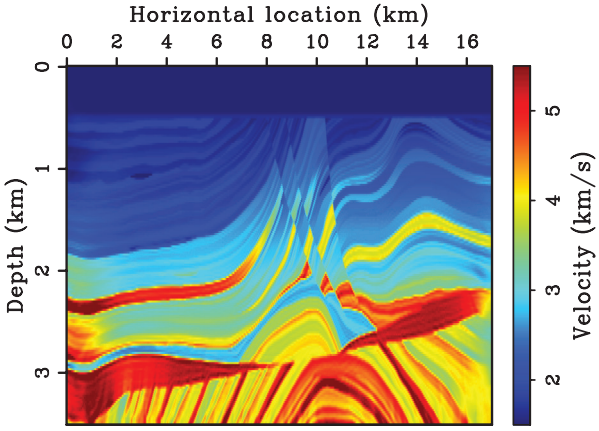

Figure 11. The inverted velocity of the Marmousi example.

iterations. Notice that the TFWI residuals are not monotonically decreasing because the model estimated using the inner iterations is not guaranteed to reduce the data-fitting residuals evaluated using the first term in equation 11.

Figures 13 and 14 provide a clear illustration of the substantial improvements in the accuracy with which the final velocity model describes the data kinematics as compared with the initial one. Figure 13 shows the migrated image corresponding to the initial velocity; most of the reflectors are out of focus and mispositioned. The kinematic errors are so large that they prevent conventional FWI from converging. Furthermore, the focusing of the reflectors in the middle of the section is so poor that conventional migration velocity analysis (either ray- or wave-equation-based) may have troubles to be bootstrapped and to start converging toward a good velocity model. In contrast, Figure 14 shows the migrated image corresponding to the final velocity. Reflectors are well focused and the structures are well imaged. In the middle of the section, even the deepest reflectors are well focused and accurately positioned.

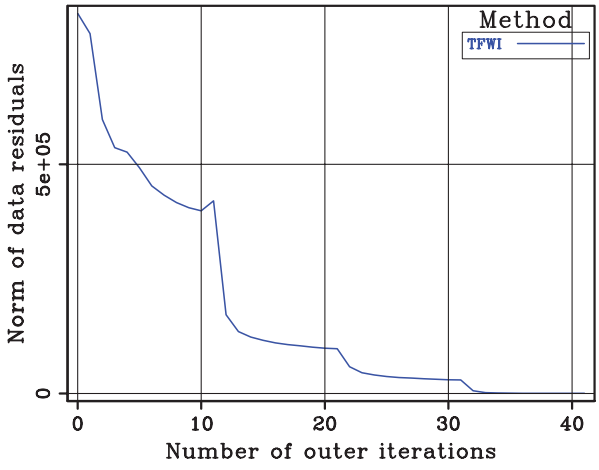

Figure 12. Norm of data-residual vector of TFWI (the first term in equation 11) as a function of the number of outer-loop iterations.

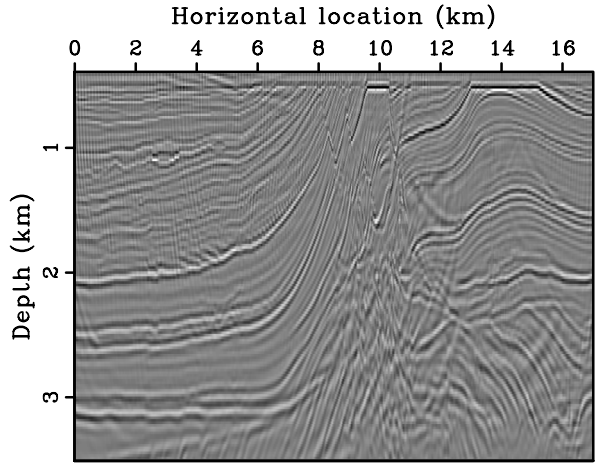

Figure 13. Migrated image using the initial velocity for the Marmousi example.

TFWI OF LONG-OFFSET DATA

To verify the capabilities of the TFWI method to invert diving waves, we tested it on a synthetic data set recorded with long offsets. The data were generated over the Caspian Sea portion of the well-known BP velocity model, as shown in Figure 15. The receiver array was assumed to be fixed at the surface, and thus data with almost 30 km long offsets were recorded. There are 297 sources with 100 m shot spacing and 1484 receivers with 20 m spacing. The source was a band-passed wavelet between 5 and 25 Hz. No energy was present below 3 Hz. The data were modeled using a constant-density finite-difference solution of the acoustic wave equation.

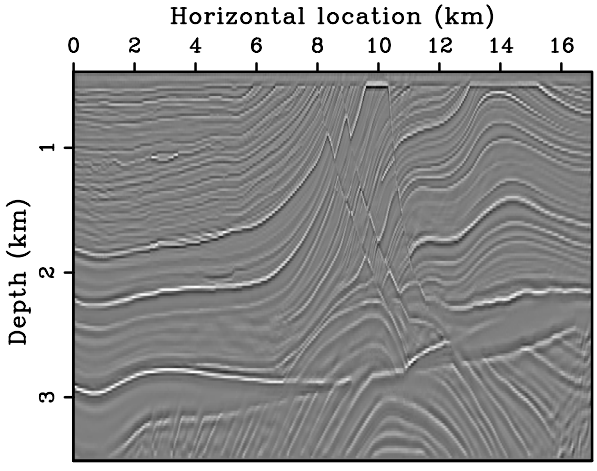

Figure 14. Migrated image using the final velocity for the Marmousi example.

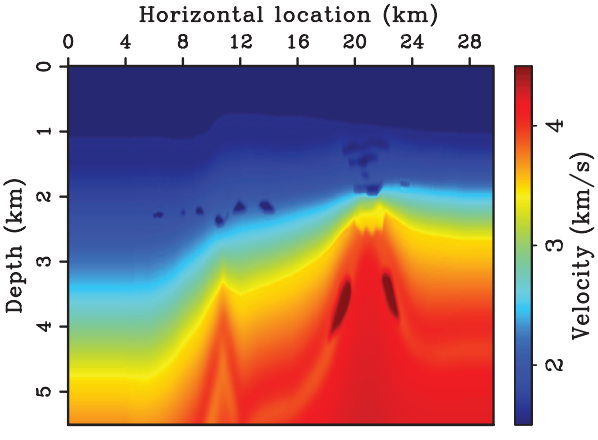

Figure 15. Portion of the BP velocity model used for the numerical test of the proposed TFWI method. The model contains both low-velocity anomalies (shallow gas) and a high-velocity anomaly on the flanks of the mud volcano.

DOI:10.1190/geo2013-0340.1

Figure 16 shows the data recorded for the shot at the horizontal coordinate of 4 km. Strong and complex diving waves and refracted arrivals are visible in the data starting from approximately 8 km offsets. These events carry useful information on the velocity, in particular, in the shallow part of the section. In this data set, they are extremely useful to define the low-velocity anomalies present around the depth of 2 km.

The starting model, shown in Figure 17, was a simple horizontally invariant model obtained by horizontally averaging the true model after the low- and high-velocity anomalies were removed. Figure 18 shows the shot gather located at the same location as the one shown in Figure 16, but modeled with the starting model. As a direct comparison of Figure 16 with Figure 18 demonstrates, the differences between the true and starting models cause large time shifts in the diving-wave arrivals. The inaccuracies of the starting model, together with the lack of low frequencies in the data, prevent conventional FWI from converging to any useful model. Figure 19 shows the result of FWI after 200 iterations, at which time the inversion process was stopped because it was not converging. Notice that the conventional FWI results are a particular case of the TFWI results when $\tau = 0$.

Figure 20 shows the result of the TFWI inversion after 35 outer-loop iterations of the nested optimization algorithm described above. As for the Marmousi example, the relative weight between the data-fitting and the model-focusing terms in the objective function (ϵ in equation 11) was determined heuristically. It was set to the value that balanced the magnitude of the two terms at the second inner iteration of the inversion. The τ-axis spanned the -0.5–0.5 s

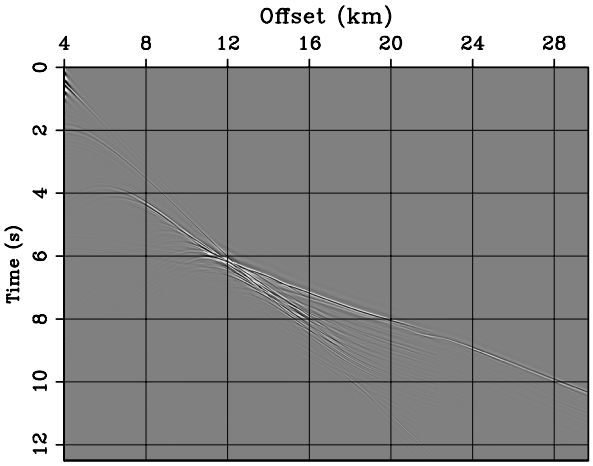

Figure 16. Leftmost shot profile recorded on the model shown in Figure 15. Notice several diving waves and refractions present in the data at offsets larger than 8 km. These events carry useful information for the estimation of the velocity anomalies present in the model.

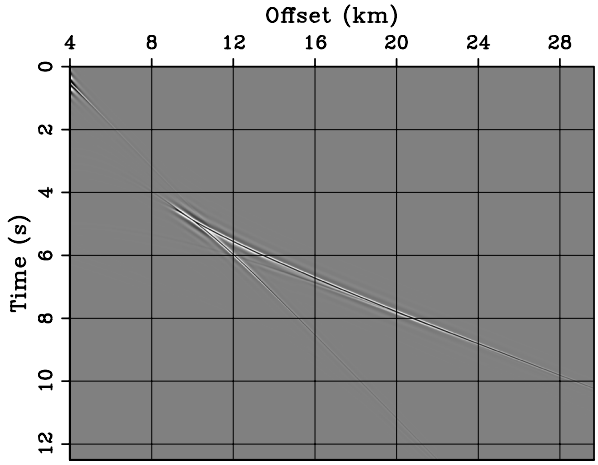

Figure 18. Shot gather modeled assuming the starting model shown in Figure 17 at the same shot location as the data shown in Figure 16. Notice the large time shifts between the diving-wave arrivals in this gather with the one shown in Figure 16.

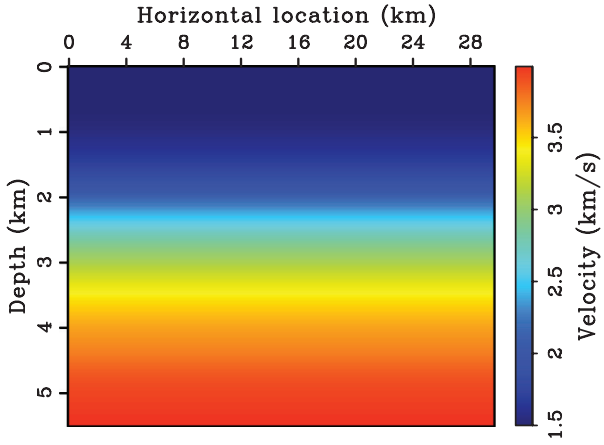

Figure 17. Starting model for the TFWI inversion. This model was obtained by horizontal averaging of the model shown in Figure 15, after the low- and high-velocity anomalies were removed. The lack of low frequencies in the data makes this model inappropriate for starting a conventional FWI inversion.

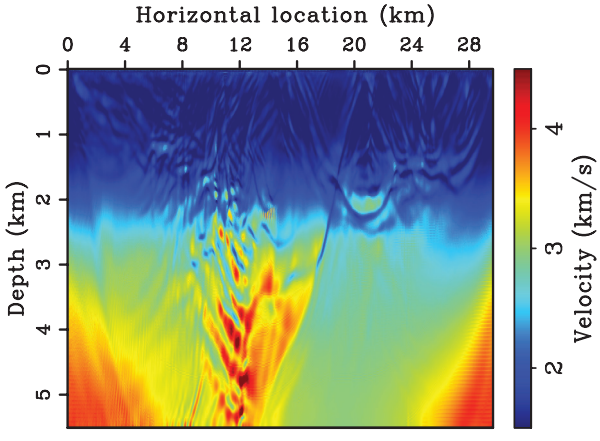

Figure 19. Estimated model after 200 iterations of conventional FWI. Because the starting model was inaccurate, and the data have no energy below 3 Hz, conventional FWI does not converge to any useful model.

range, with a 5 ms sampling. All the features of the true model are accurately reconstructed; only some faint edge artifacts are present. As from the previous example, the four velocity models are displayed using the same clip values, and consequently the same color corresponds to the same velocity value for all four figures.

Figure 21 shows the norm of the data residual of both FWI (dense dots) and TFWI (sparse circles), as expressed in equation 2 and in the first term of equation 11, respectively. For the TFWI process, the horizontal axis corresponds to number of inner iterations, in contrast with the previous example in which we plotted the number of outer iterations (Figure 12). The cost of one TFWI inner iteration is comparable to, though higher than, the cost of one FWI iteration. The plot therefore compares data residuals of models obtained with comparable computational effort. After 200 iterations, the norm of FWI residuals is approximately half of the initial and the curve is flattening out. On the contrary, the TFWI residuals are almost zero after a little more than 200 inner iterations. Notice that, as for the previous example, the TFWI residuals are not monotonically decreasing because the solution of the inner iteration is not guaranteed to reduce the residuals evaluated according to the data-fitting term of equation 11.

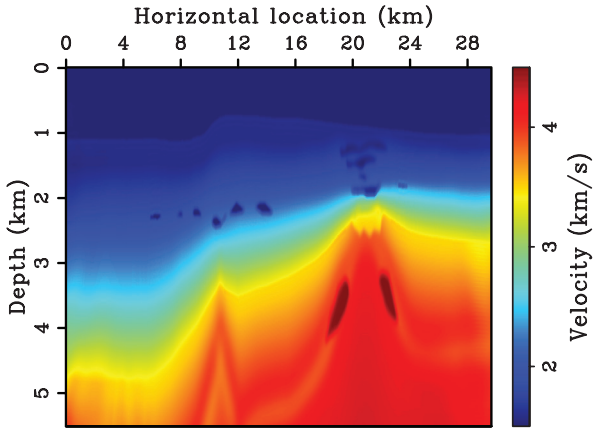

Figure 20. Estimated model after 35 iterations of the outer loop of the TFWI inversion based on time-lag extension of the velocity model. The main features of the true model are accurately reconstructed.

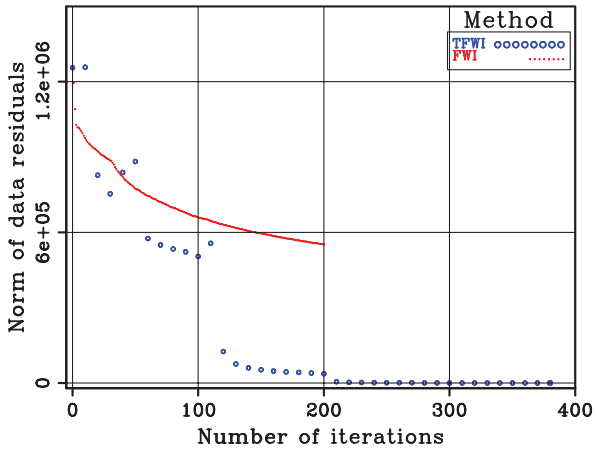

Figure 21. Norm of the data-residual vector of FWI (equation 2) and of TFWI (the first term in equation 11) as a function of the number of iterations (FWI), and inner-loop iterations (TFWI).

DISCUSSION

The synthetic data used for our tests were modeled using a constant-density finite-difference solution of the acoustic wave equation. Further tests using synthetic data modeled using the elastic wave equation and field data are outside the scope of this paper, but they are needed to further validate the proposed method.

The proposed optimization is effective to demonstrate TFWI convergence properties on 2D data sets modeled assuming complex geologic models. The application of the proposed method to 3D problems will require us to extend the velocity model to a 4D hypercube, and thus it would be computationally more expensive than the applications presented in this paper. However, it would be more efficient than an extension of the velocity model along the subsurface-offset axes, which would require at least a 5D velocity hypercube and possibly a 6D one. The increase in computational cost of TFWI over conventional FWI is mostly related to the additional floating-point operations needed to propagate the perturbed wavefields (introducing a factor of two in computational cost) and the additional data-movements needed to compute the time-lag correlations. In our numerical experiments, we measured, in average, a triplication of the run time for each iteration when we switched from FWI to TFWI.

CONCLUSIONS

The integration of FWI and WEMVA into TFWI has the potential of yielding a waveform-inversion method that robustly converges to high-resolution models using the whole bandwidth of the seismic data simultaneously. We introduced a TFWI method based on the extension of the velocity along the τ-axis. This extension is based on a linear operator capable of correctly modeling transmitted events with large time shifts, as we demonstrate by two numerical examples: The first is based on a simple 1D model, and the second is based on a 2D layered model.

To minimize the TFWI objective function, we propose a specialized nested inversion scheme. In the inner loop of this scheme, the extended velocity model is separated into its background and perturbation components. The inversion scheme performs simultaneous inversion of different model scales by mixing the gradients of the two components and then separating them in the Fourier domain.

The results of the inversion of the Marmousi model illustrate the convergence characteristics of the new method for typical reflection data. The results of the inversion of a long-offset data set recorded over the BP Caspian Sea demonstrate that the inversion method converges when both reflections and diving waves are recorded in the data and inaccuracies of the starting velocity model create large errors in the kinematics of the diving waves. When we applied conventional FWI to the full-bandwidth data starting from the same inaccurate starting model, the inversion quickly converged to the unsatisfactory local minimum.

ACKNOWLEDGMENTS

We would like to thank BP for making publicly available the velocity model of which we used a portion, and IFP and the University of Houston for generating and distributing the Marmousi models. Almomin would like the thank Saudi Aramco for supporting his Ph.D. studies at Stanford. We would like also to thank an anonymous reviewer who provided useful suggestions that substantially helped to improve the clarity and completeness of the paper.

APPENDIX A

EVALUATION OF THE GRADIENTS OF THE OBJECTIVE FUNCTION WITH RESPECT TO B AND P

The minimization of the objective function (equation 12) requires the computation of the gradients with respect to $\mathbf{b}$ and $\mathbf{p}$. These gradients can be computed by a perturbation analysis of the modeling operator $\tilde{\mathbf{L}}$. As discussed in the main text, $\tilde{\mathbf{L}}$ is evaluated by solving equations 7 and 9. Rewriting these equations in terms of $\mathbf{b}$ and $\mathbf{p}$, the incident wavefield $\mathbf{w_i}$, and the scattered wavefield $\mathbf{w_s}$, we obtain

$$[\mathbf{B}_i^2\mathbf{D_2} - \nabla^2]\mathbf{w_i} = \mathbf{f}, \tag{A-1}$$

$$[\mathbf{B}_i^2\mathbf{D_2} - \nabla^2]\mathbf{w_s} = \mathbf{P}_i^2(\tau)^{\tau}*\mathbf{D_2}\mathbf{w_i}, \tag{A-2}$$

where the diagonal matrices $\mathbf{B}_i^2$ and $\mathbf{P}_i^2$ are formed from the corresponding vectors $\mathbf{b}_i$ and $\mathbf{p}_i$ in the same way as described in the main text for the diagonal matrix $\mathbf{S}$. Introducing the perturbations $\delta\mathbf{B}^2$ and $\delta\mathbf{P}^2$ in the two model variables into equations A-1 and A-2 and introducing the corresponding perturbations in the wavefields $\delta\mathbf{w_i}$ and $\delta\mathbf{w_s}$ yields the following perturbed system:

$$[(\mathbf{B}_i^2 + \delta\mathbf{B}^2)\mathbf{D_2} - \nabla^2](\mathbf{w_i} + \delta\mathbf{w_i}) = \mathbf{f}, \tag{A-3}$$

$$[(\mathbf{B}_i^2 + \delta\mathbf{B}^2)\mathbf{D_2} - \nabla^2](\mathbf{w_s} + \delta\mathbf{w_s}) = [\mathbf{P}_i^2(\tau) + \delta\mathbf{P}^2(\tau)]^{\tau}*\mathbf{D_2}(\mathbf{w_i} + \delta\mathbf{w_i}). \tag{A-4}$$

By setting $\delta\mathbf{B}^2 = 0$ in equations A-3 and A-4, we derive the following system of equations:

$$[\mathbf{B}_i^2\mathbf{D_2} - \nabla^2]\mathbf{w_i} = \mathbf{f}, \tag{A-5}$$

$$[\mathbf{B}_i^2\mathbf{D_2} - \nabla^2]\delta\mathbf{w_s} = \delta\mathbf{P}^2(\tau)^{\tau}*\mathbf{D_2}\mathbf{w_i}, \tag{A-6}$$

which again defines $\tilde{\mathbf{L}}$ and can be used to evaluate the perturbations in the scattered wavefield $\delta\mathbf{w_s}$ and consequently in the recorded data $\delta\mathbf{d} = \mathbf{K}\delta\mathbf{w_s}$ caused by perturbations $\delta\mathbf{P}^2$.

We can derive the data-space tomographic operator $\mathbf{T}$ which relates perturbations in the scattered wavefield $\delta\mathbf{w_s}$ to perturbations in the background model $\delta\mathbf{B}^2$ by setting $\delta\mathbf{P}^2 = 0$ and neglecting the higher order terms in $\delta\mathbf{B}^2$ in equations A-3 and A-4.

This tomographic operator is equal to the sum of two operators $\mathbf{T_i}$ and $\mathbf{T_s}$. The first operator, $\mathbf{T_i}$ models perturbations in the scattered wavefield caused by perturbations in the propagation of the incident wavefield:

$$[\mathbf{B}_i^2\mathbf{D_2} - \nabla^2]\mathbf{w_i} = \mathbf{f}, \tag{A-7}$$

$$[\mathbf{B}_i^2\mathbf{D_2} - \nabla^2]\delta\mathbf{w_i} = \delta\mathbf{B}^2\mathbf{D_2}\mathbf{w_i}, \tag{A-8}$$

$$[\mathbf{B}_i^2\mathbf{D_2} - \nabla^2]\delta\mathbf{w_s} = \mathbf{P}_i^2(\tau)^{\tau}*\mathbf{D_2}\delta\mathbf{w_i}. \tag{A-9}$$

The second operator $\mathbf{T_s}$ models perturbations in the scattered wavefield caused by perturbations in the propagation of the scattered wavefield itself:

$$[\mathbf{B}_i^2\mathbf{D_2} - \nabla^2]\mathbf{w_i} = \mathbf{f}, \tag{A-10}$$

$$[\mathbf{B}_i^2\mathbf{D_2} - \nabla^2]\mathbf{w_s} = \mathbf{P}_i^2(\tau)^{\tau}*\mathbf{D_2}\mathbf{w_i}, \tag{A-11}$$

$$[\mathbf{B}_i^2\mathbf{D_2} - \nabla^2]\delta\mathbf{w_s} = \delta\mathbf{B}^2\mathbf{D_2}\mathbf{w_s}. \tag{A-12}$$

Both of these tomographic operators depend nonlinearly on $\mathbf{B}_i$ and linearly on $\mathbf{P}_i^2$. They have zero output when $\mathbf{P}_i^2$ is equal to zero; that is, at the first iteration of the inner loop because we set $\mathbf{P}_o^2 = 0$ as a starting model of the inner iterations. However, as we update the linearization at each iteration, starting from the second iteration the output of $\mathbf{T}$ becomes different from zero.

The data-space tomographic operator represented by equations A-7–A-9 and equations A-10–A-12 is analogous to the WEMVA operator (Biondi and Sava, 1999; Sava and Vlad, 2008) except that in the WEMVA operator, we keep the data fixed and vary the image; it is the other way around in this tomographic operator.

REFERENCES

Almomin, A., and B. Biondi, 2012, Tomographic full waveform inversion: Practical and computationally feasible approach: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: 10.1190/segam2012-0976.1.

Bamberger, A., G. Chavent, C. Hemon, and P. Lailly, 1982, Inversion of normal incidence seismograms: Geophysics, **47**, 757–770, doi: 10.1190/1.1441345.

Biondi, B., and A. Almomin, 2012, Tomographic full waveform inversion (TFWI) by combining full waveform inversion with wave-equation migration velocity analysis: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: 10.1190/segam2012-0275.1.

Biondi, B., and P. Sava, 1999, Wave-equation migration velocity analysis: 69th Annual International Meeting, SEG, Expanded Abstracts, 1723–1726.

Biondi, B., and W. W. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging: Geophysics, **69**, 1283–1298, doi: 10.1190/1.1801945.

Dahlen, F. A., S. H. Hung, and G. Nolet, 2000, Frechet kernels for finite-frequency traveltimes — I. Theory: Geophysical Journal International, **141**, 157–174, doi: 10.1046/j.1365-246X.2000.00070.x.

Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, **56**, 645–653, doi: 10.1190/1.1443081.

Martin, G., K. Marfurt, and S. Larsen, 2002, Marmousi–2: An updated model for the investigation of AVO in structurally complex areas: 72nd Annual International Meeting, SEG, Expanded Abstracts, 1979–1982.

Mora, P., 1989, Inversion = migration + tomography: Geophysics, **54**, 1575–1586, doi: 10.1190/1.1442625.

Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: Mathematics of Computation, **35**, no. 151, 773–782, doi: 10.1090/S0025-5718-1980-0572855-7.

Sava, P., and B. Biondi, 2004a, Wave-equation migration velocity analysis — I: Theory: Geophysical Prospecting, **52**, 593–606, doi: 10.1111/j.1365-2478.2004.00447.x.

Sava, P., and B. Biondi, 2004b, Wave-equation migration velocity analysis — II: Examples: Geophysical Prospecting, **52**, 607–623, doi: 10.1111/j.1365-2478.2004.00448.x.

Sava, P., and I. Vlad, 2008, Numeric implementation of wave-equation migration velocity analysis operators: Geophysics, **73**, no. 5, VE145–VE159, doi: 10.1190/1.2953337.

Shen, P., and W. W. Symes, 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, no. 5, VE49–VE59, doi: 10.1190/1.2972021.

Sun, D., and W. Symes, 2012, Waveform inversion via nonlinear differential semblance optimization: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: 10.1190/segam2012-1190.1.

Symes, W. W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790, doi: 10.1111/j.1365-2478.2008.00698.x.

Symes, W. W., and J. J. Carazzone, 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 654–663, doi: 10.1190/1.1443082.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266, doi: 10.1190/1.1441754.

Woodward, M., 1990, Wave equation tomography: Ph.D. thesis, Stanford University.

Yang, T., and P. Sava, 2009, Wave-equation migration velocity analysis using extended images: 79th Annual International Meeting, SEG, Expanded Abstracts, 3715–3719.

Yang, T., and P. Sava, 2010, Moveout analysis of wave-equation extended images: Geophysics, **75**, no. 4, S151–S161, doi: 10.1190/1.3460296.