

Full-waveform inversion for microseismic events using sparsity constraints

Bharath Shekar¹ and Harpreet Singh Sethi²

ABSTRACT

Full-waveform inversion (FWI) is a powerful tool that can be used to invert for microseismic event locations and the source signature because it can exploit the complete waveform information. We have developed an algorithm to invert for a spatio-temporal source function that encapsulates microseismic events with spatially localized or distributed locations and source signatures. The algorithm does not require assumptions to be made about the number or type of sources; however, it does require that the velocity model is close to the true subsurface velocity. We reformulate the conventional FWI algorithm based on the l_2 -norm data-misfit function by adding sparsity constraints using a sparsity promoting l_1 -norm as an additional regularization term to get more focused and less noise-sensitive event locations. The Orthant-Wise Limited-memory quasi-Newton algorithm is used to solve the optimization problem. It inherits the advantageous (fast convergence) properties of the limited memory Broyden-Fletcher-Goldfarb-Shanno method and can easily overcome the nondifferentiability of l_1 -norm at null positions. We determine the performance of the algorithm on noise-free and noisy synthetic data from the SEG/EAGE overthrust model.

INTRODUCTION

Microseismic events are low-magnitude earthquakes that may occur throughout the production life of hydrocarbon reservoirs. They are mainly induced by petroleum extraction and associated recovery activities (water, gas, or steam injection) or by hydraulic fracturing experiments. These events occur due to local stress changes within the rock mass and have their origin in the reservoir

or in its vicinity at small fractures. Microseismic events can help us to delineate active fault network within the reservoir because they mostly occur at pre-existing faults and fractures. By examining the change in location of these events in time, a real-time 3D picture of the fluid-front movement inside the rock mass can be achieved (Rutledge and Phillips, 2003; Maxwell et al., 2010). The determination of spatial and temporal variations of microseismic events has become an important area of research in exploration seismology. Conventional methods involve picking and analyzing P- and S-wave traveltimes (Eisner et al., 2011; Zhebel and Eisner, 2014) to invert for source parameters.

Manually picking the traveltimes can be challenging if the signal-to-noise ratio of the data is low. Recently, advances have been made in using imaging and inversion-based methods that directly use waveform information. For instance, Artman et al. (2010) use time reversal and crosscorrelation imaging conditions to locate microseismic events using multicomponent data, Xuan and Sava (2010) use time-reversal methods in the context of Bayesian inversion theory to estimate the source location from data with noise. Michel and Tsvankin (2014) use full-waveform inversion (FWI) to invert for the source parameters including moment tensors for vertical transverse isotropic medium but their algorithm requires a good initial estimate of the number of sources and their approximate locations. Kaderli et al. (2015) present an FWI-based technique to invert for source locations and source wavelet without a priori knowledge about the sources or their locations. Behura (2015a, 2015b) outlines a waveform inversion strategy that uses continuously recorded multicomponent data from microseismic monitoring to retrieve source spatial and temporal locations, along with a map of a medium property related to the P- and S-wave velocities. However, the algorithms of Kaderli et al. (2015), Behura (2015b), and Wang and Alkhalifah (2016) do require that the source function can be decomposed as a product of independent spatial and temporal functions. Sharan et al. (2016) use the linearized Bregman method to promote

Part of the work presented at the SEG 87th Annual International Meeting. Manuscript received by the Editor 21 December 2017; revised manuscript received 25 August 2018; published ahead of production 21 November 2018.

¹Indian Institute of Technology Bombay, Department of Earth Sciences, Mumbai, India. E-mail: bshekhar@iitb.ac.in (corresponding author).

²Formerly Indian Institute of Technology Bombay, Department of Earth Sciences, Mumbai, India; presently Colorado School of Mines, Center for Wave Phenomena, Golden, Colorado, USA. E-mail: hssethi@mines.edu.

© 2019 Society of Exploration Geophysicists. All rights reserved.

sparsity in FWI for microseismic event location. They demonstrate that sparsity promotion helps achieve well-focused source images.

Here, we present a methodology to invert for the spatio-temporal source function by FWI with sparsity constraints. We first formulate the inverse problem in terms of the spatio-temporal source function that includes microseismic event locations and source signatures. We then introduce the modification of the conventional l_2 -norm data-misfit objective function to include an additional regularization or sparsity-promoting term involving the l_1 -norm of the source function. Then, we review the Orthant-Wise limited-memory quasi-Newton (OWL-QN) method (Andrew and Gao, 2007), which is used to solve the optimization problem. The robustness of the algorithm is demonstrated by performing tests on noise-free and noisy data generated from the SEG/EAGE overthrust model.

METHODOLOGY

Wave propagation in a 2D acoustic medium is governed by the following equation:

$$\frac{1}{v(x, z)^2} \frac{\partial^2 u(x, z, t)}{\partial t^2} - \nabla^2 u(x, z, t) = s(x, z, t), \quad (1)$$

where $u(x, z, t)$ is the pressure wavefield; $v(x, z)$ is the wave velocity; and $s(x, z, t)$ is the spatio-temporal source. Following Kaderli et al. (2015), Wang and Alkhalifah (2016), and Sethi and Shekar (2017), it is possible to split the spatio-temporal source term as $s(x, z, t) = f(x, z)w(t)$, where $f(x, z)$ represents a purely spatial term and $w(t)$ is a purely temporal term or the source wavelet. The spatial term can correspond to spatially distributed sources or point sources at distinct spatial locations. It should be noted that while splitting the source term, it is assumed that the source wavelet is the same for all of the microearthquakes that may occur in the region of interest. The source waveform is controlled by the rise and rupture times of faults (Aki and Richards, 2002), and the assumption that the source waveform is uniform may be justified on the grounds that hydraulic fracturing is expected to lead to microearthquakes with small rise and rupture times (Eisner et al., 2013). In this paper, we consider the inversion for the composite $(s(x, z, t))$ and

Algorithm 1. OWL-QN method in FWI.

```

Choose initial point  $m_0$ 
 $S \Leftarrow \{\}$   $Y \Leftarrow \{\}$ 
for  $k = 0$  to MaxIters do
    Compute  $\nu_k = -\diamond_i \chi_s(\mathbf{m}^k)$ 
    Compute  $\mathbf{d}_k \Leftarrow \mathbf{H}_k \nu_k$  using  $S$  and  $Y$ 
     $\mathbf{p}_k \Leftarrow \pi(\mathbf{d}_k; \nu_k)$ 
    Find  $\mathbf{m}^{k+1}$  with constrained line search
    if termination condition satisfied then
        Stop and return  $\mathbf{m}^{k+1}$ 
    end if
    Update  $S$  with  $s^k = \mathbf{m}^{k+1} - \mathbf{m}^k$ 
    Update  $Y$  with  $y^k = \nabla \chi_s^{k+1} - \nabla \chi_s^k$ 
end for

```

split $(f(x, z)w(t))$ spatio-temporal source functions sampled appropriately in space and time.

We now outline our algorithm to invert for the spatio-temporal source based on waveform inversion with sparsity constraints. The conventional FWI uses an l_2 -norm data-misfit function $\chi(\mathbf{m})$:

$$\chi(m) = \frac{1}{2} \|u^{\text{pred}}(\mathbf{m}) - u^{\text{obs}}\|_2, \quad (2)$$

to minimize the misfit between the predicted $(u(\mathbf{m})^{\text{pred}})$ and observed (u^{obs}) wavefields sampled at receiver positions. The model vector \mathbf{m} is the spatio-temporal source, i.e., $s(x, z, t)$ or $f(x, z)w(t)$. To account for the fact that microearthquakes usually occur in sparse locations in a region of interest, that the source wavelet is time limited (i.e., sparse in the time domain), and to regularize the inversion, we modify the objective function $\chi(\mathbf{m})$ to $\chi_s(\mathbf{m})$ by adding a regularization term that corresponds to the l_1 -norm of the spatio-temporal source:

$$\chi_s(\mathbf{m}) = \frac{1}{2} \|u^{\text{pred}}(\mathbf{m}) - u^{\text{obs}}\|_2 + c \|\mathbf{m}\|_1, \quad (3)$$

where c is a positive scalar coefficient that controls the weight assigned to favor sparse models of $s(x, z, t)$.

Minimizing the misfit function outlined in equation 3 is challenging as the second term of the misfit function involving the l_1 -norm is nondifferentiable as any (or all) of the model parameters approach zero. Many algorithms overcome this difficulty by transforming the minimization problem into a constrained optimization problem and solving it using quasi-Newton methods (Goodman, 2004; Lee et al., 2006; Andrew and Gao, 2007). We use the OWL-QN method developed by Andrew and Gao (2007) because it has been shown to be adapted to problems with a large number of model parameters. The OWL-QN method combines the advantage of quasi-Newton algorithms (e.g., the L-BFGS algorithm; see Nocedal and Wright, 1999) with sparsity promotion by the regularization term. We shall briefly review the OWL-QN algorithm developed by Andrew and Gao (2007) in the context of the source-location problem. The main idea of the algorithm is based on the observation that the l_1 -norm when restricted to an orthant, i.e., a set of points in which each coordinate never changes sign, is differentiable and linear so that the second-order behavior is completely determined only by the data-misfit component, i.e., the first term on the right side of equation 3. This allows the construction of a quadratic approximation, valid for the particular orthant, with the inverse Hessian matrix determined from the (unregularized) data-misfit term alone.

In the following discussion, we borrow some of the notation from Andrew and Gao (2007). The signum function σ is defined as

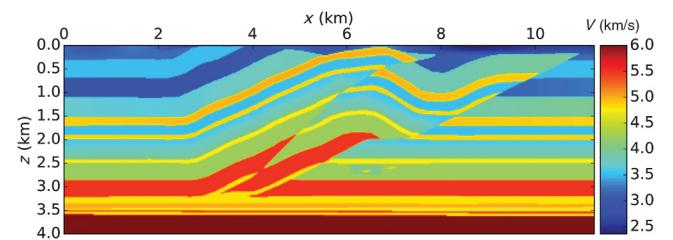


Figure 1. Original velocity model modified from the SEG over-thrust model.

$$\sigma(x) = \begin{cases} \frac{x}{|x|} & x \neq 0 \\ 0 & x = 0, \end{cases} \quad (4)$$

for real numbers x . The function $\pi(x; y)$ for real numbers x and y is defined as

$$\pi(x; y) = \begin{cases} x & \text{if } \sigma(x) = \sigma(y) \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

and it can be regarded as the projection of x onto an orthant defined by y .

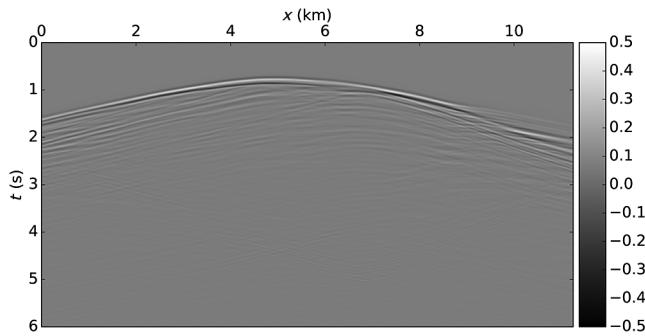


Figure 2. Noise-free synthetic data generated from a point source (with a 10 Hz Ricker wavelet) placed at (5.0 km, 2.0 km) in the velocity model from Figure 1.

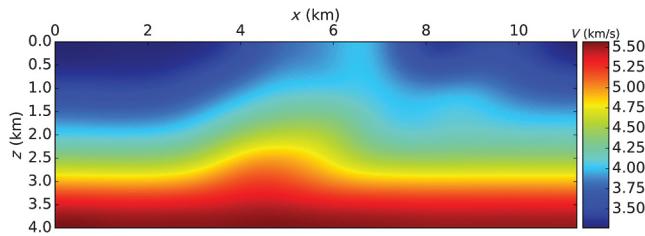


Figure 3. Velocity model used in the inversion, obtained by applying Gaussian smoothing to the velocity model from Figure 1.

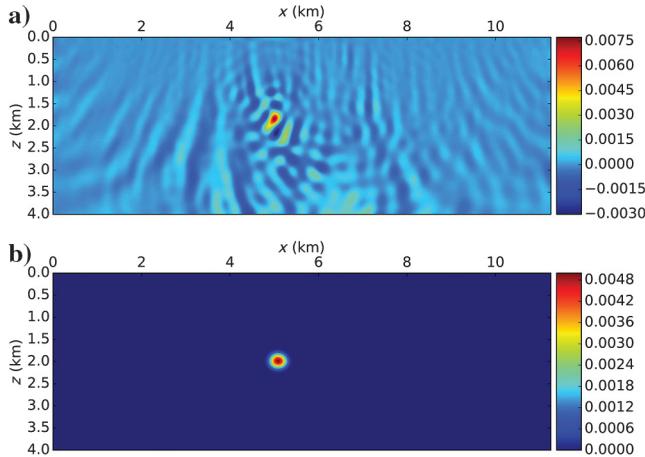


Figure 4. Inverted source spatial image $f(x, z)$ from (a) conventional FWI ($c = 0$) and (b) sparse FWI ($c = 10$). The true location of the source is (5.0 km, 2.0 km).

We adopt the definition of the “pseudo gradient” $\diamond_i \chi_s(\mathbf{m})$ of the objective function in equation 4 from Andrew and Gao (2007) valid for an orthant:

$$\diamond_i \chi_s(\mathbf{m}) = \begin{cases} \partial_i^- \chi_s & \text{if } \partial_i^- \chi_s > 0 \\ \partial_i^+ \chi_s & \text{if } \partial_i^+ \chi_s < 0 \\ 0 & \text{otherwise,} \end{cases} \quad (6)$$

where $\partial_i^- \chi_s(\mathbf{m})$ and $\partial_i^+ \chi_s(\mathbf{m})$ are the left and right partial derivatives of the data-misfit function with respect to the model parameter m_i defined as

$$\partial_i^\pm \chi_s = \nabla_{m_i} \chi + \begin{cases} c \sigma(m_i) & \text{if } m_i \neq 0 \\ \pm c & \text{if } m_i = 0, \end{cases} \quad (7)$$

where $\nabla_{m_i} \chi$ denotes the Fréchet derivative or the gradient of the data-misfit function with respect to the model parameters. We shall discuss the computation of the gradients later in this section.

The FWI algorithm with sparsity constraints on the source function is as follows:

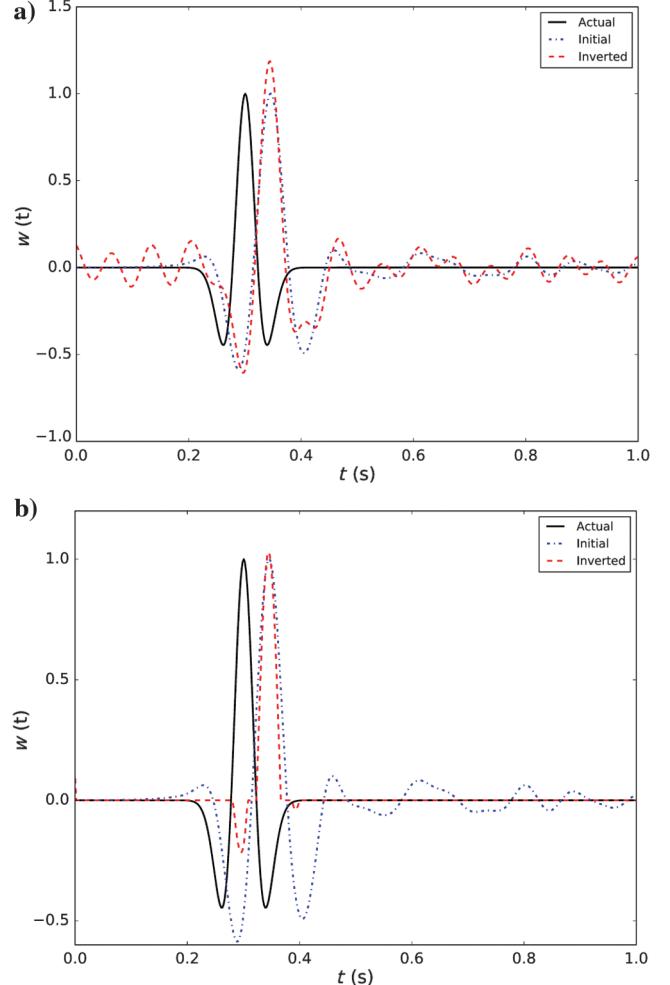


Figure 5. Source wavelets $w(t)$ from (a) conventional FWI ($c = 0$) and (b) sparse FWI ($c = 10$). The true source wavelet is plotted as a solid black line, the initial wavelet is plotted as a dashed-dotted blue line, and the inverted wavelet is plotted as a dashed red line.

The above algorithm follows the L-BFGS algorithm closely, except for a few key differences. The quantity Y depends on the gradient of the (unregularized) data-misfit term, and the approximate inverse Hessian matrix \mathbf{H} is computed as per the standard L-BFGS algorithm. A key difference is that the negative of the pseudogradient $\nu_i = -\diamond_{i\chi_s}(\mathbf{m})$ is projected onto a subspace containing the orthant. Furthermore, it is ensured that the search direction \mathbf{p}_i does not deviate much from the steepest-descent direction. This is ensured by defining \mathbf{p}_i using the projection function π , as outlined in the algorithm above.

Table 1. Coordinates (x , z) of the source locations. The first column lists the true locations, whereas the second and third columns list the locations inverted from noise-free (Figure 7) and noisy data (Figure 13) with the smooth velocity model (Figure 3). The fourth column lists the source locations inverted from noisy data (Figure 13) and the exact velocity model (Figure 1); the fifth column lists the inverted source locations from noise-free data and a 1D velocity model (Figure 19).

Sl. no.	True location	Inverted location	Inverted location	Inverted location	Inverted location
		Noise-free data	Noisy data	Exact velocity	1D velocity
1	(4.25, 1.35)	(4.20, 1.28)	(4.20, 1.28)	(4.25, 1.37)	(4.35, 1.78)
2	(5.50, 1.65)	(5.48, 1.53)	(5.45, 1.48)	(5.50, 1.63)	(5.48, 1.9)
3	(6.75, 3.00)	(6.70, 2.90)	(6.70, 2.95)	(6.70, 3.00)	(6.65, 3.45)
4	(7.50, 1.75)	(7.60, 1.98)	(7.53, 1.90)	(7.50, 1.75)	(7.35, 2.15)

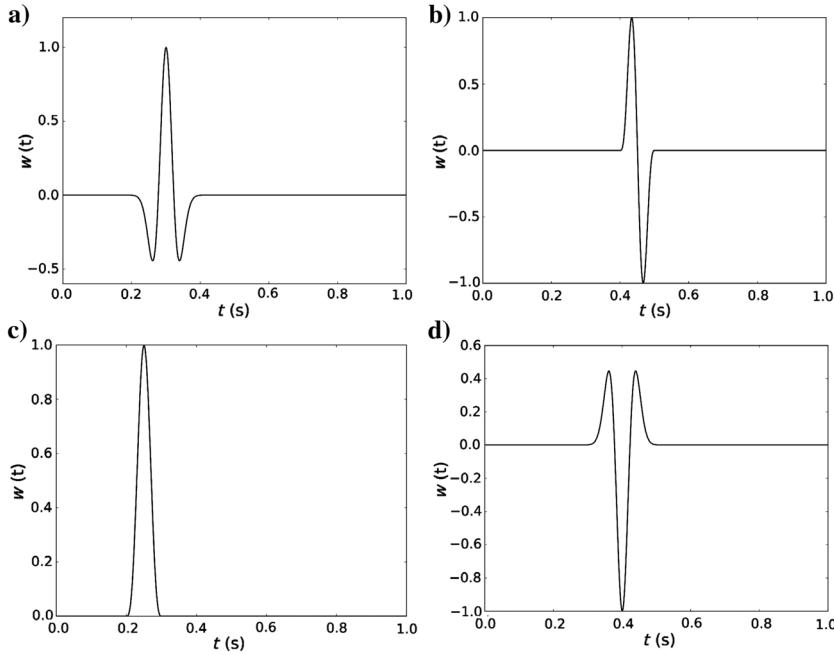


Figure 6 Source wavelets with a central frequency of 10 Hz used to test the algorithm. The wavelet in (a) is a Ricker wavelet and corresponds to the source 1, (b) is a sine-cubed wavelet for the source 2, (c) is a Fucher-Mullers wavelet for the source 3, and (d) is a Ricker wavelet for the source 4. The source serial numbers and locations are displayed in Table 1.

The explicit calculation of the gradient $\nabla_m \chi$ is computationally expensive. We use the adjoint-state method to calculate the Frechet derivatives, which reduces the cost to two forward simulations (Plessix, 2006).

The expressions for the gradient of the data-misfit function χ with respect to the spatial $f(x, z)$ and temporal $w(t)$ source parameters (Behura, 2015b; Kaderli et al., 2015) are as follows:

$$\nabla_f \chi = \int_0^T u^\dagger(t, x, z) w(t) dt \quad (8)$$

and

$$\nabla_w \chi = \int_x \int_z u^\dagger(t, x, z) f(x, z) dx dz, \quad (9)$$

where T is the total time of the recording and $u^\dagger(t, x, z)$ is the adjoint wavefield, which corresponds to back-propagation of the data residuals from the receiver locations. The above expressions imply that the gradient with respect to the source spatial parameter $f(x, z)$ is the zero-lag crosscorrelation in time between the adjoint wavefield and the source wavelet $w(t)$, and the gradient with respect to $w(t)$ is the zero-lag crosscorrelation in space between the adjoint wavefield and $f(x, z)$. Clearly, there is a cyclic relationship between $f(x, z)$ and $w(t)$. We tackle this by alternatively updating $w(t)$ and $f(x, z)$ similar to Kaderli et al. (2015).

The expression for the gradient of the data-misfit function χ with respect to the spatio-temporal source $s(x, z, t)$ derived from the adjoint-state method (see Appendix A) is

$$\nabla_s \chi = u^\dagger(t, x, z), \quad (10)$$

where $u^\dagger(t, x, z)$ is the adjoint wavefield. This implies that the gradient with respect to the source function corresponds to the data residuals back-propagated from the receiver locations. Consequently, if the initial model for the source function is chosen to be zero, the simulated data would also be identically zero, and the gradient in the first step of the inversion then corresponds to the back-propagation of the observed data. Hence, if the velocity model accounts correctly for the kinematics, the gradient with respect to the spatio-temporal source function attains its highest value corresponding to the spatial and temporal position of the focus.

NUMERICAL EXAMPLES

We shall now discuss the performance of the algorithm outlined above on synthetic data from a set of numerical experiments. In all of the experiments, we use a velocity model modified from the SEG/EAGE overthrust velocity model (Aminzadeh et al., 1997), shown in Figure 1.

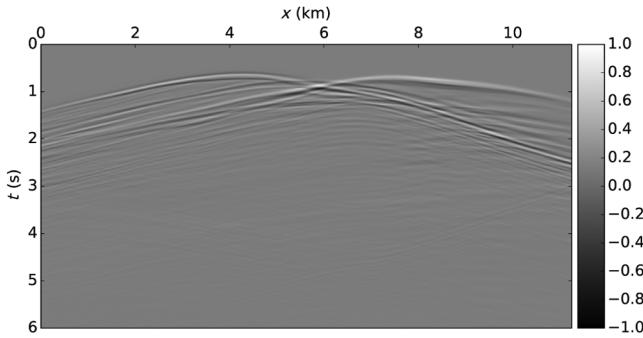


Figure 7. Noise-free synthetic data generated from four point sources with locations as shown in Table 1 and the source wavelets displayed in Figure 6.

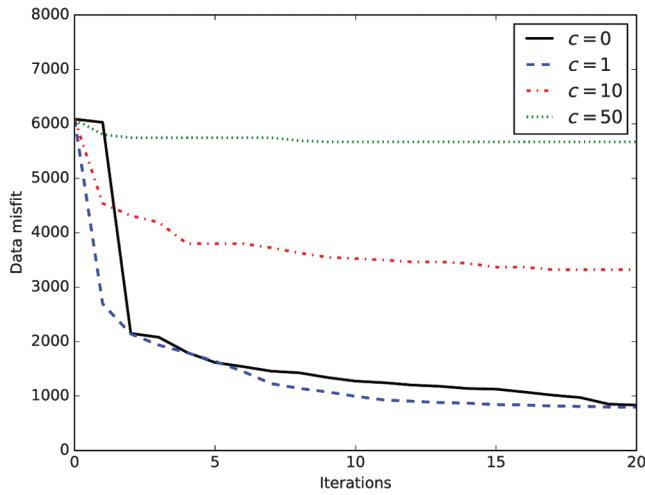


Figure 8. Data misfit as a function of the number of iterations for the regularization parameter $c = 0$ (solid black line), $c = 1$ (dashed blue line), $c = 10$ (dashed-dotted red line), and $c = 50$ (dotted green line).

The model has 160×450 (4.0×11.25 km) grid points with a spatial interval of 25 m in the x - and z -directions. The receivers are placed 25 m below from the surface with a spacing interval of 25 m in the x -direction. Acoustic-wave propagation is simulated with the finite-difference method with absorbing boundaries (Clayton and Engquist, 1977) to model the microseismic data.

Experiment 1: Inversion for the split-source function

In the first experiment, we invert for the source spatial ($f(x, z)$) and temporal ($w(t)$) functions using noise-free data generated from the overthrust model. A point source is placed at (5.0 km, 2.0 km), and the source wavelet is a Ricker wavelet with a peak frequency of 10 Hz. The resulting data generated from the acoustic finite-difference method (Clayton and Engquist, 1977) is displayed in Figure 2. Figure 3 displays the velocity model used in the inversion; it is obtained by applying Gaussian smoothing to the original velocity model. At the start of the inversion process, we assume that the spatial source function is identically zero, i.e., that $f(x, z) = 0$. The initial model for the source wavelet is calculated by back-propagating the observed data and then extracting the wavelet from the back-propagated wavefield at the spatial position corresponding to maximum energy. The spatial and temporal source functions are then alternatively updated using the inversion scheme outlined above. The inverted spatial source functions $f(x, z)$ plotted in Figure 4a and 4b correspond to the regularization parameter values of $c = 0$ and $c = 10$, respectively. A zero value for the parameter c corresponds to conventional FWI, whereas a positive value implies sparsity promotion. The strategy for the selection of the regularization parameter shall be discussed in greater detail in the next section. The sparsity-enforcing regularization causes the spatial source image in Figure 4b to be well-focused with fewer artifacts compared to the source image in Figure 4a. The inverted source wavelet using the regularization parameter $c = 0$ (conventional FWI) and $c = 10$ (sparse FWI) are compared with the actual and initial source wavelet in Figure 5a and 5b, respectively. It can be seen that the sparsity promotion has the effect of attenuating the side lobes of the initial wavelet, whereas the wavelet retrieved from the conventional FWI is more oscillatory compared with the initial wavelet. The oscillations

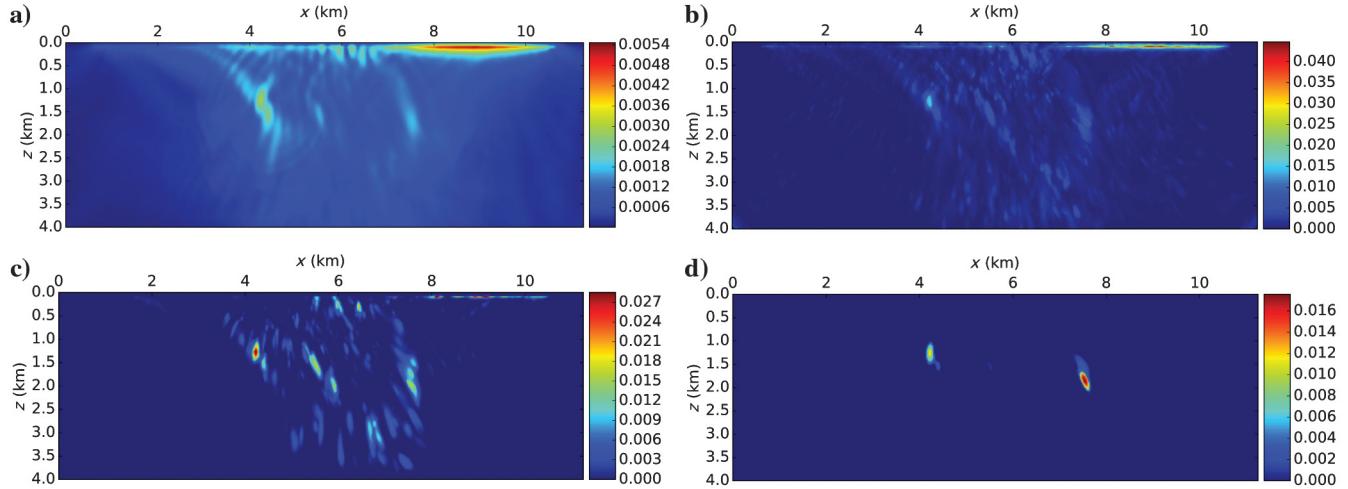


Figure 9. Source power function obtained from noise-free data and the smooth velocity model for regularization parameter (a) $c = 0$, (b) $c = 1$, (c) $c = 10$, and (d) $c = 50$.

tory behavior of the source wavelet can be attributed to the fact that the spatial source function is comparatively diffuse and the wavefield from a wider spatial extent contributes to the source wavelet.

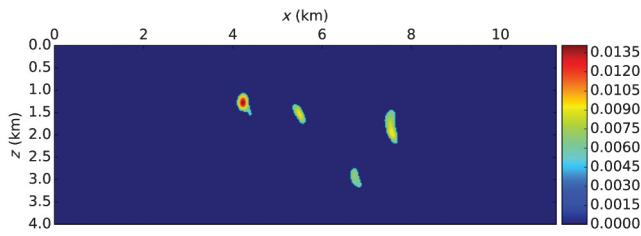


Figure 10. Isolated source images derived from the source power function shown in Figure 9c. The coordinates of the peaks are listed in Table 1.

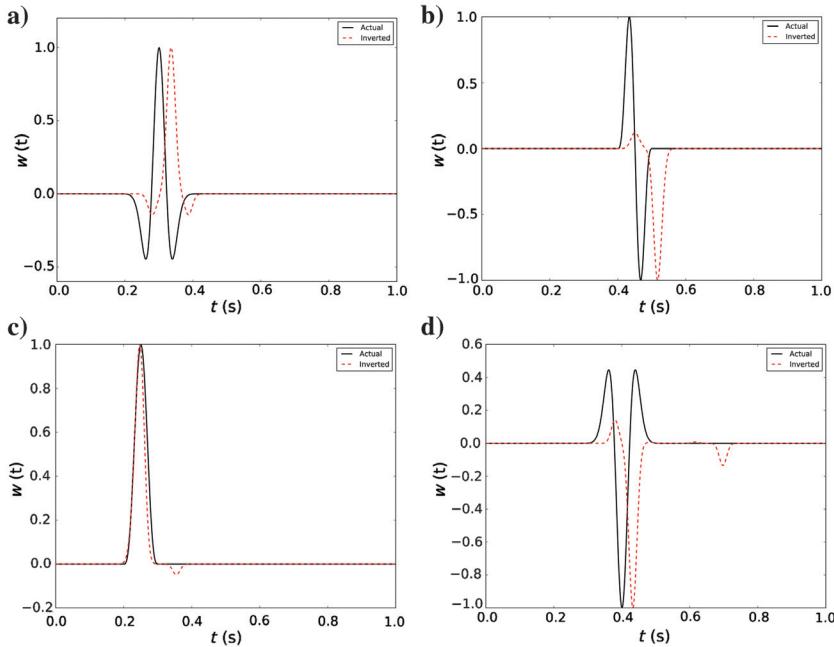


Figure 11. Source wavelets derived from the spatio-temporal source function inverted from noise-free data and the smooth velocity model. The true source wavelet is plotted as a solid black line, and the inverted wavelet is plotted as a dashed red line.

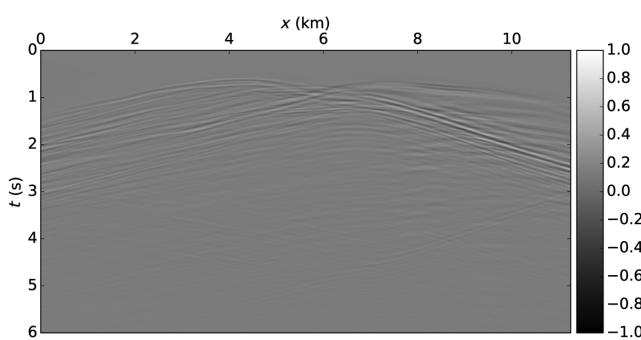


Figure 12. Data residuals corresponding to noise-free data and $c = 10$.

However, it should be noted that the peaks of the inverted wavelets from conventional and sparse FWI do not coincide with the actual wavelet; this is due to the differences between the true velocity model and that used in the inversion.

Experiment 2: Inversion for the composite source function from data generated by spatially and temporally isolated sources

We now consider the inversion for the composite spatio-temporal source function $s(x, z, t)$ for data generated from the overthrust model. In this experiment, the sources are assumed to be spatially isolated with time-limited source wavelets. We model four point sources with their locations as listed in Table 1 and the source wavelets as shown in Figure 6. Figure 7 shows the noise-free data generated by the four point sources from the SEG overthrust model (Figure 1). As in the previous experiment, the velocity model used in the inversion is a smoothed version of the overthrust model (Figure 2), and the initial value for the spatio-temporal source function is assumed to be zero, i.e., $s(x, z, t) = 0$. In the following subsections, we discuss the sensitivity of the inversion scheme to the regularization parameter c , noise, and the velocity model.

Choice of regularization parameter

Here, we illustrate the effect of the sparsity-enforcing regularization parameter c on the recovered spatio-temporal source function. Initially, a range of values between zero and approximately one-tenth of the l_2 -norm of the initial data-misfit value was chosen for c . Because the initial value for $s(x, z, t)$ is zero, the l_2 -norm of the data misfit corresponds to the l_2 -norm of the observed data. The values of c were refined by observing the l_2 -norm of the data-misfit function for a few iterations of the inversion scheme. The l_2 -norm of the data-misfit function for $c = 0, 1, 10$, and 50 is plotted against the number of iterations in Figure 8. It can be seen that increasing the magnitude of c causes the data-misfit function to flatten out for a smaller number of iterations and at higher values of the data misfit. Furthermore, for

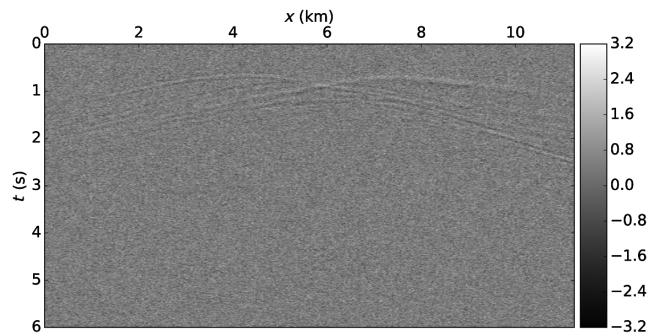


Figure 13. Noisy data generated by adding band-limited Gaussian noise to noise-free data (shown in Figure 7) from the overthrust model. The signal-to-noise ratio is 1.0.

$c > 100$, an insignificant reduction in the initial misfit value was observed (not shown here). Because the parameter c controls the sparsity of the inverted model, source functions corresponding to higher values of c will possess lesser features and consequently lead to higher data misfit when compared with the source functions for lower values of c .

Before we discuss the inversion results for the spatio-temporal source function $s(x, z, t)$, we need a framework to interpret the source function in terms of individual sources and the corresponding source wavelets. To this end, we express the average power $P(x, z)$ of the inverted source function as a function of space:

$$P(x, z) = \left[\int_0^T s^2(x, z, t) dt \right]^{0.5}. \quad (11)$$

Figure 9 displays the source power function $P(x, z)$ computed from the inverted source functions for $c = 0, 1, 10$, and 50. The value $c = 0$ corresponds to inversion without sparsity promotion, and the corresponding source power function plotted in Figure 9a is diffuse. Predictably, successively higher values of c lead to the power function to retain only the most significant features. The high-energy regions in the images of the power function (see Figure 9b–9d) contain the source locations along with areas in the model in which the back-propagated residual wavefield may have been strong (due to diffractions or focusing by the velocity model). In this experiment, we assume that connected regions higher than 90th percentile of the power function correspond to source locations. The hence isolated source image corresponding to $c = 10$ is plotted in Figure 10, and the coordinates of the peaks of the isolated source images are listed in Table 1. The inverted source locations are close to the true source locations; however, there is greater difference in the depth (z) coordinates due to the use of a smooth velocity model in the inversion. The corresponding source wavelets $w^{\text{inv}}(t)$ can be calculated as

$$w^{\text{inv}}(t) = s(x^p, z^p, t), \quad (12)$$

where (x^p, z^p) correspond to the peaks of the inverted source images. The hence recovered source wavelets are plotted in Figure 11 along with the true wavelets. The differences between the peaks of the inverted and true wavelets can be attributed to the difference between the actual velocity model and the velocity model used for the inversion, whereas the distortion in wavelet shapes is due to the sparsity enforcement and the difference in the location of the source peaks. The data residuals from the inversion with $c = 10$ are plotted in Figure 12. Comparing Figure 7 with Figure 12, we find that the predicted data from the inversion fit the first arrivals fairly well. The inversion scheme cannot account for later arrivals as a smooth velocity model is used in the inversion; consequently, reflections and multiples are not modeled in the predicted data.

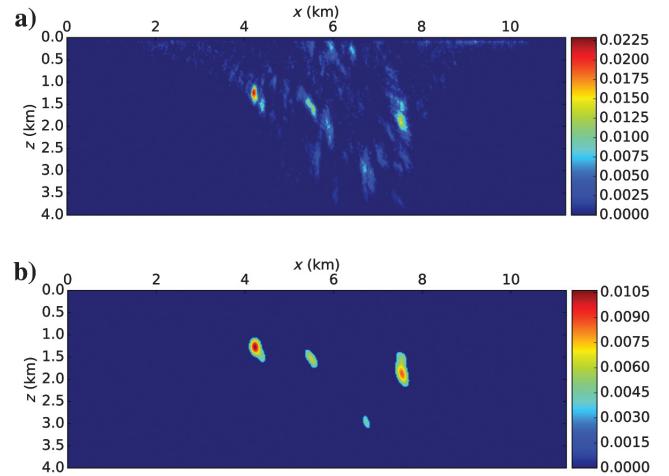


Figure 14. (a) Source power function obtained from the noisy data, smooth velocity model, and regularization parameter $c = 10$ and (b) the image with isolated sources.

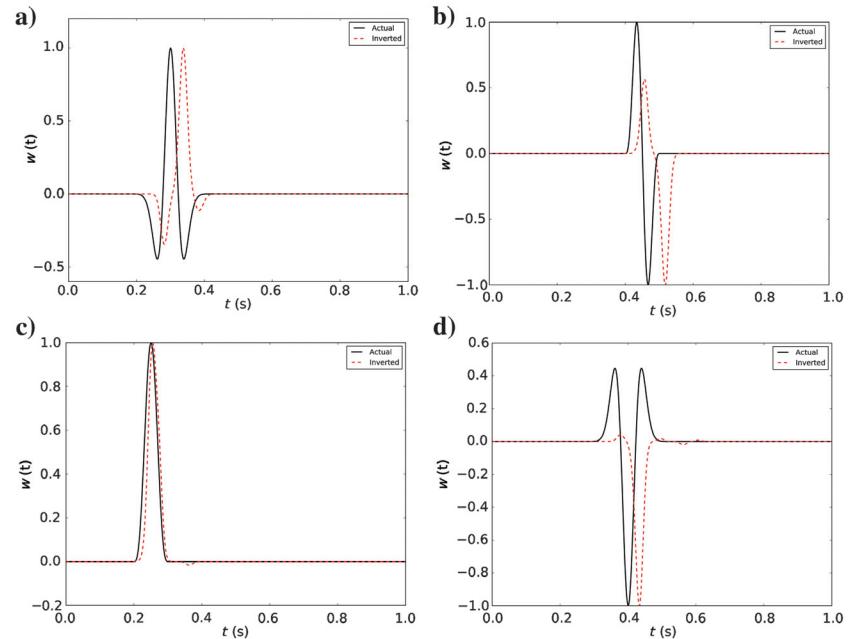


Figure 15. Source wavelets derived from the spatio-temporal source function inverted from noisy data and the smooth velocity model. The true source wavelet is plotted as a solid black line, and the inverted wavelet is plotted as a dashed red line.

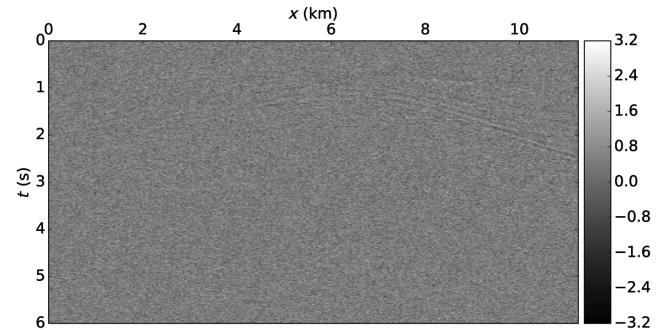


Figure 16. Data residuals for noisy data, the smooth velocity model, and $c = 10$.

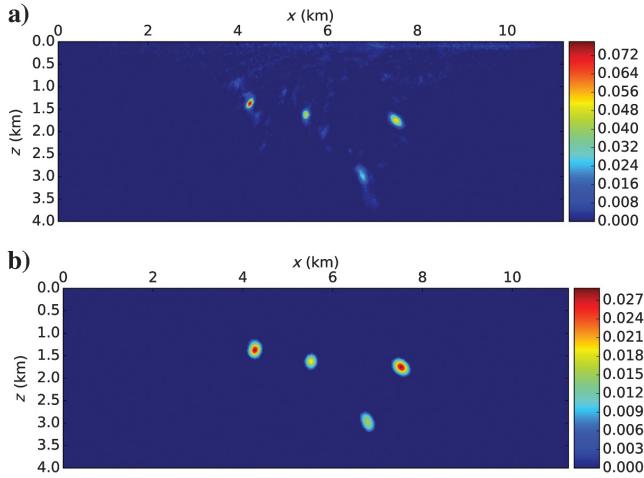


Figure 17. (a) Source power function and (b) isolated source image obtained from the inversion with the exact velocity model, noisy data, and regularization parameter $c = 10$.

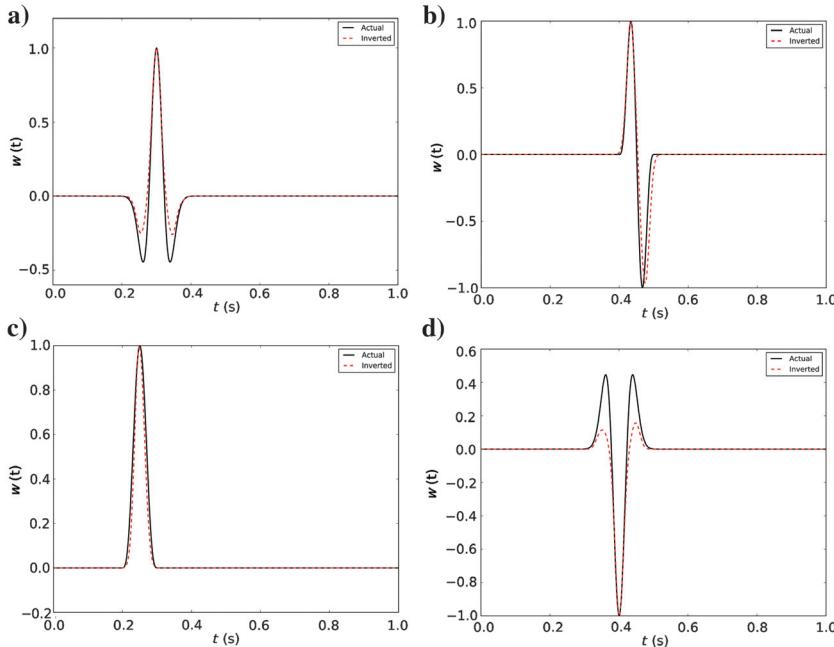


Figure 18. Source wavelets derived from the spatio-temporal source function inverted from noisy data and the exact velocity model. The true source wavelet is plotted as a solid black line, and the inverted wavelet is plotted as a dashed red line.

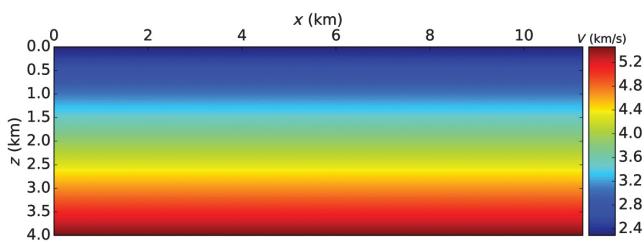


Figure 19. Initial 1D velocity model used in the inversion to demonstrate the effect of cycle skipping.

Sensitivity to noise

We now test the performance of the algorithm on noisy data generated by adding band-limited Gaussian noise to the noise-free data considered previously. The noisy data is displayed in Figure 13; it has a signal-to-noise ratio of 1.0. The power function for the inverted source function for $c = 10$ is plotted in Figure 14a, and the isolated sources are plotted in Figure 14b. The peaks of the isolated source images are listed in Table 1. To ensure that the test with noisy data is meaningful, we used the same regularization parameter ($c = 10$) and source selection criteria as for the noise-free experiment.

The recovered source locations are fairly accurate, except for the differences in depth coordinates as was observed for the noise-free case. The inverted source wavelets are plotted in Figure 15. As observed before, the shifts between the true and inverted wavelets are due to differences between the actual velocity model and the velocity model used for the inversion. The data residuals from the inversion with are plotted in Figure 16. The first arrivals are fit well by the inversion scheme.

Sensitivity to the velocity model: Exact velocity model is chosen

We now demonstrate the influence of the velocity model on the inversion algorithm. We begin by choosing the exact velocity model (Figure 7) and we invert noisy data (Figure 13) for the spatio-temporal source function. The source power function and the isolated sources are displayed in Figure 17. The peaks of the isolated source images are listed in Table 1, and the source wavelets are plotted in Figure 18. The inverted and true source locations and wavelets compare well because the exact velocity model is used in the inversion.

Sensitivity to the velocity model: The chosen velocity model is far from the exact velocity model

Like most FWI algorithms, the outlined inversion algorithm requires the velocity model to be close to the true velocity model. Large differences between the velocity models cause cycle skipping, in which the predicted data differ from the observed data by more than half a cycle, leading the inversion to converge to a local minimum.

To highlight the issue of cycle skipping, we invert for the source locations from the noise-free data (Figure 7) used before but with a 1D velocity model as shown in Figure 19. Figure 20 displays the source power function and isolated sources from the inverted source function with $c = 10$. The peaks of the isolated source images are listed in Table 1, and the source wavelets are plotted in Figure 21. There are differences in the actual and true source locations, but there is a pronounced difference between the true and inverted source wavelets due to the effect of cycle skipping. In addition, it should be noted that some of the retrieved wavelets (Figure 21a and 21c) are acausal. The residuals plotted in Figure 22 display significant energy compared with the observed data (Figure 7).

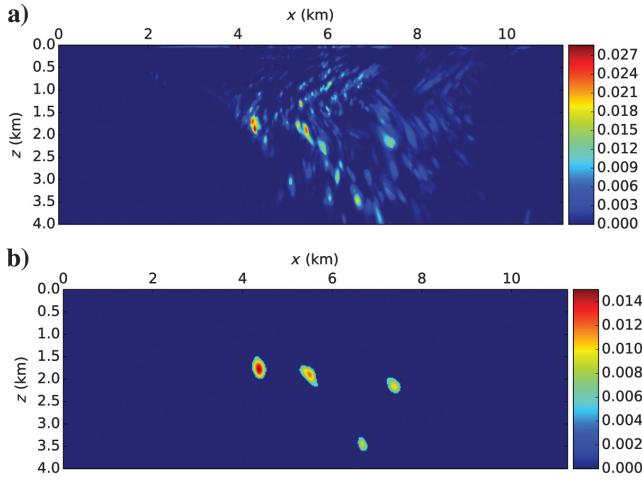


Figure 20 a. (a) Source power function obtained from inversion with the 1D velocity model as shown in Figure 19, noise-free data, and parameter $c = 10$ and (b) the image with the isolated sources.

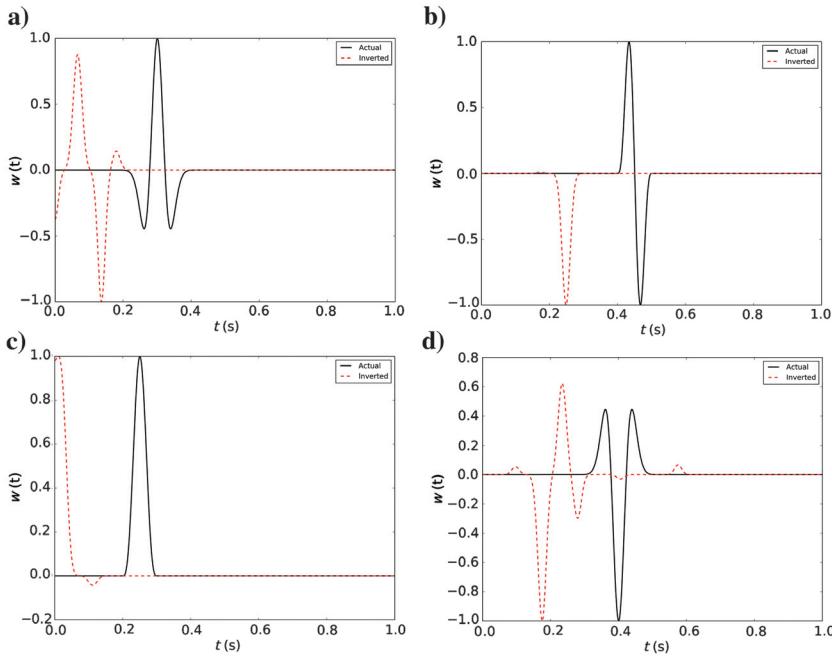


Figure 21. Source wavelets derived from the spatio-temporal source function inverted from noise-free data and with the 1D velocity model. The true source wavelet is plotted as a solid black line, and the inverted wavelet is plotted as a dashed red line.

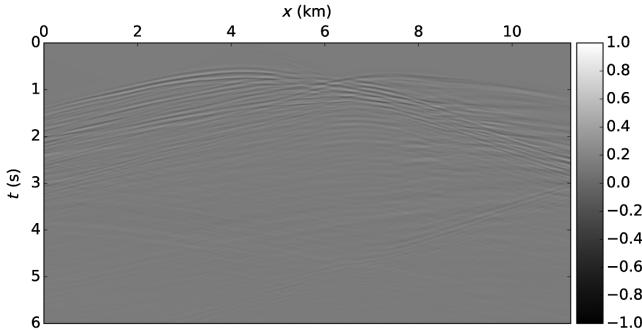


Figure 22. Data residuals for noise-free data and $c = 10$, with the 1D velocity model used in the inversion scheme.

Experiment 3: Inversion for the composite source function from data generated by spatially and temporally distributed sources

In this experiment, we test the validity of the assumption of spatio-temporal sparsity on data generated from a spatially and temporally distributed source. We place 16 point sources along neighboring grid points corresponding to a fault plane as shown in Figure 23. The source wavelets displayed in Figure 24a have a signal-to-noise ratio of one. The source wavelets are generated by adding band-limited noise to Ricker wavelets with a central frequency of 30 Hz but with the peaks shifted in time as displayed in Figure 24b. The data generated from the distributed source is shown in Figure 25. The data is visibly less noisy than the source wavelets due to the effect of wave-front healing during propagation. However, energy similar to the first arrivals is repeated multiple times in the data due to the fact that the temporal source function possesses significant energy for its whole duration.

As in the experiments above, we assume the initial values for $s(x, z, t)$ to be zero and investigate the data misfit as a function

of a large range of values for the parameter c . Figure 26 displays the data misfit as a function of the number of iterations for $c = 10, 100, 500$, and 1000 . The corresponding source-power functions are displayed in Figure 27. Although the spatially extended source is mapped with reasonable accuracy for all four values of c , higher values of the regularization parameter lead to sharper images of the sources along the fault. The source wavelet corresponding to the average of the source-power function and $c = 500$ is displayed in Figure 28. The retrieved source wavelet is sparse, and its peak is close to the average of the peak times of the Ricker wavelets displayed in Figure 24b. A significant reduction in the first-arrival energy is observed in the residuals displayed in Figure 29.

DISCUSSION

The various experiments in the preceding section demonstrate the efficacy of the inversion algorithm in source-function inversion from noise-free and noisy data, and data in which the assumption of spatio-temporal sparsity of source functions may not be satisfied. The sparsity-promoting regularization parameter may be chosen

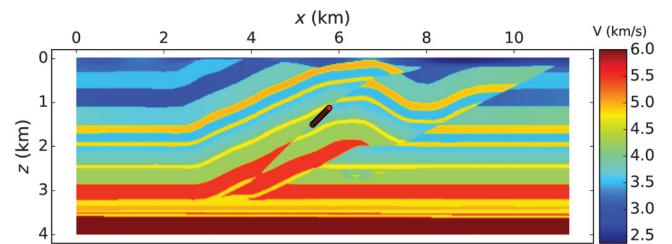


Figure 23. Velocity model modified from the SEG overthrust model. Point sources placed along the central fault are plotted as red circles.

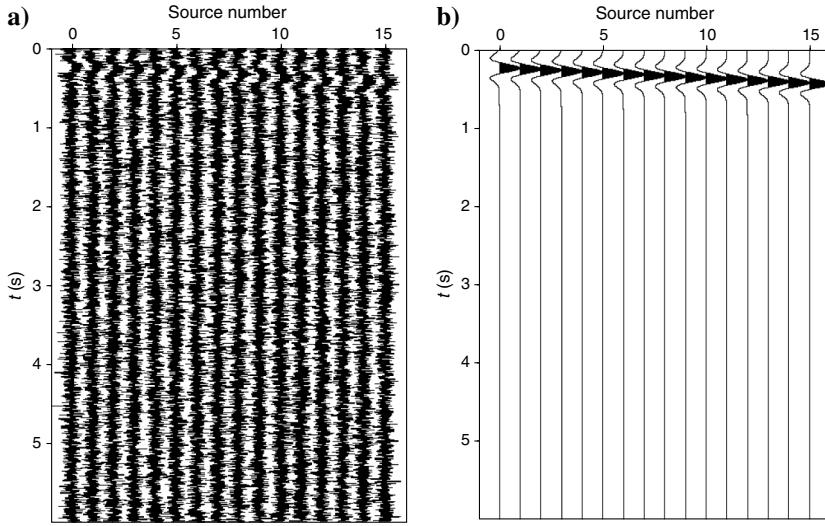


Figure 24. (a) Noisy source wavelets with a signal-to-noise ratio of one as a function of the source number obtained by adding band-limited Gaussian noise to the Ricker wavelets in (b). The source number increases along the downdip direction of the fault in Figure 23.

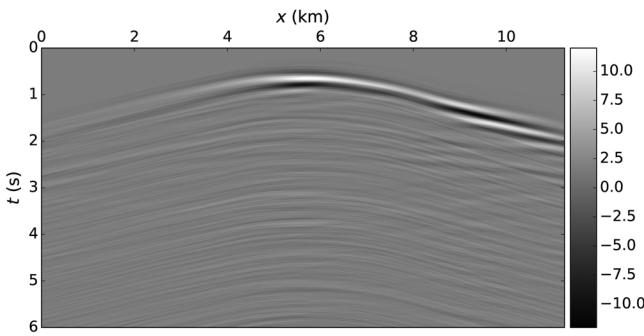


Figure 25. Data generated from the overthrust model with point sources as displayed in Figure 23 and the corresponding source wavelets as displayed in Figure 24a.

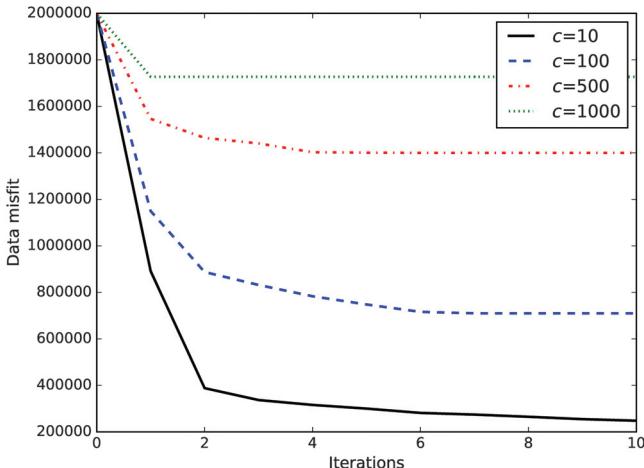


Figure 26. Data misfit as a function of number of iterations for regularization parameter $c = 10$ (solid black line), $c = 100$ (dashed blue line), $c = 500$ (dashed-dotted red line), and $c = 1000$ (dotted green line).

based on the l_2 -norm of the data-misfit function and the features expected in the spatio-temporal source function. Although the algorithm does not require the number of sources to be known in advance, the choice of regularization parameter does control the features of the spatio-temporal source function: the number of sources and the shape of the source wavelet.

The outlined method does requires that the velocity model used in the inversion is close to the true model. Because the method exclusively inverts for the spatio-temporal source function, any errors in velocity will lead to errors in the inverted source function. The cycle-skipping problem may be mitigated by using techniques such as dynamic time warping (Hale, 2013; Luo and Hale, 2013).

Furthermore, the inversion methodology does not account for reflections, and the scattered data can cause distortions in the inverted source wavelets. Although the effect of scattered data on the

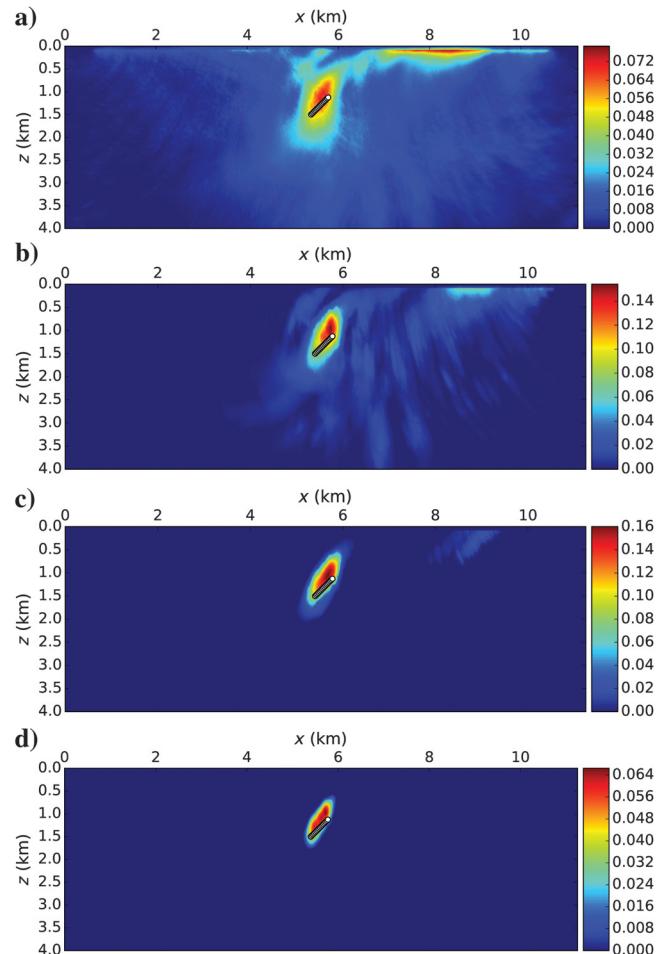


Figure 27. Source power function obtained from data shown in Figure 25 and for regularization parameter (a) $c = 10$, (b) $c = 100$, (c) $c = 500$, and (d) $c = 1000$. The true sources are plotted as white circles.

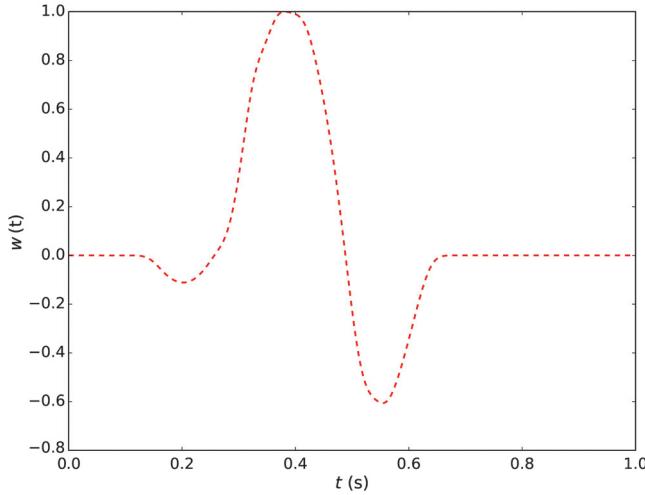


Figure 28. Source wavelet derived from the spatio-temporal source function corresponding to the peak of the power function displayed in Figure 27c, corresponding to $c = 500$.

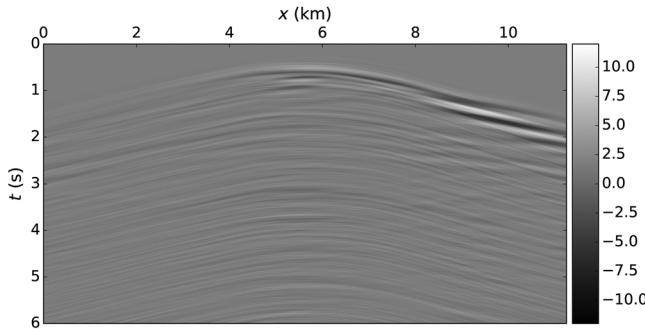


Figure 29. Data residuals for the data shown in Figure 25 and $c = 500$.

inversion can be minimized by working exclusively with transmitted data, isolating transmissions in continuously recorded data would require knowledge of the origin times of the microseismic events.

The source-detection algorithm can be further improved by incorporating shape-detection algorithms such as the Hough transform (Ballard, 1981) or measures of focusing such as Shannon's entropy function (Shustak and Landa, 2017) into the spatio-temporal source function to potentially recognize explosions or fault-plane solutions. Extending the proposed methodology to three dimensions would require enhancing the computational efficiency by using techniques such as domain decomposition or wavefield checkpointing strategies.

CONCLUSION

We have presented a strategy based on FWI to invert for the spatio-temporal source function of microearthquakes. The spatio-temporal source function can account for differences in the source locations and wavelets of distinct events. The sparsity inherent in the microseismic event location problem is used to regularize the inversion. The OWL-QN method is used to overcome the nondifferentiability

of the l_1 -norm terms in the regularized objective function. We tested the algorithm on noise-free and noisy data generated from the SEG/EAGE overthrust model, and we found that the method is robust in the presence of noise. Although the method assumes spatio-temporal sparsity of source functions, it also performs well for spatially and temporally distributed source functions. We found that the spatial source functions are well-focused and closely approximate the true locations of the point sources. The inverted source wavelets were close to the true wavelets, except for distortion due to scattering and noise.

ACKNOWLEDGMENTS

The authors would like to thank assistant editor D. Draganov, J. Behura, and two anonymous reviewers for valuable suggestions that greatly improved the manuscript. The OWL-QN algorithm is implemented by the package in <http://www.chokkan.org/software/crfsuite/>. Connected regions in the source images are found using the multidimensional image processing package in <https://docs.scipy.org/doc/scipy/reference/ndimage.html>.

DATA AND MATERIALS AVAILABILITY

Data associated with this research are available and can be obtained by contacting the corresponding author.

APPENDIX A

GRADIENT EXPRESSIONS FOR A SPATIO-TEMPORAL SOURCE

Here, we derive the expressions for the gradient of the objective function with respect to the spatio-temporal source function. The wave equation for a 2D acoustic medium can be written as

$$\frac{1}{v(x, z)^2} \frac{\partial^2 u(x, z, t)}{\partial t^2} - \nabla^2 u(x, z, t) = s(x, z, t), \quad (\text{A-1})$$

where $u(x, z, t)$ is the wavefield, $v(x, z)$ is the wave velocity, and $s(x, z, t)$ is the spatio-temporal source.

The l_2 -norm objective function defined in equation 2 can be expanded as

$$\begin{aligned} \chi(\mathbf{m}) = & \frac{1}{2} \int_0^T dt \int_{\partial\omega} (u^{\text{obs}}(x, z, t) - u^{\text{pre}}(x, z, t))^2 \\ & \times \delta(x - x_r) \delta(z - z_r) dx dz, \end{aligned} \quad (\text{A-2})$$

where $\partial\omega$ represents the acquisition surface, T is the total recording time, and (x_r, z_r) are the receiver coordinates. Following Plessix (2006), we can define the augmented functional as

$$\mathcal{L}(\tilde{u}, \tilde{\lambda}, \mathbf{m}) = \chi(\tilde{u}(\mathbf{m}), \mathbf{m}) - \langle \tilde{\lambda}, F(\tilde{u}(\mathbf{m}), \mathbf{m}) \rangle, \quad (\text{A-3})$$

where \tilde{u} is the generalized state variable, $\tilde{\lambda}$ is the adjoint state variable, $\langle \cdot \rangle$ denotes the inner product over spatial dimensions, and $F(\tilde{u}(\mathbf{m}), \mathbf{m})$ is the forward problem, i.e., the acoustic wave equation in this case. If u is a physical realization of \tilde{u} , then $F(u(\mathbf{m}), \mathbf{m}) = 0$ and $\mathcal{L}(u, \tilde{\lambda}, \mathbf{m}) = \chi(u, \mathbf{m})$. Now, the gradient of the objective function can be computed from the gradient of the augmented functional:

$$\nabla_m \chi(u, \mathbf{m}) = \nabla_m \mathcal{L}(u, \tilde{\lambda}, \mathbf{m}) + \nabla_u \mathcal{L}(u, \tilde{\lambda}, \mathbf{m}) \cdot \nabla_m u. \quad (\text{A-4})$$

We can choose $\tilde{\lambda} = \lambda$ such that $\nabla_u \mathcal{L}(u, \lambda, \mathbf{m}) = 0$. Then, the gradient of the objective function simplifies to

$$\nabla_m \chi(u, \mathbf{m}) = \nabla_m \mathcal{L}(u, \lambda, \mathbf{m}). \quad (\text{A-5})$$

For the spatio-temporal source-function inversion, the model parameter is $\mathbf{m} = s(x, z, t)$ and the augmented functional defined in equation A-2 can be expanded as

$$\begin{aligned} \mathcal{L}(u, \lambda, s) &= \frac{1}{2} \int_0^T dt \int_{\partial\omega} (u^{\text{obs}}(x, z, t) - u^{\text{pre}}(x, z, t))^2 \\ &\quad \times \delta(x - x_r) \delta(z - z_r) dx dz \\ &\quad - \int_0^T \left\langle \tilde{\lambda}, \frac{1}{v(x, z)^2} \frac{\partial^2 u}{\partial t^2} - \nabla^2 u - s(x, z, t) \right\rangle dt. \end{aligned} \quad (\text{A-6})$$

The gradient with respect to the spatial-temporal source can be calculated by substituting equation A-6 into equation A-5:

$$\nabla_s \chi(s(x, z, t)) = \lambda(x, z, t). \quad (\text{A-7})$$

Hence, the gradient with respect to the spatial-temporal source function is the adjoint variable.

We now derive the equations that govern the adjoint variable λ . The second term of equation A-6 can be simplified by using integration by parts twice:

$$\begin{aligned} \int_0^T \left\langle \tilde{\lambda}, \frac{1}{v(x, z)^2} \frac{\partial^2 u}{\partial t^2} \right\rangle dt &= \int_0^T \left\langle \frac{1}{v(x, z)^2} \frac{\partial^2 \tilde{\lambda}}{\partial t^2}, u \right\rangle dt \\ &\quad - \left\langle \frac{1}{v(x, z)^2} \frac{\partial \tilde{\lambda}(T)}{\partial t}, u(T) \right\rangle \\ &\quad + \left\langle \frac{1}{v(x, z)^2} \frac{\partial \tilde{\lambda}(0)}{\partial t}, u(0) \right\rangle + \left\langle \tilde{\lambda}(T), \frac{1}{v(x, z)^2} \frac{\partial u(T)}{\partial t} \right\rangle \\ &\quad - \left\langle \tilde{\lambda}(0), \frac{1}{v(x, z)^2} \frac{\partial u(0)}{\partial t} \right\rangle. \end{aligned} \quad (\text{A-8})$$

Imposing the condition $\nabla_u \mathcal{L}(u, \lambda, \mathbf{m}) = 0$, we get the adjoint state equation:

$$\lambda(T) = 0, \quad (\text{A-9})$$

$$\frac{\partial \lambda(T)}{\partial t} = 0, \quad (\text{A-10})$$

$$\begin{aligned} \frac{1}{v(x, z)^2} \frac{\partial^2 \lambda}{\partial t^2} - \nabla^2 \lambda &= \int_{\partial\omega} (u^{\text{obs}}(x, z, t) - u^{\text{pre}}(x, z, t)) \\ &\quad \delta(x - x_r) \delta(z - z_r) dx dz. \end{aligned} \quad (\text{A-11})$$

The above set of adjoint equations can be interpreted as the backpropagation of data residuals from receiver locations (Plessix, 2006).

REFERENCES

- Aki, K., and P. G. Richards, 2002, Quantitative seismology, 2nd ed.: University Science Books.
- Aminzadeh, F., J. Brac, and T. Kunz, 1997, 3-D salt and overthrust models: SEG, SEG/EAGE 3-D Modeling Series 1.
- Andrew, G., and J. Gao, 2007, Scalable training of L_1 -regularized log-linear models: Proceedings of the 24th International Conference on Machine Learning, 33–40.
- Artman, B., I. Podladchikov, and B. Witten, 2010, Source location using time-reverse imaging: Geophysical Prospecting, **58**, 861–873, doi: [10.1111/j.1365-2478.2010.00911.x](https://doi.org/10.1111/j.1365-2478.2010.00911.x).
- Ballard, D. H., 1981, Generalizing the Hough transform to detect arbitrary shapes: Pattern Recognition, **13**, 111–122, doi: [10.1016/0031-3203\(81\)90009-1](https://doi.org/10.1016/0031-3203(81)90009-1).
- Behura, J., 2015a, Expedited FWI of microseismic data for mapping velocity changes and hypocenters: 85th Annual International Meeting, SEG, Expanded Abstracts, 2461–2466, doi: [10.1190/segan2015-5908346.1](https://doi.org/10.1190/segan2015-5908346.1).
- Behura, J., 2015b, Method for expediently processing passive seismic field data for subsurface mapping with full-waveform inversion: U.S. 62104063, Patent Pending.
- Clayton, R., and B. Engquist, 1977, Absorbing boundary conditions for acoustic and elastic wave equation: Bulletin of the Seismological Society of America, **67**, 1529–1540.
- Eisner, L., D. Gei, M. Hallo, I. Opršal, and M. Y. Ali, 2013, The peak frequency of direct waves for microseismic events: Geophysics, **78**, no. 6, A45–A49, doi: [10.1190/geo2013-0197.1](https://doi.org/10.1190/geo2013-0197.1).
- Eisner, L., M. Thornton, and J. Griffin, 2011, Challenges for microseismic monitoring: 81st Annual International Meeting, SEG, Expanded Abstracts, 1519–1523, doi: [10.1190/1.3627491](https://doi.org/10.1190/1.3627491).
- Goodman, J., 2004, Exponential priors for maximum entropy models: ACL.
- Hale, D., 2013, Dynamic warping of seismic images: Geophysics, **78**, no. 2, S105–S115, doi: [10.1190/geo2012-0327.1](https://doi.org/10.1190/geo2012-0327.1).
- Kaderli, J., M. D. McChesney, and S. E. Minkoff, 2015, Microseismic event estimation in noisy data via full waveform inversion: 85th Annual International Meeting, SEG, Expanded Abstracts, 2596–2600, doi: [10.1190/segan2015-5867154.1](https://doi.org/10.1190/segan2015-5867154.1).
- Lee, S.-I., H. Lee, P. Abbeel, and A. Y. Ng, 2006, Efficient l1 regularized logistic regression: AAAI.
- Luo, S., and D. Hale, 2013, Separating traveltimes and amplitudes in waveform inversion: 83rd Annual International Meeting, SEG, Expanded Abstracts, 969–974, doi: [10.1190/segan2013-1284.1](https://doi.org/10.1190/segan2013-1284.1).
- Maxwell, S. C., J. Rutledge, R. Jones, and M. Fehler, 2010, Petroleum reservoir characterization using downhole microseismic monitoring: Geophysics, **75**, no. 5, 75A129–75A137, doi: [10.1190/1.3477966](https://doi.org/10.1190/1.3477966).
- Michel, O. J., and I. Tsvankin, 2014, Gradient calculation for waveform inversion of microseismic data in VTI media: Journal of Seismic Exploration, **23**, 201–207.
- Nocedal, J., and S. J. Wright, 1999, Numerical optimization: Springer.
- 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).
- Rutledge, J., and W. S. Phillips, 2003, Hydraulic stimulation of natural fractures as revealed by induced microearthquakes, Carthage Cotton Valley gas field, east Texas: Geophysics, **68**, 441–452, doi: [10.1190/1.1567214](https://doi.org/10.1190/1.1567214).
- Sethi, H. S., and B. Shekar, 2017, Full waveform inversion for microseismic events using sparsity constraints: 87th Annual International Meeting, SEG, Expanded Abstracts, 2888–2892, doi: [10.1190/segan2017-17687815.1](https://doi.org/10.1190/segan2017-17687815.1).
- Sharan, S., R. Wang, T. van Leeuwen, and F. J. Herrmann, 2016, Sparsity-promoting joint microseismic source collocation and source-time function estimation: 86th Annual International Meeting, SEG, Expanded Abstracts, 2574–2579, doi: [10.1190/segan2016-13871022.1](https://doi.org/10.1190/segan2016-13871022.1).
- Shustak, M., and E. Landa, 2017, Time reversal based detection of subsurface scatterers: 87th Annual International Meeting, SEG, Expanded Abstracts, 969–973, doi: [10.1190/segan2017-17640587.1](https://doi.org/10.1190/segan2017-17640587.1).
- Wang, H., and T. Alkhalifah, 2016, Micro-seismic imaging using a source independent full waveform inversion method: 86th Annual International Meeting, SEG, Expanded Abstracts, 2596–2600, doi: [10.1190/segan2016-13946573.1](https://doi.org/10.1190/segan2016-13946573.1).
- Xuan, R., and P. Sava, 2010, Probabilistic microearthquake location for reservoir monitoring: Geophysics, **75**, no. 3, MA9–MA26, doi: [10.1190/1.3417757](https://doi.org/10.1190/1.3417757).
- Zhebel, O., and L. Eisner, 2014, Simultaneous microseismic event localization and source mechanism determination: Geophysics, **80**, no. 1, KS1–KS9, doi: [10.1190/geo2014-0055.1](https://doi.org/10.1190/geo2014-0055.1).