LARGE SCALE HIGH-FREQUENCY SEISMIC WAVEFIELD RECONSTRUCTION, ACQUISITION VIA RANK MINIMIZATION AND SPARSITY-PROMOTING SOURCE ESTIMATION

A Dissertation
Presented to
The Academic Faculty

By

Shashin Sharan

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Earth and Atmoshpheric Sciences

Georgia Institute of Technology

December 2020

LARGE SCALE HIGH-FREQUENCY SEISMIC WAVEFIELD RECONSTRUCTION, ACQUISITION VIA RANK MINIMIZATION AND SPARSITY-PROMOTING SOURCE ESTIMATION

Ph.D. committee:

Dr. Felix J. Herrmann, Advisor
School of Earth and Atmoshpheric Sciences
Georgia Institute of Technology

Dr. Zhigang Peng
School of Earth and Atmospheric Sciences
Georgia Institute of Technology

Dr. Andrew Newman
School of Earth and Atmospheric Sciences
Georgia Institute of Technology

Dr. James H McClellan
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Alison Malcolm
Earth Sciences Department
Memorial University of Newfoundland

Date Approved: November 23, 2020

Dedicated to my loving parents and wife

ACKNOWLEDGEMENTS

First and foremost I would like to thank my PhD advisor Professor Felix J. Herrmann for his support and guidance. I am very grateful to Professor Felix J. Herrmann for giving me this opportunity to pursue my doctoral research at Seismic Laboratory for Imaging and Modeling (SLIM). I am also thankful for his believe on me and for his motivation.

My gratitude also goes to Professor Zhigang Peng, Professor Andrew Newman from Georgia Institute of Technology, Professor Michael Bostock, Professor Phil Austin from University of British Columbia for serving as my supervisory committee. I am also thankful to Professor James H. McClellan and Professor Alison Malcolm for their valuable inputs and attendance in my defense as the external examiners.

I would also like to thank my colleagues and researchers at SLIM for there support and collaboration. I am especially thankful to Dr Henryk Modzelewski, Miranda Joyce and Diana Ruiz for their professional and administrative support. I would also like to extend my gratitude to Dr Rongrong Wang and Dr Rajiv Kumar for their mentorship and discussing valuable research ideas. I would also like to thank my fellow coauthors Prof Stanislav Emelianov, Dr Tristan van Leeuwen, Dr Haneet Wason, Dr Oscar Lopez, Dr Mathias Loboutin, Yijun Zhang, Dr Diego S. Dumani for the wonderful collaboration.

My special thank goes to Nick Moldoveanu and Dr Michel Schonewille for giving me opportunities to do internships with Schlumberger. I would like to also thank Franck Li Diagon and Dr Yousif Kamil from Schlumberger for their mentorship during my internship with Schlumberger.

I would also like to thank developers of Julia and Matlab languages as well as developers of Devito, JUDI, JOLI software packages which were used in my research. I also would like to acknowledge the collaboration of the SENAI CIMATEC Supercomputing Center for Industrial Innovation, Bahia, Brazil, and the support of BG Group (now Shell) and the International Inversion Initiative Project. I would also like to acknowledge AWS

for the cloud computation support.

Finally, I would like to acknowledge support of my wonderful family and friends. My heartfelt gratitude goes to my parents and my father-in-law. My special thanks to my mother-in-law who left us all for heavenly abode in the middle of my PhD. I would like to express my special thanks to my wonderful wife Divya Rashmi for her selfless love, patience and support. Without her support I cannot imagine to come this far. I would also like to thank my dear friend Vikas for his constant support and motivation. I would also like to express my heartfelt gratitude to my Guru SriSri Ravi Shankar.

This research was partially funded by the University of British Columbia, Natural Sciences and Engineering Research Council of Canada Collaborative Research and Development Grant (DNOISE II), the SINBAD consortium, Georgia Research Alliance and Georgia Institute of Technology.

TABLE OF CONTENTS

LIST OF FIGURES

SUMMARY

Seismic data reconstruction on a dense periodic grid from seismic data acquired on a coarse grid is a common approach followed by most of the oil & gas companies. This approach allows them to save on operationally challenging and expensive dense seismic data acquisition. Dense seismic data is one of the key requirements for generating high-resolution images of earth's subsurface for exploration and production decisions. Based on the Compressive Sensing (CS) paradigm, low-rank matrix factorization based seismic data reconstruction methods are computationally cheaper and scalable to large datasets in comparison to sparsity-promotion based methods. The sparsity-promotion based methods are based on transformation in certain transform domains that can be computationally expensive for large datasets. Although, low-rank matrix factorization based methods perform well at lower frequencies, their performance degrades at higher frequencies due to increase in rank of approximating matrix. One of the contributions of this thesis is a recursively weighted matrix factorization approach to improve the quality of reconstructed data at higher frequencies. This recursively weighted approach exploits the similarity between adjacent frequency slices. Although, recursively weighted method improves the data reconstruction quality at higher frequencies, it can be computationally expensive for large scale seismic datasets. This is because of the interdependence of frequencies preventing simultaneous reconstruction of frequencies. Another contribution of this thesis is a computationally efficient recursively weighted framework for large scale dataset by parallelizing data reconstruction over rows of low-rank factors of each frequency slices. To reduce the cost and turnaround time of seismic data acquisition simultaneous source acquisition is adapted by the oil and gas industry in last few years. Another contribution of this thesis is a low-rank based method for simultaneous separation and reconstruction of seismic data on a dense periodic grid from large scale seismic data acquired with simultaneous source acquisition. Next part of this thesis focuses on accurate detection of fractures

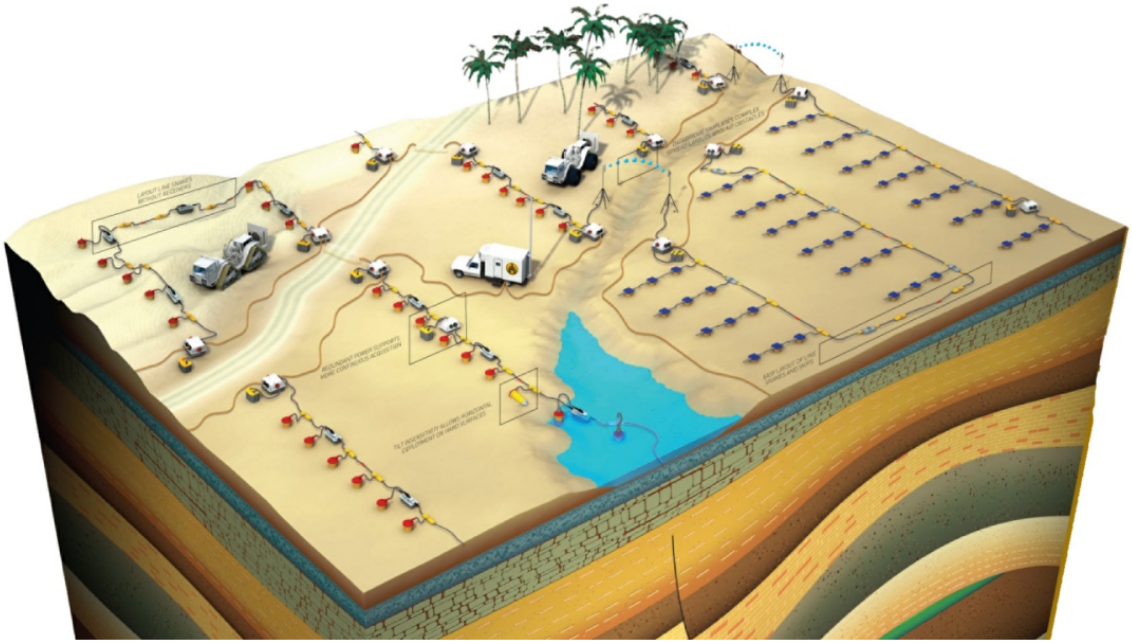
created by hydraulic fracturing in unconventional reservoirs for economical production of oil and gas. Fracturing of rocks during hydraulic fracturing gives rise to microseismic events, which are localized along these fractures. In this work, a sparsity-promoting microseismic source estimation framework is proposed to detect closely spaced microseismic sources along with estimation of their associated source-time functions from noisy microseismic data recorded by receivers along the earth's surface or along monitor wells. Detecting closely spaced microseismic events helps in delineating fractures and estimation of source-time function is useful in estimating fracture's origin in time. Also, source-time functions can be potentially useful for estimating the source-mechanism. Also, this method does not make any prior assumption on number of microseismic sources or shape of their source-time functions. Therefore, this method is useful for detecting microseismic sources with different source signatures and frequency content. Last part of this thesis focuses on sparsity-promoting photoacoustic imaging to detect photoabsorbers along with estimating the associated source-time functions. Traditional photoacoustic imaging can only estimate the locations of photoacoustic absorbers. Also, traditional methods require dense transducer coverage whereas sparsity-promotion based method can work with reduced transducer sampling reducing the overall data storage cost.

CHAPTER 1
INTRODUCTION

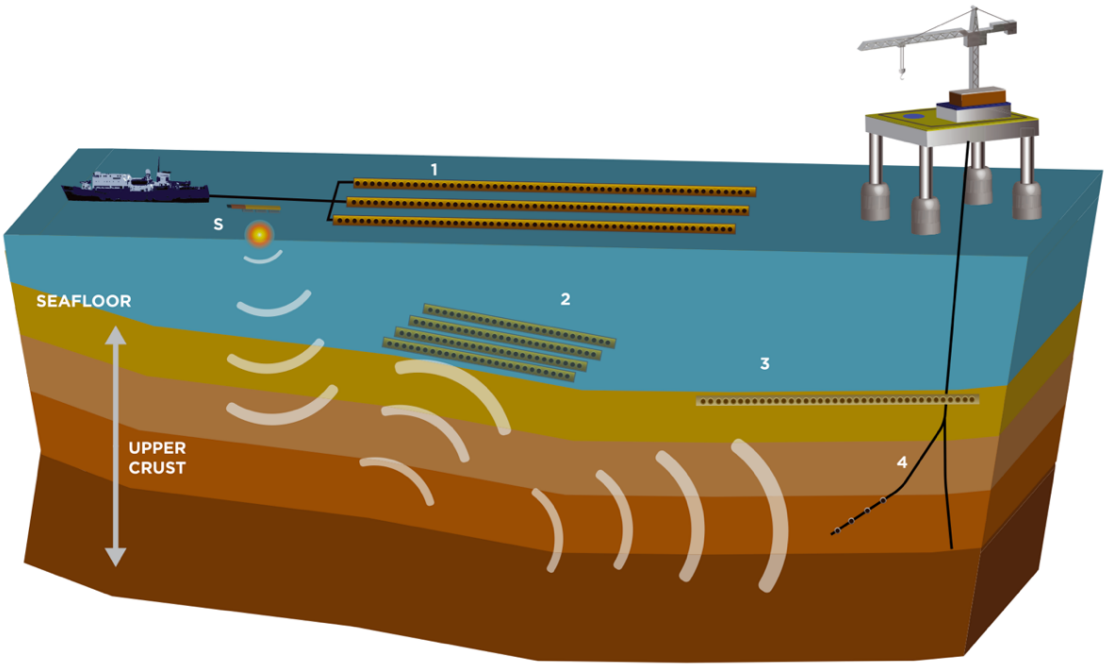
Since the last century, oil & gas has been one of the largest supplier of world's energy demand. Extraction of oil & gas from the earth's subsurface requires drilling wells. Depending on the subsurface geology and depth of the oil and gas reservoirs, drilling of these wells can be operationally complex and very expensive. For successful extraction of oil & gas and to prevent any hazardous situation, we rely on good quality images of earth's subsurface and estimation of subsurface physical properties (e.g. velocity, density etc.). These images and physical properties help in delineating prospective oil & gas reservoirs. We can obtain these images and estimate physical parameters by processing data acquired from the field.

There are different kinds of data available (e.g. gravitational, magnetic, electrical, electromagnetic, seismic etc.). Among all of these, seismic data gives the best resolution (Sheriff and Geldart 1995). Therefore, seismic data is widely used by the oil & gas industry. In seismic data acquisition, seismic sources (e.g. airguns, vibroseis etc.) send acoustic waves from earth's surface (in land seismic data acquisition)(Figure 1.1a) or from water surface (in marine seismic data acquisition)(Figure 1.1b) into the earth's subsurface. Due to changes in the physical properties (e.g. velocity, density, impedance etc.) of rocks in the subsurface, acoustic waves get reflected from interfaces and subsequently recorded by receivers (geophones or hydrophones) at the surface or on the sea surface. Seismic data recorded by the receivers contain information about the physical properties of the subsurface. We apply a sequence of processing steps on the raw seismic data to get subsurface image and estimate the physical properties of the subsurface.

We refer seismic data collected from a single shot experiment as a common-shot gather (Figure 1.2). In general, depending on seismic survey area, seismic data acquisition in-

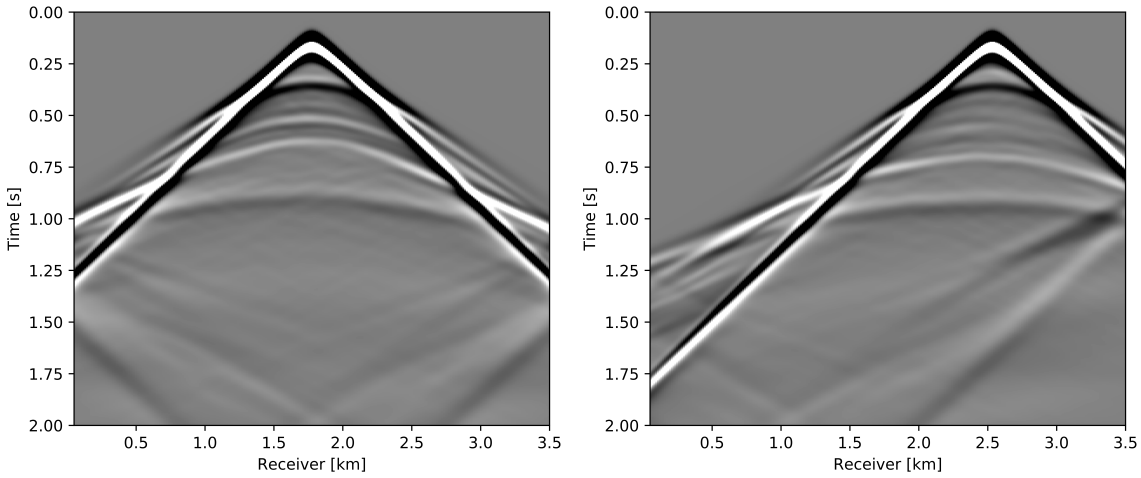


(a)



(b)

Figure 1.1: Schematic of *(a)* Land seismic data acquisition (Image courtesy: www.iongeo.com) and *(b)* marine seismic data acquisition (Caldwell and Walker 2011)



(a) (b)

Figure 1.2: Two Common-Shot Gathers

volves acquiring many such common-shot gathers (hundred of thousand to millions). To get a high resolution image of the subsurface, we require massive amounts of sources and receivers on a fine regular grid covering the whole survey area of interest. Seismic data acquisition at fine source and receiver interval is usually operationally and economically very challenging. Therefore, a general practice followed by the oil & gas industry is to acquire seismic data at coarse grids followed by data reconstruction onto fine grids.

Over the years many different seismic data reconstruction methods have been developed. Most of these methods are based on transformation to some domain such as Fourier, Radon, Wavelets, Curvelets, etc [(Bardan 1987); (Kabir and Verschuur 1995); (Villasenor, Ergas, and Donoho 1996); (Herrmann and Hennenfent 2008a)]. These transformations can be computationally demanding in case of large seismic datasets. Therefore, to avoid these expensive transformations, in recent years low-rank matrix completion [(Recht, Fazel, and Parrilo 2010b); (Aravkin, Kumar, Mansour, Recht, and Herrmann 2014)] based methods have been developed. Although, low-rank matrix completion based seismic data reconstruction methods are computationally cheaper than transform domain based methods, they perform poorly at higher frequencies. Presence of good quality high frequency seismic data is important for high resolution subsurface imaging. One of the topics of this thesis is to

improve the data reconstruction quality of low-rank matrix completion method at higher frequencies and to scale this improvement for large scale seismic datasets.

Conventional seismic data acquisition involves acquiring common-shot gathers sequentially. As we mentioned a typical industry scale seismic data acquisition involves acquiring massive amount of these common-shot gathers. This results in high turn-around time making seismic data acquisition very expensive and time-consuming. By using simultaneous source acquisition followed by deblending and reconstruction of sequential common shot gathers, we can reduce the total time and cost of acquisition. (Mansour, Wason, Lin, and Herrmann 2012) used curvelet transform based deblending and reconstruction of shot gathers. Curvelet based method can be computationally expensive for large scale (3D) seismic datasets. For this reason, one of the topics of this thesis is to incorporate low-rank matrix completion for deblending and reconstruction of common-shot gathers from seismic data acquired through simultaneous source acquisition.

The reconstructed seismic data usually goes through a series of processing steps (e.g. demultiple, imaging, FWI etc.) to obtain the final image of subsurface and also to obtain estimate of the physical properties of subsurface. Based on these images and subsurface physical properties, geological interpreters delineate different geological structures (e.g.faults, folds, unconformity, traps etc.) and potential oil & gas reservoirs. Based on the ease of extraction these oil & gas reservoirs are of two types — Conventional and Unconventional. As the name suggests conventional oil & gas reservoirs are naturally viable for extraction and upon drilling oil & gas can naturally flow through the wells (see Figure 1.3) making the production economical. On the other hand, unconventional reservoirs (see Figure 1.3) are not naturally viable for oil & gas extraction and require external stimulation to make them viable for production.

In the last few years, technological developments and population increase have resulted in an increase in demand for energy. Also, changes in the geopolitical situations and strong desire for countries to become self-dependent in terms of energy has fueled in the increased

demand for oil & gas. In last couple of decades, to meet the increased demand for energy, the energy industry have started producing oil & gas from the unconventional reservoirs. In conventional reservoirs, oil & gas are trapped in a permeable medium. On the other hand, in unconventional reservoirs, oil & gas have very low permeability and it requires external stimulation for production. This external stimulation includes creating fractures in the unconventional reservoir by injecting very high pressure fluid (includes water, chemicals and cement) through injection wells, commonly known as hydraulic fracturing or fracking (Figure 1.3). These fractures improve the permeability of the unconventional reservoirs allowing economical production of oil & gas. With recent advancement in engineering and technology, hydraulic fracturing has been frequently used across the globe for production of oil & gas from unconventional reservoirs. For example, in the year 2007 hydraulic fracturing was first used to produce oil from an unconventional reservoir in the Bakken structure in North Dakota and Montana (Wenzel 2012).

Even though hydraulic fracturing has made production of oil & gas from unconventional reservoirs viable, it poses its own environmental concerns. For example, one of the concern is the risk of these fractures interfering with nearby structures such as pre-existing wells, faults, etc. Interference with nearby faults can cause earthquakes posing significant risk for nearby population. Therefore, to prevent any hazardous situation, accurate detection of these fractures and how these fractures are originating in time is extremely important. Often these fractures generate small scale earthquakes or microseismic waves. These microseismic waves carry information (such as location, source-time function, origin time) about the microseismic sources located along the fractures. Oil & gas industry record these microseismic waves by receivers deployed on the surface or along a nearby monitor wells. One of the topics of this thesis is to use this microseismic data to detect closely spaced microseismic sources originating along these fractures along with estimation of associated source-time function in a computationally efficient way.

Similar to external stimulation in case of unconventional reservoirs that gives rise to

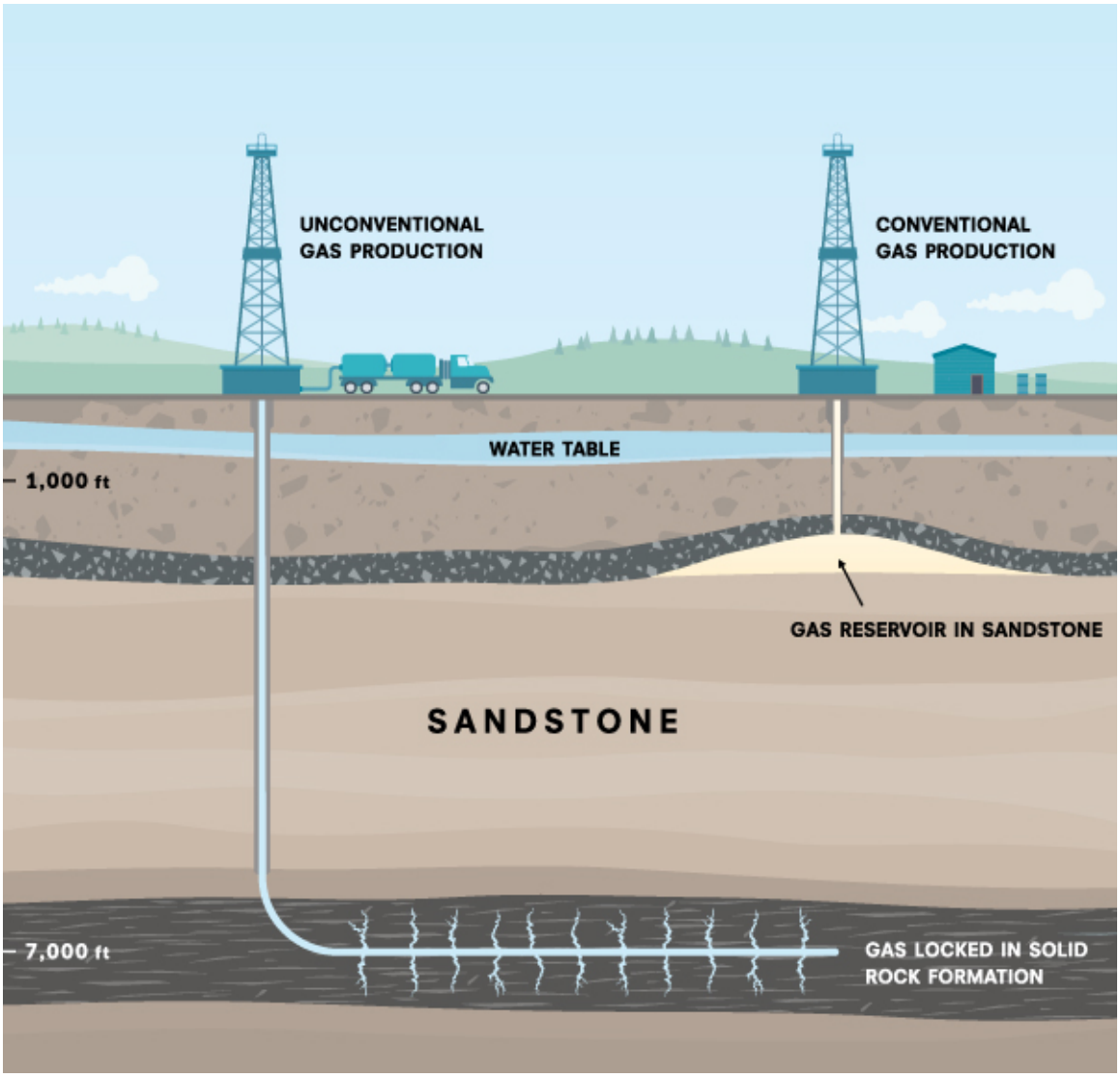


Figure 1.3: Schematic of conventional and unconventional reservoirs (Image courtesy: www.ic.gc.ca)

microseismic waves, in photoacoustic imaging cells are illuminated by electromagnetic waves giving rise to photoacoustic waves. As with microseismic waves, these photoacoustic waves carry information about the location, source-time function, and other physical properties of the photoacoustic sources. Information like source-time function, frequency content can be potentially useful for identifying different cell types. One of the topics of this thesis is to use these photoacoustic measurements to detect the location and associated source-time function.

1.1 High-frequency seismic data reconstruction

For economical extraction of oil and gas and to prevent any drilling related hazards, oil and gas industry rely on images of earth's subsurface and the physical parameters (e.g. velocity, density etc.) of earth's subsurface. We obtain images and invert for the physical parameters of earth's subsurface after series of processing steps (e.g. migration, demultiple, full waveform inversion etc.) on the raw seismic data acquired from field. Generally, these processing steps require seismic data on a dense periodic grid.

Seismic data acquisition on a dense periodic grid is challenging due to operational complexity and economical constraints. Therefore, a general practice followed by the oil and gas industry is to acquire seismic data on coarse grid followed by interpolation on a dense periodic grid. In recent years, Compressive Sensing (CS) (Donoho 2006; Candès, Romberg, and Tao 2005) based methods are developed for seismic data reconstruction (Herrmann and Hennenfent 2008a). These CS methods exploit the fact that densely sampled seismic data is sparse in some transform domain (e.g. Curvelets) and randomized subsampling destroys this sparsity. One of the biggest advantage of CS based seismic data reconstruction methods are that they can work with a smaller number of samples in comparison to the minimum number of samples required according to the Nyquist-Shannon sampling theorem. Compressive sensing can also be understood as a theoretical justification of transformed based seismic data reconstruction method (Xu, Zhang, Pham, and Lam-

baré 2005; Bardan 1987; Villasenor, Ergas, and Donoho 1996) that existed in Geophysics literature before the advent of CS. Unfortunately, for large scale realistic size seismic data these transform based methods can be very expensive.

(Recht, Fazel, and Parrilo 2010b) extended the idea of CS to matrix completion to fill the incomplete entries of a matrix from only a few known entries. Similar to the concept of sparsity, in case of matrix completion we exploit the fact that fully sampled matrix can be approximated by a low-rank matrix in some transform domain and randomized subsampling increases the rank of matrix required to approximate the subsampled matrix. (Aravkin, Kumar, Mansour, Recht, and Herrmann 2014) extended low-rank matrix factorization for seismic data reconstruction. Unlike the CS based methods, matrix-completion based methods are computationally cheaper and can be extended for realistic size 3D seismic data reconstruction (Kumar, Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015; Oropeza and Sacchi 2011; Yang, Ma, and Osher 2013).

Although low-rank matrix completion based methods perform well for seismic data reconstruction at lower frequencies, the quality of reconstructed data degrades at higher frequencies. This poor data reconstruction quality at higher frequencies is because seismic data at higher frequencies require matrices with higher ranks to approximate them. But the presence of good quality high frequency data on dense periodic grid is important for high resolution images of subsurface and detailed inversion of earth's subsurface physical parameters like velocity, density etc. Good quality subsurface images and estimation of physical parameters are important and essential for geological interpreters to delineate oil and gas reservoirs and hence minimizing the risk of drilling a dry well.

In (Aravkin, Kumar, Mansour, Recht, and Herrmann 2014), authors proposed weighted matrix completion that uses similarity between adjacent frequency slices as a prior information to improve the seismic data reconstruction quality. In (Eftekhari, Yang, and Wakin 2018) authors quantify this similarity as the largest principle angles between row and column subspaces of adjacent frequency slices, respectively. Smaller angles imply more reli-

ability on the adjacent frequency slice and higher angles mean less reliability. Although, weighted matrix completion improves the quality of reconstructed data, solving the associated optimization involves the action of an expensive to evaluate projection operator for each iteration making this method computationally expensive. One of the main contributions of this thesis is a computationally efficient formulation to solve the weighted matrix completion problem.

Although weighted matrix completion improves the data reconstruction quality, it depends on the availability of low-rank factors of adjacent frequency slices. Generally, we don't have access to these low-rank factors. Therefore, in this thesis we propose a recursively weighted matrix completion framework from lower to higher frequencies. In this recursively weighted framework, we start from low frequencies considering the fact that conventional low-rank matrix performs well at lower frequencies. This recursive approach reaps the full potential of weighted matrix completion in terms of improved reconstructed data quality at higher frequencies.

Recursively weighted approach can be computationally expensive as higher frequencies need to wait for reconstruction of all the lower frequencies. For realistic size 3D seismic data, this waiting time poses a major computational bottleneck. By 3D seismic data, we mean five-dimensional seismic data acquired by sources and receivers spread along a 2D plane. So these 5 dimensions are time, x and y coordinates of sources and receivers. Therefore, another contribution of thesis is the implementation of a computationally efficient recursively weighted approach. This approach parallelizes over individual rows of the low-rank factors making it computationally efficient. Depending on the availability of parallel computational nodes, we can scale the recursively weighted approach for large scale 3D seismic data.

1.2 Simultaneous-source acquisition

Because of operational complexity and economical considerations, seismic data acquisition is one of the major bottlenecks in the initial phase of oil & gas exploration. To avoid interference between common-shot records of two consecutive sources, conventional seismic data acquisition involves firing of sources at a fix interval of time on a uniform grid. Simultaneous source acquisition (Beasley, Chambers, and Jiang 1998; Berkhout 2008), in which more than one source fire at the same time, can make seismic data acquisition more efficient by reducing overall time of acquisition in comparison to the time taken by conventional acquisition. One of the main challenges of simultaneous source acquisition is to separate or deblend individual common-shot gathers from blended data. By using the fact that interference noise in the blended data appears as incoherent random noise in some transform domain such as common-receiver gather, (Hampson, Stefani, and Herkenhoff 2008) and (Moore, Dragoset, Ommundsen, Wilson, Ward, and Eke 2008) removed the interference noise. (Mansour, Wason, Lin, and Herrmann 2012); (Wason and Herrmann 2013) proposed Compressive Sensing (Donoho 2006; Candès, Romberg, and Tao 2005) based sparsity promoting framework for deblending. This framework exploits the fact that seismic data acquired through conventional acquisition is compressible or can be represented by a few sparse coefficients in curvelet domain (Candes and Donoho 2000) whereas interference noise in the blended data is incoherent.

One of the main requirements for success of sparsity promoting based method is randomized sampling. This sampling destroys the sparsity in the transform domain (e.g. the curvelet domain (Candes and Donoho 2000)) giving a favorable condition for recovery of fully sampled data. In case of marine seismic data acquisition randomized acquisition can be achieved by seismic sources firing at random times at random locations for a given speed of the vessel towing these seismic sources or airguns. Randomized marine seismic data acquisition can give rise to large gaps in seismic data. These large gaps are because

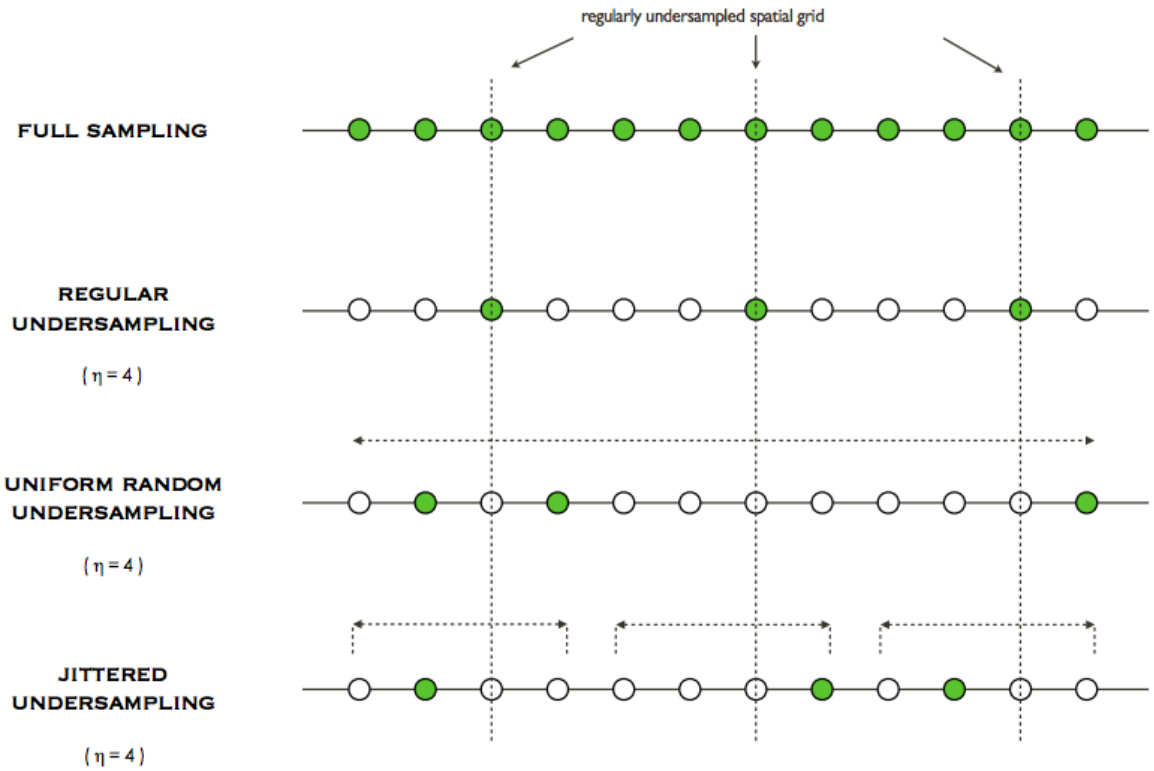


Figure 1.4: Schematic of different sampling schemes (Hennenfent and Herrmann 2008). η is the undersampling factor. For the same undersampling factor η, pure randomized sampling can lead to large gaps whereas jittered sampling controls the size of maximum gap.

of constraints on the variability of locations of seismic sources because of seismic vessels moving at a constant speed and recharge time of airguns. Unfortunately, these large gaps are not suitable for source separation using sparsity-promoting framework (Hennenfent and Herrmann 2008) since gaps bigger than spatio-temporal extent of the transform elements causes problems. To control the size of maximum gap, (Herrmann and Hennenfent 2008a) and (Hennenfent and Herrmann 2008) proposed jittered sampling scheme (Figure 1.4). (Wason and Herrmann 2013) incorporated the jittered sampling scheme to design time-jittered marine seismic data acquisition setup. This time-jittered acquisition provides randomness in acquisition and at the same time controls the size of maximum gap.

Although, sparsity-promoting approaches can reconstruct conventional seismic data

from blended data acquired using time-jitered acquisition setup, this method can be computationally challenging in terms of time and storage for large scale 3D seismic which entails datasets sampled along five dimensions. As we discussed earlier, matrix completion based methods provide good alternative in terms of computational efficiency in comparison to transform based methods for 3D seismic data reconstruction. One of the contributions of this thesis is the design of a computationally efficient matrix completion designed to scale well to industry-scale full-azimuth 3D wavefield reconstruction problems.

Seismic data acquired through conventional acquisition exhibits low-rank structure in some transform domain. Time-jittered (Herrmann and Hennenfent 2008a) acquisition setup destroys this low-rank structure. Therefore, the time-jittered acquisition setup provides a favorable condition to recover conventional seismic data using matrix-completion based method. Unlike in seismic data reconstruction, the sampling operator in case of time-jittered acquisition is not separable over frequency slices as it is a combined time-shifting and shot-jittering operator. Therefore, we cannot solve for individual frequency slices separately. Instead, we formulate our matrix-completion based source separation approach over temporal-frequency domain. By the temporal-frequency domain approach we mean that every iteration of the algorithm involves going back to time domain to calculate data residual after calculating current estimate of the low-rank factors for each frequency slice.

1.3 Microseismic source estimation

In unconventional reservoirs, hydraulic fracturing gives rise to small scale earthquakes or microseismic events causing propagation of microseismic waves that are recorded by receivers (Figure 1.5) planted along earth's surface or kept along monitor wells. Accurate detection of these microseismic sources and estimation of the associated source-time functions can help in tracking the evolution of the fractures in both space and time as these microseismic sources are localized along these fractures. Moreover, source origin time and source-time functions can potentially be used for the estimation of associated source

mechanism (Madariaga 1989). The source mechanism is helpful in estimating the moment tensor components that gave rise to microseismic events.

For last many years, (micro)seismic source estimation has been a topic of active research because of its similarity with the earthquake source estimation problem. Recently, many wave-equation based methods are developed for microseismic source estimation (McMechan 1982; Gajewski and Tessmer 2005). Wave equation based methods use information from the full waveform and hence try to exploit the true physics in order to get accurate source estimations. For example, time-reversal methods are based on simultaneously back propagating data from the receivers to focus on its source origin time and location (Fink 1997). These time-reversal methods require scanning the backpropagated volume in time and space to get the source origin time and source locations. This scanning can be very complex when there are multiple microseismic sources originating in spatial proximity and also in a close time interval giving rise to overlapping events. Also, time-reversal based methods require dense receiver and time sampling and wide receiver aperture to resolve close by sources. Due to operational complexity and economical constraints, acquiring microseismic data with dense sampling and wide aperture is usually a challenging task (Bazargani and Snieder 2016).

To avoid scanning the back propagated volume and to improve resolution of estimated locations of microseismic sources, (Nakata and Beroza 2016) proposed Geometric reverse time migration (GmRTM). GmRTM is based on back propagating data from individual receivers followed by cross-correlation. Although this method gives better resolution of locations, it comes at the additional cost of solving wave-equation for each individual receiver. GmRTM can be computationally very expensive specially in case of 3D. To avoid solving wave-equation for each receiver (Sun, Zhu, Fomel, and Song 2015) proposed Hybrid RTM (HyRTM) by simultaneously backpropagating group of neighboring receivers followed by cross-correlation. Although HyRTM improves the computational efficiency but it comes at a cost of reduced resolution. Also, selecting the number of adjacent receivers to form

groups is not a trivial problem (Nakata and Beroza 2016).

To overcome these issue such as choosing appropriate number of receivers in each group, full waveform inversion based methods such as (Wu and McMechan 1996) and (Kim, Liu, and Tromp 2011) invert for source locations, source origin time, moment tensor components with prior assumption about the shape of the source-time function. (Kaderli, McChesney, and Minkoff 2015) invert for both source location and source-time function but their method assumes the same source-time function for all the sources. Although this method can be extended to multiple sources with different source-time functions but then it requires knowledge on the number of sources a priori. Moreover, this problem is not a convex problem and can give rise to non-unique solutions. Therefore, there is no guarantee of successful recovery of locations and source-time functions.

One of the contributions of this thesis is to propose a sparsity-promoting microseismic source estimation method (Sharan, Wang, and Herrmann 2018). Our sparsity-promoting method exploits the fact that the microseismic sources are localized along fractures and has finite energy along time to simultaneously detect closely spaced (within half a dominant wavelength distance) microseismic sources along with estimation of associated source-time function. Our method does not have any assumptions on shape, origin time or number of microseismic sources. We solve the sparsity-promoting microseismic source estimation problem using modified version of the linearized Bregman algorithm (Yin, Osher, Goldfarb, and Darbon 2008; Lorenz, Schöpfer, and Wenger 2014). To make this method computationally efficient and for faster convergence, we propose a dual problem that can be solved using a quasi-Newton method such as LBFGS (Liu and Nocedal 1989).

Our sparsity-promotion based method can handle noise in the data but presence of very high amplitude noise with similar frequency range as of the microseismic signal poses challenge in successful detection of microseismic sources. Also, often due to complexity of model the amplitude information of the source-time function is not accurate. Therefore, another contribution of this thesis is a debiasing method that restores amplitude of source-

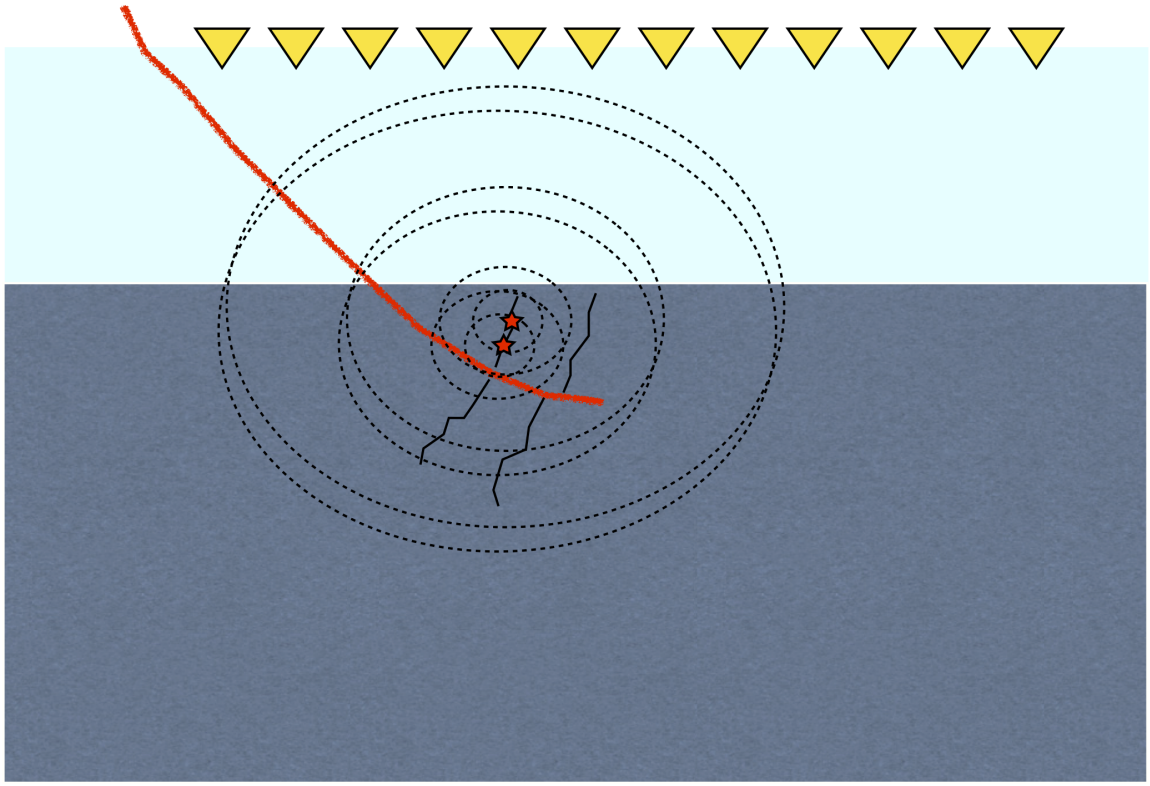


Figure 1.5: Hydraulic fracturing schematic diagram: Solid red colour curve shows the injection well through which high pressure fluid is injected creating fractures (solid black color). Fracturing gives rise to small scale earthquakes or microseismic events (red color stars) and subsequent emission of microseismic waves (dashed black color). These microseismic waves are recorded by receivers along the surface (yellow inverted triangle). Top of the figure depicts earth's surface.

time functions of multiple microseismic sources detected from noisy data.

1.4 Photoacoustic imaging

The challenges we face in the field of seismic bears a lot of similarity with the challenges we face in the field of medical imaging. For example, in the unconventional seismic, high pressure fluid injected through the injection wells creates fractures and eventually results in microseismic waves. Similarly in photoacoustic imaging, cells are illuminated by electromagnetic waves generated from laser pulse source giving rise to photoacoustic waves that are recorded by the transducers (Figure 1.6). In Photoacoustic imaging (or in general thermoacoustic imaging), electromagnetic waves are absorbed by cells and that causes heat

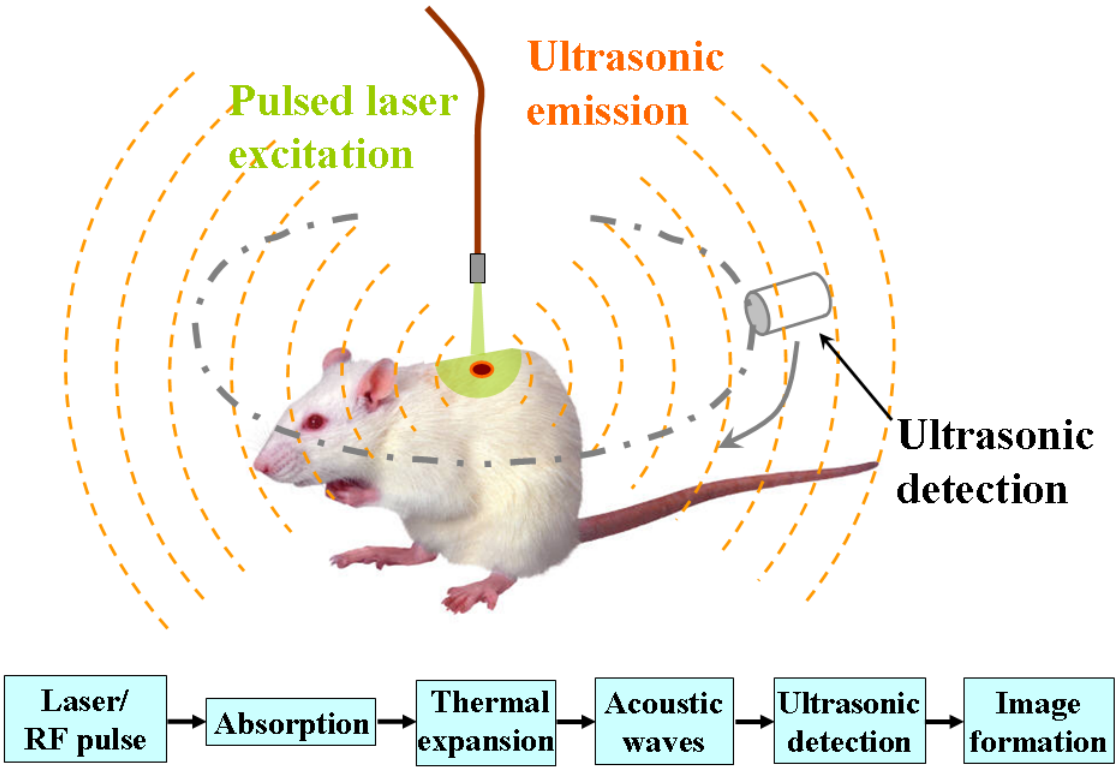


Figure 1.6: Schematic showing photoacoustic imaging (Image courtesy: Wikipedia)

deposition. This heat deposition results in thermal expansion of cells causing emission of photoacoustic waves. Photoacoustic based medical imaging techniques are of extreme importance for the detection and treatment of cancerous cells. As mentioned in (Emelianov, Li, and O'Donnell 2009), no other medical imaging methods (e.g. Xray, MRI, Ultrasound etc.) other than photoacoustic imaging has optical absorption as a contrast parameter enabling this method to image molecular properties at high resolution. For example, ultrasound imaging is based on the contrast in the mechanical parameters such as density or compressibility and therefore gives morphological information. In (Wang, Pang, Ku, and Wang 2003), authors showed potential of photoacoustic imaging in terms of high resolution imaging by showing oxygen utilization in brain of a mouse during left or right whisker stimulation.

Conventional photoacoustic imaging is based on backpropagation of photoacoustic data

from transducers, followed by extraction of back-propagated wavefield at time $t = 0$. Although, this method gives absorption map, this method does not give information about source-time function of these absorbers. These source-time functions and corresponding spectrum can be potentially useful for classifying different kinds of absorbers. For example, (Kumon, Deng, and Wang 2011) used frequency content of the photoacoustic waves for cancerous and non-cancerous region characterization. One of the contributions of this thesis is to use sparsity-promotion to image the photoabsorbers and estimate the associated source-time function.

Although we can get real time good quality images of photoabsorbers using the method of back-propagation, it requires dense sampling of transducers. Dense sampling of transducers can be expensive. In the method of back propagation, the image quality degrades with decreased sampling of transducers. Another topic of this thesis to obtain the absorption map and estimating the associated source-time function with smaller number of transducers.

1.5 Objectives

We can summarize main objectives of this thesis as follows:

1. To address the high-frequency seismic data reconstruction challenges faced by the low-rank matrix completion based seismic data reconstruction framework. Here, we use weighted matrix completion and exploit the similarity between adjacent frequency slices to improve the quality of seismic data reconstruction at higher frequencies. Another objective is to design a computationally efficient weighted matrix completion for large scale 3D seismic datasets.

2. To design low-rank matrix completion based simultaneous source seismic data acquisition and source-separation technique for decreasing the cost of marine seismic data acquisition. Here, low-rank matrix-completion provides computationally tractable

technique for source-separation from large scale 3D simultaneous source data.

3. To design a computationally feasible framework for detecting closely spaced microseismic sources along with estimation of the associated source-time functions. As these microseismic events are localized along the fractures created by fracking, therefore detection of closely spaced microseismic sources eventually helps in tracking fracture evolution in space and time. Here, we use take insights from compressive sensing and wave-equation based inversion techniques.

4. Microseismic data generated due to fracking is often heavily contaminated with high amplitude ambient noise in the same frequency range as that of microseismic signals. Presence of strong noise causes challenges in detecting location and estimating the associated source-time functions of microseismic sources. Hence, another objective of this thesis is to address this noise issue in microseismic data for detection and estimation of microseismic sources.

5. To estimate the location and source-time function of each photoabsorber using sparsity-promoting method to potentially distinguish different types of photoabsorbers or to distinguish different degree of aggregation of exogenous agents. Another objective is to decrease the storage cost by decreasing the number of transducers.

1.6 Thesis outline

This thesis consists of total 7 chapters including this introduction. In **chapter** 2, we first introduce the low-rank matrix completion based framework for large scale seismic data reconstruction. Next, we discuss the challenges associated with the low-rank matrix completion framework in reconstructing seismic data at high frequencies. Further, we introduce the recursively weighted matrix completion framework and demonstrate that by exploiting similarity between the adjacent frequency slices the recursively weighted low-rank matrix

completion can improve the reconstruction quality of missing seismic data at higher frequencies. Next, we discuss how the interdependence of frequency slices in the recursively weighted framework does not allow parallelization across frequencies and making it computationally expensive. Next we discuss how we made this recursively weighted framework computationally efficient using strategies of alternation and decoupling. We examine the effectiveness of this method on field data from Gulf of Suez and synthetic data simulated using complex geological model such as BG compass (E. Jones, A. Edgar, I. Selvage, and Crook 2012). A version of this chapter has been submitted to *Geophysics*.

In **chapter** 3, we first introduce the modification in the low-rank matrix completion formulation to separate sources from seismic data acquired by simultaneous source acquisition in computationally efficient manner. Next we apply this method on a realistic size 3D simultaneous source seismic data simulated using BG compass model to reconstruct data from individual sources. A version of this chapter was published in *SEG Technical Program Expanded Abstracts* (Kumar, Sharan, Wason, and Herrmann 2016).

In **chapter** 4, we start by introducing a sparsity-promotion based formulation to estimate the location and source-time function of microseismic sources. Next, we discuss the computational challenges associated with the algorithm in terms of resolving closely spaced microseismic sources. Further, we discuss how we made this problem computationally cheaper by solving the original sparsity-promotion problem in its dual form. By showing the application of sparsity-promoting framework on synthetic datasets generated using multiple sources in complex geological models such as BG Compass and Marmousi, we justify the selection of $\ell_{2,1}$-norm over Frobenius norm in terms of better resolution of location of closely spaced sources. A version of this chapter was published in *Geophysical Journal International* (Sharan, Wang, and Herrmann 2018).

The sparsity-promoting framework seems to perform poorly in presence of lots of noise. Therefore, in **chapter** 5 we introduce debiasing in our sparsity-promoting microseismic

source estimation framework to detect the location of microseismic sources along with estimation of source-time functions with correct amplitude from extremely noisy data. A version of this chapter was published in *SEG Technical Program Expanded Abstracts* (Sharan, Kumar, Wang, and Herrmann 2018).

In **chapter** 6, we extend the sparsity-promoting framework for photoacoustic imaging to detect the photoacoustic absorbers. Here, we modify the original sparsity formulation to take into account of the fact that unlike microseismic sources, all the photoacoustic sources have same source-time function and are activated for a very small time interval. Here, we also show effectiveness of this method with lesser number of transducers by comparing the images obtained from our method and images obtained from the k-Wave MATLAB toolbox (Treeby and Cox 2010). A version of this chapter was published in *IEEE International Ultrasonics Symposium* (Sharan, Kumar, Dumani, Louboutin, Wang, Emelianov, and Herrmann 2018).

Finally, in **chapter** 7, I present conclusions of this thesis and discuss the future research direction.

CHAPTER 2

LARGE SCALE HIGH FREQUENCY WAVEFIELD RECONSTRUCTION WITH RECURSIVELY WEIGHTED MATRIX FACTORIZATIONS

2.1 Summary

Seismic data acquisition on a regular periodic fine grid is essential for many seismic data processing steps such as migration and multiple removal. Unfortunately, acquiring seismic data on a regular periodic fine grid is challenging because of environmental and budgetary constraints. By exploiting the low-rank approximation property of fully sampled seismic data in some transform domain, low-rank matrix completion offers a scalable way to reconstruct seismic data on a regular periodic fine grid from coarsely randomly sampled data acquired in the field. While wavefield reconstruction have been applied successfully at the lower end of the spectrum, its performance deteriorates at the higher frequencies where the low-rank assumption no longer holds rendering this type of wavefield reconstruction ineffective in situations where high resolution images are desired. We overcome this shortcoming by exploiting similarities between adjacent frequency slices explicitly. During low-rank matrix factorization, these similarities translate to alignment of subspaces of the factors, a notion we propose to employ as we reconstruct monochromatic frequency slices recursively starting at the low frequencies. While this idea is relatively simple in its core, to turn this recent insight into a successful scalable wavefield reconstruction scheme for 3D seismic requires a number of important steps. First, we need to move the weighting matrices, which encapsulate the prior information from adjacent frequency slices, from the objective to the data misfit constraint. This move considerable improves the performance of the weighted low-rank matrix factorization on which our wavefield reconstructions is based. Secondly, we introduce approximations that allow us to decouple computations on

a row-by-row and column-by-column basis, which in turn allow to parallelize the alternating optimization on which our low-rank factorization relies. The combination of weighting and decoupling leads to a computationally feasible full-azimuth wavefield reconstruction scheme that scales to industry-scale problem sizes. We demonstrate the performance of the proposed parallel algorithm on a 2D field data and on a 3D synthetic dataset. In both cases our approach produces high-fidelity broadband wavefield reconstructions from severely (up to 90%) subsampled data.

2.2 Introduction

In this chapter, we consider wavefield reconstruction from randomized samples taken from a periodic grid. The reader is referred to (López, Yilmaz, and Herrmann 2016) for an off-the-grid extension of presented wavefield reconstruction methodology.

In recent years, several methods for wavefield reconstruction have been developed. Many of these methods perform wavefield reconstruction in a transformed domain involving Fourier (Xu, Zhang, Pham, and Lambaré 2005), Radon (Bardan 1987), wavelet (Villasenor, Ergas, and Donoho 1996), or curvelet (Herrmann and Hennenfent 2008a) domain. While powerful, sparsity-based wavefield reconstruction does not scale well to 3D seismic where the data volumes become prohibitively large when structure is promoted along more than three dimensions, e.g. along all four source and receiver coordinates. By exploiting low-rank properties of matrices and tensors (Kumar, Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015; Oropeza and Sacchi 2011), some of these high dimensional challenges have been overcome by building on early work of (Recht, Fazel, and Parrilo 2010b), who extended some of the ideas of compressed sensing to matrices. (Kumar, Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015) and (Da Silva and Herrmann 2015) exploited these properties and formalized matrix-/tensor-based wavefield reconstructions that are practical for large-scale seismic datasets (Kumar, Wason, Sharan, and Herrmann 2017).

As demonstrated in the work by (Kumar, Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015), low-rank matrix completion methods work well when reconstructing seismic data at the lower angular frequencies but the recovery quality degrades when we move to the higher frequencies ($> 15\,\mathrm{Hz}$). Unfortunately, techniques such as multiple elimination and migration need access to high-frequency data to create high-fidelity artefact-free high-resolution images. This is especially true when physical properties are of interest in areas of complex geology.

To meet the challenges of recovering seismic data at high frequencies, we build on earlier work by (Aravkin, Kumar, Mansour, Recht, and Herrmann 2014) and (Eftekhari, Yang, and Wakin 2018) who discussed how to improve the performance of low-rank matrix completion by including prior information in the form of weighting matrices. The weighting matrices are projections spanned by the row and column subspaces (and their complements) of a low-rank matrix factorization of a matrix that is close to the to-be-recovered matrix. As with weighted ℓ_1-norm minimization, these weighting matrices improve the wavefield recovery if the principle angle between the subspaces of the weighting matrices and the to-be-recovered matrix is small. Conceptually, this is the matrix counterpart of weighted ℓ_1-norm minimization proposed by (Mansour, Herrmann, and Yılmaz 2012). (Aravkin, Kumar, Mansour, Recht, and Herrmann 2014) and (Eftekhari, Yang, and Wakin 2018) showed that wavefield recovery via matrix completion can be improved when low-rank factorizations from adjacent frequency slices are used to define these weighting matrices. (Aravkin, Kumar, Mansour, Recht, and Herrmann 2014) used this principle assuming access to the low-rank factorization of an adjacent frequency slice using a modified version of the spectral-projected gradient algorithm (Van Den Berg and Friedlander 2008). Also, (Eftekhari, Yang, and Wakin 2018) showed that for small principle angles, these weighing matrices reduce sampling requirement for successful data reconstruction by a logarithmic factor in comparison to the sampling requirement for the conventional matrix completion method.

While the initial results on wavefield reconstruction via weighted matrix completion were encouraging, the presented approach was not very practical because it relied on having access to the weights. In addition, the optimization relied on a computationally expensive optimization algorithm. We overcome these shortcomings by proposing a parallelizable recursive method that uses a recently developed alternating minimization procedure (Xu and Yin 2013; Jain, Netrapalli, and Sanghavi 2013) proposed by (Lopez, Kumar, and Herrmann 2015). Thanks to the recursive reconstruction, as first proposed by (Zhang, Sharan, and Herrmann 2019), and the improved optimization we will demonstrate that we are able to improve the performance of our wavefield reconstruction algorithm.

The outline of this chapter is as follows. We first provide a short overview of the principles of wavefield reconstruction via matrix completion. We follow this brief exposition by describing the challenges of high-frequency wavefield reconstruction and how these challenges can be addressed through weighted matrix completion. After this introduction, we describe how to derive a formulation in factored form, which allows to drastically reduce the problem size rendering our approach practical for 3D seismic data. In particular, we describe how our algorithm can be parallelized and applied to a large-scale high-frequency seismic wavefield reconstruction problem.

2.3 Wavefield reconstruction via weighted matrix completion

According to the seminal work of (Recht, Fazel, and Parrilo 2010b), matrices that exhibit low-rank structure can be recovered from random missing entries through a nuclear norm minimization procedure. During the optimization the sum of the singular values is minimized. As long as the randomized subsampling decreases the rate of decay of the singular values, this type of minimization allows for the recovery of matrices that are well approximated by low-rank matrices when fully sampled. (Kumar, Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015) used this principle to recover frequency slices from seismic lines in the midpoint-offset domain or from 3D seismic data permuted in non-

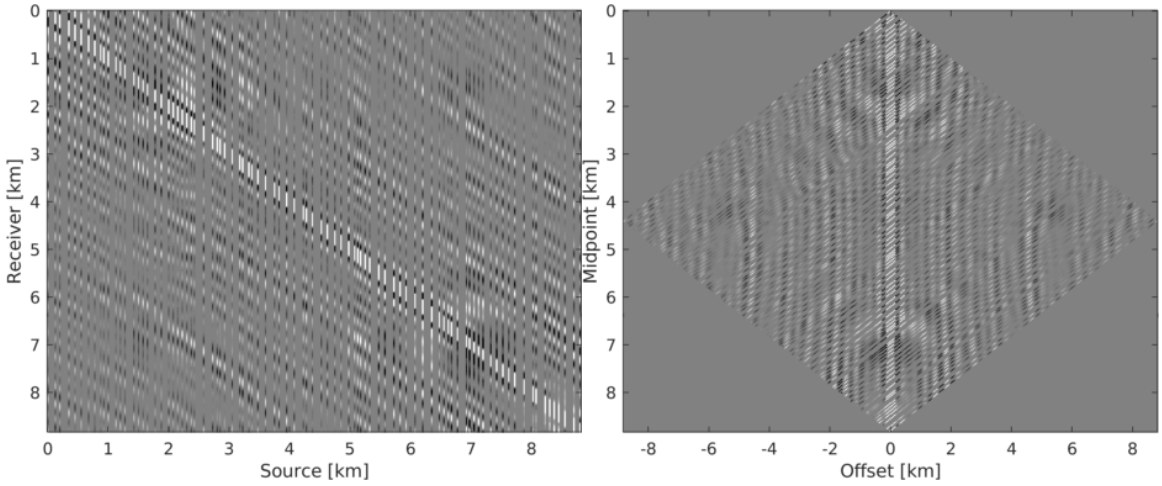
canonical form (Da Silva and Herrmann 2015). In either case, the resulting frequency slice can be approximated accurately by a low-rank matrix factorization.

To illustrate the underlying principle of wavefield reconstruction via matrix completion, we consider a $12\,\mathrm{Hz}$ monochromatic frequency slice assembled from a 2D line acquired in the Gulf of Suez. Figure 2.1, includes the real part of this frequency slice in the source-receiver and midpoint-offset domain after removing 75% of the sources via jittered subsampling (Hennenfent and Herrmann 2008). Compared to uniform random subsampling, jittered subsampling controls the maximal spatial gap between sources, which favors wavefield reconstruction. While the monochromatic data contained in these frequency slices is comparable, the behavior of the singular values before and after subsampling is very different before and after transforming to the midpoint-offset domain (juxtapose Figures 2.2a and 2.2b). The singular values for the matricization in the shot-domain (denoted by the dashed lines) decay slowly when fully sampled and fast when subsampled, which can be understood since removing rows or columns from a matrix reduces the rank. The converse is true for data in the midpoint-offset domain, which shows a fast decay of the singular values for the fully sampled data and a slow decay after randomized subsampling. The latter creates favorable conditions for recovery via matrix completion (Kumar, Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015) via

$$\mathbf{X} := \arg\min_{\mathbf{Y}} \quad \|\mathbf{Y}\|_* \quad \text{subject to} \quad \|\mathcal{A}(\mathbf{Y}) - \mathbf{B}\|_F \leq \epsilon, \tag{2.1}$$

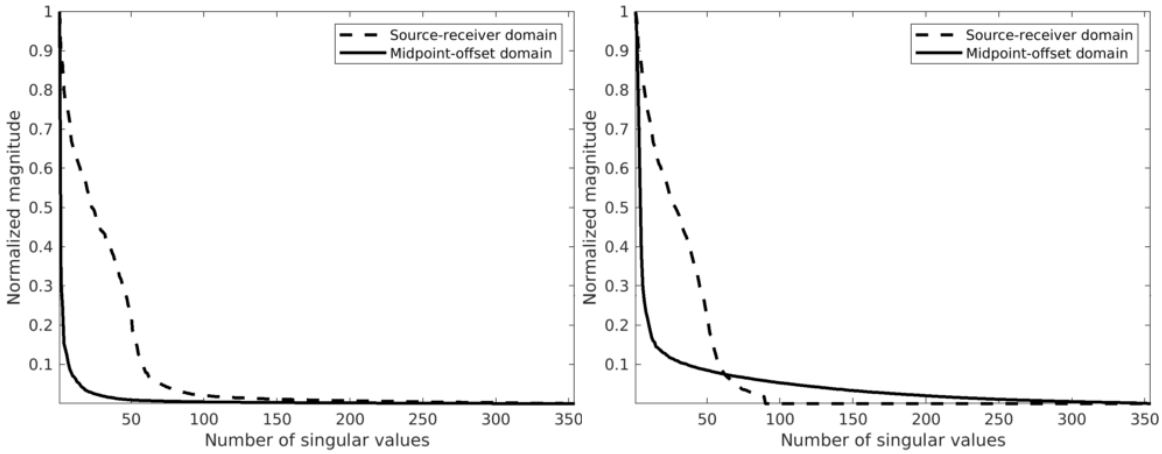
which promotes low-rank matrices.

By solving this minimization problem, we aim to recover the minimum nuclear norm ($\|\mathbf{X}\|_* = \sum \sigma_i$ with the sum running over the singular values of $\mathbf{X}$) of the complex-valued data matrix $\mathbf{X} \in \mathbb{C}^{m\times n}$, with m offsets and n midpoints. Aside from minimizing the nuclear-norm objective, the minimizer fits the observed data $\mathbf{B} \in \mathbb{C}^{m\times n}$ at the sampling locations to within some tolerance ϵ measured by the Frobenious norm—i.e.,



(a) (b)

Figure 2.1: $12.0\,\mathrm{Hz}$ frequency slice extracted from 2D seismic data acquired in Gulf of Suez. Data with 75% missing random jittered sources in *(a)* source-receiver domain and *(b)* in midpoint-offset domain.



(a) (b)

Figure 2.2: Decay of singular values for $12.0\,\mathrm{Hz}$ frequency slice in source-receiver and midpoint-offset domain for *(a)* full data and for *(b)* subsampled data with 75% missing sources.

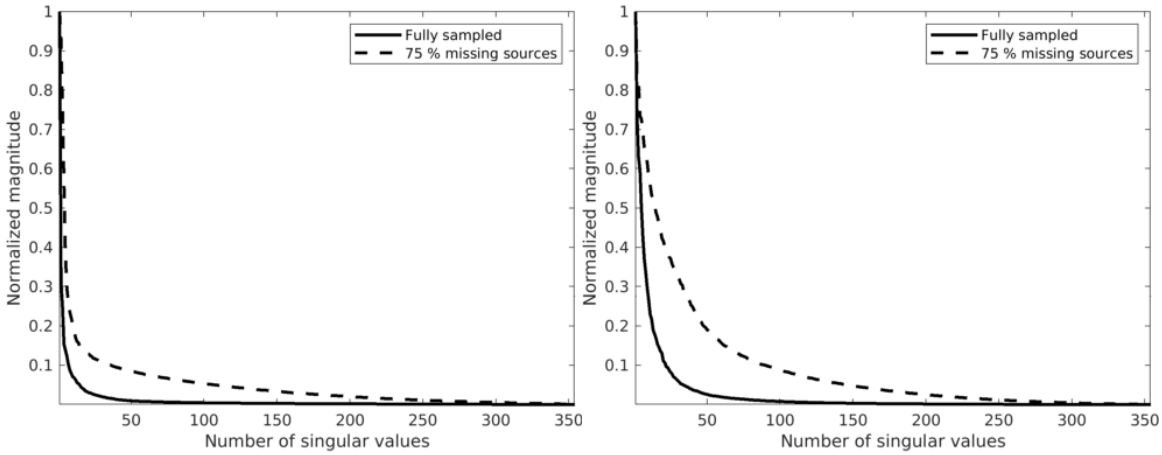
$\|\mathbf{D}\|_F = \sqrt{\sum_j \sum_k D_{jk}^2}$ for a matrix $\mathbf{D}$. In this expression, the linear operator $\mathcal{A}$ implements the sampling mask putting zeros at source (and possibly receiver) locations that are not collected in the field. Here, the matrix $\mathbf{Y}$ is the optimization variable. Equation 2.1 is similar to the classic Basis Pursuit DeNoising problem (BPDN, Van Den Berg and Friedlander 2008) and can be solved with a modified version of the SPGℓ_1 algorithm adapted for nuclear-norm minimization (Aravkin, Kumar, Mansour, Recht, and Herrmann 2014). To solve problem 2.1, SPGℓ_1 solves a series of constrained subproblems during which the nuclear-norm constraint is relaxed to fit the observed data.

2.3.1 The challenge of high-frequency wavefield recovery

Wavefield reconstruction via matrix completion (cf. problem 2.1) relies on the assumption that the singular values of monochromatic data organized in matrix decay rapidly. For the lower frequencies ($< 15.0\,\mathrm{Hz}$) this is indeed the case but unfortunately this assumption no longer holds for the higher frequencies. To illustrate this phenomenon, we compare in Figure 2.3 the decay of the singular values for the two matricizations of Figure 2.2 at $12.0\,\mathrm{Hz}$ and $60.0\,\mathrm{Hz}$. While the singular values at $12.0\,\mathrm{Hz}$ indeed decay quickly this is clearly no longer the case at $60.0\,\mathrm{Hz}$ (juxtapose solid lines in Figures 2.3a and 2.3b) where the singular values for the fully sampled data decay more slowly. This slower decay at the high frequencies is caused by the increased complexity and oscillatory behavior exhibited by data at higher temporal frequencies. Despite the fact that the randomized source subsampling slows the decay down, the slower decay of the fully sampled data leads to poor wavefield reconstruction (Figure 2.4a) and unacceptable large residuals (Figure 2.4b) at $60.0\,\mathrm{Hz}$.

2.3.2 Weighted matrix completion

As Figures 2.3, 2.4a, and 2.4b illustrate, the success of wavefield reconstruction by minimizing the nuclear norm (cf. equation 2.1) hinges on rapid decay of the singular values



(a) (b)

Figure 2.3: Singular value decay for fully sampled and subsampled data (75% missing sources) in midpoint-offset domain for *(a)* $12.0\,\mathrm{Hz}$ and *(b)* $60.0\,\mathrm{Hz}$ frequency slice

an assumption violated at the higher frequencies. This shortcoming can, at least in part, be overcome by using prior information from a related problem in the form of weights, an approach initially put forward by (Aravkin, Kumar, Mansour, Recht, and Herrmann 2014) and further theoretically analyzed by (Eftekhari, Yang, and Wakin 2018). In its original form, the weights were derived from the reconstruction of the wavefield at a neighboring temporal frequency, which leads to a significant improvement for the reconstruction and the residual plotted in Figures 2.4c and 2.4d, respectively. By applying this approach recursively from low to high frequencies, (Zhang, Sharan, and Herrmann 2019) improved the reconstruction even further judged by the quality of Figure 2.4e and the size of the residual plotted in Figure 2.4f. In this work, we further extend this result by reformulating the optimization problem and by introducing a parallel algorithm that limits communication.

We obtained the above weighted wavefield reconstructions by minimizing (Aravkin, Kumar, Mansour, Recht, and Herrmann 2014; Eftekhari, Yang, and Wakin 2018)

$$\mathbf{X} := \arg\min_{\mathbf{Y}} \quad \|\mathbf{QYW}\|_* \quad \text{subject to} \quad \|\mathcal{A}(\mathbf{Y}) - \mathbf{B}\|_F \leq \epsilon, \tag{2.2}$$

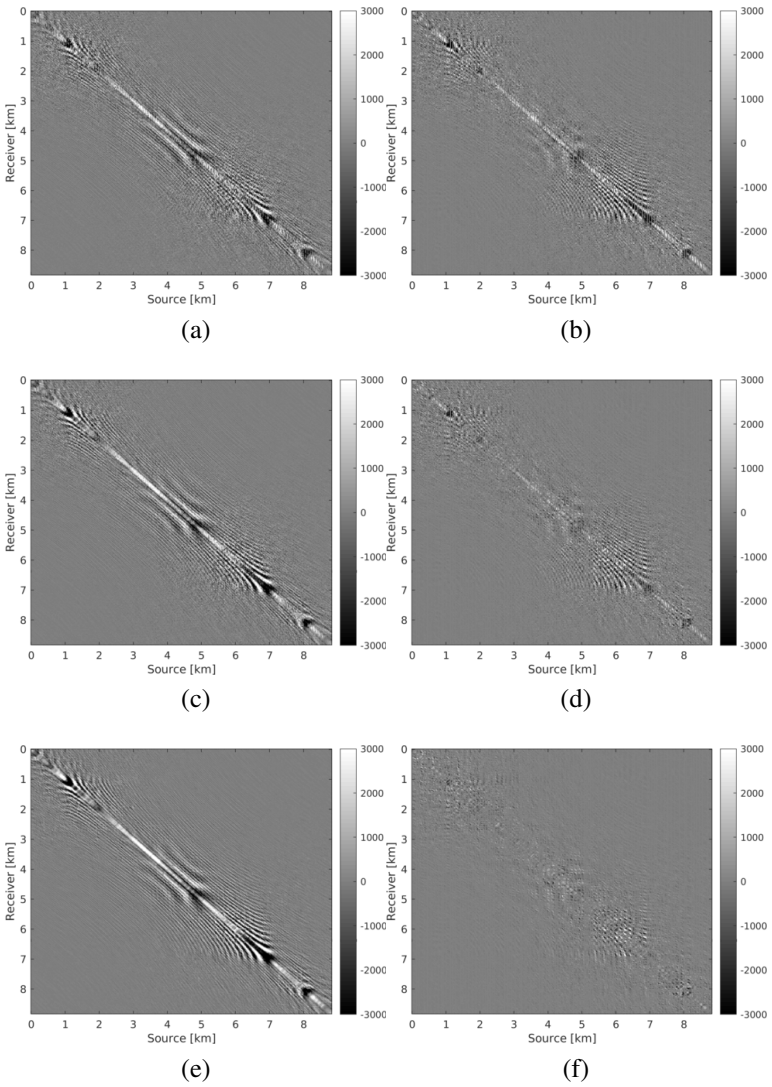


Figure 2.4: Wavefield reconstruction comparison for a 60 Hz frequency slice. *(a)* Reconstructed wavefield from 75% subsampling. *(b)* residual with a poor S/R = 2.83 dB. *(c)* Reconstructed wavefield using the recovery at the adjacent lower frequency as weights and *(d)* improved residual with S/R = 5.08 dB. *(e)* and *(f)* the same but now with the weighting scheme applied recursively with significantly improved S/R = 8.72 dB.

where the weighting matrices $\mathbf{Q} \in \mathbb{C}^{m \times m}$ and $\mathbf{W} \in \mathbb{C}^{n \times n}$ are projections given by

$$\mathbf{Q} = w_1 \mathbf{U}\mathbf{U}^H + \mathbf{U}^\perp \mathbf{U}^{\perp H} \tag{2.3}$$

and

$$\mathbf{W} = w_2 \mathbf{V}\mathbf{V}^H + \mathbf{V}^\perp \mathbf{V}^{\perp H}, \tag{2.4}$$

where the symbol ^{H} denotes complex transpose. These projections are spanned by the row and column subspaces $\mathbf{U}$, $\mathbf{V}$ and their orthogonal complements $\mathbf{U}^\perp$ and $\mathbf{V}^\perp$. Also, these subspaces $\mathbf{U}$, $\mathbf{V}$ have orthonormal columns making $\mathbf{U}\mathbf{U}^H$ and $\mathbf{V}\mathbf{V}^H$ orthogonal projections. The pair of matrices $\{\mathbf{U}, \mathbf{V}\}$ are low rank and can be obtained from the factorization of a lower adjacent frequency slice. The choice for the weights w_1 and w_2 in equations 2.3 and 2.4 depends on the similarity between the corresponding row and column subspaces of the two adjacent frequency slices. We follow (Eftekhari, Yang, and Wakin 2018) and quantify this similarity by the largest principle angle between these subspaces. The smaller this angle, the more similar the subspaces from the two adjacent frequency slices will be. In situations where the adjacent frequency slices are near orthogonal—i.e., have a near 90° angle, we choose $w_1 \uparrow 1$ and $w_2 \uparrow 1$ so that the weighting matrices $\mathbf{Q}$ and $\mathbf{W}$ become identity matrices. In that case, the weighting matrices should not add information—i.e., the solution of problem 2.2 should become equivalent to solving the original problem in equation 2.1. Conversely, when the subspaces are similar—i.e., they have an angle $\ll 90^\circ$, then the w_1 and w_2 should be chosen small such that we penalize solutions more in the orthogonal complement space. Depending on our confidence in the given factorization, we chose these weights close to one when we have little confidence and close to zero when we have more confidence.

While replacing the nuclear-norm objective in equation 2.1 by its weighted counterpart in equation 2.2 is a valid approach responsible for improvements reported in Figure 2.4, its solution involves non-trivial weighted projections (see equation 7.3 in Aravkin, Ku-

mar, Mansour, Recht, and Herrmann 2014). These computationally costly operations can be avoided by rewriting optimization problem 2.2 in a slightly different form where the weights are moved from the objective to the data constraint—i.e., we have

$$\bar{\mathbf{X}} := \arg\min_{\mathbf{Y}} \quad \|\bar{\mathbf{Y}}\|_* \quad \text{subject to} \quad \|\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{Y}}\mathbf{W}^{-1}) - \mathbf{B}\|_F \leq \epsilon. \tag{2.5}$$

In this formulation, the optimization is carried out over the new variable $\bar{\mathbf{Y}} = \mathbf{QYW}$. After solving for this variable, we recover the solution of the original problem $\mathbf{X}$ from $\bar{\mathbf{X}}$ as follows: $\mathbf{X} = \mathbf{Q}^{-1}\bar{\mathbf{X}}\mathbf{W}^{-1}$. We arrived at this formulation by using the fact that the matrices $\mathbf{Q}$ and $\mathbf{W}$ are invertible (for non-zeros weights w_1 and w_2) with inverses given by

$$\mathbf{Q}^{-1} = \frac{1}{w_1}\mathbf{U}\mathbf{U}^H + \mathbf{U}^{\perp}\mathbf{U}^{\perp H} \tag{2.6}$$

and

$$\mathbf{W}^{-1} = \frac{1}{w_2}\mathbf{V}\mathbf{V}^H + \mathbf{V}^{\perp}\mathbf{V}^{\perp H}. \tag{2.7}$$

Because we moved the weighting matrices to the data constraint, we no longer have to project onto a more complicated constraint as in (Aravkin, Kumar, Mansour, Recht, and Herrmann 2014), which results in solutions of equation 2.5 at almost the same computational costs as in the original formulation (Equation 2.1). This formulation forms the basis for our approach to wavefield reconstruction that is capable of handling the large data volumes of 3D seismic.

2.4 Scalable multi-frequency seismic wavefield reconstruction

So far, our minimization problems relied on explicit formation of the data matrix and on the singular-value decomposition (SVD, (Aravkin, Kumar, Mansour, Recht, and Herrmann 2014)) both of which are unfeasible for industry-scale 3D wavefield reconstruction problems. To address this issue, we discuss how to recast the above weighted matrix completion

approach into factored form, which has computational benefits and, as we will show below, can still be parallelized.

2.4.1 Weighted low-rank matrix factorization

To avoid computing costly SVDs, we first cast the solution of equation 2.5 into factored form:

$$\bar{\mathbf{L}}, \bar{\mathbf{R}} := \underset{\bar{\mathbf{L}}_{\#}, \bar{\mathbf{R}}_{\#}}{\arg\min} \quad \frac{1}{2} \left\| \begin{bmatrix} \bar{\mathbf{L}}_{\#} \\ \bar{\mathbf{R}}_{\#} \end{bmatrix} \right\|_F^2 \quad \text{subject to} \quad \|\mathcal{A}(\mathbf{Q}^{-1} \bar{\mathbf{L}}_{\#} \bar{\mathbf{R}}_{\#}^H \mathbf{W}^{-H}) - \mathbf{B}\|_F \leq \epsilon, \tag{2.8}$$

where $\bar{\mathbf{L}} = \mathbf{QL}$ and $\bar{\mathbf{R}} = \mathbf{WR}$. Under certain technical conditions (Candes and Recht 2009), which include choosing the proper rank r, the factored solution, $\mathbf{X} = \mathbf{L}\mathbf{R}^H$ with $\mathbf{L} = \mathbf{Q}^{-1}\bar{\mathbf{L}}$ and $\mathbf{R} = \mathbf{W}^{-1}\bar{\mathbf{R}}$, corresponds to the solution of the weighted problem included in equation 2.2. Here, the matrices $\mathbf{L} \in \mathbb{C}^{m \times r}$ and $\mathbf{R} \in \mathbb{C}^{n \times r}$ are the low-rank factors of $\mathbf{X}$. Using the property that the matrices $\mathbf{W}^H = \mathbf{W}$ and $\mathbf{Q}^H = \mathbf{Q}$ in equation 2.8 are idempotent, we replace $\mathbf{W}^{-H}$ by $\mathbf{W}^{-1}$ to avoid extra computation. Compared to the original convex formulation, equation 2.8 can be solved with alternating optimization, which is computationally efficient as evidenced from the runtimes plotted in Figure 2.5 as a function of temporal frequency. Of course, this approach only holds as long as the monochromatic data matrices can be well approximated by low rank matrices—i.e., $r \ll \min(m, n)$.

While the above weighted formulation allows us to solve the problem in factored form, it needs access to the subspaces $\{\mathbf{U}, \mathbf{V}\}$, which requires computing the full SVD (Eftekhari, Yang, and Wakin 2018). Since we cannot compute this full SVD, we instead orthogonalize the low-rank factors from adjacent frequency slices themselves by carrying out computationally cheap SVDs on the factors rather than on the full data matrix and then keeping the top r left singular vectors. This approach is justified because orthogonalizing low-rank factors allows approximating orthogonal subspaces spanned by the full frequency slice.

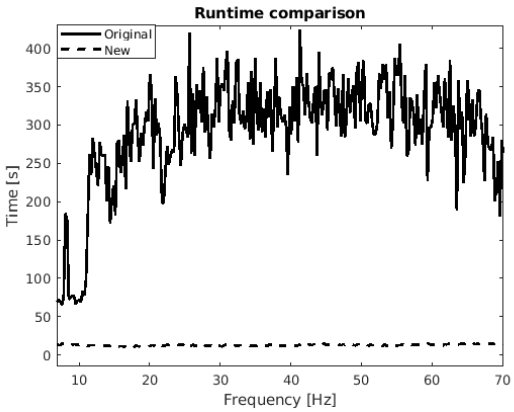


Figure 2.5: Runtime comparison plot: Solid black line shows runtime of the original weighted formulation and dashed black line shows runtime of the new weighted formulation for same number of iterations with same data residual at the end.

The results presented in Figure 2.4 were obtained in factored form and demonstrated clearly how the incorporation of weight matrices improves recovery especially when these weight matrices are calculated recursively from low to high frequencies (juxtapose Figures 2.4c, 2.4d and 2.4e, 2.4f). This improvement is due to the fact that low-frequency data matrices can be better approximated by low-rank matrices, which improves the recovery and therefore the weighted reconstruction.

2.4.2 Weighted parallel recovery

The example in Figure 2.4 made it clear that wavefield reconstruction via matrix factorization improves when including weight matrices that carry information on the row and column subspaces. However, inclusion of these weight matrices makes it more difficult to parallelize the algorithm because the parallelized alternating optimization approach by (Recht and Ré 2013) and (Lopez, Kumar, and Herrmann 2015) no longer applies straightforwardly. That approach relies on decoupled computations on a row-by-row and column-by-column basis (see Figure 2.6) in which case one alternates between minimizing the rows

via

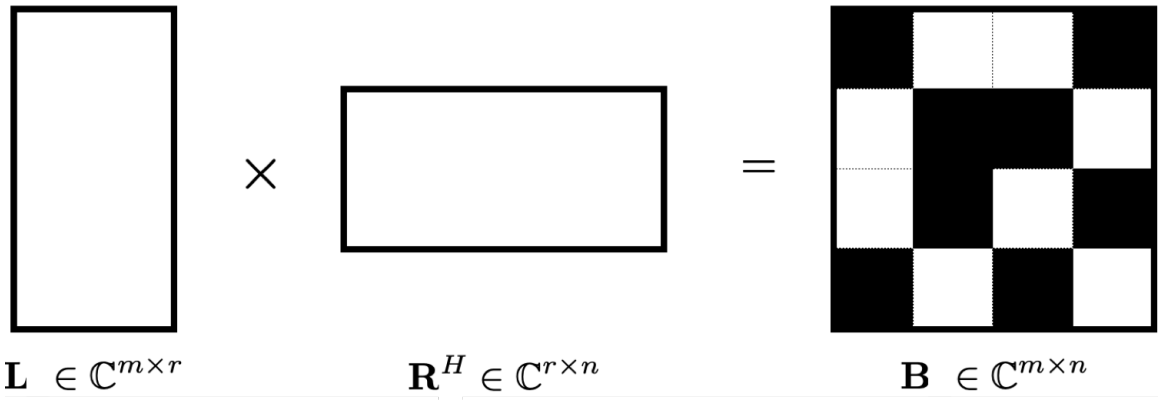
$$\mathbf{R}(l_1,:)^H := \arg\min_{\mathbf{v}} \quad \frac{1}{2}\|\mathbf{v}\|^2 \quad \text{subject to} \quad \|\mathcal{A}_{l_1}(\mathbf{L}\mathbf{v}) - \mathbf{B}(:,l_1)\| \leq \gamma \tag{2.9}$$

for $l_1 = 1 \cdots n$ and the columns via

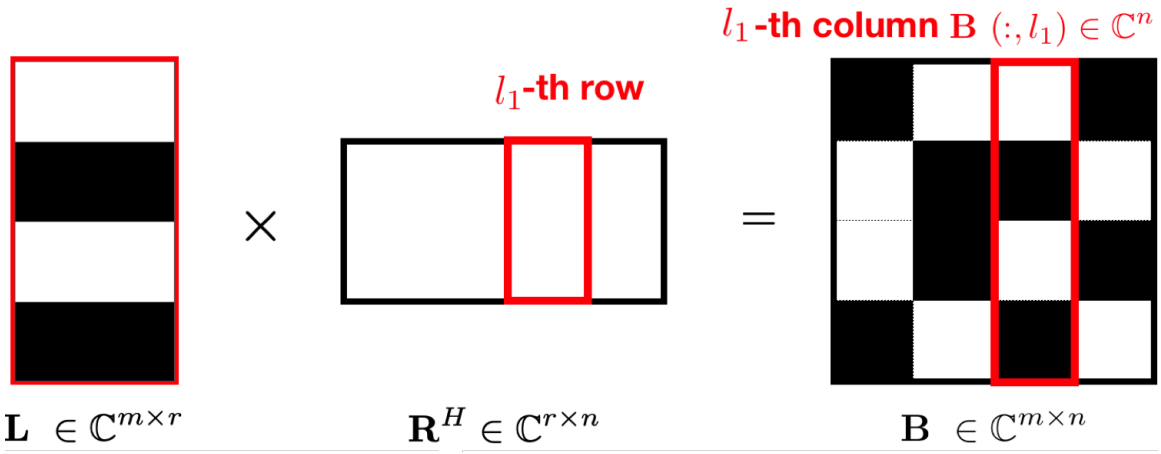
$$\mathbf{L}(l_2,:)^H := \arg\min_{\mathbf{u}} \quad \frac{1}{2}\|\mathbf{u}\|^2 \quad \text{subject to} \quad \|\mathcal{A}_{l_2}((\mathbf{R}\mathbf{u})^H) - \mathbf{B}(l_2,:)\| \leq \gamma \tag{2.10}$$

for $l_2 = 1 \cdots m$. With this approach, the rows of the right factor $\mathbf{R}$ are updated first by iterating over the rows via the index $l_1 = 1 \cdots n$. These updates are followed by updates on the rows of the left $\mathbf{L}$ factor by iterating over the rows via the index $l_2 = 1 \cdots m$. Contrary to the serial problem, these optimizations are conducted on individual vectors $\mathbf{v} \in \mathbb{C}^r$ and $\mathbf{u} \in \mathbb{C}^r$ in parallel because they decouple—i.e., the l_1, l_2th row of $\mathbf{R}, \mathbf{L}$ only involve the l_1, l_2th column/row of the observed data matrix $\mathbf{B}$ and submatrices $\mathcal{A}_{l_1}, \mathcal{A}_{l_2}$ that act on these columns/rows. To simplify notation, we introduced the symbol : to extract the l_1th column, $\mathbf{B}(:,l_1)$, or l_2th row, $\mathbf{B}(l_2,:)$. As before, we allow for the presence of noise by solving the optimizations to within a user-specified ℓ_2-norm tolerance γ.

Because the operations in equations 2.9 and 2.10 decouple, they allow for a parallel implementation that scales well for large-scale industrial 3D seismic problems. However, the decoupled formulation does not include weighting matrices limiting its usefulness for recovery problems at higher frequencies that require weighting matrices. Below we present a novel approach to ameliorate this problem in which we take equations 2.9 and 2.10 as a starting point and pre- and post multiply the data misfit terms by $\mathbf{Q}$ and $\mathbf{W}$ after including the weighting matrices as in equation 2.8. Next, we use the property that for large weights, the matrices $\mathbf{Q}$ and $\mathbf{W}$ nearly commute with the measurement operator $\mathcal{A}$—i.e., we have $\mathbf{Q}\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{X}}\mathbf{W}^{-1}) \approx \mathcal{A}(\bar{\mathbf{X}}\mathbf{W}^{-1})$ and $\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{X}}\mathbf{W}^{-1})\mathbf{W} \approx \mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{X}})$ where $\bar{\mathbf{X}}$ represents the fully sampled data matrix or its factored form. With these approximations, we arrive at



(a)



(b)

Figure 2.6: Alternating minimization and decoupling. *(a)* Solving for the low-rank factor $\mathbf{R}$ by using fixed factor $\mathbf{L}$ and observed data $\mathbf{B}$. *(b)* Solving for the l_1^{th} row of the low-rank factor $\mathbf{R}$ by using rows (in black color) of the fixed factor $\mathbf{L}$ corresponding to the non-zero entries (in black color) of the l_1^{th} column from the observed data $\mathbf{B}$.

the following weighted iterations:

$$\bar{\mathbf{R}}(l_1,:)^H := \arg\min_{\bar{\mathbf{v}}} \quad \frac{1}{2}\|\bar{\mathbf{v}}\|_2^2 \quad \text{subject to} \quad \|\mathcal{A}_{l_1}(\mathbf{Q}^{-1}\bar{\mathbf{L}}\bar{\mathbf{v}}) - \mathbf{B}_R(:,l_1)\| \leq \gamma \quad (2.11)$$

for $l_1 = 1 \cdots n$ and

$$\bar{\mathbf{L}}(l_2,:)^H := \arg\min_{\bar{\mathbf{u}}} \quad \frac{1}{2}\|\bar{\mathbf{u}}\|^2 \quad \text{subject to} \quad \|\mathcal{A}_{l_2}((\bar{\mathbf{R}}\bar{\mathbf{u}})^H\mathbf{W}^{-1}) - \mathbf{B}_L(l_2,:)\| \leq \gamma \quad (2.12)$$

for $l_2 = 1 \cdots m$. In these expressions, we replaced the incomplete data matrix by $\mathbf{B}_R = \mathbf{BW}$ and $\mathbf{B}_L = \mathbf{QB}$, respectively. This means that we pre- and post-multiply the observed monochromatic data matrix $\mathbf{B}$ with $\mathbf{Q}$ and $\mathbf{W}$ before extracting its columns or rows.

The above derivation is only valid if the above approximations involving commutations of the weight matrices with the sampling operator $\mathcal{A}$ are sufficiently accurate. To verify that these approximations are indeed justified, we compare in Figure 2.7 their accuracy by comparing plots of $\mathbf{Q}\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{X}}\mathbf{W}^{-1})$ and $\mathcal{A}(\bar{\mathbf{X}}\mathbf{W}^{-1})$ for two different values of the weights in equations 2.6 and 2.7. As expected, for the small value $w_{1,2} = 0.25$ the weighting matrix $\mathbf{Q}$ does not commute with the sampling matrix (see Figures 2.7a – 2.7c). However, for $w_{1,2} = 0.75$ the approximation is reasonably accurate (see Figures 2.7d – 2.7f). Similarly, in Figure 2.8 we compare plots of $\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{X}}\mathbf{W}^{-1})\mathbf{W}$ and $\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{X}})$ for small and large weights. As before, for smaller weights $w_{1,2} = 0.25$, the weighing matrix $\mathbf{W}$ does not commute with the sampling matrix (see Figures 2.8a – 2.8c). However, for $w_{1,2} = 0.75$ the approximation is again reasonably accurate (see Figures 2.8d – 2.8f). Remember, the weights $w_{1,2}$ reflect confidence we have in the weight matrices and are chosen small when we have confidence that the weighting matrices $\mathbf{Q}$ and $\mathbf{W}$ add information to the recovery. This means we need to select a value for the weights $w_{1,2}$ that balances between how much prior information we want to invoke and how accurate the commutation relations need to be. Choosing small weights goes at the expense of large "commutation" errors while large weights leads to small "commutation" errors but limits the inclusion of the prior informa-

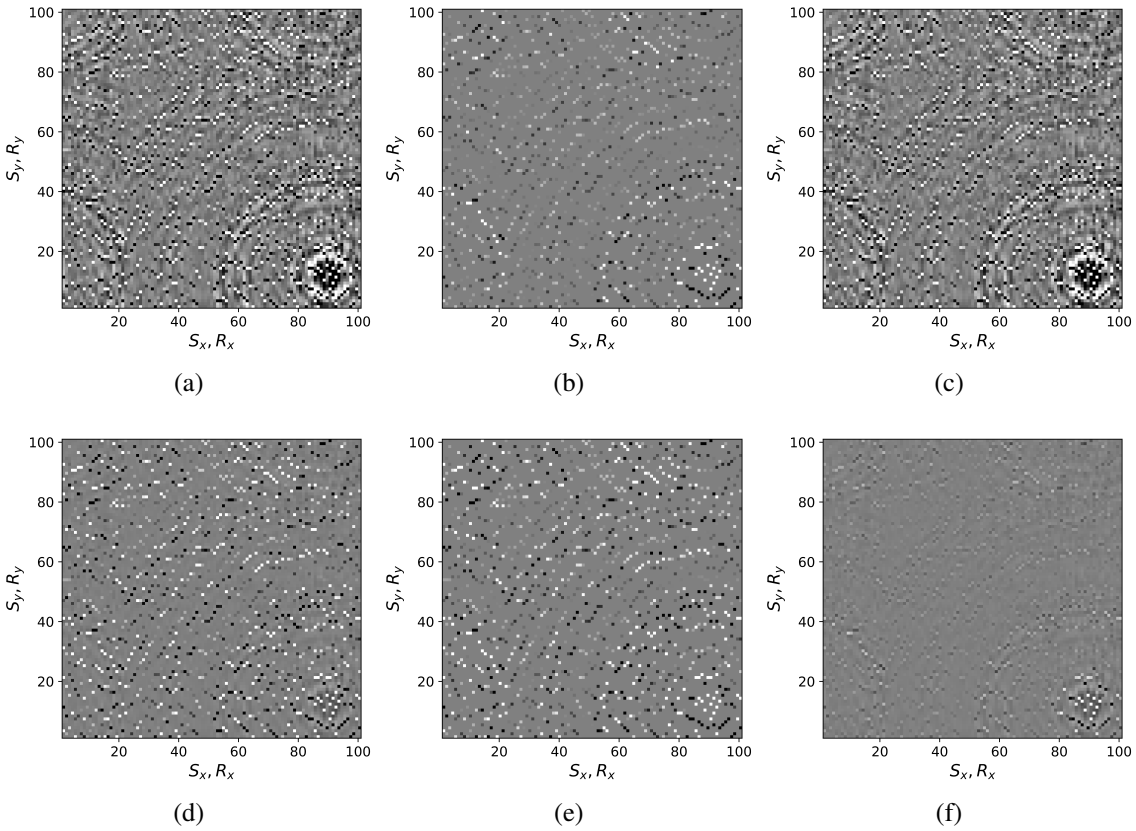


Figure 2.7: Commutation test for small and large weights. *(a)* Subset of 3D frequency slice for $\mathbf{Q}\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{X}}\mathbf{W}^{-1})$ for $w_{1,2} = 0.25$; *(b)* the same but now for $\mathcal{A}(\bar{\mathbf{X}}\mathbf{W}^{-1})$; *(c)* difference plot between *(a)* and *(b)*; *(d)-(f)* the same as *(a)-(c)* but now for $w_{1,2} = 0.75$.

tion via the weights.

Although, the decoupled equations 2.11 and 2.12 can now be parallelized over the rows of the low-rank factors $\bar{\mathbf{R}}$ and $\bar{\mathbf{L}}$, they come at additional computational cost. Unlike sparse observed data collected in the matrix $\mathbf{B}$, the data matrices $\mathbf{B}_R$ and $\mathbf{B}_L$ are dense (have all non-zero entries) because of the multiplications by $\mathbf{W}$ and $\mathbf{Q}$. However, when the weights $w_{1,2}$ are relatively large we observe that both dense matrices $\mathbf{B}_L$, $\mathbf{B}_R$ (Figure 2.9b and 2.9d) can be well approximated by the sparse observed data matrix $\mathbf{B}$ judged by the difference plots in Figure 2.9. With this approximation, equations 2.11 and 2.12 can be solved computationally efficiently.

While the above formulation allows us to carry out weighted factored wavefield recovery in parallelized form, we observed that taking inverses of $\mathbf{Q}$ and $\mathbf{W}$ in the data misfit

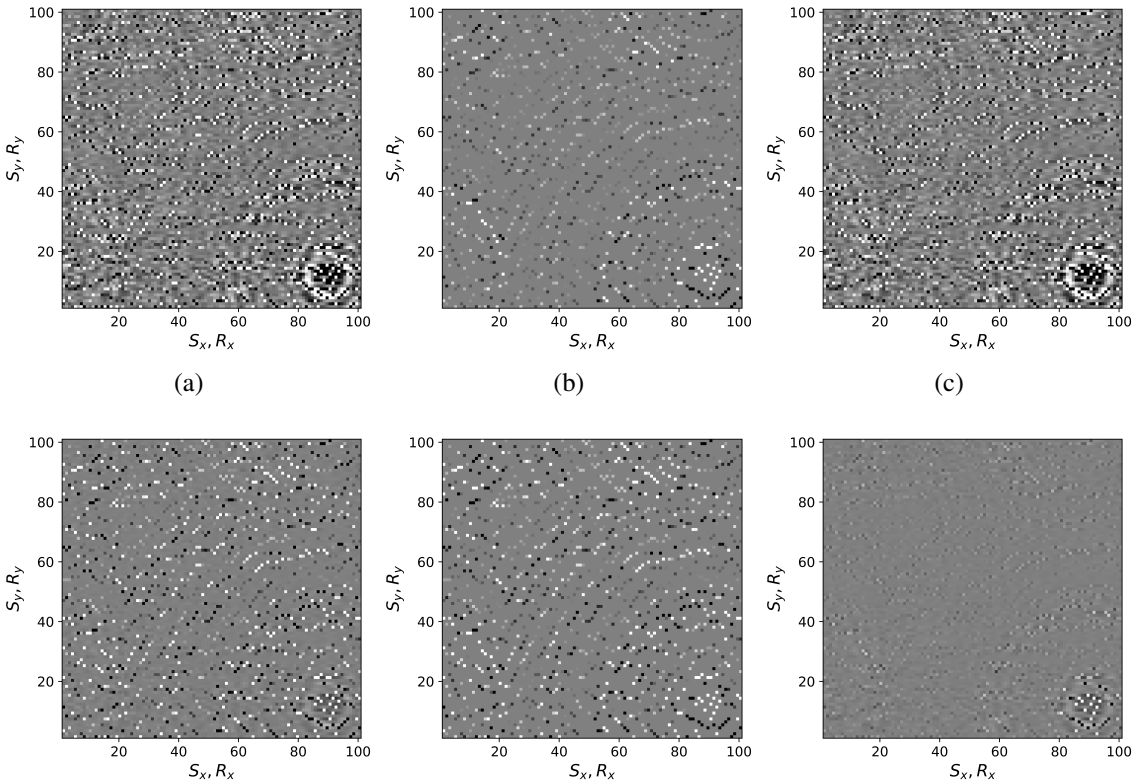


Figure 2.8: Commutation test for small and large weights. *(a)* Subset of 3D frequency slice for $\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{X}}\mathbf{W}^{-1})\mathbf{W}$ for $w_{1,2} = 0.25$; *(b)* the same but now for $\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{X}})$; *(c)* difference plot between *(a)* and *(b)*; *(d)-(f)* the same as *(a)-(c)* but now for $w_{1,2} = 0.75$.

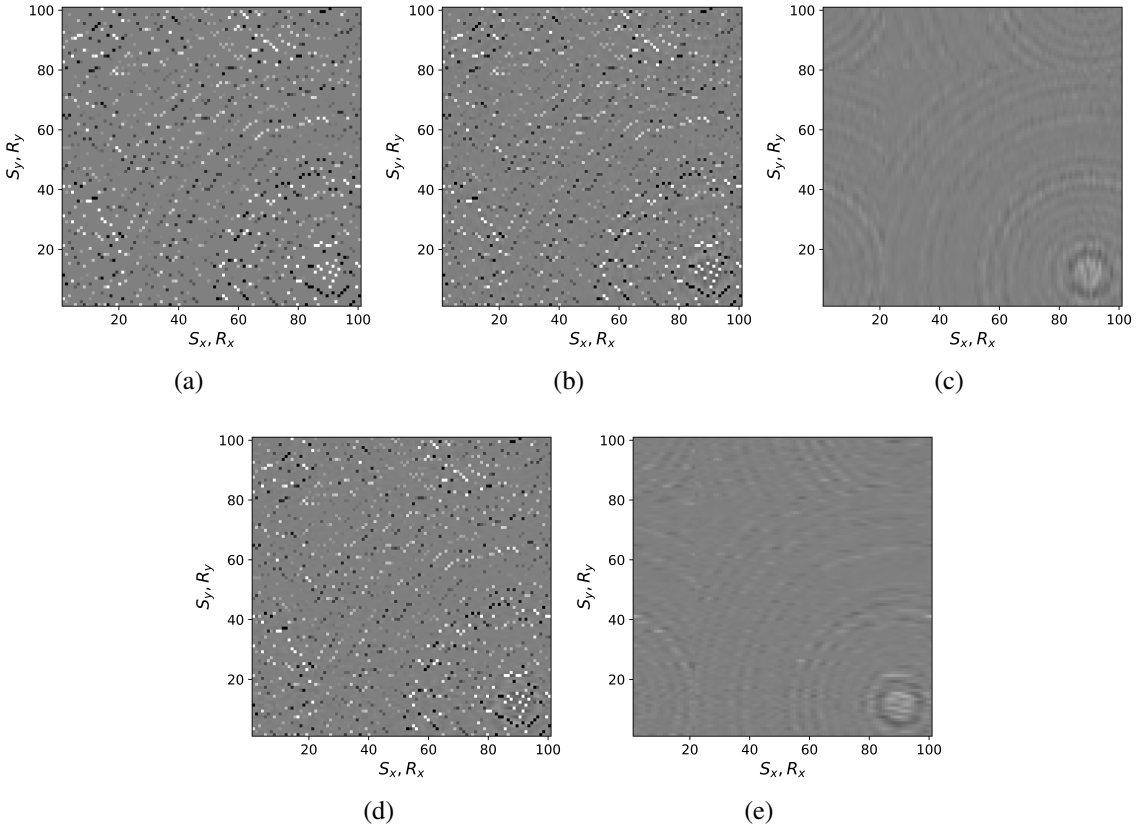


Figure 2.9: Accuracy of sparse approximation for weights $w_{1,2} = 0.75$, *(a)* Subset of 3D frequency slice for sparse observed data $\mathbf{B}$; *(b)* the same but now for the dense matrix $\mathbf{B}_L$; *(c)* difference plot between *(a)* and *(b)*; *(d)* Subset of 3D frequency slice for the dense matrix $\mathbf{B}_R$; *(e)* difference plot between *(a)* and *(d)*.

objective (see equation 2.8) leads to inferior recovery because these involve reciprocals of the weights (see equations 2.6 and 2.7). The value range of these reciprocals is no longer contained to the interval $(0, 1]$ and this can lead to numerical problems during the recovery. To circumvent this issue, we propose an alternative but equivalent form for the weighted formulation with weights defined as

$$\widehat{\mathbf{Q}} = \mathbf{U}\mathbf{U}^H + w_1\mathbf{U}^\perp\mathbf{U}^{\perp^H} = w_1\mathbf{Q}^{-1}, \tag{2.13}$$

$$\widehat{\mathbf{W}} = \mathbf{V}\mathbf{V}^H + w_2\mathbf{V}^\perp\mathbf{V}^{\perp^H} = w_2\mathbf{W}^{-1} \tag{2.14}$$

With these alternative definitions, we can as before approximate $\widehat{\mathbf{Q}}^{-1}\mathcal{A}(\widehat{\mathbf{Q}}\bar{\mathbf{X}}\widehat{\mathbf{W}})$ by $\mathcal{A}(\bar{\mathbf{X}}\widehat{\mathbf{W}})$ and $\mathcal{A}(\widehat{\mathbf{Q}}\bar{\mathbf{X}}\widehat{\mathbf{W}})\widehat{\mathbf{W}}^{-1}$ by $\mathcal{A}(\widehat{\mathbf{Q}}\bar{\mathbf{X}})$, yielding the following decoupled parallellizable equations for the factors

$$\begin{aligned} \bar{\mathbf{R}}(l_1,:)^H := \quad & \arg\min_{\bar{\mathbf{v}}} \quad \frac{1}{2}\|\bar{\mathbf{v}}\|^2 \\ & \text{subject to} \\ & \|\mathcal{A}_{l_1}(\widehat{\mathbf{Q}}\bar{\mathbf{L}}\bar{\mathbf{v}}) - w_1w_2\mathbf{B}(:,l_1)\| \leq w_1w_2\gamma \end{aligned} \tag{2.15}$$

for $l_1 = 1 \cdots n$ and

$$\begin{aligned} \bar{\mathbf{L}}(l_2,:)^H := \quad & \arg\min_{\bar{\mathbf{u}}} \quad \frac{1}{2}\|\bar{\mathbf{u}}\|^2 \\ & \text{subject to} \\ & \|\mathcal{A}_{l_2}((\bar{\mathbf{R}}\bar{\mathbf{u}})^H\widehat{\mathbf{W}}) - w_1w_2\mathbf{B}(l_2,:)\| \leq w_1w_2\gamma \end{aligned} \tag{2.16}$$

for $l_2 = 1 \cdots m$. Equations 2.15 and 2.16 form the basis for our recovery approach summarized in Algorithm 2.1 below, which corresponds to

$$\min_{\bar{\mathbf{X}}} \quad \|\bar{\mathbf{X}}\|_* \quad \text{subject to} \quad \|\mathcal{A}(\widehat{\mathbf{Q}}\bar{\mathbf{X}}\widehat{\mathbf{W}}) - w_1w_2\mathbf{B}\|_F \leq w_1w_2\epsilon, \tag{2.17}$$

which is equivalent to equation 2.5 as we show in Appendix A.

Algorithm 2.1 Weighted minimization via Alternating minimization.

Input: Observed Data $\mathbf{B}$, rank r, acquisition mask $\mathcal{A}$, priors $\widehat{\mathbf{Q}}$, $\widehat{\mathbf{W}}$, initial guess, $\bar{\mathbf{L}}^{(0)}$

1. **for** $k = 0,\, 1,\, 2, \cdots,\, N-1$ `//solve for rows of` $\bar{\mathbf{R}}$ `&` $\bar{\mathbf{L}}$ `in parallel`
2. $\left(\bar{\mathbf{R}}^{(k+1)}(l_1,:)\right)^H := \arg\min_{\bar{\mathbf{v}}} \quad \frac{1}{2}\left\|\bar{\mathbf{v}}\right\|^2 \quad \text{s. t.} \quad \|\mathcal{A}_{l_1}(\widehat{\mathbf{Q}}\bar{\mathbf{L}}^{(k)}\bar{\mathbf{v}}) - w_1 w_2 \mathbf{B}(:,l_1)\| \leq w_1 w_2 \gamma$
3. $\left(\bar{\mathbf{L}}^{(k+1)}(l_2,:)\right)^H := \arg\min_{\bar{\mathbf{u}}} \quad \frac{1}{2}\left\|\bar{\mathbf{u}}\right\|^2 \quad \text{s. t.} \quad \|\mathcal{A}_{l_2}((\bar{\mathbf{R}}^{(k+1)}\bar{\mathbf{u}})^H\widehat{\mathbf{W}}) - w_1 w_2 \mathbf{B}(l_2,:)\| \leq w_1 w_2 \gamma$
4. **end for**
5. $\mathbf{L} = \frac{1}{w_1}\widehat{\mathbf{Q}}\bar{\mathbf{L}}$
6. $\mathbf{R} = \frac{1}{w_2}\widehat{\mathbf{W}}\bar{\mathbf{R}}$

Output: Recovered wavefield in factored form $\{\mathbf{L},\, \mathbf{R}\}$.

In Algorithm 2.1, **Line 2** corresponds to solving for each row of the low-rank factor $\bar{\mathbf{R}}^{(k+1)}$ at the $(k+1)^{th}$ iteration using the estimate of low-rank factor $\bar{\mathbf{L}}^{(k)}$ from the $(k)^{th}$ iteration. Similarly, **Line 3** corresponds to solving for each row of the low-rank factor $\bar{\mathbf{L}}^{(k+1)}$ at the $(k+1)^{th}$ iteration using the estimate of low-rank factor $\bar{\mathbf{R}}^{(k+1)}$. Finally, **Lines 5** and **6** correspond to retrieving the low-rank factors $\mathbf{L}$ and $\mathbf{R}$ from $\bar{\mathbf{L}}$ and $\bar{\mathbf{R}}$, respectively.

2.5 Case studies

We now conduct a series of experiments to evaluate the accuracy of the proposed weighted wavefield reconstruction methodology. In all cases, we have access to the ground truth fully sampled data. This allows us to assess the accuracy by means of visual inspection and S/R's (Signal to noise ratio). Our examples include the seismic line from the Gulf of Suez we discussed earlier and a complex full-azimuth synthetic 3D dataset.

2.5.1 Gulf of Suez field data: 2D example

To evaluate the performance of our recursively weighted wavefield recovery method on field data, we conduct an experiment on a 2D line from the Gulf of Suez. The fully sampled split-spread dataset consists of 1024 time samples, acquired with 354 sources and 354

receivers. Time is sampled at $0.004\,\mathrm{s}$. The source-receiver spacing is $25\,\mathrm{m}$. To test our algorithm, we reconstruct this 2D line from randomly subsampled traces, which we obtain by removing 75% of the sources via optimal jittered subsampling (Herrmann and Hennenfent 2008a).

We assess the performance of recursively weighted matrix factorization by comparing wavefield recovery with and without weighting as a function of the angular frequency. To avoid the impact of noise at the low frequencies, we start the recovery at $7.0\,\mathrm{Hz}$. Since this is a small problem, we reconstruct the frequency slices by performing 150 iterations of the SPG-LR algorithm (Van Den Berg and Friedlander 2008; Aravkin, Kumar, Mansour, Recht, and Herrmann 2014) for each frequency. We compare wavefield reconstructions with and without weights the results of which are summarized in Figure 2.10. From these results we can see that above 17Hz, the wavefield reconstruction clearly benefits from including weights for reconstructions carried out with the same number of iterations but without weighting.

For comparison purposes, we also reconstruct missing data using the conventional matrix factorization method. For fairness of comparison, we once again use 150 iterations of SPG-LR algorithm for each frequency slice. As expected other than few lower frequency slices where conventional method performs well, we get improvements in signal to noise ratio (Figure 2.10) across all other frequency slices with our recursively weighted approach.

Figures 2.11c and 2.12c include the shallow and deeper parts of a reconstructed common receiver gather extracted from results based on the conventional method yielding $S/R = 6.9\,\mathrm{dB}$. In the data residual plots (Figures 2.11d and 2.12d), we observe signal leakage and noise due to reconstruction artifact in both shallow and deeper parts. By signal leakage we mean that there are coherent events in the data residual plot indicating incomplete reconstruction of data. Also, at far offsets we observe signal leakage in the difference plot. Far offset data is important for FWI (Full Waveform Inversion) purposes since it contains turning waves. On the other hand, we observe better reconstructed data in the

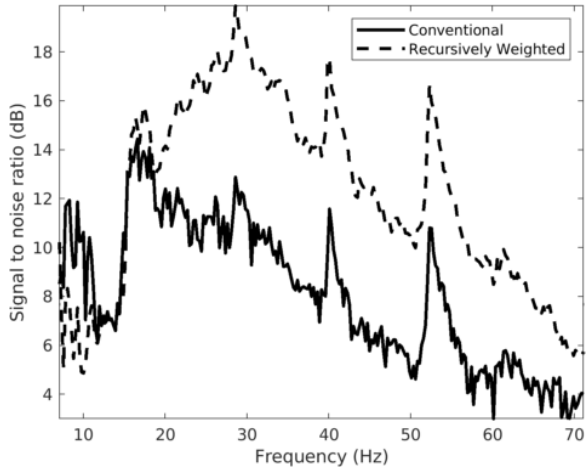


Figure 2.10: *(a)* Signal to noise ratio comparison of conventional (solid black line) and recursively weighted method (dashed black line) with ($w_{1,2} = 0.75$) for all frequencies

common receiver gather (Figures 2.11e and 2.12e) extracted from reconstructed data using the recursively weighted approach with improved S/R of $11.7\,\mathrm{dB}$. Its corresponding data residual plot (Figures 2.11f and 2.12f) shows less signal leakage in comparison to its conventional counterpart. Even at far offsets, we observe better reconstruction of signal.

2.5.2 Synthetic Compass model data: 3D example

In 2D seismic surveys, receivers only measure wavefields traveling in the vertical plane along sources and receivers. Therefore, we fail to capture reflections out of the 2D source-receiver plane. This lack of recording of out of plane scattering ultimately affects the quality of the subsurface image, especially in regions where there is strong lateral heterogeneity. To capture 3D effects most of the seismic exploration surveys are 3D nowadays during which sources and receivers are spread along the surface rather than confined to a single line. To evaluate the performance of our recursively weighted low-rank matrix factorization methodology in this more challenging 3D setting, we consider synthetic 3D data simulated on the Compass model (E. Jones, A. Edgar, I. Selvage, and Crook 2012). We choose this model because it contains velocity kickbacks, strong reflectors, and small wavelength details constrained by real well-log data collected in the North Sea. Because

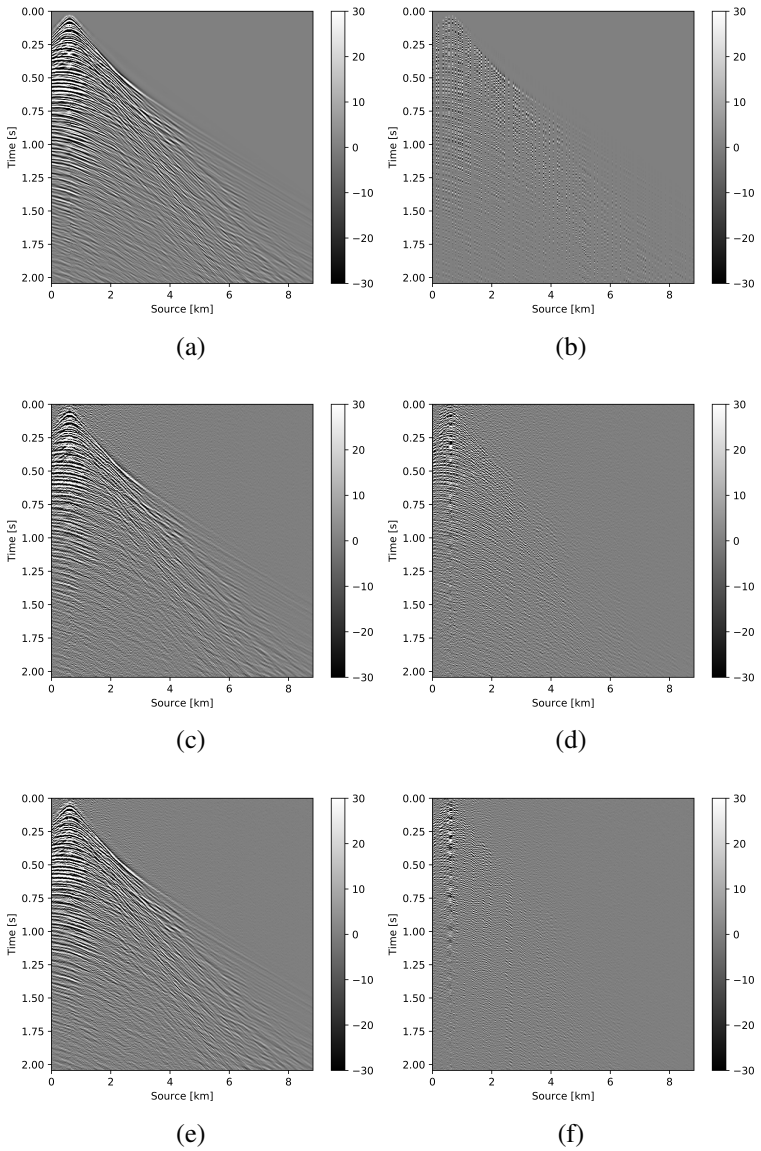
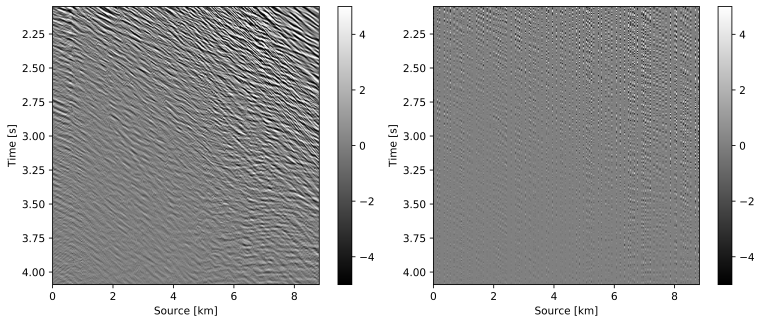
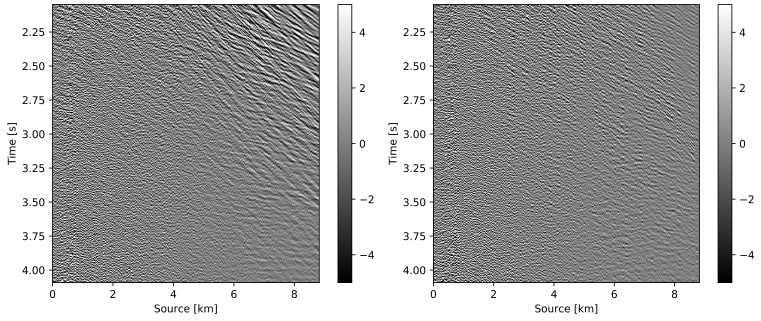


Figure 2.11: Wavefield reconstruction in common receiver gather domain in the shallow part. *(a)* True data, *(b)* Observed data with 75% missing sources. *(c)* Reconstructed data using the conventional method with $S/R = 6.9\,\mathrm{dB}$ and *(d)* corresponding difference with respect to the true data. *(e)* Reconstructed data using the recursively weighted method with $S/R = 11.7\,\mathrm{dB}$ and *(f)* corresponding difference with respect to the true data.



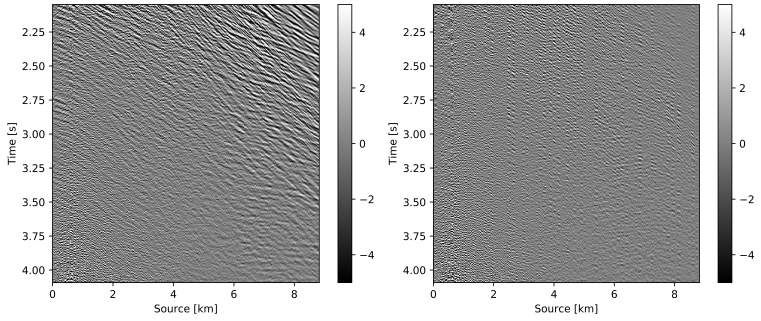
(a)

(b)



(c)

(d)



(e)

(f)

Figure 2.12: Wavefield reconstruction in common receiver gather domain in the deeper part. *(a)* True data, *(b)* Observed data with 75% missing sources. *(c)* Reconstructed data using the conventional method and *(d)* corresponding difference with respect to the true data. *(e)* Reconstructed data using the recursively weighted method and *(f)* corresponding difference with respect to the true data.

of the complexity of this dataset, which mimics marine acquisition with a towed array, we face similar challenges in wavefield reconstruction as we would face dealing with real 3D field data. The authors (Da Silva and Herrmann 2015) also used this 3D dataset to evaluate their tensor-based wavefield reconstruction algorithm based on the Hierarchical Tucker decompositions.

For this experiment, we use a subset of the total data volume of $501 \times 201 \times 201 \times 41 \times 41$ gridpoints—i.e., $n_t \times n_{rx} \times n_{ry} \times n_{sx} \times n_{sy}$ along the time, receiver x, receiver y, source x, and source y directions. Here, n_t is the number of samples along time, n_{rx}, n_{ry} are number of receivers along x and y directions respectively and n_{sx}, n_{sy} are number of sources along x and y directions respectively. In both spatial directions, the spacing between the adjacent sources is $150.0\,\mathrm{m}$ and $25.0\,\mathrm{m}$ between adjacent receivers. The sampling interval along time is $0.01\,\mathrm{s}$. To get the subsampled data, we remove 75% of the receivers from jittered locations (Herrmann and Hennenfent 2008a). We use this incomplete data as input to our recursively weighted wavefield reconstruction scheme.

Before proceeding further, let us first briefly discuss the organization of the data in which we will carry out the wavefield reconstructions. While we could in principle transform the data into the midpoint offset domain as in the 2D case, we follow (Da Silva and Herrmann 2015) and (Demanet 2006) and exploit the fact that monochromatic 3D frequency slices rearranged along the x and y-coordinates for sources and receivers can be well approximated by a low-rank factorization. In this rearrangement the data is organized as a matrix with S_x, R_x and S_y, R_y coupled along the columns and rows respectively unfolded along is coordinate directions. Here $S_{x,y}$ and $R_{x,y}$ are the source and receiver coordinates along the x and y directions. After rearrangement in this non-canonical form, the frequency slices are low-rank while data with randomly missing receivers is not (juxtapose $10.0\,\mathrm{Hz}$ frequency slices in Figures 2.13c and 2.13d and the singular value plots in Figures 2.14a and 2.14b). We choose $10\,\mathrm{Hz}$ frequency slice as the changes in the rate of decay of singular values upon sampling at lower frequencies is more prominent in com-

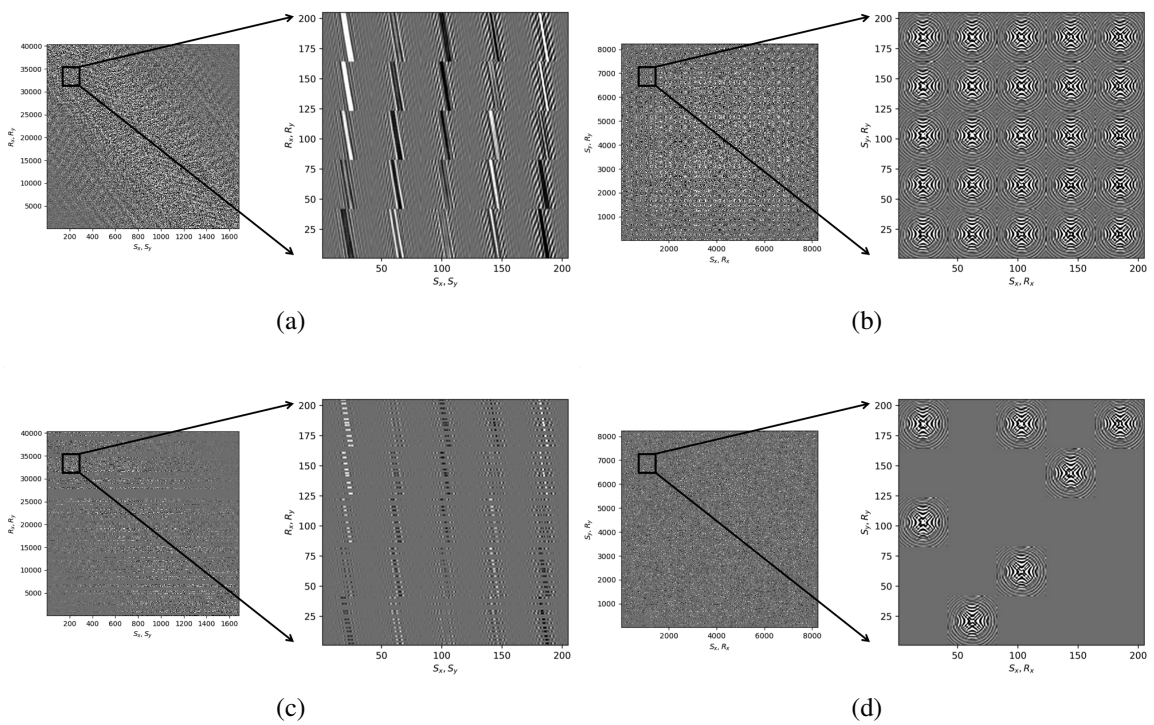
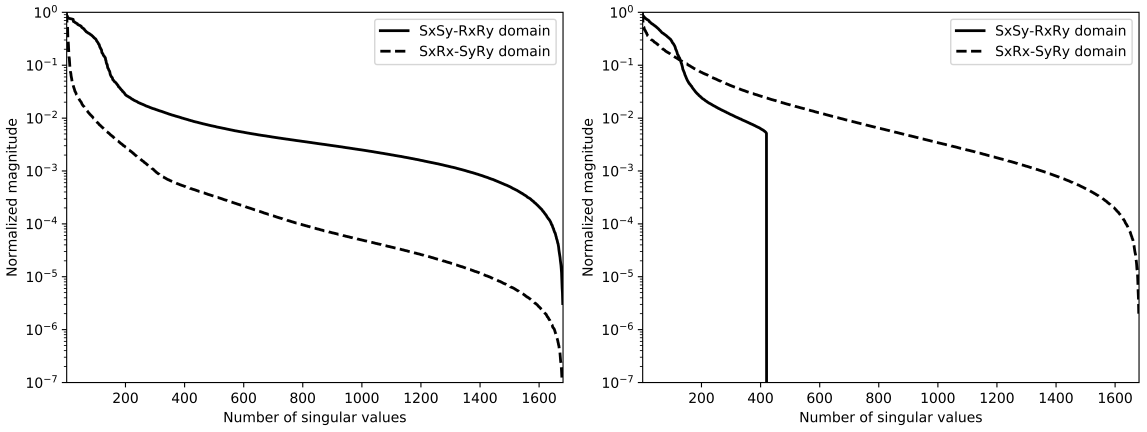


Figure 2.13: $10.0\,\mathrm{Hz}$ Frequency slice from 3D data: *(a)* True and *(c)* observed data in S_x, S_y domain with 75% missing receivers. *(b)* True and *(d)* observed data in S_x, R_x domain with 75% missing receivers. Figures in left column show full data and in right column show data zoomed in the small black box.

parison to changes we observe at higher frequencies. This frequency choice allows us to better demonstrate the reasoning behind choosing S_x, R_x and S_y, R_y domain for reconstruction. In the canonical organization, missing receivers leads to missing rows and this decreases the rank (cf. solid lines in Figure 2.14) in the non-canonical rearrangements the rank increases (cf. dashed lined in Figure 2.14). The sudden drop in the singular values in the canonical arrangement is a direct consequence of the fact that removing complete rows or columns decreases the rank. From the behavior of the singular values before and after removal of the receivers, it is clear that the simple rearrangement of the data in the non-canonical organization can serve as the transform domain in which to recover that data via weighted low-rank factorization.

As before, we now perform the full-azimuth 3D wavefield reconstruction for each fre-



(a) (b)

Figure 2.14: Singular values decay comparison for *(a)* fully sampled and *(b)* subsampled data with 75% missing receivers in S_x, S_y domain (solid black line) and S_x, R_x (dashed black line) domain for $10.0\,\mathrm{Hz}$ frequency slice

quency slice using our proposed recursively weighted low-rank matrix factorization approach. Since this is a relatively large problem, we employ the parallel framework presented in the previous section (Algorithm 2.1) for 4 alternations with 40 iterations of SPG-ℓ_2 per frequency slice. We choose these values because for a given rank parameter we observed better continuity of signals and lesser noise in the reconstructed data. In addition to setting the number of alternations, i.e. switching between Equations 2.15 and 2.16, the algorithm needs us to specify the rank of the factorization and the weights. Based on tests performed using different rank values and weights, we selected a rank of $r = 228$ and a value for the weights of $w_{1,2} = 0.75$, because they provide a good balance between quality of reconstructed data (in terms of continuity of events, lesser noise) and computational time.

To avoid noise at the very low frequencies observed due to simulation artifacts, we start our recursively weighted from $4.4\,\mathrm{Hz}$. For comparison, we also use conventional matrix completion for wavefield reconstruction. Here, we use the same number of alternations and SPG-ℓ_2 iterations as before. We also use same rank of 228. For visualization purpose we show results in a common shot gather (Figure 2.15a) extracted from $15\,\mathrm{Hz}$ frequency slice.

Here we choose higher frequency of $15\,\mathrm{Hz}$ instead of $10\,\mathrm{Hz}$ to show how the recursively weighted method is able to give better reconstruction at high frequency in comparison to reconstructed data obtained from the conventional method. In Figure 2.15b we show subsampled shot gather with 75% missing receivers. Using the conventional method we get S/R of $17.7\,\mathrm{dB}$ for the reconstructed data at $15.0\,\mathrm{Hz}$ (Figure 2.15d). Whereas, with the recursively weighted method we get improved S/R of $19.9\,\mathrm{dB}$ (Figure 2.15f). We also observe less leakage of signal and less noise in the residual plots for the data reconstructed using recursively weighted method (Figure 2.15g) in comparison to the data reconstructed using the conventional method (Figure 2.15e). From Figure 2.18a, we also observe improvement in the S/R of reconstructed data for all the frequencies with the recursively weighted method (dashed black line in Figure 2.18a) in comparison to its conventional counterpart (solid black line in Figure 2.18a). In Figures 2.16 and 2.17 we also show comparison of the recursively weighted and conventional method in time domain common shot gather at earlier and later arrivals respectively. In Figure 2.16, we also show comparison of a time slice at $1.6\,\mathrm{s}$ extracted from a 3D common shot gather. Figure 2.16a shows two common shot gathers extracted from the true data along x and y directions along with a time-slice on top left corner. Figure 2.16b shows the corresponding observed data with missing receivers. We observe improved reconstruction of signals in the common shot gather (with $S/R = 17.8\,\mathrm{dB}$) reconstructed from recursively weighted method (Figure 2.16f) in comparison to the reconstructed data from the conventional method (Figure 2.16d) with S/R of $15.3\,\mathrm{dB}$. Even in the residual plots we observe less leakage of signal with the recursively weighted method (Figure 2.16g) in comparison to its conventional counterpart (Figure 2.16e). In Figures 2.17a and 2.17b, we show the same common shot gather at later time along x and y directions extracted from true and observed data respectively. We observe noise in the data and corresponding residual (Figures 2.17c and 2.17d) reconstructed from the conventional method. Whereas, we observe better reconstruction and less noise in the data reconstructed (Figures 2.17e and 2.17f) from the recursively weighted method.

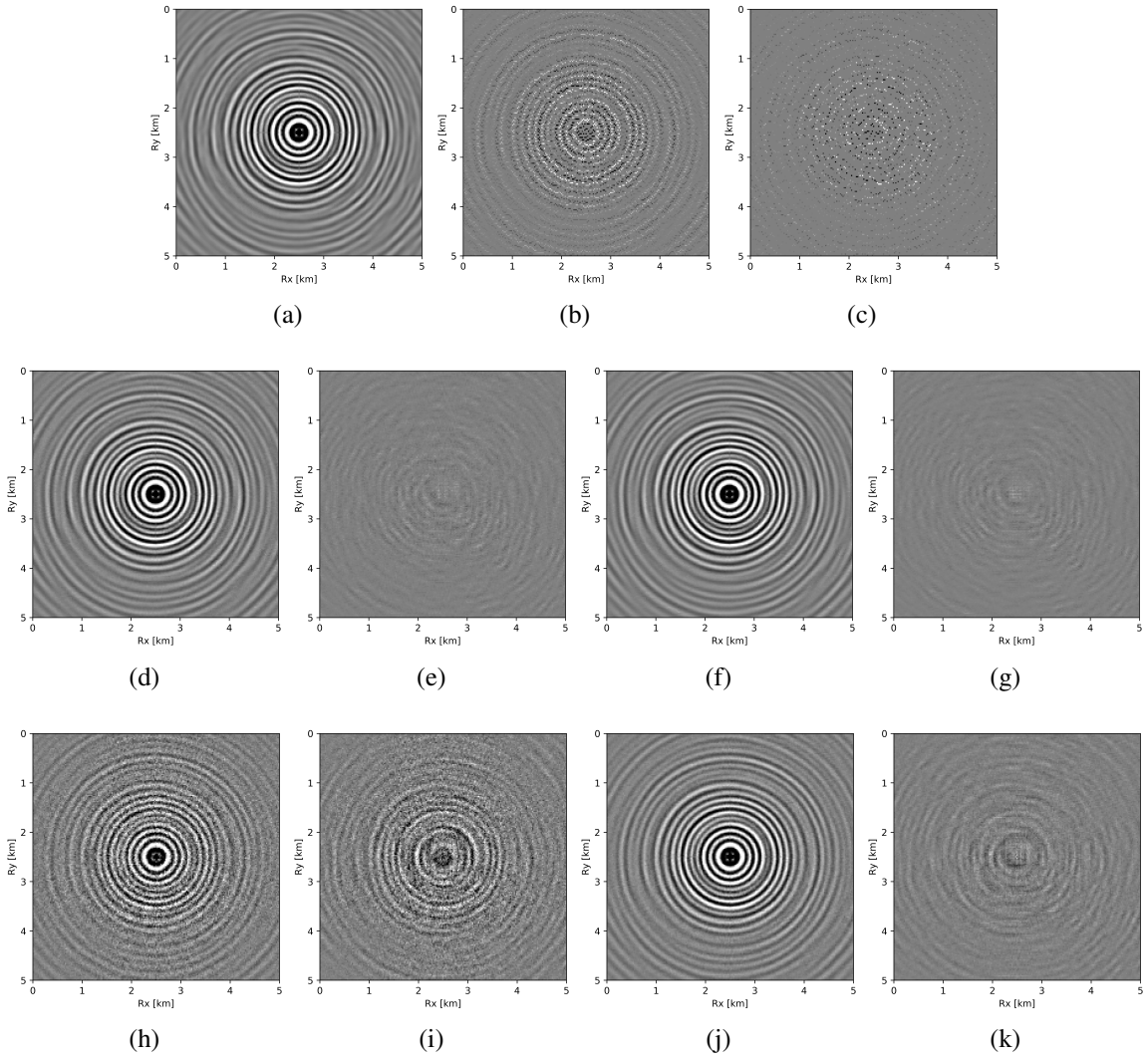


Figure 2.15: Full azimuth wavefield reconstruction comparison for $15.0\,\mathrm{Hz}$ frequency slice in common shot domain. *(a)* True frequency slice. Subsampled frequency slice with *(b)* 75% missing receivers and *(c)* 90% missing receivers. Middle row represents reconstruction using observed data with 75% missing receivers. *(d)* Reconstructed data using conventional method with $S/R = 17.7\,\mathrm{dB}$ and *(e)* corresponding data residual with respect to true data. *(f)* Reconstructed data using recursively weighted method with $S/R = 19.9\,\mathrm{dB}$ and *(g)* corresponding data residual with respect to true data. Last row represents reconstruction using observed data with 90% missing receivers. *(h)* Reconstructed data using conventional method with $S/R = 3.7\,\mathrm{dB}$ and *(i)* corresponding data residual with respect to true data. *(j)* Reconstructed data using recursively weighted method with $S/R = 12.5\,\mathrm{dB}$ and *(k)* corresponding data residual with respect to true data.

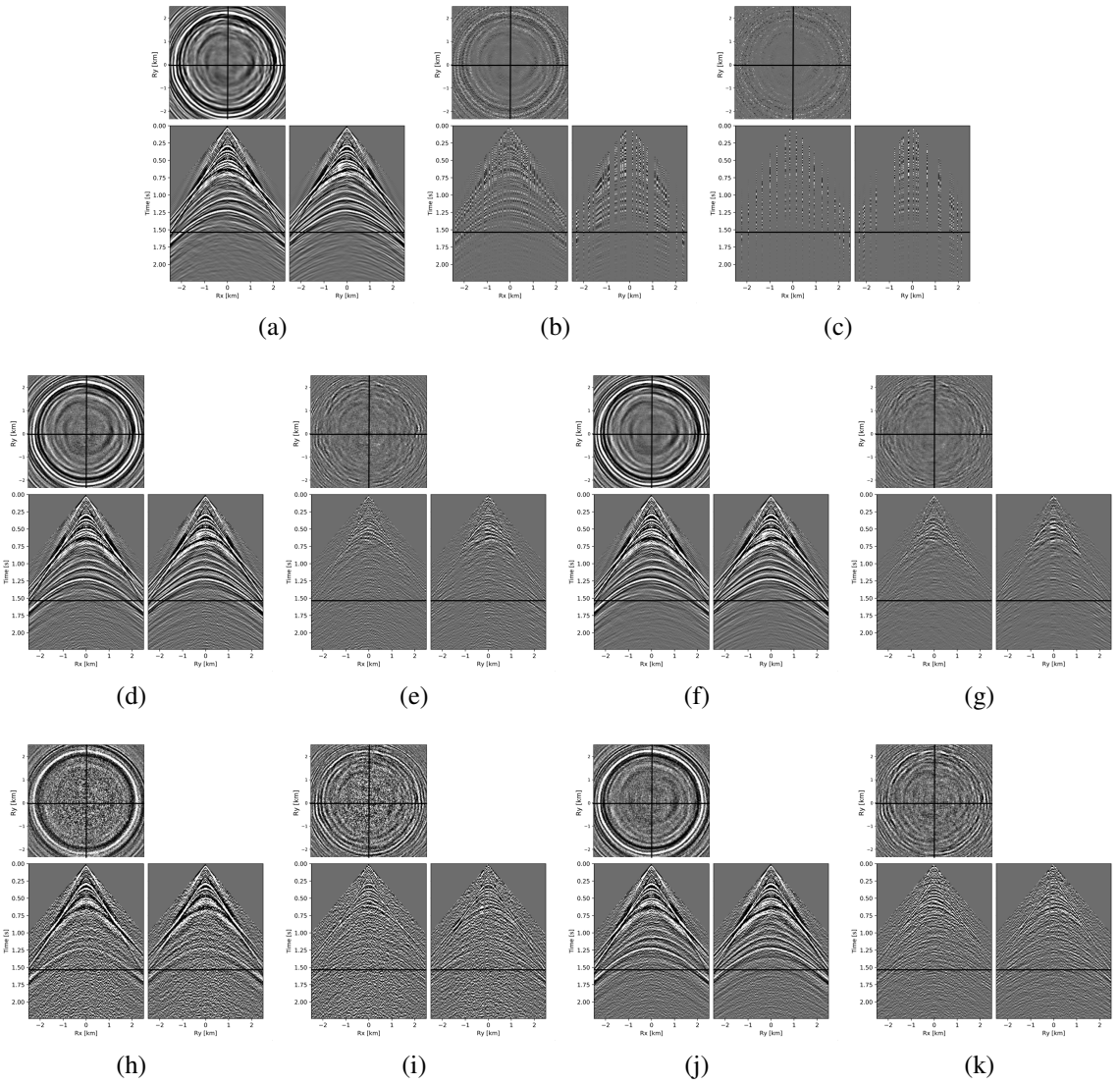


Figure 2.16: Full azimuth wavefield reconstruction in time domain for a common shot gather along with time slice at $1.6\,\mathrm{s}$. *(a)* True data. Subsampled data with *(b)* 75% missing receivers and *(c)* 90% missing receivers. Middle row represents reconstruction using observed data with 75% missing receivers. *(d)* Reconstructed data using the conventional method with $S/R = 15.3\,\mathrm{dB}$ and *(e)* corresponding data residual with respect to the true data. *(f)* Reconstructed data using the recursively weighted method with $S/R = 17.8\,\mathrm{dB}$ and *(g)* corresponding data residual with respect to the true data. Last row represents reconstruction using observed data with 90% missing receivers. *(h)* Reconstructed data using the conventional method with $S/R = 3\,\mathrm{dB}$ and *(i)* corresponding data residual with respect to the true data. *(j)* Reconstructed data using the recursively weighted method with $S/R = 10.2\,\mathrm{dB}$ and *(k)* corresponding data residual with respect to the true data.

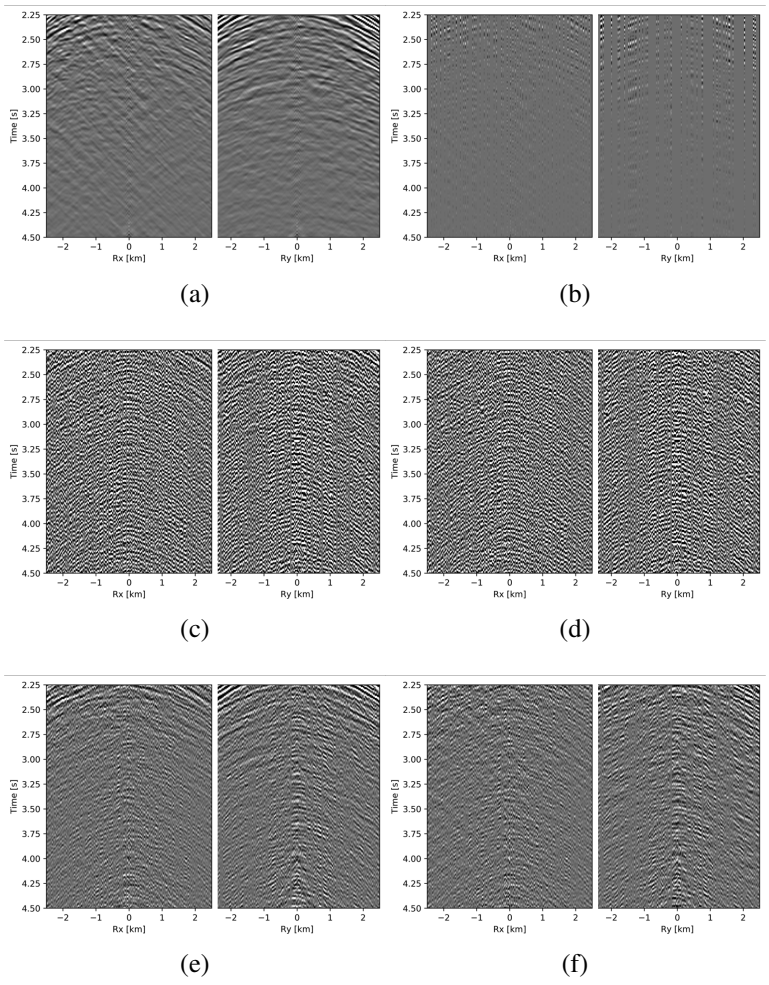
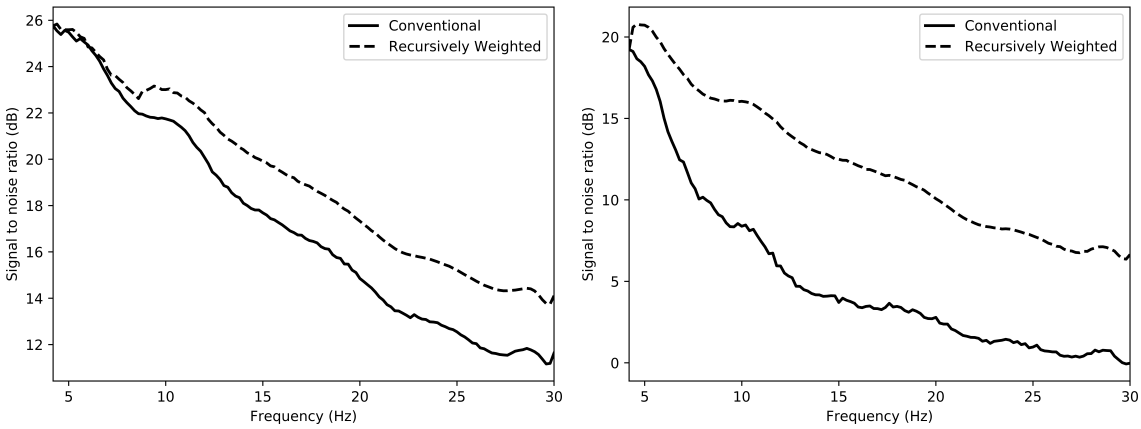


Figure 2.17: Full azimuth wavefield reconstruction in time domain for a common shot gather (deeper section). *(a)* True data. *(b)* Subsampled data with 75% missing receivers. *(c)* Reconstructed data using the conventional method and *(d)* corresponding difference with respect to the true data. *(e)* Reconstructed data using the recursively weighted method and *(f)* corresponding difference with respect to the true data.



(a) (b)

Figure 2.18: Signal to noise ratio comparison of conventional (solid black line) and recursively weighted ($w_{1,2} = 0.75$) method (dashed black line) for all the frequencies for *(a)* 75% and *(b)* 90% missing receiver scenarios.

2.5.3 BG synthetic 3D data with 90% missing receivers

Next we test the ability of the recursively weighted method with a reduced number of samples. We subsample the BG synthetic 3D data by 90% using jitter subsampling, i.e. we use only 10% of receivers for wavefield reconstruction. We use 4 alternations and 40 inner iterations of SPG-ℓ_2 in each alternation per frequency slice for both conventional and recursively weighted method. We use rank parameter of 228 for all the frequency slices. Like before we arrive at these values by inspecting the quality of reconstructed data based on the continuity of signal and attenuated noise in the reconstructed data. To avoid noise at lower frequencies we start recursively weighted method from $4.4\,\mathrm{Hz}$. As evident from the signal to noise ratio plot (Figure 2.18b), we observe improvement in data reconstruction quality across all the frequency slices using the recursively weighted method (dashed black line in Figure 2.18b) in comparison to its conventional counterpart (solid black line in Figure 2.18b). In a common shot gather extracted from a frequency slice at $15\,\mathrm{Hz}$, we observe better continuity and less noise in the reconstructed wavefield (Figure 2.15j) in comparison to the reconstruction obtained from its conventional counterpart (Figure 2.15h). We observe more leakage of signal in the data residual with the conventional method (Figure 2.15i) in

comparison to the data residual obtained from recursively weighted method (Figure 2.15k). In Figure 2.16 we compare data reconstruction in time domain using conventional (Figures 2.16h and 2.16i) and recursively weighted method (Figures 2.16j and 2.16k). We again observe better data reconstruction and reduced data residual with the recursively weighted method in comparison to reconstruction obtained from the conventional method.

2.6 Discussion

From the above case studies it is clear that recursively weighted low-rank matrix completion provides several benefits over conventional method. The reconstructed data preserves the signals and continuity of events even at high frequencies and at deeper sections where amplitude is very weak. We arrive at these results by exploiting the fact that fully sampled seismic data can be approximated by a low-rank matrix in some transform domain and randomized/jittered sampling degrades (or negatively affects) this low-rank property. We also exploit the fact that there is some degree of similarity between the subspaces of adjacent frequency slices of the seismic data.

Our method, which uses recursively weighted low-rank matrix completion, outperforms its conventional counterpart in terms of quality of the reconstructed data specially at higher frequencies. At higher frequencies conventional low-rank matrix completion performs poorly because these increasingly complex matrices eventually violate our low-rank assumption. As we mentioned earlier, good quality high frequency content in the data is important for high-resolution imaging of earth's subsurface and also for inversion of earth's physical parameters with fine details.

Weighted matrix completion was first introduced by (Aravkin, Kumar, Mansour, Recht, and Herrmann 2014) to improve the seismic data reconstruction quality of the conventional matrix completion framework. Here we have exploited the potential of weighted method by recursively reconstructing data from low to high frequencies. Also, we have made the original weighted method formulation proposed by (Aravkin, Kumar, Mansour, Recht, and

Herrmann 2014) computationally efficient by switching the weights from objective to data misfit constraint function.

Similarity between the adjacent frequency slices and appropriate choice of weights play key role in the success of recursively weighted method. Conventional method is easily parallelized over frequencies making it computationally very efficient. Whereas, the interdependence between frequency slices in the recursively weighted method does not allow us to parallelize the recursively weighted method over frequencies. This poses computational challenge especially for large scale 3D datasets. By using the strategies of alternating minimization and decoupling we have made the recursively weighted method computationally efficient for higher weights. Depending on the availability of computational resources, the recursively weighted method can be efficiently applied to large scale 3D datasets. Our parallel weighted framework partially exploits the benefits of weighted low-rank matrix factorization since it can be parallelized only for higher weights. Despite this, our numerical experiments demonstrate improvements in the reconstructed data quality across all the frequencies for 3D seismic data generated on a geologically complicated velocity model resembling part of true Earth's subsurface. To exploit the full benefits of weighted method for large 3D datasets, our future work will focus on extending this methodology to exploit parallelism even for smaller weight values.

By directly using the low-rank factors from a subsequent previous frequency slice to calculate the weight matrix, our recursively weighted framework avoids taking SVDs of the complete dataset to calculate its row and column subspaces. Our SVD free parallel weighted framework can be applied to industry scale large seismic datasets. With the advent of cloud computing there are plenty of computational resources available. But the main issue is to use these resources to optimize both the turnaround time and the budget. Therefore, next steps will be to re-engineer the weighted framework to efficiently use the cloud based computational resources using the ideas of serverless computing. For example, (Witte, Louboutin, Modzelewski, Jones, Selvage, and Herrmann 2019) designed serverless

computing architecture to perform large scale 3D seismic imaging.

Both the datasets used for experiments have sources and receivers on a uniform grid but in reality this is not the situation. Because of environmental and operational constraints sources and receivers are often shifted from the uniform grid. If we do not take into account of this shift in our reconstruction framework then we can encounter poor performance of the reconstruction framework. By incorporating an extra operator (López, Yilmaz, and Herrmann 2016) corresponding to these shifts from the uniform grid we can apply our weighted framework to field data recorded on a non-uniform grid.

2.7 Conclusions

While successful at the low to midrange frequencies, wavefield reconstruction based on matrix factorization fails at the higher frequency where seismic data is no longer low rank. We overcome this problem by exploiting similarities between low-rank factorizations of adjacent monochromatic frequency slices organized in a form that reveals the underlying low-rank structure of, the for budgetary and physical reasons not accessible, fully sampled data. During matrix factorization these similarities take the form of alignment of the subspaces in which the low-rank factors live. By introducing weight matrices that project these factors onto the nearby subspace of the adjacent frequency, the performance of the low-rank matrix factorization improves if this beneficial feature is used recursively starting at the lower frequencies. However, turning this approach into an algorithm that scales to industry-scale wavefield reconstruction problems for full-azimuth data requires a number of additional important steps. First, we need to avoid costly projections onto weighted constraints. We accomplish this by moving the weights to the data misfit. This simple reformulation results in an equivalent formulation, which is computationally significantly faster. Secondly, while the recursively applied weighting matrices improve the performance for the high frequencies, the introduction of these matrices does not allow for a row-by-row and column-by-column parallelization of the alternating minimization procedure we em-

ploy to carry out the matrix factorizations on which our low-rank wavefield reconstruction is based. We overcome this problem by balancing the emphasis we put on information from adjacent frequency slices with our ability to decouple the operations so that the algorithm can be parallelized. By means of carefully selected examples on a 2D field dataset and on a full-azimuth 3D dataset, we demonstrate the ability of the proposed algorithm to handle high frequencies. We also show that the proposed algorithm scales well to 3D problems with large percentages of traces missing. From these results, we argue that the proposed approach could be a valuable alternative to transform-based methods that are force to work on small multidimensional patches.

CHAPTER 3

TIME-JITTERED MARINE ACQUISITION—A RANK-MINIMIZATION APPROACH FOR 5D SOURCE SEPARATION

3.1 Summary

Simultaneous source marine acquisition has been recognized as an economic way of improving spatial sampling and speedup acquisition time, where a single- (or multiple-) source vessel fires at jittered source locations and time instances. Consequently, the acquired simultaneous data volume is processed to separate the overlapping shot records resulting in densely sampled data volume. It has been shown in the past that the simultaneous source acquisition design and source separation process can be setup as a compressed sensing problem, where conventional seismic data is reconstructed from simultaneous data via a sparsity-promoting optimization formulation. While the recovery quality of separated data is reasonably well, the recovery process can be computationally expensive due to transform-domain redundancy. In this chapter, we present a computationally tractable rank-minimization algorithm to separate simultaneous data volumes. The proposed algorithm is suitable for large-scale seismic data, since it avoids singular-value decompositions and uses a low-rank based factorized formulation instead. Results are illustrated for simulations of simultaneous time-jittered continuous recording for a 3D ocean-bottom cable survey.

3.2 Introduction

Simultaneous source marine acquisition mitigates the challenges posed by conventional marine acquisition in terms of sampling and survey efficiency, since more than one shot can be fired at the same time (Beasley, Chambers, and Jiang 1998; Kok and Gillespie 2002;

Berkhout 2008; Beasley 2008; Hampson, Stefani, and Herkenhoff 2008). The final objective of source separation is to get interference-free shot records. (Wason and Herrmann 2013) have shown that the challenge of separating simultaneous data can be addressed through a combination of tailored single- (or multiple-) source simultaneous acquisition design and curvelet-based sparsity-promoting recovery. The idea is to design a pragmatic time-jittered marine acquisition scheme where acquisition time is reduced and spatial sampling is improved by separating overlapping shot records and interpolating jittered coarse source locations to fine source sampling grid. While the proposed sparsity-promoting approach recovers densely sampled conventional data reasonably well, it poses computational challenges since curvelet-based sparsity-promoting methods can become computationally intractable—in terms of speed and memory storage—especially for large-scale 5D seismic data volumes.

Recently, nuclear-norm minimization based methods have shown the potential to overcome the computational bottleneck (Kumar, Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015), hence, these methods are successfully used for source separation (Maraschini, Dyer, Stevens, and Bird 2012; Cheng and Sacchi 2013; Kumar, Wason, and Herrmann 2015). The general idea is that conventional seismic data can be well approximated in some rank-revealing transform domain where the data exhibit low-rank structure or fast decay of singular values. Therefore, in order to use nuclear-norm minimization based algorithms for source separation, the acquisition design should increase the rank or slow the decay of the singular values. In (Kumar, Wason, and Herrmann 2015) we used nuclear-norm minimization formulation to separate simultaneous data acquired from an over/under acquisition design, where the separation is performed on each monochromatic data matrix independently. However, by virtue of the design of the simultaneous time-jittered marine acquisition we formulate a nuclear-norm minimization formulation that works on the temporal-frequency domain—i.e., using all monochromatic data matrices together. One of the computational bottlenecks of working with the nuclear-norm minimization formulation

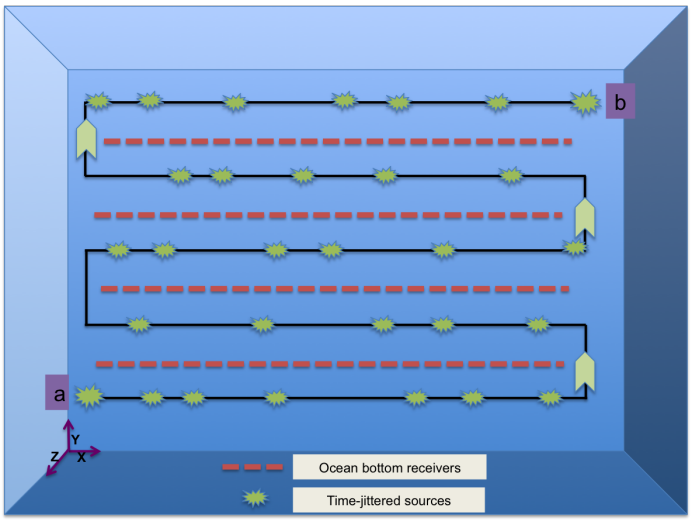


Figure 3.1: Aerial view of the 3D time-jittered marine acquisition. Here, we consider one source vessel with two airgun arrays firing at jittered times and locations. Starting from point a, the source vessel follows the acquisition path shown by black lines and ends at point b. The receivers are placed at the ocean bottom (red dashed lines).

is the computation of singular values. Therefore, in this chapter we combine the modified nuclear-norm minimization approach with the factorization approach recently developed by (Lee, Recht, Salakhutdinov, Srebro, and Tropp 2010). The experimental results on a synthetic 5D data set demonstrate successful implementation of the proposed methodology.

3.3 Methodology

Simultaneous source separation problem can be perceived as a rank-minimization problem. In this chapter, we follow the time-jittered marine acquisition setting proposed by (Wason and Herrmann 2013), where a single source vessel sails across an ocean-bottom array firing two airgun arrays at jittered source locations and time instances with receivers recording continuously (Figure 3.1). This results in a continuous time-jittered simultaneous data volume.

Conventional 5D seismic data volume can be represented as a tensor $\mathbf{D} \in \mathbb{C}^{n_f \times n_{rx} \times n_{sx} \times n_{ry} \times n_{sy}}$, where (n_{sx}, n_{sy}) and (n_{rx}, n_{ry}) represents number of sources and receivers along x, y coor-

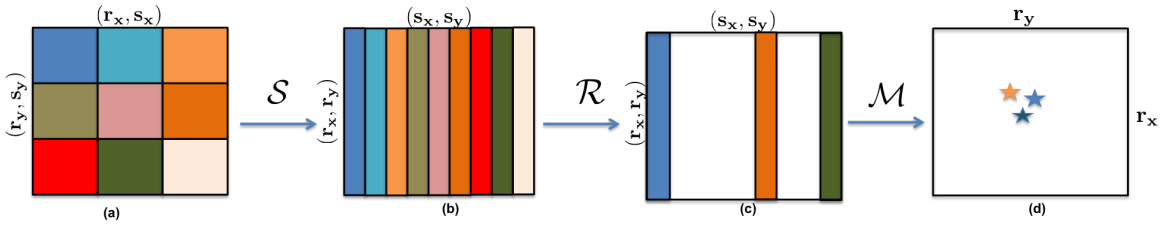


Figure 3.2: Schematic representation of the sampling-transformation operator $\mathcal{A}$ during the forward operation. The adjoint of the operator $\mathcal{A}$ follows accordingly. (a, b, c) represent a monochromatic data slice from conventional data volume and (d) represents a time slice from the continuous data volume.

dinates and n_f represents number of frequencies. The aim is to recover the data volume $\mathbf{D}$ from the continuous time-domain simultaneous data volume $\mathbf{b} \in \mathbb{C}^{n_T \times n_{rx} \times n_{ry}}$ by finding a minimum rank solution $\mathbf{D}$ that satisfies the system of equations $\mathcal{A}(\mathbf{D}) = \mathbf{b}$. Here, $\mathcal{A}$ represents a linear sampling-transformation operator, $n_T < (n_t \times n_{sx} \times n_{sy})$ is the total number of time samples in the continuous time-domain simultaneous data volume, n_t is the total number of time samples in the conventional seismic data. Note that the operator $\mathcal{A}$ maps $\mathbf{D}$ to a lower dimensional simultaneous data volume $\mathbf{b}$ since the acquisition process superimposes shot records shifted with respect to their firing times. The sampling-transformation operator $\mathcal{A}$ is defined as $\mathcal{A} = \mathcal{MRS}$, where the operator $\mathcal{S}$ permutes the tensor coordinates from $(n_{rx}, n_{sx}, n_{ry}, n_{sy})$ (rank-revealing domain, i.e., Figure 3.2 a) to $(n_{rx}, n_{ry}, n_{sx}, n_{sy})$ (standard acquisition ordering, i.e., Figure 3.2 b) and its adjoint reverses this permutation. The restriction operator $\mathcal{R}$ subsamples the conventional data volume at jittered source locations (Figure 3.2 c), the sampling operator $\mathcal{M}$ maps the conventional subsampled temporal-frequency domain data to the simultaneous time-domain data (Figure 3.2 d). Note that Figure 3.2 d represents a time slice from the continuous (simultaneous) data volume where the stars represent locations of jittered sources in the simultaneous acquisition.

Rank-minimization formulations require that the target data set should exhibit a low-rank structure or fast decay of singular values. Consequently, the sampling-restriction $\mathcal{MR}$ operation should increase the rank or slow the decay of singular values. As we know, there

is no unique notion of rank for tensors, therefore, we can choose the rank of different matricizations of $\mathbf{D}$ (Kreimer and Sacchi 2012) where the idea is to create the matrix $\mathbf{D}^{(i)}$ by group the dimensions of $\mathbf{D}^{(i)}$ specified by i and vectorize them along the rows while vectorizing the other dimensions along the columns. In this work, we consider the matricization proposed by (Silva and Herrmann 2013), where $i = (n_{sx}, n_{sy})$—i.e., placing both source coordinates along the columns (Figure 3.3a), or $i = (n_{rx}, n_{sx})$—i.e., placing receiver-x and source-x coordinates along the columns (Figure 3.3b). As we see in Figure 3.3e, the matricization $i = (n_{sx}, n_{sy})$ has higher rank or slow decay of the singular values (solid red curve) compared to the matricization $i = (n_{rx}, n_{sx})$ (solid blue curve). The sampling-restriction operator removes random columns in the matricization $i = (n_{sx}, n_{sy})$ (Figure 3.3c), as a result the overall singular values decay faster (dotted red curve). This is because missing columns put the singular values to zero, which is opposite to the requirement of rank-minimization algorithms. On the other hand, the sampling-restriction operator removes random blocks in the matricization $i = (n_{rx}, n_{sx})$ (Figure 3.3d), hence, slowing down the decay of the singular values (dotted blue curve). This scenario is much closer to the matrix-completion problem ((Recht, Fazel, and Parrilo 2010a)), where samples are removed at random points in a matrix. Therefore, we address the source separation problem by exploiting low-rank structure in the matricization $i = (n_{rx}, n_{sx})$.

Since rank-minimization problems are NP hard and therefore computationally intractable, (Recht, Fazel, and Parrilo 2010a) showed that solutions to rank-minimization problems can be found by solving a nuclear-norm minimization problem. (Silva and Herrmann 2013) showed that for seismic data interpolation the sampling operator $\mathcal{M}$ is separable, hence, data can be interpolated by working on each monochromatic data tensor independently. Since in continuous time-jittered marine acquisition, the sampling operator $\mathcal{M}$ is nonseparable as it is a combined time-shifting and shot-jittering operator, we can not perform source separation independently over different monochromatic data tensors. Therefore, we formulate the nuclear-norm minimization formulation over the

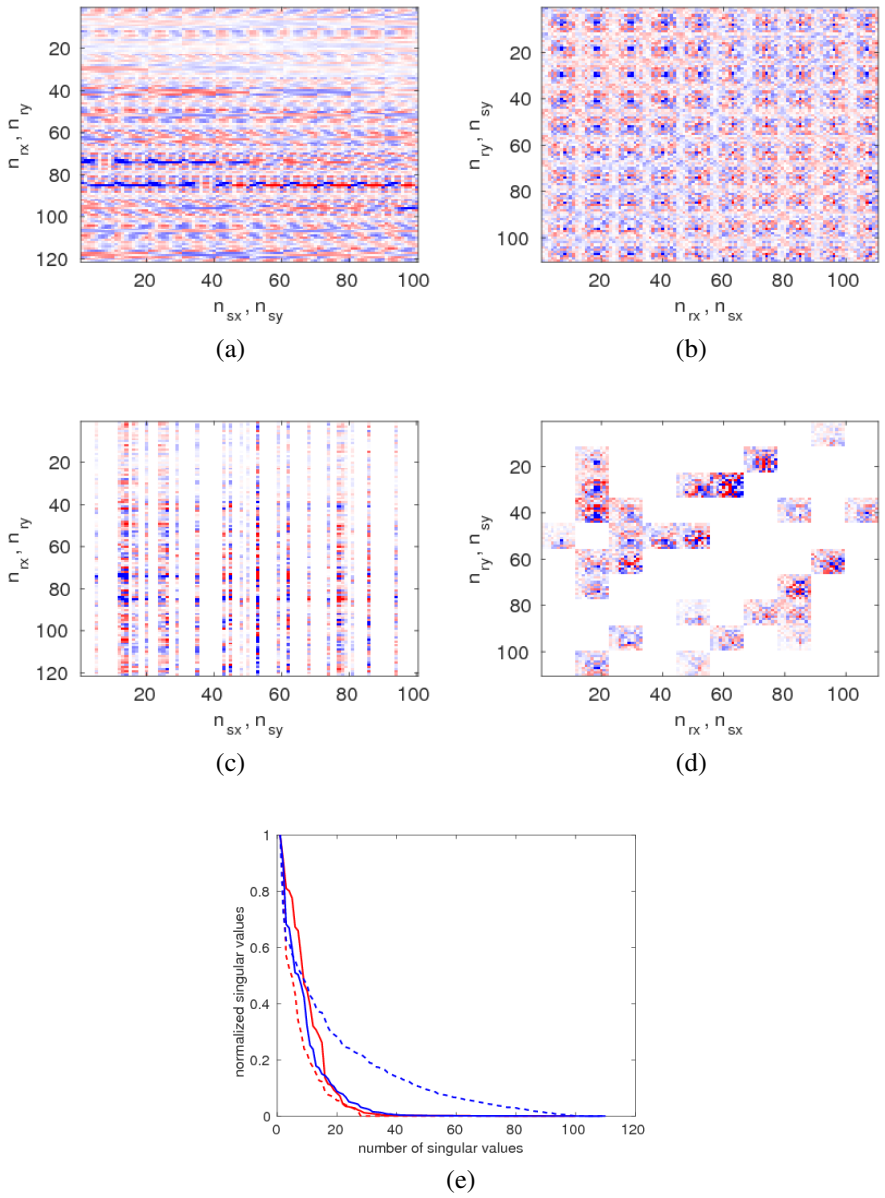


Figure 3.3: Monochromatic slice at 10.0 Hz. Fully sampled data volume and simultaneous data volume matricized as (a, c) $i = (n_{sx}, n_{sy})$, and (b, d) $i = (n_{rx}, n_{sx})$. (e) Decay of singular values. Notice that fully sampled data organized as $i = (n_{sx}, n_{sy})$ has slow decay of the singular values (solid red curve) compared to the $i = (n_{rx}, n_{sx})$ organization (solid blue curve). However, the sampling-restriction operator slows the decay of the singular values in the $i = (n_{rx}, n_{sx})$ organization (dotted blue curve) compared to the $i = (n_{sx}, n_{sy})$ organization (dotted red curve), which is a favorable scenario for the rank-minimization formulation.

temporal-frequency domain as follows:

$$\min_{\mathbf{D}} \sum_{j}^{n_f} \|\mathbf{D}_j^{(i)}\|_* \quad \text{subject to} \quad \|\mathcal{A}(\mathbf{D}) - \mathbf{b}\|_2 \leq \epsilon, \tag{3.1}$$

where $\sum_j^{n_f} \|.\|_* = \sum_j^{n_f} \|\boldsymbol{\sigma}_j\|_1$ and $\boldsymbol{\sigma}_j$ is the vector of singular values for each monochromatic data matricization. One of the main drawbacks of the nuclear-norm minimization problem is that it involves computation of the singular-value decomposition (SVD) of the matrices, which is prohibitively expensive for large-scale seismic data. Therefore, we avoid the direct approach to nuclear-norm minimization problem and follow a factorization-based approach (Rennie and Srebro 2005; Lee, Recht, Salakhutdinov, Srebro, and Tropp 2010; Recht and Ré 2011). The factorization-based approach parametrizes each monochromatic data matrix $\mathbf{D}^{(i)}$ as a product of two low-rank factors $\mathbf{L}^{(i)} \in \mathbb{C}^{(n_{rx} \cdot n_{sx}) \times k}$ and $\mathbf{R}^{(i)} \in \mathbb{C}^{(n_{ry} \cdot n_{sy}) \times k}$ such that, $\mathbf{D}^{(i)} = \mathbf{L}^{(i)} \mathbf{R}^{(i)^H}$, where k represents the rank of the underlying matrix and ^{H} represents the Hermitian transpose. Note that tensors $\mathbf{L}, \mathbf{R}$ can be formed by concatenating each matrix $\mathbf{L}^{(i)}, \mathbf{R}^{(i)}$, respectively. The optimization scheme can then be carried out using the tensors $\mathbf{L}, \mathbf{R}$ instead of $\mathbf{D}$, thereby significantly reducing the size of the decision variable from $n_{rx} \times n_{ry} \times n_{sx} \times n_{sy} \times n_f$ to $2k \times n_{rx} \times n_{sx} \times n_f$ when $k \leq n_{rx} \times n_{sx}$. Following (Rennie and Srebro 2005), the sum of the nuclear norm obeys the relationship:

$$\sum_{j}^{n_f} \|\mathbf{D}_j^{(i)}\|_* \leq \sum_{j}^{n_f} \frac{1}{2} \|\mathbf{L}_j^{(i)} \mathbf{R}_j^{(i)}\|_F^2,$$

where $\|\cdot\|_F^2$ is the Frobenius norm of the matrix (sum of the squared entries).

3.4 Experiments & Results

We test the efficacy of our method by simulating a synthetic 5D data set using the BG Compass velocity model (provided by the BG Group) which is a geologically complex and realistic model. We also quantify the cost savings associated with simultaneous acquisition in terms of an improved spatial-sampling ratio defined as a ratio between the spatial grid

interval of observed simultaneous time-jittered acquisition and the spatial grid interval of recovered conventional acquisition. The speed-up in acquisition is measured using the survey-time ratio (STR), proposed by (Berkhout 2008), which measures the ratio of time of conventional acquisition and simultaneous acquisition.

Using a time-stepping finite-difference modelling code provided by Chevron, we simulate a conventional 5D data set of dimensions $2501 \times 101 \times 101 \times 40 \times 40$ ($n_t \times n_{rx} \times n_{ry} \times n_{sx} \times n_{sy}$) over a survey area of approximately $4\,\mathrm{km} \times 4\,\mathrm{km}$. Conventional time-sampling interval is $4.0\,\mathrm{ms}$, source- and receiver-sampling interval is $6.25\,\mathrm{m}$. We use a Ricker wavelet with central frequency of $15.0\,\mathrm{Hz}$ as source function. Figure 3.4a shows a conventional common-shot gather. Applying the sampling-transformation operator ($\mathcal{A}$) to the conventional data generates approximately 65 minutes of 3D continuous time-domain simultaneous seismic data, 30 seconds of which is shown in Figure 3.4b. By virtue of the design of the simultaneous time-jittered acquisition, the simultaneous data volume $\mathbf{b}$ is 4-times subsampled compared to conventional acquisition. Consequently, the spatial sampling of recovered data is improved by a factor of 4 and the acquisition time is reduced by the same factor.

Simply applying the adjoint of the sampling operator $\mathcal{M}$ to simultaneous data $\mathbf{b}$ results in strong interferences from other sources as shown in Figure 3.4c. Therefore, to recover the interference-free conventional seismic data volume from the simultaneous time-jittered data, we solve the factorization based nuclear-norm minimization formulation. We perform the source separation for a range of rank k values of the two low-rank factors $\mathbf{L}^{(i)}, \mathbf{R}^{(i)}$ and find that $k = 100$ gives the best signal-to-noise ratio (SNR) of the recovered conventional data. Figure 3.4d shows the recovered shot gather, with an SNR of $20.8\,\mathrm{dB}$, and the corresponding residual is shown in Figure 3.4e. As illustrated, we are able to separate the shots along with interpolating the data to the finer grid of $6.25\,\mathrm{m}$. To establish that we loose very small coherent energy during source separation, we intensify the amplitudes of the residual plot by a factor of 8 (Figure 3.4e). The late arriving events, which are often weak in energy,

are also separated reasonably well. Computational efficiency of the rank-minimization approach—in terms of the memory storage—in comparison to the curvelet-based sparsity-promoting approach is approximately 7.2 when compared with 2D curvelets and 24 when compared with 3D curvelets.

3.5 Conclusions

We propose a factorization based nuclear-norm minimization formulation for simultaneous source separation and interpolation of 5D seismic data volume. Since the sampling-transformation operator is nonseparable in the simultaneous time-jittered marine acquisition, we formulate the factorization based nuclear-norm minimization problem over the entire temporal-frequency domain, contrary to solving each monochromatic data matrix independently. We show that the proposed methodology is able to separate and interpolate the data to a fine underlying grid reasonably well. The proposed approach is computationally memory efficient in comparison to the curvelet-based sparsity-promoting approach.

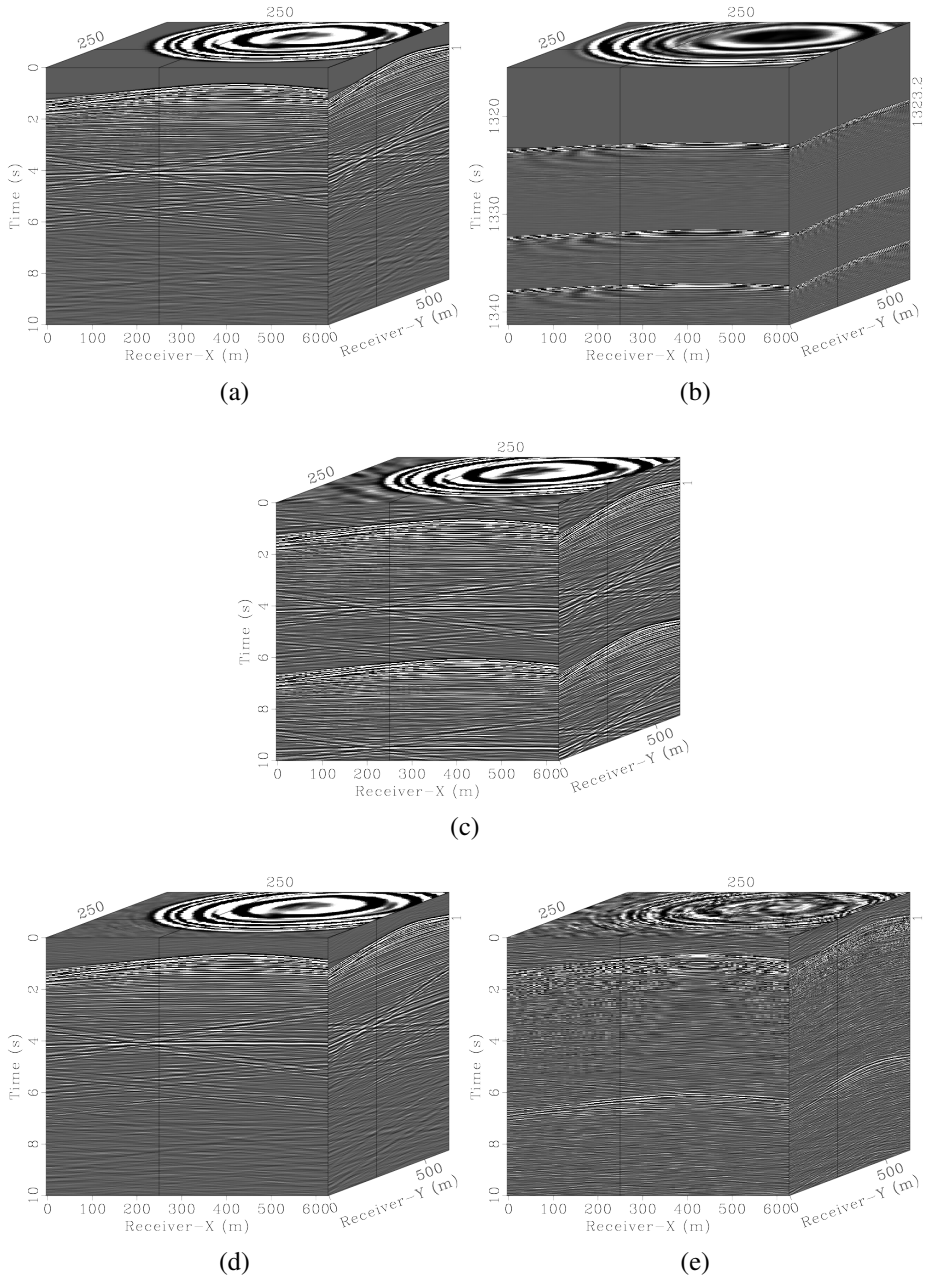


Figure 3.4: Source separation recovery. A shot gather from the (a) conventional data; (b) a section of 30 seconds from the continuous time-domain simultaneous data (b); (c) recovered data by applying the adjoint of the sampling operator $\mathcal{M}$; (d) data recovered via the proposed formulation (SNR = $20.8\,\mathrm{dB}$); (e) difference of (a) and (d) where amplitudes are magnified by a factor of 8 to illustrate a very small loss in coherent energy.

CHAPTER 4

FAST SPARSITY-PROMOTING MICROSEISMIC SOURCE ESTIMATION

4.1 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.

4.2 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 chapter, 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.

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 and Muir 1973) , who used the ℓ_1-norm to handle outliers in the data followed by (Taylor, Banks, and McCoy 1979), (Santosa and Symes 1986), and (Donoho and 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 and Rauhut 2008; Sharan, Wang, Leeuwen, and Herrmann 2016; Gao and 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 and 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 and 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 and 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.

To avoid computationally expensive scanning, in (Nakata and Beroza 2016) and (Sun, Zhu, Fomel, and Song 2015) authors combined back-propagation and cross-correlation based methods to get improved spatial resolution of sources. In (Sun, Xue, Fomel, Zhu, and Nakata 2016) authors 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 and McMechan 1996) and (Kim, Liu, and Tromp 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, McChesney, and Minkoff 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, Alkhalifah, Wu, and Sun 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 and 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, Albera, Bertin, and Gribonval

2016) and propose a sparsity-promoting method (Sharan, Wang, Leeuwen, and Herrmann 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, Albera, Bertin, and Gribonval 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 and 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.

4.3 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, Romberg, and Tao 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.

4.3.1 The forward model

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

$$\mathbf{d} = \mathcal{F}[\mathbf{m}](\mathbf{Q}), \tag{4.1}$$

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 Figure 4.1a, which we generated with the acquisition setup of Figure 4.1b. We obtained this blurred image (Figure 4.2a) 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 Figure 4.2a.

4.3.2 Minimal energy solutions

Equation 4.1 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 equation 4.1 as a constrained optimization problem:

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

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 equation 4.1 by minimizing energy of the unknown source wavefield is inadequate as we can see by juxtaposing Figures 4.2a and 4.2b. While it is relatively easy to state the forward model (equation 4.1), 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.

4.3.3 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 and 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 and 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, Romberg, and Tao 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

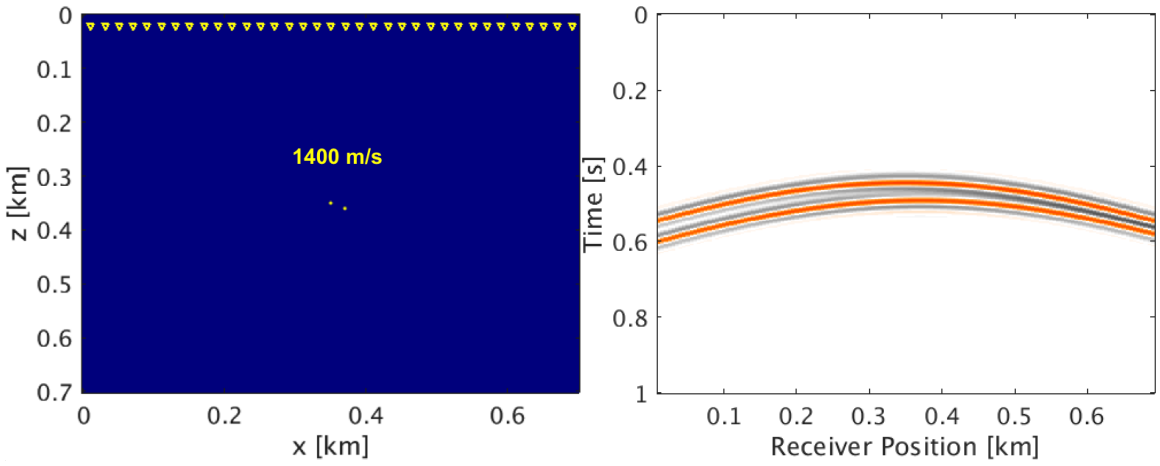
source or else the solution would not have been sparse. We can observe this phenomenon in Figure 4.2b where the intensity (sum of absolute values over time) of the minimal-energy solution is spread out spatially rather than being focussed onto the point sources that were responsible for generating the observed data according to equation 4.1. 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, Albera, Bertin, and Gribonval 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 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 Figures 4.2b and 4.3d). Below, we present algorithmic details on how these results can be accomplished at greatly reduced computational costs.

4.4 Sparsity-promoting source localization

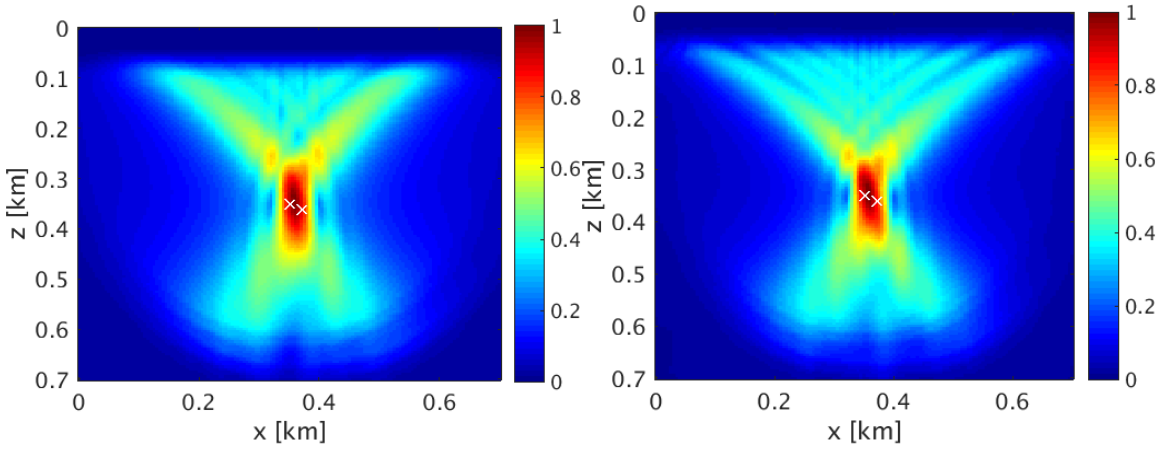
The observation that microseismic event locations are sparse in space supports the above theoretical arguments to impose the sparsity-promoting ℓ_1-norm along the spatial coordinates. Because stress drops are finite during these micro-seismic events, the emitted energy will be finite as well justifying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

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



(a) (b)

Figure 4.1: 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 $20.0\,\mathrm{m}$ and separated by $5.0\,\mathrm{m}$. The yellow dots indicate the location of 2 microseismic sources. *(b)* Microseismic data generated by two microseismic sources.



(a) (b)

Figure 4.2: *(a)* Back-propagated source image, *(b)* Intensity plot of the minimal energy solution (cf. equation LS_ϵ). White colour crosses indicate the actual location of two point sources.

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 $\mathbf{m}$, equation $\text{LS}_{\epsilon}^{2,1}$ takes a form similar to the classic Basis Pursuit Denoising (BPDN) problem (Chen, Donoho, and Saunders 1998; Berg and 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.

4.4.1 Linearized Bregman algorithm

Motivated by recent successful application of the linearized Bregman method (Yin, Osher, Goldfarb, and Darbon 2008; Lorenz, Schöpfer, and Wenger 2014) to sparsity-promoting least-squares migration (Herrmann, Tu, and Esser 2015; Witte, Louboutin, Luporini, Gorman, and Herrmann 2019), 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

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

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 equation $\text{LBR}^{2,1}_{\mu,\epsilon}$ approaches the solution of equation $\text{LS}^{2,1}_{\epsilon}$ (Osher, Mao, Dong, and Yin 2010). Because the objective in equation $\text{LBR}^{2,1}_{\mu,\epsilon}$ now contains a quadratic term, problem $\text{LBR}^{2,1}_{\mu,\epsilon}$ becomes easier to solve (Boyd and 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, Osher, Goldfarb, and Darbon 2008), Algorithm 4.1 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 and Rauhut 2008; Gao and Sacchi 2017). Despite this important difference, the resulting algorithm still shares the simplicity of the original linearized Bregman as we see from Algorithm 4.1.

Algorithm 4.1 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, Wenger, Schöpfer, and Magnor 2014), we can cheaply calculate a dynamic step length

$$t_k = \frac{\|\mathcal{F}[\mathbf{m}](\mathbf{Q}_k) - \mathbf{d}\|^2}{\|\mathcal{F}[\mathbf{m}]^\top(\mathcal{F}[\mathbf{m}](\mathbf{Q}_k) - \mathbf{d})\|_F^2} \tag{4.2}$$

that guarantees Algorithm 4.1 to converge (Lorenz, Schöpfer, and Wenger 2014). In this expression, we reserve the symbol $^\top$ for the matrix transpose.

During each iteration of Algorithm 4.1, 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:

$$\Pi_\epsilon(\mathbf{r}) = \max\left\{0, 1 - \frac{\epsilon}{\|\mathbf{r}\|}\right\} \mathbf{r}. \tag{4.3}$$

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 and Pesquet 2011) on a matrix $\mathbf{C}$ is defined by

$$\text{Prox}_{\mu\ell_{2,1}}(\mathbf{C}) := \arg\min_{\mathbf{B}} \quad \|\mathbf{B}\|_{2,1} + \frac{1}{2\mu}\|\mathbf{C} - \mathbf{B}\|_F^2\,, \tag{4.4}$$

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

$$\left(\text{Prox}_{\mu\ell_{2,1}}(\mathbf{C})\right)_{i,j} = S(\mathbf{C}_{i,j}, \frac{\mu}{\|\mathbf{C}_{i,:}\|^2}), \tag{4.5}$$

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

$$S(a,T) = \max\{|a| - T, 0\}\text{sign}(a) \tag{4.6}$$

is the soft thresholding operator. From equation 4.5, 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 4.1, let us first briefly explain

what the different steps of Algorithm 4.1 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 4.1 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 equations LS_ϵ and $\text{LBR}^{2,1}_{\mu,\epsilon}$ (cf. Figures 4.2b and 4.3d) clearly shows the ability of sparsity-promotion to locate closely spaced microseismic sources.

4.4.2 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 equation $\text{LBR}^{2,1}_{\mu,\epsilon}$ 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 Figure 4.3 where we aim to recover two events separated by $22.0\,\mathrm{m}$ (Figure 4.1a), which is very close to half of the dominant wavelength of $23.0\,\mathrm{m}$. 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 (Figure 4.1a), is relatively simple, its solution is much more challenging because of the poor conditioning of $\mathcal{F}[\mathrm{m}]$. 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 4.1 after k iterations. We find that after 200 iterations, we have a normalized residual of less than 0.05 for $\mu = 8\text{e-}5$. For this value of μ we are not able to resolve the two sources (Figure 4.3a). Even if we increase $\mu = 8\text{e-}4$ and run for 600 iterations, the sources are still not resolved (Figure 4.3b). Only by increasing μ to $8\text{e-}3$ we are able to get a reasonable result (Figure 4.3c) but at the expense of running 4900 iterations of the linearized Bregman Algorithm 4.1, 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 equation $\mathrm{LBR}^{2,1}_{\mu,\epsilon}$ is strongly convex, it has a smooth dual objective (Shalev-shwartz and 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 problem with Limited-memory BFGS (L-BFGS) (Liu and Nocedal

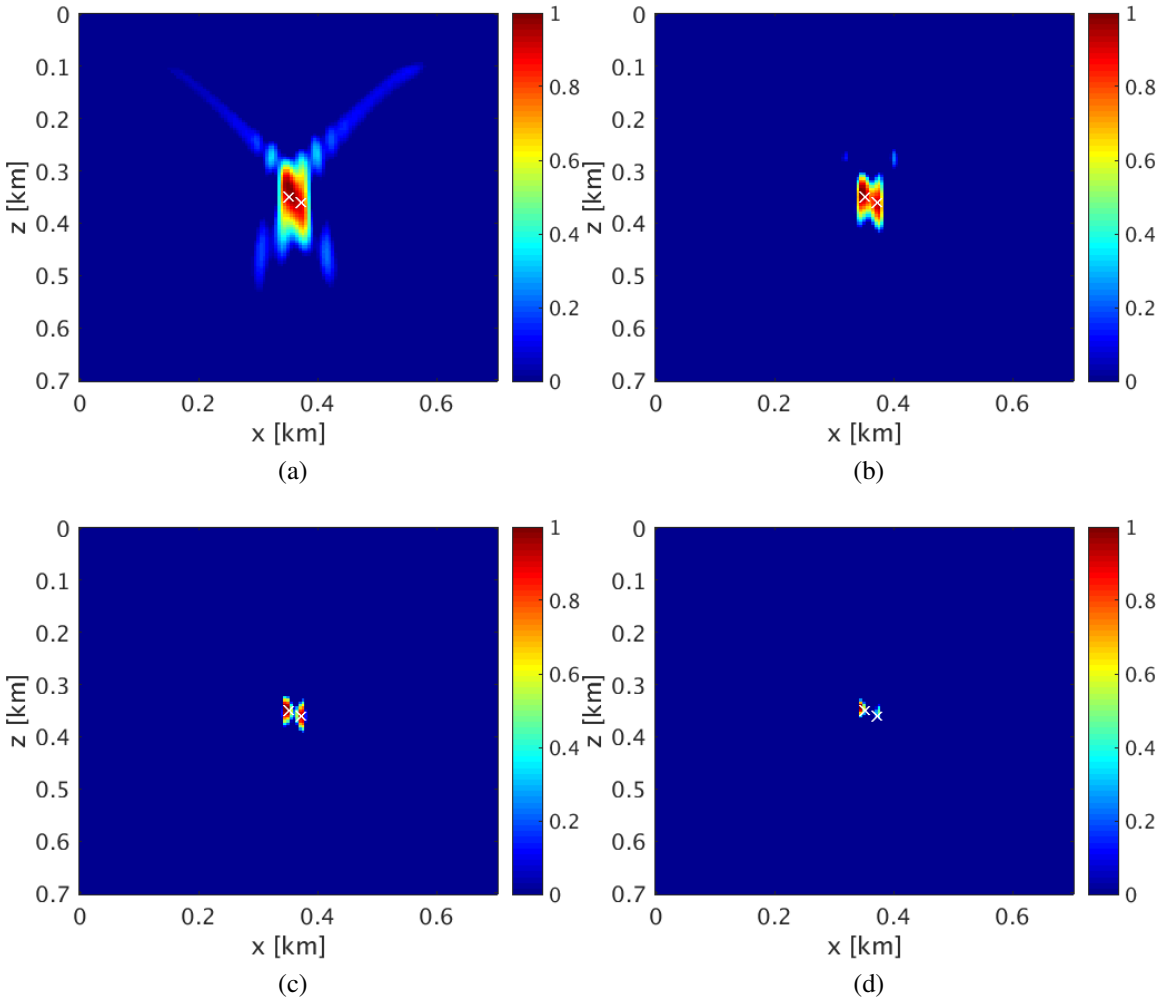


Figure 4.3: Estimated intensity plots. *(a)* After 200 iterations of linearized Bregman with $\mu = 8\text{e-}5$, *(b)* after 600 iterations of linearized Bregman with $\mu = 8\text{e-}4$ and *(c)* after 4900 iterations of linearized Bregman with $\mu = 8\text{e-}3$. *(d)* After only 10 preconditioned iterations of the dual formulation for linearized Bregman. Iterations of the dual formulation cost about the same as iterations of the primal problem (Algorithm 4.1). White colour crosses indicate the actual locations of microseismic sources

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 B)

$$\min_{\mathbf{y}} \quad f(\mathbf{y}) = -\left\{\Psi(\mathbf{y}) - \epsilon\|\mathbf{y}\|_2\right\}, \tag{4.7}$$

where

$$\Psi(\mathbf{y}) = \min_{\mathbf{Q}} \quad \|\mathbf{Q}\|_{2,1} + \frac{1}{2\mu}\|\mathbf{Q}\|_F^2 + \mathbf{y}^\top\big(\mathcal{F}[\mathbf{m}](\mathbf{Q}) - \mathbf{d}\big), \tag{4.8}$$

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 B) that equals

$$\nabla\Psi(\mathbf{y}) = \mathbf{d} - \mathcal{F}[\mathbf{m}]\big(\text{Prox}_{\mu\ell_{2,1}}(\mu\mathcal{F}[\mathbf{m}]^\top(\mathbf{y})\big). \tag{4.9}$$

Using the dual objective function $f(\mathbf{y})$ from equation 4.7 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

$$\nabla f(\mathbf{y}) = -\left\{\nabla\Psi(\mathbf{y}) - \epsilon\mathbf{y}/\|\mathbf{y}\|_2\right\}. \tag{4.10}$$

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 4.1, 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 4.2. The cost per iteration of this algorithm are approximately the same as those of Algorithm 4.1.

Algorithm 4.2 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 equation 4.7 and its gradient in equation 4.10. **Line** 4 corresponds to computing the primal solution by back propagating the dual solution with the adjoint of the forward modeling operator (cf. equation 4.1), 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 $\mathbf{I}(\mathbf{x}) = \text{vec}^{-1}\left(\sum_t \mid \mathbf{Q}(\mathbf{x}, t) \mid\right)$.

4.4.3 Preconditioning in 2D

As for the primal formulation (Algorithm 4.1), each iteration of L-BFGS in Algorithm 4.2 requires solving at least one forward and an adjoint wave equation. Although, Algorithm 4.2 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 and Williamson 1995). This causes

waveforms and amplitudes to differ significantly from the waveforms and amplitudes of a true point source. Following (Herrmann, Brown, Erlangga, and Moghaddam 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

$$\begin{aligned} \min_{\mathbf{Q}} \quad & \|\mathbf{Q}\|_{2,1} + \frac{1}{2\mu}\|\mathbf{Q}\|_F^2 \\ & \text{subject to} \\ & \|\mathbf{M}_L^{-1}\left(\mathcal{F}[\mathbf{m}](\mathbf{Q}) - \mathbf{d}\right)\|_2 \leq \gamma \end{aligned} \tag{4.11}$$

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 4.11 with L-BFGS.

If we apply our dual formulation along with the preconditioning to the source localization problem discussed in the last section (Figures 4.1a & 4.1b), we are now able to resolve close by sources (Figure 4.3d) in only 10 iterations. Figures 4.4a and 4.4b 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 Figures 4.4a and 4.4b). 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 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 Figure 4.5 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

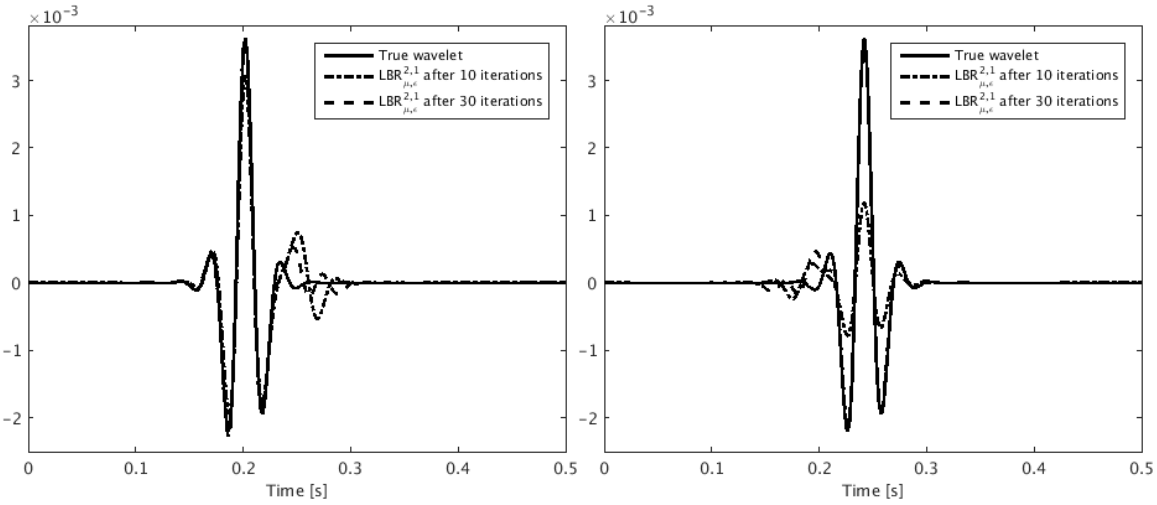


Figure 4.4: Source-time function comparison: True source-time functions (solid black), recovered source-time functions with 10 iterations (denoted by dash dot line) and with 30 iterations (dashed) of the accelerated dual formulation at the first in *(a)* and second source location in *(b)* from left to right. Estimated source-time functions are corrected by a factor of approximately 10 for comparison purposes.

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.

4.5 Numerical experiments

To compare the performance of our sparsity-promoting approach with the method of back propagation or minimum energy (cf. equation LS_ϵ), we conduct a series of numerical experiments on parts of the synthetic BG compass (Figure 4.6a) and Marmousi (Figure 4.10a)

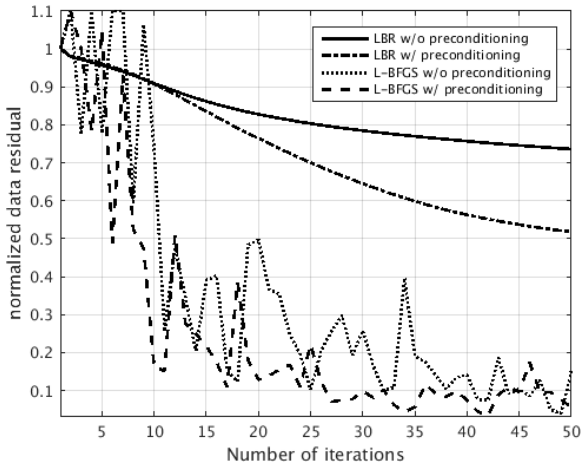


Figure 4.5: Convergence comparison. Decay of normalized ℓ_2 norm of data residual with linearized Bregman without preconditioning (solid black lines), with preconditioning (denoted by dash dot line) and L-BFGS without (denoted by dotted line) and with preconditioning (dashed).

velocity models. Both are sampled at $5\,\mathrm{m}$. 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, Aravkin, Leeuwen, and Herrmann 2012) was designed to test full-waveform inversion (Leeuwen and Herrmann 2012; Li, Esser, and Herrmann 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, Bourget, Lailly, Poulet, Ricarte, and Versteeg 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 and Beroza 2016) and (Li and Baan 2016) to test their methods.

For reasons stated earlier, we use a $2\mathrm{D}$ acoustic finite difference modeling code (Louboutin, Witte, Lange, Kukreja, Luporini, Gorman, and Herrmann 2017) to simulate microseismic data with a record length of $1\,\mathrm{s}$. Compared to conventional active-source seismic, microseismic events can be considered as low in amplitude and high in frequency earthquake-like events (Kamei, Nakata, and Lumley 2015). This motivates our choice of using the earthquake source model by (Madariaga 1989). Following (Nakata and Beroza

2016); (Sun, Xue, Fomel, Zhu, and Nakata 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 $20\,\mathrm{m}$ sampled at $10\,\mathrm{m}$. To speed up the iterations, we precondition both experiments. To more closely mimic field data, we add bandwidth limited noise ($5.0\,\mathrm{Hz}$ to $45.0\,\mathrm{Hz}$) to synthetic data we generate from the velocity models plotted in Figures 4.6a and 4.10a.

4.5.1 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 $3.21\,\mathrm{dB}$ (Figure 4.6c) generated from the complex velocity model plotted in Figure 4.6a with an aperture of almost $2.1\,\mathrm{km}$. 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 Figures 4.6a and 4.6b. 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 Figure 4.6a has a size of $2.25\,\mathrm{km} \times 0.915\,\mathrm{km}$, which corresponds to 451×184 grid points.

To make the source localization problem more challenging, we place two microseismic sources, with source-time functions of central frequency $30\,\mathrm{Hz}$ (solid black plots in Figures 4.8a and 4.8b) close to the high velocity unconformity at depths of $485\,\mathrm{m}$ and $490\,\mathrm{m}$ and separated by approximately $20.0\,\mathrm{m}$ well within Abbe's diffraction limit. We indicated these source locations by the two black dots in Figure 4.6a. We included results for the true and smoothed velocity models in the left and right columns of Figure 4.7. To arrive at these results, we used $\mu = 2\text{e-}2$, 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 (Figures 4.7a and 4.7b). 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. Figures 4.8a and 4.8b 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.

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 Figures 4.9a and 4.9b the normalized data residuals and the normalized Earth Movers Distance (EMD, (Rubner, Tomasi, and Guibas 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 defined as

$$\begin{aligned} \min_{\mu \geq 0} \quad & \int d(x,y)\mu(x,y)\mathrm{d}x\mathrm{d}y \\ \text{subject to} \quad & \int \mu(x,y)\mathrm{d}y = p(x) \\ & \int \mu(x,y)\mathrm{d}x = q(y) \end{aligned} \tag{4.12}$$

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

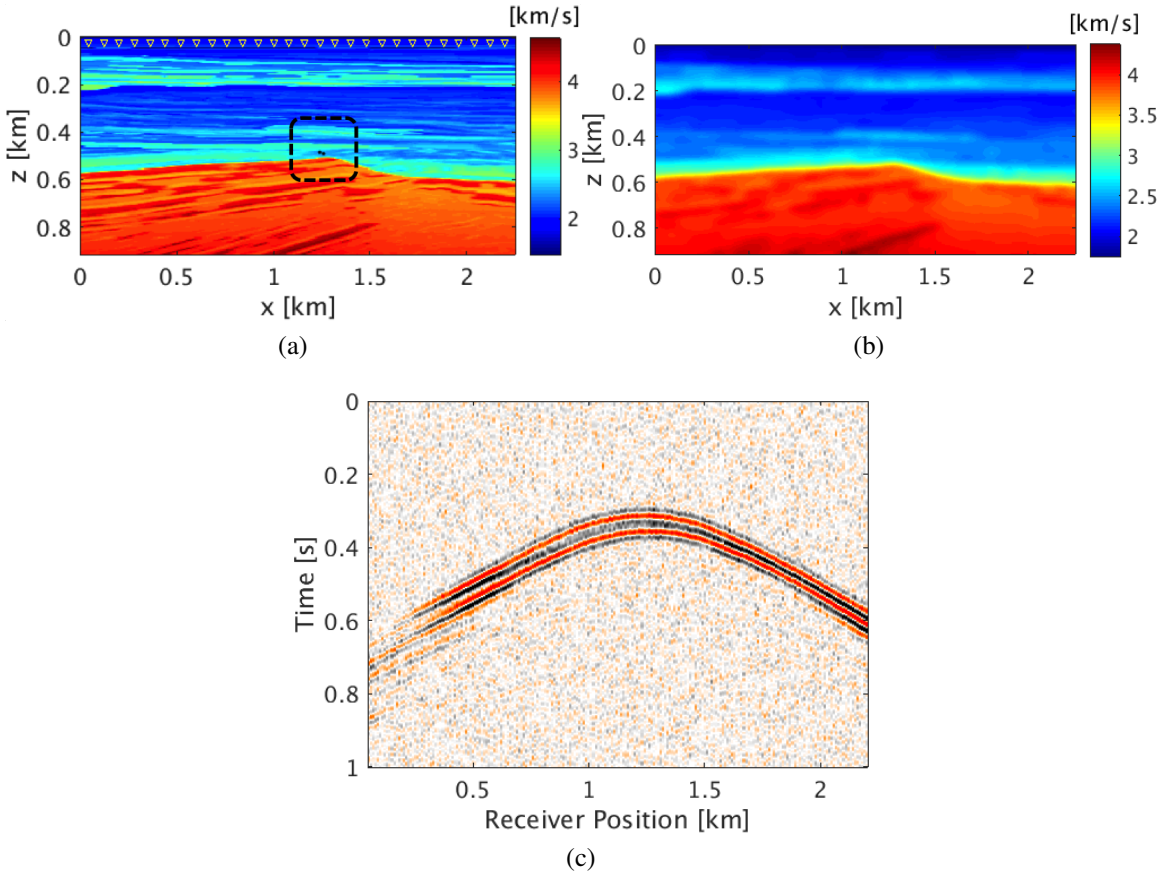


Figure 4.6: BG model source localization experiment. *(a)* True velocity model with acquisition geometry and microseismic source locations superimposed. The inverted yellow triangles indicate receivers buried at a depth of $20.0\,\mathrm{m}$ and sampled at $10.0\,\mathrm{m}$. The black dots indicate 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 ($5.0\,\mathrm{Hz}$ to $45.0\,\mathrm{Hz}$) yielding a S/R of only $3.21\,\mathrm{dB}$.

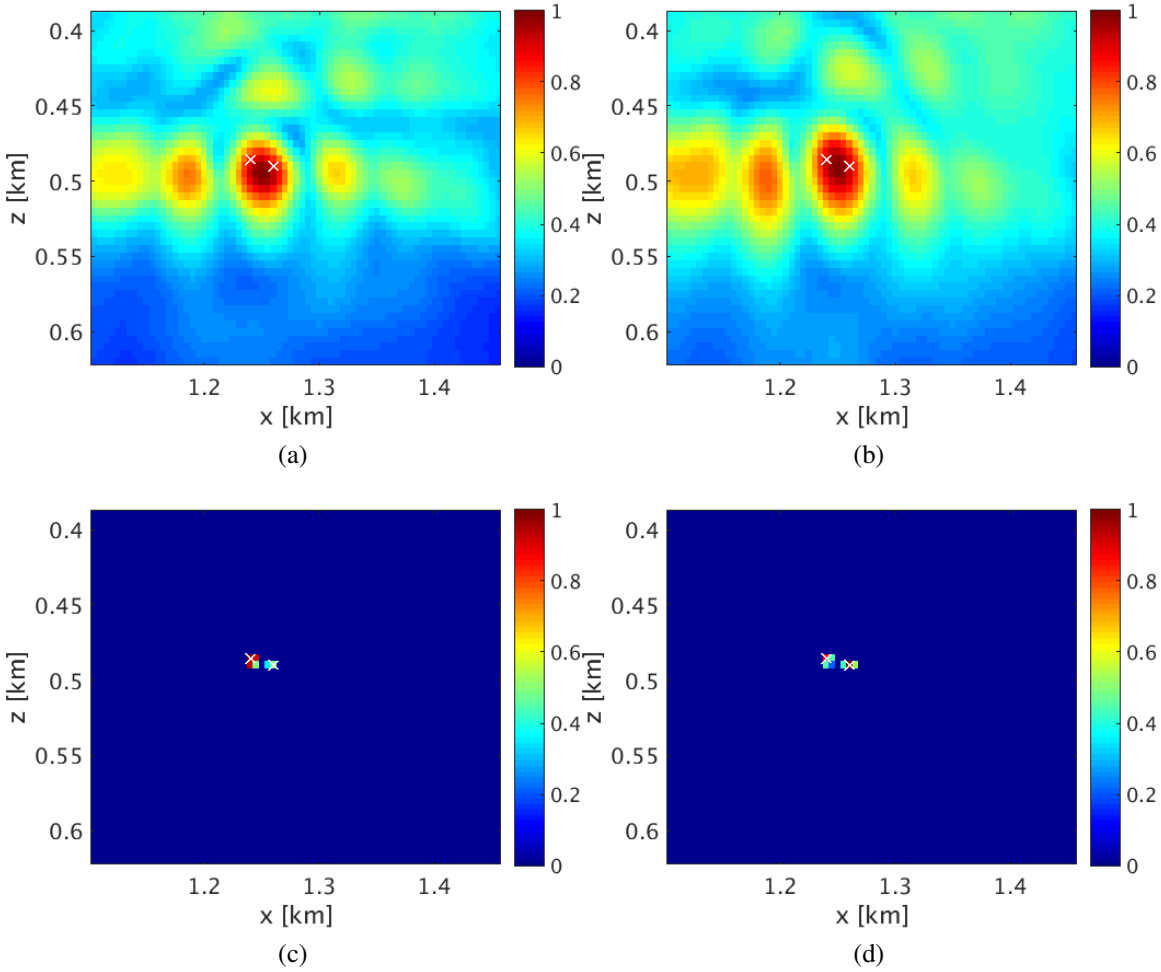


Figure 4.7: Zoomed intensity plots for the true (left column) and smoothed (right column) velocity models obtained by solving LS_ϵ with 300 iterations and plotted in *(a,b)*; and by solving $\mathrm{LBR}^{2,1}_{\mu,\epsilon}$ for $\mu = 2\text{e-}2$ with only ten iterations and plotted in *(c,d)*. The white crosses indicate the actual spatial locations of the two microseismic sources. Notice the improved resolution.

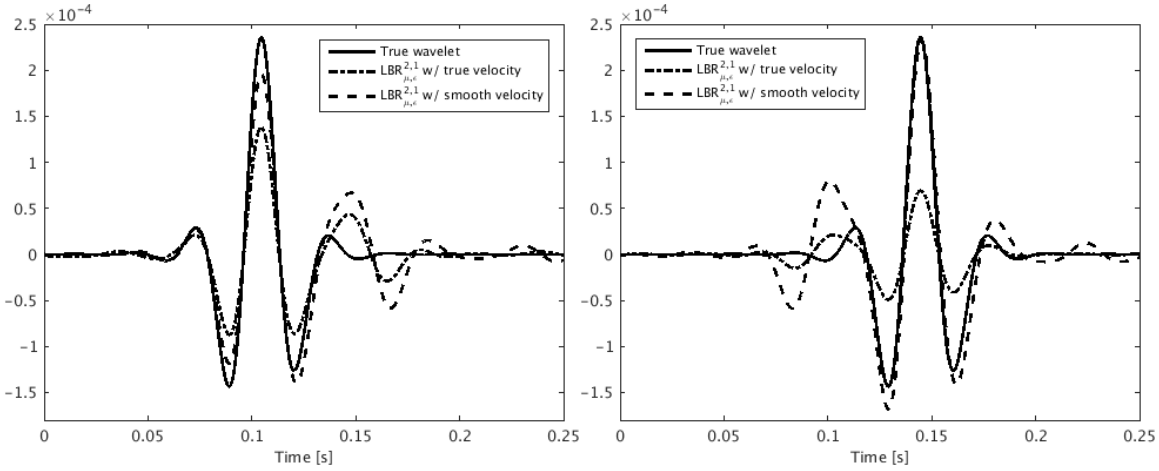


Figure 4.8: Comparison between the actual (solid black) and estimated source-time functions for the true velocity model (depicted by dash dot line) and the smoothed velocity model (dashed) at the two source locations plotted in *(a)* and *(b)* from left to right. The estimated source-time functions are corrected by a factor of 2 for comparison purposes.

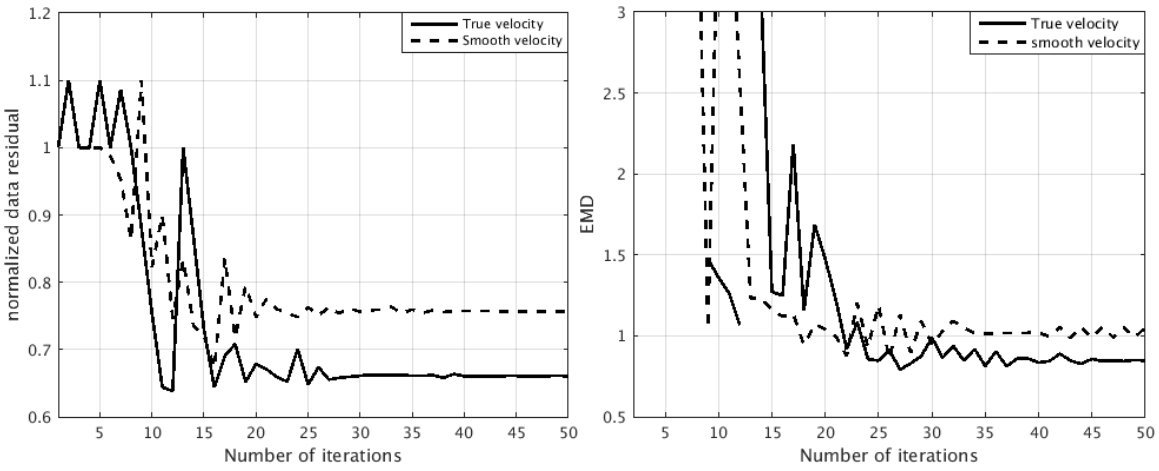
it difficult to quantitatively measure the performance of our algorithm as a function of the (normalized) model error in the ℓ_2-norm. Because point sources modeled by Dirac distributions are not ℓ_2 integrable, the energy norm does not lend itself well to measure distances between delta-like functions. 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, Osher, and Fatemi 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 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 Figures 4.9a and 4.9b, the solid and dashed lines in plots for the normalized residuals and EMDs correspond to inversions with the true and smoothed velocity models (cf. Figures 4.6a and 4.6b). 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.

4.5.2 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 = $3.5\,\mathrm{dB}$, Figure 4.10d) with an aperture of almost $3.0\,\mathrm{km}$. 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 $22.0)\,\mathrm{Hz}$. To test the validity of our approach, we work with a subset of



(a) (b)

Figure 4.9: 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.

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 Figure 4.10a). 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 (Figure 4.10b) 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. Figures 4.11a and 4.11b show the estimated intensity plots obtained after back propagation using the smoothed velocity model depicted in Figure 4.10b. 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.

Figures 4.11c and 4.11d 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 $\text{LBR}^{2,1}_{\mu,\epsilon}$) with $\mu = 9\text{e-}4$ clearly resolve the individual events (Figures 4.11e and 4.11f) 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 $9\text{e-}3$, we further improve the resolution (Figures 4.11g and 4.11h) 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 $11\,\text{m}$) 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. Figure 4.12 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 Figure 4.12. *(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 Figure 4.13 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.

4.6 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

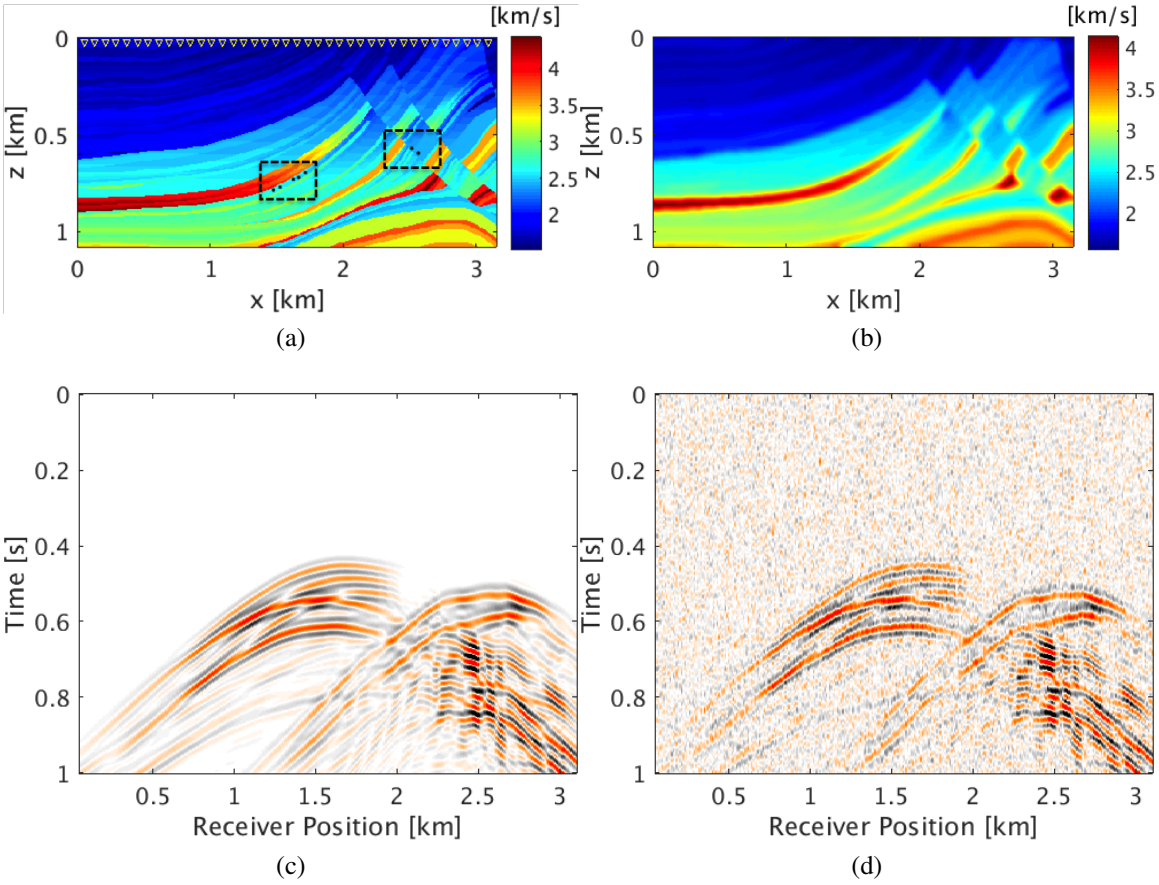


Figure 4.10: Marmousi model experiment. *(a)* True velocity model with acquisition geometry and the microseismic source locations superimposed. The inverted yellow triangles indicate receivers buried at a depth of $20.0\,\mathrm{m}$ and separated by $10.0\,\mathrm{m}$ with an aperture of almost $3.0\,\mathrm{km}$. 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 ($5.0\,\mathrm{Hz}$ to $45.0\,\mathrm{Hz}$) Gaussian random noise to the noise-free simulated data.

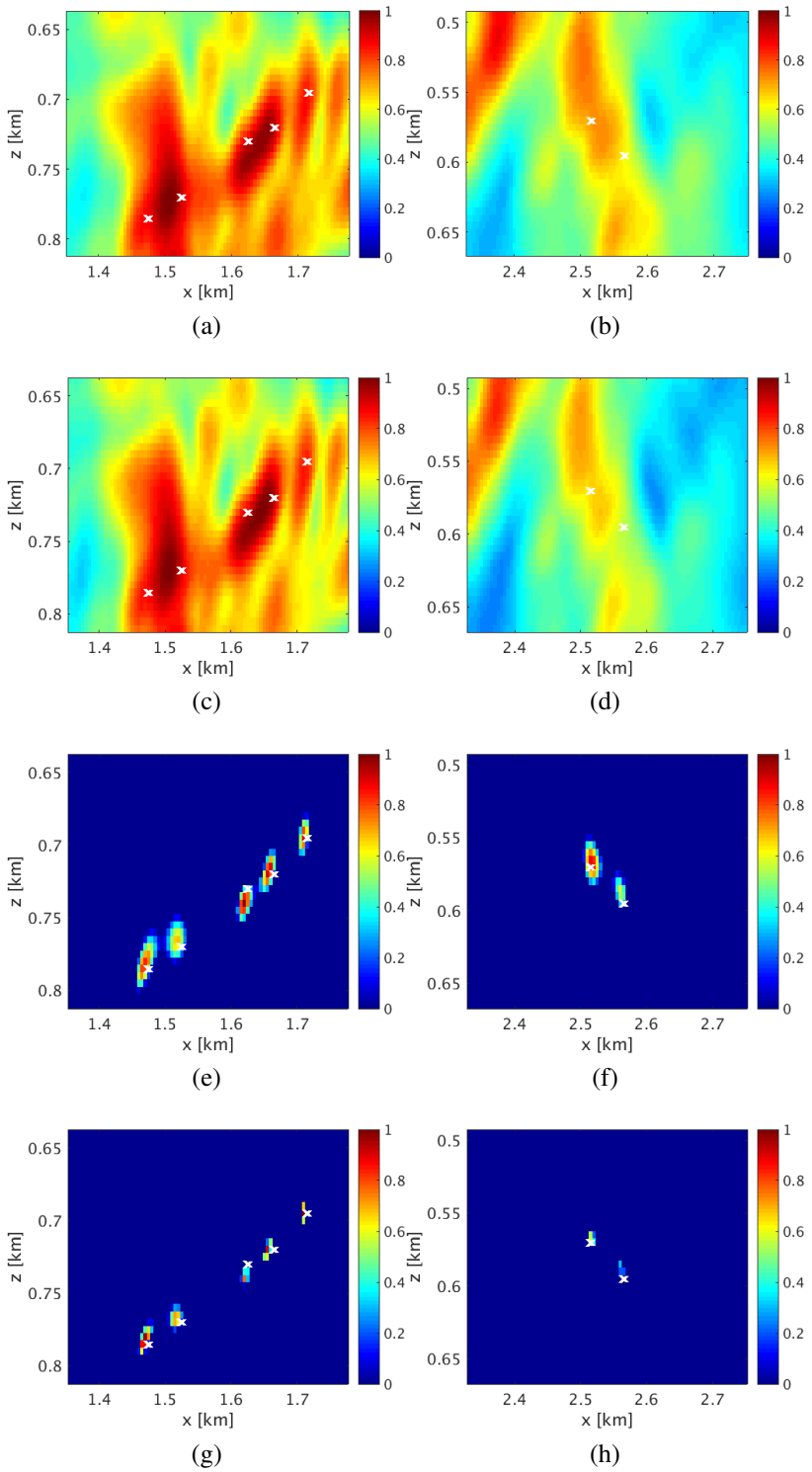


Figure 4.11: Zoomed intensity plots near cluster 1 (left column) and cluster 2 (right column) from left to right for the smoothed velocity models obtained via back propagation in *(a, b)*; solving LS_ϵ with 300 iterations in *(c, d)*; and by solving $\text{LBR}^{2,1}_{\mu,\epsilon}$ for $\mu = 9\text{e-}4$ with only 10 iterations in *(e, f)*; and by solving $\text{LBR}^{2,1}_{\mu,\epsilon}$ for $\mu = 9\text{e-}3$ with 30 iterations in *(g, h)*. The white colour crosses indicate the actual spatial locations of the microseismic sources.

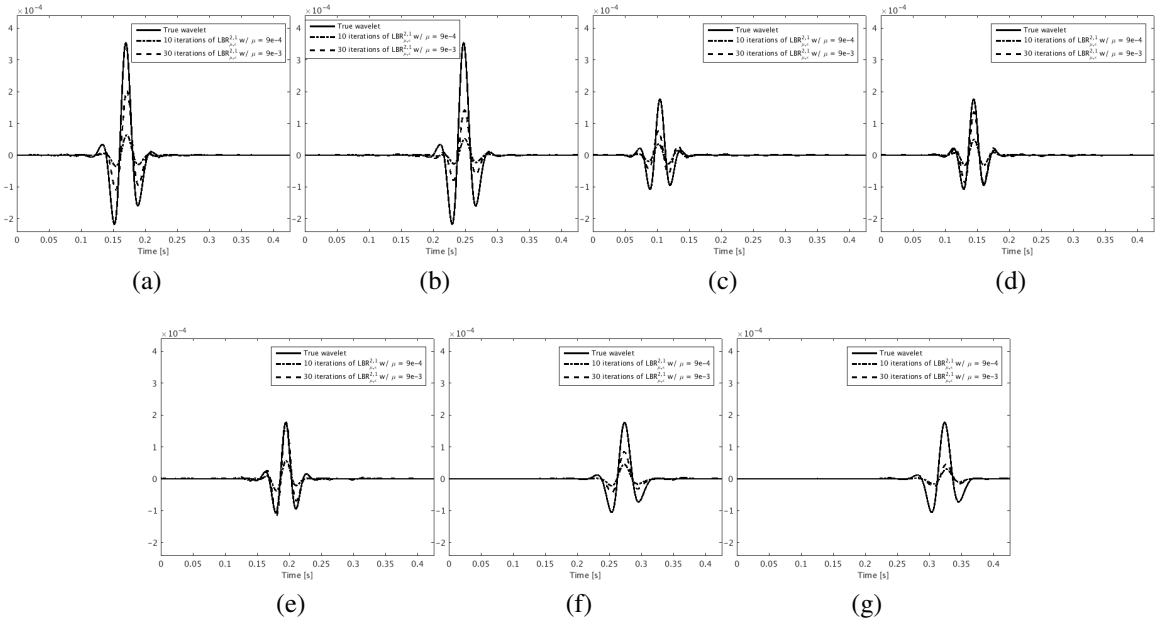


Figure 4.12: Comparison between the actual source-time functions (solid black) and the estimated source-time functions after 10 (denoted by dash dot line) and 30 (dashed) iterations of our method at the seven locations in *(a)*, *(b)*, *(c)*, *(d)*, *(e)*, *(f)* and *(g)* from left to right. The peak 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 $25.0\,\mathrm{Hz}$, in *(c)*, *(d)* and *(e)* $30.0\,\mathrm{Hz}$ and in *(f)* and *(g)* it is $22.0\,\mathrm{Hz}$. Estimated source-time functions are corrected by a factor of 6.4 for comparison purposes.

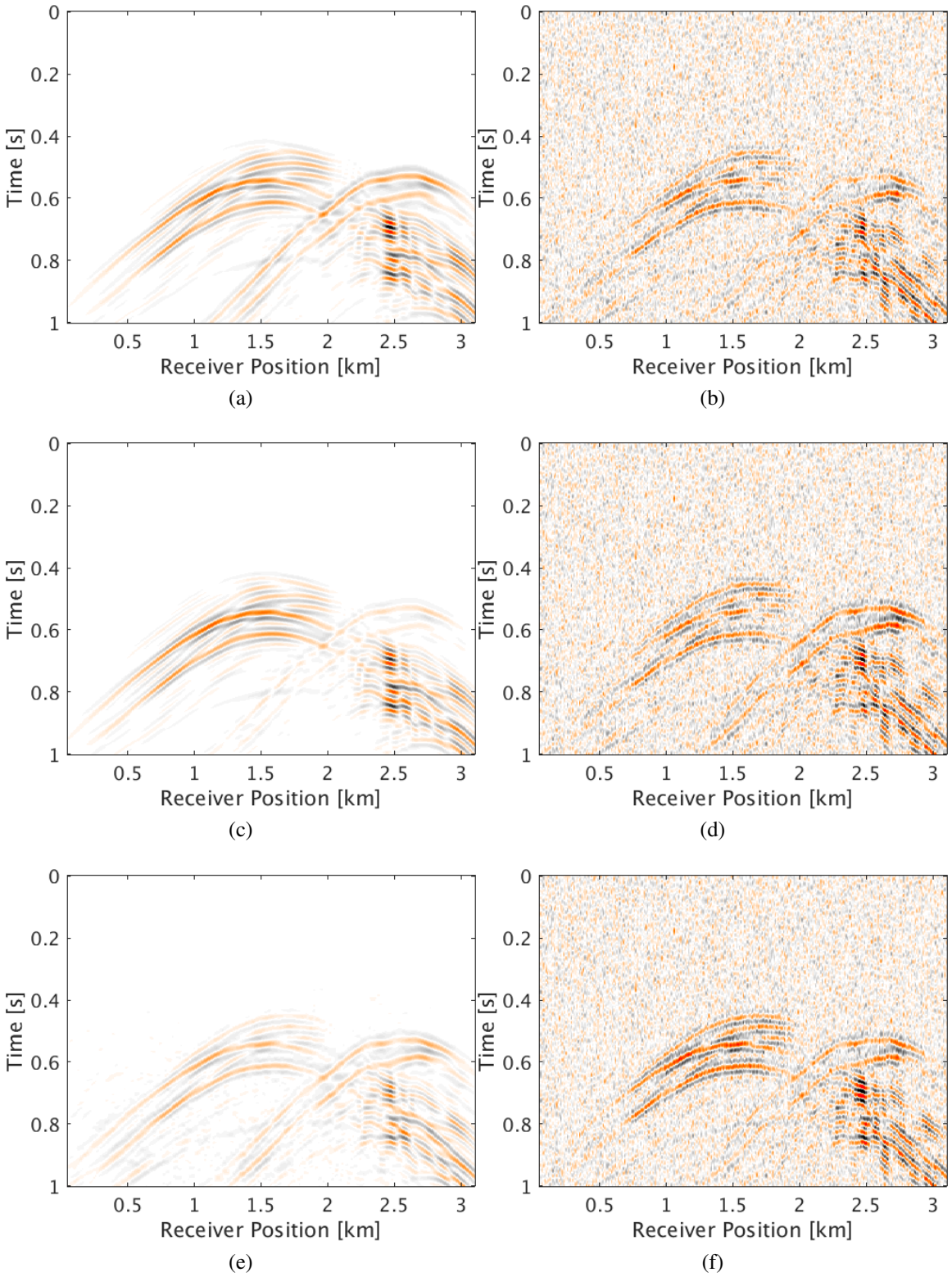


Figure 4.13: Modeled data (left column) and data residual (right column) in *(a,b)* by solving $\text{LBR}^{2,1}_{\mu,\epsilon}$ for $\mu = 9\text{e-}4$ with only 10 iterations; in *(c, d)* by solving $\text{LBR}^{2,1}_{\mu,\epsilon}$ for $\mu = 9\text{e-}3$ with 30 iterations; in *(e, f)* by solving LS_{ϵ} with 300 iterations. All plots are on the same color scale.

resolves closely 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 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 the events. We expect that including an amplitude term in our preconditioner in combination with an additional debiasing (Sharan, Kumar, Wang, and Herrmann 2018) step will remedy some of these amplitude issues. Debiasing involves running a few iterations of LSQR (Paige and 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, Romberg, and Tao 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, Xue, Fomel, Zhu, and Nakata 2016) and (Wu and 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.

4.7 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.

CHAPTER 5

A DEBIASING APPROACH TO MICROSEISMIC

5.1 Summary

Microseismic data is often used to locate fracture locations and their origin in time created by fracking. Although surface microseismic data can have large apertures and is easier to acquire than the borehole data, it often suffers from poor signal to noise ratio (S/R). Poor S/R poses a challenge in terms of estimating the correct location and source-time function of a microseismic source. In this work, we propose a denoising step in combination with a computationally cheap debiasing based approach to locate microseismic sources and to estimate their source-time functions with correct amplitude from extremely noisy data. Through numerical experiments, we demonstrate that our method can work with closely spaced microseismic sources with source-time functions of different peak amplitudes and frequencies. We have also shown the ability of our method with the smooth velocity model.

5.2 Introduction

To make unconventional reservoirs economical for the production of oil & gas, hydraulic fracturing is a common practice adapted by the oil & gas industry. During hydraulic fracturing fractures are created, which give rise to microseismic events. To make drilling decisions and to prevent hazardous situations, we need accurate information on the location and temporal evolution of these fractures. Because microseismic waves carry important information about fracture's location and origin time, the microseismic data recorded at surface or along a monitor well is often used to locate these fractures (Maxwell 2014).

Because of the operational ease and option to cover wide aperture, surface receivers are widely used (Duncan and Eisner 2010; Lakings, Duncan, Neale, and Theiner 2006). But the

microseismic data recorded along the surface comes at a cost of poor signal to noise ratio (S/R) in comparison to the data recorded along a monitor well. This is because we record more ambient noise (Forghani-Arani, Willis, Haines, Batzle, and Davidson 2012). Moreover, microseismic waves suffer from attenuation while travelling large distance from subsurface to the surface receivers (Maxwell, Raymer, Williams, and Primiero 2013), which makes them difficult to observe in noisy data. Thus, low S/R of surface microseismic data poses a big challenge in terms of estimation of accurate location and the origin time of microseismic sources. For example, travel-time picking based methods rely on accurate picking of first arrivals of P and S-phases. When the noise levels are high, it becomes difficult to accurately pick these first arrivals (Bolton and Masters 2001). Sometimes, the weak signal is not even visible to be picked.

In (Sharan, Wang, Leeuwen, and Herrmann 2016) & (Sharan, Wang, and Herrmann 2017), we proposed a computationally efficient sparsity promotion based method to invert for the microseismic source wavefield from which we extract locations and source-time functions of closely spaced microseismic sources. Our method works with noisy data with S/R as low as -1 dB, but performs poorly as the S/R decreases further. Moreover, while our sparsity-promotion based method is able to give us a good estimate of the shape of the source-time function, its amplitude is often incorrect. To overcome these limitations, we propose a debiasing approach to handle the noise and correctly estimate the amplitude of the source-time function. Our proposed approach consists of two main steps. The first step involves curvelet based denoising along with sparsity promotion based microseismic source inversion to detect the location of the microseismic sources. Since noise and signal has different morphological behaviour in curvelet domain (Candes and Donoho 2000) it is easy to separate the noise in curvelet domain (Herrmann and Hennenfent 2008b; Neelamani, Baumstein, Gillard, Hadidi, and Soroka 2008; Kumar, Moldoveanu, and Herrmann 2017). In the second step, we use the estimated source location and perform a wave-equation based debiasing step to get the source-time function with correct amplitudes.

This chapter is organized as follows. We first discuss the challenges in terms of detecting the location and estimating the source-time function of microseismic sources in the presence of strong incoherent noise in the observed data. Next, we explain the basics of curvelet transform and the steps we are using to denoise the noisy observed data. Subsequently, we explain the debiasing step to get the correct amplitude of the source-time function. Finally, we show the efficacy of the proposed approach on a noisy dataset generated on a complex subset of the Marmousi model (Brougois, Bourget, Lailly, Poulet, Ricarte, and Versteeg 1990).

5.3 Methodology

Fracturing of rocks during fracking causes emission of microseismic waves. The microseismic events causing this emission are mostly localized along the fracture tips. Therefore, we assume these microseismic sources to be sparse in space. In (Sharan, Wang, Leeuwen, and Herrmann 2016), we exploited the fact that microseismic sources are sparse in space and have finite energy along time to solve

$$\min_{\mathbf{Q}} \quad \|\mathbf{Q}\|_{2,1} \quad \text{subject to} \quad \|\mathcal{F}[\mathbf{m}](\mathbf{Q}) - \mathbf{d}\|_2 \leq \epsilon, \tag{5.1}$$

where $\mathbf{Q} \in \mathbb{R}^{n_x \times n_t}$, with n_x being the size of spatial grid and n_t being the number of time samples, is a matrix representing the complete microseismic source field containing the spatial temporal distribution of different microseismic sources —i.e. the (i, j) entry in $\mathbf{Q}_{i,j} = q(x_i, t_j)$. $\mathcal{F}[\mathbf{m}] = \mathbf{P}\mathcal{A}[\mathbf{m}]^{-1}$ is the linear operator modeling the 2D time-domain acoustic wave-equation. The linear operator $\mathbf{P}$ restricts the wavefield to the receivers. $\mathcal{A}[\mathbf{m}]$ is the 2D finite-difference time stepping operator parametrized by the squared slowness $\mathbf{m}$ of the medium. The minimizaton problem 5.1 aims to find such a $\mathbf{Q}$, which has a minimum ℓ_1-norm in space and has a minimum ℓ_2-norm in time while fitting the observed data $\mathbf{d}$ within the noise level ϵ. As mentioned earlier, sparseness of microseismic source wavefield

$\mathbf{Q}$ in space justifies the choice of ℓ_1-norm in space and finite energy of these microseismic sources justifies the choice of ℓ_2-norm in time in Equation 5.1.

Problem 5.1 has a form very similar to the classic Basis Pursuit Denoising (BPDN) problem (Chen, Donoho, and Saunders 1998; Berg and Friedlander 2008). In (Sharan, Wang, Leeuwen, and Herrmann 2016), we proposed a new algorithm tailored to solve a slightly modified version of the problem 5.1 for the situations when the forward modeling operator is ill conditioned and computationally expensive. We will discuss about this new algorithm in the next section.

5.3.1 Linearized Bregman algorithm

Motivated by the recent successful application of linearized Bregman algorithm (Yin, Osher, Goldfarb, and Darbon 2008; Lorenz, Schöpfer, and Wenger 2014) to solve sparsity promoting least-squares migration problem (Herrmann, Tu, and Esser 2015), in (Sharan, Wang, Leeuwen, and Herrmann 2016), we proposed to solve

$$\min_{\mathbf{Q}} \quad \|\mathbf{Q}\|_{2,1} + \frac{1}{2\mu}\|\mathbf{Q}\|_F^2 \quad \text{subject to} \quad \|\mathcal{F}[\mathbf{m}](\mathbf{Q}) - \mathbf{d}\|_2 \leq \epsilon, \tag{5.2}$$

which is strongly convex and a relaxed form of the original classic BPDN problem 5.1. $\|.\|_F$ is the Frobenius norm and μ acts as a trade-off parameter between sparsity given by $\ell_{2,1}$-norm term and the Frobenious norm term. When $\mu \uparrow \infty$, then Equation 5.2 is equivalent to solving the original BPDN problem 5.1. Solving Equation 5.2 can be achieved through a simple algorithm with few tuning parameters (Sharan, Wang, Leeuwen, and Herrmann 2016). We estimate the location of microseismic sources as outliers in the intensity plot calculated as $\mathbf{I}(\mathbf{x}) = \text{vec}^{-1}\left(\sum_t \mid \mathbf{Q}(\mathbf{x}, t) \mid\right)$ from the inverted source field $\mathbf{Q}$, where $\text{vec}^{-1}(\cdot)$ reshapes a vector into its original matrix form. The temporal variation of the inverted source field $\mathbf{Q}$ at the estimated source locations give the source-time function of microseismic sources.

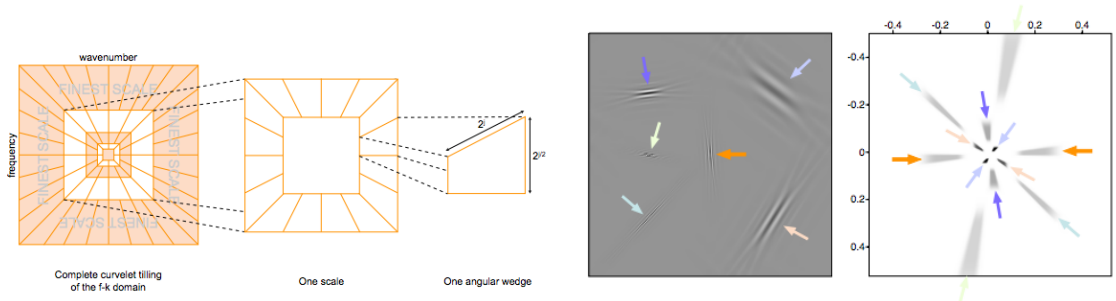
To avoid fitting noise in the data, every iteration of linearized Bregman algorithm involves projecting the data residual $\mathbf{r}$ (i.e. difference between predicted and observed data) on the ℓ_2 ball of size ϵ (Sharan, Wang, Leeuwen, and Herrmann 2016)

$$\Pi_\epsilon(\mathbf{r}) = \max\left\{0, 1 - \frac{\epsilon}{\|\mathbf{r}\|}\right\}\mathbf{r}. \tag{5.3}$$

Linearized Bregman algorithm performs well in locating microseismic sources and estimating their source-time function when the data has low to moderate levels of noise. But this is not always the case, microseismic data can be very noisy— i.e. $\epsilon \gg \|\mathbf{r}\|$. Higher value of ϵ implies that the projection of data residual in Equation 5.3 will give a vector with all zeros. Therefore, linearized Bregman fails to update the source field $\mathbf{Q}$ at every iteration. By using a smaller value of ϵ instead of the actual noise level, projection in the Equation 5.3 works and we can get source field $\mathbf{Q}$ updated in every iteration of linearized Bregman algorithm. But using a smaller value of ϵ instead of actual noise level in data means linearized Bregman algorithm will invert for such a source field $\mathbf{Q}$ which will also fit the noise. Hence, the inverted source field will have many false sources as we observe this phenomenon in the numerical experiment section. To avoid the above mentioned situations with very noisy data, we propose to incorporate curvelet based denoising step prior to applying linearized Bregman algorithm to invert for the location followed by a wave-equation based debiasing approach to get the source-time function with correct amplitude.

5.3.2 Curvelet based denoising

Microseismic signals recorded by the surface receivers are very weak in amplitude and are often contaminated with ambient noise that have similar or higher amplitude level than the amplitude of the microseismic signal present in the data (Forghani-Arani, Willis, Haines, Batzle, and Davidson 2012). Also, the frequency range of the ambient noise is very similar to that of the microseismic signal (St-Onge and Eaton 2011). This makes signal and



(a) (b)

Figure 5.1: Properties of the curvelet transform (Source: (Herrmann and Hennenfent 2008b))

noise separation very difficult and eventually causes problems in detecting the microseismic sources and estimating their source-time functions. Curvelet transform is a multi-scale and multi-directional transform (Candes and Donoho 2000), that maps seismic data into angular wedges of different scales in the 2D Fourier domain (Figure 5.1). This property of the curvelet transform helps in separating signals components based on their location, dip and scaling in the transform domain. Therefore, curvelet transform has been succefully used for incoherent (Herrmann and Hennenfent 2008b; Neelamani, Baumstein, Gillard, Hadidi, and Soroka 2008) and coherent noise attenuation (Kumar, Moldoveanu, and Herrmann 2017; Lin and Herrmann 2013).

Motivated by prior successful application of curvelets, we propose following steps for denoising:

Algorithm 5.1 Denoising with curvelets.

1. Noisy data $\mathbf{d}$, forward curvelet transform operator $\mathbf{C}$, Threshold parameter λ `//Input`
2. $\mathbf{b} = \mathbf{C}\mathbf{d}$ `//Forward curvelet transform`
3. $[\mathbf{sb}, \mathbf{idx}] = \text{Sort}(|\ \mathbf{b}\ |)$ `//Sorting in descending order`
   where $\mathbf{idx}$ stores the indices of sorted curvelet coefficients
4. $(\mathbf{e})_h = \sqrt{\frac{\sum_{i=1}^{h} \mathbf{sb}_i^2}{\sum_{i=1}^{l} \mathbf{sb}_i^2}}$ `//normalized cumulative energy`
   where l is the length of $\mathbf{sb}$
5. Find the smallest index p such that $(\mathbf{e})_p \geq \lambda$
6. $\mathbf{R} = \mathbf{C}^\top(\mathbf{idx}(1:p), :)$ `//New inverse curvelet transform operator`
7. $\mathbf{b}_{dn} = (\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top\mathbf{d}$ `//Solving the normal equation`
8. $\mathbf{d}_{dn} = \Re(\mathbf{R}\mathbf{b}_{dn})$ `//denoising`

Line 2 in the Algorithm 5.1 corresponds to forward curvelet transform of the noisy microseismic data $\mathbf{d}$ (Figure 5.2a) in the curvelet domain (Figure 5.2b). The indices in Figure 5.2b are arranged from coarse to fine scale. **Line 3** corresponds to sorting of the absolute value of curvelet coefficients $\mathbf{sb}$ in descending order. We also store the indices of the sorted curvelet coefficients in $\mathbf{idx}$. **Line 4** corresponds to computing the square root of normalized cumulative energy of the sorted curvelet coefficients. **Line 5** corresponds to finding the smallest index p in vector $\mathbf{e}$ at which the square root of the normalized cumulative energy of sorted curvelet coefficient exceeds or is equal to the threshold λ. In **line 6**, we form a subset $\mathbf{R} \subseteq \mathbf{C}^\top$ of the inverse curvelet transform operator. Columns of $\mathbf{R}$ correspond to the curvelet coefficients whose square root of the normalized cumulative energy in $\mathbf{e}$ is greater than or equal to the threshold λ. **Line 7** corresponds to solving the normal equation to get the debiased and denoised curvelet coefficients $\mathbf{b}_{dn}$ effectuated by the new inverse curvelet transform operator $\mathbf{R}$. Debiasing neutralizes the shrinkage effect of thresholding and preserves energy. Figure 5.2c shows absolute value of the denoised and debiased curvelet coefficients $\mathbf{b}_{dn}$ mapped to the corresponding location of the noisy curvelet coefficients $\mathbf{b}$. **Line 8** corresponds to taking inverse curvelet transform of the denoised curvelet coefficients $\mathbf{b}_{dn}$ and taking its real part to get the denoised microseismic data (Figure 5.2d). We choose the threshold parameter as large as possible but for which we do not see any primary leakage in the difference plot.

The curvelet based denoising involves very few forward and inverse curvelet transform, which makes the proposed denoising method computationally cheap. To detect the location of microseismic sources from the denoised data $\mathbf{d}_{dn}$ in a computationally efficient manner, we use the accelerated version of linearized Bregman algorithm (Sharan, Wang, and Herrmann 2017) along with a 2D left preconditioner (Herrmann, Brown, Erlangga, and Moghaddam 2009).

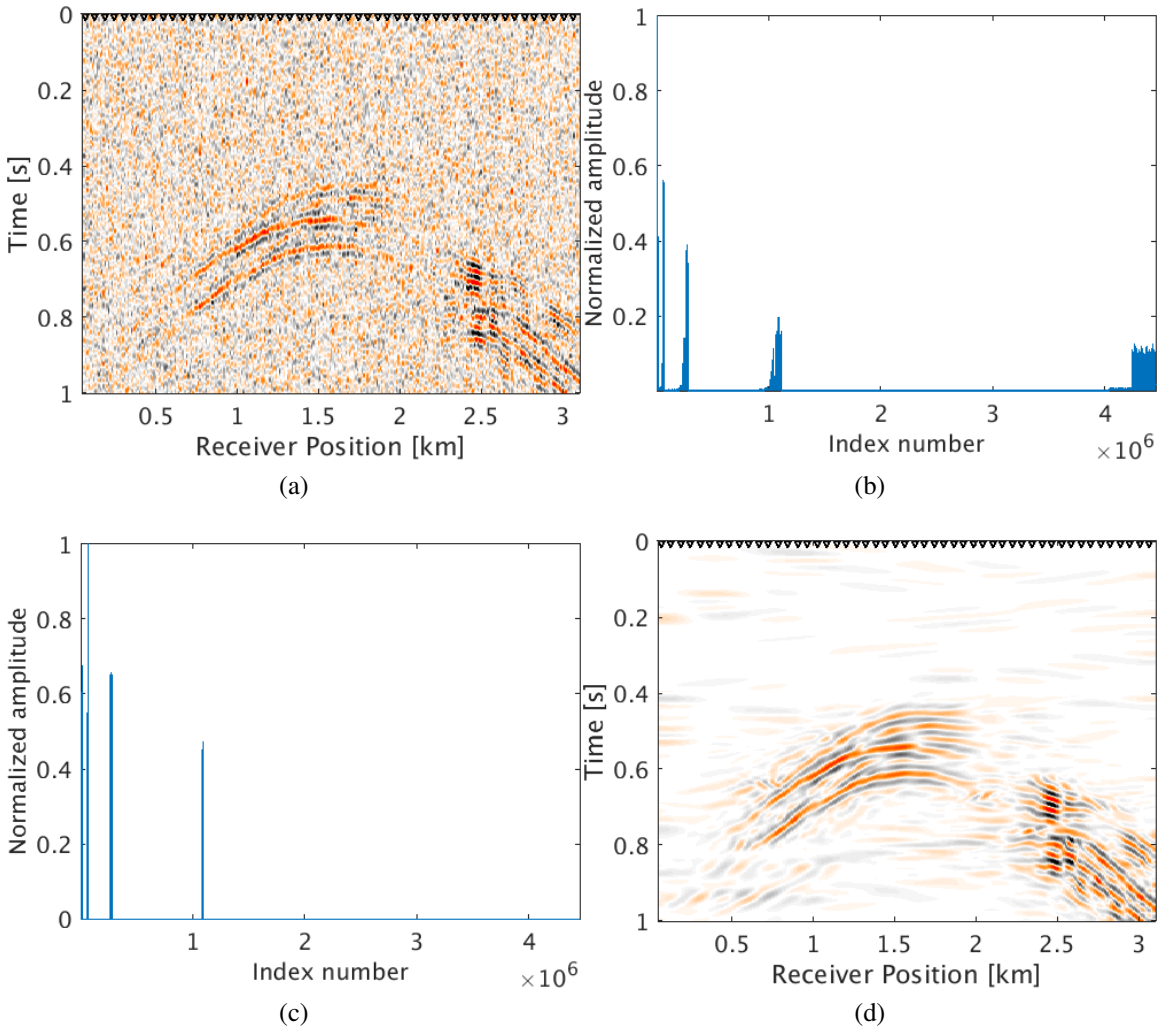


Figure 5.2: Curvelet denoising schematic. *(a)* Noisy microseismic data with S/R = -5.70 dB. *(b)* Absolute value noisy curvelet coefficient. *(c)* Absolute value debiased curvelet coefficients. *(d)* Denoised microseismic data with S/R = 5.3 dB.

5.3.3 Debiasing of the source-time function

Given the location of microseismic sources, next step is to estimate the correct amplitude of source-time function. To achieve this, we use the forward modeling operator $\mathcal{F}[\mathbf{m}]$ and estimate source locations to fit the noisy data $\mathbf{d}$ within some tolerance level. We use noisy data to avoid any kind of amplitude errors introduced in the approximated data by denoising. We now solve a debiasing problem using least squares as

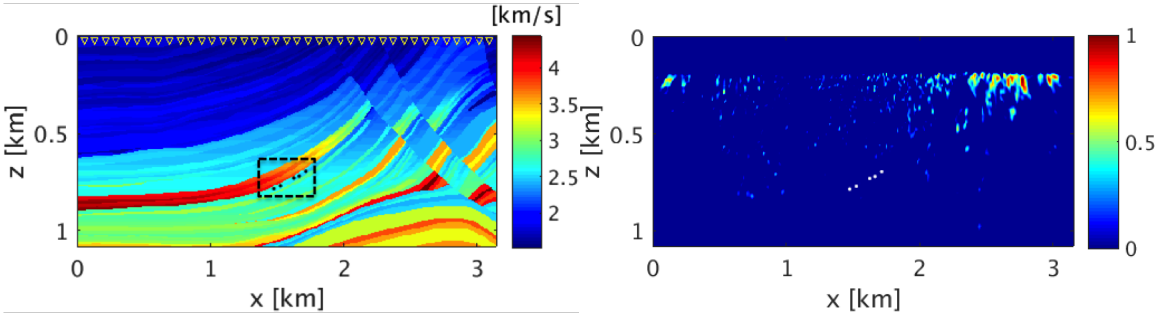
$$\tilde{\mathbf{W}} = \underset{\mathbf{W}\in\mathbb{R}^{n_t\times n}}{\arg\min} \quad \|\mathcal{F}[\mathbf{m}](\mathbf{H}\mathbf{W}^\top) - \mathbf{d}\|, \tag{5.4}$$

where $\mathbf{H} \in \mathbb{R}^{n_x\times n}$ is a matrix, with n being the number of detected microseismic sources, whose i^{th} column is $\mathbf{h}_i$,which corresponds to the location of the i^{th} source. $\mathbf{h}_i$ is a spatial delta function $\delta(\mathbf{x}-\mathbf{x}_i)$ with $\mathbf{x}_i$ being the location of i^{th} microseismic source. Equation 5.4 solves for the unknown matrix $\mathbf{W}$ whose i^{th} column corresponds to the source-time function of i^{th} microseismic source. We run only a few iterations of the unconstrained problem 5.4 to avoid overfitting the noise in the data.

5.4 Numerical Experiments

To demonstrate the effectiveness of our method for data with high noise level (S/R = -7.30 dB), we performed a numerical experiment. We used 2D acoustic finite-difference modeling code (Louboutin, Witte, Lange, Kukreja, Luporini, Gorman, and Herrmann 2017) to generate microseismic data of record length $1.0\,\mathrm{s}$. To make the experimental setup more realistic, we used 5 microseismic sources of different amplitudes (differ by a factor of 2) and dominant frequencies $(30.0$, and $25.0)$ Hz activating at a small time interval with overlapping source-time functions to generate the microseismic data. Because of its geological complexity, we chose a part of Marmousi model with dimensions $3.15\,\mathrm{km} \times 1.08\,\mathrm{km}$ (631×217 points) to perform the experiment. We place 5 microseismic sources (indicated by black dots in Figure 5.3a) in low velocity layer to generate the data. The adjacent

sources are separated by half a dominant wavelength. We use surface receivers placed at a depth of $20.0\,\mathrm{m}$ from the top surface to record the data. To get noisy data (Figure 5.4a) we add random noise ($5.0\,\mathrm{Hz}$ to $40.0\,\mathrm{Hz}$) to the noise free data. We use kinematically correct smooth velocity model to invert for microseismic source field in the experiment. As expected, our method performs poorly without curvelet denoising and gives an intensity plot (Figure 5.3b) that is not very informative. This is because of the presence of lots of false sources in the estimated intensity plot (Figure 5.3b). Therefore, we apply the proposed curvelet based denoising steps to the noisy data (Figure 5.4a) to get denoised data (Figure 5.4b) with improved S/R of 3.5 dB. The difference plot (Figure 5.4c) between noisy (Figure 5.4a) and the denoised data (Figure 5.4b) shows that we do not loose any coherent signal with the proposed denoising method. With only 10 iterations of accelerated version of linearized Bregman algorithm, we are now able to locate all the 5 microseismic sources (Figure 5.4d) from this denoised data (Figure 5.4b). The white colour crosses in the estimated intensity plot correspond to the actual location of microseismic sources. All the outliers are located near the actual location of microseismic sources. To get the correct source-time function (blue colour plot in Figures 5.5a to 5.5e), we perform debiasing by least-squares using the estimated source location. We use the noisy data to perform this debiasing. Denoising helps us to get the correct source location and the debiasing step helps us to recover the source-time function with correct scaling. We further compare the source-time function (blue colour plot in Figures 5.5a to 5.5e) estimated by proposed approach to the source-time function estimated by (Sharan, Wang, and Herrmann 2017) (red colour plot in Figures 5.5a to 5.5e). For visualization purpose, we scale the wavelets displayed in red color by a factor of 40. Thus, the proposed method can estimate the location and source-time function with correct scaling for microseismic data acquired in extremely noisy environment.



(a) (b)

Figure 5.3: *(a)* Acquisition geometry with velocity model. Inverted yellow colour triangles indicate receivers buried at a depth of $20.0\,\mathrm{m}$ & separated by $10.0\,\mathrm{m}$. Black dots indicate the location of two microseismic sources. *(b)* Estimated intensity plot from noisy data without denoising. White colour dots indicate actual location of microseismic sources.

5.5 Conclusions

We proposed a debiasing based approach to estimate the location and source-time function of microseismic sources with correct amplitude from data with very low S/R. We showed ability of the proposed method to resolve microseismic events even when the sources are spatially close and have overlapping source-time functions. The proposed method is also computationally cheap as it requires very few forward and backward curvelet transforms along with few iterations of the accelerated version of linearized Bregman. Also, the source-time function estimation requires only a very few least-squares iterations. In future, we would like to apply PCA based denoising techniques to deal with different types of noise such as ground roll, source side noise etc.

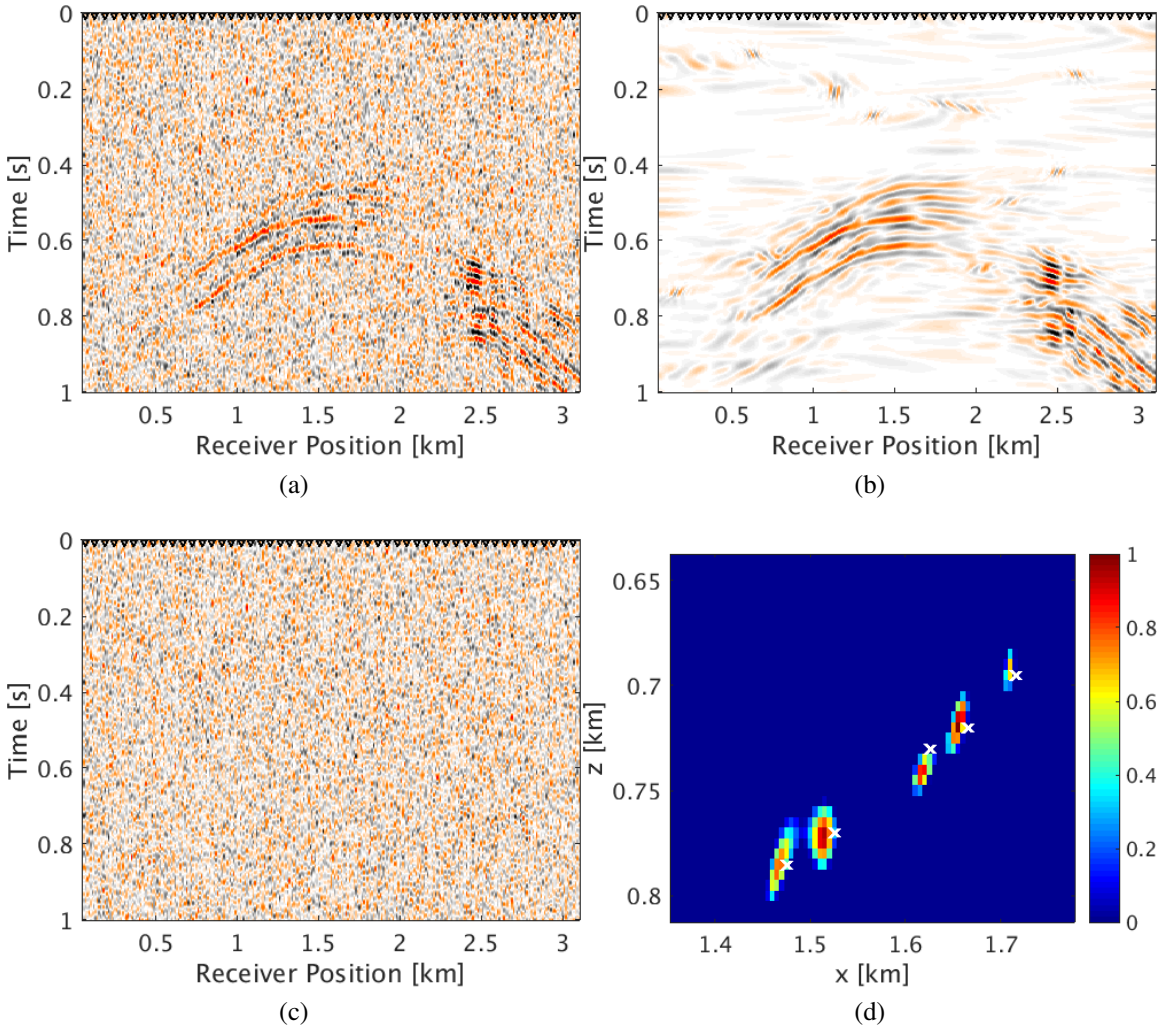


Figure 5.4: Noisy microseismic data and estimated intensity plots(zoomed): noisy data with *(a)* S/R = -7.3 dB. *(b)* Denoised data using curvelet based denoising with improved S/R of 3.5 dB. *(c)* Data difference plots after denoising. *(d)* Estimated intensity plot. White colour crosses indicate the true location of microseismic sources.

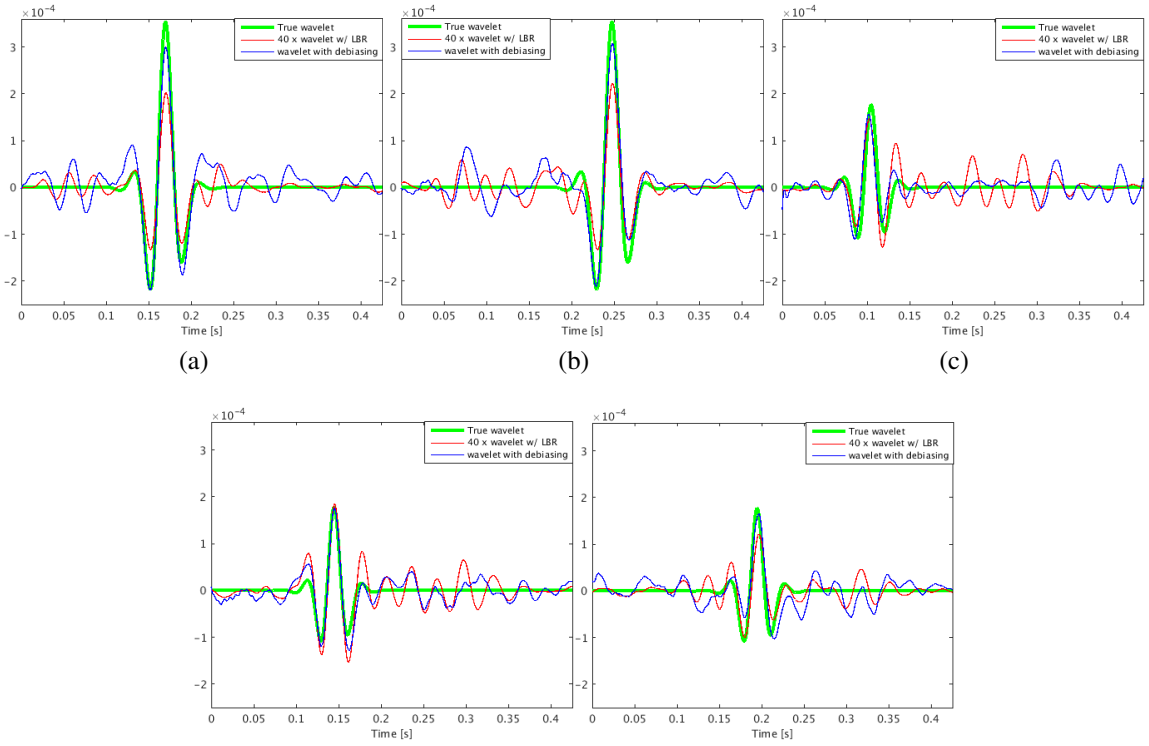


Figure 5.5: Source-time function comparison: Comparison of the true source-time functions (solid green) with source-time function (blue color) estimated by proposed method. We also perform comparison with the source-time function (amplified by 40 times) estimated using the approach proposed in (Sharan, Wang, and Herrmann 2017) (solid magenta) at locations *(a)* 1, *(b)* 2, *(c)* 3, *(d)* 4, *(e)* 5 from LtoR in Figure 5.3a. Dominant frequency of source-time functions at (from LtoR) locations 1 and 2 is 25.0 Hz, at locations 3, 4 and 5 dominant frequency is 30.0 Hz.

CHAPTER 6
SPARSITY-PROMOTING PHOTOACOUSTIC IMAGING WITH SOURCE ESTIMATION

6.1 Summary

Photoacoustics has emerged as a high-contrast imaging modality that provides optical absorption maps inside of tissues, therefore complementing morphological information of conventional ultrasound. The laser-generated photoacoustic waves are usually envelope-detected, thus disregarding the specific waveforms generated by each photoabsorber. Here we propose a sparsity-promoting image reconstruction method that allows the estimation of each photoabsorber's source-time function. Preliminary studies showed the ability to reconstruct the optical absorption map of an in silico vessel phantom. By using a sparsity-promoting imaging method, absorption maps and source-time functions can still be recovered even in situations where the number of transducers is decreased by a factor of six. Moreover, the recovery is able to attain higher resolution than conventional beamforming methods. Because our method recovers the source-time function of the absorbers, it could potentially also be used to distinguish different types of photoabsorbers, or the degree of aggregation of exogenous agents, under the assumption that these would generate different source-time functions at the moment of laser irradiation.

6.2 Introduction

Applications of compressive sensing techniques in ultrasound and photoacoustic imaging have seen an increase in the last decade (Liebgott, Basarab, Kouame, Bernard, and Friboulet 2012; Provost and Lesage 2009; Kruizinga, Meulen, Fedjajevs, Mastik, Springeling, Jong, Bosch, and Leus 2017). These techniques enable the estimation of waveform char-

acteristics that are not readily available via conventional beamforming methods. Through the use of convex optimization techniques that promote sparsity, the frequency content of the ultrasound waves can be recovered beyond the Nyquist limit—i.e., from recordings of the induced wavefield that are well below the spatial Nyquist frequency.

In photoacoustic imaging, the optical absorption maps are usually created from envelope-detected waves (Park, Karpiouk, Aglyamov, and Emelianov 2008). While conventional beamforming techniques offer advantages, including real-time image reconstruction, they often disregard the frequency content of the measured waves. Additionally, the available frequency content is always band-limited by the receiving transducers. In ultrasound imaging, several studies have shown that the frequency content of raw pressure waves can provide information of tissue composition, in applications such as ophthalmic, rectal, and intravascular imaging (Lizzi, Greenebaum, Feleppa, Elbaum, and Coleman 1983; Feleppa, Kalisz, Sokil-Melgar, Lizzi, Tian Liu, Rosado, Shao, Fair, Yu Wang, Cookson, Reuter, and Heston 1996; Lizzi 1997; Moore, Spencer, Salter, Kearney, Shaw, Starkey, Fitzgerald, Erbel, Lange, McDicken, Sutherland, and Fox 1998). Similar concepts have been explored in photoacoustics (Kumon, Deng, and Wang 2011), denoting that there is a potential use of the photoacoustic wave frequency content. The study of RF photoacoustic techniques could further enhance the functional and molecular imaging capabilities of this modality (Emelianov, Li, and O'Donnell 2009).

Recently, a sparsity-promoting algorithm for the estimation of full-wave source-time functions was developed as a method to localize microseismic events for geological exploration (Sharan, Wang, and Herrmann 2018). In photoacoustic imaging, the localization of strong photoabsorbers in tissue represents an analog problem albeit that in microseismic the source firing times differ for each source and are not known. The fact that these source-time functions are unknown make the microseismic problem more challenging because it leads to a large non-trivial nullspace—i.e., there exist so-called non-radiating sources that correspond to sources that do not contribute to wavefield measured at the receivers. Con-

ventional photoacoustic imaging overcomes this problem by assuming the absorbers to fire synchronously. As a result, high-fidelity images can be obtained by time reversal, followed by extraction of the back propagated wavefield at $t = 0$. While this method has proven to be highly succesful, it relies on dense sampling and does not provide information on the source-time function of the different absorbers. Our sparsity-promoting method on the other hand, is able to handle the large null space and as such can handle sparse samplings of the wavefield while providing information on the source-time functions. The latter could potentially be used for tissue and contrast agent characterization.

Our contribution is organized as follows. First, we provide an overview of our methodology including statement of the problem, solution by linearized Bregman, and acceleration with a dual formulation. Next, we demonstrate our method on a phantom in a constant velocity model, yielding an estimate for the phantom and the source-time function. We follow this result by a series of experiments where the number of transducers (receivers) is reduced significantly. We compare our results to those obtained with time reversal. We conclude by showing an example with a background velocity model that varies strongly.

6.3 Methodology

Contrary to conventional time-reversal methods, our unknown is a wavefield across the domain of interest and a function of time without making assumptions on the source waveform. Since the sources are of short duration firing at approximately the same time, we restrict the unknown wave by putting its entries to zero after a user specified number of time samples. Since the number of absorbers is small, we assume sparsity in space and finite energy along time. The latter makes sense because the photoabsorbers emit a finite amount of energy. To image the wavefield induced by these sparsely distributed absorbers, we solve

$$\min_{\mathbf{Q}\in\mathcal{T}_\tau} \quad \|\mathbf{Q}\|_{2,1} \quad \text{subject to} \quad \|\mathcal{F}[\mathbf{m}](\mathbf{Q}) - \mathbf{d}\|_2 \leq \epsilon, \tag{6.1}$$

with $\mathbf{Q} \in \mathbb{R}^{n_x \times n_t}$ being the unknown source wavefield, which we restrict to a user defined duration τ—i.e, $\mathcal{T}_\tau = \{\mathbf{Q} \,|\, \mathbf{Q}(\cdot, t) = 0,\, t > \tau\}$. The source wavefield in (6.1) consists of n_x grid points in space and n_t time samples. The matrix $\mathcal{F}[\mathbf{m}]$ represents the acoustic forward modeling operator parametrized by $\mathbf{m}$ the discretized squares slowness. Slowness is defined as the inverse of the acoustic wave speed.

Solving (6.1) corresponds to minimizing the $\ell_{2,1}$-norm of the unknown source wavefield $\mathbf{Q}$ while fitting the observed data $\mathbf{d}$ within ϵ, which depends on the noise level. After solving for $\mathbf{Q}$, we detect the location of sources as outliers in the intensity plot $\mathbf{I}(\mathbf{x}) = \mathrm{vec}^{-1}\left(\mathbf{Q}(\mathbf{x}, t = t_0)\right)$, where $\mathrm{vec}^{-1}(\cdot)$ reshapes the vector into its original matrix form and t_0 is the firing time (typically the maximum of the source-time function). We finally estimate the associated source-time functions as the temporal variation extracted from the estimated source wavefield $\mathbf{Q}$ at the detected source locations.

6.3.1 Linearized Bregman Algorithm

Equation (6.1) is similar to the classic basis pursuit denoising (BPDN) problem (Chen, Donoho, and Saunders 1998) involving a sparsity-promoting objective and a data constraint. With the recent successful application of the linearized Bregman algorithm (Yin, Osher, Goldfarb, and Darbon 2008), (Lorenz, Schöpfer, and Wenger 2014) to seismic imaging (Witte, Louboutin, Luporini, Gorman, and Herrmann 2019) and microseismic source estimation problems (Sharan, Wang, and Herrmann 2018), we use this method to solve BPDN via a relaxed form that is strongly convex. Following (Sharan, Wang, and Herrmann 2018), we minimize

$$\min_{\mathbf{Q} \in \mathcal{T}_\tau} \quad \|\mathbf{Q}\|_{2,1} + \frac{1}{2\mu}\|\mathbf{Q}\|_F^2 \quad \text{subject to} \quad \|\mathcal{F}[\mathbf{m}](\mathbf{Q}) - \mathbf{d}\|_2 \leq \epsilon, \tag{6.2}$$

where μ is a tradeoff parameter between the sparsity $\ell_{2,1}$-norm and the Frobenius norm $\|.\|_F$. As $\mu \uparrow \inf$, (6.2) becomes equivalent to (6.1). Therefore, for large enough μ, the solution of (6.2) approaches the solution of (6.1), which in principle should give us a high

resolution photoabsorber image. Unfortunately, for increasing values of μ, the linearized Bregman algorithm requires more iterations making this method prohibitively expensive because each iteration requires one forward and one time-reversed simulation (Sharan, Wang, and Herrmann 2018). We denote the adjoint by the superscript $^\top$.

6.3.2 Acceleration using a dual formulation

Reference (Yin 2010) showed that solving the original problem (6.2) through linearized Bregman iterations is equivalent to solving its dual formulation through gradient descent steps. Therefore, we use gradient descent acceleration method such as spectral projected gradient (SPG) method to accelerate the convergence of the linearized Bregman algorithm. To arrive at this accelerated algorithm, we derive the Fenchel's dual of problem (6.2) yielding

$$\min_{\mathbf{y}} \quad f(\mathbf{y}) = -\left\{\Psi(\mathbf{y}) - \epsilon\|\mathbf{y}\|_2\right\}, \tag{6.3}$$

where $\mathbf{y}$ is the dual variable and

$$\Psi(\mathbf{y}) = \min_{\mathbf{Q}\in\mathcal{T}_\tau} \quad \|\mathbf{Q}\|_{2,1} + \frac{1}{2\mu}\|\mathbf{Q}\|_F^2 + \mathbf{y}^\top\big(\mathcal{F}[\mathbf{m}](\mathbf{Q}) - \mathbf{d}\big), \tag{6.4}$$

is a value function of the dual variable $\mathbf{y}$. The dual objective function gives the minima of the objective function defined in (6.3) as a function of the dual variable $\mathbf{y}$. We derive the gradient $\nabla f(\mathbf{y})$ of the dual objective function $f(\mathbf{y})$ as

$$\nabla f(\mathbf{y}) = -\left\{\nabla\Psi(\mathbf{y}) - \epsilon\mathbf{y}/\|\mathbf{y}\|_2\right\}, \tag{6.5}$$

where

$$\nabla\Psi(\mathbf{y}) = \mathbf{d} - \mathcal{F}[\mathbf{m}]\big(\text{Prox}_{\mu\ell_{2,1}}(\mu\mathcal{F}[\mathbf{m}]^\top(\mathbf{y})\big) \tag{6.6}$$

is the gradient of the value function $\Psi(\mathbf{y})$. The proximal operator in this expression is equivalent to a thresholding operation and its action on a matrix $\mathbf{C}$ is defined as:

$$\text{Prox}_{\mu\ell_{2,1}}(\mathbf{C}) := \arg\min_{\mathbf{B}} \quad \|\mathbf{B}\|_{2,1} + \frac{1}{2\mu}\|\mathbf{C}-\mathbf{B}\|_F^2 \,. \tag{6.7}$$

The main steps of the linearized Bregman with acceleration are summarized in Algorithm (6.1).

Algorithm 6.1 Acceleration with SPG.

1. Data $\mathbf{d}$, slowness square $\mathbf{m}$, number of iterations l `//Input`
2. Initialize dual variable $\mathbf{y} = 10^{-3}\mathbf{d}$
3. $\hat{\mathbf{y}} = \text{SPG}(f(\mathbf{y}), \nabla f(\mathbf{y}), \mathbf{y}, l)$ `//Dual solution`
4. $\mathbf{Q} = \text{Prox}_{\mu\ell_{2,1}}(\mu\mathcal{F}[\mathbf{m}]^\top(\mathbf{y}))$ `//Primal solution`

6.4 Experiments

To demonstrate our ability to jointly image and estimate the source-time function, we performed simulation experiments in acoustic medium using JUDI (Witte, Louboutin, Kukreja, Luporini, Lange, Gorman, and Herrmann 2019) and Devito (Louboutin, Lange, Luporini, Kukreja, Witte, Herrmann, Velesko, and Gorman 2019) - (Luporini, Lange, Louboutin, Kukreja, Hückelheim, Yount, Witte, Kelly, Gorman, and Herrmann 2018), a high performance finite difference partial differential equation solver and the k-Wave MATLAB toolbox (Treeby and Cox 2010). To avoid problems with the boundaries of the domain, we use derivatives of the Gaussian as the source-time function. In this way, we avoid low frequencies to enter into the solution so we can limit the size of the absorbing boundary layer. By casting photoacoustic imaging as a sparsity-promoting imaging problem, we are able to estimate the source-time function from poorly sampled wavefields. To illustrate this, we first carry out an imaging experiment in a constant velocity model that is densely sampled and show that we can indeed estimate the source-time function. Next, we compare images obtained with the proposed method and with the back-propagation

method for increasingly poor sampling, followed by an example in a strongly heterogeneous medium. For all experiments, we use Devito to generate data. For the time reversal method, we generate data using k-Wave MATLAB toolbox, which has a broader frequency content in comparison to the data we use.

6.4.1 Image and source-function recovery from dense data

Figure 6.1a shows the actual location of photoabsorbers (solid white color) and white color dots indicate the transducers. The background is constant with an acoustic wave velocity of $1500\,\mathrm{m/s}$. We obtain a high resolution image in Figure 6.1b with the proposed method. We are also able to reconstruct the source-time functions (red color plots in Figures 6.1c and 6.1d), which is close to the true source-time function (cf. Figures 6.1c and 6.1d).

6.4.2 Image recovery from subsampled data

Figures 6.2a and 6.2b contain photoabsorber images we reconstructed with the proposed method from transducers sampled at every two degrees and every six degrees. Although we see some background noise when the transducer sampling becomes poor, the proposed method is still able to get a reasonable high resolution image in comparison to the image obtained using the back-propagation method of the k-wave toolbox (cf. Figures 6.2c and 6.2d). As the sparsity of transducers increases, we observe blurring of the images obtained with time reversal.

6.4.3 Image recovery in strongly heterogenous media

In this experiment, we show imaging results and source-time reconstruction using a kinematically correct smooth velocity model and data simulated with the hard model in Figure 6.3a with a background wave speed of $1500\,\mathrm{m/s}$ and two lobes with $1575\,\mathrm{m/s}$ and $1650\,\mathrm{m/s}$, repectively. We obtain the image plotted in Figure 6.3c with a smooth velocity model with the proposed method. Although, the image obtained with our method is

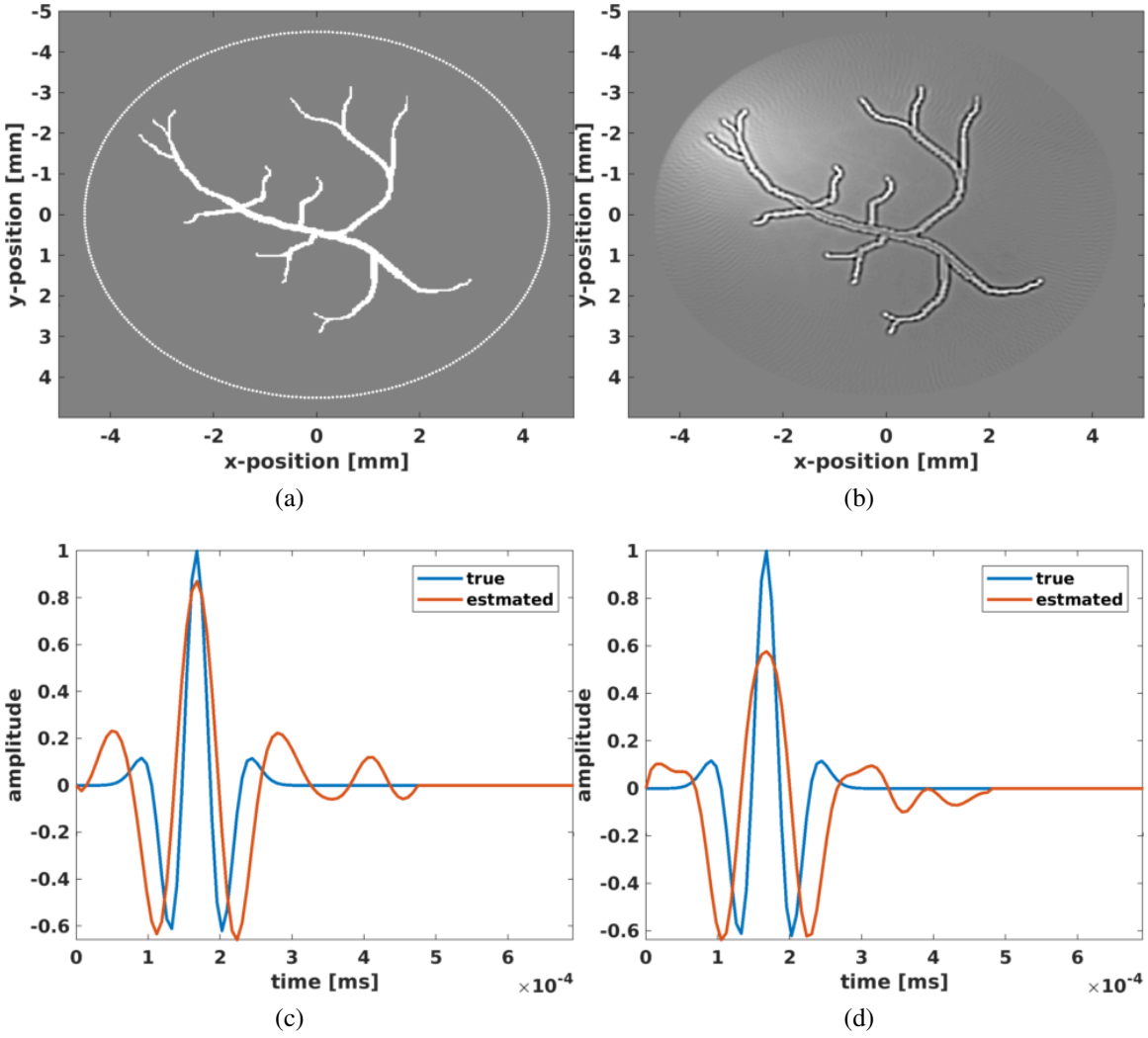


Figure 6.1: (a) Data acquisition with locations of the photoabsorbers denoted by the solid white color phantom and transducers by white color dots (b) Image reconstructed from fully sampled data using our proposed method. (c) and (d) show a source-time function comparison for two locations.

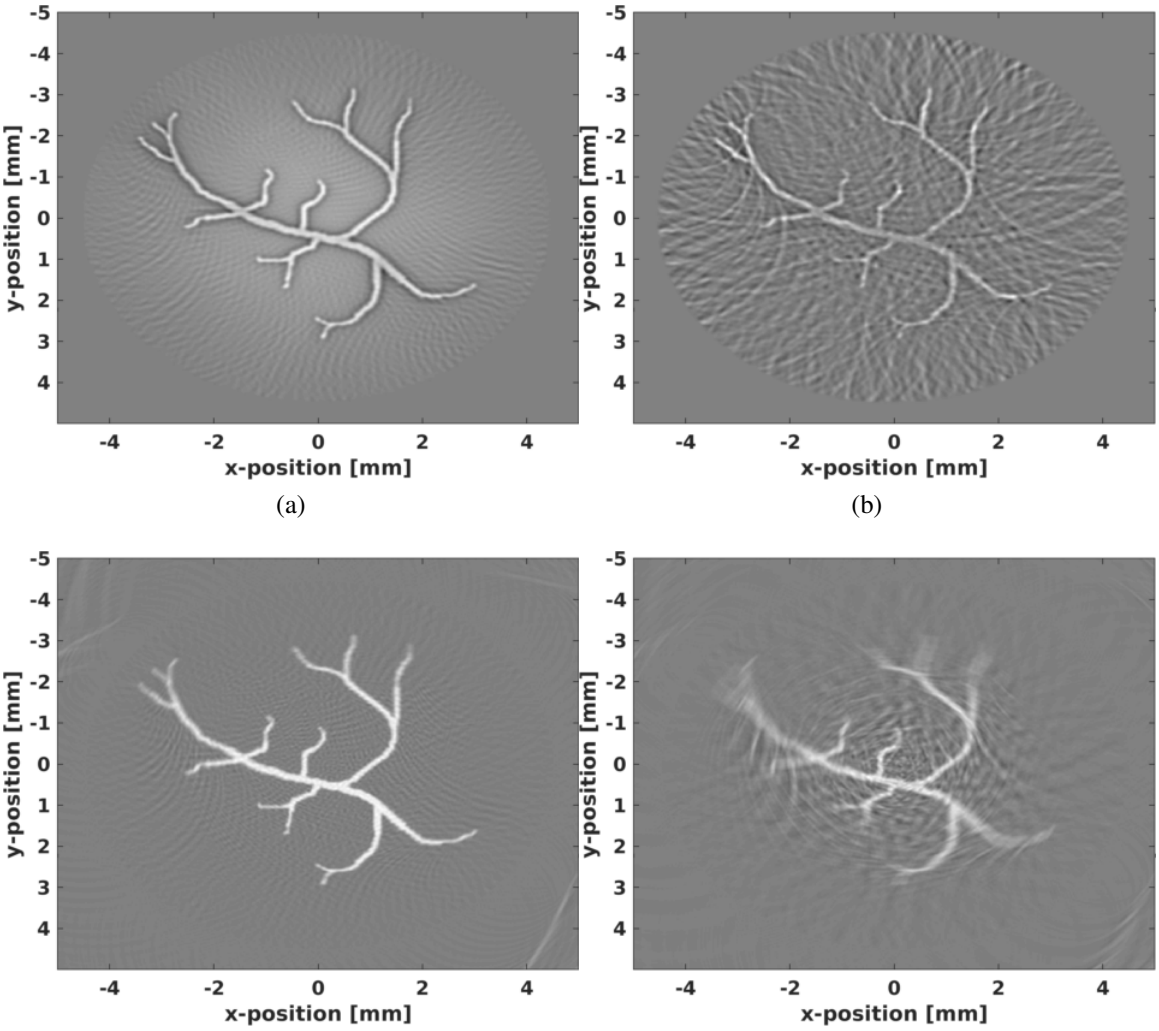


Figure 6.2: Receiver sparsity experiment: Reconstructed image with transducers at every (a) 2 degrees and (b) 6 degrees using proposed method. Image with transducers at every (c) 2 degrees and (d) 6 degrees using k-wave toolbox

somewhat lower in resolution in comparison to the images obtained with time reversal (Figure 6.3d). This is because of the higher frequency content of the k-wave data. Still, the proposed method gives a good estimate for the source-time function when using the original hard model and the smoothed model (cf. red and yellow lines in Figure 6.3b). The amplitudes of the estimated source-time functions can be corrected with a debiasing step.

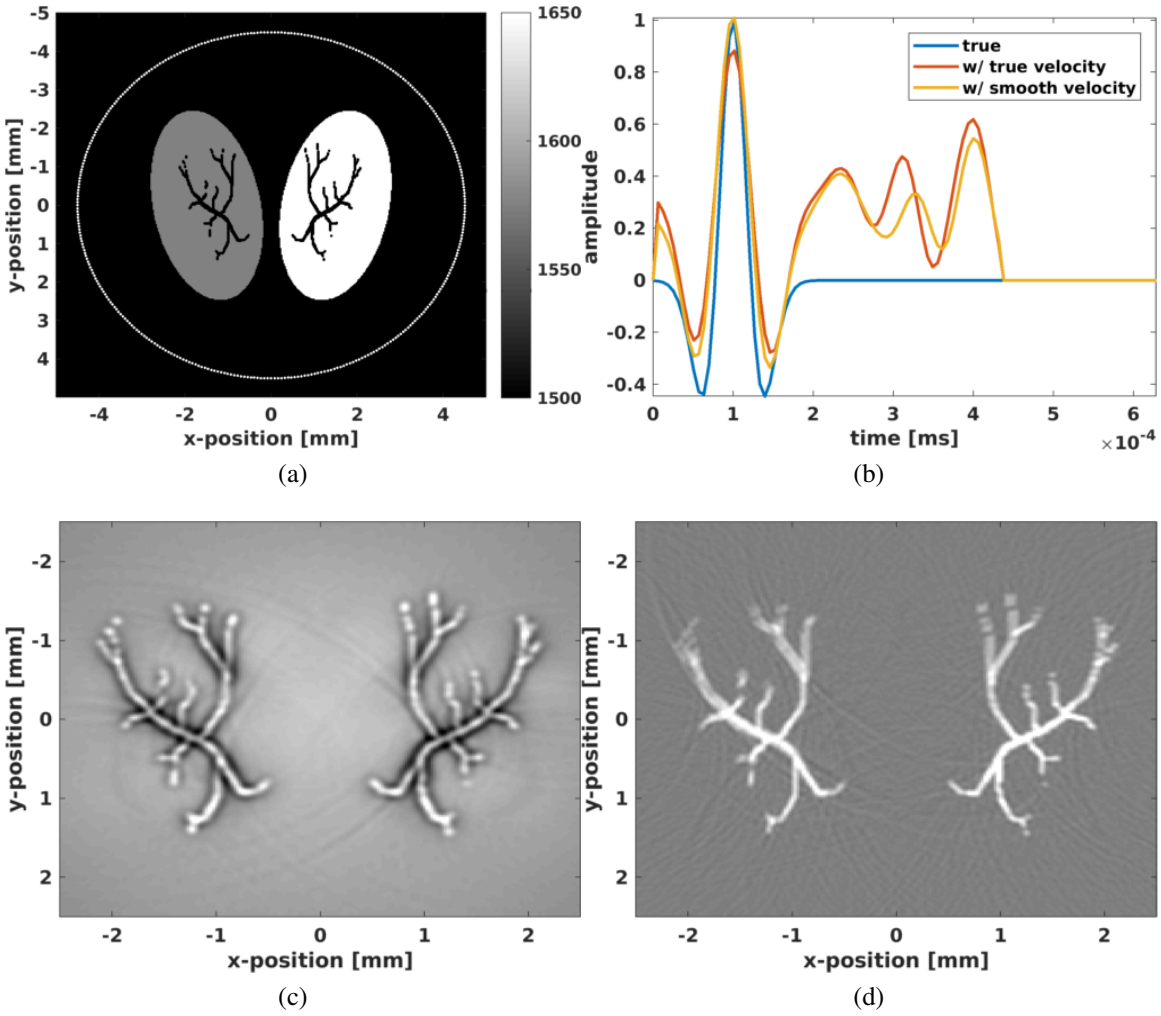


Figure 6.3: Heterogeneous medium experiment: (a) Heterogeneous medium with hard discontinuities. (b) Source-time functions true (blue) and estimated with the true (red) and smoothed velocity (orange). Reconstructed photoabsorber image (zoomed) using smooth velocity with (c) proposed method and with (d) k-wave.

6.5 Discussion and conclusions

Photoacoustic imaging is a powerful method, which in part derives its performance by more or less complete elimination of the source-time function from the inversion procedure. While this obviously has the advantage of reducing the size of the unknowns, it does not allow us to study the frequency-dependence of the absorption mechanism locally. In addition, imaging by a single time-reversed imaging step relies on dense sampling, which strains the acquisition system.

Aside from being able to obtain more information on the source mechanism, our formulation is also robust with respect to poor sampling while it allows for media with strongly varying velocities. In the latter case, we only need access to a smooth velocity model that is kinematically correct. We achieve robustness with respect to the sampling by virtue of promoting structure on the image. For now, we only imposed sparsity, which may not be reasonable when photoabsorbers form objects that exhibit spatial continuity such as the used phantom. Still, our method is capable of producing images for poor sampling because the subsampling artifacts are not sparse, which is paramount to the success of compressive sensing. Despite the less than ideal assumptions—i.e., the sources are not sparsely distributed Dirac deltas, our approach produces reasonable results while opening a perspective of being able to estimate the frequency-dependence of the sources, which may allow us to distinguish between different types of photoabsorbers.

CHAPTER 7
CONCLUSION

For better drilling decisions and to prevent any hazardous situations, oil and gas companies rely on high resolution images and high resolution estimation of physical properties of earth's subsurface. These high resolution images and estimation of high resolution physical properties rely on dense seismic data at higher frequencies. This thesis has contributed a computationally efficient recursively weighted matrix completion framework to reconstruct dense seismic data at higher frequencies. Also, another contribution of this thesis is a computationally efficient low-rank based simultaneous source acquisition for large scale 5D seismic datasets. Accurate detection of fractures along with estimation of there evolution in time is very important to prevent any hazardous situations while fracking. Therefore, another contribution of this thesis is a fast sparsity-promoting framework for accurate detection of closely spaced microseismic sources along with estimation of associated source-time functions. By resolving closely spaced microseismic sources along the fractures, we are able to estimate accurate spatio-temporal evolution of these fractures. To get additional information from photoabsorbers in terms of their spectral properties and to get a good quality map of these photoabsorbers even with poor receiver sampling, we have proposed sparsity-promoting photoacoustic imaging in this thesis.

7.1 High frequency seismic data reconstruction using recursively weighted matrix completion

In the first part of this thesis (chapter 2), we proposed a recursively weighted matrix completion method to improve the quality of seismic data reconstruction at higher frequencies. Moreover, conventional matrix completion methods perform well at lower frequencies. Rows and column subspaces of adjacent frequencies exhibit some degree of similarity.

Based on these two facts, we recursively reconstruct frequency slices using weighted matrix completion from lower to higher frequencies. Good performance of conventional matrix completion at lower frequencies helps us in getting good prior information for the recursively weighted method. Also, careful selection of the weight parameters, indicating the correlation between row and column subspaces of adjacent frequencies, helps in exploiting full potential of recursively weighted method.

The original weighted matrix completion formulation (Aravkin, Kumar, Mansour, Recht, and Herrmann 2014) involves taking a computationally expensive projection operation per iteration. To avoid taking these expensive projections, we solve an equivalent version of the original weighted formulation problem by shifting the weights to data constraints of the original weighted matrix completion problem. With above modifications, we observed significant speed up (almost 20 to 25 times) in performance of the weighted formulation compared to the original weighted formulation.

The conventional matrix completion method can be embarrassingly parallel across frequency slices making it computationally efficient and scalable to large-scale data volumes. Unfortunately, the interdependence between frequency slices does not allow parallelization across frequencies in the recursively weighted method. This interdependence between frequency slices in the weighted method poses a computational challenge for large scale problems. To mitigate this computational challenge, we incorporated strategies of alternation and decoupling to solve for individual rows of low-rank factors in parallel using the recursively weighted method. We achieve this parallel formulation by assuming that weight matrices at higher weights are near diagonal matrices. Therefore, our current parallel recursively weighted framework is able to partially exploit the benefits of recursively weighted method.

We have also shown application of our recursively weighted framework on a 5D synthetic seismic data and a 2D field data. Through these experiments we demonstrated how recursively weighted method is able to improve the quality of reconstructed seismic data in

comparison to the conventional method.

7.2 5D time-jittered marine acquisition using low-rank

In chapter 3 of this thesis, we proposed a large scale 5D time-jittered seismic data acquisition using low-rank matrix completion. Our method is based on the fact that frequency slices of dense seismic data on a uniform grid exhibits low-rank property in certain transform domains (e.g. midpoint-offset domain for 2D). We exploit the fact that time-jittered acquisition destroys this low-rank property for simultaneous separation and interpolation of shots on a dense grid. For large scale 5D time-jittered acquisition, transform based methods (e.g. Fourier, Radon, Curvelets etc.) can be computationally expensive. Therefore, our method based on low-rank matrix factorizations provides a computationally efficient alternative to transform-based methods.

Because of overlap of shots, the sampling operator in case of time-jittered marine acquisition is a combined time-shift and shot-jitter operator. Therefore, unlike in wavefield reconstruction, the sampling operator in case of time-jittered marine acquisition is not separable over frequencies. Hence, we cannot solve for individual frequency slices separately. Our method takes this fact into account and therefore solves the source-separation problem in the temporal-frequency domain. By temporal-frequency domain approach we mean that every iteration involves going back and forth from frequency to time-domain and time to frequency domain to calculate data residual in time-domain and low-rank factors in frequency domain respectively.

Through experiments on synthetic 5D data we have demonstrated that our method can simultaneously recover and interpolate shot gathers from a time-jittered data. Here, 5D data is associated with seismic data acquired with 3D acquisition and 5 dimensions stand for time, x and y coordinates of source and receivers. A software release will be also available for the 5D time-jittered marine acquisition. In addition to be computationally fast, our method is also memory efficient. Instead of storing complete reconstructed data

volume, we can only store the low-rank factors for each frequency slice.

7.3 Sparsity-promoting microseismic source estimation

In chapter 4 of this thesis, we proposed a sparsity-promotion based method to simultaneously detect closely spaced microseismic sources and estimate the associated source-time function. Microseismic events generated by hydraulic fracturing are active for a short duration and mostly localized along fractures. In other words, these microseismic sources are sparse in space and have finite energy in time. Based on these two assumptions, we proposed a sparsity-promoting framework that aims to estimate the complete source-wavelfield, which is sparse in space and has finite energy along time. Given a sufficiently accurate subsurface velocity model, the resulting source-wavefield also minimizes the difference between the observed and predicted microseismic data. Since our method is able to estimate the complete source-wavefield, it can detect multiple closely spaced microseismic sources with different source-time functions and frequency content without prior assumptions on the number of microseismic sources or shape of the the source-time functions.

Because of the ease of implementation of the linearized Bregman algorithm and the fact that this algorithm solves a convex problem, we chose this algorithm to solve our sparsity-promoting problem. To resolve closely spaces microseismic sources even within a half of the dominant wavelength, linearized Bregman algorithm requires a large number of iterations because of the requirement of higher threshold values. Each iteration of linearized Bregman algorithm requires solving at least one forward and adjoint wave-equation. Therefore, linearized-Bregman algorithm can become computationally expensive for large datasets and also when sources are in close vicinity. To avoid this cost, we proposed to solve the dual of the original sparsity-promoting problem, which can be solved using quasi-Newton methods such as L-BFGS. By solving the dual problem, where the unknown is the much smaller observed wavefield at the receivers rather than the full wavefield everywhere, we achieve much faster convergence in comparison to the original problem.

To address the issue of ambient noise in microseismic data, we proposed a debiasing approach in chapter 5. Our debiasing approach allowed us to detect microseismic sources and estimate the associated source-time functions from microseismic data heavily contaminated with noise. We incorporated a Curvelet based denoising approach in sparsity-promoting framework to handle extreme noise issue, which otherwise sparsity-promoting framework was unable to handle.

We demonstrated the effectiveness of our sparsity-promoting framework on synthetic microseismic data generated using complex geological models (e.g. Marmousi, BG Compass) mimicking the complexity of earth's subsurface. Through these experiments, we demonstrated that our sparsity-promoting framework can detect multiple closely spaced microseismic sources along with estimating overlapping source-time functions. Also, we showed that our method can detect these sources located in complex geological structures such as faults, unconformity etc.

7.4 Sparsity-promoting photoacoustic imaging

In the last part of this thesis (chapter 6), we proposed a sparsity-promoting photoacoustic imaging method. In comparison to the conventional photoacoustic imaging, which focuses on estimating the locations of photoabsorbers, our method based on sparsity-promotion can simultaneously estimate location and source-time functions of photoabsorbers. By estimating the source-time function, our method allows us to determine the frequency dependence of photoabsorbers and potentially use this frequency dependence to classify different kinds of photoabsorbers.

The traditional photoacoustic imaging method is based on back-propagation of data from transducers. To obtain a good quality image from back-propagation, we need dense coverage of transducers. Dense coverage of transducers can be expensive in terms of data storage and operational cost of so many transducers. Whereas, our sparsity-promoting method produces good quality images along with additional information about source-time

functions even with smaller number of transducers.

Through synthetic experiments, we also demonstrated the application of our method to locate the photoabsorbers along a phantom representing a blood vessel. Through these experiments we also demonstrated that our method can work with poor receiver sampling. Also, our method is independent of the shape of source-time functions.

7.5 Future research directions

Some of the ideas for the future research direction are as follows:

1. Our parallel version of recursively weighted method is computationally efficient and allows us to implement it for large scale seismic data. But the current version of this parallel framework can only work with higher weights. By using higher weights we are only able to partially exploit the benefits for weighted method for large datasets. Therefore, a future research direction would be to develop a computationally efficient version of recursively weighted method for smaller weights to get maximum benefits of the recursively weighted method.

2. As we discussed, the overlap of shots in the time-jittered acquisition does not allow us to reconstruct different frequency slices separately. This poses a challenge in incorporating weighted matrix completion for time-jittered acquisition. Another research direction would be to incorporate weights in the time-jittered acquisition to improve the quality of reconstructed data.

3. The sampling-transformation operators in over-under and simultaneous long offset acquisition allow to work separately on different frequency slices. Therefore, a research direction could be to implement weighted matrix completion for these kinds of acquisition to get improved quality of reconstructed data.

4. The sparsity-promoting framework requires storing complete spatio-temporal source-wavefield. For a large scale 3D data storing the complete source-wavefield can be

very expensive. A future research direction can be to incorporate check-pointing strategy in the existing framework. The check-pointing would allow us to store only a few snapshots or checkpoints rather than storing the complete source-wavefield.

5. The sparsity-promoting photoacoustic imaging assumes photoabsorbers to be localized in space. But these photoabsorbers are not necessarily point sources. Rather they can be located along a plane (e.g. along a blood vessel) with sharp boundaries. To get better image of these planes we can use Total-variation (TV) norm (Rudin, Osher, and Fatemi 1992) in space rather than using the ℓ_1 norm. TV-norm is defined as ℓ_1 norm of derivatives. By using TV norm we can get better image of the planes containing the photoabsorbers.

Appendices

APPENDIX A

In this section, we justify our parallel implementation of the weighted matrix completion problem. Beginning at equation 2.2, our original weighted program, we will arrive at equations 2.15 and 2.16 which specify our implemented parallel counterpart.

Recall equation 2.2

$$\mathbf{X} := \arg\min_{\mathbf{Y}} \quad \|\mathbf{QYW}\|_* \quad \text{subject to} \quad \|\mathcal{A}(\mathbf{Y}) - \mathbf{B}\|_F \leq \epsilon.$$

Because this is a convex program and $\mathbf{Q}$,$\mathbf{W}$ are invertible when $w_1, w_2 > 0$, we can show that

$$\mathbf{QXW} := \arg\min_{\mathbf{Y}} \quad \|\mathbf{Y}\|_* \quad \text{subject to} \quad \|\mathcal{A}(\mathbf{Q}^{-1}\mathbf{YW}^{-1}) - \mathbf{B}\|_F \leq \epsilon, \tag{A1}$$

where

$$\mathbf{Q}^{-1} = \frac{1}{w_1}\mathbf{U}\mathbf{U}^H + \mathbf{U}^{\perp}\mathbf{U}^{\perp^H}$$

and

$$\mathbf{W}^{-1} = \frac{1}{w_2}\mathbf{V}\mathbf{V}^H + \mathbf{V}^{\perp}\mathbf{V}^{\perp^H}.$$

From a numerical perspective, we wish to avoid implementing the operators $\mathbf{Q}^{-1}$, $\mathbf{W}^{-1}$ due to the factors w_1^{-1}, w_2^{-1} which may be large and cause algorithmic instability. Instead, by multiplying both sides of the constraint of equation A1 by $w_1 w_2$ we obtain the equivalent program

$$\mathbf{QXW} := \arg\min_{\mathbf{Y}} \quad \|\mathbf{Y}\|_* \quad \text{subject to} \quad \|\mathcal{A}(\widehat{\mathbf{Q}}\mathbf{Y}\widehat{\mathbf{W}}) - w_1 w_2 \mathbf{B}\|_F \leq w_1 w_2 \epsilon, \tag{A2}$$

where we have defined

$$\widehat{\mathbf{Q}} = \mathbf{U}\mathbf{U}^H + w_1\mathbf{U}^\perp\mathbf{U}^{\perp^H}$$

and

$$\widehat{\mathbf{W}} = \mathbf{V}\mathbf{V}^H + w_2\mathbf{V}^\perp\mathbf{V}^{\perp^H}.$$

Choosing a rank parameter r, we now apply a factorization approach and solve

$$\begin{aligned} \bar{\mathbf{L}}, \bar{\mathbf{R}} := \quad & \underset{\bar{\mathbf{L}}_\#, \bar{\mathbf{R}}_\#}{\arg\min} \quad \frac{1}{2}\left\| \begin{bmatrix} \bar{\mathbf{L}}_\# \\ \bar{\mathbf{R}}_\# \end{bmatrix} \right\|_F^2 \\ & \text{subject to} \\ & \|\mathcal{A}(\widehat{\mathbf{Q}}\bar{\mathbf{L}}_\#\bar{\mathbf{R}}_\#^H\widehat{\mathbf{W}}) - w_1w_2\mathbf{B}\|_F \leq w_1w_2\epsilon, \end{aligned} \tag{A3}$$

which gives the approximation $\bar{\mathbf{L}}\bar{\mathbf{R}}^H \approx \mathbf{Q}\mathbf{X}\mathbf{W}$. Given an initial left factor estimate, $\bar{\mathbf{L}}^0$, we proceed with a block coordinate descent (Xu and Yin 2013) approach which at the k-th iteration solves

$$\bar{\mathbf{R}}^k := \underset{\bar{\mathbf{R}}_\#}{\arg\min} \quad \|\bar{\mathbf{R}}_\#\|_F^2 \quad \text{subject to} \quad \|\mathcal{A}(\widehat{\mathbf{Q}}\bar{\mathbf{L}}^{k-1}\bar{\mathbf{R}}_\#^H\widehat{\mathbf{W}}) - w_1w_2\mathbf{B}\|_F \leq w_1w_2\epsilon, \tag{A4}$$

and upon output switches to optimize over the left factor

$$\bar{\mathbf{L}}^k := \underset{\bar{\mathbf{L}}_\#}{\arg\min} \quad \|\bar{\mathbf{L}}_\#\|_F^2 \quad \text{subject to} \quad \|\mathcal{A}(\widehat{\mathbf{Q}}\bar{\mathbf{L}}_\#(\bar{\mathbf{R}}^k)^H\widehat{\mathbf{W}}) - w_1w_2\mathbf{B}\|_F \leq w_1w_2\epsilon. \tag{A5}$$

After k iterations, we obtain estimate $\bar{\mathbf{L}}^k(\bar{\mathbf{R}}^k)^H \approx \mathbf{Q}\mathbf{X}\mathbf{W}$.

Our next goal is to approximately solve problems A5 and A4 in a distributed manner, to be implemented in a parallel computing architecture. To this end, we apply our approximate commutative property, i.e., $\mathcal{A}(\widehat{\mathbf{Q}}\mathbf{Y}\widehat{\mathbf{W}}) \approx \mathcal{A}(\widehat{Q}\mathbf{Y})\widehat{\mathbf{W}}$ and $\mathcal{A}(\widehat{\mathbf{Q}}\mathbf{Y}\widehat{\mathbf{W}}) \approx \widehat{\mathbf{Q}}\mathcal{A}(\mathbf{Y}\widehat{\mathbf{W}})$ for

large values of w_1 and w_2. Using these approximations, we obtain

$$\bar{\mathbf{L}}^k \approx \underset{\bar{\mathbf{L}}_\#}{\arg\min} \quad \|\bar{\mathbf{L}}_\#\|_F^2 \quad \text{subject to} \quad \|\widehat{\mathbf{Q}}\mathcal{A}(\bar{\mathbf{L}}_\#(\bar{\mathbf{R}}^k)^H\widehat{\mathbf{W}}) - w_1w_2\widehat{\mathbf{Q}}\widehat{\mathbf{Q}}^{-1}\mathbf{B}\|_F \leq w_1w_2\epsilon. \tag{A6}$$

Define $\widehat{\mathbf{B}}_L = \widehat{\mathbf{Q}}^{-1}\mathbf{B}$. Using the inequality property $\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|\|\mathbf{B}\|_F$ for any two matrices, where $\|\circ\|$ is the spectral norm, in the constraint, we see that

$$\begin{aligned}\|\widehat{\mathbf{Q}}\left(\mathcal{A}(\bar{\mathbf{L}}_\#(\bar{\mathbf{R}}^k)^H\widehat{\mathbf{W}}) - w_1w_2\widehat{\mathbf{B}}_L\right)\|_F &\leq \|\widehat{\mathbf{Q}}\|\|\mathcal{A}(\bar{\mathbf{L}}_\#(\bar{\mathbf{R}}^k)^H\widehat{\mathbf{W}}) - w_1w_2\widehat{\mathbf{B}}_L\|_F \\ &= \|\mathcal{A}(\bar{\mathbf{L}}_\#(\bar{\mathbf{R}}^k)^H\widehat{\mathbf{W}}) - w_1w_2\widehat{\mathbf{B}}_L\|_F.\end{aligned}$$

The last equality holds since $\|\hat{\mathbf{Q}}\| = \max\{1, w_1\} = 1$. Therefore, if we instead solve

$$\tilde{\mathbf{L}}^k := \underset{\bar{\mathbf{L}}_\#}{\arg\min} \quad \|\bar{\mathbf{L}}_\#\|_F^2 \quad \text{subject to} \quad \|\mathcal{A}(\bar{\mathbf{L}}_\#(\bar{\mathbf{R}}^k)^H\widehat{\mathbf{W}}) - w_1w_2\widehat{\mathbf{B}}_L\|_F \leq w_1w_2\epsilon, \tag{A7}$$

we expect $\tilde{\mathbf{L}}^k \approx \bar{\mathbf{L}}^k$ due to approximate commutativity and therefore $\tilde{\mathbf{L}}^k$ is feasible for A6 . A similar argument can be established for the right factor, where we solve

$$\tilde{\mathbf{R}}^k := \underset{\bar{\mathbf{R}}_\#}{\arg\min} \quad \|\bar{\mathbf{R}}_\#\|_F^2 \quad \text{subject to} \quad \|\mathcal{A}(\widehat{\mathbf{Q}}\tilde{\mathbf{L}}^{k-1}\bar{\mathbf{R}}_\#^H) - w_1w_2\widehat{\mathbf{B}}_R\|_F \leq w_1w_2\epsilon, \tag{A8}$$

with $\widehat{\mathbf{B}}_R = \mathbf{B}\widehat{\mathbf{W}}^{-1}$.

The main advantage in computing iterates $\tilde{\mathbf{R}}^k, \tilde{\mathbf{L}}^k$, rather than $\bar{\mathbf{R}}^k, \bar{\mathbf{L}}^k$, is that these programs allow for a distributed implementation. The data matrices $\widehat{\mathbf{B}}_R$ and $\widehat{\mathbf{B}}_L$ in equations A8 and A7 are dense (have all non-zero entries) making computation expensive. However, when the weights $w_{1,2}$ are relatively large we observe that both dense matrices $\widehat{\mathbf{B}}_R$, $\widehat{\mathbf{B}}_L$ can be well approximated by the sparse observed data matrix $\mathbf{B}$. This leads to subproblems 2.15 and 2.16 and concludes our derivation.

APPENDIX B

We provide here complete derivation of equation (4.7). We rewrite the primal source estimation problem

$$\min_{\mathbf{Q}} \quad \|\mathbf{Q}\|_{2,1} + \frac{1}{2\mu}\|\mathbf{Q}\|_F^2 \quad \text{s.t.} \quad \|\mathcal{F}[\mathbf{m}](\mathbf{Q}) - \mathbf{d}\|_2 \leq \epsilon, \tag{B1}$$

into the equivalent unconstrained form

$$\min_{\mathbf{Q}} \quad \|\mathbf{Q}\|_{2,1} + \frac{1}{2\mu}\|\mathbf{Q}\|_F^2 + \iota_{\|\mathcal{F}[\mathbf{m}](\mathbf{Q})-\mathbf{d}\|_2 \leq \epsilon}, \tag{B2}$$

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 (B2), we obtain the Fenchel's dual of (B1)

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

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, Ma, and Goldfarb 2013) to our $\ell_{2,1}$ norm, we see that both $\Psi(\mathbf{y})$ and $\nabla\Psi(\mathbf{y})$ have close-form representations

$$\Psi(\mathbf{y}) = \Phi_\mu(\mu\mathcal{F}[\mathbf{m}]^T(\mathbf{y})) - \frac{\mu}{2}\|\mathcal{F}[\mathbf{m}]^T(\mathbf{y})\|_2^2 + \mathbf{d}^T\mathbf{y} \tag{B3}$$

and

$$\nabla\Psi(\mathbf{y}) = \mathbf{d} - \mathcal{F}[\mathbf{m}](\text{Prox}_{\mu\ell_{2,1}}(\mu\mathcal{F}[\mathbf{m}]^T(\mathbf{y}))), \tag{B4}$$

where

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

APPENDIX C

C.1 Permissions to use copyrighted material

The content of chapter 2 is still in review in *Geophysics* journal and no transfer of copyright is signed. Copyright remains with the author.

The content of chapters 3 and 5 was published as a technical article in SEG Expanded Abstract. As stated at `https://library.seg.org/page/policies/open-access`, it grants author permission to reuse it in thesis.

The content of chapter 4 was published as a technical article in *Geophysical Journal International*. As stated at `https://academic.oup.com/journals/pages/access_purchase/rights_and_permissions/publication_rights`, it grants author permission to reuse it in thesis.

The content of chapter 6 was published as a conference proceeding in 2018 IEEE International Ultrasonics Symposium and author has the permission from IEEE to reuse it in thesis.

REFERENCES

[1] R. E. Sheriff and L. P. Geldart, *Exploration seismology*. Cambridge university press, 1995.

[2] J. Caldwell and C. Walker, "An overview of marine seismic operations," *The international association of oil and gas producers*, Apr. 2011.

[3] V Bardan, "Trace interpolation in seismic data processing," *Geophysical Prospecting*, vol. 35, no. 4, pp. 343–358, 1987.

[4] M. N. Kabir and D. Verschuur, "Restoration of missing offsets by parabolic radon transform1," *Geophysical Prospecting*, vol. 43, no. 3, pp. 347–368, 1995.

[5] J. D. Villasenor, R. Ergas, and P. Donoho, "Seismic data compression using high-dimensional wavelet transforms," in *Proceedings of Data Compression Conference-DCC'96*, IEEE, 1996, pp. 396–405.

[6] F. J. Herrmann and G. Hennenfent, "Non-parametric seismic data recovery with curvelet frames," *Geophysical Journal International*, vol. 173, no. 1, pp. 233–248, 2008.

[7] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[8] A. Aravkin, R. Kumar, H. Mansour, B. Recht, and F. J. Herrmann, "Fast methods for denoising matrix completion formulations, with applications to robust seismic data interpolation," *SIAM Journal on Scientific Computing*, vol. 36, no. 5, S237–S266, 2014.

[9] H. Mansour, H. Wason, T. T. Lin, and F. J. Herrmann, "Randomized marine acquisition with compressive sampling matrices," *Geophysical Prospecting*, vol. 60, no. 4, pp. 648–662, Jul. 2012.

[10] C. Wenzel, *A case study - hydraulic fracturing geography: The case of the Eagle Ford shale, TX, USA*, 2012.

[11] D. L. Donoho, "Compressed sensing," *IEEE T. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[12] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2005.

[13] S. Xu, Y. Zhang, D. Pham, and G. Lambaré, "Antileakage fourier transform for seismic data regularization," *GEOPHYSICS*, vol. 70, no. 4, pp. V87–V95, 2005.

[14] R. Kumar, C. D. Silva, O. Akalin, A. Y. Aravkin, H. Mansour, B. Recht, and F. J. Herrmann, "Efficient matrix completion for seismic data reconstruction," *Geophysics*, vol. 80, no. 05, pp. V97–V114, Sep. 2015.

[15] V. Oropeza and M. Sacchi, "Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis," *Geophysics*, vol. 76, no. 3, pp. V25–V32, 2011.

[16] Y. Yang, J. Ma, and S. Osher, "Seismic data reconstruction via matrix completion," *Inverse Problems Imaging*, vol. 7, p. 1379, 2013.

[17] A. Eftekhari, D. Yang, and M. B. Wakin, "Weighted matrix completion and recovery with prior subspace information," *IEEE Transactions on Information Theory*, vol. 64, no. 6, pp. 4044–4071, 2018.

[18] C. J. Beasley, R. E. Chambers, and Z. Jiang, "A new look at simultaneous sources," 1, vol. 17, SEG, 1998, pp. 133–135.

[19] A. Berkhout, "Changing the mindset in seismic data acquisition," *The Leading Edge*, vol. 27, no. 7, pp. 924–938, 2008.

[20] G. Hampson, J. Stefani, and F. Herkenhoff, "Acquisition using simultaneous sources," *The Leading Edge*, vol. 27, no. 7, pp. 918–923, 2008.

[21] I. Moore, B. Dragoset, T. Ommundsen, D. Wilson, C. Ward, and D. Eke, "Simultaneous source separation using dithered sources," in *78th Annual SEG meeting*, SEG Technical Program Expanded Abstracts, pp. 2806-2810, 2008.

[22] H. Wason and F. J. Herrmann, "Time-jittered ocean bottom seismic acquisition," in *83rd Annual SEG meeting*, 2013, pp. 1–6.

[23] E. J. Candes and D. L. Donoho, "Curvelets: A surprisingly effective nonadaptive representation for objects with edges," DTIC Document, Tech. Rep., 2000.

[24] G. Hennenfent and F. J. Herrmann, "Simply denoise: Wavefield reconstruction via jittered undersampling," *GEOPHYSICS*, vol. 73, no. 3, pp. V19–V28, 2008.

[25] R. Madariaga, *Seismic source: Theory*. Springer US, 1989, pp. 1129–1133, ISBN: 978-0-387-30752-7.

[26] G. A. McMechan, "Determination of source parameters by wavefield extrapolation," *Geophysical Journal of the Royal Astronomical Society*, vol. 71, no. 3, 613–628, 1982.

[27] D. Gajewski and E. Tessmer, "Reverse modelling for seismic event characterization," *Geophysical Journal International*, vol. 163, no. 1, 276–284, 2005.

[28] M. Fink, "Time reversed acoustics," *Physics Today*, vol. 50, no. 3, 34–40, 1997.

[29] F. Bazargani and R. Snieder, "Optimal source imaging in elastic media," *Geophysical Journal International*, vol. 204, no. 2, pp. 1134–1147, 2016.

[30] N. Nakata and G. C. Beroza, "Reverse time migration for microseismic sources using the geometric mean as an imaging condition," *Geophysics*, vol. 81, no. 2, K551–K560, 2016.

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

[32] Y. Wu and G. A. McMechan, "Elastic full-waveform inversion for earthquake source parameters," *Geophysical Joournal International*, vol. 127, no. 1, pp. 61–74, 1996.

[33] Y. Kim, Q. Liu, and J. Tromp, "Adjoint centroid-moment tensor inversions," *Geophysical Journal International*, vol. 186, no. 1, pp. 264–278, 2011.

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

[35] S. Sharan, R. Wang, and F. J. Herrmann, "Fast sparsity-promoting microseismic source estimation," *Geophysical Journal International*, vol. 216, no. 1, pp. 164–181, Oct. 2018.

[36] W Yin, S Osher, D Goldfarb, and J Darbon, "Bregman Iterative Algorithms for ℓ_1-minimization with Applications to Compressed Sensing," *SIAM Journal of Imaging Sciences*, vol. 1, no. 1, 143–168, 2008.

[37] D. A. Lorenz, F. Schöpfer, and S. Wenger, "The Linearized Bregman Method via Split Feasibility Problems: Analysis and Generalizations," *SIAM Journal of Imaging Sciences*, vol. 7, no. 2, 1237–1262, 2014.

[38] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Mathematical Programming*, vol. 45, no. 1, 503–528, 1989.

[39] S. Y. Emelianov, P.-C. Li, and M. O'Donnell, "Photoacoustics for molecular imaging and therapy.," *Physics today*, vol. 62 8, pp. 34–39, 2009.

[40] X. Wang, Y. Pang, G. Ku, and L. Wang, "Noninvasive laser-induced photoacoustic tomography for structural and functional in vivo imaging of the brain," *Nat Biotechnol*, vol. 21, no. 7, pp. 803–806, 2003.

[41] R. E. Kumon, C. X. Deng, and X. Wang, "Frequency-domain analysis of photoacoustic imaging data from prostate adenocarcinoma tumors in a murine model," *Ultrasound in Medicine Biology*, vol. 37, no. 5, pp. 834 –839, 2011.

[42] C. E. Jones, J. A. Edgar, J. I. Selvage, and H. Crook, "Building complex synthetic models to evaluate acquisition geometries and velocity inversion technologies," in *EAGE Annual Conference Proceedings*, (EAGE, Copenhagen), Jun. 2012.

[43] R. Kumar, S. Sharan, H. Wason, and F. J. Herrmann, "Time-jittered marine acquisition–-a rank-minimization approach for 5D source separation," in *SEG Technical Program Expanded Abstracts*, (SEG, Dallas), Oct. 2016, pp. 119–123.

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

[45] B. E. Treeby and B. T. Cox, "k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields," *Journal of Biomedical Optics*, vol. 15, no. 2, pp. 1 –12, 2010.

[46] S. Sharan, R. Kumar, D. S. Dumani, M. Louboutin, R. Wang, S. Emelianov, and F. J. Herrmann, "Sparsity-promoting photoacoustic imaging with source estimation," in *2018 IEEE International Ultrasonics Symposium (IUS)*, (IEEE IUS, Kobe), Oct. 2018, pp. 206–212.

[47] O. López, R. K. O. Yilmaz, and F. J. Herrmann, "Off-the-grid low-rank matrix recovery and seismic data reconstruction," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 658–671, 2016.

[48] C. Da Silva and F. J. Herrmann, "Optimization on the hierarchical tucker manifold – applications to tensor completion," *Linear Algebra and its Applications*, vol. 481, pp. 131–173, 2015.

[49] R. Kumar, H. Wason, S. Sharan, and F. J. Herrmann, "Highly repeatable 3d compressive full-azimuth towed-streamer time-lapse acquisition — a numerical feasibility study at scale," *The Leading Edge*, vol. 36, no. 8, pp. 677–687, 2017.

[50] H. Mansour, F. J. Herrmann, and Özgür Yılmaz, "Improved wavefield reconstruction from randomized sampling via weighted one-norm minimization," *GEOPHYSICS*, vol. 78, no. 5, pp. V193–V206, 2012.

[51] E. Van Den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.

[52] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1758–1789, 2013.

[53] P. Jain, P. Netrapalli, and S. Sanghavi, "Low-rank matrix completion using alternating minimization," in *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing*, ACM, 2013, pp. 665–674.

[54] O. Lopez, R. Kumar, and F. J. Herrmann, "Rank minimization via alternating optimization: Seismic data interpolation," in *EAGE Annual Conference Proceedings*, (EAGE, Madrid), Jun. 2015.

[55] Y. Zhang, S. Sharan, and F. J. Herrmann, "High-frequency wavefield recovery with weighted matrix factorizations," in *SEG Technical Program Expanded Abstracts*, SEG, SEG, 2019, pp. 3959–3963.

[56] E. J. Candes and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, p. 717, 2009.

[57] B. Recht and C. Ré, "Parallel stochastic gradient algorithms for large-scale matrix completion," *Mathematical Programming Computation*, vol. 5, 201–226, May 2013.

[58] L. Demanet, "Curvelets, wave atoms, and wave equations," (PhD), PhD thesis, California Institute of Technology, May 2006.

[59] P. A. Witte, M. Louboutin, H. Modzelewski, C. Jones, J. Selvage, and F. J. Herrmann, "Event-driven workflows for large-scale seismic imaging in the cloud," in *SEG Technical Program Expanded Abstracts*, SEG, SEG, 2019, pp. 3984–3988.

[60] R. de Kok and D. Gillespie, "A universal simultaneous shooting technique," EAGE, Eur. Ass. of Geosc. and Eng., Expanded abstracts, 2002.

[61] C. J. Beasley, "A new look at marine simultaneous sources," *The Leading Edge*, vol. 27, no. 7, pp. 914–917, 2008.

[62] M. Maraschini, R. Dyer, K. Stevens, and D. Bird, "Source separation by iterative rank reduction - theory and applications," in *74th EAGE Conference and Exhibition*, 2012.

[63] J. Cheng and M. D. Sacchi, "Separation of simultaneous source data via iterative rank reduction," in *SEG Technical Program Expanded Abstracts*, 2013, pp. 88–93.

[64] R. Kumar, H. Wason, and F. J. Herrmann, "Source separation for simultaneous towed-streamer marine acquisition –- a compressed sensing approach," *Geophysics*, vol. 80, no. 06, WD73–WD88, Nov. 2015, (Geophysics).

[65] J. Lee, B. Recht, R. Salakhutdinov, N. Srebro, and J. Tropp, "Practical large-scale optimization for max-norm regularization," in *Advances in Neural Information Processing Systems, 2010*, 2010.

[66] N. Kreimer and M. D. Sacchi, "A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation," *GEOPHYSICS*, vol. 77, no. 3, pp. V113–V122, 2012.

[67] C. D. Silva and F. J. Herrmann, "Hierarchical Tucker tensor optimization - applications to 4D seismic data interpolation," in *EAGE Annual Conference Proceedings*, Jun. 2013.

[68] B. Recht, M Fazel, and P. Parrilo, "Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization.," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[69] J. D. M. Rennie and N. Srebro, "Fast maximum margin matrix factorization for collaborative prediction," in *Proceedings of the 22nd international conference on Machine learning*, ser. ICML '05, Bonn, Germany: ACM, 2005, pp. 713–719.

[70] B. Recht and C. Ré, "Parallel stochastic gradient algorithms for large-scale matrix completion," in *Optimization Online*, 2011.

[71] J. F. Claerbout and F. Muir, "Robust modeling with erratic data," *Geophysics*, vol. 38, no. 5, pp. 826–844, 1973.

[72] H. L. Taylor, S. Banks, and J. McCoy, "Deconvolution with the ℓ_1 norm," *Geophysics*, vol. 44, pp. 39–52, 1979.

[73] F. Santosa and W. Symes, "Linear Inversion of Band-Limited Reflection Seismograms," *SIAM Journal on Scientific and Statistical Computing*, vol. 7, no. 4, pp. 1307–1330, 1986.

[74] D. L. Donoho and B. F. Logan, "Signal Recovery and the Large Sieve," *SIAM Journal on Applied Mathematics*, vol. 52, no. 2, pp. 577–591, 1992.

[75] M. Fornasier and H. Rauhut, "Recovery algorithms for vector-valued data with joint sparsity constraints," *SIAM Journal on Numerical Analysis*, vol. 46, no. 2, pp. 577–613, 2008.

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

[77] W. Gao and M. Sacchi, "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, Jun. 2017.

[78] R. P. Porter and A. J. Devaney, "Holography and the inverse source problem," *J. Opt. Soc. Am.*, vol. 72, no. 3, pp. 327–330, 1982.

[79] R. E. Musafir, "On non-radiating sources," *Journal of Sound and Vibration*, vol. 332, no. 17, pp. 3947 –3955, 2013.

[80] E. Abbe, "Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung," *Archiv fur Mikroskopische Anatomie*, vol. 9, no. 1, 413–418, 1873.

[81] J. Sun, Z. Xue, S. Fomel, T. Zhu, and N. Nakata, "Full waveform inversion of passive seismic data for sources and velocities," in *86th Annual SEG meeting*, SEG Technical Program Expanded Abstracts, pp. 1405-1410, 2016.

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

[83] Z. Wu and T. Alkhalifah, "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, Jun. 2017.

[84] R Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, no. 3, pp. 276–280, 1986.

[85] S. Kitić, L. Albera, N. Bertin, and R. Gribonval, "Physics-driven inverse problems made tractable with cosparse regularization," *IEEE Transactions on Signal Processing*, vol. 64, no. 2, 335–348, 2016.

[86] T. Alkhalifah, "An acoustic wave equation for anisotropic media," *Geophysics*, vol. 65, no. 4, pp. 1239–1250, 2000.

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

[88] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[89] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, Jan. 2008.

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

[91] P. A. Witte, M. Louboutin, F. Luporini, G. J. Gorman, and F. J. Herrmann, "Compressive least-squares migration with on-the-fly fourier transforms," *GEOPHYSICS*, vol. 84, no. 5, R655–R672, 2019.

[92] S. Osher, Y. Mao, B. Dong, and W. Yin, "Fast linearized Bregman iterations for compressive sensing and sparse denoising," *Communications in Mathematical Sciences*, vol. 8, no. 1, pp. 93–111, 2010.

[93] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2009, ISBN: 978-0-521-83378-3.

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

[95] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. New York: Springer, 2011, pp. 185–212, ISBN: 978-1-4419-9569-8.

[96] M. Kowalski and B. Torrésani, "Sparsity and persistence: Mixed norms provide simple signal models with dependent coefficients," *Signal, Image and Video Processing*, vol. 3, no. 3, pp. 251–264, 2009.

[97] W. Yin, "Analysis and Generalizations of the Linearized Bregman Method," *SIAM Journal of Imaging Sciences*, vol. 3, no. 4, 856–877, 2010.

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

[99] Z.-M. Song and P. R. Williamson, "Frequency-domain acoustic-wave modeling and inversion of crosshole data: Part 1-2.5d modeling method," *Geophysics*, vol. 60, no. 3, pp. 784–795, 1995.

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

[101] X. Li, A. Y. Aravkin, T. van Leeuwen, and F. J. Herrmann, "Fast randomized full-waveform inversion with compressive sensing," *Geophysics*, vol. 77, no. 3, A13–A17, May 2012.

[102] T. van Leeuwen and F. J. Herrmann, "Fast waveform inversion without source-encoding," *Geophysical Prospecting*, vol. 61, no. s1, pp. 10–19, 2012.

[103] X. Li, E. Esser, and F. J. Herrmann, "Modified gauss-newton full-waveform inversion explained — why sparsity-promoting updates do matter," *Geophysics*, vol. 81, no. 3, R125–R138, 2016.

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

[105] Z. Li and M. van der Baan, "Microseismic event localization by acoustic time reversal extrapolation," *Geophysics*, vol. 81, no. 3, KS123–KS134, 2016.

[106] M. Louboutin, P. A. Witte, M. Lange, N. Kukreja, F. Luporini, G. Gorman, and F. J. Herrmann, "Full-waveform inversion - part 1: Forward modeling," *The Leading Edge*, vol. 36, no. 12, pp. 1033–1036, Dec. 2017.

[107] R. Kamei, N. Nakata, and D. Lumley, "Introduction to microseismic source mechanisms," *The Leading Edge*, vol. 34, no. 8, 876–880, 2015.

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

[109] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1, pp. 259 –268, 1992.

[110] C. C. Paige and M. A. Saunders, "LSQR: Sparse linear equations and least squares problems," *ACM Transactions on Mathematical Software*, vol. 8, no. 2, pp. 195–209, 1982.

[111] S. Maxwell, *2. Hydraulic-fracturing Concepts*, ser. Distinguished Instructor Series 17. Society of Exploration Geophysicists, 2014, ch. 2, pp 15–30, ISBN: 978-1-56080-316-4.

[112] P. M. Duncan and L. Eisner, "Reservoir characterization using surface microseismic monitoring," *GEOPHYSICS*, vol. 75, no. 5, A139–A146, 2010.

[113] J. D. Lakings, P. M. Duncan, C. Neale, and T. Theiner, "Surface based microseismic monitoring of a hydraulic fracture well stimulation in the barnett shale," in *SEG Technical Program Expanded Abstracts*, 2006, pp. 605–608.

[114] F. Forghani-Arani, M. Willis, S. Haines, M. Batzle, and M. Davidson, "Analysis of passive surface-wave noise in surface microseismic data and its implications," in *SEG Technical Program Expanded Abstracts*, SEG, SEG, 2012, pp. 1493–1498.

[115] S. Maxwell, D. Raymer, M. Williams, and P. Primiero, "Microseismic signal loss from reservoir to surface," CSEG, CSEG, 2013.

[116] H. Bolton and G. Masters, "Travel times of p and s from the global digital seismic networks: Implications for the relative variation of p and s velocity in the mantle," *Journal of Geophysical Research: Solid Earth*, vol. 106, no. B7, pp. 13 527–13 540, 2001.

[117] S. Sharan, R. Wang, and F. J. Herrmann, "High-resolution fast microseismic source collocation and source time-function estimation," in *SEG Technical Program Expanded Abstracts*, SEG, vol. 35, SEG, 2017, pp. 2778–2783.

[118] F. J. Herrmann and G. Hennenfent, "Non-parametric seismic data recovery with curvelet frames," *Geophysical Journal International*, vol. 173, pp. 233–248, Apr. 2008.

[119] R. Neelamani, A. I. Baumstein, D. G. Gillard, M. T. Hadidi, and W. L. Soroka, "Coherent and random noise attenuation using the curvelet transform," *The Leading Edge*, vol. 27, no. 2, pp. 240–248, 2008.

[120] R. Kumar, N. Moldoveanu, and F. J. Herrmann, "Denoising high-amplitude cross-flow noise using curvelet-based stable principle component pursuit," in *EAGE Annual Conference Proceedings*, (EAGE, Paris), Jun. 2017.

[121] A. St-Onge and D. W. Eaton, "Noise examples from two microseismic datasets," *CSEG RECORDER*, vol. 36, no. 8, 46–49, 2011.

[122] T. T. Lin and F. J. Herrmann, "Robust estimation of primaries by sparse inversion via one-norm minimization," *Geophysics*, vol. 78, no. 3, R133–R150, May 2013.

[123] H. Liebgott, A. Basarab, D. Kouame, O. Bernard, and D. Friboulet, "Compressive sensing in medical ultrasound," in *2012 IEEE International Ultrasonics Symposium*, 2012, pp. 1–6.

[124] J. Provost and F. Lesage, "The application of compressed sensing for photoacoustic tomography," *IEEE T. Med Imaging*, vol. 28, no. 4, pp. 585–594, 2009.

[125] P. Kruizinga, P. van der Meulen, A. Fedjajevs, F. Mastik, G. Springeling, N. de Jong, J. G. Bosch, and G. Leus, "Compressive 3d ultrasound imaging using a single sensor," *Science Advances*, vol. 3, no. 12, 2017.

[126] S. Park, A. B. Karpiouk, S. R. Aglyamov, and S. Y. Emelianov, "Adaptive beamforming for photoacoustic imaging using linear array transducer," in *2008 IEEE Ultrasonics Symposium*, 2008, pp. 1088–1091.

[127] F. L. Lizzi, M. Greenebaum, E. J. Feleppa, M. Elbaum, and D. J. Coleman, "Theoretical framework for spectrum analysis in ultrasonic tissue characterization," *J Acoust. Soc. Am.*, vol. 73, no. 4, pp. 1366–1373, 1983.

[128] E. J. Feleppa, A. Kalisz, J. B. Sokil-Melgar, F. L. Lizzi, Tian Liu, A. L. Rosado, M. C. Shao, W. R. Fair, Yu Wang, M. S. Cookson, V. E. Reuter, and W. D. W. Heston, "Typing of prostate tissue by ultrasonic spectrum analysis," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 43, no. 4, pp. 609–619, 1996.

[129] F. L. Lizzi, "Ultrasonic scatterer-property images of the eye and prostate," in *1997 IEEE Ultrasonics Symposium Proceedings*, vol. 2, 1997, 1109–1117 vol.2.

[130] M. P. Moore, T Spencer, D. M. Salter, P. P. Kearney, T. R. D. Shaw, I. R. Starkey, P. J. Fitzgerald, R Erbel, A Lange, N. W. McDicken, G. R. Sutherland, and K. A. A. Fox, "Characterisation of coronary atherosclerotic morphology by spectral analysis of radiofrequency signal: In vitro intravascular ultrasound study with histological and radiological validation," *Heart*, vol. 79, no. 5, pp. 459–467, 1998.

[131] P. A. Witte, M. Louboutin, N. Kukreja, F. Luporini, M. Lange, G. J. Gorman, and F. J. Herrmann, "A large-scale framework for symbolic implementations of seismic inversion algorithms in Julia," *Geophysics*, vol. 84, no. 3, F57–F71, 2019.

[132] M. Louboutin, M. Lange, F. Luporini, N. Kukreja, P. A. Witte, F. J. Herrmann, P. Velesko, and G. J. Gorman, "Devito (v3.1.0): An embedded domain-specific language for finite differences and geophysical exploration," *Geoscientific Model Development*, vol. 12, no. 3, pp. 1165–1187, 2019.

[133] F. Luporini, M. Lange, M. Louboutin, N. Kukreja, J. Hückelheim, C. Yount, P. A. Witte, P. H. J. Kelly, G. J. Gorman, and F. J. Herrmann, *Architecture and performance of Devito, a system for automated stencil computation*, `https://arxiv.org/abs/1807.03032`, Computing Research Repository (arXiv CoRR), 2018. (visited on 07/21/2018).

[134] B. Huang, S. Ma, and D. Goldfarb, "Accelerated Linearized Bregman Method," *Journal of Scientific Computing*, vol. 54, no. 2, 428–453, 2013.