

A new optimization approach for source-encoding full-waveform inversion

Peyman P. Moghaddam¹, Henk Keers², Felix J. Herrmann³, and Wim A. Mulder⁴

ABSTRACT

Waveform inversion is the method of choice for determining a highly heterogeneous subsurface structure. However, conventional waveform inversion requires that the wavefield for each source is computed separately. This makes it very expensive for realistic 3D seismic surveys. Source-encoding waveform inversion, in which the sources are modeled simultaneously, is considerably faster than conventional waveform inversion but suffers from artifacts. These artifacts can partly be removed by assigning random weights to the source wavefields. We found that the misfit function, and therefore also its gradient, for source-encoding waveform inversion is an unbiased random estimation of the misfit function used in conventional waveform

inversion. We found a new method of source-encoding waveform inversion that takes into account the random nature of the gradients used in the optimization. In this new method, the gradient at each iteration is a weighted average of past gradients such that the most recent gradients have the largest weights with exponential decay. This way we damped the random fluctuations of the gradient by incorporating information from the previous iterations. We compared this new method with existing source-encoding waveform inversion methods as well as conventional waveform inversion and found that the model misfit reduction is faster and smoother than those of existing source-encoding waveform inversion methods, and it approaches the model misfit reduction obtained in conventional waveform inversion.

INTRODUCTION

Since the early work of [Tarantola \(1984, 1986\)](#), full-waveform inversion (FWI) has received little attention. This changed relatively recently, when the oil and service industries started to use this method on a large scale ([Virieux and Operto, 2009](#); [Kapoor et al., 2010](#); [Vigh et al., 2010](#)). FWI's ability to directly estimate the velocity model with minimal preprocessing and to use most of the recorded data is appealing. Two of the main drawbacks of FWI are the requirement to have an accurate initial model and its computational cost. If the initial velocity model is not accurate, FWI is likely to converge to the nearest local minimum, which can be quite far from the global minimum. We will not address this important subject but instead assume that a reasonably accurate initial velocity model is available and that computational cost is the main problem, especially in 3D. In conventional FWI ([Tarantola, 1984, 1986](#); [Mora,](#)

[1987](#); [Cruse et al., 1990](#); [Pratt and Worthington, 1990](#); [Sun and McMechan, 1992](#); [Bunks et al., 1995](#); [Pratt et al., 1996, 1998, 2001](#); [Djikpéssé and Tarantola, 1999](#); [Shipp and Singh, 2002](#); [Causse, 2002](#); [Operto et al., 2004](#); [Mulder and Plessix, 2004](#); [Sirgue and Pratt, 2004](#); [Soubier et al., 2008](#); [Symes, 2008](#); [Virieux and Operto, 2009](#)), the seismic simulations are performed for each individual seismic source separately. Therefore, the cost of conventional FWI is proportional to the number of sources. In this paper, we consider a significant reduction of the calculation time using source encoding ([Krebs et al., 2009](#); [Li and Herrmann, 2010](#); [Moghaddam and Herrmann, 2010](#); [van Leeuwen et al., 2011](#); [Haber et al., 2012](#); [Li et al., 2012](#)). Source encoding uses a linear combination of all shots, with random weights assigned to each shot. [Krebs et al. \(2009\)](#) initially perform source encoding as part of FWI by applying a +1 or -1 weight randomly to each shot. This

Manuscript received by the Editor 13 March 2012; revised manuscript received 15 January 2013; published online 10 April 2013.

¹Delft University of Technology, Faculty of Civil Engineering and Geosciences, Department of Geotechnology, Delft, The Netherlands. E-mail: p.poormoghaddam@tudelft.nl; wim.mulder@shell.com.

²University of Bergen, Faculty of Mathematics and Natural Sciences, Department of Earth Science, Bergen, Norway. E-mail: henk.keers@geo.uib.no.

³The University of British Columbia, Faculty of Science, Department of Earth, Ocean and Atmospheric Sciences, Vancouver, British Columbia, Canada. E-mail: fherrmann@eos.ubc.ca.

⁴Delft University of Technology, Faculty of Civil Engineering and Geosciences, Department of Geotechnology, Delft, The Netherlands and Shell Global Solutions International BV, Rijswijk, The Netherlands. E-mail: wim.mulder@shell.com.

© 2013 Society of Exploration Geophysicists. All rights reserved.

method can be further extended because Moghaddam and Herrmann (2010), van Leeuwen et al. (2011), and Li et al. (2012) find that applying any random zero mean weights to each shot per frequency achieves the same result.

There are two open questions in the source-encoding approach: (1) What is the optimum choice for the encoding sequence? and (2) Which is the best optimization method for the solution of the problem? The first question has received wide attention (Boonyasiriwat and Schuster, 2010; Gao et al., 2010; Habashy et al., 2010; Symes, 2010; Wang and Goo, 2010). The second question has received much less attention, despite its importance (Aravkin et al., 2012; van Leeuwen and Herrmann, 2012). In an approach similar to Haber et al. (2012), van Leeuwen et al. (2011), and Aravkin et al. (2012), we prove mathematically that the misfit function for source encoded inversion is an unbiased random estimation of the true misfit function when all the shots are incorporated. The same arguments hold true for the gradient of the misfit function. Because the gradient of the FWI cost function is randomized, the step toward the solution is randomized as well. However, most authors (Krebs et al., 2009; Boonyasiriwat and Schuster, 2010; Symes, 2010) still use deterministic optimization solutions such as steepest descents or quasi-Newton approaches for the randomized FWI problem and do not consider the fact that the search direction in each iteration of the solver is random.

Stochastic optimization (Goldberg, 1989; Spall, 1992) refers to a class of iterative methods that takes the randomness of the search directions into account. They correct for the fact that the steps are random. This is done by either (1) averaging over all the past updated models, (2) averaging over past gradients, or (3) averaging the Hessian operator (Goldberg, 1989; Spall, 1992, 2003; Hoos and Stutzle, 2004). In this paper, we borrow some ideas from the area of stochastic optimization and propose our own method for the solution of the source-encoding FWI problem. We show that stochastic optimization methods perform better than the conventional optimization methods (van Leeuwen et al., 2011). We call our stochastic optimization method integrated stochastic gradient descent (iSGD), and we show its superior performance over conventional optimization.

We start with the formulation of conventional FWI, followed by an explanation of source-encoding FWI. Then, an overview of stochastic optimization and the algorithm for source-encoding FWI is presented. The Marmousi model (Bourgeois et al., 1991) serves as a test case for the various methods. We end with a summary of the main conclusions and a discussion of future directions.

FULL-WAVEFORM INVERSION

FWI in the frequency domain can be defined as the minimization of the cost function

$$J(\sigma) = \frac{1}{2} \|\mathbf{d} - \mathbf{D}\mathbf{u}(\sigma)\|_2^2, \quad (1)$$

where \mathbf{d} denotes the observed data and \mathbf{D} is the detection operator that samples the wavefield \mathbf{u} at the receiver for a given shot in a slowness model $\sigma(\mathbf{x})$ as a function of position \mathbf{x} . The minimization of the cost function is subject to a discretized solution of the scalar wave equation assuming a constant density

$$\omega^2 \sigma^2 \mathbf{u} + \Delta \mathbf{u} + \mathbf{f} = 0, \quad (2)$$

where $\mathbf{u}(\omega, \mathbf{x})$ is the wavefield, $\mathbf{f}(\omega, \mathbf{x})$ is the source function in the frequency domain, ω is the angular frequency, and Δ is the Laplacian operator.

To minimize $J(\sigma)$, we need to calculate its gradient with respect to the slowness σ . The gradient with respect to the slowness $\sigma(\mathbf{x})$ at some point is

$$\partial J(\sigma) / \partial \sigma = -\text{Re}\{(\partial \mathbf{u} / \partial \sigma)^H \mathbf{D}^T [\mathbf{d} - \mathbf{D}\mathbf{u}(\sigma)]\}, \quad (3)$$

where $\text{Re}\{\cdot\}$ denotes the real part, $(\cdot)^H$ is the Hermitian or conjugate transpose of a matrix, and $(\cdot)^T$ is the transpose. The computation of the sensitivity matrix $\partial \mathbf{u} / \partial \sigma$ can be avoided by using the adjoint-state method (Lions and Magenes, 1972; Lailly, 1983; Tarantola, 1984; Giles et al., 2003; Plessix, 2006; Virieux and Operto, 2009). Having calculated the gradient of the cost function in equation 3, the cost function can be minimized using one of various optimization methods such as the limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) method (Byrd et al., 1995; Mulder and Plessix, 2004; Nocedal and Wright, 2006; Plessix, 2009), a preconditioned conjugate gradient method (Ravaut et al., 2004), Gauss-Newton (Virieux and Operto, 2009), and so on. The result for conventional FWI in this paper is obtained with the LBFGS method. Currently, the determination of the best optimization strategy for FWI is an area of active research.

Following Virieux and Operto (2009), the step length η for a computed update direction $\delta \sigma$ can be found by an additional demigration and solving

$$\min_{\eta} \frac{1}{2} \|\mathbf{d} - \mathbf{D}\mathbf{u}(\sigma + \eta \delta \sigma)\|_2^2. \quad (4)$$

Instead of computing the wavefield $\mathbf{u}(\sigma + \eta \delta \sigma)$, we use the Born approximation

$$\omega^2 \sigma^2 \tilde{\mathbf{u}} + \Delta \tilde{\mathbf{u}} + 2\sigma \delta \sigma \omega^2 \mathbf{u} = 0 \quad (5)$$

to find $\tilde{\mathbf{u}}$ and approximate $\mathbf{u}(\sigma + \eta \delta \sigma) \simeq \mathbf{u}(\sigma) + \eta \tilde{\mathbf{u}}$. The solution of equation 4 then becomes

$$\hat{\eta} = \frac{\mathbf{a}^T \mathbf{b}}{\mathbf{b}^T \mathbf{b}}, \quad (6)$$

with $\mathbf{a} = \mathbf{d} - \mathbf{D}\mathbf{u}(\sigma)$ and $\mathbf{b} = \mathbf{D}\tilde{\mathbf{u}}$.

SOURCE-ENCODING FWI

In source-encoding FWI (Moghaddam and Herrmann, 2010; van Leeuwen et al., 2011), the inversion is performed on a linear combination of all shots, called a supershot, rather than on each individual shot separately and summing the result. Each individual shot contributes to the supershot with a random weight that changes for each iteration during the optimization. Note that the receiver positions should be the same for all shots. This condition is relaxed in the recent work of van Leeuwen and Herrmann (2012) and Choi and Alkhalifah (2011). For each single frequency, the supershot form of FWI has the cost function

$$J_s(\sigma) = \frac{1}{2} \|\mathbf{M}[\mathbf{d} - \mathbf{D}\mathbf{u}(\sigma)]\|_2^2, \quad (7)$$

with \mathbf{d} being the observed data for all shots and \mathbf{u} being the wavefield for these shots. The random weighting matrix associated with the supershot experiment can be written as

$$\mathbf{M} = \mathbf{w} \otimes \mathbf{I}, \quad (8)$$

where \mathbf{I} is the identity matrix of size $N_r \times N_r$, with N_r being the number of receivers. $\mathbf{w}_{1 \times N}$ (N is the number of shots) is a random Gaussian matrix with entries drawn from a normal distribution with zero mean and standard deviation 1, i.e., $\mathbf{N}(0, 1)$, and \otimes is the Kronecker product. Basically, a supershot is a linear combination of all shots with a random weight associated with each shot. The cost function in equation 7 becomes

$$\begin{aligned} J_s(\sigma) &= \frac{1}{2N_s} [\mathbf{d} - \mathbf{D}\mathbf{u}(\sigma)]^T \mathbf{M}^T \mathbf{M} [\mathbf{d} - \mathbf{D}\mathbf{u}(\sigma)] \\ &= \frac{1}{2N_s} [\mathbf{d} - \mathbf{D}\mathbf{u}(\sigma)]^T \mathbf{I} \otimes \mathbf{w}^T \mathbf{w} \otimes \mathbf{I} [\mathbf{d} - \mathbf{D}\mathbf{u}(\sigma)]. \end{aligned} \quad (9)$$

The expected value of $\mathcal{E}\{\mathbf{w}^T \mathbf{w}\} = N_s \mathbf{I}$. The expected value of the cost function in equation 7 satisfies

$$\begin{aligned} \mathcal{E}\{J_s(\sigma)\} &= \frac{1}{2} [\mathbf{d} - \mathbf{D}\mathbf{u}(\sigma)]^T [\mathbf{d} - \mathbf{D}\mathbf{u}(\sigma)] \\ &= \frac{1}{2} \sum_{i=1}^N \|\mathbf{d}_i - \mathbf{D}\mathbf{u}_i(\sigma)\|_2^2, \end{aligned} \quad (10)$$

with N being the number of shots; \mathbf{d}_i the recorded data; and $\mathbf{u}_i(\sigma)$ the wavefield, for the i th shot (see also Haber et al. [2012] and Aravkin et al. [2012]). This is the cost function associated with all the shots. The above shows that each randomized realization of the cost function is an unbiased random estimate of the cost function associated with conventional full waveform inversion in which the shots are modeled sequentially. The gradient of the randomized cost function has the same property.

STOCHASTIC OPTIMIZATION

The minimization method for randomized inversion is called stochastic optimization and is based on machine learning methods (Goldberg, 1989; Spall, 1992, 2003; Hoos and Stutzle, 2004; Yu et al., 2010). In the stochastic optimization framework, the search direction in each iteration of the inverse problem is considered to be a random but educated guess that should lead toward the true solution. However, special measures are needed to ensure that this randomized optimization method will converge to the true solution (van Leeuwen et al., 2011).

Source-encoding FWI uses the randomized supershot experiment, in which the randomized weights are different for each iteration. As discussed in the previous section, the expectation of each realization of the gradient of the function at each iteration can be shown to be equal to the gradient obtained when all the shots are treated independently and not randomly. This makes stochastic optimization methods appealing because they are constructed using the fact that the gradient toward the solution is a random perturbation of the actual gradient. Table 1 describes the source-encoding FWI method, assuming that $J(\sigma, \mathbf{w}_i)$ is the cost function associated with the i th set of random weights \mathbf{w}_i .

In this section, we briefly explain two conventional optimization methods, steepest descent and LBFSG, for the source-encoding problem. Then we review a recent stochastic optimization solver that is called online LBFSG (oLBFSG) (Schraudolph et al., 2007; Yu et al., 2010) and is based on smoothing the Hessian. Finally, we propose our own stochastic optimization method that we call iSGD. Contrary to the recently proposed method by Aravkin et al. (2012), we do not rely on controlling the random error by growing the number of randomized supershots, but instead we damp the random fluctuations of the gradient by incorporating information from previous iterations in the updates.

Stochastic gradient descent

Stochastic gradient descent (Schraudolph et al., 2007; Sunebag et al., 2009) takes the following form:

$$\sigma_{k+1} = \sigma_k - \eta_k \nabla J(\sigma_k, \mathbf{w}_k), \quad (11)$$

where k is the iteration number, η_k is the step length, J is the misfit function, σ_k is the model at iteration k , and \mathbf{w}_k is the current randomized weight. The step length η_k can be obtained by a line search.

Stochastic LBFSG

Stochastic LBFSG is regular LBFSG with randomized gradients. Each step of the LBFSG algorithm has the following form:

$$\sigma_{k+1} = \sigma_k - \eta_k \mathbf{H}_k \nabla J(\sigma_k, \mathbf{w}_k), \quad (12)$$

where η_k is the step length, \mathbf{H}_k is the inverse Hessian matrix updated in each iteration,

$$\mathbf{H}_{k+1} = \mathbf{V}_k^T \mathbf{H}_k \mathbf{V}_k + \rho_k \mathbf{s}_k \mathbf{s}_k^T, \quad (13)$$

with $\rho_k = \frac{1}{\mathbf{y}_k^T \mathbf{s}_k}$, $\mathbf{V}_k = \mathbf{I} - \rho_k \mathbf{y}_k \mathbf{s}_k^T$, and $\mathbf{s}_k = \sigma_{k+1} - \sigma_k$, $\mathbf{y}_k = \nabla J(\sigma_{k+1}, \mathbf{w}_k) - \nabla J(\sigma_k, \mathbf{w}_k)$. Note that for construction of \mathbf{y}_k , we use the same random weighting \mathbf{w}_k for the current gradient $J(\sigma_{k+1})$ and the previous one at k . This is crucial for the convergence of the LBFSG method with source encoding (Schraudolph et al., 2007). An alternative approach suggested by Guitton and Diaz (2012) is to keep the weights and the encoding sequence constant during several iterations.

The LBFSG routine is carried out in two steps. First, the vectors $\mathbf{y}_{k-m}, \dots, \mathbf{y}_k$ and $\mathbf{s}_{k-m}, \dots, \mathbf{s}_k$ at the latest $m+1$ iterations are calculated, where $m=10$ is the buffer size for the LBFSG routine. Second, the routine that updates the LBFSG direction (see Table 2) is applied, with $\mathbf{H}_k^0 = (\mathbf{y}_k^T \mathbf{s}_k) / (\mathbf{y}_k^T \mathbf{y}_k)$.

Table 1. Algorithm for source-encoding FWI.

STATE Set: $\sigma = \sigma_0$, initial model;
WHILE $\|\nabla_{\sigma} J(\sigma, \mathbf{w}_i)\| \geq \epsilon$
STATE $\{\nabla_{\sigma} J(\sigma, \mathbf{w}_i)\} \leftarrow$ Compute the gradient for a new randomized supershot
STATE $\sigma \leftarrow$ Update the model with the new gradient using stochastic optimization
END WHILE

Stochastic oLBFGS

The oLBFGS method (Schraudolph et al., 2007; Yu et al., 2010) is different from regular LBFGS in two ways. It uses $\mathbf{y}_k = \nabla J(\mathbf{m}_{k+1}, \mathbf{w}_k) - \nabla J(\mathbf{m}_k, \mathbf{w}_k) + \lambda \mathbf{s}_k$ with an additional term $\lambda \mathbf{s}_k$ for better convergence. Step $\mathbf{r} \leftarrow \mathbf{H}_k^0 \mathbf{q}$ in Table 2 is replaced by

$$\mathbf{r} = \frac{\mathbf{q}}{\min(k, m)} \sum_{i=1}^{\min(k, m)} \frac{\mathbf{s}_{k-i}^T \mathbf{y}_{k-i}}{\mathbf{y}_{k-i}^T \mathbf{y}_{k-i}}. \quad (14)$$

Schraudolph et al. (2007) and Yu et al. (2010) claim that these changes smooth the Hessian and average away sampling noise by diagonal loading and therefore minimize the random behavior of the search direction. In our simulations, we set $\lambda = 0.1 \|\mathbf{J}(\mathbf{m}_0, \mathbf{w}_0)\|_2^2 / \|\mathbf{m}_0\|_2^2$.

Integrated stochastic gradient descent

Because the gradients are randomized, we propose averaging of the gradients over the past iterations. However, simple averaging is not a good idea for two reasons. First, the current gradient probably provides a more accurate description of the current update than the

Table 2. LBFGS algorithm, inner loop.

```

 $\mathbf{q} \leftarrow \nabla J(\mathbf{m}_k, \mathbf{w}_k)$ 
FOR  $i = k - 1$  to  $k - m$ 
STATE  $\alpha_i \leftarrow \rho_i \mathbf{s}_i^T \mathbf{q}$ 
STATE  $\mathbf{q} \leftarrow \mathbf{q} - \alpha_i \mathbf{y}_i$ 
END FOR
STATE  $\mathbf{r} \leftarrow \mathbf{H}_k^0 \mathbf{q}$ 
FOR  $i = k - m$  to  $k - 1$ 
STATE  $\beta \leftarrow \rho_i \mathbf{y}_i^T \mathbf{r}$ 
STATE  $\mathbf{r} \leftarrow \mathbf{r} + \mathbf{s}_i (\alpha_i - \beta)$ 
END FOR
STATE stop with result  $\mathbf{r} = \mathbf{H}_k \nabla J(\mathbf{m}_k, \mathbf{w}_k)$ 

```

gradients of past iterations. Second, the convergence of stochastic gradient descent decreases when we average over previous model iterations as proposed by van Leeuwen et al. (2011). Although this latter averaging method guarantees convergence, it slows down the convergence, which we would like to avoid.

To overcome this slowdown, we suggest a weighted averaging of the past gradients as a good candidate for the true gradient. We explored various schemes for weighted averages and found that exponential weighting among previous gradients gave the best performance. Therefore, we suggest an exponential weighting for the averaging. In this approach, called iSGD, the iteration step takes into account the past history of the gradients and has the form

$$\sigma_{k+1} = \sigma_k - \eta_k \overline{\nabla J(\sigma_k)}, \quad (15)$$

with $\overline{\nabla J(\sigma_k)}$ defined as

$$\overline{\nabla J(\sigma_k)} = \frac{\sum_{i=k-m}^k e^{\alpha(i-k)} \nabla J(\sigma_i, \mathbf{w}_i)}{\sum_{i=k-m}^k e^{\alpha(i-k)}}. \quad (16)$$

Here, m is the number of earlier steps. Note that for $\alpha = 0$, $\overline{\nabla J(\sigma_k)}$ is an average of the gradients over past m iterations. We set the value of $m = 10$ during our simulations. As before, the value of η_k is obtained by a line search. We will determine the optimum value of α heuristically in the next section.

EXAMPLES

Subset of the Marmousi model

We perform a test for constant-density acoustic FWI on a subset of the Marmousi model shown in Figure 1. The subset is 5-km long and 3-km deep with a 10-m grid spacing. The source wavelet for all the shots is a Ricker wavelet with a 10-Hz central frequency.

The synthetic data are generated with 125 shots at lateral positions between 40 and 4960 m with a 40-m shot spacing at 50-m depth and with 250 receivers at lateral positions between 20 and 4990 m and a 20-m receiver spacing at 20-m depth. The acquisition is a fixed spread with all the receivers at the same location for each shot. The total recording time is 3.6 s with a time sampling of

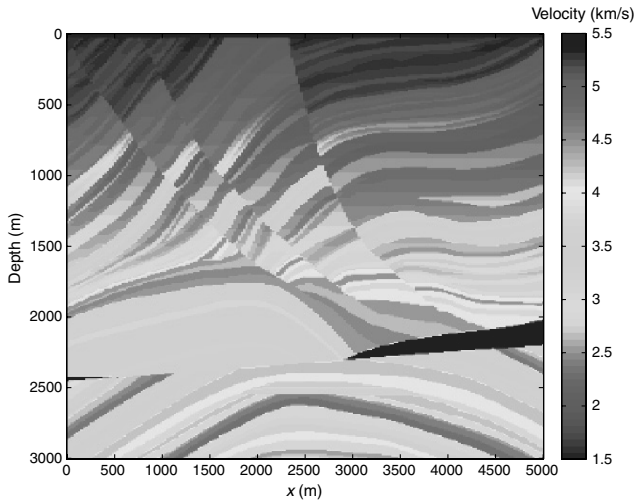


Figure 1. Subset of the Marmousi model.

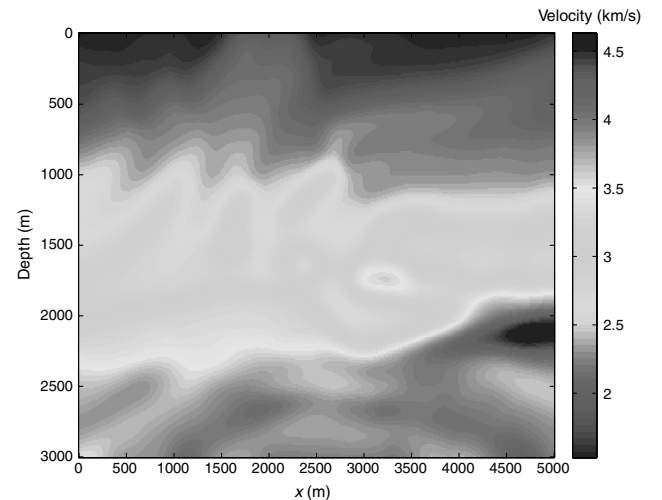


Figure 2. Initial model.

0.9 ms, which is a smooth version of the true velocity model. We first invert for the velocity model using conventional waveform inversion, with all sequential shots separately, in the frequency band 3–30 Hz with a 0.5-Hz increment. We then run the source-encoding optimization methods on the same data. We stopped the computation after 1/40th times the calculation time of the conventional approach, which has 125 separate shots, and considered the error in the reconstructed model. The reduction time of 1/40 is chosen heuristically. We stopped the computation after the misfit reduction of the conventional (or “sequential”) waveform inversion was more than 90. The misfit reduction of our new method iSGD at this iteration was 85. It is not entirely true that each additional shot adds another Helmholtz inversion to the computational cost of conventional FWI because we can save the LU decomposition and apply it to multiple sources at the same time. However, in the case of large 2D models or 3D models, it is not possible to do this. Our evaluation of speed up in this paper is based on the assumption that the inverse of the operator is applied implicitly to each shot separately. Figure 1

shows the subset of the model used for our tests. The initial model for the inversion is displayed in Figure 2. To evaluate the quality of the results, we calculate the relative least-squares error (RLSE) = $(\|\delta\mathbf{m} - \bar{\delta\mathbf{m}}\|_2^2 / \|\bar{\delta\mathbf{m}}\|_2^2)$ with $\bar{\delta\mathbf{m}}$ the reference result shown in Figure 3, which is a region of the true model with a good source and receiver coverage, and $\delta\mathbf{m}$ the result of each experiment in that region. Figure 4 depicts the inversion result for all shots, treated separately, after 30 iterations of the LBFGS method with RLSE = 0.1.

Figures 5 to 8 show the inversion results for the various stochastic optimization methods described in the previous section. The required computational time for each of these results is up to 40 times smaller than that for the conventional result of Figure 4. This reduction is obtained by decreasing the number of shots from 125 to 1 — i.e., one random supershot, and increasing the number of source-encoding optimizer iterations from 30 to 95.

Figure 5 shows the inversion results for the stochastic gradient descent algorithm with RLSE = 0.54, Figure 6 shows those for

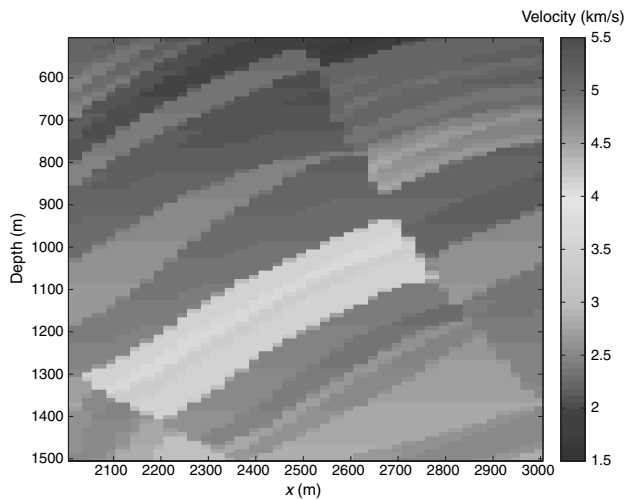


Figure 3. A subset of the velocity model in which the RLSE is calculated.

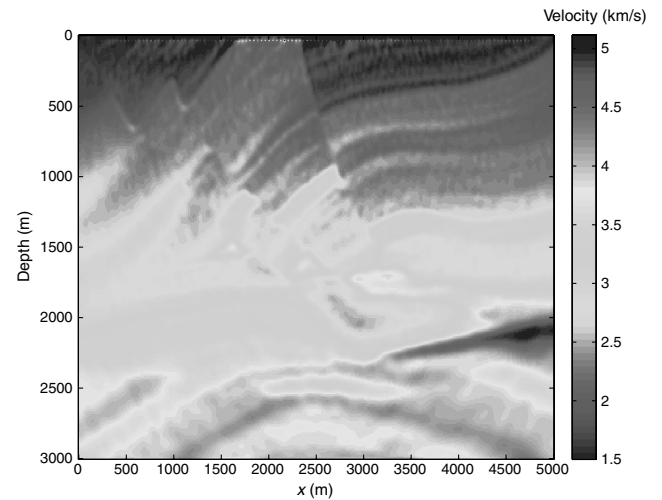


Figure 5. Inverted model based on the stochastic gradient descent method (RLSE = 0.54).

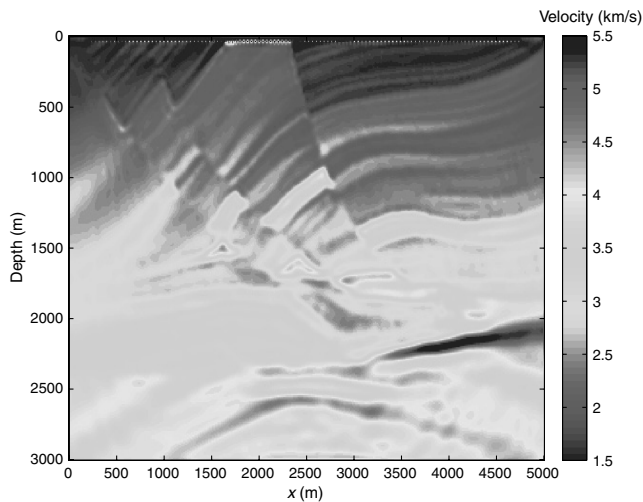


Figure 4. Inverted model after 30 iterations of LBFGS, all shots; 54 frequencies were used from 3 to 30 Hz with a 0.5-Hz increment.

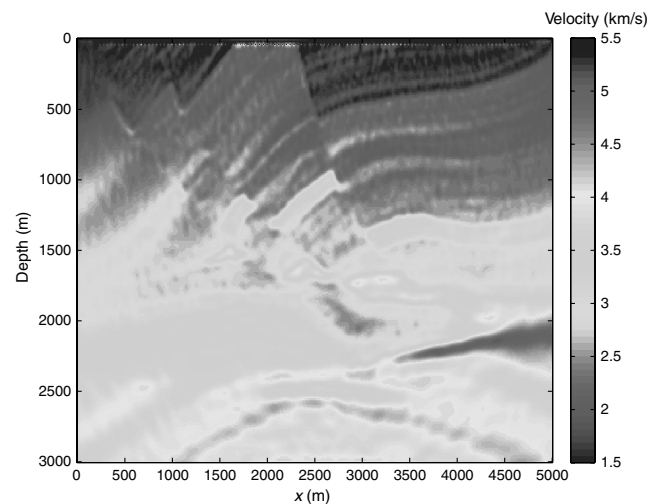


Figure 6. Inverted model using stochastic LBFGS (RLSE = 0.39).

stochastic LBFGS with RLSE = 0.39, and Figure 7 shows the inversion results for oLBFGS with RLSE = 0.22. To find the optimum value of the history parameter α for the iSGD method, we run a test over a range of values for α . Table 3 lists the resulting RLSE. The RLSE for the stochastic gradient descent algorithm is 0.54. The results in Table 3 show that a reduction in the calculation time of a factor of 40 is obtained when compared to conventional FWI.

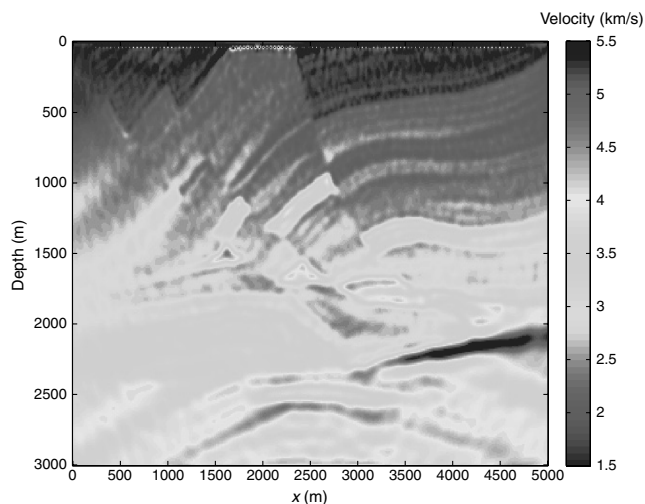


Figure 7. Inverted model using the oLBFGS method (RLSE = 0.22).

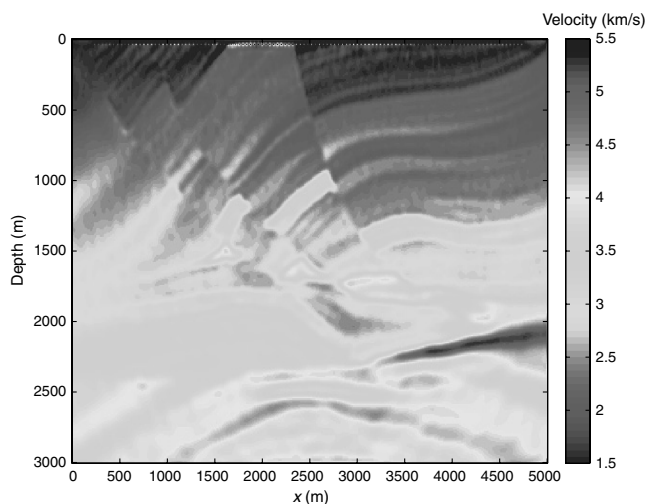


Figure 8. Inverted model using the iSGD method (RLSE = 0.15).

Table 3. Signal-to-noise ratios, $RLSE = \|\delta m - \overline{\delta m}\|_2^2 / \|\overline{\delta m}\|_2^2$ for reconstructions with the iSGD algorithm for different exponential weighting α .

α	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
RLSE	0.52	0.41	0.25	0.21	0.19	0.15	0.18	0.22	0.29	0.36

Figure 8 displays the inversion result obtained with the iSGD algorithm for $\alpha = 0.5$ with RLSE = 0.17. Figure 9 plots RLSE versus the number of iterations for different optimization methods. We observe from this figure that the iSGD method has the smallest error RLSE among the various stochastic optimization methods. Therefore, it is our preferred choice.

One might think that the result in Figure 7 is better than that of Figure 8 because the velocity under the faults is more accurate. However, this method has a higher RLSE. The reason for this is that the amount of crosstalk left is larger with LBFGS than with iSGD, as is evident from the figures.

To investigate whether the optimum value of α depends on frequency, we perform another test. We divide the frequency band 0–30 Hz into five different subbands, and we invert the model for each subband separately, using the initial model shown in Figure 2. As we see from Figure 10, the RLSE for each subband is generally smallest around $\alpha = 0.5$, suggesting that this value is a reasonable choice for our proposed stochastic inversion method.

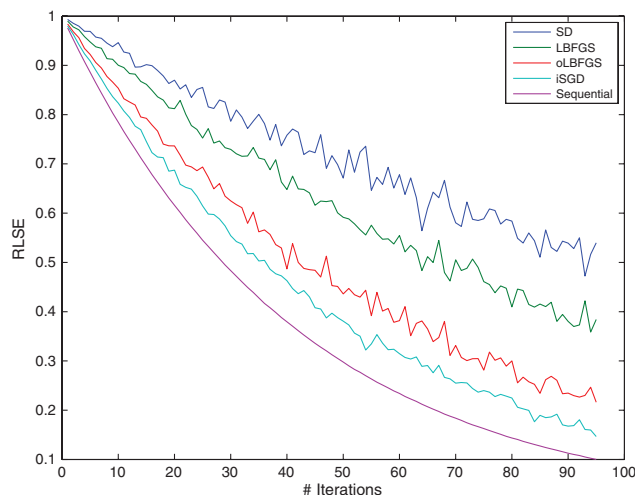


Figure 9. RLSE versus number of iterations for different optimization methods.

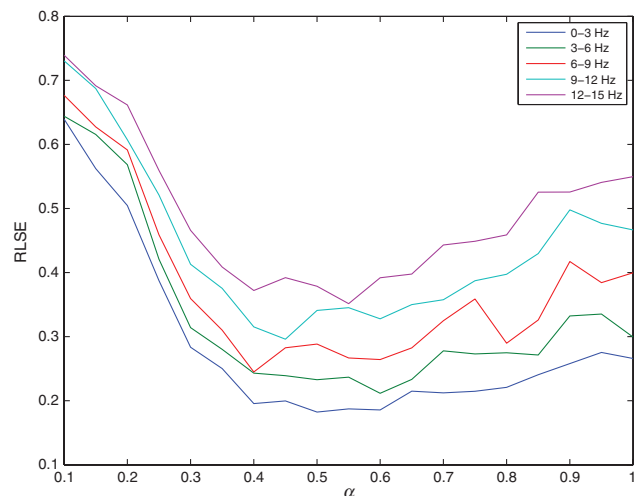


Figure 10. RLSE versus α for different frequency band inversion.

CONCLUSIONS

The computational cost for 3D FWI with thousands of shots is expensive, even on today's computers. Source-encoding FWI reduces the number of forward-modeling runs (i.e., n_f), whereas accuracy can be preserved by increasing the number of iterations (i.e., n_i). Thus source-encoding FWI is n_f/n_i times faster than conventional FWI. As in practice with a large number of sources (typically more than 100), n_f is considerably larger than n_i , which makes source-encoding FWI a viable alternative to conventional FWI.

Here, we showed that each realization of the misfit, as well as its gradient, in a source-encoding scheme is an unbiased estimate of the misfit when all sources are treated independently. This property is essential for stochastic optimization methods.

We considered various stochastic optimization methods for the solution of source-encoding FWI. The optimization methods performed differently in terms of their convergence rate. We proposed a new stochastic optimization algorithm, the iSGD method, which exponentially weights and averages the gradients of the past iterations. We showed that this performs better than known methods in terms of convergence rate, smoothness of the misfit reduction, and accuracy of the solution in an example.

ACKNOWLEDGMENTS

This work was partially supported by the Seismic Laboratory of Imaging and Modeling at the University of British Columbia, Vancouver, Canada, the Geoscience Institute of the University of Bergen, Norway, and the Department of Geotechnology at Delft University of Technology, the Netherlands.

REFERENCES

- Aravkin, A. Y., M. P. Friedlander, F. J. Herrmann, and T. van Leeuwen, 2012, Robust inversion, dimensionality reduction, and randomized sampling: *Mathematical Programming*, **134**, 101–125, doi: [10.1007/s10107-012-0571-6](https://doi.org/10.1007/s10107-012-0571-6).
- Boonyasiriwat, C., and G. T. Schuster, 2010, 3D multisource full-waveform inversion using dynamic random phase encoding: 80th Annual International Meeting, SEG, Expanded Abstracts, 1044–1049.
- Bourgeois, A., M. Bourget, P. Lailly, M. Poulet, P. Ricarte, and R. Versteeg, 1991, Marmousi, model and data: Presented at EAEG workshop on practical aspects of seismic data inversion.
- Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: *Geophysics*, **60**, 1457–1473, doi: [10.1190/1.1443880](https://doi.org/10.1190/1.1443880).
- Byrd, R. H., P. Lu, J. Nocedal, and C. Y. Zhu, 1995, A limited memory algorithm for bound constrained optimization: *SIAM Journal on Scientific Computing*, **16**, 1190–1208, doi: [10.1137/0916069](https://doi.org/10.1137/0916069).
- Causse, E., 2002, Avoiding pitfalls of least-squares inversion by using the energy-flux error criterion: *Geophysical Prospecting*, **50**, 317–327, doi: [10.1046/j.1365-2478.2002.00320.x](https://doi.org/10.1046/j.1365-2478.2002.00320.x).
- Choi, Y., and T. Alkhalifah, 2011, Source-independent time-domain waveform inversion using convolved wavefields: Application to the encoded multi-source waveform inversion: *Geophysics*, **76**, no. 5, R125–R134, doi: [10.1190/geo2010-0210.1](https://doi.org/10.1190/geo2010-0210.1).
- Cruse, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola, 1990, Robust elastic nonlinear waveform inversion: Application to real data: *Geophysics*, **55**, 527–538, doi: [10.1190/1.1442864](https://doi.org/10.1190/1.1442864).
- Djokpéssé, H., and A. Tarantola, 1999, Multiparameter L1 norm waveform fitting: Interpretation of Gulf of Mexico reflection seismograms: *Geophysics*, **64**, 1023–1035, doi: [10.1190/1.1444611](https://doi.org/10.1190/1.1444611).
- Gao, F., A. Atle, and P. Williamson, 2010, Full waveform inversion using deterministic source encoding: 80th Annual International Meeting, SEG, Expanded Abstracts, 1013–1017.
- Giles, M. B., M. C. Duta, J. Müller, and N. A. Pierce, 2003, Algorithm developments for discrete adjoint methods: *AIAA Journal*, **41**, 198–205, doi: [10.2514/2.1961](https://doi.org/10.2514/2.1961).
- Goldberg, D., 1989, Genetic algorithms in search, optimization, and machine learning: Addison-Wesley.
- Guittou, A., and E. Diaz, 2012, Attenuating crosstalk noise with simultaneous source full waveform inversion: *Geophysical Prospecting*, **60**, 759–768, doi: [10.1111/j.1365-2478.2011.01023.x](https://doi.org/10.1111/j.1365-2478.2011.01023.x).
- Habashy, T. M., A. Abubakar, G. Pan, and A. Belani, 2010, Full-waveform seismic inversion using the source-receiver compression approach: 80th Annual International Meeting, SEG, Expanded Abstracts, 1023–1028.
- Haber, E., M. Chung, and F. J. Herrmann, 2012, An effective method for parameter estimation with pde constraints with multiple right-hand sides: *SIAM Journal on Optimization*, **22**, 739–757, doi: [10.1137/11081126X](https://doi.org/10.1137/11081126X).
- Hoos, H., and T. Stutzle, 2004, Stochastic local search: Foundations and applications: Morgan Kaufmann/Elsevier.
- Kapoor, S., N. Moldoveanu, D. Vigh, and H. Li, 2010, The quest for subsalt imaging — The next chapter: 72nd Annual International Conference and Exhibition, EAGE, Extended Abstracts, K041.
- Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse, 2009, Fast full-wavefield seismic inversion using encoded sources: *Geophysics*, **74**, no. 6, WCC177–WCC188, doi: [10.1190/1.3230502](https://doi.org/10.1190/1.3230502).
- Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migration: Presented at Conference on Inverse Scattering, Theory and Applications, SIAM, 206–220.
- Li, X., A. Y. Aravkin, T. van Leeuwen, and F. J. Herrmann, 2012, Fast randomized full-waveform inversion with compressive sensing: *Geophysics*, **77**, no. 3, A13–A17, doi: [10.1190/geo2011-0410.1](https://doi.org/10.1190/geo2011-0410.1).
- Li, X., and F. J. Herrmann, 2010, Full-waveform inversion from compressively recovered model updates: 80th Annual International Meeting, SEG, Expanded Abstracts, 1029–1033.
- Lions, J., and E. Magenes, 1972, Nonhomogeneous boundary value problems and applications: Springer Verlag.
- Moghaddam, P. P., and F. J. Herrmann, 2010, Randomized full-waveform inversion: A dimensionality-reduction approach: 80th Annual International Meeting, SEG, Expanded Abstracts, 977–982.
- Mora, P., 1987, Nonlinear two-dimensional elastic inversion of multioffset seismic data: *Geophysics*, **52**, 1211–1228, doi: [10.1190/1.1442384](https://doi.org/10.1190/1.1442384).
- Mulder, W. A., and R.-E. Plessix, 2004, A comparison between one-way and two-way wave-equation migration: *Geophysics*, **69**, 1491–1504, doi: [10.1190/1.1836822](https://doi.org/10.1190/1.1836822).
- Nocedal, J., and S. Wright, 2006, Numerical optimization: Springer Verlag.
- Operto, S., C. Ravaut, L. Improta, J. Virieux, A. Herrero, and P. Dell'Aversana, 2004, Quantitative imaging of complex structures from dense wide-aperture seismic data by multi-scale traveltime and waveform inversions: A case study: *Geophysical Prospecting*, **52**, 625–651, doi: [10.1111/j.1365-2478.2004.00452.x](https://doi.org/10.1111/j.1365-2478.2004.00452.x).
- Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: *Geophysical Journal International*, **167**, 495–503, doi: [10.1111/j.1365-246X.2006.02978.x](https://doi.org/10.1111/j.1365-246X.2006.02978.x).
- Plessix, R.-E., 2009, Three-dimensional frequency-domain full-waveform inversion with an iterative solver: *Geophysics*, **74**, no. 6, WCC149–WCC157, doi: [10.1190/1.3211198](https://doi.org/10.1190/1.3211198).
- Pratt, R., Z. Song, P. Williamson, and M. Warner, 1996, Two-dimensional velocity models from wide-angle seismic data by waveform inversion: *Geophysical Journal International*, **124**, 323–340, doi: [10.1111/j.1365-246X.1996.tb07023.x](https://doi.org/10.1111/j.1365-246X.1996.tb07023.x).
- Pratt, R. G., R.-E. Plessix, and W. A. Mulder, 2001, Seismic waveform tomography: Crosshole results from the Nimr field in Oman: 71st Annual International Meeting, SEG, Expanded Abstracts, 714–717.
- Pratt, R. G., C. Shin, and G. Hicks, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: *Geophysical Journal International*, **133**, 341–362, doi: [10.1046/j.1365-246X.1998.00498.x](https://doi.org/10.1046/j.1365-246X.1998.00498.x).
- Pratt, R. G., and M. H. Worthington, 1990, Inverse theory applied to multi-source cross-hole tomography. Part I: Acoustic wave-equation method: *Geophysical Prospecting*, **38**, 287–310, doi: [10.1111/j.1365-2478.1990.tb01846.x](https://doi.org/10.1111/j.1365-2478.1990.tb01846.x).
- Ravaut, C., S. Operto, L. Improta, J. Virieux, A. Herrero, and P. Dell'Aversana, 2004, Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: Application to a thrust belt: *Geophysical Journal International*, **159**, 1032–1056, doi: [10.1111/j.1365-246X.2004.02442.x](https://doi.org/10.1111/j.1365-246X.2004.02442.x).
- Schraudolph, N. N., J. Yu, and S. Günter, 2007, A stochastic quasi-Newton method for online convex optimization: Presented at 11th International Conference on Artificial Intelligence and Statistics.
- Shipp, R., and S. Singh, 2002, Two-dimensional full waveform inversion of wide-aperture marine seismic streamer data: *Geophysical Journal International*, **151**, 325–344, doi: [10.1046/j.1365-246X.2002.01645.x](https://doi.org/10.1046/j.1365-246X.2002.01645.x).
- Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: *Geophysics*, **69**, 231–248, doi: [10.1190/1.1649391](https://doi.org/10.1190/1.1649391).
- Soubrier, F., A. Haidar, L. Giraud, S. Operto, and J. Virieux, 2008, Frequency-domain full-waveform modeling using a hybrid direct-iterative solver based on a parallel domain decomposition method: A tool for 3D

- full-waveform inversion? 78th Annual International Meeting, SEG, Expanded Abstracts, 2147–2151.
- Spall, J., 1992, Multivariate stochastic approximation using a simultaneous perturbation gradient approximation: *IEEE Transactions on Automatic Control*, **37**, 332–341, doi: [10.1109/9.119632](https://doi.org/10.1109/9.119632).
- Spall, J., 2003, Introduction to stochastic search and optimization: Wiley.
- Sun, R., and G. McMechan, 1992, 2-D full-wavefield inversion for wide-aperture, elastic, seismic data: *Geophysical Journal International*, **111**, 1–10, doi: [10.1111/j.1365-246X.1992.tb00550.x](https://doi.org/10.1111/j.1365-246X.1992.tb00550.x).
- Sunehag, P., J. Trumpf, S. V. N. Vishwanathan, and N. Schraudolph, 2009, Variable metric stochastic approximation theory: Presented at 11th International Conference on Artificial Intelligence and Statistics, 560–566.
- Symes, W., 2008, Migration velocity analysis and waveform inversion: *Geophysical Prospecting*, **56**, 765–790, doi: [10.1111/j.1365-2478.2008.00698.x](https://doi.org/10.1111/j.1365-2478.2008.00698.x).
- Symes, W., 2010, Source synthesis for waveform inversion: 80th Annual International Meeting, SEG, Expanded Abstracts, 1018–1022.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, 1259–1266, doi: [10.1190/1.1441754](https://doi.org/10.1190/1.1441754).
- Tarantola, A., 1986, Strategy for nonlinear elastic inversion of seismic reflection data: *Geophysics*, **51**, 1893–1903, doi: [10.1190/1.1442046](https://doi.org/10.1190/1.1442046).
- van Leeuwen, T., A. Y. Aravkin, and F. J. Herrmann, 2011, Seismic waveform inversion by stochastic optimization: *International Journal of Geophysics*, **2011**, 1–18, doi: [10.1155/2011/689041](https://doi.org/10.1155/2011/689041).
- van Leeuwen, T., and F. J. Herrmann, 2012, Fast waveform inversion without source encoding: *Geophysical Prospecting*, early view, doi: [10.1111/j.1365-2478.2012.01096.x](https://doi.org/10.1111/j.1365-2478.2012.01096.x).
- Vigh, D., B. Starr, J. Kapoor, and H. Li, 2010, 3D full waveform inversion on a Gulf of Mexico WAZ data set: 80th Annual International Meeting, SEG, Expanded Abstracts, 957–961.
- Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74**, no. 6, WCC1–WCC26, doi: [10.1190/1.3238367](https://doi.org/10.1190/1.3238367).
- Wang, B., and J. Goo, 2010, Fast full waveform inversion of multi-shot seismic data: 80th Annual International Meeting, SEG, Expanded Abstracts, 1055–1058.
- Yu, J., S. Vishwanathan, S. Günter, and N. N. Schraudolph, 2010, A quasi-Newton approach to nonsmooth convex optimization problems in machine learning: *Journal of Machine Learning Research*, **11**, 1145–1200.