submitted to *Geophys. J. Int.*

Fast sparsity-promoting microseismic source estimation

Shashin Sharan[1,4], Rongrong Wang[2] and Felix J. Herrmann[1,3]

[1] *Seismic Laboratory for Imaging and Modeling (SLIM), the University of British Columbia*

[2] *now at Michigan State University*

[3] *now at School of Computational Science and Engineering, Georgia Institute of Technology*

[4] *now at School of Earth and Atmospheric Sciences, Georgia Institute of Technology*

Email: ssharan31@gatech.edu

SUMMARY

Microseismic events are generated during hydraulic fracturing of unconventional reservoirs and carry information on fracture locations and the origin times associated with these microseismic events. For drilling purposes and to prevent hazardous situations, we need to have accurate knowledge on the fracture locations as well as on their size, and their density. Because microseismic waves can travel far distances, microseismic data collected at the surface and or in boreholes can help us to monitor hydraulic fracturing. While so-called back propagation or time-reversal methods are able to focus recorded energy back onto the sources when a reasonable velocity model is available, these methods suffer from blurring especially in situations where the data acquisition suffers from lack of aperture, sparse sampling, and noise. As a result, these methods typically cannot resolve sources in close proximity, a desired feature since we need this information if we want to follow the fracture evolution in space and time. In that situation, we need to estimate the locations and the associated source-time functions for closely spaced microseismic sources along the active

fractures. To overcome the limitations of time-reversal methods, we propose a wave-equation based inversion approach where we invert for the complete source wavefield in both space and time. By promoting sparsity on the source wavefield in space, we negate the effects of non-radiating sources during the inversion and obtain high-resolution intensity plots and high-fidelity estimates for the source-time functions. We obtain these results relatively quickly by accelerating the linearized Bregman method with a dual formulation. Through experiments, we demonstrate that our method is computationally feasible, robust to noise, and works for closely spaced sources with overlapping source-time functions in complex geological settings.

Key words: Waveform inversion; Joint inversion; Induced seismicity.

1 INTRODUCTION

Unlike conventional reservoirs, unconventional plays are not naturally viable for economic production of oil and gas. They require stimulation by injecting high-pressure fluid causing fractures in the rocks. These fractures make the rock more permeable, hence, the extraction of oil and gas becomes feasible. For drilling purposes, and to prevent potentially hazardous situations or interference between wells, we need to have reliable information on the fracture locations, their density, and their propagation history. Microseismic waves, generated by fracturing in principle carry this information. In this work, we present a new wave-equation based inversion methodology that focuses observed pressure data back onto possibly closely spaced source locations enabling us to extract accurate location and origin time information. This latter information can subsequently be used to invert for the source mechanism (Madariaga 1989).

There exists an extensive literature on (micro)seismic source estimation, an active research topic that is met with many challenges forcing researchers to make certain sometimes unrealistic assumptions or to limit their inversions to certain aspects of the complete source mechanism, e.g. they estimate the source-time function or location but not both. Our approach aims to estimate both and derives from the assumption that microseismic sources are spatially sparsely distributed. Based on that assumption, we come up with a sparsity-promoting source inversion method for the acoustic wave equation that recovers the full source wavefield from data collected at the surface. While sparsity promotion and ℓ_1-norm minimization are ideas that have a long history in geophysics—we only have to think of the seminal work by Claerbout

& Muir (1973) , who used the ℓ_1-norm to handle outliers in the data followed by Taylor et al. (1979), Santosa & Symes (1986), and Donoho & Logan (1992) who used the ℓ_1-norm as an objective in seismic deconvolution–the use of mixed $\ell_{2,1}$-norm, also known as group sparsity promotion (Fornasier & Rauhut 2008; Sharan et al. 2016; Gao & Sacchi 2017), is relatively new. In this approach, sparsity is promoted along space while seeking solutions that have finite energy along time. Compared to other methods, our approach does not require picking of phases and it makes no other assumptions other than considering P-waves only.

In recent years, wave-equation based methods are being developed to address the incomplete physics and fragile nature of traveltime picks and inversions. While the premise of this approach is to estimate the source location as well as the origin time (possibly of multiple events), this approach is met by many challenges that include computational demand, large data volumes, limitations on the attainable resolution, and the existence of so-called non-radiating sources (Porter & Devaney 1982; Musafir 2013). Despite these challenges, there exists a wealth of different wave-equation based approaches that can be traced all the way back to early work by McMechan (1982), who proposed a wavefield extrapolation technique to back propagate observed data to focus the seismic source energy back onto its origin in space and time. In its original form, this method used the acoustic wave equation. Gajewski & Tessmer (2005) used a similar reverse-time modeling technique but extended it to elastic data. Even though this wavefield extrapolation maps the recorded waves back towards the sources, it needs to scan the resulting volume for points of maximal focusing to find the source locations and origin times. As expected, this combination of focusing and scanning is complicated by complex velocity models; by multiple microseismic events that occur in close vicinity at near coincident origin times; and by limited frequency content, aperture, and spatial sampling. Moreover, Abbe's diffraction limit (Abbe 1873) states that retro-focusing with a single back propagation cannot resolve sources within half a dominant wavelength. While potentially more powerful than travel-time based approaches, time-reversal techniques do require wide receiver aperture and adequate receiver and time sampling. Otherwise, the resolving power of time-reversal methods degrades quickly (Bazargani & Snieder 2016). Due to practical acquisition considerations, there are often situations where the receiver aperture or the receiver and time sampling are inadequate and time-reversal methods may fail especially when there is also noise.

Nakata & Beroza (2016) avoid computationally expensive scanning by reducing the scanning volume from 4D to 3D while at the same time increasing the spatial resolution of the imaged sources. These authors accomplish this by back propagating data from each receiver

individually, a method known as Geometric-mean Reverse-Time Migration (GmRTM). According to Nakata & Beroza (2016), this results in improved focussing because each received trace originates from the same seismic source. Despite its reduced scanning, GmRTM comes at a significant additional cost of having to solve wave equations for each receiver separately. This can be computationally demanding especially in 3D. To reduce these computational costs, Sun et al. (2015) proposed Hybrid Reverse-Time Migration (HyRTM). This method involves grouping of neighboring receivers and simultaneous back propagation of receiver data for each group, followed by a cross-correlation of back propagated data for each receiver group. If the velocity model is sufficiently accurate, this cross-correlation peaks at time zero. While less costly, this method reduces the resolving power, since it involves simultaneous back propagation of receiver data in each group. Moreover, it is not trivial to choose the right set of receivers for each group (Nakata & Beroza 2016). Finally Sun et al. (2016) mentioned that back propagation and cross-correlation alone may not preserve the phase information. For this reason, these authors propose to invert for the complete source wavefields using Least-Squares Time-Reversal Imaging (LSTRI). To recover correct amplitude and phase information, LSTRI introduces a sliding window in time and a threshold parameter to locate the back propagated wavefield sparsely. However, the success of this method depends on careful design of the window and of the threshold value.

To overcome issues with window design and parameter selection, our approach builds on ideas from full-waveform inversion (FWI) and LSTRI without making assumptions on the shape of the source-time function as in Wu & McMechan (1996) and Kim et al. (2011), who use FWI to invert for source locations, origin times, and moment tensor components without solving for the source-time function. Our work also differs from recent work by Kaderli et al. (2015), who alternate between inverting for the source-time function and the spatial distribution of the source wavefield. Even though this approach can be extended to multiple sources, it needs to know the number of sources a priori. Even with this knowledge, this method offers no guarantees of successful recovery because the optimization problem it solves is no longer convex. While Song et al. (2017) more recently removed the need to know the number of sources beforehand, they need to assume the same source-time function for all sources, an assumption also made by Wu & Alkhalifah (2017).

Instead of making assumptions on the number of events, which could lead to methods that attain super resolution with algorithms such as statistical multiple signal classification (MUSIC, Schmidt (1986)), we follow recent work by Kitić et al. (2016) and propose a sparsity-promoting method (Sharan et al. 2016) to invert for the complete source wavefield. This allows

us to jointly locate microseismic sources and their source-time function. By including sparsity in the objective along the spatial coordinates and energy along time, we mitigate as in Kitić et al. (2016) some of the issues related to the null space and lack of resolution. These can lead to non-radiating sources and cross talk between close by sources. Following our early contributions, Gao & Sacchi (2017) extended our approach that uses techniques from convex optimization to the elastic case. The contribution of this work is to make the method fast and computationally feasible by finding a dual formulation, which greatly reduces the number of iterations needed to converge in situations of interest—i.e., where microseismic sources are closely located. While our method can also be extended to the elastic case, we focus on the use of acoustic data to detect multiple events—i.e., determine their locations and origin times. We do this as part of an envisioned detection-estimation procedure where after detection, the estimated location and origin-time information is used as input in a second estimation step during which moment tensor inversions are carried out. In this way, we avoid the often computationally prohibitive costs of elastic wavefield simulations over large domains.

The outline is as follows: We first further motivate our formulation where we jointly estimate the source-time function and spatial distribution of microseismic sources. We numerically demonstrate that the source localization problem cannot be resolved by simple back propagation or by inversion with a ℓ_2-norm objective. Next, we show how including a sparsity-promoting objective addresses this problem. To more quickly solve the source wavefield estimation problem with a sparsity-promoting objective, we introduce the linearized Bregman algorithm. To make our method computationally feasible, we introduce a dual formulation and an appropriate preconditioner. With these we are able to reduce the computational costs drastically. Through several numerical experiments, we evaluate the performance of our method for situations of increasing complexity including noise and realistic geology.

2 MOTIVATION FOR SPARSITY-PROMOTING SOURCE ESTIMATION

When reasonable reliable prior knowledge on the velocity model is available, time-reversal methods (Fink 1997) can be an important tool to focus receiver wavefields back to their correct source location and origin time without having prior knowledge on the number of sources and their associated source-time functions. However, as we mentioned earlier, time-reversal techniques cannot resolve closely spaced sources within half a wavelength (Abbe 1873).

Motivated by the lack of attainable resolution of time-reversal methods, we propose an alternative formulation that addresses this issue of low resolution. To simplify our message and to avoid non-essential computational challenges for the elastic case, our method works

with acoustic data only. We justify this choice by making the assumption that the acoustic wave equation and elastic wave equation are kinematically the same for P-waves (Alkhalifah 2000); that P-waves dominate; and that differences in amplitudes will not too much impede our ability to detect seismic events by inverting for the source wavefield, followed by deriving location and origin time information. As we mentioned before, this information can then be used in moment tensor inversions (Madariaga 1989).

Sparsity can serve as a strong prior allowing for solutions of inverse problems that were previously thought unsolvable. For instance, in the field of Compressive Sensing (Candès et al. 2005; Donoho 2006) densely sampled signals are reconstructed from sub-Nyquist sampling by making use of the fact that natural signals are often sparse or compressible. In our context, this corresponds to seismic records with relatively few large events among many negligible smaller events. In that situation, we can locate the relatively strong events by promoting sparsity. Before we discuss how to promote sparsity, we first briefly describe the forward problem that describes how microseismic events are mapped to the data and the challenges one encounters when inverting this forward map.

2.1 The forward model

We start our formulation by expressing microseismic data, collected in the vector $\mathbf{d}$, as

$$\mathbf{d} = \mathcal{F}[\mathbf{m}](\mathbf{Q}), \tag{1}$$

where $\mathcal{F}[\mathbf{m}] = P\mathcal{A}[\mathbf{m}]^{-1}$ is the linear forward operator given by the time-domain acoustic forward modeling operator. The linear operator P restricts propagated wavefields to the receiver locations. The operator $\mathcal{A}[\mathbf{m}]$ implements a finite-difference approximation of the wave equation and acts on a vectorized wavefield, followed by a matricization. This operator is parameterized by an assumed to be known background squared slowness $\mathbf{m}$, defined as the reciprocal of the square of the acoustic wavespeed. The matrix $\mathbf{Q} \in \mathbb{R}^{n_x \times n_t}$, with n_x the size of the spatial grid and n_t the number of time samples, contains the spatial-temporal distribution of the source wavefield—i.e. the $(i,j)^{th}$ entry in $\mathbf{Q}_{i,j} = q(\mathbf{x}_i, t_j)$ corresponds in 2D to the source wavefield sampled at the spatial position $\mathbf{x}_i = (x_i,\, z_i)$ and $t_j = (j-1)\Delta t$ with Δt the sample interval in time. In our problem, the discretized source wavefield $\mathbf{Q}$ is the unknown that needs to be recovered from discrete data typically collected at the surface.

A quick way to recover the source wavefield $\mathbf{Q}$ is to back propagate the measured data. However, this approach suffers from poor focusing as we observe when we back propagate the microseismic data plotted in Fig. 1a, which we generated with the acquisition setup of Fig. 1b.

We obtained this blurred image (Fig. 2a) for the spatial distribution of the sources by summing at each gridpoint the absolute value along time of the back-propagated source wavefield. As we know, applying the adjoint of $\mathcal{F}[\mathbf{m}]$ does not correspond to inverting the forward map, which is partly responsible for the observed blurring in Fig. 2a.

2.2 Minimal energy solutions

Eq. 1 clearly has a null space because we only collect data at the receivers typically located at the surface while the source wavefield $\mathbf{Q}$ lives everywhere in the subsurface. To overcome possible instabilities related to this null space, we cast the solution of eq. 1 as a constrained optimization problem:

$$\begin{aligned} \underset{\mathbf{Q}}{\text{minimize}} \quad & \|\mathbf{Q}\|_F \\ \text{subject to} \quad & \\ \|\mathcal{F}[\mathbf{m}](\mathbf{Q}) - \mathbf{d}\|_2 \leq \epsilon, & \end{aligned} \tag{LS$_\epsilon$}$$

that includes the Frobenius norm $\|.\|_F$—i.e., the ℓ_2-norm of the vectorized data, in its objective. By minimizing this objective, we find a source wavefield such that it has the smallest energy while fitting the observed data $\mathbf{d}$ to within a user specified parameter ϵ that depends on the noise level. This constrained formulation is the time-domain counterpart of the time-harmonic constrained optimization problems that Devaney (2012) discusses.

Unfortunately, inverting eq. 1 by minimizing energy of the unknown source wavefield is inadequate as we can see by juxtaposing Figs 2a and 2b. While it is relatively easy to state the forward model (eq. 1), which relates the source wavefield to the observed data, inverting this relationship is hampered by the large null space of $\mathcal{F}$. As a result of this null space, which so-called non-radiating components may enter into the solution. These non-radiating sources do not contribute to wavefields measured on a surface outside some finite region enclosing the sources and lead to blurring and artifacts.

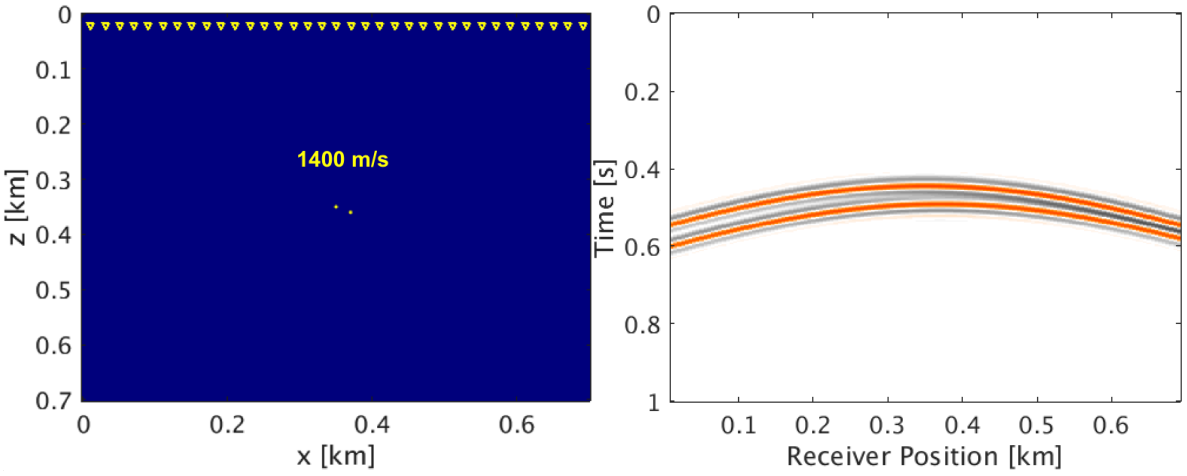
2.3 Problem with non-radiating sources

It is well known that the inverse source problem governed by the acoustic equation has infinitely many solutions when the source distribution lies in a closed volume and receivers are placed on the boundary or outside of the volume (Porter & Devaney 1982; Musafir 2013). This non-uniqueness is due to the existence of so-called non-radiating sources which lie in the null space of the forward modeling operator. In theory, these silent sources can never be detected or recovered by receivers outside a domain of radius R unless internal probing

is performed. Mathematically speaking, non-radiating sources are vectors in the kernel of the forward operator $\mathcal{A}$, therefore they do not contribute to the observations. Let $P_{\ker \mathcal{A}}$ be the projection operator onto the kernel. For any source distribution $\mathbf{Q}$, the decomposition $\mathbf{Q} = P_{\ker \mathcal{A}}\mathbf{Q} + (1 - P_{\ker \mathcal{A}})\mathbf{Q}$ implies that $\mathbf{Q}$ consists of a non-radiating part $P_{\ker \mathcal{A}}\mathbf{Q}$ and its orthogonal complement $(1 - P_{\ker \mathcal{A}})\mathbf{Q}$, called the radiating part. For any $\bar{\mathbf{Q}} \in \ker \mathcal{A}$, $\mathbf{Q} + \bar{\mathbf{Q}}$ generates the same wavefields as $\mathbf{Q}$ would have so these two wavefields are indistinguishable based on the observations alone. As a result, without prior knowledge, one can only recover the so-called quotient space of $\mathbf{Q}$ modulo $\ker(\mathcal{A})$ instead of $\mathbf{Q}$ itself. In Porter & Devaney (1982), the authors showed that the solution that minimizes the ℓ_2-norm energy is unique. However, due to the orthogonality of the radiating (row space) and non-radiating (kernel space) parts, the ℓ_2 minimizer recovers only the radiating part $(1 - P_{\ker \mathcal{A}})\mathbf{Q}$, and thus becomes inaccurate if the actual source contains non-radiating energy, i.e., $P_{\ker \mathcal{A}}\mathbf{Q} \neq \mathbf{0}$.

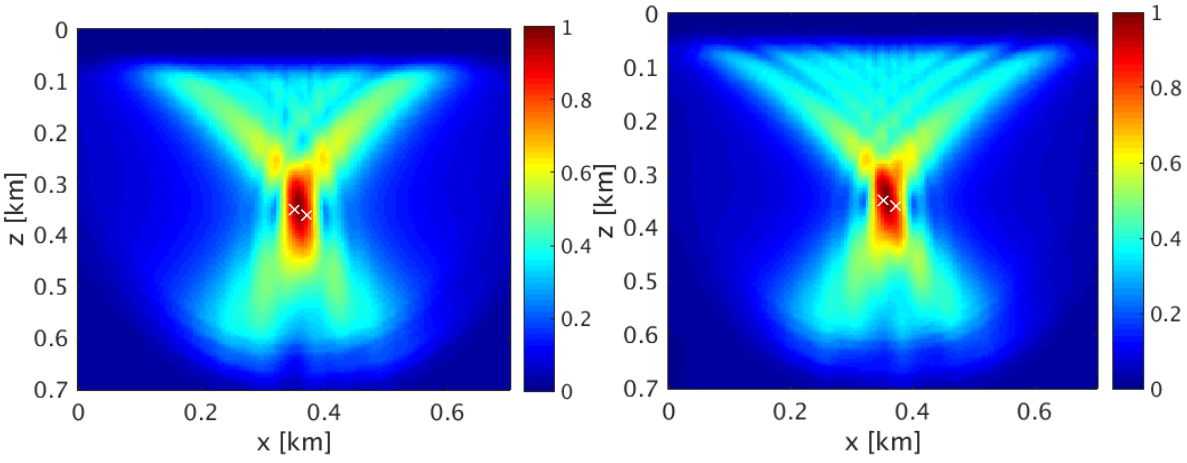
To overcome this problem, we make the assumption that the spatial distribution of microseismic events can be represented by a superposition of point sources suggesting that time snapshots of the microseismic source wavefields are sparse. Stress changes induced by applied hydraulic pressure cause the creation of fractures. This fracturing process is accompanied by emissions of localized microseismic energy at the fracture tips. Because fractures are interspersed sparsely in space, we argue that microseismic event locations are sparse as well. As demonstrated by Compressive Sensing (Candès et al. 2005; Donoho 2006), minimal energy solutions of underdetermined systems are almost never sparse and hence a sparsity-promoting norm such as the ℓ_1-norm is needed to replace the ℓ_2-norm. By reconstructing a sparse source solution, one then not only recovers the radiating part (i.e., the minimal energy solution), but also the non-radiating part of the source or else the solution would not have been sparse. We can observe this phenomenon in Fig. 2b where the intensity (sum of absolute values over time) of the minimal-energy solution is spread out spatially rather than being focussed onto the point sources that were responsible for generating the observed data according to eq. 1. This phenomenon is predicted by the above theoretical arguments and corresponds to the non-radiating part of the sources. Because we are missing these non-radiating contributions, we are not able to recover the point sources.

Following recent work by Kitić et al. (2016), who worked on source localization problems in room acoustics, and the argument that sparse solutions almost never coincide with minimal energy solutions, we propose to minimize the sparsity-promoting ℓ_1 norm of the sources along the space coordinates, while still minimizing the ℓ_2 norm in time, as we assume the sources to be of finite energy but not localized in time. If we apply this combined $\ell_{2,1}$-norm defined



(a) (b)

Figure 1. Nearby sources experiment. *(a)* Velocity model with acquisition geometry and microseismic source locations. The inverted yellow color triangles indicate receivers buried at a depth of 20.0 m and separated by 5.0 m. The yellow dots indicate the location of 2 microseismic sources. *(b)* Microseismic data generated by two microseismic sources.



(a) (b)

Figure 2. *(a)* Back-propagated source image, *(b)* Intensity plot of the minimal energy solution (cf. eq. LS_ϵ). White colour crosses indicate the actual location of two point sources.

as $\|\mathbf{Q}\|_{2,1} = \sum_i \mid (\sum_j \mathbf{Q}_{i,j}^2)^{\frac{1}{2}} \mid$ for an unknown source wavefield $\mathbf{Q}$, we are able to recover the individual point sources accurately (juxtapose Figs 2b and 3d). Below, we present algorithmic details on how these results can be accomplished at greatly reduced computational costs.

3 SPARSITY-PROMOTING SOURCE LOCALIZATION

The observation that microseismic event locations are sparse in space supports the above theoretical arguments to impose the sparsity-promoting ℓ_1-norm along the spatial coordinates. Because stress drops are finite during these micro-seismic events, the emitted energy will be finite as well justifying the use of the ℓ_2-norm along time. To exploit sparsity in space and finite-energy along time, we recast eq. LS_ϵ into

$$\begin{aligned} \underset{\mathbf{Q}}{\text{minimize}} \quad & \|\mathbf{Q}\|_{2,1} \\ \text{subject to} \quad & \\ \|\mathcal{F}[\mathbf{m}](\mathbf{Q}) - \mathbf{d}\|_2 \leq \epsilon. & \end{aligned} \tag{$\text{LS}_\epsilon^{2,1}$}$$

By solving problem $\text{LS}_\epsilon^{2,1}$, we find a source wavefield $\mathbf{Q}$ that has the smallest ℓ_1-norm in space and the smallest ℓ_2 -norm in time (denoted by the norm $\|\cdot\|_{2,1}$) while fitting the observed data $\mathbf{d}$ to within ϵ. From the estimated source wavefield $\mathbf{Q}$, we find the location of the microseismic sources by finding outliers in the intensity plot. We calculate these by taking the absolute sum of the estimated source wavefield along time at each point in space —i.e., we have $\mathbf{I}(\mathbf{x}) = \text{vec}^{-1}\left(\sum_t \mid \mathbf{Q}(\mathbf{x}, t) \mid\right)$ with $\text{vec}^{-1}(\cdot)$ representing a reshape into the original matrix form. After finding the source locations from the outliers in $\mathbf{I}(\mathbf{x})$, we calculate the source-time functions associated with these events by extracting the time variations of the wavefield at these locations.

For a given background slowness squared vector $\mathbf{m}$, eq. $\text{LS}_\epsilon^{2,1}$ takes a form similar to the classic Basis Pursuit Denoising (BPDN) problem (Chen et al. 1998; van den Berg & Friedlander 2008) known from Compressive Sensing and sparsity-promoting inverse problems. In the next section, we will introduce a new algorithm tailored to solve a modified form of the original BPDN problem $\text{LS}_\epsilon^{2,1}$ for situations where $\mathcal{F}[\mathbf{m}]$ is large, expensive to compute, and ill-conditioned.

3.1 Linearized Bregman algorithm

Motivated by recent successful application of the linearized Bregman method (Yin et al. 2008; Lorenz et al. 2014) to sparsity-promoting least-squares migration (Herrmann et al. 2015; Witte et al. 2018), we make the objective strongly convex by relaxing the $\|.\|_{2,1}$ norm to a sum of

$\|.\|_{2,1}$ and the Frobenius norm $\|.\|_F^2$ —i.e., we solve

$$\begin{aligned} \underset{\mathbf{Q}}{\text{minimize}} \quad & \|\mathbf{Q}\|_{2,1} + \frac{1}{2\mu}\|\mathbf{Q}\|_F^2 \\ \text{subject to} \quad & \\ & \|\mathcal{F}[\mathbf{m}](\mathbf{Q}) - \mathbf{d}\|_2 \leq \epsilon, \end{aligned} \tag{LBR$^{2,1}_{\mu,\epsilon}$}$$

where the parameter μ controls the trade-off between the sparsifying $\ell_{2,1}$-norm and the strong convexity. When $\mu \uparrow \infty$, the solution of eq. LBR$^{2,1}_{\mu,\epsilon}$ approaches the solution of eq. LS$^{2,1}_{\epsilon}$ (Osher et al. 2010). Because the objective in eq. LBR$^{2,1}_{\mu,\epsilon}$ now contains a quadratic term, problem LBR$^{2,1}_{\mu,\epsilon}$ becomes easier to solve (Boyd & Vandenberghe 2009), because it is strongly convex. Another advantage of this approach is that the linearized Bregman algorithm permits a simple three line algorithm for its solution with tuning parameters that are arguably easy to choose. Compared to the original linearized Bregman method designed to solve ℓ_1 minimization problems (Yin et al. 2008), Algorithm 1 can be considered as an extension to the "matrix" case where the ℓ_1 and ℓ_2 norm are enforced jointly as in group sparsity problems (Fornasier & Rauhut 2008; Gao & Sacchi 2017). Despite this important difference, the resulting algorithm still shares the simplicity of the original linearized Bregman as we see from Algorithm 1.

Algorithm 1 Microseismic source collocation with linearized Bregman.

1.`Data` $\mathbf{d}$ `&  slowness square` $\mathbf{m}$ `//Input`

2. **for** k=1,2,...

3. $\quad \mathbf{V}_k = \mathcal{F}[\mathbf{m}]^\top(\Pi_\epsilon(\mathcal{F}[\mathbf{m}](\mathbf{Q}_k) - \mathbf{d}))$ `//adjoint solve`

4. $\quad \mathbf{Z}_{k+1} = \mathbf{Z}_k - t_k\mathbf{V}_k$ `//auxiliary variable update`

5. $\quad \mathbf{Q}_{k+1} = \text{Prox}_{\mu\ell_{2,1}}(\mathbf{Z}_{k+1})$ `//sparsity promotion`

6. **end**

7. $\mathbf{I}(\mathbf{x}) = \text{vec}^{-1}\left(\sum_t \mid \mathbf{Q}(\mathbf{x}, t) \mid\right)$ `//Intensity plot`

According to Lorenz et al. (2014), we can cheaply calculate a dynamic step length

$$t_k = \frac{\|\mathcal{F}[\mathbf{m}](\mathbf{Q}_k) - \mathbf{d}\|^2}{\|\mathcal{F}[\mathbf{m}]^\top(\mathcal{F}[\mathbf{m}](\mathbf{Q}_k) - \mathbf{d})\|_F^2} \tag{2}$$

that guarantees Algorithm 1 to converge (Lorenz et al. 2014). In this expression, we reserve the symbol $^\top$ for the matrix transpose.

During each iteration of Algorithm 1, we subtract from the residue $\mathbf{r} \equiv \mathcal{F}[\mathbf{m}](\mathbf{Q}_k) - \mathbf{d}$ its projection onto the ℓ_2 ball of size ϵ using the following operator:

$$\Pi_\epsilon(\mathbf{r}) = \max\left\{0, 1 - \frac{\epsilon}{\|\mathbf{r}\|}\right\}\mathbf{r}. \tag{3}$$

After projecting the residual at iteration k on this ℓ_2-norm ball, we apply the proximal operator

$\text{Prox}_{2,1}$ to the updated auxiliary variable $\mathbf{Z}_{k+1}$ to get a new sparse estimate for the source wavefield $\mathbf{Q}_{k+1}$. The action of this proximal operator (Combettes & Pesquet 2011) on a matrix $\mathbf{C}$ is defined by

$$\text{Prox}_{\mu\ell_{2,1}}(\mathbf{C}) := \arg\min_{\mathbf{B}} \|\mathbf{B}\|_{2,1} + \frac{1}{2\mu}\|\mathbf{C}-\mathbf{B}\|_F^2\,, \tag{4}$$

which involves an elementwise soft thresholding (Kowalski & Torrésani 2009)—i.e., we have for the $(i,j)^{th}$ element of $\mathbf{C}$

$$\big(\text{Prox}_{\mu\ell_{2,1}}(\mathbf{C})\big)_{i,j} = S(\mathbf{C}_{i,j}, \frac{\mu}{\|\mathbf{C}_{i,:}\|^2}), \tag{5}$$

where $\mathbf{C}_{i,:}$ refers to the ith row of $\mathbf{C}$ and where

$$S(a,T) = \max\{|a|-T,0\}\text{sign}(a) \tag{6}$$

is the soft thresholding operator. From eq. 5, we see that the threshold parameter for a proximal operator depends on the tradeoff parameter μ, as well as on the ℓ_2 norm of the rows of $\mathbf{C}$, $\mathbf{C}_{i,:}$. As a result, we use different threshold levels for different rows. In the context of our problem, locations with high source energy along time have smaller threshold parameters in comparison to the threshold parameters of locations with lower source energy along time. Therefore, locations with higher source energy are more likely to survive after thresholding.

Before we explain how to apply and accelerate Algorithm 1, let us first briefly explain what the different steps of Algorithm 1 mean physically. **Line 3** entails a back projection of the residual projected onto the ℓ_2-norm ball of size ϵ. The resulting matrix $\mathbf{V}_k$ is subsequently used to update the auxiliary variable $\mathbf{Z}_k$ in **line 4**. To focus the resulting estimate for the source wavefield, we apply the proximal operator in **line 5**, which is designed to sparsify the new source wavefield $\mathbf{Q}_k$ along the spatial coordinates while minimizing the energy along time.

While Algorithm 1 is relatively simple to implement, the ability to recover closely located point sources may come at a high computational price because of the high costs of evaluating many $\mathcal{F}[\mathbf{m}]$'s and $\mathcal{F}^\top[\mathbf{m}]$'s each of which entails solving the wave equation over the whole domain. These costs are compounded by the fact that $\mathcal{F}[\mathbf{m}]$ has increasingly bad condition numbers as the size of the domain and the temporal frequency content increases. Having said this, juxtaposition of recoveries via eqs LS_ϵ and $\text{LBR}_{\mu,\epsilon}^{2,1}$ (cf. Figs 2b and 3d) clearly shows the ability of sparsity-promotion to locate closely spaced microseismic sources.

3.2 Dual formulation and acceleration with quasi-Newton

As we discussed earlier, microseismic sources are typically localized along filamentous structures generated by hydraulic fracturing. To track the evolution of these fractures amongst both

space and time, we aim to locate the individual microseismic events at an as high as possible resolution. This need for an accurate map of the fracture propagation calls for recovery of closely spaced events.

While we can expect under certain circumstances eq. $\text{LBR}^{2,1}_{\mu,\epsilon}$ to recover events well within half a wavelength, we are only able to obtain these results with linearized Bregman for a relatively high value of the trade-off parameter μ, as higher values of μ better approximate the behaviour of the original $\ell_{2,1}$ minimization problem $\text{LS}^{2,1}_{\epsilon}$. This problem is known to promote sparsity and to attain super resolution under certain conditions. Unfortunately, increasing the trade-off parameter μ increases the number of iterations needed to convergence drastically as we can see in Fig. 3 where we aim recover two events separated by 22.0 m (Fig. 1a), which is very close to half of the dominant wavelength of 23.0 m. By dominant wavelength, we mean the wavelength corresponding to the central frequency of the source-time functions of microseismic sources. To make the recovery more challenging, we allow the corresponding source-time functions to partially overlap. Also, the constant velocity medium makes the recovery of these two microseismic sources a bit more challenging because of the relative low diversity of wavenumbers in the absence of scattering from inhomogeneities.

Even though the scenario that we seek to resolve, in which two microseismic sources are closely spaced (Fig. 1a), is relatively simple, its solution is much more challenging because of the poor conditioning of $\mathcal{F}[\mathbf{m}]$. As a result, we should expect a need for large μ's, to emphasize the sparsity, and as a consequence many iterations. To confirm this behavior, let us consider the normalized data residual, $\|\mathcal{F}[\mathbf{m}](\mathbf{Q}_k)-\mathbf{d}\|/\|\mathbf{d}\|$, with $\mathbf{Q}_k$ the solution of Algorithm 1 after k iterations. We find that after 200 iterations, we have a normalized residual of less than 0.05 for $\mu =$ 8e-5. For this value of μ we are not able to resolve the two sources (Fig. 3a). Even if we increase $\mu =$ 8e-4 and run for 600 iterations, the sources are still not resolved (Fig. 3b). Only by increasing μ to 8e-3 we are able to get a reasonable result (Fig. 3c) but at the expense of running 4900 iterations of the linearized Bregman Algorithm 1, which is too many for the method to be viable in practice since every iteration requires solving at least one forward and adjoint wave equation.

To overcome this problem of too slow convergence, we follow the work by Yin (2010) who propose a dual formulation for linearized Bregman with the aim to converge faster. Since eq. $\text{LBR}^{2,1}_{\mu,\epsilon}$ is strongly convex, it has a smooth dual objective (Shalev-shwartz & Singer 2006) whose gradient decent steps correspond to the linearized Bregman of the original Basis pursuit denoising problem. By virtue of this smoothness, we can accelerate the solution of the dual

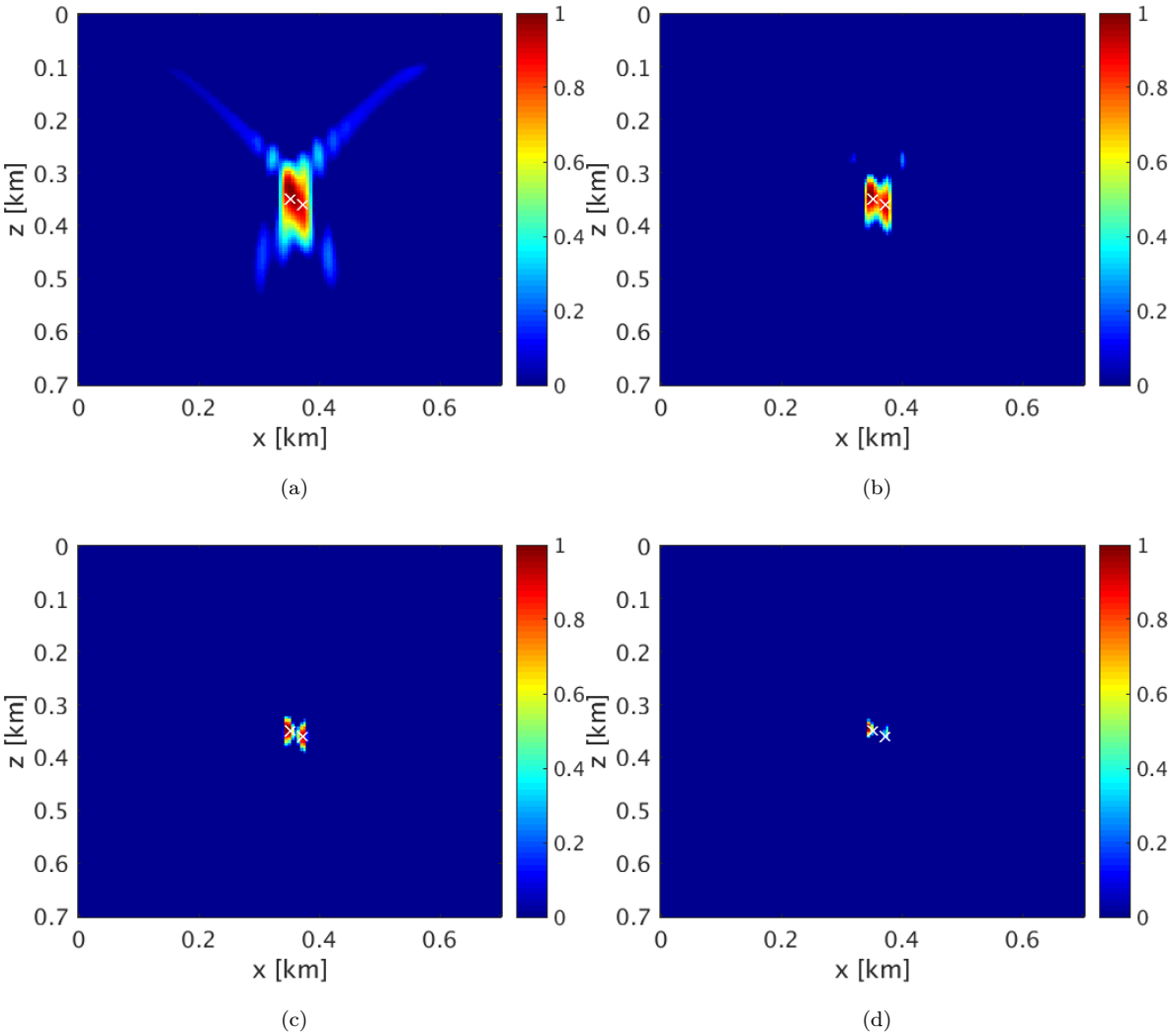


Figure 3. Estimated intensity plots. *(a)* After 200 iterations of linearized Bregman with μ = 8e-5, *(b)* after 600 iterations of linearized Bregman with μ = 8e-4 and *(c)* after 4900 iterations of linearized Bregman with μ = 8e-3. *(d)* After only 10 preconditioned iterations of the dual formulation for linearized Bregman. Iterations of the dual formulation cost about the same as iterations of the primal problem (Algorithm 1). White colour crosses indicate the actual locations of microseismic sources

problem with Limited-memory BFGS (L-BFGS) (Liu & Nocedal 1989), an approach also taken by Yin (2010).

To arrive at the accelerated form of linearized Bregman, we write down its Fenchel's dual—–i.e., we have (more details in the Appendix)

$$\underset{\mathbf{y}}{\text{minimize}} \quad f(\mathbf{y}) = -\left\{\Psi(\mathbf{y}) - \epsilon\|\mathbf{y}\|_2\right\}, \tag{7}$$

where

$$\Psi(\mathbf{y}) = \underset{\mathbf{Q}}{\text{minimize}} \quad \|\mathbf{Q}\|_{2,1} + \frac{1}{2\mu}\|\mathbf{Q}\|_F^2 + \mathbf{y}^\top\big(\mathcal{F}[\mathbf{m}](\mathbf{Q}) - \mathbf{d}\big), \tag{8}$$

is a value function that depends on the dual variable $\mathbf{y}$. Notice that this dual variable $\mathbf{y}$ is a vector that has dimensions of the data $\mathbf{d}$, which is much smaller than the dimensions of the unknown source wavefield $\mathbf{Q}$ that lives everywhere and not only at the receiver locations. The value of $f(\mathbf{y})$ itself traces the minimal value of the above objective as a function of the dual variable. The value function Ψ is differentiable and permits a closed form derivative (as explained in the Appendix) that equals

$$\nabla\Psi(\mathbf{y}) = \mathbf{d} - \mathcal{F}[\mathbf{m}]\big(\text{Prox}_{\mu\ell_{2,1}}(\mu\mathcal{F}[\mathbf{m}]^\top(\mathbf{y})\big). \tag{9}$$

Using the dual objective function $f(\mathbf{y})$ from eq. 7 and the derivative $\nabla\Psi(\mathbf{y})$ of the value function $\Psi(\mathbf{y})$ we are able to compute the gradient of $f(\mathbf{y})$ via

$$\nabla f(\mathbf{y}) = -\left\{\nabla\Psi(\mathbf{y}) - \epsilon\mathbf{y}/\|\mathbf{y}\|_2\right\}. \tag{10}$$

Aside from being differentiable with respect to $\mathbf{y}$, the dual formulation has the advantage that $\mathbf{y}$ lives in a much smaller space. This allows us to use L-BFGS, which needs access to previous gradient vectors to approximate the inverse of the quasi-Newton Hessian. Compared to the primal formulation cf. Algorithm 1, where we would have needed access to multiple instances of the full source wavefield, the dual formulation only needs storage of vectors of the size of the observed data, which is feasible. This is because the size of the observed data $(n_r \times n_t)$, with n_r being the number of receivers, is much smaller than the size of the full source wavefield $(n_x \times n_t)$. The resulting method is summarized in Algorithm 2. The cost per iteration of this algorithm are approximately the same as those of Algorithm 1.

Algorithm 2 Acceleration with L-BFGS.

1. Data $\mathbf{d}$, slowness square $\mathbf{m}$, number of iterations l `//Input`
2. Initialize dual variable $\mathbf{y} = 10^{-3}\mathbf{d}$
3. $\hat{\mathbf{y}} = \text{L-BFGS}(f(\mathbf{y}), \nabla f(\mathbf{y}), \mathbf{y}, l)$ `//Dual solution`
4. $\mathbf{Q} = \text{Prox}_{\mu\ell_{2,1}}(\mu\mathcal{F}[\mathbf{m}]^\top(\mathbf{y}))$ `//Primal solution`
5. $\mathbf{I}(\mathbf{x}) = \text{vec}^{-1}\left(\sum_t \mid \mathbf{Q}(\mathbf{x}, t) \mid\right)$ `//Intensity plot`

This algorithm includes the following main steps: **Line 2** applies a scaling of the observed data to better numerically condition the problem. The solution does not depend critically on the particular value of this scaling. **Line** 3 solves the dual problem with $l \ll k$ iterations of L-BFGS for the dual objective function in eq. 7 and its gradient in eq. 10. **Line** 4 corresponds

to computing the primal solution by back propagating the dual solution with the adjoint of the forward modeling operator (cf. eq. 1), followed by applying a scaling by μ, and an additional application of the Prox operator. As before, we calculate the intensity plot in **Line** 5 via $I(\mathbf{x}) = \sum_t \mid \mathbf{Q}(\mathbf{x}, t) \mid$.

3.3 Preconditioning in 2D

As for the primal formulation (Algorithm 1), each iteration of L-BFGS in Algorithm 2 requires solving at least one forward and an adjoint wave equation. Although, Algorithm 2 is known to converge in far fewer iterations compared to the original linearized Bregman algorithm, the sheer size and computational costs of our problem call for further reduction in the number of iterations. Recognizing that the slow convergence is mainly due to the relatively large condition number of the forward modeling operator $\mathcal{F}$, we apply a left scaling to partially address this issue and further accelerate the convergence of the proposed method. In two dimensions, a point source implicitly assumes a line source extending infinitely in the out of plane direction (Song & Williamson 1995). This causes waveforms and amplitudes to differ significantly from the waveforms and amplitudes of a true point source. Following Herrmann et al. (2009), we introduce a correction via a symmetric half-differentiation along the time direction as a left preconditioner—i.e., we modify the original problem $\text{LBR}^{2,1}_{\mu,\epsilon}$ into

$$\begin{aligned} \underset{\mathbf{Q}}{\text{minimize}} \quad & \|\mathbf{Q}\|_{2,1} + \frac{1}{2\mu}\|\mathbf{Q}\|_F^2 \\ & \text{subject to} \\ & \|\mathbf{M}_L^{-1}\left(\mathcal{F}[\mathbf{m}](\mathbf{Q}) - \mathbf{d}\right)\|_2 \leq \gamma \end{aligned} \tag{11}$$

with $\mathbf{M}_L^{-1} := \partial_{|t|}^{1/2}$ with $\partial_{|t|}^{1/2} = \mathbf{F}^{-1}|\omega|^{1/2}\mathbf{F}$ and where $\mathbf{F}$ is the temporal Fourier transform and $\mathbf{F}^{-1}$ its inverse. As before, the parameter γ depends on the ℓ_2-norm of the noise of the preconditioned data. Again, we solve problem 11 with L-BFGS.

If we apply our dual formulation along with the preconditioning to the source localization problem discussed in the last section (Figs 1a & 1b), we are now able to resolve close by sources (Fig. 3d) in only 10 iterations. Figs 4a and 4b show the corresponding source-time functions estimated by our method (denoted by the dash dot line) and compared to the true wavelet (denoted by solid lines). Except for the "side-lobe artifacts" and incorrect amplitudes, we resolve both the source locations and source-time functions accurately. When we increase the number of iterations to 30, the side lobes decrease (compare the dashed and dot dashed plots in Figs 4a and 4b). Because our method still suffers from an amplitude scaling ambiguity, we correct the estimated source-time functions by a constant factor of approximately 10. This

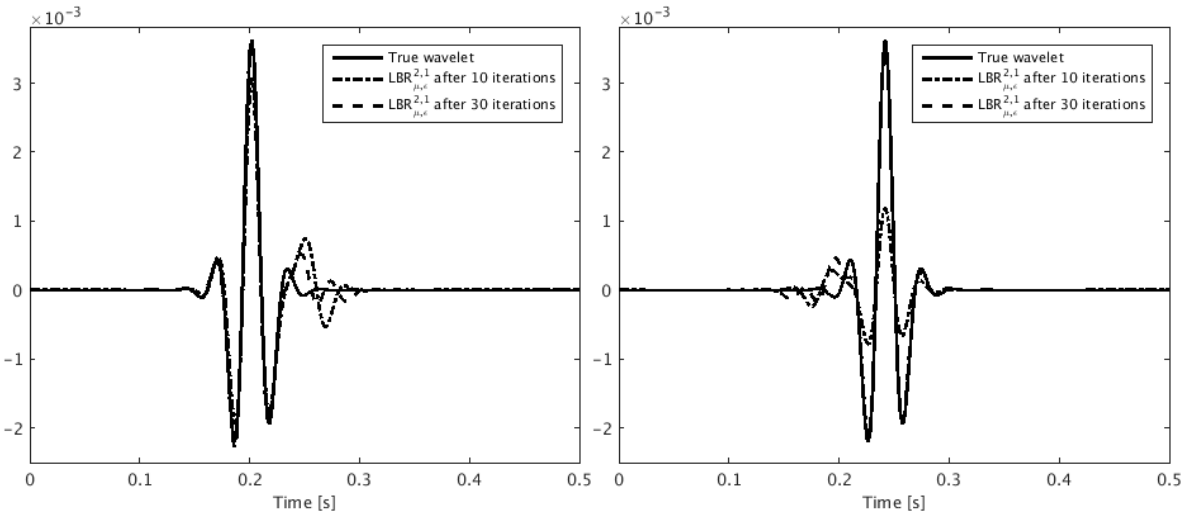


Figure 4. Source-time function comparison: True source-time functions (solid black), recovered source-time functions with 10 iterations (denoted by dash dot line) and with 30 iterations (dashed) of the accelerated dual formulation at the first in *(a)* and second source location in *(b)* from left to right. Estimated source-time functions are corrected by a factor of approximately 10 for comparison purposes.

factor corresponds to the smallest ratio of the maxima of the true and the corresponding recovered source-time functions.

To better understand the gains in computational efficiency and inversion improvements we achieve with our preconditioned dual formulation, we plot in Fig. 5 the normalized data residuals as a function of the number of iterations k for the linearized Bregman without (solid black line) and with preconditioning (denoted by dash dot line), and for L-BFGS without (denoted by dot symbol) and with preconditioning (dashed black line).

From this example, we clearly observe improved convergence for linearized Bregman with our preconditioner compared to without a preconditioner. We achieve more drastic improvements without and with preconditioned L-BFGS for the later (> 6) iterations. After only 10 iterations, we reach a satisfactory decrease in the residual compared to the 4900 iterations we need for linearized Bregman without preconditioning. The wild fluctuations in the beginning are an artifact of the dual formulation. These fluctuations are less for the preconditioned dual formulation. These fluctuations are of no concern because they damp out rapidly.

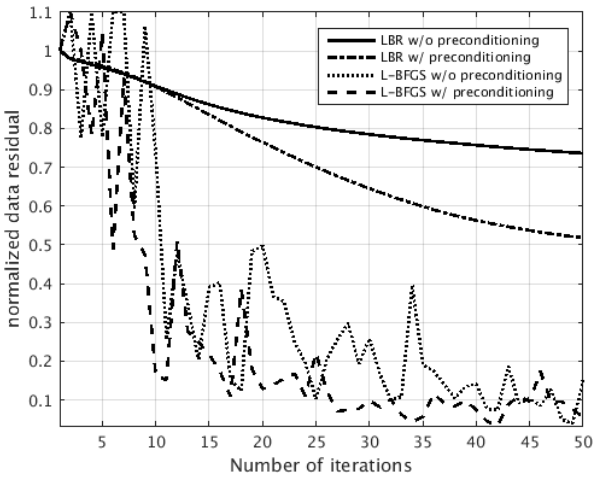


Figure 5. Convergence comparison. Decay of normalized ℓ_2 norm of data residual with linearized Bregman without preconditioning (solid black lines), with preconditioning (denoted by dash dot line) and L-BFGS without (denoted by dotted line) and with preconditioning (dashed).

4 NUMERICAL EXPERIMENTS

To compare the performance of our sparsity-promoting approach with the method of back propagation or minimum energy (cf. eq. LS_ϵ), we conduct a series of numerical experiments on parts of the synthetic BG compass (Fig. 6a) and Marmousi (Fig. 10a) velocity models. Both are sampled at 5 m. We selected these models because they represent different, geologically complex, representative geological areas. With its velocity kickback and well-constrained variability, the BG compass model (Li et al. 2012) was designed to test full-waveform inversion (van Leeuwen & Herrmann 2012; Li et al. 2016). We decided to use this model because it contains realistic small-scale variability and a large velocity jump that may challenge our method. We also consider the Marmousi model (Brougois et al. 1990), which is a more widely studied and representative of a geology with faults, which may also challenge our method. We also chose this model because it was recently used by Nakata & Beroza (2016) and Li & van der Baan (2016) to test their methods.

For reasons stated earlier, we use a 2D acoustic finite difference modeling code (Louboutin et al. 2017) to simulate microseismic data with a record length of 1 s. Compared to conventional active-source seismic, microseismic events can be considered as low in amplitude and high in frequency earthquake-like events (Kamei et al. 2015). This motivates our choice of using the earthquake source model by Madariaga (1989). Following Nakata & Beroza (2016); Sun et al. (2016), we also do not include a free-surface. To avoid interference with the absorbing boundary condition at the surface, we place our receivers at a depth of 20 m sampled at 10 m. To speed

up the iterations, we precondition both experiments. To more closely mimic field data, we add bandwidth limited noise (5.0 Hz to 45.0 Hz) to synthetic data we generate from the velocity models plotted in Figs 6a and 10a.

4.1 Two-source experiment in BG compass model

To validate the proposed method, we test our source-localization approach on noisy data with a signal-to-noise ratio (S/R) of only 3.21 dB (Fig. 6c) generated from the complex velocity model plotted in Fig. 6a with an aperture of almost 2.1 km. To compare the performance for a noisy but ideal situation where we model the wave physics correctly with a more realistic scenario where we only have access to a smooth non-reflective velocity model we invert this noisy data with the true and smoothed, but kinematically correct, velocity models plotted in Figs 6a and 6b. For the smoothed model, we only capture the kinematics and we miss wavenumber diversity, which should lead to inferior results. The subset of the BG compass model as shown in Fig. 6a has a size of 2.25 km $\times$ 0.915 km, which corresponds to 451 $\times$ 184 grid points.

To make the source localization problem more challenging, we place two microseismic sources, with source-time functions of central frequency 30 Hz (solid black plots in Figs 8a and 8b) close to the high velocity unconformity at depths of 485 m and 490 m and separated by approximately 20.0 m well within Abbe's diffraction limit. We indicated these source locations by the two black dots in Fig. 6a. We included results for the true and smoothed velocity models in the left and right columns of Fig. 7. To arrive at these results, we used μ = 2e-2, which we found after extensive parameter testing. Despite the strong and sharp unconformity below the two microseismic sources, our sparsity-promoting approach produces an intensity plot with two well resolved delta-like sources for both the actual and the smoothed velocity models after only ten iterations. As expected, the same cannot be said for the results obtained with LS_ϵ (Figs 7a and 7b). Even after 300 iterations, the events remain completely blurred due to the existence of non-radiating sources. Aside from producing a high-resolution intensity plot, our method also provides good estimates for the source-time functions. Figs 8a and 8b depict the true and estimated source-time functions for both source locations again obtained with the true and smoothed velocity models. To facilitate comparison, we again correct the amplitudes of the estimated wavelets by a factor of approximately 4.5. We obtained this factor as we described before.

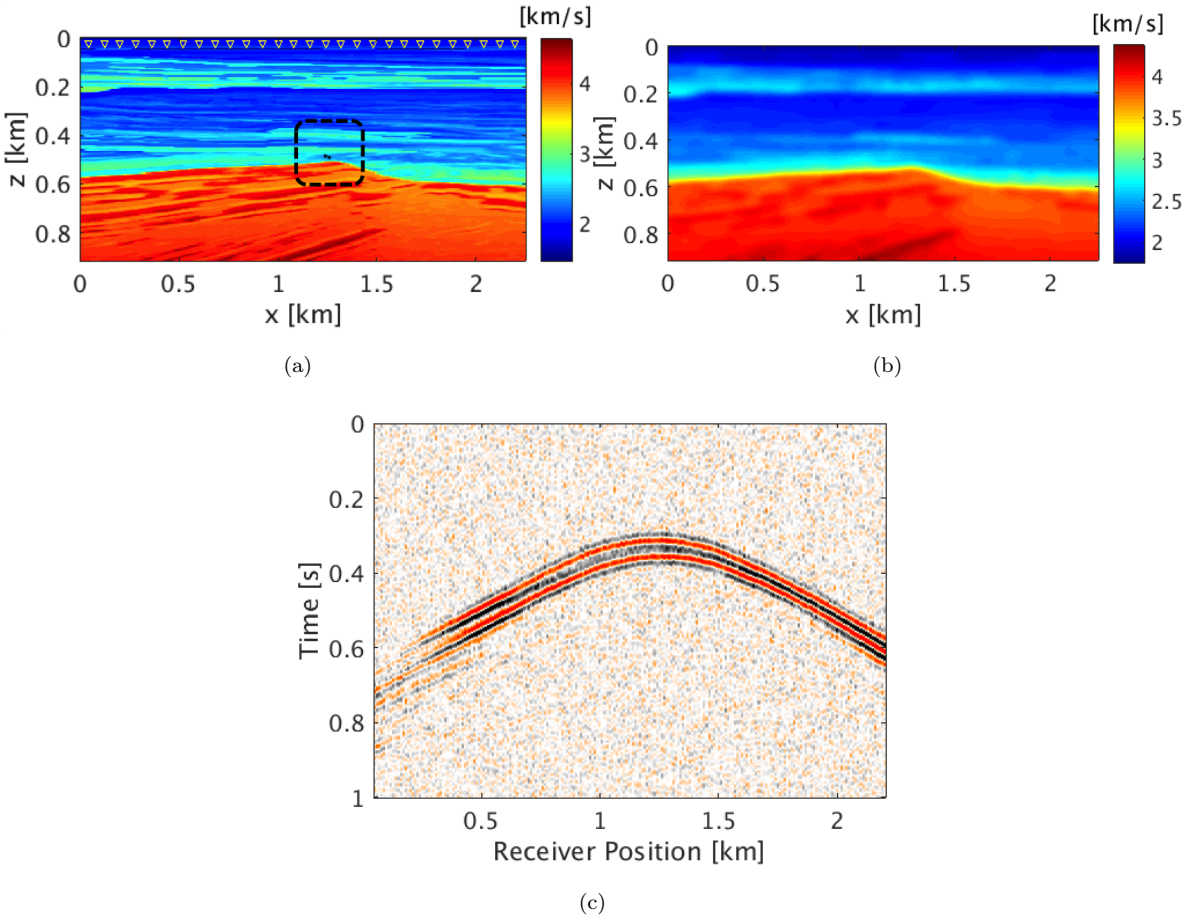


Figure 6. BG model source localization experiment. *(a)* True velocity model with acquisition geometry and microseismic source locations superimposed. The inverted yellow triangles indicate receivers buried at a depth of 20.0 m and sampled at 10.0 m. The black dots indicate the location of 2 nearby microseismic sources. *(b)* Kinematically correct smoothed velocity model used during the inversion. *(c)* Noisy simulated microseismic data generated by the two microseismic sources in the true velocity model. The synthetic data is contaminated with bandwidth limited random noise (5.0 Hz to 45.0 Hz) yielding a S/R of only 3.21 dB.

4.1.1 Earth Movers Distance — a more suitable performance metric

To further analyze the computational efficiency and inversion improvements we achieve with our preconditioned dual formulation, we plot in Figs 9a and 9b the normalized data residuals and the normalized Earth Movers Distance (EMD, Rubner et al. (1998)) between the estimated and the true intensity plots as a function of the number of iterations k. We propose to use the EMD because it measures the distance between two probability distributions $p(x)$ and $q(x)$ in terms of how much "mass" one has to move to map these two distributions into each other. This distance is better suited to measure the distance between delta-like distributions and is

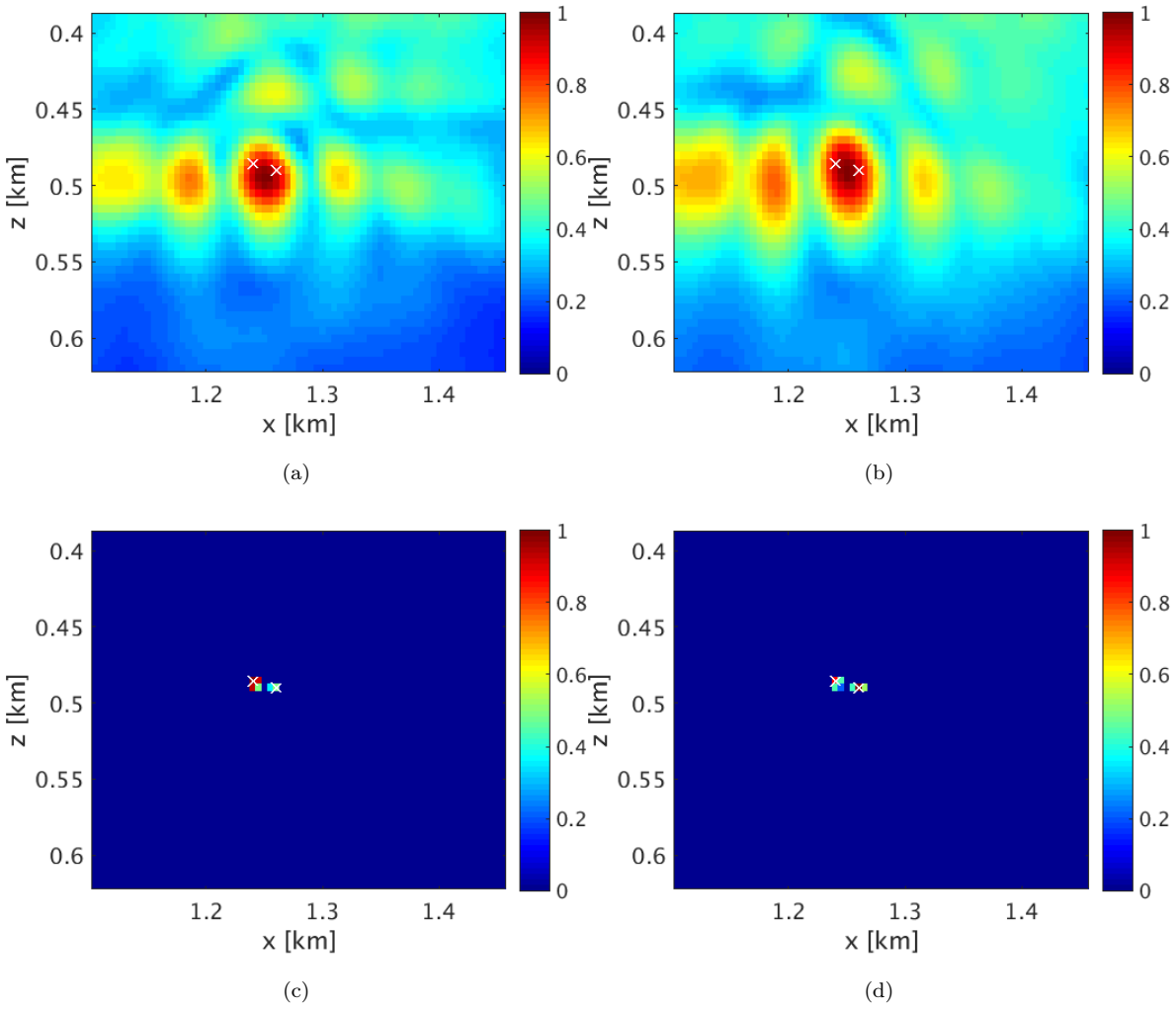


Figure 7. Zoomed intensity plots for the true (left column) and smoothed (right column) velocity models obtained by solving LS_ϵ with 300 iterations and plotted in *(a,b)*; and by solving $\mathrm{LBR}^{2,1}_{\mu,\epsilon}$ for $\mu =$ 2e-2 with only ten iterations and plotted in *(c,d)*. The white crosses indicate the actual spatial locations of the two microseismic sources. Notice the improved resolution.

defined as

$$\begin{aligned} &\min_{\mu \geq 0} \int d(x,y)\mu(x,y)\mathrm{d}x\mathrm{d}y \\ &\text{subject to } \int \mu(x,y)\mathrm{d}y = p(x) \\ &\qquad\qquad \int \mu(x,y)\mathrm{d}x = q(y) \end{aligned} \tag{12}$$

where $d(x,y) = \|x - y\|_2$.

In our experiments, the true intensity plots correspond to discrete delta functions located at the true source locations. This choice for the spatial distribution of the sources makes it difficult to quantitatively measure the performance of our algorithm as a function of the

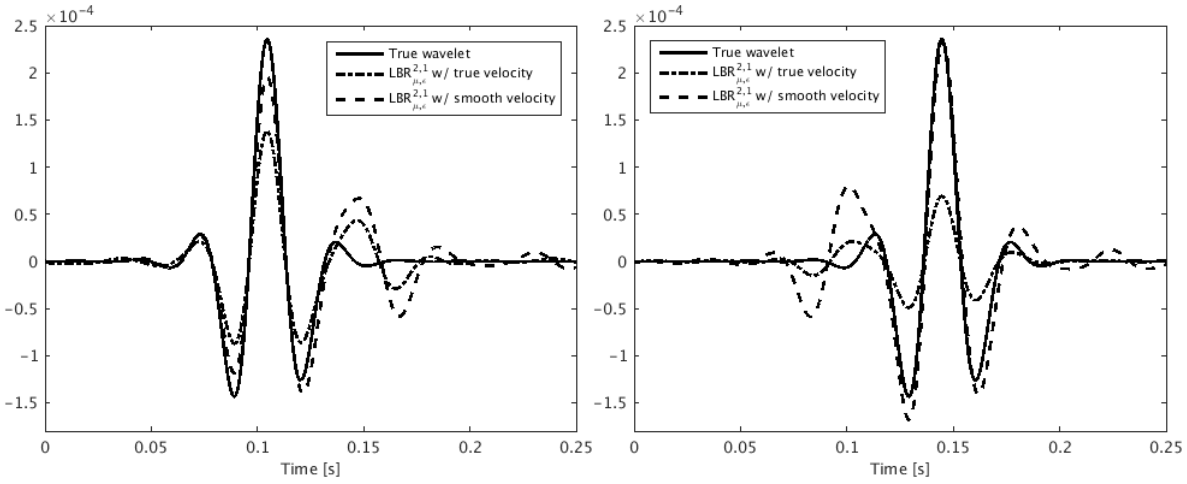
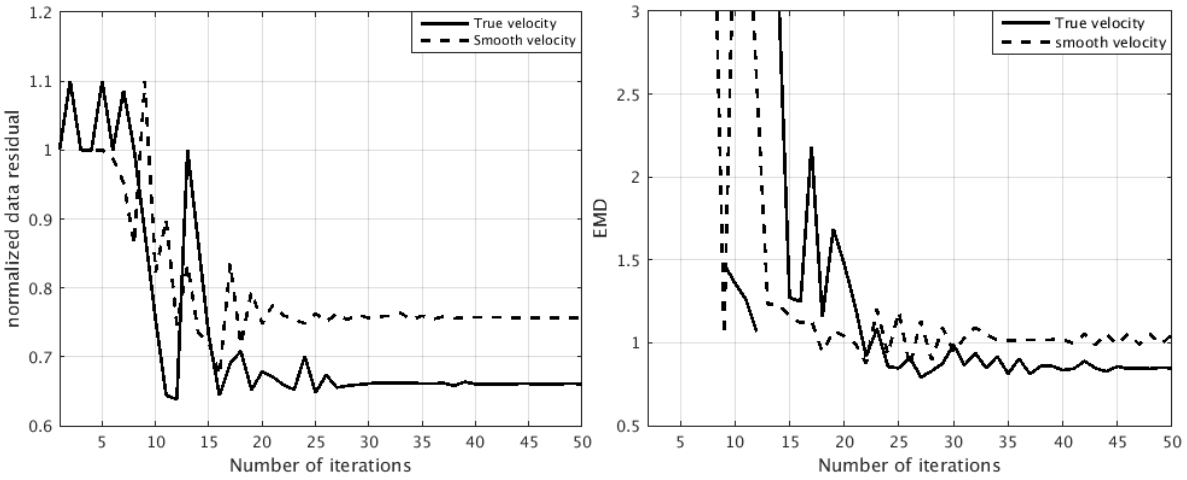


Figure 8. Comparison between the actual (solid black) and estimated source-time functions for the true velocity model (depicted by dash dot line) and the smoothed velocity model (dashed) at the two source locations plotted in *(a)* and *(b)* from left to right. The estimated source-time functions are corrected by a factor of 2 for comparison purposes.

(normalized) model error in the ℓ_2-norm. Because point sources modeled by Dirac distributions are not ℓ_2 integrable, the energy norm does not lend itself well to measure distances between delta-like functions. This norm would basically report little progress even though our algorithm produces intensity plots with delta-like sources at roughly the correct locations. To overcome this difficulty, we use the EMD metric instead. We argue that the EMD is a more suitable metric than the Euclidean distance. The non-integrability of the ℓ_2-norm suggests the use of the Total-Variation (TV) norm instead. In 1D, this norm is defined as the ℓ_1-norm of the derivative (Rudin et al. 1992). Unfortunately, the TV-norm is also unable to give us a meaningful metric because the TV distance between two shifted but identical delta-like pulses with normalized energy equals 2 as long as their distance is not zero, $l \neq 0$. For this reason, the TV-norm cannot inform us how far delta-like pulses are separated both in location and focus. On the other hand, when the pulses are energy normalized and identical, the EMD measures the distance l between these pulses exactly. When the recovered pulse is blurred, the EMD also comes up with a sensible number. For instance, if the original pulse is a delta centered at x_0 and the inverted pulse has a Gaussian bell shape, centered at x_0 with standard deviation σ, then the distance in TV norm is again 2 irrespective how large the width of the blurred pulse becomes. On the other hand, the EMD is proportional to σ with a constant depending only on the dimension of the model. These properties make the EMD highly suitable as a distance



(a) (b)

Figure 9. Convergence comparison: decay of *(a)* normalized data residuals and *(b)* EMD using true (solid) and smoothed velocity models (dashed). Notice the gaps in the EMD because of the thresholding nonlinearity of $\mathrm{LBR}^{2,1}_{\mu,\epsilon}$. As expected, the results for the true model lead to smaller residuals and a better EMD for the later iterations.

measure as long as the recovered intensities can be considered as probability distributions. Since our images concern intensity plots, this is always the case as long we normalize them.

In Figs 9a and 9b, the solid and dashed lines in plots for the normalized residuals and EMDs correspond to inversions with the true and smoothed velocity models (cf. Figs 6a and 6b). We can make the following observations from these plots. First, the plots for the true model vary more wildly during the early iterations of both methods. This can be understood because the inversion is working with a modeling operator that produces a more complex wavefield. Moreover, the thresholding nonlinearity can lead to removal of one of the sources at which point the EMD fails. This explains the gaps in the plots for the EMD, which are not unexpectedly more severe for the true velocity model. Second, our algorithm produces as we predicted the best results for both the normalized residual and the EMD when we use the true velocity model. Third, both the normalized residuals and EMD converge after about 25 iterations. The fact that the normalized residual saturates is consistent with the noiselevel.

4.2 Multiple source-cluster experiment in Marmousi model

To test our method in a more realistic scenario involving multiple events, we perform our inversion on two clusters of closely spaced microseismic events on a noisy record (with an S/R = 3.5 dB, Fig. 10d) with an aperture of almost 3.0 km. Each cluster contains contributions from multiple microseismic events. Moreover, the events differ by factor of 2 in amplitude and

have overlapping source-time functions with central frequencies of (30.0 , 25.0 , and 22.0) Hz. To test the validity of our approach, we work with a subset of the Marmousi model with a size of $3.15\,\mathrm{km} \times 1.08\,\mathrm{km}$, which corresponds to 631×217 grid points. Contrary to the BG compass model, the Marmousi model has a dominant fault structure and strongly contrasting layers. To make the joint source localization problem more challenging, we place seven microseismic sources in two clusters, five of which are located in the low-velocity zone with adjacent sources within half a dominant wavelength (see Fig. 10a). The other two sources are located near the fault zone and are also separated within half a dominant wavelength.

As we mentioned earlier, we never know the true velocity model and we assume to have access to a kinematically correct but smoothed version (Fig. 10b) instead. After generating synthetic data for the sources in the true model, our task is again to detect the location of the microseismic sources and estimate the corresponding source-time functions. Figs 11a and 11b show the estimated intensity plots obtained after back propagation using the smoothed velocity model depicted in Fig. 10b. As before, the white crosses indicate the actual location of the microseismic sources. From these results, we observe that back propagation has a strong smearing effect on the estimated intensity plots making it difficult to accurately locate the spatial source positions. In addition, there are also some spurious outliers in these results produced by back propagation, which can lead to undesired false detections of microseismic sources.

Figs 11c and 11d include inversion results according to problem LS_ϵ for 300 iterations. As before, the existence of non-radiating sources renders these results unsuitable for interpretation. The results are too blurred and there are too many spurious outliers in the intensity plots. While these inverted results clearly fail, results obtained with our method (by minimizing $\mathrm{LBR}^{2,1}_{\mu,\epsilon}$) with μ = 9e-4 clearly resolve the individual events (Figs 11e and 11f) in both clusters despite the fact that these events have overlapping source-time functions firing within half of the dominant time period. We arrive at these results in just ten iterations. By increasing the value of μ to 9e-3, we further improve the resolution (Figs 11g and 11h) at the cost of 20 more iterations. Compared to the two other methods, our sparsity promoting results allows us to accurately estimate the source locations albeit occasionally with a small (up to 11 m) spatial shift, which is due to the fact that we only have access to a smoothed velocity model, which does not allow us to fully capture all wave propagation effects.

In addition to finding accurate estimates for the microseismic source locations, our method also gives us access to the corresponding source-time functions — a feature difficult if not impossible to achieve with most other methods. Fig. 12 summarizes our results and leads

to the following observations: *(i)* Even though there is some to-be-expected source crosstalk, which leads to artifacts, the estimated source-time functions accurately recover the shape, origin time, and frequency content of the different source-time functions. We obtain these results by applying a spectral smoothing aimed at removing some of the source crosstalk. *(ii)* As expected, the quality and amplitudes of the estimated source-time functions improve with increasing μ—i.e., compare the results obtained with 10 iterations (denoted by the dash dot line) to the results obtained with 30 iterations (dashed lines) in Fig. 12. *(iii)* While the amplitudes are recovered to within a scale factor (we corrected all source-time functions with a factor of 6.4), the estimates do carry some relative amplitude information—i.e. the first two events are indeed larger in amplitude. However, the recovered amplitudes contain the imprint of propagation effects and the fact that sparsity-promotion often leads to biasing of the recovered amplitudes. Once again, we get this scaling factor of 6.4 by taking the minimum among the ratios of peak amplitudes of the true and the estimated source-time functions.

As a final check, we also compare in Fig. 13 data and data residuals modeled with the source wavefield estimates yielded by the two inversion methods, both of which seek to fit the observed data. Since none of the source estimation algorithms converged completely, we observe coherent energy in the data residuals. We cannot expect this coherent energy to completely disappear because we only used a smoothed velocity model during the inversion, which means that (multiple) reflections are not included in the modeling.

5 DISCUSSION

The above synthetic case studies reveal that our sparsity-promoting method solves for the complete source wavefield for spatially sparsely distributed microseismic sources. The estimated source wavefield preserves the locations and the source-time functions of multiple microseismic sources activated within close spatial proximity with partially overlapping source-time functions. We arrive at these results without prior assumptions on the shape of the source-time functions and the location of the sources.

Our method, which uses sparsity-promotion, outperforms methods based on back propagation or on energy minimization because it reconstructs the source wavefield near the actual source locations only by virtue of the thresholding nonlinearity. As a result, our method leads to better interpretable results thanks to the improved resolution, which is essential in situations where the microseismic sources are within close proximity—i.e., within half a dominant wavelength and time period of each other. In this situation, both back propagation and minimum-energy solution fail to resolve nearby sources whereas our method resolves closely

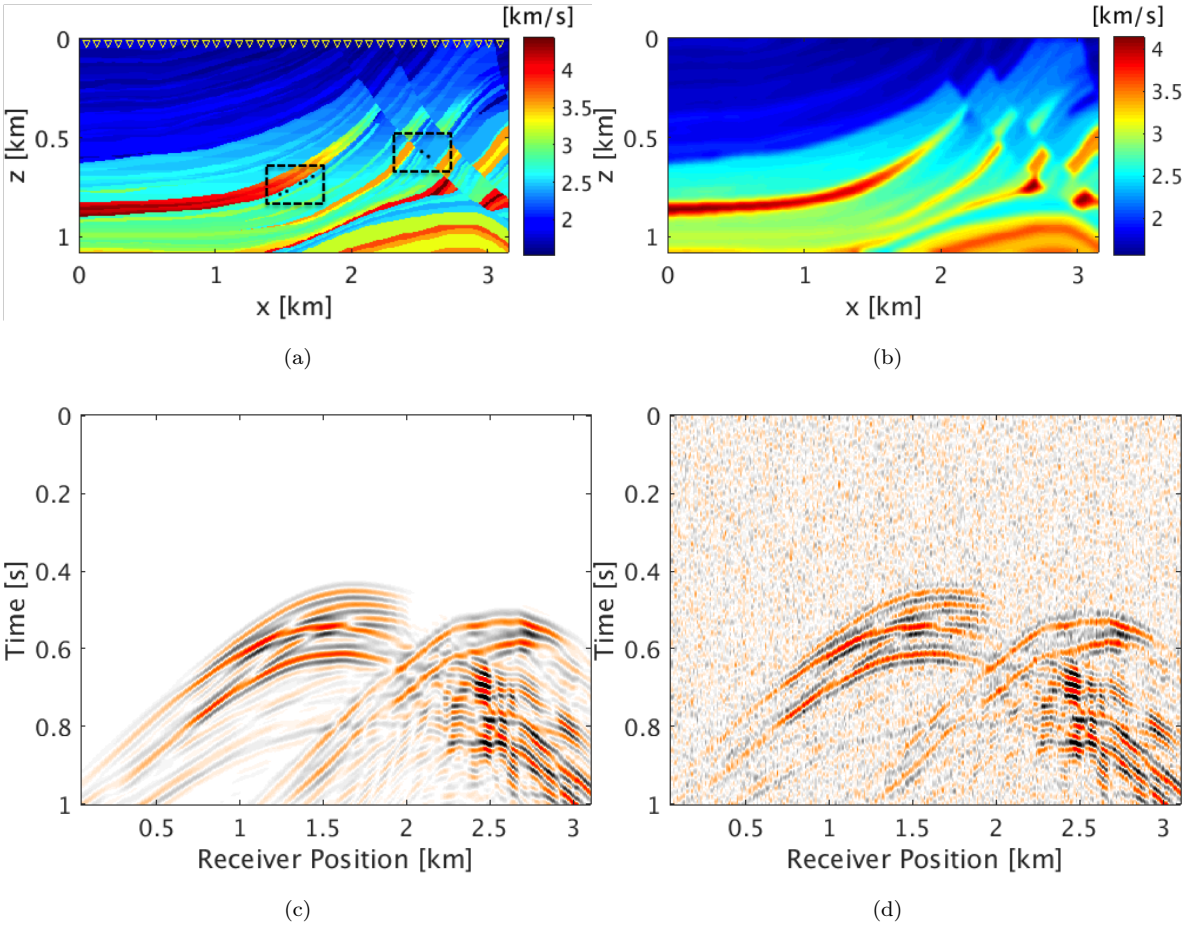


Figure 10. Marmousi model experiment. *(a)* True velocity model with acquisition geometry and the microseismic source locations superimposed. The inverted yellow triangles indicate receivers buried at a depth of 20.0 m and separated by 10.0 m with an aperture of almost 3.0 km. The black dots indicate the location of seven microseismic sources. *(b)* Kinematically correct smoothed velocity model used during the inversion. *(c)* Noise free microseismic data generated by seven microseismic sources in the true velocity model. *(d)* Noisy microseismic data (S/R = 3.5 dB) generated by adding bandwidth limited (5.0 Hz to 45.0 Hz) Gaussian random noise to the noise-free simulated data.

spaced microseismic sources relatively well. Because of this ability, we are also capable to estimate the corresponding source-time functions relatively well albeit some to-be-expected artifacts remain because of source interference. These important features make our method potentially suitable for tracking the spatial temporal evolution of microseismic fractures.

Our observations hold for idealized and relatively simple situations and carry over to source-estimation problems in complex velocity models involving up to seven sources with different source signatures. Our method successfully locates microseismic sources in all cases while preserving relative amplitude and frequency information of the underlying source-time

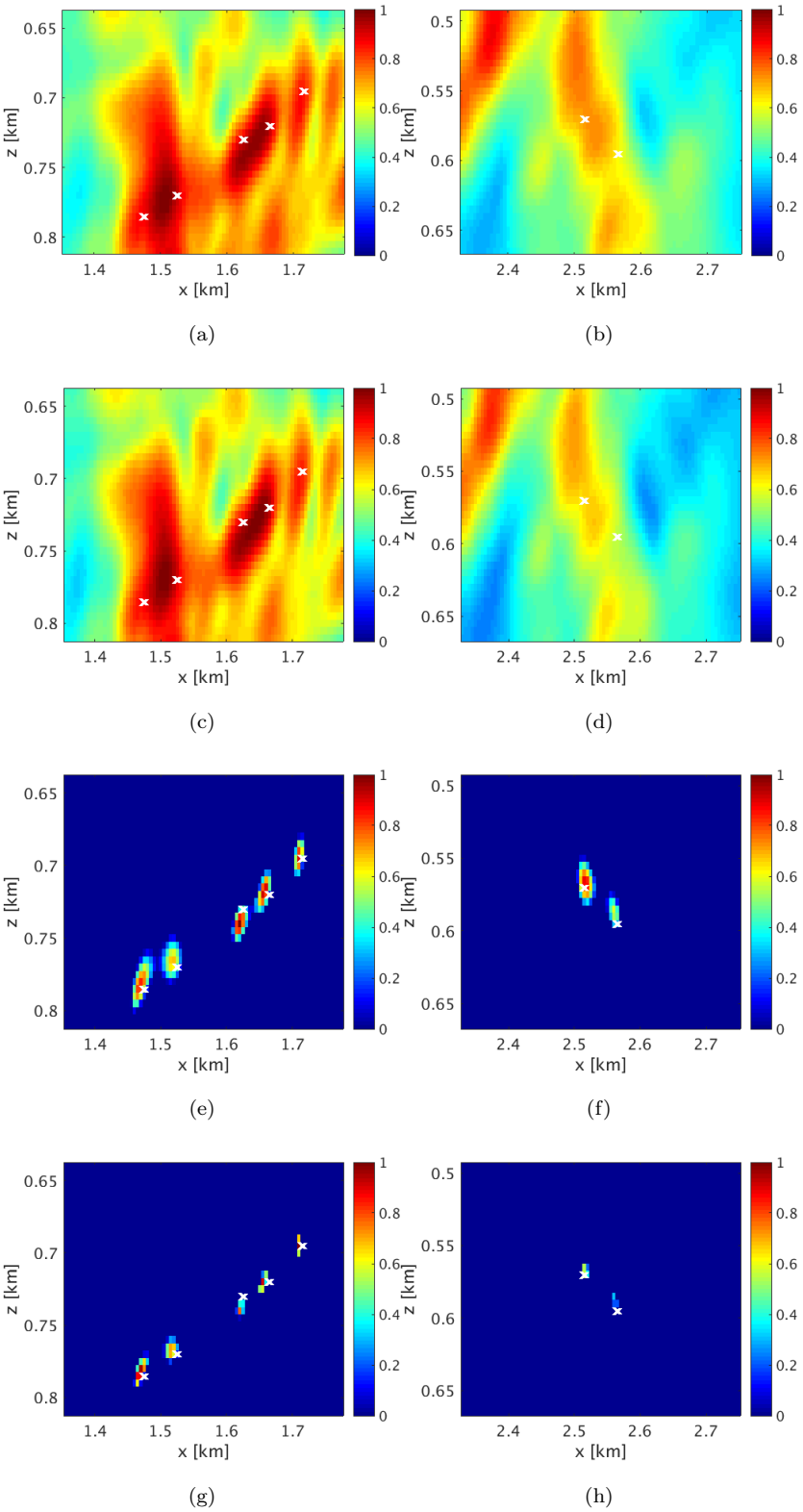


Figure 11. Zoomed intensity plots near cluster 1 (left column) and cluster 2 (right column) from left to right for the smoothed velocity models obtained via back propagation in *(a, b)*; solving LS_ϵ with 300 iterations in *(c, d)*; and by solving $\mathrm{LBR}^{2,1}_{\mu,\epsilon}$ for μ = 9e-4 with only 10 iterations in *(e, f)*; and by solving $\mathrm{LBR}^{2,1}_{\mu,\epsilon}$ for μ = 9e-3 with 30 iterations in *(g, h)*. The white colour crosses indicate the actual spatial locations of the microseismic sources.

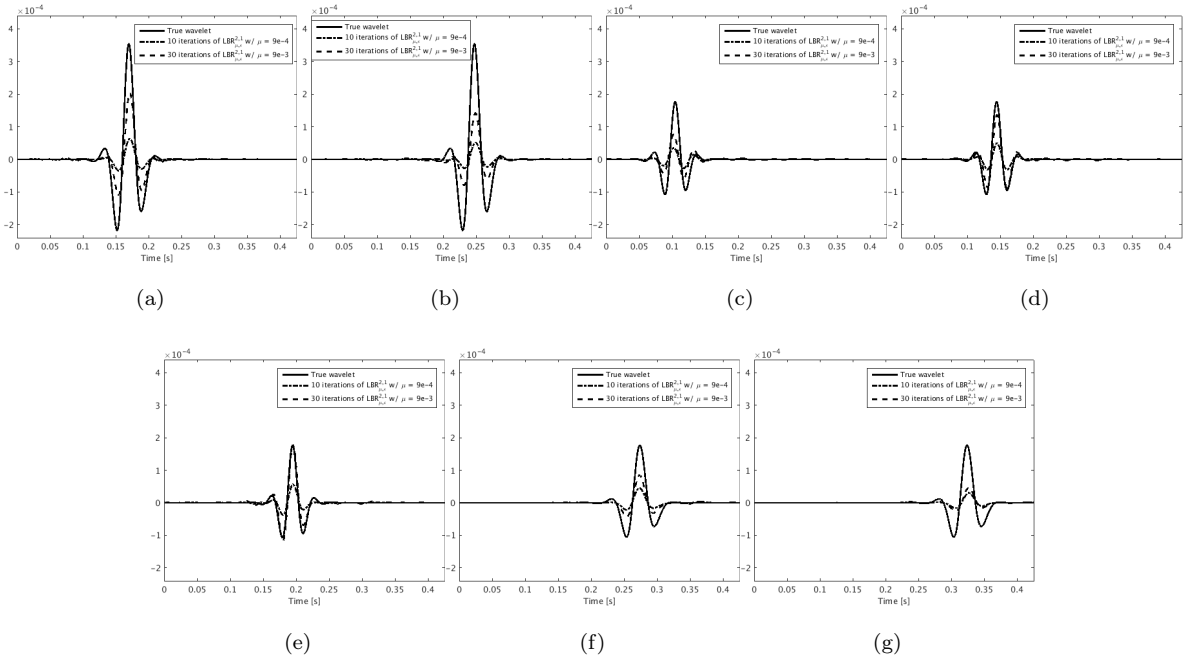


Figure 12. Comparison between the actual source-time functions (solid black) and the estimated source-time functions after 10 (denoted by dash dot line) and 30 (dashed) iterations of our method at the seven locations in *(a)*, *(b)*, *(c)*, *(d)*, *(e)*, *(f)* and *(g)* from left to right. The peak amplitude of true source-time functions in *(a)* and *(b)* is approximately twice the peak amplitude of true source-time functions at other locations. The dominant frequency of the source-time functions in *(a)* and *(b)* is 25.0 Hz, in *(c)*, *(d)* and *(e)* 30.0 Hz and in *(f)* and *(g)* it is 22.0 Hz. Estimated source-time functions are corrected by a factor of 6.4 for comparison purposes.

functions. Our method works with smoothed velocity models without prior knowledge on the number of sources, their temporal distribution, and prior information on their spatial distribution aside from being sparsely distributed. The only assumption we made inverting noisy data generated from the true velocity models is that we have access to a smoothed version of the true velocity model. By promoting sparse solutions in space with finite-energy along time, we greatly improve the resolution and overcome issues with non-radiating sources. This apparent null-space associated with these non-radiating sources renders inversions that minimize the energy in the objective only useless. By including sparsity promotion along the spatial coordinates, we are able to resolve multiple close-by sources near faults and velocity lows with a resolution that encroaches on Abbe's diffraction limit.

While the recovered intensity plots and source-time functions carry information on the magnitude and temporal frequency content of the events, the recovered amplitudes suffer. Aside from a global scaling ambiguity, we find differences in amplitude related to the depth of

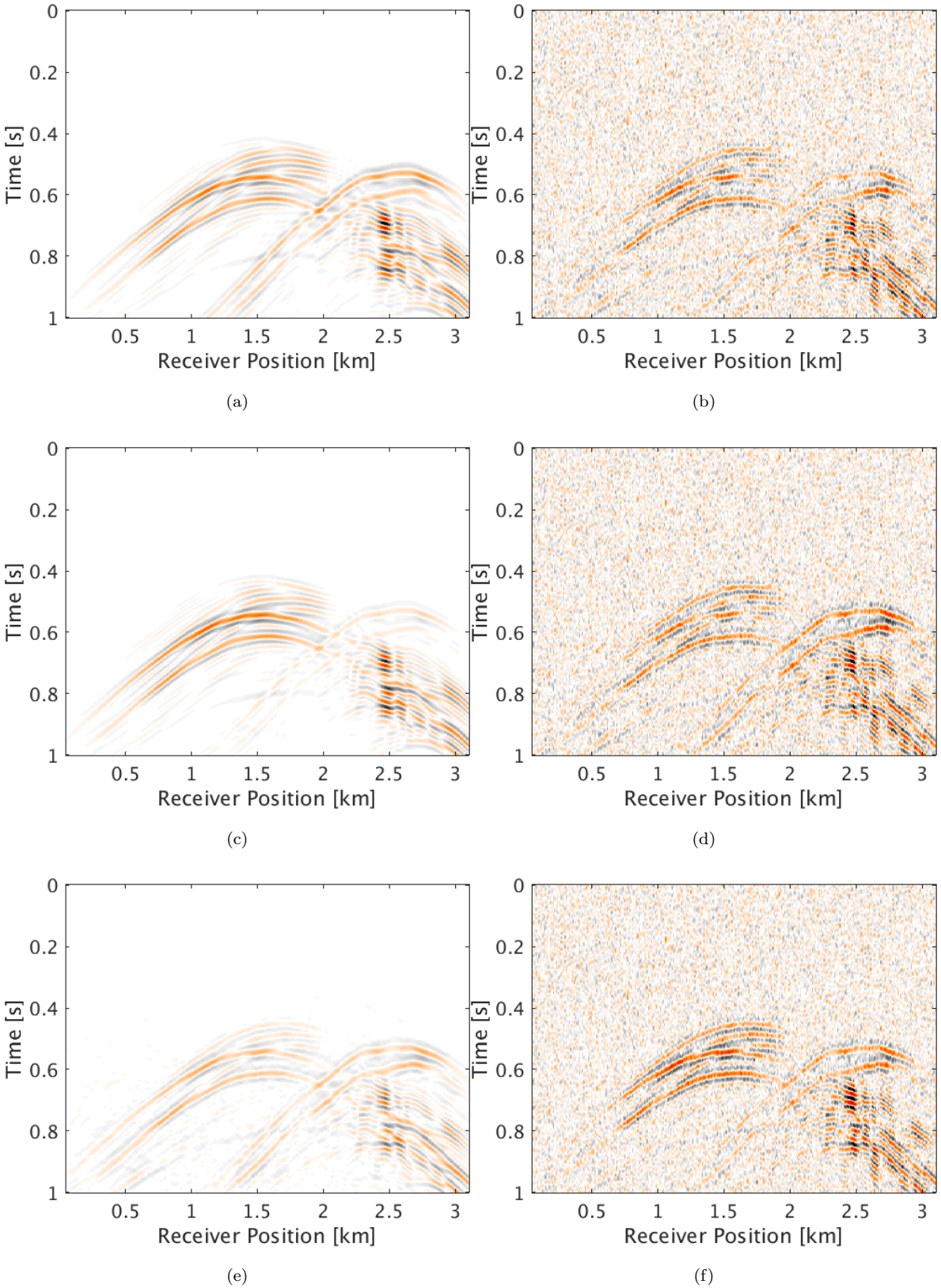


Figure 13. Modeled data (left column) and data residual (right column) in *(a,b)* by solving $\text{LBR}^{2,1}_{\mu,\epsilon}$ for $\mu =$ 9e-4 with only 10 iterations; in *(c, d)* by solving $\text{LBR}^{2,1}_{\mu,\epsilon}$ for $\mu =$ 9e-3 with 30 iterations; in *(e, f)* by solving LS_ϵ with 300 iterations. All plots are on the same color scale.

the events. We expect that including an amplitude term in our preconditioner in combination with an additional debiasing (Sharan et al. 2018) step will remedy some of these amplitude issues. Debiasing involves running a few iterations of LSQR (Paige & Saunders 1982) on the support—i.e. on the entries of the recovered wavefield that survive the threshold. For a fixed trade off parameter μ, the detectability of smaller events depends on the noise level, the acquisition geometry, and the velocity model. Moreover, for larger μ more iterations are needed to bring in smaller entries. While the detectability of compressible signals is well understood for very specific situations in Compressive Sensing (Candès et al. 2005; Donoho 2006), deriving similar rigorous theoretical results for seismic source localization with sparsity promotion is still largely an open problem.

The fact that we are able to recover the individual source-time functions with manageable crosstalk is also encouraging because it could provide us with additional information on the source mechanisms underlying the fracturing process itself.

To arrive at these results, we made two major assumptions namely access to a (smooth) velocity model and a simplified source mechanism. Looking at the work of Sun et al. (2016) and Wu & Alkhalifah (2017), we envision that our framework can relatively easily be extended to include velocity model updates. As for the simplified source mechanism, we argue that our method will likely be relatively robust with respect to different source mechanisms. We base this assertion on the fact that our parameterization by the acoustic wave equation accounts for the leading order wave physics and that more detailed information on the source mechanism can be obtained during a later inversion step. Our approach is useful because moment tensor inversions typically need estimates of the approximate source locations and origin times as input and the proposed method will be able to provide these as we have demonstrated.

6 CONCLUSIONS

By exploiting the fact that microseismic sources are localized along fractures that are sparsely distributed and emit finite energy along time, we arrive at a sparsity promoting formulation capable of estimating microseismic event locations and their associated source-time functions in the presence of noise. When we are given a smoothed velocity model, our method recovers close-by microseismic events with partially overlapping source-time functions without relying on prior information other than having access to a smoothed velocity model. By switching to a dual formulation, we are able to significantly speed up convergence of the inversion, rendering our methodology computationally feasible and therefore more practical. We demonstrate that the improved convergence is essential when we want to recover microseismic events that are

encroaching on or slightly within Abbe's diffraction limit of half a wavelength. For two complex velocity models, we were able to demonstrate the viability of our approach for situations where most other source localization methods would fail by lack of prior information on either the source locations or origin times. Our method needs neither to recover the full source wavefield from noisy data collected at the surface.

ACKNOWLEDGMENTS

We would like to acknowledge Dr. Aleksandr Aravkin (University of Washington) for the useful discussion about the linearized Bregman algorithm. This research was carried out as part of the SINBAD project with the support of the member organizations of the SINBAD Consortium.

REFERENCES

Abbe, E., 1873. Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung, *Arch. Mikrosk. Anat.*, **9**(1), 413–418.

Alkhalifah, T., 2000. An acoustic wave equation for anisotropic media, *Geophysics*, **65**(4), 1239–1250.

Bazargani, F. & Snieder, R., 2016. Optimal source imaging in elastic media, *Geophys. J. Int.*, **204**(2), 1134–1147.

Boyd, S. & Vandenberghe, L., 2009. *Convex Optimization*, Cambridge University Press.

Brougois, A., Bourget, M., Lailly, P., Poulet, M., Ricarte, P., & Versteeg, R., 1990. Marmousi, model and data, in *EAEG Workshop-Practical Aspects of Seismic Data Inversion*.

Candès, E. J., Romberg, J., & Tao, T., 2005. Stable signal recovery from incomplete and inaccurate measurements, *Commun. Pur. Appl. Math.*, **59**(8), 1207–1223.

Chen, S. S., Donoho, D. L., & Saunders, M. A., 1998. Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.*, **20**(1), 33–61.

Claerbout, J. F. & Muir, F., 1973. Robust modeling with erratic data, *Geophysics*, **38**(5), 826–844.

Combettes, P. L. & Pesquet, J.-C., 2011. *Proximal Splitting Methods in Signal Processing*, pp. 185–212, Springer, New York.

Devaney, A. J., 2012. *Mathematical Foundations of Imaging, Tomography and Wavefield Inversion*, Cambridge University Press.

Donoho, D. L., 2006. Compressed sensing, *IEEE T. Inform. Theory*, **52**(4), 1289–1306.

Donoho, D. L. & Logan, B. F., 1992. Signal Recovery and the Large Sieve, *SIAM J. Appl. Math.*, **52**(2), 577–591.

Fink, M., 1997. Time reversed acoustics, *Phys. Today*, **50**(3), 34–40.

Fornasier, M. & Rauhut, H., 2008. Recovery algorithms for vector-valued data with joint sparsity constraints, *SIAM J. Numer. Anal.*, **46**(2), 577–613.

Gajewski, D. & Tessmer, E., 2005. Reverse modelling for seismic event characterization, *Geophys. J. Int.*, **163**(1), 276–284.

Gao, W. & Sacchi, M., 2017. Microseismic event localization via least-squares full waveform inversion with group sparsity constraints, in *Proceedings of the 79th EAGE Conference and Exhibition*, Paris, 11-15 June 2017, Expanded Abstract.

Herrmann, F. J., Brown, C. R., Erlangga, Y. A., & Moghaddam, P. P., 2009. Curvelet-based migration preconditioning and scaling, *Geophysics*, **74**, A41–A46.

Herrmann, F. J., Tu, N., & Esser, E., 2015. Fast "online" migration with compressive sensing, in *Proceedings of the 77th EAGE Conference and Exhibition*, Madrid, 1-4 June 2015, Expanded Abstract.

Huang, B., Ma, S., & Goldfarb, D., 2013. Accelerated Linearized Bregman Method, *J. Sci. Comput.*, **54**(2), 428–453.

Kaderli, J., McChesney, M. D., & Minkoff, S. E., 2015. Microseismic event estimation in noisy data via full waveform inversion, in *85th Annual SEG meeting*, SEG Technical Program Expanded Abstracts, pp. 1159–1164.

Kamei, R., Nakata, N., & Lumley, D., 2015. Introduction to microseismic source mechanisms, *The Leading Edge*, **34**(8), 876–880.

Kim, Y., Liu, Q., & Tromp, J., 2011. Adjoint centroid-moment tensor inversions, *Geophys. J. Int.*, **186**(1), 264–278.

Kitić, S., Albera, L., Bertin, N., & Gribonval, R., 2016. Physics-driven inverse problems made tractable with cosparse regularization, *IEEE T. Signal Proces.*, **64**(2), 335–348.

Kowalski, M. & Torrésani, B., 2009. Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients, *Signal Image Video P.*, **3**(3), 251–264.

Li, X., Aravkin, A. Y., van Leeuwen, T., & Herrmann, F. J., 2012. Fast randomized full-waveform inversion with compressive sensing, *Geophysics*, **77**(3), A13–A17.

Li, X., Esser, E., & Herrmann, F. J., 2016. Modified gauss-newton full-waveform inversion explained — why sparsity-promoting updates do matter, *Geophysics*, **81**(3), R125–R138.

Li, Z. & van der Baan, M., 2016. Microseismic event localization by acoustic time reversal extrapolation, *Geophysics*, **81**(3), KS123–KS134.

Liu, D. C. & Nocedal, J., 1989. On the limited memory BFGS method for large scale optimization, *Math. Program.*, **45**(1), 503–528.

Lorenz, D. A., Schöpfer, F., & Wenger, S., 2014. The Linearized Bregman Method via Split Feasibility Problems: Analysis and Generalizations, *SIAM J. Imaging Sci.*, **7**(2), 1237–1262.

Lorenz, D. A., Wenger, S., Schöpfer, F., & Magnor, M., 2014. A sparse Kaczmarz solver and a linearized Bregman method for online compressed sensing, *ArXiv e-prints*.

Louboutin, M., Witte, P. A., Lange, M., Kukreja, N., Luporini, F., Gorman, G., & Herrmann, F. J., 2017. Full-waveform inversion - part 1: forward modeling, *The Leading Edge*, **36**(12), 1033–1036.

Madariaga, R., 1989. *Seismic source: Theory*, Springer US.

McMechan, G. A., 1982. Determination of source parameters by wavefield extrapolation, *Geophys. J. Roy. Astr. S.*, **71**(3), 613–628.

Musafir, R. E., 2013. On non-radiating sources, *J. Sound and Vib.*, **332**(17), 3947 – 3955, Philip Doak Commemorative Issue.

Nakata, N. & Beroza, G. C., 2016. Reverse time migration for microseismic sources using the geometric mean as an imaging condition, *Geophysics*, **81**(2), K551–K560.

Osher, S., Mao, Y., Dong, B., & Yin, W., 2010. Fast linearized Bregman iterations for compressive sensing and sparse denoising, *Commun. Math. Sci.*, **8**(1), 93–111.

Paige, C. C. & Saunders, M. A., 1982. LSQR: Sparse linear equations and least squares problems, *ACM Trans. Math. Softw.*, **8**(2), 195–209.

Porter, R. P. & Devaney, A. J., 1982. Holography and the inverse source problem, *J. Opt. Soc. Am.*, **72**(3), 327–330.

Rubner, Y., Tomasi, C., & Guibas, L. J., 1998. A metric for distributions with applications to image databases, in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pp. 59–66.

Rudin, L. I., Osher, S., & Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms, *Physica D.*, **60**(1), 259 – 268.

Santosa, F. & Symes, W., 1986. Linear Inversion of Band-Limited Reflection Seismograms, *SIAM J. Sci. Stat. Comp.*, **7**(4), 1307–1330.

Schmidt, R., 1986. Multiple emitter location and signal parameter estimation, *IEEE T. Antenn. Propag.*, **AP-34**(3), 276–280.

Shalev-shwartz, S. & Singer, Y., 2006. Convex repeated games and fenchel duality, in *Advances in Neural Information Processing Systems 19*, MIT Press.

Sharan, S., Wang, R., van Leeuwen, T., & Herrmann, F. J., 2016. Sparsity-promoting joint microseismic source collocation and source-time function estimation, in *86th Annual SEG meeting*, SEG Technical Program Expanded Abstracts, pp. 2574-2579.

Sharan, S., Kumar, R., Wang, R., & Herrmann, F. J., 2018. A debiasing approach to microseismic, in *88th Annual SEG meeting*, SEG Technical Program Expanded Abstracts, pp. 2942-2946.

Song, C., Alkhalifah, T., Wu, Z., & Sun, B., 2017. Nonstationary filter used in microseismic-source imaging, in *87th Annual SEG meeting*, SEG Technical Program Expanded Abstracts, pp. 2898-2902.

Song, Z.-M. & Williamson, P. R., 1995. Frequency-domain acoustic-wave modeling and inversion of crosshole data: Part 1-2.5d modeling method, *Geophysics*, **60**(3), 784–795.

Sun, J., Zhu, T., Fomel, S., & Song, W., 2015. Investigating the possibility of locating microseismic sources using distributed sensor networks, in *85th Annual SEG meeting*, SEG Technical Program Expanded Abstracts, pp. 2485-2490.

Sun, J., Xue, Z., Fomel, S., Zhu, T., & Nakata, N., 2016. Full waveform inversion of passive seismic data for sources and velocities, in *86th Annual SEG meeting*, SEG Technical Program Expanded

Abstracts, pp. 1405-1410.

Taylor, H. L., Banks, S., & McCoy, J., 1979. Deconvolution with the ℓ_1 norm, *Geophysics*, **44**, 39–52.

van den Berg, E. & Friedlander, M. P., 2008. Probing the Pareto frontier for basis pursuit solutions, *SIAM J. Sci. Comput.*, **31**(2), 890–912.

van Leeuwen, T. & Herrmann, F. J., 2012. Fast waveform inversion without source-encoding, *Geophys. Prospect.*, **61**(s1), 10–19.

Witte, P. A., Louboutin, M., Luporini, F., Gorman, G. J., & Herrmann, F. J., 2018. Compressive least-squares migration with on-the-fly fourier transforms, Submitted to Geophysics on June 28, 2018.

Wu, Y. & McMechan, G. A., 1996. Elastic full-waveform inversion for earthquake source parameters, *Geophys. J. Int.*, **127**(1), 61–74.

Wu, Z. & Alkhalifah, T., 2017. A New Wave Equation Based Source Location Method with Full-waveform Inversion, in *Proceedings of the 79th EAGE Conference and Exhibition*, Paris, 12-15 June 2017, Expanded Abstract.

Yin, W., 2010. Analysis and Generalizations of the Linearized Bregman Method, *SIAM J. Imaging. Sci.*, **3**(4), 856–877.

Yin, W., Osher, S., Goldfarb, D., & Darbon, J., 2008. Bregman Iterative Algorithms for ℓ_1-minimization with Applications to Compressed Sensing, *SIAM J. Imaging. Sci.*, **1**(1), 143–168.

7 APPENDIX A

We provide here complete derivation of eq. (7). We rewrite the primal source estimation problem

$$\min_{\mathbf{Q}} \quad \|\mathbf{Q}\|_{2,1} + \frac{1}{2\mu}\|\mathbf{Q}\|_F^2 \quad \text{s.t.} \quad \|\mathcal{F}[\mathbf{m}](\mathbf{Q}) - \mathbf{d}\|_2 \leq \epsilon, \tag{A-1}$$

into the equivalent unconstrained form

$$\min_{\mathbf{Q}} \quad \|\mathbf{Q}\|_{2,1} + \frac{1}{2\mu}\|\mathbf{Q}\|_F^2 + \iota_{\|\mathcal{F}[\mathbf{m}](\mathbf{Q})-\mathbf{d}\|_2 \leq \epsilon}, \tag{A-2}$$

where ι is the support function defined as

$$\iota_C(x) = \begin{cases} 0 & \text{for } x \in C \\ \infty & \text{for } x \notin C \end{cases}.$$

Plugging the following identity

$$\iota_{\|\mathcal{F}[\mathbf{m}](\mathbf{Q})-\mathbf{d}\|_2 \leq \epsilon} = \max_{\mathbf{y}} \langle \mathbf{y}, \mathcal{F}[\mathbf{m}](\mathbf{Q}) - \mathbf{d} \rangle - \epsilon \|\mathbf{y}\|_2$$

into (A-2), we obtain the Fenchel's dual of (A-1)

$$\max_{\mathbf{y}} \min_{\mathbf{Q}} \quad \|\mathbf{Q}\|_{2,1} + \frac{1}{2\mu}\|\mathbf{Q}\|_F^2 + \langle \mathbf{y}, \mathcal{F}[\mathbf{m}](\mathbf{Q}) - \mathbf{d} \rangle - \epsilon \|\mathbf{y}\|_2 \equiv \Psi(\mathbf{y}) - \epsilon \|\mathbf{y}\|_2$$

where

$$\Psi(\mathbf{y}) = \min_{\mathbf{Q}} \quad \|\mathbf{Q}\|_{2,1} + \frac{1}{2\mu}\|\mathbf{Q}\|_F^2 + \langle \mathbf{y}, \mathcal{F}[\mathbf{m}](\mathbf{Q}) - \mathbf{d}\rangle$$

Applying similar arguments as in Huang et al. (2013) to our $\ell_{2,1}$ norm, we see that both $\Psi(\mathbf{y})$ and $\nabla\Psi(\mathbf{y})$ have close-form representations

$$\Psi(\mathbf{y}) = \Phi_\mu(\mu\mathcal{F}[\mathbf{m}]^T(\mathbf{y})) - \frac{\mu}{2}\|\mathcal{F}[\mathbf{m}]^T(\mathbf{y})\|_2^2 + \mathbf{d}^T\mathbf{y} \tag{A-3}$$

and

$$\nabla\Psi(\mathbf{y}) = \mathbf{d} - \mathcal{F}[\mathbf{m}](\mathrm{Prox}_{\mu\ell_{2,1}}(\mu\mathcal{F}[\mathbf{m}]^T(\mathbf{y}))), \tag{A-4}$$

where

$$\Phi_\mu(\mathbf{X}) = \min_{\mathbf{W}} \quad \{\|\mathbf{W}\|_{2,1} + \frac{1}{2\mu}\|\mathbf{W} - \mathbf{X}\|_F^2\}.$$