Enabling large-scale seismic data acquisition, processing and waveform-inversion via rank-minimization

by

Rajiv Kumar

B.Sc., Hindu College, Delhi University, 2006

M.Sc., Indian Institute of Technology, Bombay, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Geophysics)

The University Of British Columbia
(Vancouver)

August 2017

Abstract

In this thesis, I adapt ideas from the field of compressed sensing to mitigate the computational and memory bottleneck of seismic processing workflows such as missing-trace interpolation, source separation and wave-equation based inversion for large-scale 3- and 5-D seismic data. For interpolation and source separation using rank-minimization, I propose three main ingredients, namely a rank-revealing transform domain, a subsampling scheme that increases the rank in the transform domain, and a practical large-scale data-consistent rank-minimization framework, which avoids the need for expensive computation of singular value decompositions. We also devise a wave-equation based factorization approach that removes computational bottlenecks and provides access to the kinematics and amplitudes of full-subsurface offset extended images via actions of full extended image volumes on probing vectors, which I use to perform the amplitude-versus- angle analyses and automatic wave-equation migration velocity analyses on complex geological environments. After a brief overview of matrix completion techniques in Chapter 1, we propose a singular value decomposition (SVD)-free factorization based rank-minimization approach for large-scale matrix completion problems. Then, I extend this framework to deal with large-scale seismic data interpolation problems, where I show that the standard approach of partitioning the seismic data into windows is not required, which use the fact that events tend to become linear in these windows, while exploiting the low-rank structure of seismic data. Carefully selected synthetic and realistic seismic data examples validate the efficacy of the interpolation framework. Next, I extend the SVD-free rank-minimization approach to remove the seismic cross-talk in simultaneous source acquisition. Experimental results verify that source separation using the SVD-free rank-minimization approaches are comparable to the sparsity-promotion based techniques; however, separation via rank-minimization is significantly faster and memory efficient. We further introduce a matrix-vector formulation to form full-subsurface extended image volumes, which removes the storage and computational bottleneck found in the convention methods. I demonstrate that the proposed matrix-vector formulation is used to form different image gathers with which amplitude-versus-angle and wave-equation migration velocity analyses is performed, without requiring prior information on the geologic dips. Finally, I conclude the thesis by outlining potential future research directions and extensions of the thesis work.

Lay Summary

In this thesis, I have developed fast computational techniques for large-scale seismic applications using SVD-free factorization based rank-minimization approach for missing-trace interpolation and source separation. The proposed framework is built upon the existing knowledge of compressed sensing as a successful signal recovery paradigm and outlined the necessary components of a low-rank domain, a rank-increasing sampling scheme, and a SVD-free rank-minimizing optimization scheme for successful missing-trace interpolation. I also proposed a matrix-vector formulation that avoids the explicit storage of full-subsurface offset extended image volumes and removes the expensive loop over shots found in the conventional extended imaging.

Preface

All of my thesis work herein presented is carried out under the supervision of Dr. Felix J. Herrmann, in the Seismic Laboratory for Imaging and Modeling at the University of British Columbia.

Chapter 1 is prepared by me. Figures related to seismic data survey designs are taken from the publicly available seismic literature with citations added at the appropriate locations in the chapter.

A version of Chapter 2 was published in 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, 36(5):S237-S266, 2014. F. Herrmann developed the idea of SVD-free matrix completion techniques for seismic data. B. Recht was involved in the initial theoretical discussion on matrix completion techniques. I collaborated with A. Aravkin and H. Mansour in the theoretical developments of SVD-free matrix completion framework along with its robust and weighting extensions. A. Aravkin coded the SPG-LR framework. I coded all the experiments and prepared all the figures along with the parameters related to the experiments in this chapter. This manuscript was mainly written by A. Aravkin and H. Mansour, and i contributed to the numerical and experimental section. F. Herrmann and B. Recht were involved in the final manuscript editing.

A version of Chapter 3 was published in 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, 80(05): V97-V114, 2015. All the authors were involved in the initial layout of this work and manuscript edits. I, along with C. D. Silva, was responsible for the manuscript and examples. I performed all the matrix completion examples using SPG-LR along with the real data case study and comparison to sparsity-promotion based techniques. O. Akalin performed the Jellyfish comparison example and C. D. Silva performed the ADMM and LmaFit comparison examples.

A version of Chapter 4 was published in R. Kumar, H. Wason, and F. J. Herrmann, " Source separation for simultaneous towed-streamer marine acquisition: a compressed sensing approach", Geophysics, 80(06): WD73-WD88, 2015. H. Wason identified the problem of source separation techniques in seismic literature. I, along with H. Wason, developed the theoretical framework to solve the source separation problem using compressed sensing. I coded the underlying solver

using rank-minimization based techniques to solve the source separation problem. F. Herrmann proposed to incorporate the HSS techniques in the rank-minimization framework. I performed all the experiments using the rank-minimization based techniques, and, H. Wason performed all the experiments using the sparsity-promotion based techniques. I along with H. Wason wrote the manuscript and F. Herrmann was involved in the final manuscript editing. H. Wason and I made an equal contribution (approximately 50%) in developing this research idea for large-scale seismic data applications. This chapter appears as Chapter 7 in H. Wason's dissertation. H. Wason has also granted permission for this chapter to appear in my dissertation.

A version of Chapter 5 was published in R. Kumar, S. Sharan, H. Wason, and F. J. Herrmann, "Time-jittered marine acquisition: a rank-minimization approach for 5D source separation", SEG Technical Program Expanded Abstracts, 2016, p. 119-123. This was my original contribution wherein i proposed the mathematical formulation, coded it and performed the verification examples. S. Sharan helped in simulation of continuous time-domain data. I wrote the abstract where all the coauthors were involved in the abstracts editing.

A version of Chapter 6 was published in T. van Leeuwen, R. Kumar, F. J. Herrmann, "Enabling affordable omnidirectional subsurface extended image volumes via probing", Geophysical Prospecting, 1365-2478, 2016. T. van Leeuwen and F. Herrmann proposed the idea of probing techniques for extended-images. T. van Leeuwen provided the initial framework to compute the extended image volumes. I coded the AVA, 3D extended image volume and the wave-equation migration velocity analysis examples. I, along with T. van Leeuwen, wrote the initial manuscript, and F. Herrmann provided substantial editing of the manuscript during the review process.

MATLAB and its parallel computing toolbox has been used to prepare all the examples in this thesis. The ℓ_1 solver is a public toolbox presented by Ewout van den Berg and Michael P. Friedlander, whereas, SPG-LR solver is coded by Aleksandr Y. Aravkin. The Curvelet toolbox is presented by Emmanuel Candes, Laurent Demanet, David Donoho and Lexing Ying.

Table of Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I want to express my sincere gratitude to my advisor Professor Dr. Felix J. Herrmann. I am deeply indebted for his continuous support of my research, for believing in me, his patience, motivation, and immense knowledge that have made my PhD experience productive. He has taught me, both consciously and unconsciously, the best practices in conducting scientific research.

I would also like to thank my committee members, Professor Dr. Eldad Haber, Professor Dr. Chen Greif, and Professor Dr. Ozgur Yilmaz for serving on my supervisory committee and for generously offering their time, support, and invaluable advice.

I would like to express my heartiest thanks to Dr. Aleksandr Aravkin, Dr. Tristan van Leeuwen, Dr. Hassan Mansour and Dr. Rongrong Wang for their mentorships and friendship, sharing their expertise and experiences, and for providing valuable feedback throughout my research. I specially would like to thank our late post-doc Dr. Ernie Esser (1980-2015) for his friendship, guidance and late evening conversations on research ideas. I miss cycling with him around Vancouver.

I would like to show my special appreciation to Henryk Modzelewski and Miranda Joyce for their support, friendship and generous time during my stay at the SLIM lab. I would like to express my special thanks to them for being great souls. They are always ready to help with a smile. I also would like to thank Manjit Dosanjh and Ian Hanlon for their support and help during my first year in the SLIM-group.

I am forever thankful to my colleagues at the SLIM Lab for their friendship and support, and for creating a cordial working environment. Many thanks to Haneet Wason, Xiang Li, Ning Tu, Shashin Sharan, Felix Oghenekohwo, Curt Da Silva, Oscar Lopez, Zhilong Fang, Art Petrenko, Luz Anglica Caudillo Mata, Tim Lin, Brendan Smithyman, Bas Peter, Ali Alfaraj and other members of the SLIM-group.

My grateful appreciation goes to Dr. James Rickett for giving me the opportunity to do an internship with Schlumberger. I also would like to thank Dr. Can Evren Yarman and Dr. Ivan Vasconcelos for mentorship at Schlumberger, and for creating an enjoyable working environment.

Many thanks to Dr. Eric Verschuur for providing the Gulf of Suez dataset, which I used in Chapters 2 and 4, to PGS for providing the North Sea dataset, and to Chevron for providing the Gulf of Mexico dataset that I used in Chapter 3. I also would like to thank the BG group for

providing the synthetic 3D compass velocity model and 5D seismic data that I used in Chapters 2, 3 and 5. Many thanks to the authors of IWave, SPGℓ_1, SPG-LR, Madagascar, the Marmousi velocity model and the BP salt model, which I used throughout the thesis. 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 and the International Inversion Initiative Project.

Finally, I acknowledge the people who mean a lot to me: my family, my brother and sisters for their continuous and unparalleled love, help and support. My heartfelt regard goes to my father-in-law and mother-in-law for their love and moral support. I owe thanks to a very special person, my wife Monika, for making countless sacrifices to help me get to this point; and my son Rutva for abiding my ignorance and continually providing the requisite breaks from research with his innocent smiles. Monika's unconditional love and support helped me get through this period in the most positive way, and I dedicate this milestone to her.

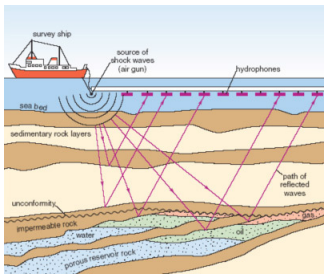
This work was in part financially supported by the NSERC Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD project with support from the following organizations: BG Group, BGP, CGG, Chevron, ConocoPhillips, DownUnder GeoSolutions, Hess, Petrobras, PGS, Sub-Salt Solutions, Schlumberger, and Woodside.

Chapter 1

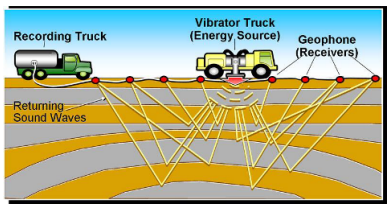
Introduction

Exploration geophysics is an applied branch of geophysics whose aim is to predict the physical properties of the subsurface of the earth, along with the anomalies, using data acquired over the surface of the earth. This data ranges from seismic, gravitational, magnetic, electrical and electromagnetic. Seismic methods are widely used in all the major oil and gas exploration activities around the world because of the better resolution capability of small scale anomalies compared to other existing geophysical methods. Although the acquisition principles are identical for land and marine environments, the operational details such as the geometry and type of receivers systems, the density of measurements made over a given area, and the type of sensors used differ between the two environments [Caldwell and Walker, January 24 2017]. All seismic surveys involve application of a seismic energy source at discrete surface locations, such as a vibroseis truck / shot-hole dynamite on land / air-guns at sea (Figure 1.1). In this thesis, I focus my investigation on the issues related to marine seismic data acquisition, processing and inversion. Figure 1.2 illustrates the different receiver geometries used in marine seismic surveying, while Figure 1.3 provides a list of the different types of surveys.

As outlined in [Caldwell and Walker, January 24 2017], in towed streamer acquisition, a cable is towed or streamed behind a moving vessel that contains the hydrophones, where the length of the streamer can vary between 3 and 12 kilometres depending upon the depth of the geological target of interest. In ocean bottom surveys, the recording system contains a hydrophone and possibly 3-component geophones at each recording location. In vertical seismic profiling, 3-component geophones are placed along vertical and/or horizontal wells. Seismic surveys are often acquired either along single lines or over an area, termed 2D and 3D seismic acquisition, respectively. In 2D acquisition, the sources and receivers are placed along a single sail line below the sea-surface with the underlying assumption being that the reflections are generated in the same 2D vertical plane lying below the sail line. The processing of 2D seismic data generates a 2D image of the subsurface with detailed geological features, i.e., a map of the locations where the acoustic properties of the earth change, which lie beneath the sail line, hence, the name 2D [Caldwell and Walker, January



(a)



(b)

Figure 1.1: Schematic representation of (a) marine and (b) land seismic data acquisition. Source [Enjolras, January 24 2017, RigZone, January 24 2017]

24 2017]. Figure 1.4 shows a standard 2D survey, where the 2D lines are placed on a grid. While the 2D surveys are economical, the gaps between the receiver lines are in kilometres, which makes interpretation of the subsurface problematic if there are strong lateral variations in the earth in the cross-line direction. In that case, 2D assumption fails and we are not able to produce the correct image of the subsurface. This limitation is overcome using 3D acquisition (Figure 1.4), where seismic surveys are acquired over an area resulting in a 3D image of the subsurface, hence the term 3D seismic. Although 3D surveys capture the geological features in detail, they are very expensive because they involve greater investment than 2D surveying in terms of logistics, turnaround acquisition time and sophisticated equipment. Apart from 2D and 3D, seismic surveys are often acquired repeatedly over a producing hydrocarbon field, known as 4D surveys (time-lapse), where the interval between the surveys can be in the order of months or years. The objective of 4D surveys is to estimate reservoir changes as a result of production and/or injection of fluid in the reservoir, by comparing different datasets acquired over a period of time [Caldwell and Walker, January 24 2017].

Seismic data acquisition results in millions of recorded traces with reflection events generated at interfaces between rock layers in the subsurface having different rock properties. Each trace displays the data associated with its common depth point as a continuous function of pressure oscillating on either side of a zero amplitude line (Figure 1.5a). The amplitude of the wiggle is relative to how large the change in rock properties is between two layers. By Society of Exploration Geophysicist (SEG) convention [AAPG, January 24 2017], a reflection event is displayed as a positive peak (polarity is positive) if it is generated from an increase in acoustic impedance, such

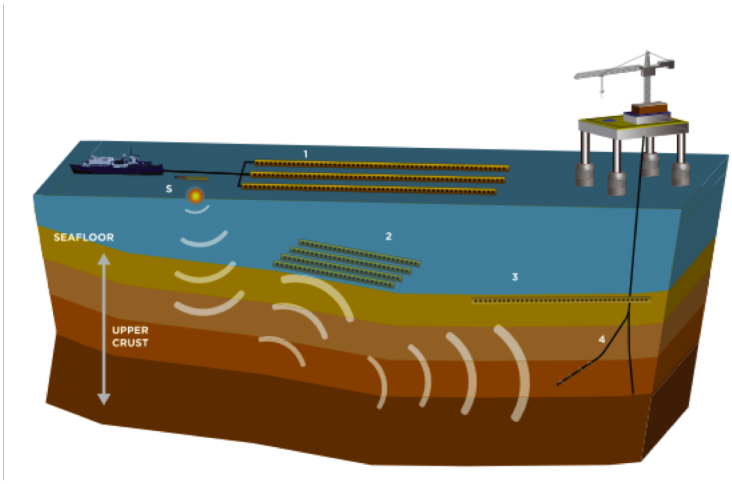


Figure 1.2: Illustration of various marine acquisition geometries, namely towed-streamer (1), an ocean bottom geometry (2), buried seafloor array (3), and Vertical Seismic Profile (4). All the seismic surveys involve a source (S), which is typically an airgun for marine survey, and receivers (black dots) that are mainly hydrophones and/or 3-component geophones. Source Caldwell and Walker [January 24 2017].

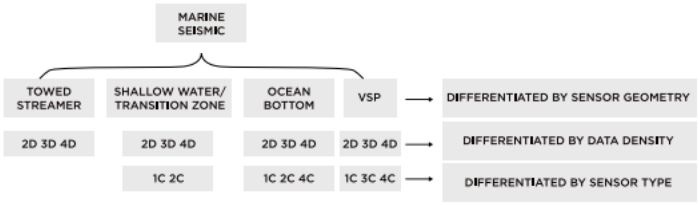


Figure 1.3: This table summarizes the different types of marine seismic surveys. Source Caldwell and Walker [January 24 2017].

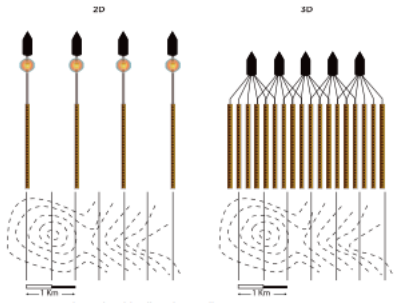


Figure 1.4: Here, we illustrate the basic difference between the 2D and 3D survey geometry. The area covered by the two surveys is exactly identical, as suggested by the dashed contour lines. Source Caldwell and Walker [January 24 2017].

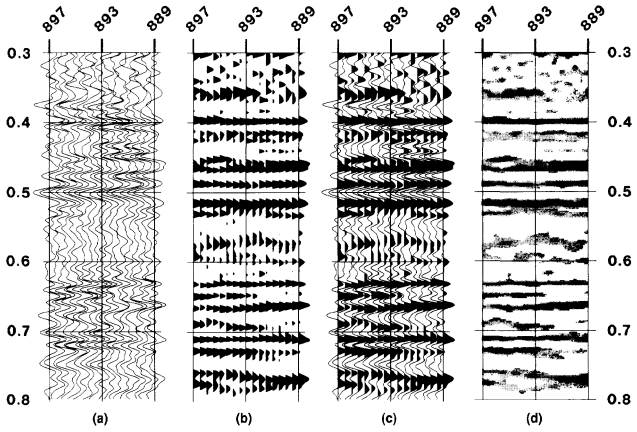


Figure 1.5: Various types of seismic displays: (a) wiggle trace, (b) variable area, (c) variable area wiggle trace, and (d) variable density. Copyright: Conoco Inc.

as from a slow velocity shale to a high velocity dolomite. If a reflection event is generated from the decrease in acoustic impedance, then it is displayed as a trough and the polarity is negative. This convention is called normal polarity. In seismic processing, this reflection energy is mapped to an image and/or attributes of the underground geological structures that are used to infer the physical rock properties.

In realistic seismic data acquisition, each sub-surface point is sampled multiple times to increase the *fold*, where fold is a measure of the redundancy of common midpoint seismic data equal to the number of offset receivers that record a given data point or in a given bin. Fold improves the signal-to-noise ratio and reduces the random environmental noise in the data during the stacking (summation) process to produce the image of the subsurface. Various factors, such as depth and thickness of the zone of interest, surface conditions and topography, play an important role in the design of the seismic acquisition layout. Sources and receivers grid points are controlled by the bin size, which determine how often you sample the subsurface. Bin sizes are smaller in the target-area of interest where the aim is to get a higher resolution image of the subsurface. Apart from the bin size, the frequency spectrum of the data controls the vertical resolution, since higher frequencies have shorter wavelengths and provide a more detailed image of the subsurface.

1.1 Problem statement

Realistically, conventional oil and gas fields are increasingly difficult to explore and produce, calling for more complex wave-equation based inversion (WEI) algorithms requiring dense long-offset samplings and wide-azimuthal coverage. Due to budgetary and/or physical constraints, seismic data acquisition involves coarser sampling (sub-sampling) along either sources or receivers, i.e., seismic data along spatial sampling grids are typically sampled below Nyquist because of cost and certain physical constraints. However, some of the seismic data processing and imaging techniques such as

surface related multiple estimation, amplitude-versus-offset (AVO) and amplitude-versus-azimuth (AVAz) analyses require densely sampled seismic data to avoid acquisition related artifacts in the inverted model of the subsurface. To mitigate these artifacts, we rely on seismic data interpolation methods that result in dense periodically sampled data preferably at or above Nyquist. I included Figures 1.6 (a ,b) to show a fully sampled and 50% subsampled common-receiver gather extracted from a 2D seismic data acquisition where sources and receivers are placed below the sea-surface. The objective of different types of interpolation algorithms are to estimate the missing traces from an undersampled dataset with minimal loss of coherent energy as shown in Figures 1.6 (c, d) and (e,f).

The practitioner also proposed to acquire simultaneous source surveys to reduce costs by reducing acquisition time and environmental impact (shorter disturbance of an area). Simultaneous acquisition also mitigates sampling related issues and improves the quality of seismic data, wherein single and/or multiple source vessels fire sources at near-simultaneous or slightly random times, resulting in overlapping shot records (also known as blending). In general, there are two different types of marine source surveys, namely static and dynamic surveys. During static surveys, sources are towed behind the source vessels and receivers are fixed at the ocean-floor, whereas during dynamic surveys, both sources and receivers are towed behind the source vessels. The current paradigm for simultaneous towed-streamer marine acquisition (dynamic geometry) incorporates *low-variability* in source firing times—i.e., $0 \leq 1$ or 2 seconds, since both the sources and receivers are moving. Figures 1.7 and 1.8 show two instances of dynamic geometry, namely simultaneous long offset [Long et al., 2013] and over/under [Hill et al., 2006, Moldoveanu et al., 2007, Lansley et al., 2007, Long, 2009, Hegna and Parkes, 2012, Torben Hoy, 2013], where we can see the low degree of randomness in the overlapping shots in simultaneous data that presents a challenging case for source separation according to compressed sensing. For static geometry, [Wason and Herrmann, 2013b, Li et al., 2013] proposed an alternate sampling strategy for simultaneous acquisition (time-jittered marine) leveraging ideas from compressed sensing (CS), where a single source vessel sails across an ocean-bottom array continuously firing two airgun arrays at jittered source locations and time instances with receivers recording continuously (Figure 1.9). This results in a continuous time-jittered simultaneous recording of seismic data (Figure 1.9). Simultaneous acquisitions are economically viable, since overall acquisition becomes compressed, where we record the overlapping shot records. While these cost savings are very welcome, especially in the current downturn, subsequent seismic data processing and imaging workflows expect recordings where the data is collected for sequential shots. So our task is to turn continuous recordings with overlapping shots into sequential recording with the least amount of artifacts and loss of energy at late arrival time. Therefore, a practical (simultaneous) source separation technique is required, which aims to recover unblended (non-overlapping) data—as acquired during conventional acquisition—from simultaneous data. However, all the interpolation and source separation workflows for 3D seismic data acquisition results in exponential growth in data volumes because the recovered seismic data contains traces on the order of millions,

and prohibitive demands on computational resources. Given the data size volumes and resources, one of the key challenges is to extract meaningful information from this huge dataset (which may in the near future grow as large as on the order of petabytes) in computationally efficient ways, i.e., to reduce the turnaround time of each of the processing and inversion steps, apart from storing the massive volumes of interpolated and separated seismic data on disks.

1.2 Objectives

The primary focus of this thesis is to propose fast computational techniques that are practical and robust for large-scale seismic data processing, namely missing-trace interpolation, source separation, and wave-equation based migration velocity analysis. The main objectives of this work are:

- to identify three basic principles of compressed sensing [Donoho, 2006b] for recovering seismic data volumes using rank-minimization based techniques, namely a low-rank transform domain, a rank-increasing sampling scheme, and a practical SVD-free rank-minimizing optimization scheme for large-scale seismic data processing;
- to decrease the cost of large-scale simultaneous source acquisition for towed streamers and ocean-bottom surveys by leveraging ideas from compressed sensing via randomization in source locations and time instances followed by (SVD-free) computationally efficient source separation;
- to derive a computationally feasible two-way wave-equation based factorization principle that gives us access to the kinematics and amplitudes of full subsurface offset extended image volumes without carrying out explicit cross-correlations between source and receiver wavefields for each shot.

1.3 Contributions

To our knowledge this work represents the first instance where seismic data processing, namely missing-trace interpolation and source separation, is performed using full 5D seismic data volume by avoiding windowed based operations. In this work, I show the benefits of large-scale SVD-free framework in terms of the computational time and memory requirements. We also propose a novel matrix-vector formulation to extract information from the full-subsurface offset extended image volumes that overcome prohibitive computational and storage costs of forming full-subsurface offset extended image volumes, which cannot be formed using conventional extended imaging workflows. I show that the purpose matrix-vector formulation offers new perspectives on the design and implementation of workflows that exploit information embedded in various types of subsurface extended images. I further demonstrate the benefits of this matrix-vector formulation to obtain local information, tied to individual subsurface points that can serve as quality control for velocity analyses

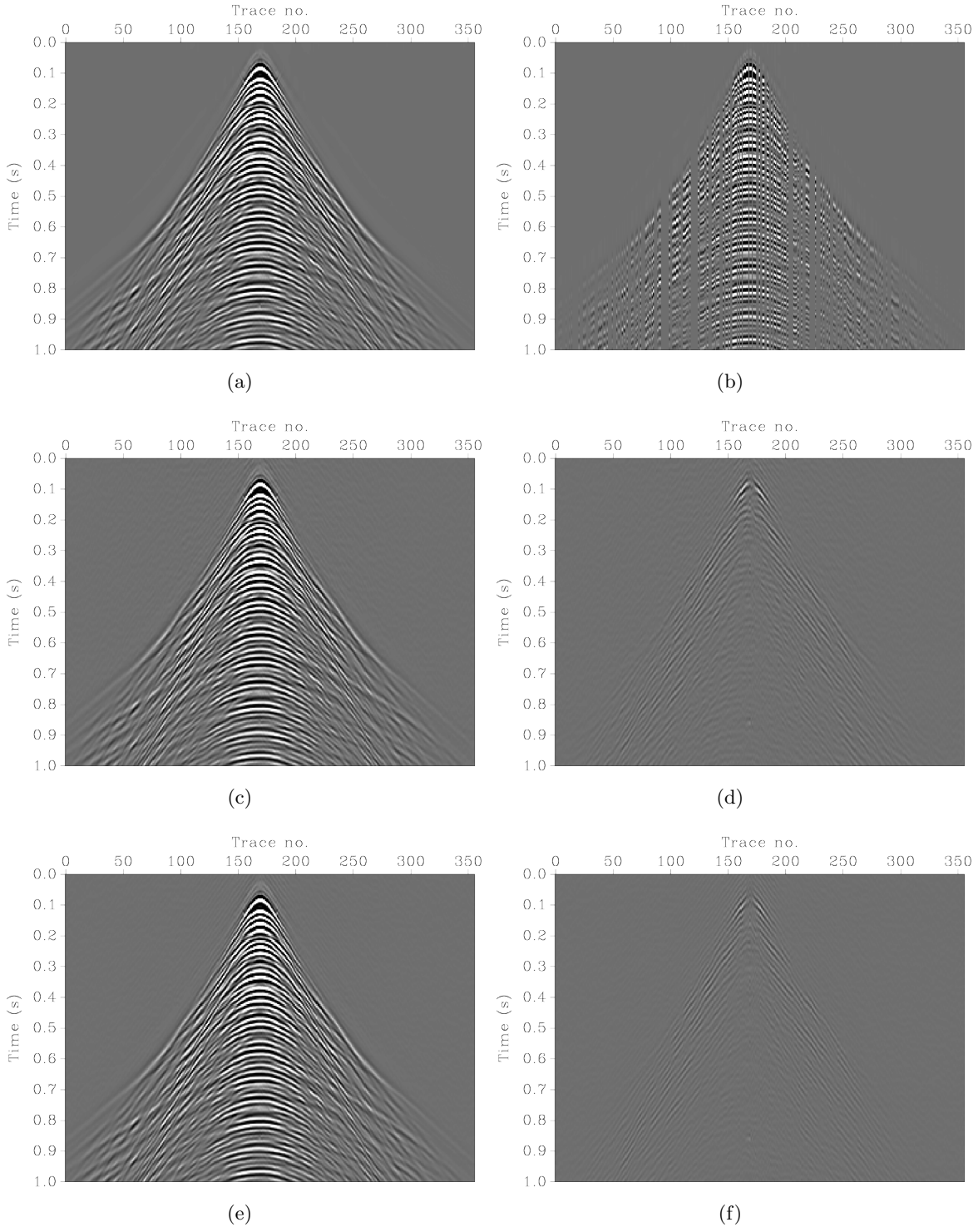
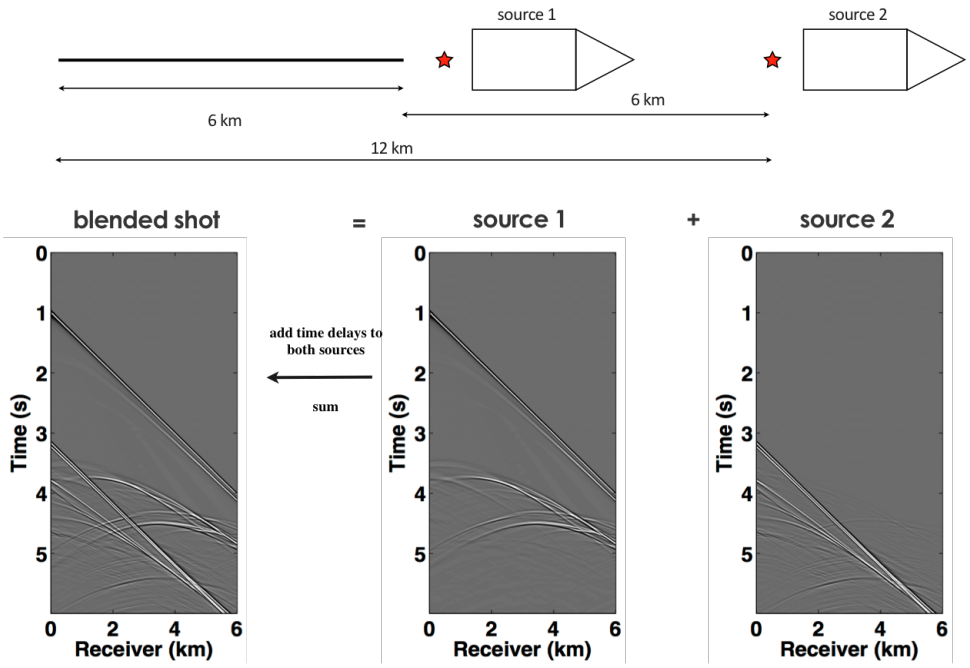


Figure 1.6: Common-receiver gather. (a) Fully sampled and (b) 50% subsampled. The final goal is to recover the fully-sampled data from the subsampled data with minimal loss of coherent energy. (c, e) Reconstruction results from two different types of interpolation and (d, f) corresponding residual plots. We can see that the interpolation results in (e, f) are better than (c, d) because the energy loss is small, especially at the cusp of the common-receiver gather.



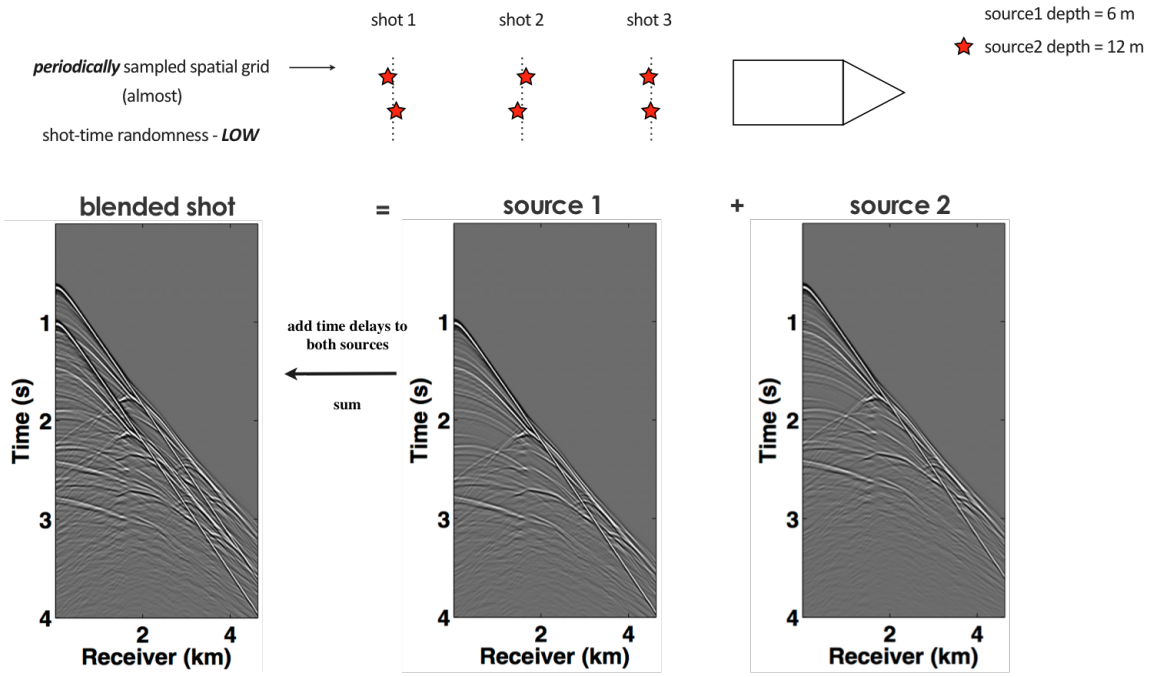
(a)

Figure 1.7: Simultaneous long-offset acquisition, where an extra source vessel is deployed, sailing one spread-length ahead of the main seismic vessel (see Chapter 4 for more details). We record overlapping shot records in the field and separate them into non-overlapping shots using source separation based techniques.

or as input to localized amplitude-versus-offset analyses, or global information that can be used to drive automatic velocity analyses without requiring prior information on the geologic dip.

1.4 Outline

In Chapters 2 and 3, we first introduce the underlying theory of matrix completion and its SVD-free approach for large-scale missing-trace interpolation problem. Next, we outline three practical principles for using low-rank optimization techniques to recover missing seismic data which are built upon theoretical ideas from Compressed Sensing. We further address the computational challenges of using the matrix-based techniques for seismic data reconstruction, where we propose to use either a (SVD-free) factorization based rank-minimization framework with the Pareto curve approach or the factorization-based parallel matrix completion framework dubbed Jellyfish. We also examine the popular approach of windowing a large data volume into smaller data volumes to be processed in parallel, and empirically demonstrate how such a process does not respect the inherent redundancy present in the data, degrading reconstruction quality as a result. Finally, I demonstrate these observations on carefully selected real 2D seismic surveys and synthetic 3D seismic surveys simulated using a complex velocity model provided by the BG Group. We also show



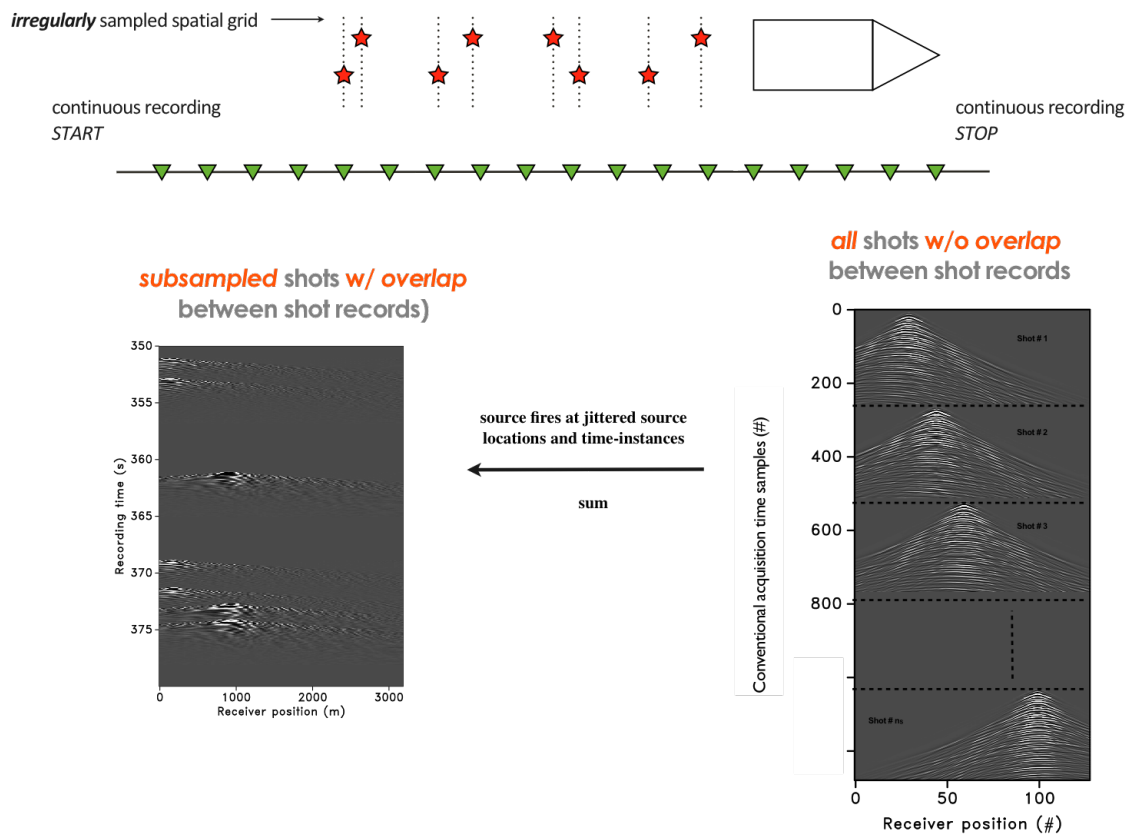
(a)

Figure 1.8: Over/Under acquisition is an instance of low-variability in source firing times, i.e, two sources are firing within 1 (or 2) seconds (see Chapter 4 for more details).

the computational advantages of matrix completion techniques over sparsity-promoting techniques and tensor-based missing-trace reconstruction techniques.

In Chapter 4, I extend the low-rank optimization based approaches to perform source separation in simultaneous towed-streamer marine acquisition, where I modified the matrix completion formulation to separate multiple sources acquired simultaneously. We address the challenge of source separation for simultaneous towed-streamer acquisitions via two compressed sensing based approaches, namely sparsity-promotion and rank-minimization. We exploit the sparse structure of seismic data in the curvelet domain and low-rank structure of seismic data in the midpoint-offset domain. I further incorporate the Hierarchical Semi-Separable matrix representation in rank-minimization framework to exploit the low-rank structure of seismic data at higher frequencies. Finally, we illustrate the performance of both the sparsity-promotion and rank-minimization based techniques by simulating two simultaneous towed-streamer acquisition scenarios: over/under and simultaneous long offset. A field data example from the Gulf of Suez for the over/under acquisition scenario is also included. I further compare these two techniques with the NMO-based median filtering type approach.

In Chapter 5, I use rank-minimization based techniques to present a computationally tractable algorithm to separate simultaneous time-jittered continuous recording for a 3D ocean-bottom cable survey. First, I formulate a factorization based rank-minimization formulation that works on the



(a)

Figure 1.9: Time-jittered marine continuous acquisition, 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

temporal-frequency domain using all monochromatic data matrices together. Then, I show the efficacy of proposed framework on a synthetic 3D seismic survey simulated on a complex geological velocity model provided by the BG Group.

In Chapter 6, we exploit the redundancy in extended image volumes, by first arranging the extended image volume as a matrix, followed by probing this matrix in a way that avoids explicit storage and removes the customary and expensive loop over shots found in conventional extended imaging. As a result, we end up with a matrix-vector formulation where I form different image gathers and use them to perform amplitude-versus-angle and wave-equation migration velocity analyses, without requiring prior information on the geologic dips. Next, I show how this factorization can be used to extract information on local geological dips and radiation patterns for AVA purposes, and how full-subsurface offset extended image volumes can be used to carry out automatic WEMVA. Each application is illustrated by carefully selected stylized numerical examples on 2- and 3-D velocity models.

In Chapter 7, I conclude the work presented in this thesis and propose future research directions.

Chapter 2

Fast methods for denoising matrix completion formulations, with applications to robust seismic data interpolation.

2.1 Summary

Recent SVD-free matrix factorization formulations have enabled rank minimization for systems with millions of rows and columns, paving the way for matrix completion in extremely large-scale applications, such as seismic data interpolation.

In this paper, we consider matrix completion formulations designed to hit a target data-fitting error level provided by the user, and propose an algorithm called LR-BPDN that is able to exploit factorized formulations to solve the corresponding optimization problem. Since practitioners typically have strong prior knowledge about target error level, this innovation makes it easy to apply the algorithm in practice, leaving only the factor rank to be determined.

Within the established framework, we propose two extensions that are highly relevant to solving practical challenges of data interpolation. First, we propose a weighted extension that allows known subspace information to improve the results of matrix completion formulations. We show how this weighting can be used in the context of frequency continuation, an essential aspect to seismic data interpolation. Second, we propose matrix completion formulations that are robust to large measurement errors in the available data.

We illustrate the advantages of LR-BPDN on collaborative filtering problem using the Movie-

A version of this chapter has been published in SIAM Journal on Scientific Computing, 2014, vol 36, pages S237-S266.

Lens 1M, 10M, and Netflix 100M datasets. Then, we use the new method, along with its robust and subspace re-weighted extensions, to obtain high-quality reconstructions for large scale seismic interpolation problems with real data, even in the presence of data contamination.

2.2 Introduction

Sparsity- and rank-regularization have had significant impact in many areas over the last several decades. Sparsity in certain transform domains has been exploited to solve underdetermined linear systems with applications to compressed sensing Donoho [2006a], Candès and Tao [2006], natural image denoising/inpainting Starck et al. [2005], Mairal et al. [2008], Mansour et al. [2010], and seismic image processing Herrmann and Hennenfent [2008a], Neelamani et al. [2010], Herrmann et al. [2012a], Mansour et al. [2012b]. Analogously, low-rank structure has been used to efficiently solve matrix completion problems, such as the Netflix Prize problem, along with many other applications, including control, system identification, signal processing, and combinatorial optimization Fazel [2002], Recht et al. [2010b], Candès et al. [2011], and seismic data interpolation and denoising Oropeza and Sacchi [2011].

Regularization formulations for both types of problems introduce a regularization functional of the decision variable, either by adding an explicit penalty to the data-fitting term

$$\min_x \rho(\mathcal{A}(x) - b) + \lambda||x||, \tag{QP$_\lambda$}$$

or by imposing constraints

$$\min_x \rho(\mathcal{A}(x) - b) \quad \text{s.t. } ||x|| \leq \tau \ . \tag{LASSO$_\tau$}$$

In these formulations, x may be either a matrix or a vector, $\|\cdot\|$ may be a sparsity or low-rank promoting penalty such as the ℓ_1 norm $\|\cdot\|_1$ or the matrix nuclear norm $\|\cdot\|_*$, $\mathcal{A}$ may be any linear operator that predicts the observed data vector b of size $p \times 1$, and $\rho(\cdot)$ is typically taken to be the 2-norm.

These approaches require the user to provide regularization parameters whose values are typically not known ahead of time, and otherwise may require fitting or cross-validation procedures.

The alternate formulation

$$\min_x ||x|| \quad \text{s.t. } \rho(\mathcal{A}(x) - b) \leq \eta. \tag{BPDN$_\eta$}$$

has been successfully used for the sparse regularization of large scale systems Berg and Friedlander [2008], and proposed for nuclear norm regularization Berg and Friedlander [2011]. This formulation requires the user to provide an acceptable error bound in the data fitting domain (BPDN$_\eta$), and is preferable for many applications, especially when practitioners know (or are able to estimate) an approximate data error level. We refer to (BPDN$_\eta$), (QP$_\lambda$) and (LASSO$_\tau$) as *regularization*

formulations, since all three limit the space of feasible solutions by considering the nuclear norm of the decision variable.

A practical implementation of (BPDN$_\eta$) for large scale matrix completion problems is difficult because of the large size of the systems of interest, which makes SVD-based approaches intractable. For example, seismic inverse problems work with 4D data volumes, and matricization of such data creates structures whose size is a bottleneck for standard low-rank interpolation approaches. Fortunately, a growing literature on factorization-based rank optimization approaches has enabled matrix completion formulations for (QP$_\lambda$) and (LASSO$_\tau$) approaches for extremely large-scale systems that avoids costly SVD computations Rennie and Srebro [2005b], Lee et al. [2010b], Recht and Ré [2011]. These formulations are non-convex, and therefore do not have the same convergence guarantees as convex formulations for low-rank factorization. In addition, they require an *a priori* rank specification, adding a rank constraint to the original problem. Nonetheless, factorized formulations can be shown to avoid spurious local minima, so that if a local minimum is found, it will correspond to the global minimum in the convex formulation, provided the chosen factor rank was high enough. In addition, computational methods for factorized formulations are more efficient, mainly because they can completely avoid SVD (or partial SVD) computations. In this paper, we extend the framework of Berg and Friedlander [2011] to incorporate matrix factorization ideas, enabling the (BPDN$_\eta$) formulation for rank regularization of large scale problems, such as seismic data interpolation.

While formulations in Berg and Friedlander [2008, 2011] choose ρ in (BPDN$_\eta$) to be the quadratic penalty, recent extensions Aravkin et al. [2013a] allow more general penalties to be used. In particular, robust convex (see e.g. Huber [1981]) and nonconvex penalties (see e.g. Lange et al. [1989], Aravkin et al. [2012]) can be used to measure misfit error in the (BPDN$_\eta$) formulation. We incorporate these extensions into our framework, allowing matrix completion formulations that are robust to data contamination.

Finally, subspace information can be used to inform the matrix completion problem, analogously to how partial support information can be used to improve the sparse recovery problem Friedlander et al. [2011]. This idea is especially important for seismic interpolation, where *frequency continuation* is used. We show that subspace information can be incorporated into the proposed framework using reweighting, and that the resulting approach can improve recovery SNR in a frequency continuation setting. Specifically, subspace information obtained at lower frequencies can be incorporated into reweighted formulations for recovering data at higher frequencies.

To summarize, we design factorization-based formulations and algorithms for matrix completion that

1. Achieve a specified target misfit level provided by the user (i.e. solve (BPDN$_\eta$)).
2. Achieve recovery in spite of severe data contamination using robust cost functions ρ in (BPDN$_\eta$)
3. Incorporate subspace information into the inversion using re-weighting.

The paper proceeds as follows. In section 2.3, we briefly discuss and compare the formulations (QP_λ), (LASSO_τ), and (BPDN_η). We also review the SPGℓ_1 algorithm Berg and Friedlander [2008] to solve (BPDN_η), along with recent extensions for (BPDN_η) formulations developed in Aravkin et al. [2013a]. In section 2.4, we formulate the convex relaxation for the rank optimization problem, and review SVD-free factorization methods. In section 2.5, we extend analysis from Burer and Monteiro [2003] to characterize the relationship between local minima of rank-optimization problems and their factorized counterparts in a general setting that captures all formulations of interest here. In section 2.6, we propose an algorithm that combines matrix factorization with the approach developed by Berg and Friedlander [2008, 2011], Aravkin et al. [2013a]. We develop the robust extensions in section 2.7, and reweighting extensions in section 2.8. Numerical results for both the Netflix Prize problem and for seismic trace interpolation of real data are presented in section 2.9.

2.3 Regularization formulations

Each of the three formulations (QP_λ), (LASSO_τ), and (BPDN_η) controls the tradeoff between data fitting and a regularization functional using a regularization parameter. However, there are important differences between them.

From an optimization perspective, most algorithms solve (QP_λ) or (LASSO_τ), together with a continuation strategy to modify τ or λ, see e.g., Figueiredo et al. [2007], Berg and Friedlander [2008]. There are also a variety of methods to determine optimal values of the parameters; see e.g. Giryes et al. [2011] and the references within. However, from a modeling perspective (BPDN_η) has a significant advantage, since the η parameter can be directly interpreted as a *noise floor*, or a threshold beyond which noise is commensurate with the data. In many applications, such as seismic data interpolation, scientists have good prior knowledge of the noise floor. In the absence of such knowledge, one still wants an algorithm that returns a reasonable solution given a fixed computational budget, and some formulations for solving (BPDN_η) satisfy this requirement.

van den Berg and Friedlander Berg and Friedlander [2008] proposed the SPGℓ_1 algorithm for optimizing (BPDN_η) that captures the features discussed above. Their approach solves (BPDN_η) using a series of inexact solutions to (LASSO_τ). The bridge between these problems is provided by the *value function* $v : \mathbb{R} \to \mathbb{R}$

$$v(\tau) = \min_x \rho(\mathcal{A}(x) - b) \quad \text{s.t. } \|x\| \leq \tau \ , \tag{2.1}$$

where the particular choice of $\rho(\cdot) = \|\cdot\|^2$ was made in Berg and Friedlander [2008, 2011]. The graph of $v(\tau)$ is often called the *Pareto curve*. The (BPDN_η) problem can be solved by finding the

root of $v(\tau) = \eta$ using Newton's method:

$$\tau^{k+1} = \tau^k - \frac{v(\tau) - \eta}{v'(\tau)} \,, \tag{2.2}$$

and the quantities $v(\tau)$ and $v'(\tau)$ can be approximated by solving (LASSO$_\tau$) problems. In the context of sparsity optimization, (BPDN$_\eta$) and (LASSO$_\tau$) are known to be equivalent for certain values of parameters τ and η. Recently, these results were extended to a much broader class of formulations (see [Aravkin et al., 2013a, Theorem 2.1]). Indeed, convexity of ρ is not required for this theorem to hold, and instead *activity* of the constraint at the solution plays a key role. The main hypothesis requires that the constraint is active at any solution $\overline{x}$, i.e. $\rho(b - \mathcal{A}(\overline{x})) = \sigma$, and $\|\overline{x}\| = \tau$.

For any ρ, $v(\tau)$ is non-increasing, since larger τ allow a bigger feasible set. For any convex ρ in (2.1), $v(\tau)$ is convex by inf-projection [Rockafellar and Wets, 1998, Proposition 2.22]. When ρ is also differentiable, it follows from [Aravkin et al., 2013a, Theorem 5.2] that $v(\tau)$ is differentiable, with derivative given in closed form by

$$v'(\tau) = -\|\mathcal{A}^* \nabla \rho(b - \mathcal{A}\bar{x})\|_d \,, \tag{2.3}$$

where $\mathcal{A}^*$ is the adjoint to the operator $\mathcal{A}$, $\|\cdot\|_d$ is the dual norm to $\|\cdot\|$, and $\bar{x}$ solves LASSO$_\tau$. For example, when the norm $\|\cdot\|$ in (2.1) is the 1-norm, the dual norm is the infinity norm, and (2.3) evaluates to the maximum absolute entry of the gradient. In the matrix case, $\|\cdot\|$ is typically taken to be the nuclear norm, and then $\|\cdot\|_d$ is the spectral norm, so (2.3) evaluates to the maximum singular value of $\mathcal{A}^* \nabla \rho(r)$.

To design effective optimization methods, one has to be able to evaluate $v(\tau)$, and to compute the dual norm $\|\cdot\|_d$. Evaluating $v(\tau)$ requires solving a sequence of optimization problems (2.1), for the sequence of τ given by (2.2).

A key idea that makes the approach of Berg and Friedlander [2008] very useful in practice is solving LASSO problems inexactly, with increasing precision as the overarching Newton's method proceeds. The net computation is therefore much smaller than what would be required if one solved a set of LASSO problems to a pre-specified tolerance. For large scale systems, the method of choice is typically a first-order method, such as spectral projected gradient where after taking a step along the negative gradient of the mismatch function $\rho(\mathcal{A}(x) - b)$, the iterate is projected onto the norm ball $\|\cdot\| \leq \tau$. Fast projection is therefore a necessary requirement for tractable implementation, since it is used in every iteration of every subproblem.

With the inexact strategy, the convergence rate of the Newton iteration (2.2) may depend on the conditioning of the linear operator $\mathcal{A}$ [Berg and Friedlander, 2008, Theorem 3.1]. For well-conditioned problems, in practice one can often observe only a few (6-10) (LASSO$_\tau$) problems to find the solution for (BPDN$_\eta$) for a given η. As the optimization proceeds, (LASSO$_\tau$) problems for

larger τ warm-start from the solution corresponding to the previous τ.

2.4 Factorization approach to rank optimization

We now consider (BPDN$_\eta$) in the specific context of rank minimization. In this setting, $\|\cdot\|$ is taken to be the nuclear norm, where for a matrix $X \in \mathbb{R}^{n\times m}$, $\|X\|_* = \|\sigma\|_1$, where σ is the vector of singular values. The dual norm in this case is $\|\sigma\|_\infty$, which is relatively easy to find for very large systems.

Unfortunately, solving the optimization problem in (2.1) is much more difficult. For the large system case, this requires repeatedly projecting onto the set $\|X\|_* \leq \tau$, which which means repeated SVD or partial SVD computations. This is not feasible for large systems.

Factorization-based approaches allow matrix completion for extremely large-scale systems by avoiding costly SVD computations Rennie and Srebro [2005b], Lee et al. [2010a], Recht and Ré [2011]. The main idea is to parametrize the matrix X as a product,

$$X = LR^T \,, \tag{2.4}$$

and to optimize over the factors L, R. If $X \in \mathbb{R}^{n\times m}$, then $L \in \mathbb{R}^{n\times k}$, and $R \in \mathbb{R}^{m\times k}$. The decision variable therefore has dimension $k(n+m)$, rather than nm; giving tremendous savings when $k \ll m, n$. The asymptotic computational complexity of factorization approaches is the same as that of partial SVDs, as both methods are dominated by an O(nmk) cost; the former having to form $X = LR^T$, and the latter computing partial SVDs, at every iteration. However, in practice the former operation is much simpler than the latter, and factorization methods outperform methods based on partial SVDs. In addition, factorization methods keep an explicit bound on the rank of all iterates, which might otherwise oscillate, increasing the computational burden.

Using representation (2.4), the projection problem in (LASSO$_\tau$) is trivial. For the nuclear norm, we have Rennie and Srebro [2005b]

$$\|X\|_* = \inf_{X=LR^T} \frac{1}{2} \left\| \begin{bmatrix} L \\ R \end{bmatrix} \right\|_F^2 \,, \tag{2.5}$$

and therefore for any partiular L, R, we have

$$\|X\|_* = \|LR^T\|_* \leq \frac{1}{2} \left\| \begin{bmatrix} L \\ R \end{bmatrix} \right\|_F^2 \,. \tag{2.6}$$

The nuclear norm is not the only formulation that can be factorized. Lee et al. [2010b] have recently introduced the max norm, which is closely related to the nuclear norm and has been successfully used for matrix completion.

2.5 Local minima correspondence between factorized and convex formulations

All of the algorithms we propose for matrix completion are based on the factorization approach described above. Even though the change of variables $X = LR^T$ makes the problem nonconvex, it turns out that for a surprisingly general class of problems, this change of variables does not introduce any extraneous local minima, and in particular any local minimum of the factorized (non-convex) problem corresponds to a local (and hence global) minimum of the corresponding un-factorized convex problem. This result appeared in [Burer and Monteiro, 2003, Proposition 2.3] in the context of semidefinite programming (SDP); however, it holds in general, as the authors point out [Burer and Monteiro, 2003, p. 431].

Here, we state the result for a broad class of problems, which is general enough to capture all of our formulations of interest. In particular, the continuity of the objective function is the main hypothesis required for this correspondence. It is worthwhile to emphasize this, since in Section 2.7, we consider smooth non-convex robust misfit penalties for matrix completion, which give impressive results (see figure 2.11).

For completeness, we provide a proof in the appendix.

Theorem 1 (General Factorization Theorem) *Consider an optimization problem of the form*

$$\begin{aligned} \min_{Z \succeq 0} \quad & f(Z) \\ s.t. \quad & g_i(Z) \leq 0 \quad i = 1, \ldots, n \\ & h_j(Z) = 0 \quad j = 1, \ldots, m \\ & \operatorname{rank}(Z) \leq r, \end{aligned} \tag{2.7}$$

where $Z \in \mathbb{R}^{n \times n}$ is positive semidefinite, and f, g_i, h_i are continuous. Using the change of variable $Z = SS^T$, take $S \in \mathbb{R}^{n \times r}$, and consider the problem

$$\begin{aligned} \min_{S} \quad & f(SS^T) \\ s.t. \quad & g_i(SS^T) \leq 0 \quad i = 1, \ldots, n \\ & h_j(SS^T) = 0 \quad j = 1, \ldots, m \end{aligned} \tag{2.8}$$

Let $\bar{Z} = \bar{S}\bar{S}^T$, where $\bar{Z}$ is feasible for (2.7)*. Then $\bar{Z}$ is a local minimum of* (2.7) *if and only if $\bar{S}$ is a local minimum of* (2.8)*.*

At first glance, Theorem 1 seems restrictive to apply to a recovery problem for a generic X, since it is formulated in terms of a PSD variable Z. However, we show that all of the formulations of interest can be expressed this way, due to the SDP characterization of the nuclear norm.

It was shown in [Recht et al., 2010b, Sec. 2] that the nuclear norm admits a semi-definite

programming (SDP) formulation. Given a matrix $X \in \mathbb{R}^{n\times m}$, we can characterize the nuclear norm $\|X\|_*$ in terms of an auxiliary matrix positive semidefinite marix $Z \in \mathbb{R}^{(n+m)\times(n+m)}$

$$\begin{aligned} &\|X\|_* = \min_{Z\succeq 0} \tfrac{1}{2}\mathrm{Tr}(Z) \\ &\text{subject to } Z_{1,2} = Z_{2,1}^T = X \ , \end{aligned} \tag{2.9}$$

where $Z_{1,2}$ is the upper right $n \times m$ block of Z, and $Z_{2,1}$ is the lower left $m \times n$ block. More precisely, the matrix Z is a symmetric positive semidefinite matrix having the structure

$$Z = \begin{bmatrix} L \\ R \end{bmatrix} \begin{bmatrix} L^T & R^T \end{bmatrix} = \begin{bmatrix} LL^T & X \\ X^T & RR^T \end{bmatrix}, \tag{2.10}$$

where L and R have the same rank as X, and $\mathrm{Tr}(Z) = \|L\|_F^2 + \|R\|_F^2$.

Using characterization (2.9)-(2.10), we can show that a broad class of formulations of interest in this paper are in fact problems in the class characterized by Theorem 1.

Corollary 1 (General Matrix Lasso) *Any optimization problem of the form*

$$\begin{aligned} \min_X \quad & f(X) \\ s.t. \quad & \|X\|_* \leq \tau \\ & \mathit{rank}(X) \leq r \end{aligned} \tag{2.11}$$

where f is continuous has an equivalent problem in the class of problems (2.7) *characterized by Theorem 1.*

Proof 1 *Using* (2.9)*, write* (2.11) *as*

$$\begin{aligned} \min_{Z\succeq 0} \quad & f(\mathcal{R}(Z)) \\ s.t. \quad & \mathrm{Tr}(Z) \leq \tau \\ & \mathit{rank}(Z) \leq r, \end{aligned} \tag{2.12}$$

where $\mathcal{R}(Z)$ extracts the upper right $n \times m$ block of Z. It is clear that if $\mathit{rank}(Z) \leq r$, then $\mathit{rank}(X) \leq r$, so every solution feasible for the problem in Z is feasible for the problem in X by (2.9)*. On the other hand, we can use the SVD of any matrix X of rank r to write $X = LR^T$, with $\mathit{rank}(L) = \mathit{rank}(R) = r$, and then the matrix Z in* (2.10) *has rank r, contains X in its upper right hand corner, and has as its trace the nuclear norm of X. In particular, if $X = U\Sigma V^T$, we can use $L = U\sqrt{\Sigma}$, and $R = V\sqrt{\Sigma}$ to get this representation. Therefore, every feasible point for the X problem has a corresponding Z.*

2.6 LR-BPDN algorithm

The factorized formulations in the previous section have been used to design several algorithms for large scale matrix completion and rank minimization Lee et al. [2010b], Recht and Ré [2011]. However, all of these formulations take the form (QP_λ) or (LASSO_τ). The (LASSO_τ) formulation enjoys a natural relaxation interpretation, see e.g. Herrmann et al. [2012b]; on the other hand, a lot of work has focused on methods for λ-selection in (QP_λ) formulations, see e.g. Giryes et al. [2011]. However, both formulations require some identification procedure of the parameters λ and τ.

Instead, we propose to use the factorized formulations to solve the (BPDN_η) problem by traversing the Pareto curve of the nuclear norm minimization problem. In particular, we integrate the factorization procedure into the SPGℓ_1 framework, which allows to find the minimum rank solution by solving a sequence of factorized (LASSO_τ) subproblems (2.14). The cost of solving the factorized (LASSO_τ) subproblems is relatively cheap and the resulting algorithm takes advantage of the inexact subproblem strategy in Berg and Friedlander [2008].

For the classic nuclear norm minimization problem, we define

$$v(\tau) = \min_X \|\mathcal{A}(X) - b\|_2^2 \quad \text{s.t. } \|X\|_* \leq \tau \;, \tag{2.13}$$

and find $v(\tau) = \eta$ using the iteration (2.2).

However, rather than parameterizing our problem with X, which requires SVD for each projection, we use the factorization formulation, exploiting Theorem 1 and Corollary 1. Specifically, when evaluating the value function $v(\tau)$, we solve the corresponding factorized formulation

$$\min_{L,R} \|\mathcal{A}(LR^T) - b\|_2^2 \quad \text{s.t. } \frac{1}{2}\left\|\begin{bmatrix} L \\ R \end{bmatrix}\right\|_F^2 \leq \tau \tag{2.14}$$

using decision variables L, R with a fixed number of k columns each.

By Theorem 1 and Corollary 1, any local solution to this problem corresponds to a local solution of the true LASSO problem, subject to a rank constraint $\text{rank}(X) \leq k$. We use residual $\mathcal{A}(\overline{LR}^T) - b$ reconstructed from (2.14) to evaluate both $v(\tau)$ and its derivative $v'(\tau)$. When the rank of L, R is large enough, a local minimum of (2.14) corresponds to a local minimum of (2.13), and for any convex ρ, every local minimum of (LASSO_τ) is also a global minimum. When the rank of the factors L and R is smaller than the rank of the optimal LASSO solution, the algorithm looks for local minima of the rank-constrained LASSO problem. Unfortunately, we cannot guarantee that the solutions we find are local minima for (2.14), rather than simply stationary points. Nonetheless, this approach works quickly and reliably in practice, as we show in our experiments.

Problem (2.14) is optimized using the spectral projected gradient algorithm. The gradient is easy to compute, and the projection requires rescaling all entries of L, R by a single value, which is fast, simple, and parallelizable.

To evaluate $v'(\tau)$, we use the formula (2.3) for the Newton step corresponding to the original (convex) problem in X; this requires computing the spectral norm (largest singular value) of

$$\mathcal{A}^*(b - \mathcal{A}(\bar{L}\bar{R}^T)) \ ,$$

where $\mathcal{A}^*$ is the adjoint of the linear operator $\mathcal{A}$, while $\bar{L}$ and $\bar{R}$ are the solutions to (2.14). The largest singular value of the above matrix can be computed relatively quickly using the power method. Again, at every update requiring $v(\tau)$ and $v'(\tau)$, we are assuming here that our solution $\bar{X} = \bar{L}\bar{R}^T$ is close to a local minimum of the true LASSO problem, but we do not have theoretical guarantees of this fact.

2.6.1 Initialization

The factorized LASSO problem (2.14) has a stationary point at $L = 0, R = 0$. This means that in particular, we cannot initialize from this point. Instead, we recommend initializing from a small random starting point. Another possibility is trying to jump start the algorithm, for example using the initialization technique of [Jain et al., 2013, Algorithm 1]. One can compute the partial SVD of the adjoint of the linear operator $\mathcal{A}$ on the observed data:

$$USV^T = \mathcal{A}^*b$$

Then L and R are initialized as

$$L = U\sqrt{S}, \quad R = V\sqrt{S}.$$

This initialization procedure can sometimes result in faster convergence over random initialization. Compared to random initialization, this method has the potential to reduce the runtime of the algorithm by 30-40% for smaller values of η, see Table 2.1. The key feature of any initialization procedure is to ensure that the starting value of

$$\tau_0 = \frac{1}{2}\begin{bmatrix} L_0 \\ R_0 \end{bmatrix}$$

is *less* than the solution to the root finding problem for (BPDN$_\eta$), $v(\tau) = \eta$.

2.6.2 Increasing k on the fly

In factorized formulations, the user must specify a factor rank. From a computational perspective, it is better that the rank stay small; however if it is too small, it may be impossible to solve (BPDN$_\eta$) to a specified error level η. For some classes of problems, where the true rank is known ahead of time (see e.g. Candès et al. [2013]), one is guaranteed that a solution will exist for a given rank.

Table 2.1: Summary of the computational time (in seconds) for LR-BPDN, measuring the effect of random versus smart ([Jain et al., 2013, Algorithm 1]) initialization of L and R for factor rank k and relative error level η for (BPDN$_\eta$). Comparison performed on the 1M MovieLens Dataset. Type of initialization had almost no effect on quality of final reconstruction.

	Random initialization			
k	10	20	30	50
η=0.5	3.54	5.46	4.04	8.31
η=0.3	11.90	6.14	8.42	20.84
η=0.2	86.53	107.88	148.12	166.92

	Smart initialization			
k	10	20	30	50
η=0.5	5.01	5.75	6.84	8.24
η=0.3	11.15	18.88	12.38	21.02
η=0.2	58.78	84.29	95.07	114.21

However, if necessary, factor rank can be adjusted on the fly within our framework.

Specifically, adding columns to L and R can be done on the fly, since

$$\begin{bmatrix} L & l \end{bmatrix} \begin{bmatrix} R & r \end{bmatrix}^T = LR^T + lr^T \, .$$

Moreover, the proposed framework for solving (BPDN$_\eta$) is fully compatible with this strategy, since the underlying root finding is blind to the factorization representation. Changing k only affects iteration (2.2) through $v(\tau)$ and $v'(\tau)$.

2.6.3 Computational efficiency

One way of assessing the cost of LR-BPDN is to compare the computational cost per iteration of the factorization constrained LASSO subproblems (2.14) with that of the nuclear norm constrained LASSO subproblems (2.13). We first consider the cost for computing the gradient direction. A gradient direction for the factor L in the factorized algorithm is given by

$$g_L = \mathcal{A}^* \left(\mathcal{A}(LR^T) - b \right) R,$$

with g_R taking a similar form. Compare this to a gradient direction for X

$$g_X = \mathcal{A}^* \left(\mathcal{A}(X) - b \right).$$

First, we consider the cost incurred in working with the residual and decision variables. While both methods must compute the action of $\mathcal{A}^*$ on a vector, the factorized formulation must modify factors L, R (at a cost of $O(k(n+m))$ and re-form the matrix $X = LR^T$ (at a cost of at most $O(knm)$, for every iteration and line search evaluation. Since $\mathcal{A}$ is a sampling matrix for the applications of interest, it is sufficient to form only the entries of X that are sampled by $\mathcal{A}$, thus reducing the cost to $O(kp)$, where p is the dimension of the measurement vector b. The sparser the sampling

operator $\mathcal{A}$, the greater the savings. Standard approaches update an explicit decision variable X, at a cost of $O(nm)$, for every iteration and line search evaluation. If the fraction sampled is smaller than the chosen rank k, the factorized approach is actually cheaper than the standard method. It is also important to note that standard approaches have a memory footprint of $O(mn)$, simply to store the decision variable. In contrast, the memory used by factorized approaches are dominated by the size of the observed data.

We now consider the difference in cost involved in the projection. The main benefit for the factorized formulation is that projection is done using the Frobenius norm formulation (2.14), and so the cost is $O(k(n+m))$ for every projection. In contrast, state of the art implementations that compute full or partial SVDs in order to accomplish the projection (see e.g. Jain et al. [2010], Becker et al. [2011]) are dominated by the cost of this calculation, which is (in the case of partial k-SVD) $O(nmk)$, assuming without loss of generality that $k \leq \min(m, n)$.

While the complexity of both standard and factorized iterations is dominated by the term $O(mnk)$, in practice forming $X = LR^T$ from two factors with k columns each is still cheaper than computing a k-partial SVD of X. This essentially explains why factorized methods are faster. While it is possible to obtain further speed up for standard methods using inexact SVD computations, the best reported improvement is a factor of two or three Lin and Wei [2010]. To test our approach against a similar approach that uses Lanczos to compute partial SVDs, we modified the projection used by the SPGL1 code to use this acceleration. We compare against this accelerated code, as well as against TFOCS Becker et al. [2011] in section 2.9 (see Table 2.7).

Finally, both standard and factorized versions of the algorithm require computing the maximum singular value in order to compute $v'(\tau)$. The analysis in section 2.5 shows that if the chosen rank of the factors L and R is larger than or equal to the rank of the global minimizers of the nuclear norm LASSO subproblems, then any local minimizer of the factorized LASSO subproblem corresponds to a global minimizer for the convex nuclear norm LASSO formulation. Consequently, both formulations will have similar of Pareto curve updates, since the derivates are necessarily equal at any global minimum whenever ρ is strictly convex[1].

2.7 Robust formulations

Robust statistics Huber [1981], Maronna et al. [2006] play a crucial role in many real-world applications, allowing good solutions to be obtained in spite of data contamination. In the linear and nonlinear regression setting, the least-squares problem

$$\min_X \|F(X) - b\|_2^2$$

[1]It is shown in Aravkin et al. [2013a] that for any differentiable convex ρ, the dual problem for the residual $r = b - \mathcal{A}x$ has a unique solution. Therefore, *any* global minimum for (LASSO$_\tau$) guarantees a unique residual when ρ is strictly convex, and the claim follows, since the derivative only depends on the residual.

corresponds to the maximum likelihood estimate of X for the statistical model

$$b = F(X) + \epsilon \,, \tag{2.15}$$

where ϵ is a vector of i.i.d. Gaussian variables. Robust statistical approaches relax the Gaussian assumption, allowing other (heavier tailed) distributions to be used. Maximum likelihood estimates of X under these assumptions are more robust to data contamination. Heavy-tailed distributions, in particular the Student's t, yield formulations that are more robust to outliers than convex formulations Lange et al. [1989], Aravkin et al. [2012]. This corresponds to the notion of a re-descending *influence function* Maronna et al. [2006], which is simply the derivative of the negative log likelihood. The relationship between densities, penalties, and influence functions is shown in Figure 2.1. Assuming that ϵ has the Student's t density leads to the maximum likelihood estimation problem

$$\min_X \rho(F(x) - b) := \sum_i \log(\nu + (F(X)_i - b_i)^2), \tag{2.16}$$

where ν is the Student's t degree of freedom.

A general version of (BPDN$_\eta$) was proposed in Aravkin et al. [2013a], allowing different penalty funtionals ρ. The root-finding procedure of Berg and Friedlander [2008] was extended in Aravkin et al. [2013a] to this more general context, and used for root finding for both convex and noncovex ρ (e.g. as in (2.16)).

The (BPDN$_\eta$) formulation for any ρ do not arise directly from a maximum likelihood estimator of (2.15), because they appear in the constraint. However, we can think about penalties ρ as *agents* who, given an error budget η, distribute it between elements of the residual. The strategy that each agent ρ will use to accomplish this task can be deduced from tail features evident in Figure 2.1. Specifically, the cost of a large residual is prohibitively expensive for the least squares penalty, since its *cost* is commensurate with that of a very large number of small residuals. For example, $(10\alpha)^2 = 100\alpha^2$; so a residual of size 10α is worth as much as 100 residuals of size α to the least squares penalty. Therefore, a least squares penalty will never assign a single residual a relatively large value, since this would quickly use up the entire error budget. In contrast, $|10\alpha| = 10|\alpha|$, so a residual of size 10α is worth only 10 residuals of size α when the 1-norm penalty is used. This penalty is likely to grant a few relatively large errors to certain residuals, if this resulted in a better fit. For the penalty in (2.16), it is easy to see that the cost of a residual of size 10α can be worth fewer than 10 residuals of size α, and specific computations depend on ν and actual size of α. A nonconvex penalty ρ, e.g. the one in (2.16), allows large residuals, as long as the majority of the remaining residuals are fit well.

From the discussion in the previous paragraph, it is clear that robust penalties are useful as *constraints* in (BPDN$_\eta$), and can cleverly distribute the allotted error budget η, using it for outliers while fitting good data. The LR-BPDN framework proposed in this paper captures the robust

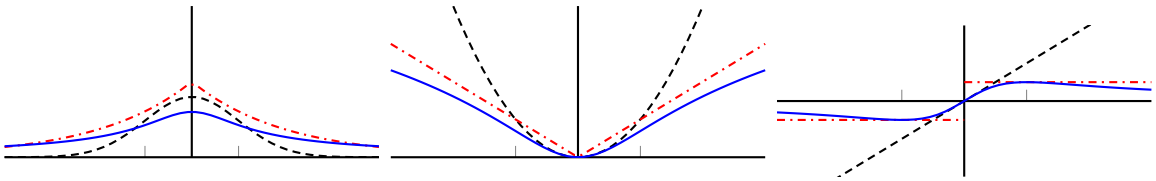


Figure 2.1: Gaussian (black dashed line), Laplace (red dashdotted line), and Student's t (blue solid line); Densities (left plot), Negative Log Likelihoods (center plot), and Influence Functions (right plot). Student's t-density has heavy tails, a non-convex log-likelihood, and re-descending influence function.

extension, allowing robust data interpolation in situations when some of available data is heavily contaminated. To develop this extension, we follow Aravkin et al. [2013a] to define the generalized value function

$$v_\rho(\tau) = \min_X \rho(\mathcal{A}(X) - b) \quad \text{s.t.} \ \|X\|_* \leq \tau \ , \tag{2.17}$$

and find $v_\rho(\tau) = \eta$ using the iteration (2.2). As discussed in section 2.3, for any *convex* smooth penalty ρ,

$$v'_\rho(\tau) = -\|\mathcal{A}^* \nabla \rho(\bar{r})\|_2 \ , \tag{2.18}$$

where $\|\cdot\|_2$ is the spectral norm, and $\bar{r} = \mathcal{A}(\bar{X}) - b$ for optimal solution $\bar{X}$ that achieves $v_\rho(\tau)$. For smooth non-convex ρ, e.g. (2.16), we still use (2.18) in iteration (2.2).

As with standard least squares, we use the factorization formulation to avoid SVDs. Note that Theorem 1 and Corollary 1 hold for any choice of penalty ρ. When evaluating the value function $v_\rho(\tau)$, we actually solve

$$\min_{L,R} \rho(\mathcal{A}(LR^T) - b) \quad \text{s.t.} \ \frac{1}{2} \left\| \begin{bmatrix} L \\ R \end{bmatrix} \right\|_F^2 \leq \tau \ . \tag{2.19}$$

For any smooth penalty ρ, including (2.16), a stationary point for this problem can be found using the projected gradient method.

2.8 Reweighting

Every rank-k solution $\bar{X}$ of (BPDN$_\eta$) lives in a lower dimensional subspace of $\mathbb{R}^{n \times m}$ spanned by the $n \times k$ row and $m \times k$ column basis vectors corresponding to the nonzero singular values of $\bar{X}$. In certain situations, it is possible to estimate the row and column subspaces of the matrix X either from prior subspace information or by solving an initial (BPDN$_\eta$) problem.

In the vector case, it was shown that prior information on the support (nonzero entries) can be incorporated in the ℓ_1-recovery algorithm by solving the weighted-ℓ_1 minimization problem. In this case, the weights are applied such that solutions with large nonzero entries on the support

estimate have a lower cost (weighted ℓ_1 norm) than solutions with large nonzeros outside of the support estimate Friedlander et al. [2011].

In the matrix case, the support estimate is replaced by estimates of the row and column subspace bases $U_0 \in \mathbb{R}^{n\times k}$ and $V_0 \in \mathbb{R}^{m\times k}$ of the largest k singular values of X. Let the matrices $\widetilde{U} \in \mathbb{R}^{n\times k}$ and $\widetilde{V} \in \mathbb{R}^{m\times k}$ be estimates of U_0 and V_0, respectively.

The weighted nuclear norm minimization problem can be formulated as follows:

$$\min_X ||QXW||_* \quad \text{s.t. } \rho(\mathcal{A}(X) - b) \leq \eta, \tag{wBPDN$_\eta$}$$

where $Q = \omega\widetilde{U}\widetilde{U}^T + \widetilde{U}^\perp\widetilde{U}^{\perp T}$, $W = \omega\widetilde{V}\widetilde{V}^T + \widetilde{V}^\perp\widetilde{V}^{\perp T}$, and ω is some constant between zero and one. Here, we use the notation $\widetilde{U}^\perp \in \mathbb{R}^{n\times n-k}$ to refer to the orthogonal complement of $\widetilde{U}$ in $\mathbb{R}^{n\times n}$, and similarly for $\widetilde{V}^\perp$ in $\mathbb{R}^{m\times m}$. The matrices Q and W are weighted projection matrices of the subspaces spanned by $\widetilde{U}$ and $\widetilde{V}$ and their orthogonal complements. Therefore, minimizing $||QXW||_*$ penalizes solutions that live in the orthogonal complement spaces more when $\omega < 1$.

Note that matrices Q and W are invertible, and hence the reweighed LASSO problem still fits into the class of problems characterized by Theorem 1. Specifically, we can write any objective $f(X)$ subject to a reweighted nuclear norm constraint as

$$\begin{aligned} \min \quad & f(Q^{-1}\mathcal{R}(Z)W^{-1}) \\ \text{s.t.} \quad & \text{Tr}(Z) \leq \tau \ , \end{aligned} \tag{2.20}$$

where as in Corollary 1, $\mathcal{R}(Z)$ extracts the upper $n \times m$ block of Z (see (2.10)). A factorization similar to (2.14) can then be formulated for the (wBPDN$_\eta$) problem in order to optimize over the lower dimensional factors $L \in \mathbb{R}^{n\times k}$ and $R \in \mathbb{R}^{m\times k}$.

In particular, we can solve a sequence of (LASSO$_\tau$) problems

$$\min_{L,R} \|\mathcal{A}(LR^T) - b\|_2^2 \quad \text{s.t. } \frac{1}{2}\left\|\begin{bmatrix} QL \\ WR \end{bmatrix}\right\|_F^2 \leq \tau \ , \tag{2.21}$$

where Q and W are as defined above. Problem (2.21) can also be solved using the spectral projected gradient algorithm. However, unlike to the non-weighted formulation, the projection in this case is nontrivial. Fortunately, the structure of the problem allows us to find an efficient formulation for the projection operator.

2.8.1 Projection onto the weighted frobenius norm ball

The projection of a point (L, R) onto the weighted Frobenius norm ball $\frac{1}{2}\left(\|QL\|_F^2 + \|WR\|_F^2\right) \leq \tau$ is achieved by finding the point $(\widetilde{L}, \widetilde{R})$ that solves

$$\min_{\hat{L},\hat{R}} \frac{1}{2}\left\|\begin{bmatrix}\hat{L}-L\\ \hat{R}-R\end{bmatrix}\right\|_F^2 \quad \text{s.t.} \quad \frac{1}{2}\left\|\begin{bmatrix}Q\hat{L}\\ W\hat{R}\end{bmatrix}\right\|_F^2 \leq \tau.$$

The solution to the above problem is given by

$$\begin{aligned}\widetilde{L} &= \left((\mu\omega^2+1)^{-1}\widetilde{U}\widetilde{U}^T + (\mu+1)^{-1}\widetilde{U}^{\perp}\widetilde{U}^{\perp T}\right)L\\ \widetilde{R} &= \left((\mu\omega^2+1)^{-1}\widetilde{V}\widetilde{V}^T + (\mu+1)^{-1}\widetilde{V}^{\perp}\widetilde{V}^{\perp T}\right)R,\end{aligned}$$

where μ is the Lagrange multiplier that solves $f(\mu) \leq \tau$ with $f(\mu)$ given by

$$\begin{aligned}f(\mu) = \frac{1}{2}\mathrm{Tr}\Big[\Big(&\frac{\omega^2}{(\mu\omega^2+1)^2}\widetilde{U}\widetilde{U}^T + \frac{1}{(\mu+1)^2}\widetilde{U}^{\perp}\widetilde{U}^{\perp T}\Big)LL^T\\ &+\Big(\frac{\omega^2}{(\mu\omega^2+1)^2}\widetilde{V}\widetilde{V}^T + \frac{1}{(\mu+1)^2}\widetilde{V}^{\perp}\widetilde{V}^{\perp T}\Big)RR^T\Big].\end{aligned} \tag{2.22}$$

The optimal μ that solves equation (2.22) can be found using the Newton iteration

$$\mu^{(t)} = \mu^{(t-1)} - \frac{f(\mu^{(t-1)}) - \tau}{\nabla f(\mu^{(t-1)})},$$

where $\nabla f(\mu)$ is given by

$$\begin{aligned}\mathrm{Tr}\Big[\Big(&\frac{-2\omega^4}{(\mu\omega^2+1)^2}\widetilde{U}\widetilde{U}^T + \frac{-2}{(\mu+1)^3}\widetilde{U}^{\perp}\widetilde{U}^{\perp T}\Big)LL^T\\ +\Big(&\frac{-2\omega^4}{(\mu\omega^2+1)^3}\widetilde{V}\widetilde{V}^T + \frac{-2}{(\mu+1)^3}\widetilde{V}^{\perp}\widetilde{V}^{\perp T}\Big)RR^T\Big].\end{aligned}$$

2.8.2 Traversing the pareto curve

The design of an effective optimization method that solves (wBPDN$_\eta$) requires 1) evaluating problem (2.21), and 2) computing the dual of the weighted nuclear norm $\|QXW\|_*$.

We first define a gauge function $\kappa(x)$ as a convex, nonnegative, positively homogeneous function such that $\kappa(0) = 0$. This class of functions includes norms and therefore includes the formulations described in (wBPDN$_\eta$) and (2.21). Recall from section 2.3 that taking a Newton step along the Pareto curve of (wBPDN$_\eta$) requires the computation of the derivative of $v(\tau)$ as in (2.3). Therefore,

we also define the polar (or dual) of κ as

$$\kappa^o(x) = \sup_w \{w^T x \mid \kappa(w) \leq 1\}. \tag{2.23}$$

Note that if κ is a norm, the polar reduces to the dual norm.

To compute the dual of the weighted nuclear norm, we follow Theorem 5.1 of Berg and Friedlander [2011] which defines the polar (or dual) representation of a weighted gauge function $\kappa(\Phi x)$ as $\kappa^o(\Phi^{-1}x)$, where Φ is an invertible linear operator. The weighted nuclear norm $\|QXW\|_*$ is in fact a gauge function with invertible linear weighting matrices Q and W. Therefore, the dual norm is given by

$$(\|Q(\cdot)W\|_*)_d(Z) := \|Q^{-1}ZW^{-1}\|_\infty.$$

2.9 Numerical experiments

We test the performance of LR-BPDN on two example applications. In section 2.9.1, we consider the Netflix Prize problem, which is often solved using rank minimization Funk [2006], Gross [2011], Recht and Ré [2011]. Using MovieLens 1M, 10M, and Netflix 100M datasets, we compare and discuss advantages of different formulations, compare our solver against state of the art convex (BPDN$_\eta$) solver SPGℓ_1, and report timing results. We show that the proposed algorithm is orders of magnitude faster than the best convex (BPDN$_\eta$) solver.

In section 2.9.2, we apply the proposed methods and extensions to seismic trace interpolation, a key application in exploration geophysics Sacchi et al. [1998], where rank regularization approaches have recently been used successfully Oropeza and Sacchi [2011]. In section 2.9.2, we include an additional comparison of LR-BPDN with classic SPGℓ_1 as well as with TFOCS Becker et al. [2011] for small matrix completion and seismic data interpolation problems. Then, using real data collected from the Gulf of Suez, we show results for robust completion in section 2.9.2, and present results for the weighted extension in section 2.9.2.

2.9.1 Collaborative filtering

We tested the performance of our algorithm on completing missing entries in the MovieLens (1M), (10M), and Netflix (100M) datasets, which contain anonymous ratings of movies made by Netflix users. The ratings are on an integer scale from 1 to 5. The ratings matrix is not complete, and the goal is to infer the values in the unseen test set. In order to test our algorithm, we further subsampled the available ratings by randomly removing 50% of the known entries. We then solved the (BPDN$_\eta$) formulation to complete the matrix, and compared the predicted (P) and actual (A) removed entries in order to assess algorithm performance. We report the signal-to-noise ratio (SNR) and root

means square error (RMSE):

$$\text{SNR} = 20 \log \left(\frac{\|A\|_F}{\|P - A\|_F} \right), \quad \text{RMSE} = \|P - A\|_F / \|A\|_0$$

for different values of η in the (BPDN$_\eta$) formulation.

Since our algorithm requires pre-defining the rank of the factors L and R, we perform the recovery with ranks $k \in \{10, 20, 30, 50\}$. Table 2.2 shows the reconstruction SNR for each of the ranks k and for a relative error $\eta \in \{0.5, 0.3, 0.2\}$ (the data mismatch is reduced to a fraction η of the initial error). The last row of table 2.2 shows the recovery for an unconstrained low-rank formulation, using the work and software of Vandereycken [2013]. This serves as an interesting baseline, since the rank k of the Riemannian manifold in the unconstrained formulation functions as a regularizer. It is clear that for small k, we get good results without additional functional regularization; however, as k increases, the quality of the rank k solution decays without further constraints. In contrast, we get better results as the rank increases, because we consider a larger model space, but solve the BPDN formulation each time. This observation demonstrates the importance of the nuclear norm regularization, especially when the underlying rank of the problem is unknown.

Table 2.3 shows the timing (in seconds) used by all methods to obtain solutions. There are several conclusions that can be readily drawn. First, for error-level constrained problems, a tighter error bound requires a higher computational investment by our algorithm, which is consistent with the original behavior of SPGℓ_1 Berg and Friedlander [2008]. Second, the unconstrained problem is easier to solve (using the Riemmanian manifolds approach of Vandereycken [2013]) than a constrained problem of the same rank; however, it is interesting to note that as the rank of the representation increases, the unconstrained Riemmanian approach becomes more expensive than the constrained problem for the levels η considered, most likely due to second-order methods used by the particular implementation of Vandereycken [2013].

Table 2.4 shows the value of $\|X\|_*$ of the reconstructed signal corresponding to the settings in Table 2.2. While the interpretation of the η values are straightforward (they are fractions of the initial data error), it is much more difficult to predict ahead of time which value of τ one may want to use when solving (LASSO$_\tau$). This illustrates the *modeling* advantage of the (BPDN$_\eta$) formulation: it requires only the simple parameter η, which is an estimate of the (relative) noise floor. Once η is provided, the algorithm (not the user) will instantiate (LASSO$_\tau$) formulations, and find the right value τ that satisfies $v(\tau) = \eta$. When no estimate of η is available, our algorithm can still be applied to the problem, with $\eta = 0$ and a fixed computational budget (see Table 2.5.)

Table 2.5 shows a comparison between classic SPGℓ_1, accelerated with a Lanczos-based truncated SVD projector, against the new solver, on the MovieLens (10M) dataset, for a fixed budget of 100 iterations. Where the classic solver takes over six hours, the proposed method finishes in less than a minute. For a problem of this size, explicit manipulation of X as a full matrix of size 10K

Table 2.2: Summary of the recovery results on the MovieLens (1M) data set for factor rank k and relative error level η for (BPDN$_\eta$). SNR in dB (higher is better) listed in the left table, and RMSE (lower is better) in the right table. The last row in each table gives recovery results for the non-regularized data fitting factorized formulation solved with Riemannian optimization (ROPT). Quality *degrades* with k due to overfitting for the non-regularized formulation, and improves with k when regularization is used.

η \ k	10	20	30	50
0.5	5.93	5.93	5.93	5.93
0.3	10.27	10.27	10.26	10.27
0.2	12.50	12.54	12.56	12.56
ROPT	11.16	8.38	6.01	2.6

η \ k	10	20	30	50
0.5	1.89	1.89	1.89	1.89
0.3	1.14	1.14	1.15	1.14
0.2	0.88	0.88	0.88	0.88
ROPT	1.03	1.42	1.87	2.77

Table 2.3: Summary of the computational timing (in seconds) on the MovieLens (1M) data set for factor rank k and relative error level η for (BPDN$_\eta$). The last row gives computational timing for the non-regularized data fitting factorized formulation solved with Riemannian optimization.

η \ k	10	20	30	50
0.5	5.0	5.7	6.8	8.2
0.3	11.1	18.8	12.3	21.0
0.2	58.7	84.2	95.0	114.2
ROPT	14.9	43.5	98.4	327.3

by 20K is computationally prohibitive. Table 2.6 gives timing and reconstruction quality results for the Netflix (100M) dataset, where the full matrix is 18K by 500K when fully constructed.

2.9.2 Seismic missing-trace interpolation

In exploration seismology, large-scale data sets (approaching the order of petabytes for the latest land and wide-azimuth marine acquisitions) must be acquired and processed in order to determine the structure of the subsurface. In many situations, only a subset of the complete data is acquired due to physical and/or budgetary constraints. Recent insights from the field of compressed sensing allow for deliberate subsampling of seismic wavefields in order to improve reconstruction quality and reduce acquisition costs Herrmann and Hennenfent [2008a]. The acquired subset of the complete data is often chosen by randomly subsampling a dense regular periodic source or receiver grid. Interpolation algorithms are then used to reconstruct the dense regular grid in order to perform

Table 2.4: Nuclear-norms of the solutions $X = LR^T$ for results in Table 2.2, corresponding to τ values in (LASSO$_\tau$). These values are found automatically via root finding, but are difficult to guess ahead of time.

η \ k	5	10	30	50
0.5	5.19e3	5.2e3	5.2e3	5.2e3
0.3	9.75e3	9.73e3	9.76e3	9.74e3
0.2	1.96e4	1.96e4	1.93e4	1.93e4

Table 2.5: Classic SPGL1 (using Lanczos based truncated SVD) versus LR factorization on the MovieLens (10M) data set (10000×20000 matrix) shows results for a fixed iteration budget (100 iterations) for 50% subsampling of MovieLens data. SNR, RMSE and computational time are shown for $k = 5, 10, 20$.

		MovieLens (10M)		
	k	5	10	20
SPGℓ_1	SNR (dB)	11.32	11.37	11.37
	RMSE	1.02	1.01	1.01
	time (sec)	**22680**	**93744**	**121392**
LR	SNR (dB)	11.87	11.77	11.72
	RMSE	0.95	0.94	0.94
	time (sec)	**54.3**	**48.2**	**47.5**

additional processing on the data such as removal of artifacts, improvement of spatial resolution, and key analysis, such as imaging.

In this section, we apply the new rank-minimization approach, along with weighted and robust extensions, to the trace-interpolation problem for two different seismic acquisition examples. We first describe the structure of the datasets, and then present the transform we use to cast the interpolation as a rank-minimization problem.

The first example is a real data example from the Gulf of Suez. Seismic data are organized into *seismic lines*, where N_r receivers and N_s sources are collocated in a straight line. Sources are deployed sequentially, and receivers collect each shot record[2] for a period of N_t time samples. The Gulf of Suez data contains $N_s = 354$ sources, $N_r = 354$ receivers, and $N_t = 1024$ with a sampling interval of 0.004s, leading to a shot duration of 4s and a maximum temporal frequency of 125 Hz. Most of the energy of the seismic line is preserved when we restrict the spectrum to the 12-60Hz frequency band. Figs. 2.2(a) and (b) illustrate the 12Hz and 60Hz frequency slices in the source-

[2]Data collection performed for several sources taken with increasing or decreasing distance between sources and receivers.

Table 2.6: LR method on the Netflix (100M) data set (17770×480189 matrix) shows results for 50% subsampling of Netflix data. SNR, computational time and RMSE are shown for factor rank k and relative error level η for (BPDN$_\eta$).

		Netflix (100M)		
	k	2	4	6
η=0.5	SNR (dB)	7.37	7.03	7.0
	RMSE	1.60	1.67	1.68
	time (sec)	**236.5**	**333.0**	**335.0**
η=0.4	SNR (dB)	8.02	7.96	7.93
	RMSE	1.49	1.50	1.50
	time (sec)	**315.2**	**388.6**	**425.0**
η=0.3	SNR (dB)	10.36	10.32	10.35
	RMSE	1.14	1.14	1.14
	time (sec)	**1093.2**	**853.7**	**699.7**

receiver domain, respectively. Columns in these frequency slices represent the monochromatic response of the earth to a fixed source and as a function of the receiver coordinate. In order to simulate missing traces, we apply a subsampling mask that randomly removes 50% of the sources, resulting in the subsampled frequency slices illustrated in Figs. 2.2 (c) and (d).

State of the art trace-interpolation schemes transform the data into sparsifying domains, for example using the Fourier Sacchi et al. [1998] and curvelet Herrmann and Hennenfent [2008a] transforms. The underlying *sparse structure* of the data is then exploited to recover the missing traces. The approach proposed in this paper allows us to instead exploit the low-rank *matrix structure* of seismic data, and to design formulations that can achieve trace-interpolation using matrix-completion strategies.

The main challenge in applying rank-minimization for seismic trace-interpolation is to find a *transform domain* that satisfies the following two properties:

1. Fully sampled seismic lines have low-rank structure (quickly decaying singular values)

2. Subsampled seismic lines have high rank (slowly decaying singular values).

When these two properties hold, rank-penalization formulations allow the recovery of missing traces. To achieve these aims, we use the transformation from the source-receiver (s-r) domain to the midpoint-offset (m-h). The conversion from (s-r) domain to (m-h) domain is a coordinate transformation, with the midpoint is defined by m = $\frac{1}{2}$(s+r) and the half-offset is defined by h = $\frac{1}{2}$(s-r).[3] This transformation is illustrated by transforming the 12Hz and 60Hz source-receiver domain frequency slices in Figs. 2.2(a) and (b) to the midpoint-offset domain frequency slices in Figs. 2.2(e)

[3] In mathematical terms, the transformation from (s-r) domain to (m-h) domain represents a tight frame.

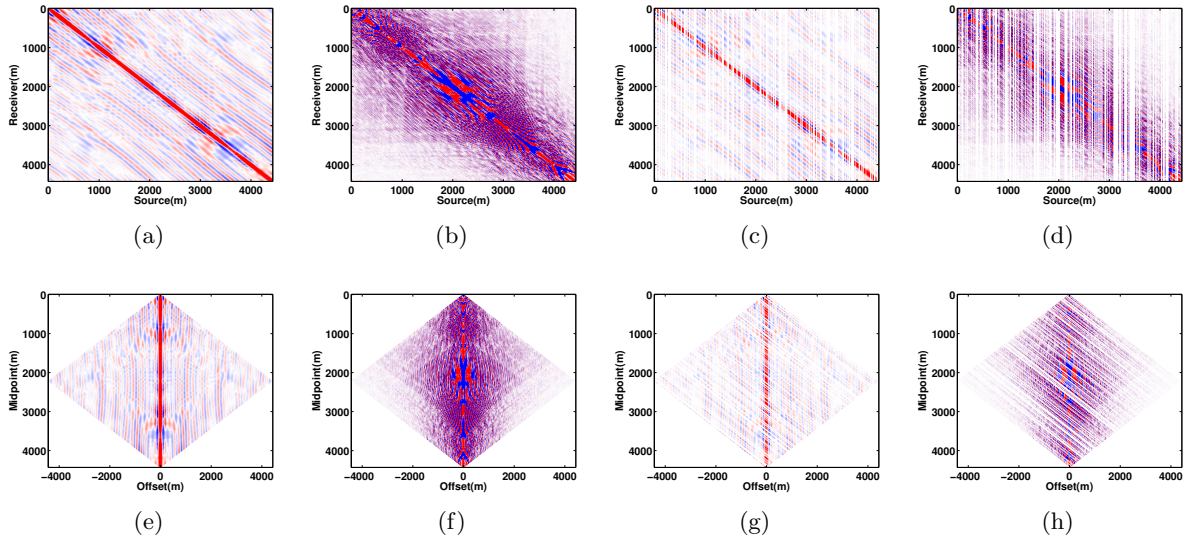


Figure 2.2: Frequency slices of a seismic line from Gulf of Suez with 354 shots, 354 receivers. Full data for (a) low frequency at 12 Hz and (b) high frequency at 60 Hz in s-r domain. 50% Subsampled data for (c) low frequency at 12 Hz and (d) high frequency at 60 Hz in s-r domain. Full data for (e) low frequency at 12 hz and (f) high frequency at 60 Hz in m-h domain. 50% subsampled data for (g) low frequency at 12 Hz and (h) high frequency at 60 Hz in m-h domain.

and (f). The corresponding subsampled frequency slices in the midpoint-offset domain are shown in Figs. 2.2(g) and (h).

To show that the midpoint-offset transformation achieves aims 1 and 2 above, we plot the decay of the singular values of both the 12Hz and 60Hz frequency slices in the source-receiver domain and in the midpoint-offset domain in Figs. 2.3 (a) and (c). Notice that the singular values of both frequency slices decay faster in the midpoint-offset domain, and that the singular value decay is slower for subsampled data in Figs. 2.3 (b) and (d).

Let X denote the data matrix in the midpoint-offset domain and let R be the subsampling operator that maps Figs. 2.2 (e) and (f) to Figs. 2.2(g) and (h). Denote by $\mathcal{S}$ the transformation operator from the source-receiver domain to the midpoint-offset domain. The resulting measurement operator in the midpoint-offset domain is then given by $\mathcal{A} = R\mathcal{S}^H$.

We formulate and solve the matrix completion problem (BPDN$_\eta$) to recover a seismic line from Gulf of Suez in the (m-h) domain. We first performed the interpolation for the frequency slices at 12Hz and 60Hz to get a good approximation of the lower and higher limit of the rank value. Then, we work with all the monochromatic frequency slices and adjust the rank within the limit while going from low- to high-frequency slices. We use 300 iterations of LR for all frequency slices. Figures 2.4(a) and (b) show the recovery and error plot for the low frequency slice at 12

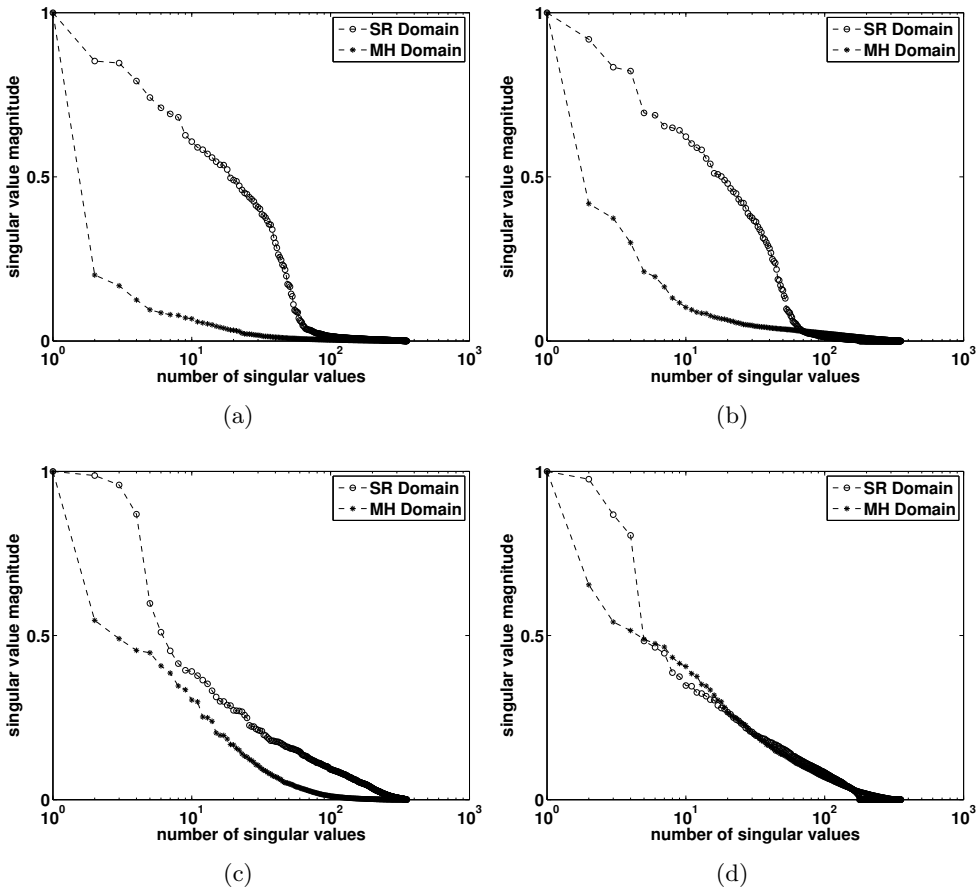


Figure 2.3: Singular value decay of fully sampled (a) low frequency slice at 12 Hz and (c) high frequency slice at 60 Hz in (s-r) and (m-h) domains. Singular value decay of 50% subsampled (b) low frequency slice at 12 Hz and (d) high frequency data at 60 Hz in (s-r) and (m-h) domains. Notice that for both high and low frequencies, decay of singular values is faster in the fully sampled (m-h) domain than in the fully sampled (s-r) domain, and that subsampling does not significantly change the decay of singular value in (s-r) domain, while it destroys fast decay of singular values in (m-h) domain.

Hz, respectively. Figures 2.4(c) and (d) show the recovery and error plot for the high frequency slice at 60 Hz, respectively. Figures 2.5 shows a common-shot gather section after missing-trace interpolation from Gulf of Suez data set. We can clearly see that we are able to recapture most of the missing traces in the data (Figures 2.5c), also evident from residual plot (Figures 2.5d).

In the second acquisition example, we implement the proposed formulation on the 5D synthetic seismic data (2 source dimension, 2 receiver dimension, 1 temporal dimension) provided by BG Group. We extract a frequency slice at 12.3Hz to perform the missing-trace interpolation, where the size of the to-be recovered matrix is 400×400 receivers spaced by 25m and 68×68 sources spaced

by 150m. Due to the low spatial frequency content of the data at 12.3 Hz, we further subsample the data in receiver coordinates by a factor of two to speed up the computation. We apply sub-sampling masks that randomly remove 75% and 50% of the shots. In case of 4D, we have two choices of matricization Silva and Herrmann [2013c], Demanet [2006a], as shown in Figures 2.6(a,b), where we can either place the (Receiver x, Receiver y) dimensions in the rows and (Source x, Source y) dimensions in the columns, or (Receiver y, Source y) dimensions in the rows and (Receiver x, Source x) dimensions in the columns. We observed the faster decay of singular value decay as shown in Figure 2.7(a,b), for each of these strategies. We therefore selected the transform domain to be the permutation of source and receivers coordinates, where matricization of each 4D monochromatic frequency slices is done using (Source x, Receiver x) and (Source y, Receiver y) coordinates. We use rank 200 for the interpolation, and run the solver for a maximum of 1000 iterations. The results after interpolation are shown in Figures 2.8 and 2.9 for 75% and 50% missing data, respectively. We can see that when 75% of data is missing, we start losing coherent energy (Figures 2.8c). With 50% missing data, we capture most of the coherent energy (Figures 2.9c). We also have higher SNR values for recovery in case of 50% compared to 75% missing data.

To illustrate the importance of the nuclear-norm regularization, we solved the interpolation problem using a simple least-squares formulation on the same seismic data set from Gulf of Suez. The least squares problem was solved using the L, R factorization structure, thereby implicitly enforcing a rank on the recovered estimate (i.e, formulation (2.14) was optimized without the τ-constraint). The problem was then solved with the factors L and R having a rank $k \in \{5, 10, 20, 30, 40, 50, 80, 100\}$. The reconstruction SNRs comparing the recovery for the regularized and non-regularized formulations are shown in Fig. 2.10. The Figure shows that the performance of the non-regularized approach decays with rank, due to overfitting. The regularized approach, in contrast, obtains better recovery as the factor rank increases.

Comparison with classical nuclear-norm formulation

To illustrate the advantage of proposed matrix-factorization formulation (which we refer to as LR below) over classical nuclear-norm formulation, we compare the reconstruction error and computation time with the existing techniques. The most natural baseline is the SPGℓ_1 algorithm Berg and Friedlander [2011] applied to the classic nuclear norm (BPDN$_\eta$) formulation, where the decision variable is X, the penalty function ρ is the 2-norm, and the projection is done using the SVD. This example tests the classic (BPDN$_\eta$) formulation against the LR extension proposed in this paper. The second comparison is with the TFOCSBecker et al. [2011], which is a library of first-order methods for a variety of problems with explicit examples written by the authors for the (BPDN$_\eta$) formulation. The TFOCS approach to (BPDN$_\eta$) relies on a proximal function for the nuclear norm, which, similar to projection, requires computing SVDs or partial SVDs.

The comparisons are done using three different data sets. In the first example, we interpolated

missing traces of a monochromatic slice (of size 354×354), extracted from Gulf of Suez data set. We subsampled the frequency slice by randomly removing the 50% of shots and performed the missing-trace interpolation in the midpoint-offset (m-h) domain. We compares the SNR, computation time and iterations for a fixed set of η. The rank of the factors was set to 28. The seismic example in Table 2.7 shows the results. Both the classic SPGℓ_1 algorithm and LR are faster than TFOCS. In the quality of recovery, both SPGℓ_1 and LR have better SNR than TFOCS. In this case LR is faster than SPGℓ_1 by a factor of 15 (see Table 2.7).

In the second example, we generated a rank 10 matrix of size 100×100. We subsampled the matrix by randomly removing 50% of the data entries. The synthetic low-rank example in Table 2.7 shows the comparison of SNR and computational time. The rank of the factors was set to be the true rank of the original data matrix for this experiment. The LR formulation proposed in this paper is faster than classic SPGℓ_1, and both are faster than TFOCS. As the error threshold tightens, TFOCS requires a large number of iterations to converge. For a small problem size, LR and the classic SPGℓ_1 perform comparably. When operating on decision variables with the correct rank LR gave uniformly better SNR results than classic SPGℓ_1 and TFOCS, and the improvement was significant for lower error thresholds.[4] In reality, we do not know the rank value in advance. To make a fair comparison, we used MovieLens (10M) dataset, where we subsampled the available ratings by randomly removing 50% of the known entries. In this example, we fixed the number of iterations to 100 and compared the SNR and computational time (Table 2.5) for multiple ranks, $k = 5, 10, 20$. It is evident that the we get better SNR in case of LR, also the computational speed of LR is significantly faster then the classic SPGℓ_1.

Simultaneous missing-trace interpolation and denoising

To illustrate the utility of robust cost functions, we consider a situation where observed data are heavily contaminated. The goal here is to simultaneously denoise interpolate the data. We work with same seismic line from Gulf of Suez. To obtain the observed data, we apply a sub-sampling mask that randomly removes 50% of the shots, and to simulate contamination, we replace another 10% of the shots with large random errors, whose amplitudes are three times the maximum amplitude present in the data. In reality, we know the sub-sampling mask but we do not know the behaviour and amplitude of noise. In this example, we formulate and solve the robust matrix completion problem (BPDN$_\eta$), where the cost ρ is taken to be the penalty (2.16); see section 2.7 for the explanation and motivation. As in the previous examples, the recovery is done in the (m-h) domain. We implement the formulation in the frequency domain, where we work with monochromatic frequency slices, and adjust the rank and ν parameter while going from low to high frequency slices. Figure 2.11 compares the recovery results with and without using a robust

[4]We tested this hypothesis by re-running the experiment with higher factor rank. For example, selecting factor rank to be 40 gives SNRs of 16.5, 36.7, 42.4, 75.3 for the corresponding η values for the synthetic low-rank experiment in 2.7.

Table 2.7: TFOCS versus classic SPGℓ_1 (using direct SVDs) versus LR factorization. **Synthetic low rank** example shows results for completing a rank 10, 100×100 matrix, with 50% missing entries. SNR, Computational time and iterations are shown for $\eta = 0.1, 0.01, 0.005, 0.0001$. Rank of the factors is taken to be 10. **Seismic** example shows results for matrix completion a low-frequency slice at 10 Hz, extracted from the Gulf of Suez data set, with 50% missing entries. SNR, Computational time and iterations are shown for $\eta = 0.2, 0.1, 0.09, 0.08$. Rank of factors was taken to be 28.

		Synthetic low rank			
η		0.1	0.01	0.005	0.0001
TFOCS	SNR (dB)	17.2	36.3	56.2	76.2
	time (s)	24	179	963	2499
	iteration	1151	8751	46701	121901
SPGℓ_1	SNR (dB)	14.5	36.4	39.2	76.2
	time (s)	4.9	17.0	17.2	61.1
	iteration	12	46	47	152
LR	SNR (dB)	16.5	36.7	42.7	76.2
	time (s)	0.6	0.5	0.58	0.9
	iteration	27	64	73	119

		Seismic			
η		0.2	0.1	0.09	0.08
TFOCS	SNR (dB)	13.05	17.4	17.9	18.5
	time (s)	593	3232	4295	6140
	iteration	1201	3395	3901	4451
SPGℓ_1	SNR (dB)	12.8	17.0	17.4	17.9
	time (s)	30.4	42.8	32.9	58.8
	iteration	37	52	40	73
LR	SNR (dB)	13.1	17.1	17.4	18.0
	time (s)	1.6	2.9	3.2	4.0
	iteration	38	80	87	113

penalty function. The error budget plays a significant role in this example, and we standardized the problems by setting the *relative error* to be 20% of the initial error, so that the formulations are comparable.

We can clearly see that the standard least squares formulation is unable to recover the true solution. The intuitive reason is that the least squares penalty is simply unable to budget large errors to what should be the outlying residuals. The Student's t penalty, in contrast, achieves a good recovery in this extreme situation, with an SNR of 17.9 DB. In this example, we used 300 iterations of SPGℓ_1 for all frequency slices.

Re-weighting

Re-weighting for seismic trace interpolation was recently used in Mansour et al. [2012a] to improve the interpolation of subsampled seismic traces in the context of sparsity promotion in the curvelet domain. The weighted ℓ_1 formulation takes advantage of curvelet support overlap across adjacent frequency slices.

Analogously, in the matrix setting, we use the weighted rank-minimization formulation (wBPDN$_\eta$) to take advantage of correlated row and column subspaces for adjacent frequency slices. We first demonstrate the effectiveness of solving the (wBPDN$_\eta$) problem when we have accurate subspace information. For this purpose, we compute the row and column subspace bases of the fully sampled low frequency (11Hz) seismic slice and pass this information to (wBPDN$_\eta$) using matrices Q and

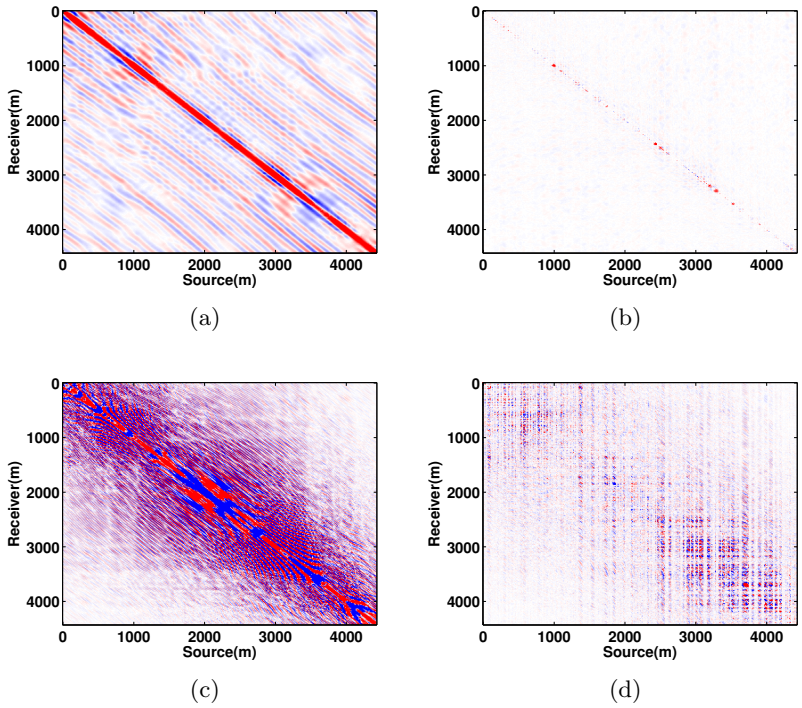


Figure 2.4: Recovery results for 50% subsampled 2D frequency slices using the nuclear norm formulation. (a) Interpolation and (b) residual of low frequency slice at 12 Hz with SNR = 19.1 dB. (c) Interpolation and (d) residual of high frequency slice at 60 Hz with SNR = 15.2 dB.

W. Figures 2.12(a) and (b) show the residual of the frequency slice with and without weighting. The reconstruction using the (wBPDN$_\eta$) problem achieves a 1.5dB improvement in SNR over the non-weighted (BPDN$_\eta$) formulation.

Next, we apply the (wBPDN$_\eta$) formulation in a practical setting where we do not know subspace bases ahead of time, but learn them as we proceed from low to high frequencies. We use the row and column subspace vectors recovered using (BPDN$_\eta$) for 10.75 Hz and 15.75 Hz frequency slices as subspace estimates for the adjacent higher frequency slices at 11 Hz and 16 Hz. Using the (wBPDN$_\eta$) formulation in this way yields SNR improvements of 0.6dB and 1dB, respectively, over (BPDN$_\eta$) alone. Figures 2.13(a) and (b) show the residual for the next higher frequency without using the support and Figures 2.13(c) and (d) shows the residual for next higher frequency with support from previous frequency. Figure 2.14 shows the recovery SNR versus frequency for weighted and non-weighted cases for a range of frequencies from 9 Hz to 17 Hz.

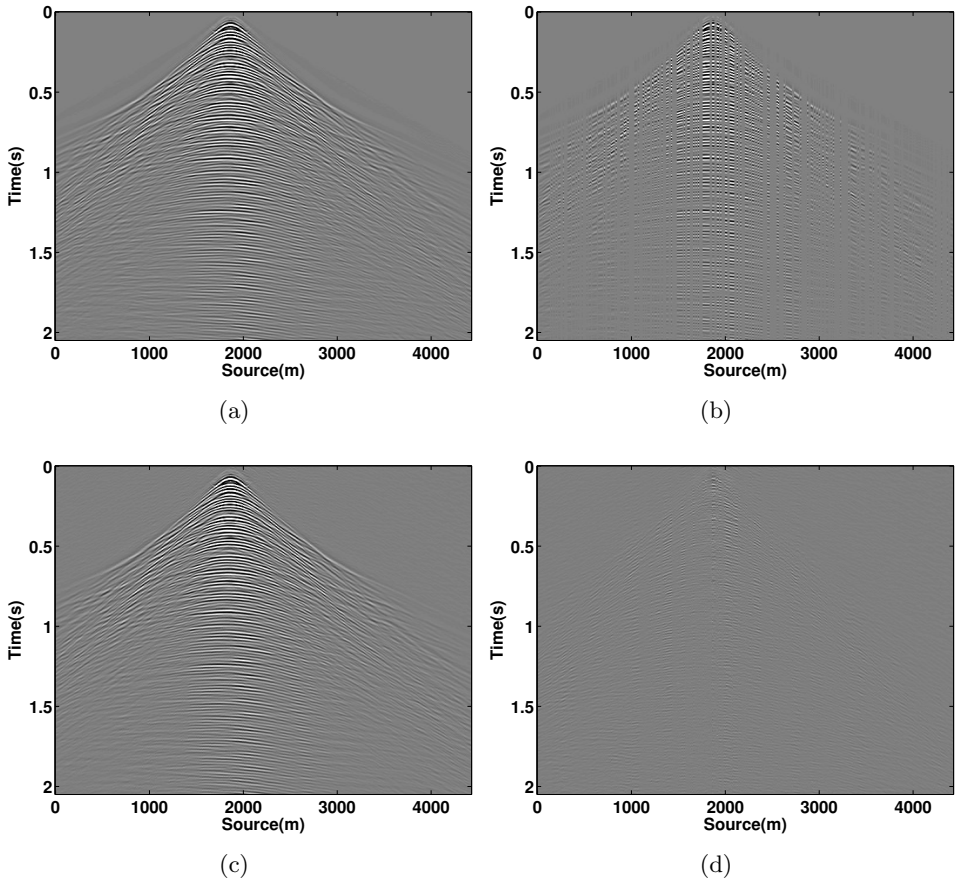


Figure 2.5: Missing trace interpolation of a seismic line from Gulf of Suez. (a) Ground truth. (b) 50% subsampled common shot gather. (c) Recovery result with a SNR of 18.5 dB. (d) Residual.

2.10 Conclusions

We have presented a new method for matrix completion. Our method combines the Pareto curve approach for optimizing (BPDN$_\eta$) formulations with SVD-free matrix factorization methods.

We demonstrated the modeling advantages of the (BPDN$_\eta$) formulation on the Netflix Prize problem, and obtained high-quality reconstruction results for the seismic trace interpolation problem. Comparison with state of the art methods for the (BPDN$_\eta$) formulation showed that the factorized formulation is faster than both TFOCS and classic SPGℓ_1 formulations that rely on the SVD. The presented factorized approach also has a small memory imprint and does not rely on SVDs, which makes this method applicable to truly large-scale problems.

We also proposed two extensions. First, using robust penalties ρ in (BPDN$_\eta$), we showed that simultaneous interpolation and denoising can be achieved in the extreme data contamination case, where 10% of the data was replaced by large outliers. Second, we proposed a weighted

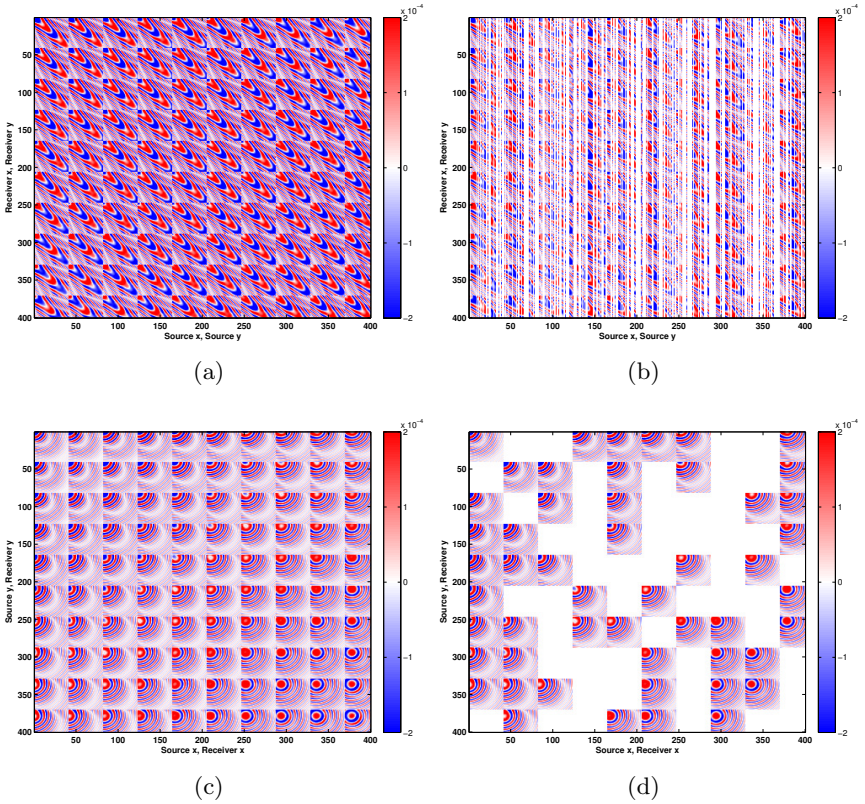


Figure 2.6: Matricization of 4D monochromatic frequency slice. Top: (Source x, Source y) matricization. Bottom: (Source x, Receiver x) matricization. Left: Fully sampled data; Right: Subsampled data.

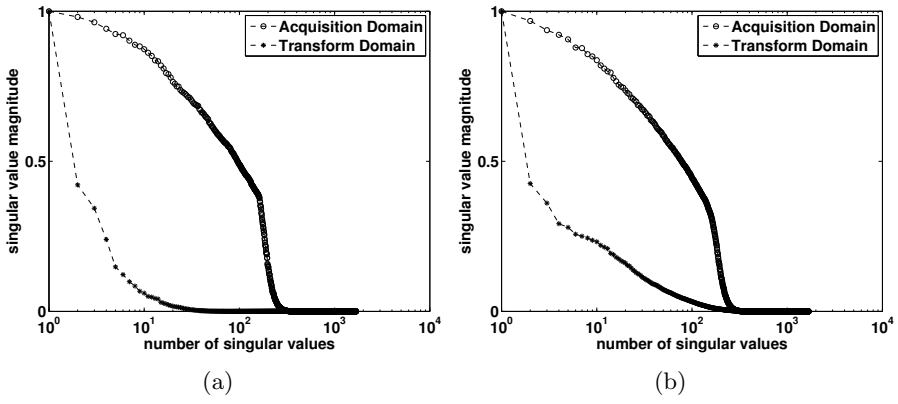


Figure 2.7: Singular value decay in case of different matricization of 4D monochromatic frequency slice. Left: Fully sampled data; Right: Subsampled data.

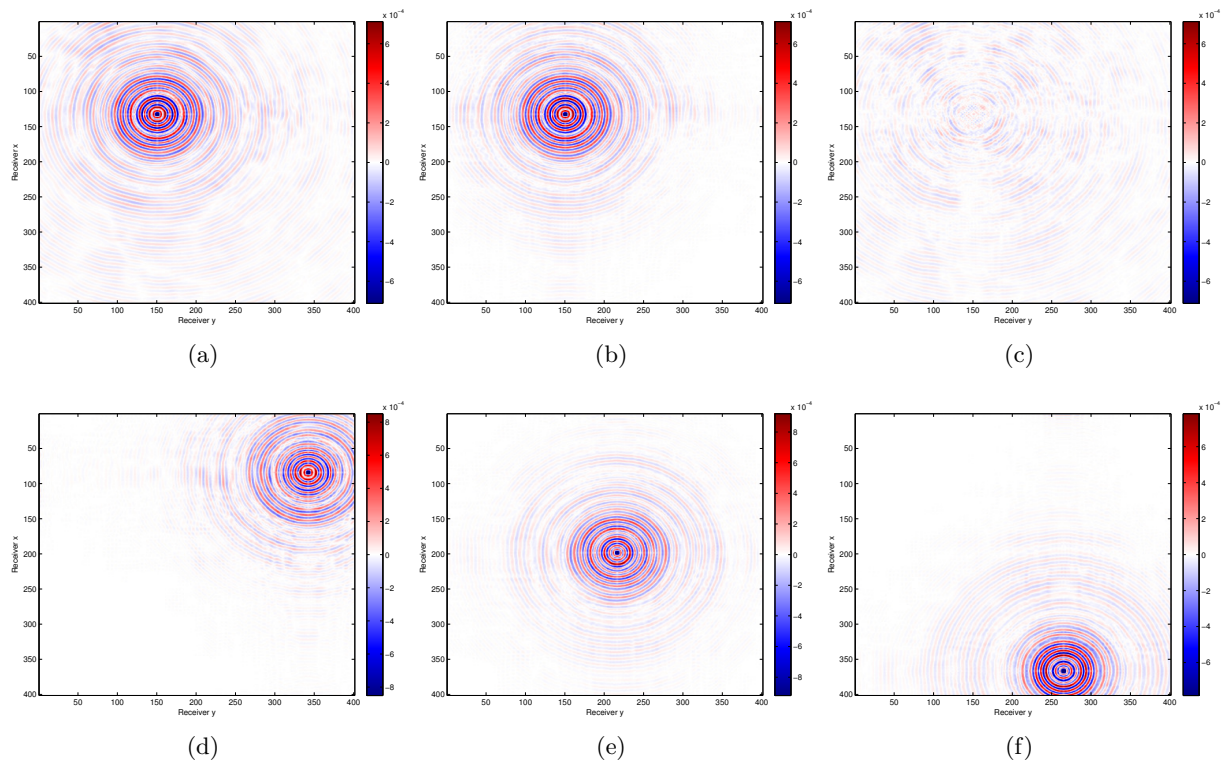


Figure 2.8: Missing-trace interpolation of a frequency slice at 12.3Hz extracted from 5D data set, 75% missing data. (a,b,c) Original, recovery and residual of a common shot gather with a SNR of 11.4 dB at the location where shot is recorded. (d,e,f) Interpolation of common shot gathers at the location where no reference shot is present.

extension (wBPDN$_\eta$), and used it to incorporate subspace information we learned on the fly to improve interpolation in adjacent frequencies.

2.11 Appendix

Proof of Theorem 1 Recall [Burer and Monteiro, 2003, Lemma 2.1]: if $SS^T = KK^T$, then $S = KQ$ for some orthogonal matrix $Q \in \mathbb{R}^{r\times r}$. Next, note that the objective and constraints of (2.8) are given in terms of SS^T, and for any orthogonal $Q \in \mathbb{R}^{r\times r}$, we have $SQQ^TS^T = SS^T$, so $\bar{S}$ is a local minimum of (2.8) if and only if $\bar{S}Q$ is a local minimum for all orthogonal $Q \in \mathbb{R}^{r\times r}$.

If $\bar{Z}$ is a local minimum of (2.7), then any factor $\bar{S}$ with $\bar{Z} = \bar{S}\bar{S}^T$ is a local minimum of (2.8). Otherwise, we can find a better solution $\tilde{S}$ in the neighborhood of $\bar{S}$, and then $\tilde{Z} := \tilde{S}\tilde{S}^T$ will be a feasible solution for (2.7) in the neighborhood of $\bar{Z}$ (by continuity of the map $S \to SS^T$).

We prove the other direction by contrapositive. If $\bar{Z}$ is not a local minimum for (2.7), then you can find a sequence of feasible solutions Z_k with $f(Z^k) < f(\bar{Z})$ and $Z_k \to \bar{Z}$. For each k,

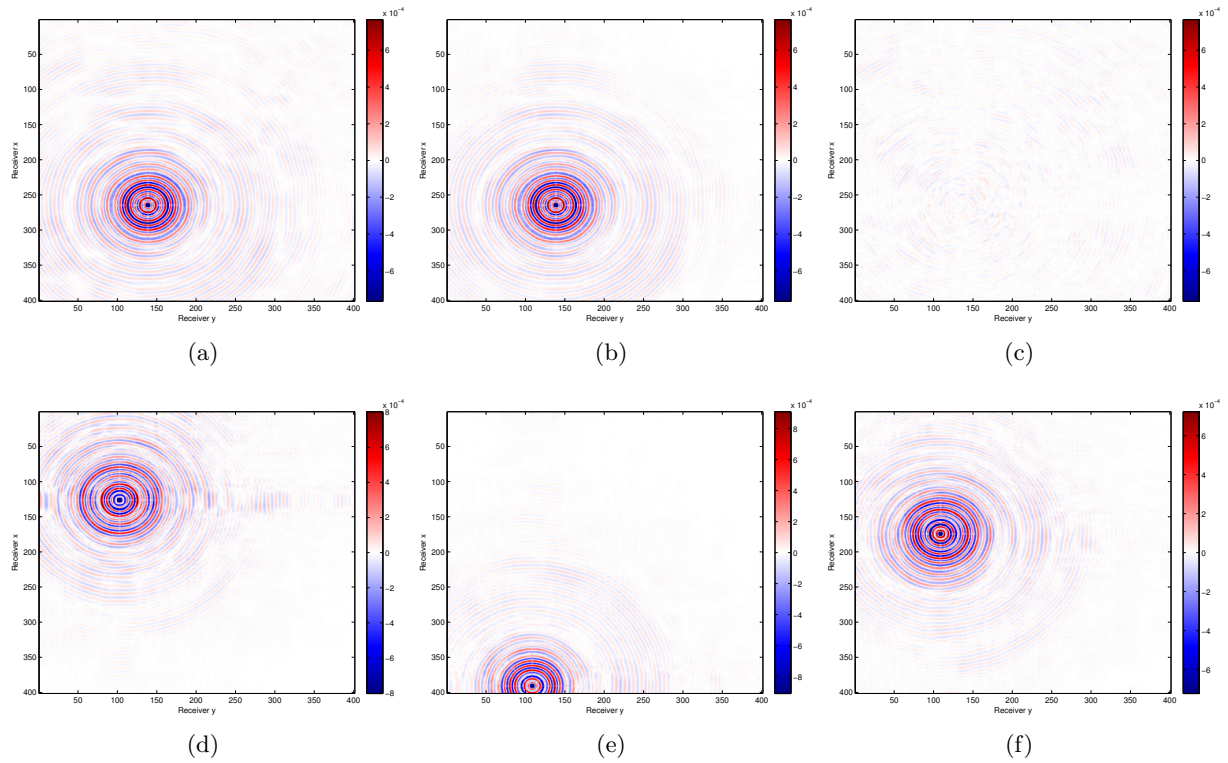
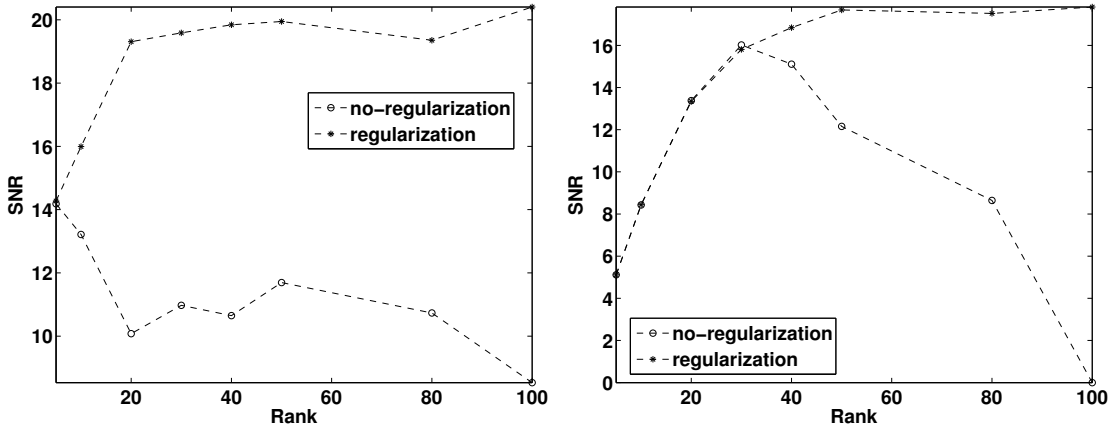


Figure 2.9: Missing-trace interpolation of a frequency slice at 12.3Hz extracted from 5D data set, 50% missing data. (a,b,c) Original, recovery and residual of a common shot gather with a SNR of 16.6 dB at the location where shot is recorded. (d,e,f) Interpolation of common shot gathers at the location where no reference shot is present.

write $Z_k = S_k S_k^T$. Since Z_k are all feasible for (2.7), so S_k are feasible for (2.8). By assumption $\{Z_k\}$ is bounded, and so is S_k; we can therefore find a subsequence of $S_j \to \tilde{S}$ with $\tilde{S}\tilde{S}^T = \bar{Z}$, and $f(S_j S_j^T) < f(\tilde{S}\tilde{S}^T)$. In particular, we have $\bar{Z} = \bar{S}\bar{S}^T = \tilde{S}\tilde{S}^T$, and $\tilde{S}$ is not a local minimum for (2.8), and therefore (by previous results) $\bar{S}$ cannot be either.



(a) (b)

Figure 2.10: Comparison of regularized and non-regularized formulations. SNR of (a) low frequency slice at 12 Hz and (b) high frequency slice at 60 Hz over a range of factor ranks. Without regularization, recovery quality decays with factor rank due to over-fiting; the regularized formulation improves with higher factor rank.

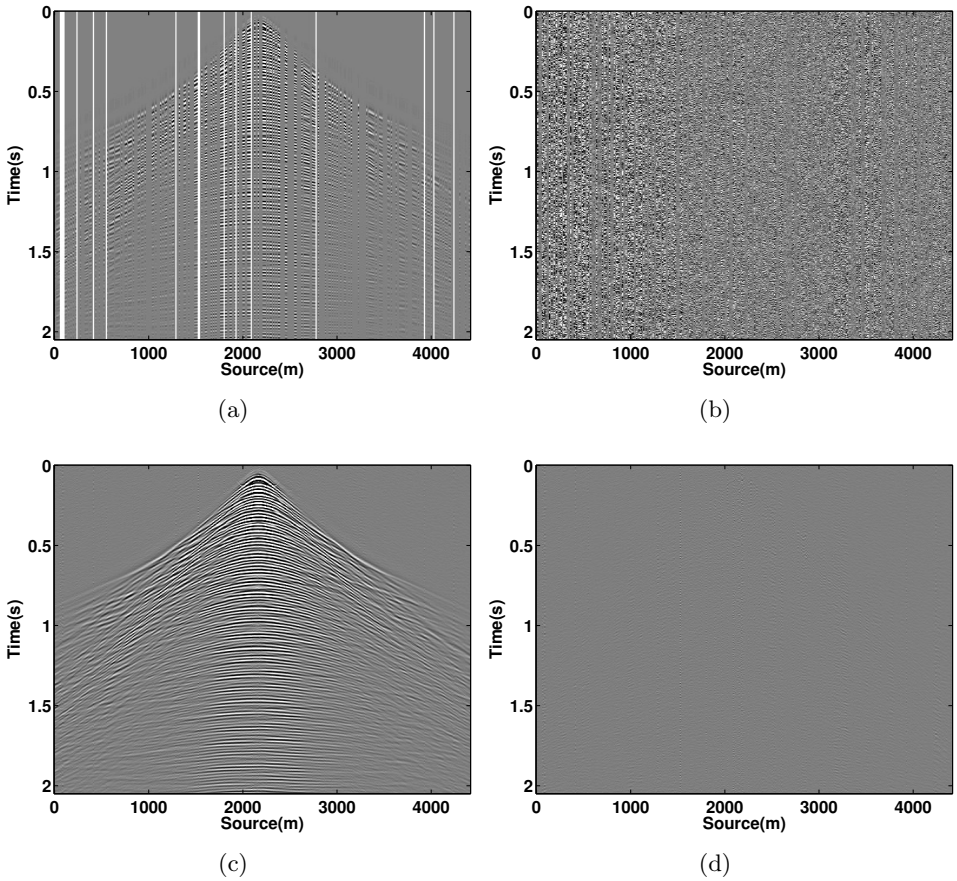


Figure 2.11: Comparison of interpolation and denoising results for the Student's t and least-squares misfit function. (a) 50% subsampled common receiver gather with another 10 % of the shots replaced by large errors. (b) Recovery result using the least-squares misfit function. (c,d) Recovery and residual results using the student's t misfit function with a SNR of 17.2 dB.

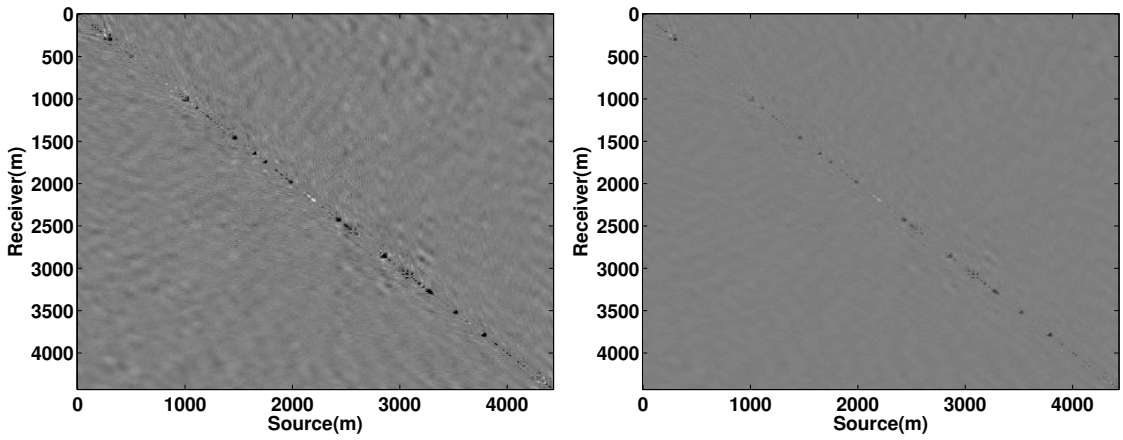


Figure 2.12: Residual error for recovery of 11 Hz slice (a) without weighting and (b) with weighting using true support. SNR in this case is improved by 1.5 dB.

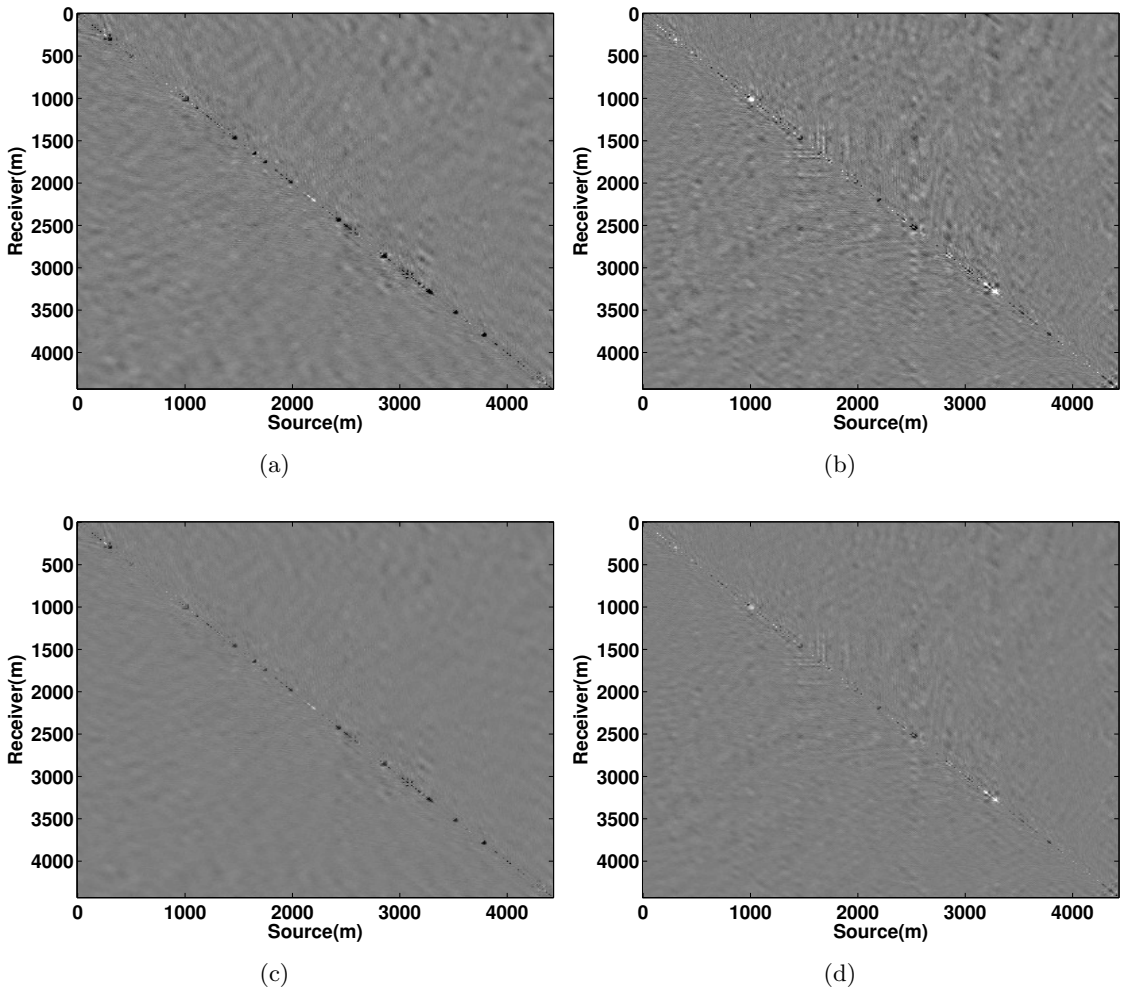


Figure 2.13: Residual of low frequency slice at 11 Hz (a) without weighing (c) with support from 10.75 Hz frequency slice. SNR is improved by 0.6 dB. Residual of low frequency slice at 16 Hz (b) without weighing (d) with support from 15.75 Hz frequency slice. SNR is improved by 1dB. Weighting using learned support is able to improve on the unweighted interpolation results.

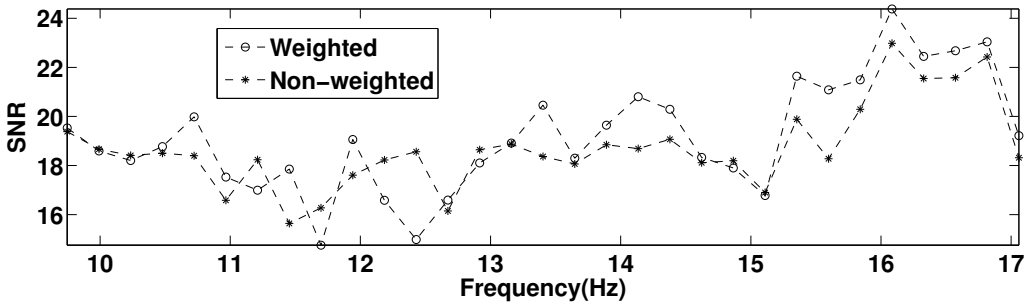


Figure 2.14: Recovery results of practical scenario in case of weighted factorized formulation over a frequency range of 9-17 Hz. The weighted formulation outperforms the non-weighted for higher frequencies. For some frequency slices, the performance of the non-weighted algorithm is better, because the weighted algorithm can be negatively affected when the subspaces are less correlated.

Chapter 3

Efficient matrix completion for seismic data reconstruction

3.1 Summary

Despite recent developments in improved acquisition, seismic data often remains undersampled along source and receiver coordinates, resulting in incomplete data for key applications such as migration and multiple prediction. We interpret the missing-trace interpolation problem in the context of matrix completion and outline three practical principles for using low-rank optimization techniques to recover seismic data. Specifically, we strive for recovery scenarios wherein the original signal is low rank and the subsampling scheme increases the singular values of the matrix. We employ an optimization program that restores this low rank structure to recover the full volume. Omitting one or more of these principles can lead to poor interpolation results, as we show experimentally. In light of this theory, we compensate for the high-rank behaviour of data in the source-receiver domain by employing the midpoint-offset transformation for 2D data and a source-receiver permutation for 3D data to reduce the overall singular values. Simultaneously, in order to work with computationally feasible algorithms for large scale data, we use a factorization-based approach to matrix completion, which significantly speeds up the computations compared to repeated singular value decompositions without reducing the recovery quality. In the context of our theory and experiments, we also show that windowing the data too aggressively can have adverse effects on the recovery quality. To overcome this problem, we carry out our interpolations for each frequency independently while working with the entire frequency slice. The result is a computationally efficient, theoretically motivated framework for interpolating missing-trace data. Our tests on realistic two- and three-dimensional seismic data sets show that our method compares favorably, both in terms of computational speed and recovery quality, to existing curvelet-based

A version of this chapter has been published in Geophysics, 2015, vol 80, pages V97-V114.

and tensor-based techniques.

3.2 Introduction

Coarsely sampled seismic data creates substantial problems for seismic applications such as migration and inversion [Canning and Gardner, 1998, Sacchi and Liu, 2005]. In order to mitigate acquisition related artifacts, we rely on interpolation algorithms to reproduce the missing traces accurately. The aim of these interpolation algorithms is to reduce acquisition costs and to provide densely sampled seismic data to improve the resolution of seismic images and mitigate subsampling related artifacts such as aliasing. A variety of methodologies, each based on various mathematical techniques, have been proposed to interpolate seismic data. Some of the methods require transforming the data into different domains, such as the Radon [Bardan, 1987, Kabir and Verschuur, 1995], Fourier [Duijndam et al., 1999, Sacchi et al., 1998, Curry, 2009, Trad, 2009] and curvelet domains [Herrmann and Hennenfent, 2008a, Sacchi et al., 2009, Wang et al., 2010]. The CS approach exploits the resulting sparsity of the signal, i.e. small number of nonzeros [Donoho, 2006a] in these domains. In the CS framework, the goal for effective recovery is to first find a representation in which the signal of interest is sparse, or well-approximated by a sparse signal, and where the the mask encoding missing traces makes the signal much less sparse. Hennenfent and Herrmann [2006b], Herrmann and Hennenfent [2008a] successfully applied the ideas of CS to the reconstruction of missing seismic traces in the curvelet domain.

More recently, rank-minimization-based techniques have been applied to interpolating seismic data [Trickett et al., 2010, Oropeza and Sacchi, 2011, Kreimer and Sacchi, 2012c,a, Yang et al., 2013]. Rank minimization extends the theoretical and computational ideas of CS to the matrix case (see Recht et al. [2010a] and the references within). The key idea is to exploit the low-rank structure of seismic data when organized as a matrix, i.e. a small number of nonzero singular values or quickly decaying singular values. Oropeza and Sacchi [2011] identified that seismic temporal frequency slices organized into a block Hankel matrix, under ideal conditions, is a matrix of rank k, where k is the number of different plane waves in the window of analysis. These authors showed that additive noise and missing samples increase the rank of the block Hankel matrix, and the authors presented an iterative algorithm that resembles seismic data reconstruction with the method of projection onto convex sets, where they use a low-rank approximation of the Hankel matrix via the randomized singular value decomposition [Liberty et al., 2007, Halko et al., 2011b, Mahoney, 2011] to interpolate seismic temporal frequency slices. While this technique may be effective for interpolating data with a limited number of distinct dips, first, the approach requires embedding the data into an even larger space where each dimension of size n is mapped to a matrix of size $n \times n$, so a frequency slice with 4 dimensions becomes a Hankel tensor with 8 dimensions. Second, the process involves partitioning the input data in to smaller subsets that can be processed independently. As we know the theory of matrix completion is predicated upon the notion of an $m \times n$ matrix

being *relatively* low rank in order to ensure successful recovery. That is, the ratio of rank of the matrix to the ambient dimension, $\min(m, n)$, should be small for rank-minimizing techniques to be successful in recovering the matrix from appropriately subsampled data. With the practice of windowing, we are inherently *increasing* the relative rank by decreasing the ambient dimension. Although mathematically desirable due to the seismic signal being stationary in sufficiently small windows, the act of windowing from a matrix rank point of view can lead to lower quality results, as we will see later in experiments. Choosing window sizes apriori is also a difficult task, as it is not altogether obvious how to ensure that the resulting sub-volume is approximately a plane-wave. Previously proposed methods for automatic window size selection include Sinha et al. [2005], Wang et al. [2011] in the context of time-frequency analysis.

Other than the Hankel transformation, Yang et al. [2013] used a texture-patch based transformation of the data, initially proposed by Schaeffer and Osher [2013], to exploit the low-rank structure of seismic data. They showed that seismic data can be expressed as a combination of a few textures, due to continuity of seismic data. They divided the signal matrix into small $r \times r$ submatrices, which they then vectorized in to the columns of a matrix with r^2 rows using the same ordering, and approximated the resulting matrix using low rank techniques. Although experimentally promising, this organization has no theoretically motivated underpinning and its performance is difficult to predict as a function of the submatrix size. The authors proposed two algorithms to solve this matrix completion problem, namely accelerated proximal gradient method (APG) and low-rank matrix fitting (LMaFit). APG does not scale well to large scale seismic data because it involves repeated singular value decompositions, which are very expensive. LMaFit, on the other hand, parametrizes the matrix in terms of two low-rank factors and uses nonlinear successive-over-relaxation to reconstruct the seismic data, but *without* penalizing the nuclear norm of the matrix. As shown in Aravkin et al. [2014a], without a nuclear norm penalty, choosing an incorrect rank parameter k can lead to *overfitting* of the data and degrading the interpolated result. Moreover, Mishra et al. [2013] demonstrates the poor performance of LMaFit, both in terms of speed and solution quality, compared to more modern matrix completion techniques that penalize the nuclear norm.

Another popular approach to seismic data interpolation is to exploit the multi-dimensional nature of seismic data and parametrize it as a *low-rank tensor*. Many of the ideas from low rank matrices carry over to the multidimensional case, although there is no unique extension of the SVD to tensors. It is beyond the scope of this paper to examine all of the various tensor formats in this paper, but we refer to a few tensor-based seismic interpolation methods here. Kreimer and Sacchi [2012a] stipulates that the seismic data volume of interest is well captured by a $k-$rank Tucker tensor and subsequently propose a projection on to non-convex sets algorithm for interpolating missing traces. Silva and Herrmann [2013a] develop an algorithm for interpolating Hierarchical Tucker tensors, which are similar to Tucker tensors but have much smaller dimensionality. Trickett et al. [2013] proposes to take a structured outer product of the data volume, using a tensor ordering

similar to Hankel matrices, and performs tensor completion in the CP-Parafac tensor format. The method of Kreimer et al. [2013], wherein the authors consider a nuclear norm-penalizing approach in each matricization of the tensor, that is to say, the reshaping of the tensor, along each dimension, in to a matrix.

These previous CS-based approaches, using sparsity or rank-minimization, incur computational difficulties when applied to large scale seismic data volumes. Methods that involve redundant transforms, such as curvelets, or that add additional dimensions, such as taking outer products of tensors, are not computationally tractable for large data volumes with four or more dimensions. Moreover, a number of previous rank-minimization approaches are based on heuristic techniques and are not necessarily adequately grounded in theoretical considerations. Algorithmic components such as parameter selection can significantly affect the computed solution and "hand-tuning" parameters, in addition to incurring unnecessary computational overhead, may lead to suboptimal results [Owen and Perry, 2009, Kanagal and Sindhwani, 2010].

3.2.1 Contributions

Our contributions in this work are three-fold. First, we outline a practical framework for recovering seismic data volumes using matrix and tensor completion techniques built upon the theoretical ideas from CS. In particular, understanding this framework allows us to determine apriori when the recovery of signals sampled at sub-Nyquist will succeed or fail and provides the principles upon which we can design practical experiments to ensure successful recovery. The ideas themselves have been established for some time in the literature, albeit implicitly by means of the somewhat technical conditions of CS and matrix completion. We explicitly describe these ideas on a high level in a qualitative manner in the hopes of broadening the accessibility of these techniques to a wider audience. These principles are all equally necessary in order for CS-based approaches of signal recovery to succeed and we provide examples of how recovery can fail if one or more of these principles are omitted.

Second, we address the computational challenges of using these matrix-based techniques for seismic-data reconstruction, since traditional rank minimization algorithms rely on computing the singular value decomposition (SVD), which is prohibitively expensive for large matrices. To overcome this issue we propose to use either a fast optimization approach that combines the (SVD-free) matrix factorization approach recently developed by Lee et al. [2010a] with the Pareto curve approach proposed by Berg and Friedlander [2008] and the factorization-based parallel matrix completion framework dubbed Jellyfish [Recht and Ré, 2013]. We demonstrate the superior computational performances of both of these approaches compared to the tensor-based interpolation of Kreimer et al. [2013] as well as traditional curvelet-based approaches on realistic 2D and 3D seismic data sets.

Third, we examine the popular approach of *windowing* a large data volume in to smaller data

volumes to be processed in parallel and empirically demonstrate how such a process does not respect the inherent redundancy present in the data, degrading reconstruction quality as a result.

3.2.2 Notation

In this paper, we use lower case boldface letters to represent vectors (i.e. one-dimensional quantities), e.g., $\mathbf{b}, \mathbf{f}, \mathbf{x}, \mathbf{y}, \ldots$. We denote matrices and tensors using upper case boldface letters, e.g., $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \ldots$ and operators that act on vectors, matrices, or tensors will be denoted using calligraphic upper case letters, e.g., $\mathcal{A}$. 2D seismic volumes have one source and one receiver dimensions, denoted $x_{\text{src}}, x_{\text{rec}}$, respectively, and time, denoted t. 3D seismic volumes have two source dimensions, denoted $x_{\text{src}}, y_{\text{src}}$, two receiver dimensions, denoted $x_{\text{rec}}, y_{\text{rec}}$, and time t. We also denote midpoint and offset coordinates as $x_{\text{midpt}}, x_{\text{offset}}$ for the x-dimensions and similarly for the y-dimensions.

The Frobenius norm of a $m \times n$ matrix $\mathbf{X}$, denoted as $\|\mathbf{X}\|_F$, is simply the usual $\ell 2$ norm of $\mathbf{X}$ when considered as a vector, i.e., $\|\mathbf{X}\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} \mathbf{X}_{ij}^2}$. We write the SVD of $\mathbf{X}$ as $\mathbf{X} = USV^H$, where U and V are orthogonal and $S = \text{diag}(s_1, s_2, \ldots, s_r)$ is a block diagonal matrix of singular values, $s_1 \geq s_2 \geq \cdots \geq s_r \geq 0$. The matrix $\mathbf{X}$ has rank k when $s_k > 0$ and $s_{k+1} = s_{k+2} = \cdots = s_r = 0$. The nuclear norm of $\mathbf{X}$ is defined as $\|\mathbf{X}\|_* = \sum_{i=1}^{r} s_i$.

We will use the matricization operation freely in the text below, which reshapes a tensor in to a matrix along specific dimensions. Specifically, if $\mathbf{X}$ is a temporal frequency slice with dimensions $x_{\text{src}}, y_{\text{src}}, x_{\text{rec}}, y_{\text{rec}}$ indexed by $i = 1, \ldots, 4$, the matrix $\mathbf{X}^{(i)}$ is formed by vectorizing the ith dimension along the rows and the remaining dimensions along the columns. Matricization can also be performed not only along singleton dimensions, but also with groups of dimensions. For example, $\mathbf{X}^{(i)}$ with $i = x_{\text{src}}, y_{\text{src}}$ places the x and y source dimensions along the columns and the remaining dimensions along the columns.

3.3 Structured signal recovery

In this setting, we are interested in completing a matrix $\mathbf{X}$ when we only view a subset of its entries. For instance, in the 2D seismic data case, $\mathbf{X}$ is typically a frequency slice and missing shots correspond to missing columns from this matrix. Matrix completion arises as a natural extension of Compressive Sensing ideas to recovering two dimensional signals. Here we consider three core components of matrix completion.

1. **Signal structure - low rank**
   Compressed Sensing is a theory that is deals with recovering vectors $\mathbf{x}$ that are sparse, or have a few nonzeros. For a matrix $\mathbf{X}$, a direct analogue for sparsity in a signal $\mathbf{x}$ is sparsity in the *singular values* of $\mathbf{X}$. We are interested in the case where the singular values of $\mathbf{X}$ decay quickly, so that $\mathbf{X}$ is well approximated by a rank k matrix. The set of all rank-k matrices has low dimensionality compared to the ambient space of $m \times n$ matrices, which will allow

us to recover a low rank signal from sufficiently incoherent measurements.

When our matrix $\mathbf{X}$ has slowly decaying singular values, i.e. is high rank, we consider transformations that promote quickly decaying singular values, which will allow us to recover our matrix in another domain.

Since we are sampling points from our underlying matrix $\mathbf{X}$, we want to make sure that $\mathbf{X}$ is not too "spiky" and is sufficiently "spread out". If our matrix of interest was, for instance, the matrix of all zeros with one nonzero entry, we could not hope to recover this matrix without sampling the single, nonzero entry. In the seismic case, given that our signals of interest are composed of oscillatory waveforms, they are rarely, if ever, concentrated in a single region of, say, (source,receiver) space.

2. **Structure-destroying sampling operator**
   Since our matrix has an unknown but small rank k, we will look for the matrix $\mathbf{X}$ of smallest rank that fits our sampled data, i.e., $\mathcal{A}(\mathbf{X}) = \mathbf{B}$, for a subsampling operator $\mathcal{A}$. As such, we need to employ subsampling schemes that increase the rank or decay of the singular values of the matrix. That is to say, we want to consider sampling schemes that are incoherent with respect to the left and right singular vectors. Given a subsampling operator $\mathcal{A}$, the worst possible subsampling scheme for the purposes of recovery would be removing columns (equivalently, rows) from the matrix, i.e. $\mathcal{A}(\mathbf{X}) = \mathbf{X}\mathbf{I}_k$, where $\mathbf{I}_k$ is a subset of the columns of identity matrix. Removing columns from the matrix can never allow for successful reconstruction because this operation lowers the rank, and therefore the original matrix $\mathbf{X}$ is no longer the matrix of smallest rank that matches the data (for instance, the data itself would be a candidate solution).

   Unfortunately, for, say, a 2D seismic data frequency slice $\mathbf{X}$ with sources placed along the columns and receivers along the rows, data is often acquired with missing sources, which translates to missing columns of $\mathbf{X}$. Similarly, periodic subsampling can be written as $\mathcal{A}(\mathbf{X}) = \mathbf{I}_k^T\mathbf{X}\mathbf{I}_{k'}$, where $\mathbf{I}_k, \mathbf{I}_{k'}$ are subsets of the columns of the identity matrix. A similar consideration shows that this operator lowers the rank and thus rank minimizing interpolation will not succeed in this sampling regime.

   The problematic aspect of the aforementioned sampling schemes is that they are separable with respect to the matrix. That is, if $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^H$ is the singular value decomposition of $\mathbf{X}$, the previously mentioned schemes yield a subsampling operator of the form $\mathcal{A}(\mathbf{X}) = \mathbf{C}\mathbf{X}\mathbf{D}^H = (\mathbf{C}\mathbf{U})\mathbf{S}(\mathbf{D}\mathbf{V})^H$, for some matrices $\mathbf{C}, \mathbf{D}$. In the compressed sensing context, this type of sampling is coherent with respect to the left and right singular vectors, which is an unfavourable recovery scenario.

   The incoherent sampling considered in the matrix completion literature is that of uniform random sampling, wherein the individual entries of $\mathbf{X}$ are sampled from the matrix with equal

probability [Candès and Recht, 2009, Recht, 2011]. This particular sampling scheme, although theoretically convenient to analyze, is impractical to implement in the seismic context as it corresponds to removing (source, receiver) pairs from the data. Instead, we will consider non-separable transformations, i.e., transforming data from the source-receiver domain to the midpoint-offset domain, under which the missing sources operator is incoherent. The resulting transformations will simultaneously increase the decay of the singular values of our original signal, thereby lowering its rank, and slow the decay of the singular values of the subsampled signal, thereby creating a favourable recovery scenario.

3. **Structure-promoting optimization program**
   Since we assume that our target signal $\mathbf{X}$ is low-rank and that subsampling *increases* the rank, the natural approach to interpolation is to find the matrix of *lowest* possible rank that agrees with our observations. That is, we solve the following problem for $\mathcal{A}$, our measurement operator, and $\mathbf{B}$, our subsampled data, up to a given tolerance σ,

$$\begin{aligned} &\underset{\mathbf{X}}{\text{minimize}} \ \|\mathbf{X}\|_* \\ &\text{subject to } \|\mathcal{A}(\mathbf{X}) - \mathbf{B}\|_F \leq \sigma. \end{aligned} \tag{3.1}$$

   Similar to using the ℓ_1 norm in the sparse recovery case, minimizing the nuclear norm promotes low-rank structure in the final solution. Here we refer to this problem as Basis Pursuit Denoising ($BPDN_\sigma$).

In summary, these three principles are all necessary for the recovery of subsampled signals using matrix completion techniques. Omitting any one of these three principles will, in general, cause such methods to fail, which we will see in the next section. Although this framework is outlined for matrix completion, a straightforward extension of this approach also applies to the tensor completion case.

3.4 Low-rank promoting data organization

Before we can apply matrix completion techniques to interpolate $\mathbf{F}$, our unvectorized frequency slice of fully-sampled data, we must deal with the following issues. First, in the original (src, rec) domain, the missing sources operator, $\mathcal{A}$, removes columns from $\mathbf{F}$, which sets the singular values to be set to zero at the end of the spectrum, thereby decreasing the rank.

Second, $\mathbf{F}$ itself also has high rank, owing to the presence of strong diagonal entries (zero offset energy) and subsequent off-diagonal oscillations. Our previous theory indicates that naively applying matrix completion techniques in this domain will yield poor results. Simply put, we are missing two of the prerequisite signal recovery principles in the (src, rec) domain, which we can see by plotting the decay of singular values in Figure 3.1. In light of our previous discussion, we will

examine different transformations under which the missing sources operator increases the singular values of our data matrix and hence promotes recovery in an alternative domain.

3.4.1 2D seismic data

In this case, we use the *Midpoint-Offset* transformation, which defines new coordinates for the matrix as

$$x_{\text{midpt}} = \frac{1}{2}(x_{\text{src}} + x_{\text{rec}})$$
$$x_{\text{offset}} = \frac{1}{2}(x_{\text{src}} - x_{\text{rec}}).$$

This coordinate transformation rotates the matrix $\mathbf{F}$ by 45 degrees and is a tight frame operator with a nullspace, as depicted in Figure 3.2. If we denote this operator by $\mathcal{M}$, then $\mathcal{M}^*\mathcal{M} = I$, so transforming from (src, rec) to (midpt, offset) to (src, rec) returns the original signal, but $\mathcal{M}\mathcal{M}^* \neq I$, so the transformation from (midpt, offset) to (src, rec) and back again does not return the original signal. By using this transformation, we move the strong diagonal energy to a single column in the new domain, which mitigates the slow singular value decay in the original domain. Likewise, the restriction operator $\mathcal{A}$ now removes super-/sub-*diagonals* from $\mathbf{F}$ rather than columns, demonstrated in Figure 3.2, which results in an overall increase in the singular values, as seen in Figure 3.1, placing the interpolation problem in a favourable recovery scenario as per the previous section. Our new optimization variable is $\tilde{\mathbf{X}} = \mathcal{M}(\mathbf{X})$, which is the data volume in the midpoint-offset domain, and our optimization problem is therefore

$$\begin{aligned} \underset{\tilde{\mathbf{X}}}{\text{minimize}} \ & \|\tilde{\mathbf{X}}\|_* \\ \text{s.t.} \ & \|\mathcal{A}\mathcal{M}^*(\tilde{\mathbf{X}}) - \mathbf{B}\|_F \leq \sigma. \end{aligned}$$

3.4.2 3D seismic data

Unlike in the matrix-case, there is no unique generalization of the SVD to tensors and as a result, there is no unique notion of *rank* for tensors. Instead, can consider the rank of different matricizations of $\mathbf{F}$. Instead of restricting ourselves to matricizations $\mathbf{F}^{(i)}$ where $i = x_{src}, y_{src}, x_{rec}, y_{rec}$, we consider the case where $i = \{x_{\text{src}}, y_{\text{src}}\}, \{x_{\text{src}}, x_{\text{rec}}\}, \{x_{\text{src}}, y_{\text{rec}}\}, \{y_{\text{src}}, x_{\text{rec}}\}$. Owing to the reciprocity relationship between sources and receivers in $\mathbf{F}$, we only need to consider two different matricizations of $\mathbf{F}$, which are depicted in Figure 3.3 and Figure 3.4. As we see in Figure 3.5, the $i = (x_{rec}, y_{rec})$ organization, that is, placing both receiver coordinates along the rows, results in a matrix that has high rank and the missing sources operator removes columns from the matrix, decreasing the rank as mentioned previously. On the other hand, the $i = (y_{src}, y_{rec})$ matriciza-

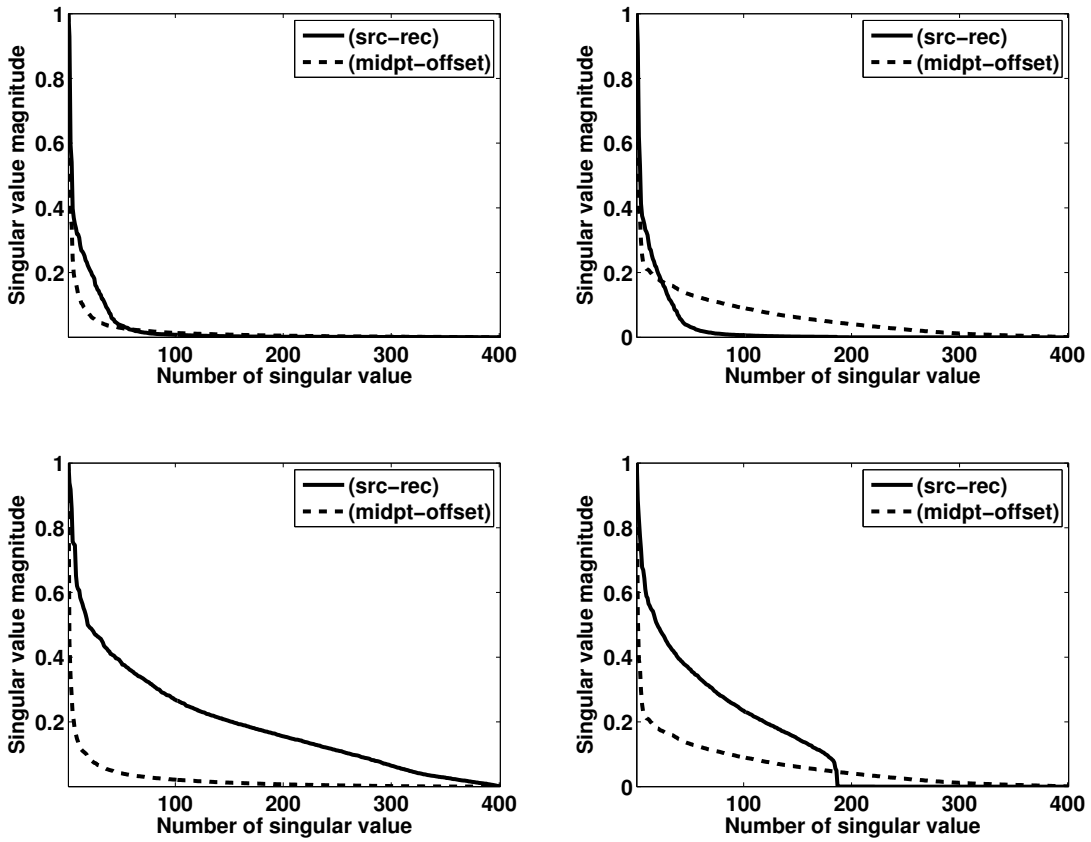


Figure 3.1: Singular value decay in the source-receiver and midpoint-offset domain. *Left* : fully sampled frequency slices. *Right* : 50% missing shots. *Top*: low frequency slice. *Bottom*: high frequency slice. Missing source subsampling *increases* the singular values in the (midpoint-offset) domain instead of *decreasing* them in the (src-rec) domain.

tion yields fast decay of the singular values for the original signal and a subsampling operator that causes the singular values to increase. This scenario is much closer to the idealized matrix completion sampling, which would correspond to the nonphysical process of randomly removing $(x_{src}, y_{src}, x_{rec}, y_{rec})$ points from $\mathbf{F}$. We note that this data organization has been considered in the context of solution operators of the wave equation in Demanet [2006b], which applies to our case as our data volume $\mathbf{F}$ is the restriction of a Green's function to the acquisition surface.

3.5 Large scale data reconstruction

In this section, we explore the modifications necessary to extend matrix completion to 3D seismic data and compare this approach to an existing tensor-based interpolation technique. Matrix-

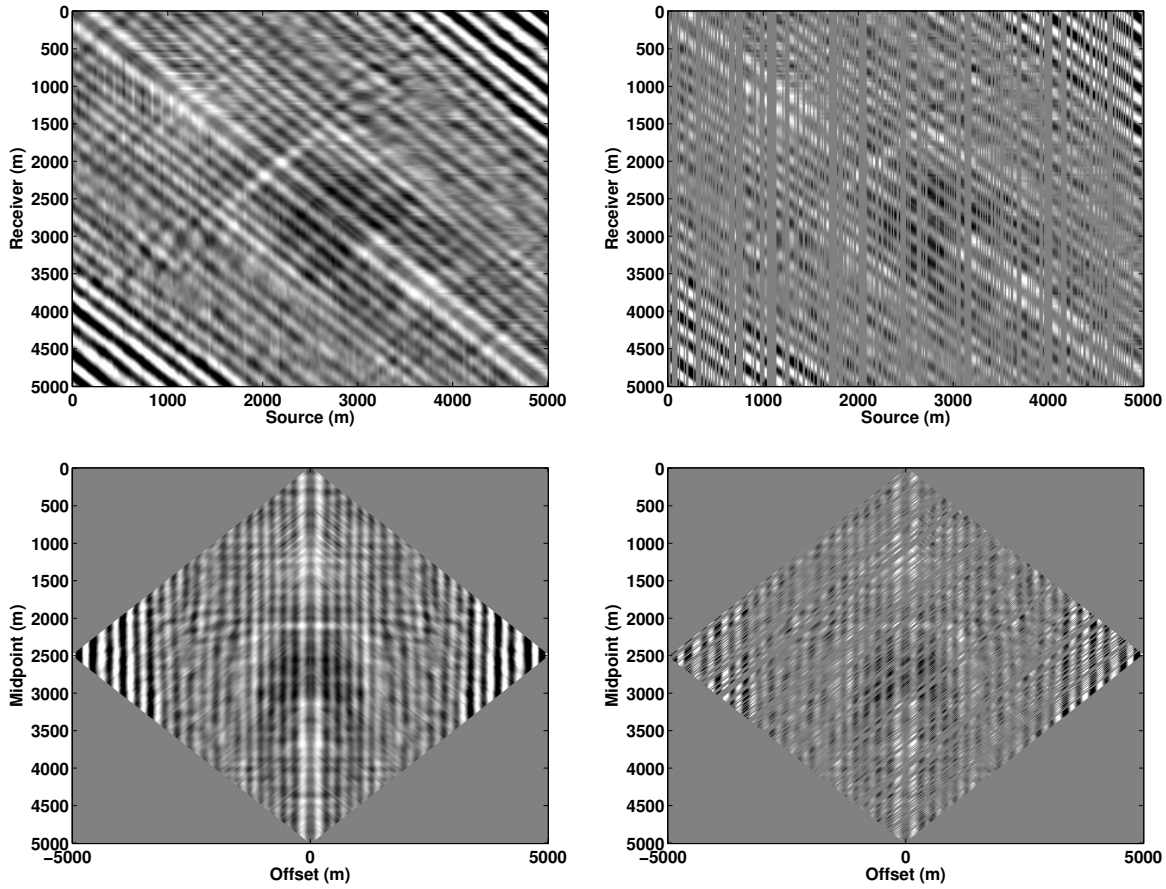


Figure 3.2: A frequency slice from the the seismic dataset from Nelson field. *Left*: Fully sampled data. *Right*: 50% subsampled data. *Top*: Source-receiver domain. *Bottom*: Midpoint-offset domain.

completion techniques, after some modification, easily scale to interpolate large, multidimensional seismic volumes.

3.5.1 Large scale matrix completion

For the matrix completion approach, the limiting component for large scale data is that of the *nuclear norm projection*. As mentioned in Aravkin et al. [2014a], the projection on to the set $\|\mathbf{X}\|_* \leq \tau$ requires the computation of the SVD of $\mathbf{X}$. The main computational costs of computing the SVD of a $n \times n$ matrix has computational complexity $O(n^3)$, which is prohibitively expensive when $\mathbf{X}$ has tens of thousands or even millions of rows and columns. On the assumption that $\mathbf{X}$ is approximately low-rank at a given iteration, other authors such as Stoll [2012] compute a partial SVD using a Krylov approach, which is still cost-prohibitive for large matrices.

We can avoid the need for the expensive computation of SVDs via a well known *factorization* of the nuclear norm. Specifically, we have the following characterization of the nuclear norm, due

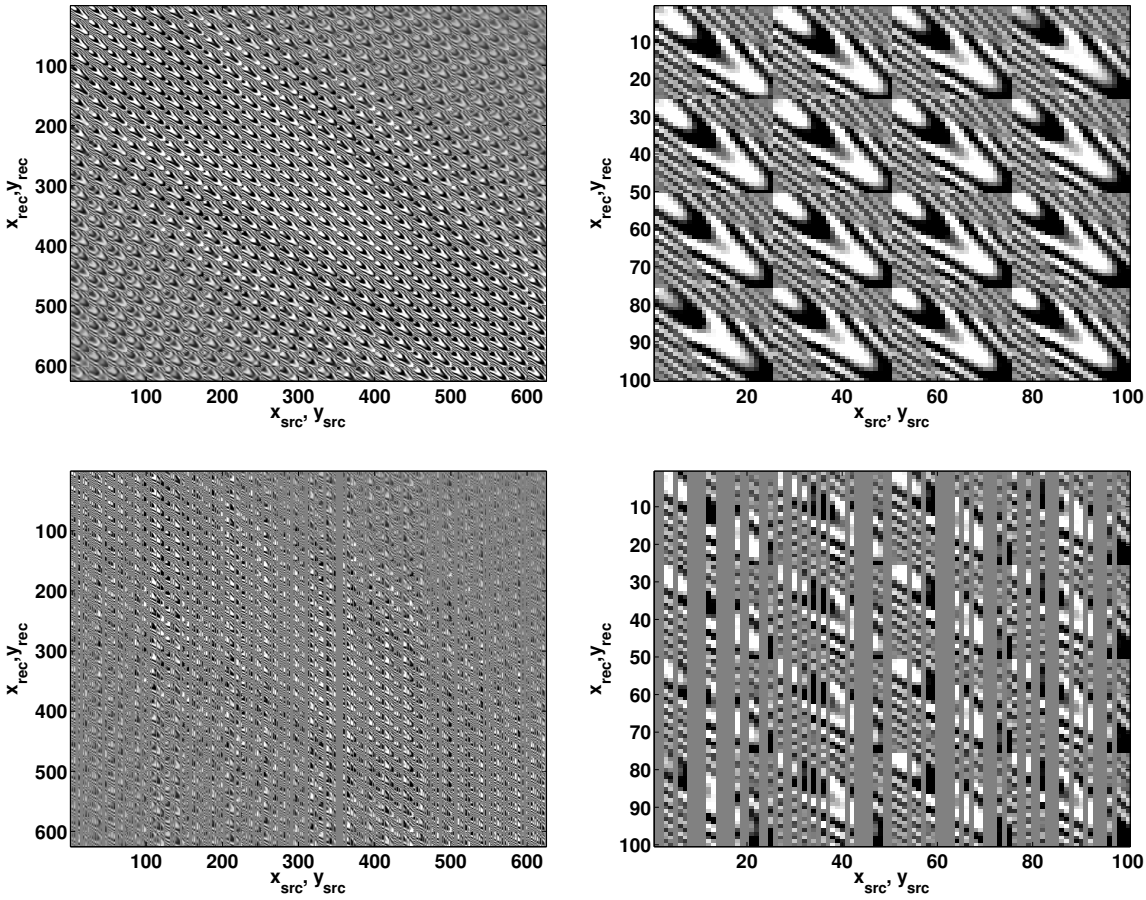


Figure 3.3: (x_{rec}, y_{rec}) matricization. *Top*: Full data volume. *Bottom*: 50% missing sources. *Left*: Fully sampled data. *Right*: Zoom plot

to Srebro [2004],

$$\|\mathbf{X}\|_* = \underset{\mathbf{L},\mathbf{R}}{\text{minimize}} \ \frac{1}{2}(\|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2)$$
$$\text{subject to } \mathbf{X} = \mathbf{L}\mathbf{R}^T.$$

This allows us to write $\mathbf{X} = \mathbf{L}\mathbf{R}^T$ for some placeholder variables $\mathbf{L}$ and $\mathbf{R}$ of a prescribed rank k. Therefore, instead of projecting on to $\|\mathbf{X}\|_* \leq \tau$, we can instead project on to the *factor* ball $\frac{1}{2}(\|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2) \leq \tau$. This factor ball projection only involves computing $\|\mathbf{L}\|_F^2$, $\|\mathbf{R}\|_F^2$ and scaling the factors by a constant, which is substantially cheaper than computing the SVD of $\mathbf{X}$.

Equipped with this factorization approach, we can still use the basic idea of SPGℓ_1 to flip the objective and the constraints. The resulting subproblems for solving $BPDN_\sigma$ can be solved *much* more efficiently in this factorized form, while still maintaining the quality of the solution. The resulting algorithm is dubbed SPG-LR by Aravkin et al. [2014a]. This reformulation allows us to

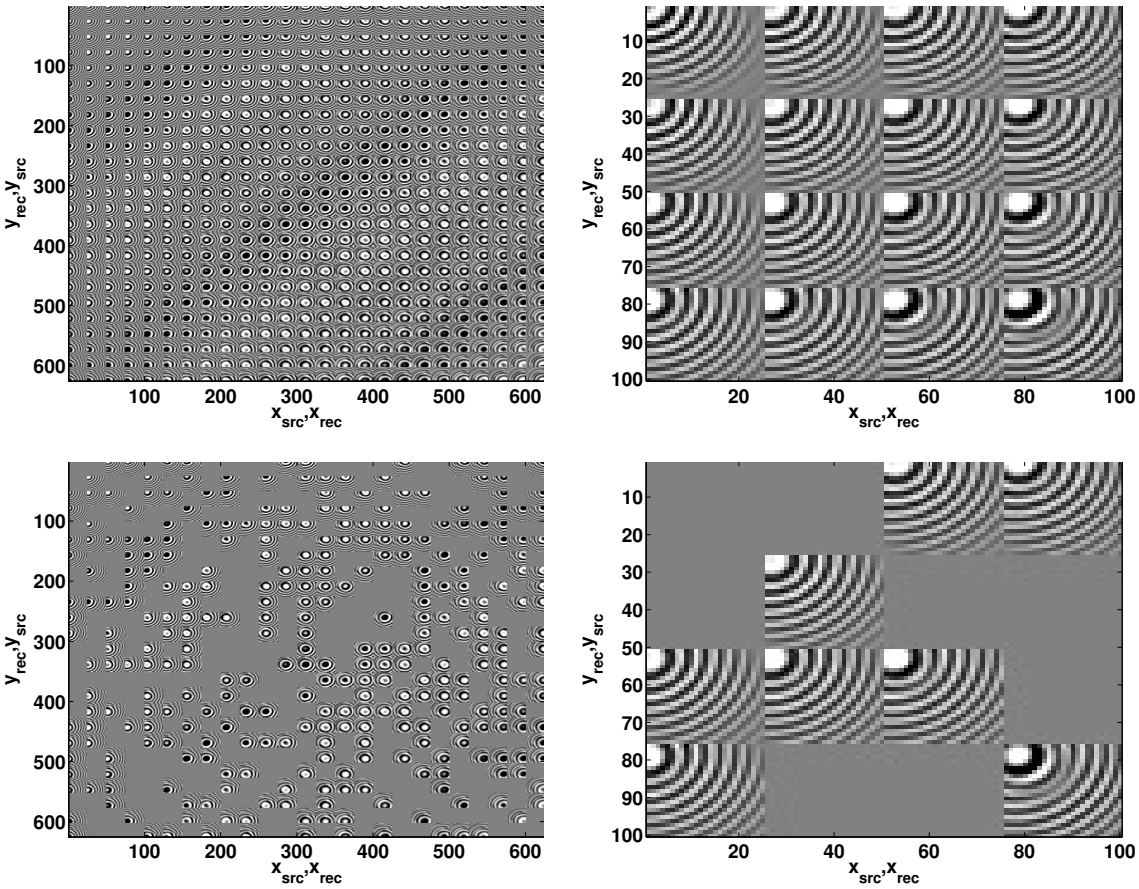


Figure 3.4: (y_{src}, y_{rec}) matricization. *Top*: Fully sampled data. *Bottom*: 50% missing sources. *Left*: Full data volume. *Right*: Zoom plot. In this domain, the sampling artifacts are much closer to the idealized 'pointwise' random sampling of matrix completion.

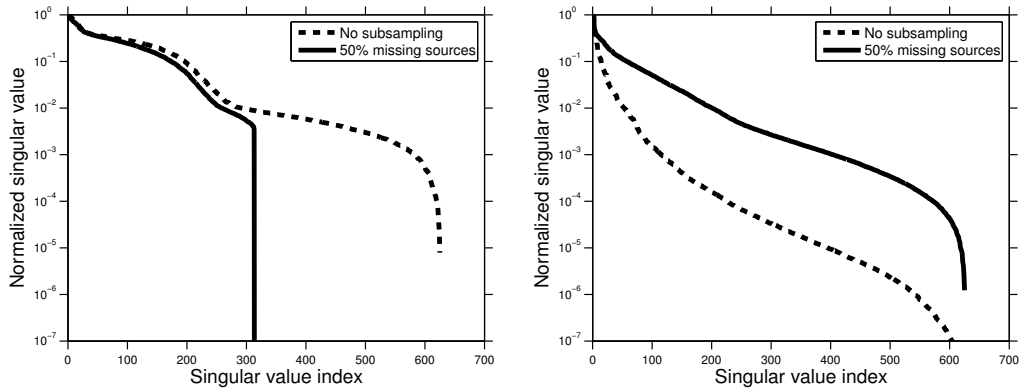


Figure 3.5: Singular value decay (normalized) of the *Left*: $(x_{\mathrm{rec}}, y_{\mathrm{rec}})$ matricization and *Right*: $(y_{\mathrm{src}}, y_{\mathrm{rec}})$ matricization for full data and 50% missing sources.

apply these matrix completion techniques to large scale seismic data interpolation.

This factorization turns the convex subproblems for solving $BPDN_\sigma$ posed in terms of $\mathbf{X}$ into a nonconvex problem in terms of the variables $\mathbf{L}, \mathbf{R}$, so there is a possibility for local minima or non-critical stationary points to arise when using this approach. As it turns out, as long as the prescribed rank k is larger than the rank of the optimal $\mathbf{X}$, any local minima encountered in the factorized problem is actually a global minimum [Burer and Monteiro, 2005, Aravkin et al., 2014a]. The possibility of non-critical stationary points is harder to discount, and remains an open problem. There is preliminary analysis indicating that initializing $\mathbf{L}$ and $\mathbf{R}$ so that $\mathbf{L}\mathbf{R}^T$ is sufficiently close to the true $\mathbf{X}$ will ensure that this optimization program will converge to the true solution [Sun and Luo, 2014]. In practice, we initialize $\mathbf{L}$ and $\mathbf{R}$ randomly with appropriately scaled Gaussian random entries, which does not noticeably change the recovery results across various random realizations.

An alternative approach to solving the factorized $BPDN_\sigma$ is to relax the data constraint of Equation (3.1) in to the objective, resulting in the QP_λ formulation,

$$\min_{\mathbf{L},\mathbf{R}} \frac{1}{2}\|\mathcal{A}(\mathbf{L}\mathbf{R}^H) - \mathbf{B}\|_F^2 + \lambda(\|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2). \tag{3.2}$$

The authors in Recht and Ré [2013] exploit the resulting *independance* of various subblocks of the $\mathbf{L}$ and $\mathbf{R}$ factors to create a partitioning scheme that updates components of these factors in parallel, resulting in a parallel matrix completion framework dubbed Jellyfish. By using this Jellyfish approach, each QP_λ problem for fixed λ and fixed internal rank k can be solved very efficiently and cross-validation techniques can choose the optimal λ and rank parameters.

3.5.2 Large scale tensor completion

Following the approach of Kreimer et al. [2013], which applies the method developed in Gandy et al. [2011] to seismic data, we can also exploit the *tensor* structure of a frequency slice $\mathbf{F}$ for interpolating missing traces.

We now stipulate that each matricization $\mathbf{F}^{(i)}$ for $i = 1, \ldots, 4$ has low-rank. We can proceed in an analogous way to the matrix completion case by solving the following problem

$$\begin{aligned} &\underset{\mathbf{F}}{\text{minimize}} \sum_{i=1}^{4} \|\mathbf{F}^{(i)}\|_* \\ &\text{subject to } \|\mathcal{A}(\mathbf{F}) - \mathbf{B}\|_2 \leq \sigma, \end{aligned}$$

i.e. look for the tensor $\mathbf{F}$ that has *simultaneously* the lowest rank in each matricization $\mathbf{F}^{(i)}$ that fits the subsampled data $\mathbf{B}$. In the case of Kreimer et al. [2013], this interpolation is performed in the $(x_{\text{midpt}}, y_{\text{midpt}}, x_{\text{offset}}, y_{\text{offset}})$ domain on each frequency slice, which we also employ in our later experiments.

To solve this problem, the authors in Kreimer et al. [2013] use the Douglas-Rachford variable

splitting technique that creates 4 additional copies of the variable $\mathbf{F}$, denoted $\mathbf{X}_i$, with each copy corresponding to each matricization $\mathbf{F}^{(i)}$. This is an inherent feature of this approach to solve convex optimization problems with coupled objectives/constraints and thus cannot be avoided or optimized away. The authors then use an Augmented Lagrangian approach to solve the *decoupled* problem

$$\begin{aligned} \underset{\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}_3,\mathbf{X}_4,\mathbf{F}}{\text{minimize}} \; & \sum_{i=1}^{4} \|\mathbf{X}_i\|_* + \lambda \|\mathcal{A}(\mathbf{F}) - \mathbf{B}\|_2^2 \\ \text{subject to } & \mathbf{X}_i = \mathbf{F}^{(i)} \text{ for } i = 1, \ldots, 4. \end{aligned} \tag{3.3}$$

The resulting problem is convex, and thus has a unique solution. We refer to this method as the alternating direction method of multipliers (ADMM) tensor method. This variable splitting technique can be difficult to implement for realistic problems, as the tensor $\mathbf{F}$ is often unable to be stored fully in working memory. Given the large number of elements of $\mathbf{F}$, creating at minimum four extraneous copies of $\mathbf{F}$ can quickly overload the storage and memory of even a large computing cluster. Moreover, there are theoretical and numerical results that state that this problem formulation is in fact no better than imposing the nuclear norm penalty on a *single* matricization of $\mathbf{F}$, at least in the case of Gaussian measurements [Oymak et al., 2012, Signoretto et al., 2011]. We shall see a similar phenomenon in our subsequent experiments.

Penalizing the nuclear norm in this fashion, as in all methods that use an explicit nuclear norm penalty, scales very poorly as the problem size grows. When our data $\mathbf{F}$ has four or more dimensions, the cost of computing the SVD of one of its matricizations easily dominates the overall computational costs of the method. Applying this operation four times per iteration in the above problem, as is required due to the variable splitting, prevents this technique from performing efficiently for large realistic problems.

3.6 Experiments

We perform seismic data interpolation on five different data sets. In case of 2D, the first data set, which is a shallow-water marine scenario, is from the Nelson field provided to us by PGS. The Nelson data set contains 401×401 sources and receivers with the temporal sampling interval of 0.004s. The second synthetic data set is from the Gulf of Mexico (GOM) and is provided to us by the Chevron. It contains 3201 sources and 801 receivers with a spatial interval of 25m. The third data set is simulated on a synthetic velocity model (see Berkhout and Verschuur [2006]) using IWave [Symes et al., 2011a]. An anticline salt structure over-lies the target, i.e., a fault structure. A seismic line is modelled using a fixed-spread configuration where sources and receivers are placed at an interval of 15m. This results in a data set of 361×361 sources and receivers.

Our 3D examples consist of two different data sets. The first data set is generated on a synthetic

single-layer model. This data set has 50 sources and 50 receivers and we use a frequency slice at 4 Hz. This simple data set allows us to compare the running time of the various algorithms under consideration. The Compass data set is provided to us by the BG Group and is generated from an unknown but geologically complex and realistic model. We selected a few 4D monochromatic frequency slices from this data set at 4.68, 7.34, and 12.3Hz. Each monochromatic frequency slice has 401×401 receivers spaced by 25m and 68×68 sources spaced by 150m. In all the experiments, we initialize $\mathbf{L}$ and $\mathbf{R}$ using random numbers.

3.6.1 2D seismic data

In this section, we compare matrix-completion based techniques to existing curvelet-based interpolation for interpolating 2D seismic data. For details on the curvelet-based reconstruction techniques, we refer to [Herrmann and Hennenfent, 2008a, Mansour et al., 2013]. For concreteness, we concern ourselves with the missing-sources scenario, although the missing-receivers scenario is analogous. In all the experiments, we set the data misfit parameter σ to be equal to $\eta\|\mathbf{B}\|_F$ where $\eta \in (0, 1)$ is the fraction of the input data energy to fit.

Nelson data set

Here, we remove 50%, 75% of the sources, respectively. For the sake of comparing curvelet-based and rank-minimization based reconstruction methods on identical data, we first interpolate a single 2D frequency slice at 10 Hz. When working with frequency slices using curvelets, Mansour et al. [2013] showed that the best recovery is achieved in the midpoint-offset domain, owing to the increased curvelet sparsity. Therefore, in order to draw a fair comparison with the matrix-based methods, we perform curvelet-based and matrix-completion based reconstruction in the midpoint-offset domain.

We summarize these results of interpolating a single 2D frequency slice in Table 3.1. Compared to the costs associated to applying the forward and adjoint curvelet transform, SPG-LR is much more efficient and, as such, this approach significantly outperforms the ℓ_1-based curvelet interpolation. Both methods perform similarly in terms of reconstruction quality for low frequency slices, since these slices are well represented both as a sparse superposition of curvelets and as a low-rank matrix. High frequency data slices, on the other hand, are empirically high rank, which can be shown explicitly for a homogeneous medium as a result of Lemma 2.7 in Engquist and Ying [2007], and we expect matrix completion to perform less well in this case, as high frequencies contains oscillations away from the zero-offset. On the other hand, these oscillations can be well approximated by low-rank values in localized domains. To perform the reconstruction of seismic data in the high frequency regime, Kumar et al. [2013a] proposed to represent the matrix in the Hierarchical semi-separable (HSS) format, wherein data is first windowed in off-diagonal and diagonal blocks and the diagonal blocks are recursively partitioned. The interpolation is then performed on each subset separately. In the interest of brevity, we omit the inclusion of this approach here.

Additionally, since the high frequency slices are very oscillatory, they are much less sparse in the curvelet dictionary.

Owing to the significantly faster performance of matrix completion compared to the curvelet-based method, we apply the former technique to an entire seismic data volume by interpolating each frequency slice in the 5-85Hz band. Figures 3.6 show the interpolation results in case of 75% missing traces. In order to get the best rank values to interpolation the full seismic line, we first performed the interpolation for the frequency slices at 10 Hz and 60 Hz. The best rank value we get for these two slices is 30 and 60 respectively. Keeping this in mind, we work with all of the monochromatic frequency slices and adjust the rank linearly from 30 to 60 when moving from low to high frequencies. The running time is 2 h 18 min using SPG-LR on a 2 quad-core 2.6GHz Intel processor with 16 GB memory and implicit multithreading via LAPACK libraries. We can see that we have low-reconstruction error with little coherent energy in the residual when 75% of the sources are missing. Figure 3.7 shows the qualitative measurement of recovery for all frequencies in the energy-band. We can further mitigate such coherent residual energy by exploiting additional structures in the data such as symmetry, as in Kumar et al. [2014].

Remark

It is important to note that if we omit the first two principles of matrix completion by interpolating the signal in the source-receiver domain, as discussed previously, we obtain very poor results, as shown in Figure 3.8. Similar to CS-based interpolation, choosing an appropriate transform-domain for matrix and tensor completion is vital to ensure successful recovery.

Table 3.1: Curvelet versus matrix completion (MC). **Real data** results for completing a frequency slice of size 401×401 with 50% and 75% missing sources. *Left*: 10 Hz (low frequency), *right*: 60 Hz (high frequency). SNR, computational time, and number of iterations are shown for varying levels of $\eta = 0.08, 0.1$.

		Curvelets		**MC**	
η		0.08	0.1	0.08	0.1
50%	SNR (dB)	18.2	17.3	**18.6**	17.7
	time (s)	1249	1020	15	**10**
	iterations	123	103	191	124
75%	SNR (dB)	**13.5**	13.2	13.0	13.3
	time (s)	1637	1410	8.5	**8**
	iterations	162	119	105	104

		Curvelets		**MC**	
η		0.08	0.1	0.08	0.1
50%	SNR (dB)	10.5	10.4	**12.5**	12.4
	time (s)	1930	1549	19	**13**
	iteration	186	152	169	118
75%	SNR (dB)	6.0	5.9	6.9	**7.0**
	time (s)	3149	1952	15	**10**
	iteration	284	187	152	105

Gulf of mexico data set

In this case, we remove 80% of the sources. Here, we perform the interpolation on a frequency spectrum of 5-30hz. Figure 3.10 shows the comparison of the reconstruction error using rank-

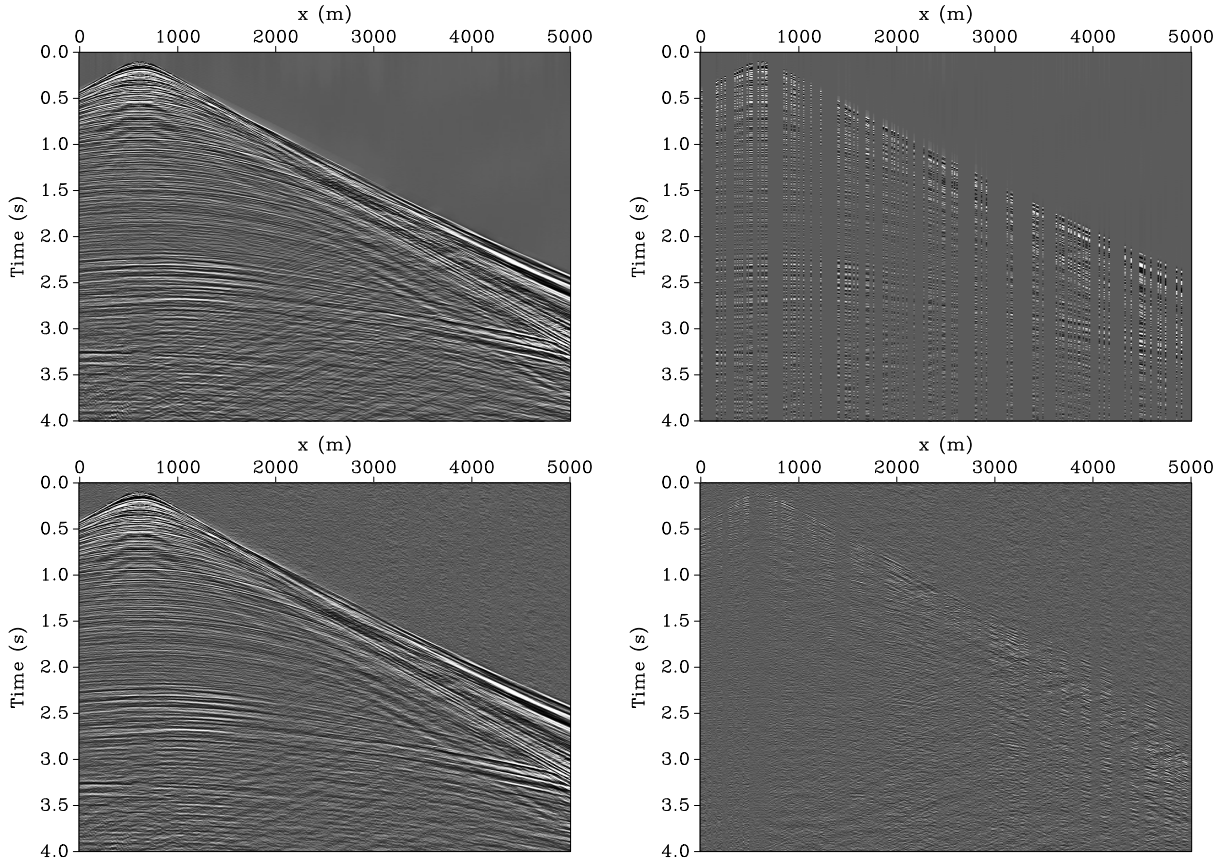


Figure 3.6: Missing-trace interpolation. *Top* : Fully sampled data and 75% subsampled common receiver gather. *Bottom* Recovery and residual results with a SNR of 9.4 dB.

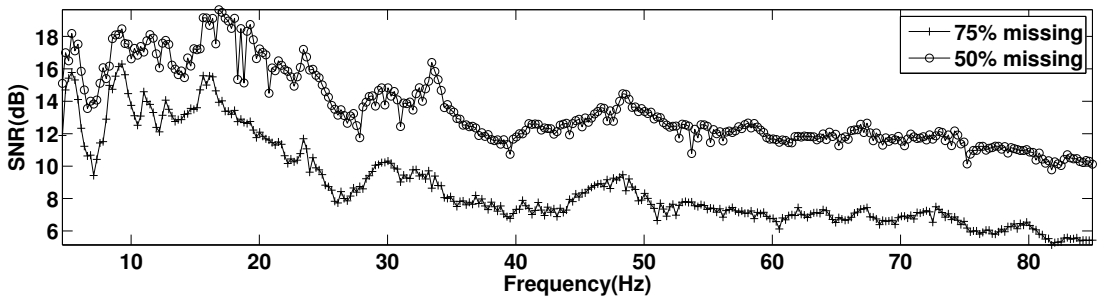


Figure 3.7: Qualitative performance of 2D seismic data interpolation for 5-85 Hz frequency band for 50% and 75% subsampled data.

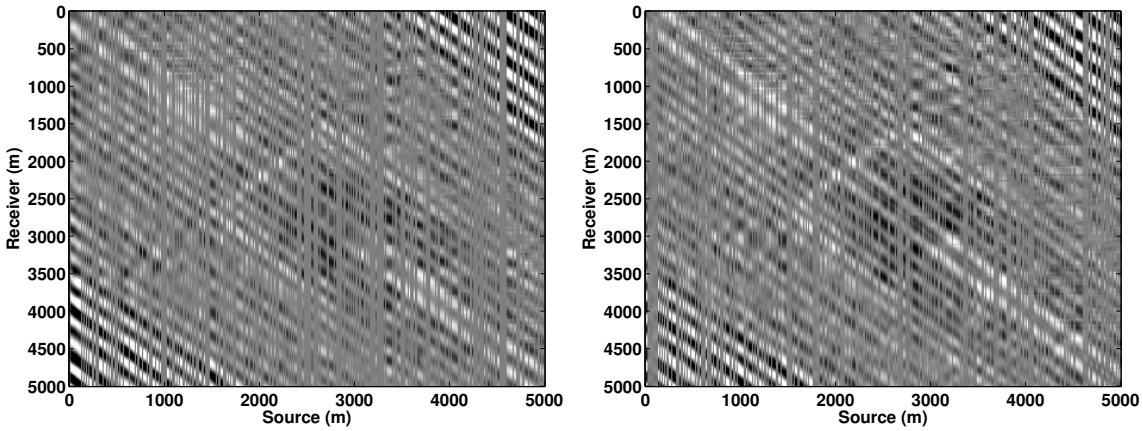


Figure 3.8: Recovery results using matrix-completion techniques. *Left*: Interpolation in the source-receiver domain, low-frequency SNR 3.1 dB. *Right*: Difference between true and interpolated slices. Since the sampling artifacts in the source-receiver domain do *not* increase the singular values, matrix completion in this domain is unsuccesful. This example highlights the necessity of having the appropriate principles of low-rank recovery in place before a seismic signal can be interpolated effectively.

minimization based approach for a frequency slice at 7Hz and 20 Hz. For visualization purposes, we only show a subset of interpolated data corresponding to the square block in Figure 3.9, but we interpolate the monochromatic slice over all sources and receivers. Even in the highly sub-sampled case of 80%, we are still able to recover to a high SNR of 14.2 dB, 10.5dB, respectively, but we start losing coherent energy in the residual as a result of the high-subsampling ratio. These results indicate that even in complex geological environments, low-frequencies are still low-rank in nature. This can also be seen since, for a continuous function, the smoother the function is (i.e., the more derivatives it has), the faster its singular values decay (see, for instance, Chang and Ha [1999]). For comparison purposes, we plot the frequency-wavenumber spectrum of the 20Hz frequency slice in Figure 3.11 along with the corresponding spectra of matrix with 80% of the sources removed periodically and uniform randomly. In this case, the volume is approximately three times aliased in the bandwidth of the original signal for periodic subsampling, while the randomly subsampled case has created noisy aliases. The average sampling interval for both schemes is the same. As shown in this figure, the interpolated matrix has a significantly improved spectrum compared to the input. Figure 3.12 shows the interpolation result over a common receiver gather using rank-minimization based techniques. In this case, we set the rank parameter to be 40 and use the same rank for all the frequencies. The running time on a single frequency slice in this case is 7 min using SPG-LR and 1320 min using curvelets.

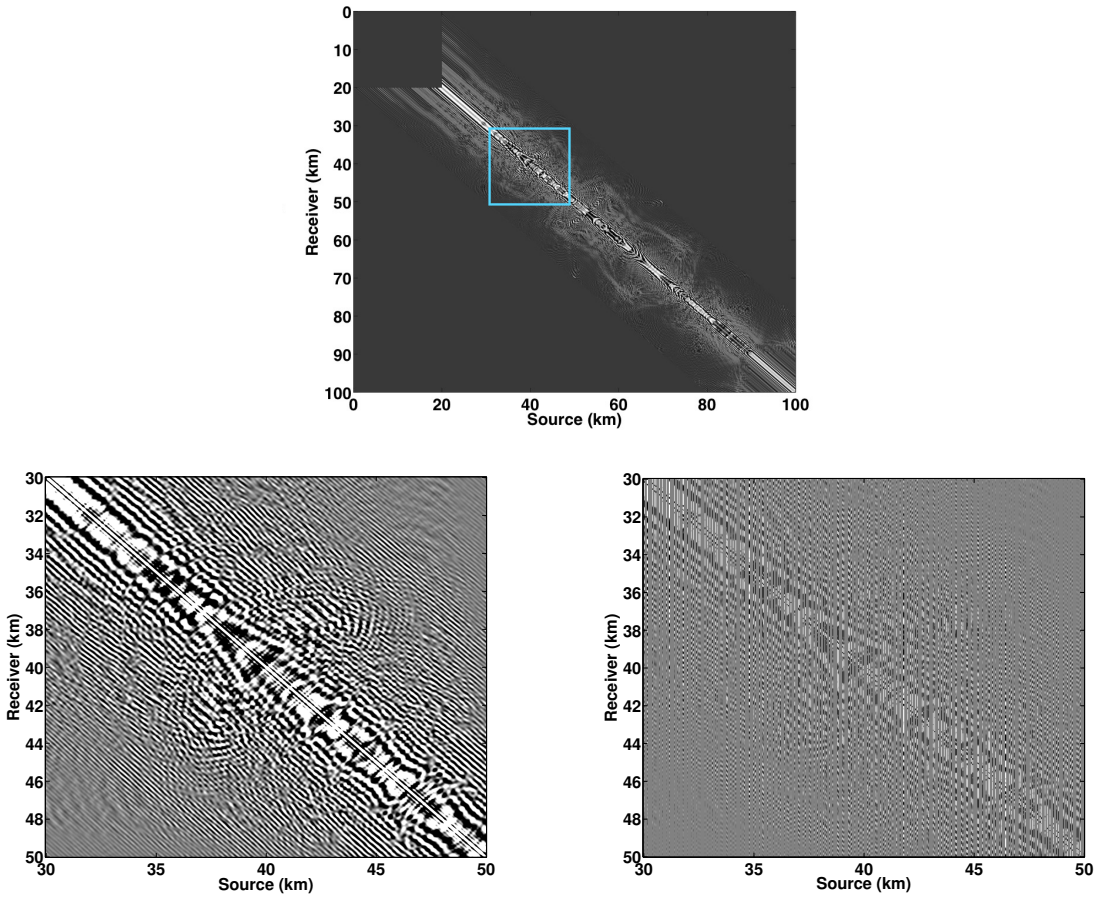


Figure 3.9: Gulf of Mexico data set. *Top*: Fully sampled monochromatic slice at 7 Hz. *Bottom left*: Fully sampled data (zoomed in the square block). *Bottom right*: 80% subsampled sources. For visualization purpose, the subsequent figures only show the interpolated result in the square block.

Synthetic fault model

In this setting, we remove 80% of the sources and display the results in Figure 3.13. For simplicity, we only perform rank-minimization based interpolation on this data set. In this case we set the rank parameter to be 30 and used the same for all frequencies. Even though the presence of faults make the geological environment complex, we are still able to successfully reconstruct the data volume using rank-minimization based techniques, which is also evident in the low-coherency of the data residual (Figure 3.13).

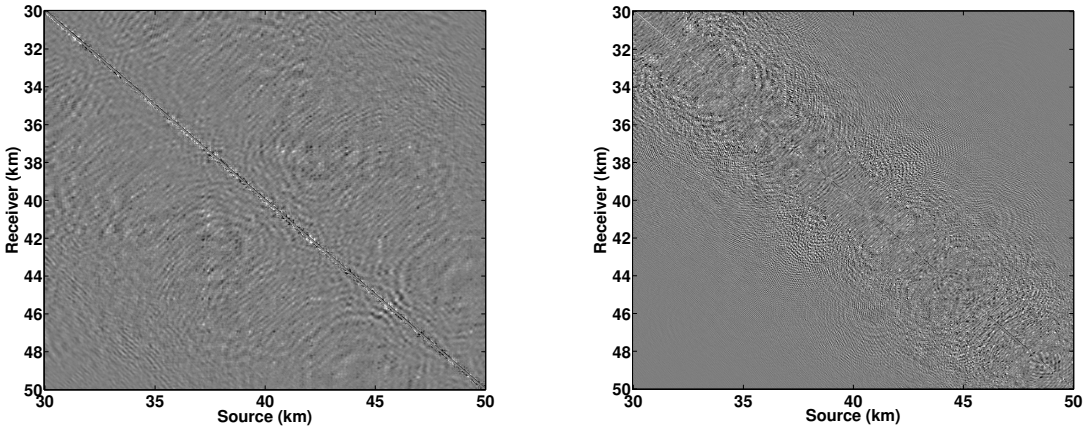


Figure 3.10: Reconstruction errors for frequency slice at 7Hz (left) and 20Hz (right) in case of 80% subsampled sources. Rank-minimization based recovery with a SNR of 14.2 dB and 11.0 dB respectively.

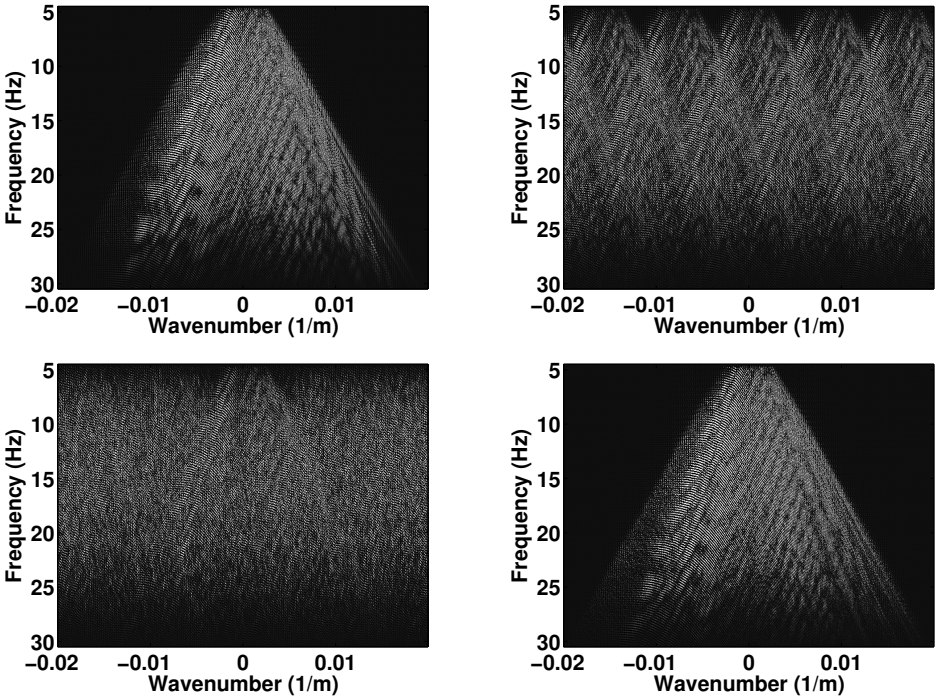


Figure 3.11: Frequency-wavenumber spectrum of the common receiver gather. *Top left*: Fully-sampled data. *Top right*: Periodic subsampled data with 80% missing sources. *Bottom left*: Uniform-random subsampled data with 80% missing sources. *Bottom Right*: Reconstruction of uniformly-random subsampled data using rank-minimization based techniques. While periodic subsampling creates aliasing, uniform-random subsampling turns the aliases in to incoherent noise across the spectrum.

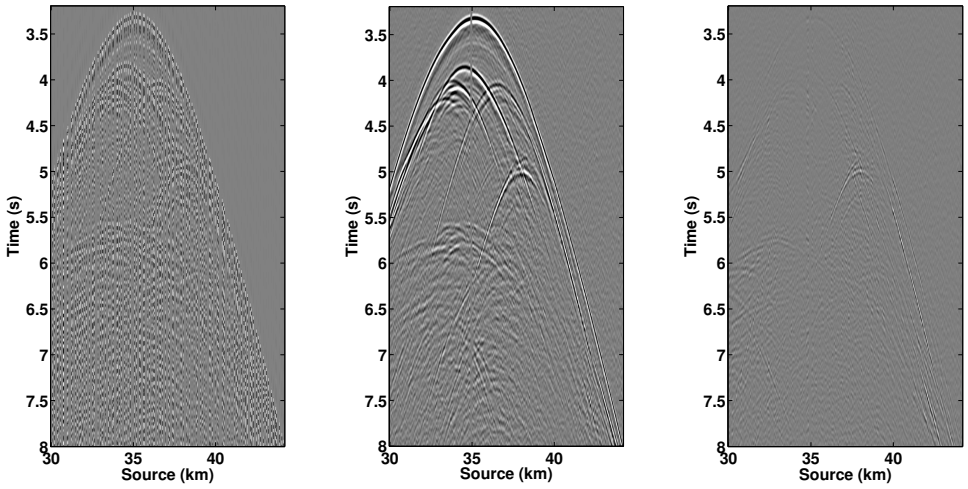


Figure 3.12: Gulf of Mexico data set, common receiver gather. *Left* : Uniformly-random subsampled data with 80% missing sources. *Middle* : Reconstruction results using rank-minimization based techniques (SNR = 7.8 dB). *Right* : Residual.

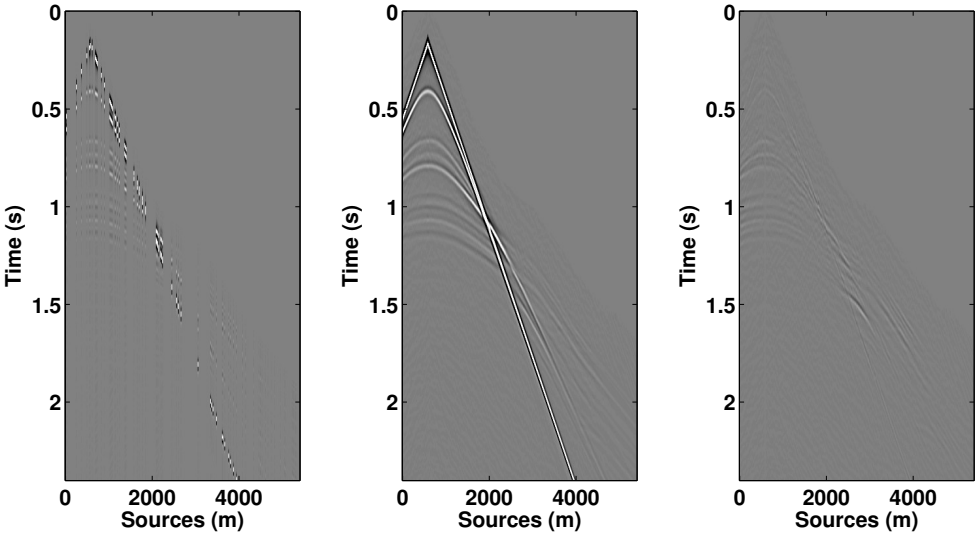


Figure 3.13: Missing-trace interpolation (80% sub-sampling) in case of geological structures with a fault. *Left*: 80% sub-sampled data. *Middle*: after interpolation (SNR = 23 dB). *Right*: difference.

3.6.2 3D seismic data

Single-layer reflector data

Before proceeding to a more realistically sized data set, we first test the performance of the SPG-LR matrix completion and the tensor completion method of Kreimer et al. [2013] on a small, synthetic data set generated from a simple, single-reflector model. We only use a frequency slice at 4 Hz. We normalize the volume to unit norm and randomly remove 50% of the sources from the data.

For the alternating direction method of multipliers (ADMM) tensor method, we complete the data volumes in the midpoint-offset domain, which is the same domain used in Kreimer et al. [2013]. In the context of our framework, we note that the midpoint-offset domain for recovering 3D frequency slices has the same recovery-enhancing properties as for recovering 2D frequency slices, as mentioned previously. Specifically, missing source sampling tends to *increase* the rank of the individual source and receiver matricizations in this domain, making completion via rank-minimization possible in midpoint-offset compared to source-receiver. In the original source-receiver domain, removing (x_{src}, x_{rec}) points from the tensor does *not* increase the singular values in the x_{src} and x_{rec} matricizations and hence the reconstruction quality will suffer. On the other hand, for the matrix completion case, the midpoint-offset conversion is a tight frame that acts on the left and right singular vectors of the matricized tensor $\mathbf{F}^{(x_{\text{src}}, x_{\text{rec}})}$ and thus does not affect the rank for this particular matricization. Also in this case, we consider the effects of windowing the input data on interpolation quality and speed. We let ADMM-w denote the ADMM method with a window size of w with an additional overlap of approximately 20%. In our experiments, we consider $w = 10$ (small windows), $w = 25$ (large windows), and $w = 50$ (no windowing).

In the ADMM method, the two parameters of note are λ, which control the relative penalty between data misfit and nuclear norm, and β, which controls the speed of convergence of the individual matrices $X^{(i)}$ to the tensor $\mathbf{F}$. The λ, β parameters proposed in Kreimer et al. [2013] do not appear to work for our problems, as using the stated parameters penalizes the nuclear norm terms too much compared to the data residual term, resulting in the solution tensor converging to $\mathbf{X} = 0$. Instead, we estimate the optimal λ, β parameters by cross validation, which involves removing 20% of the 50% known sources, creating a so-called "test set", and using the remaining data points as input data. We use various combinations of λ, β to solve Problem 3.3, using 50 iterations, and compare the SNR of the interpolant on the test set in order to determine the best parameters, i.e. we estimate the optimal λ, β without reference to the unknown entries of the tensor. Owing to the large computational costs of the "no window" case, we scan over the values of λ increasing exponentially and fix $\beta = 0.5$. For the windowed cases, we scan over exponentially increasing values of λ, β for a single window and use the estimated λ, β for interpolating the other windows. For the SPG-LR, we set our internal rank parameter to be 20 and allow the algorithm to run for 1000 iterations. As shown in Aravkin et al. [2014a], as long as the chosen rank is sufficiently

Method	SNR	Solve time	Parameter selection time	Total time
SPG-LR	25.5	0.9	N/A	0.9
ADMM - 50	20.8	87.4	320	407.4
ADMM - 25	16.8	4.4	16.4	20.8
ADMM - 10	10.9	0.1	0.33	0.43

Table 3.2: Single reflector data results. The recovery quality (in dB) and the computational time (in minutes) is reported for each method. The quality suffers significantly as the window size decreases due to the smaller redundancy of the input data, as discussed previously.

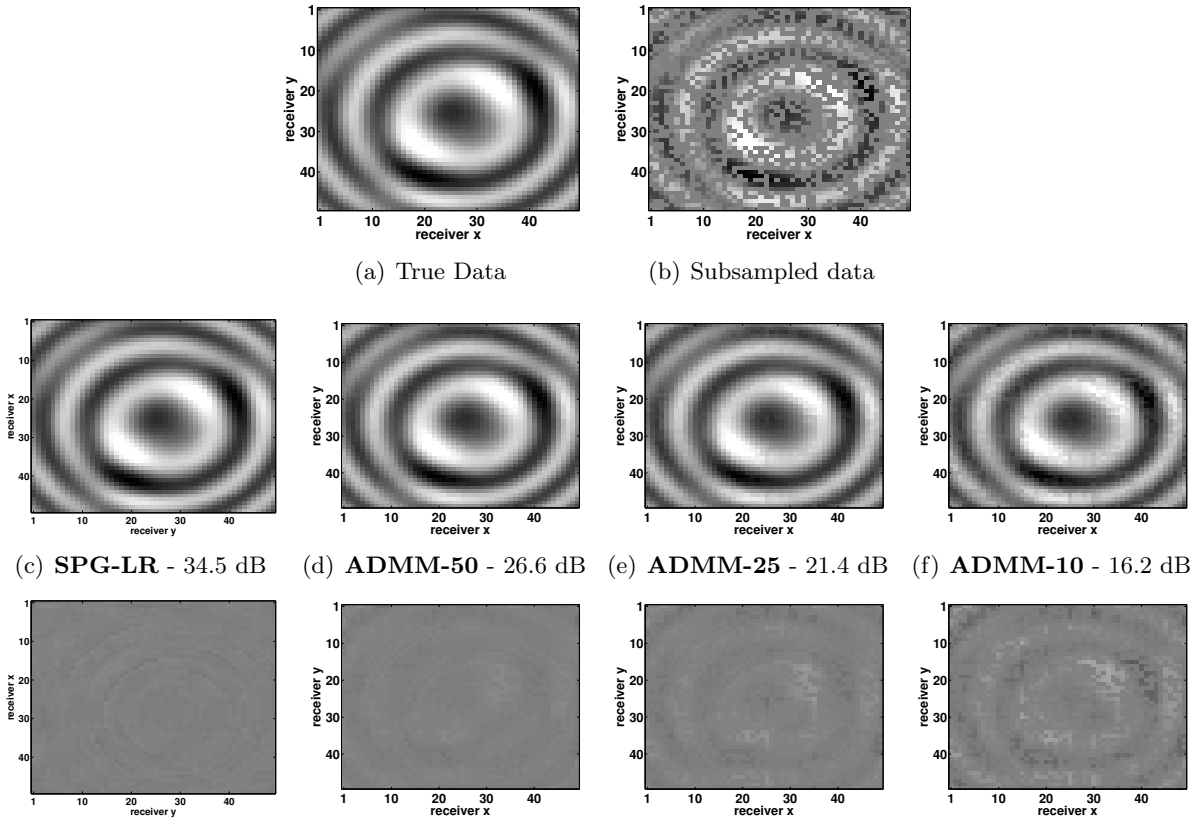
large, further increasing the rank parameter will not significantly change the results. We summarize our results in Table 3.2 and display the results in Figure 3.14.

Even disregarding the time spent selecting ideal parameters, SPG-LR matrix completion drastically outperforms the ADMM method on this small example. The tensor-based, per-dimension windowing approach also degrades the overall reconstruction quality, as the algorithm is unable to take advantage of the redundancy of the full data volume once the windows are sufficiently small. There is a very prominent tradeoff between recovery speed and reconstruction quality as the size of the windows become smaller, owing to the expensive nature of the ADMM approach itself for large data volumes and the inherent redundancy in the full data volume that makes interpolation possible which is decreased when windowing.

BG compass data

Owing to the smoothness of the data at lower frequencies, we uniformly downsample the individual frequency slices in the receiver coordinates without introducing aliasing. This reduces the overall computational complexity while simultaneously preserving the recovery quality. The 4.64Hz, 7.34Hz and 12.3Hz slices were downsampled to 101×101, 101×101 and 201×201 receiver grids, respectively. For these problems, the data was subsampled along the source coordinates by removing 25%, 50%, and 75% of the shots.

In order to apply matrix completion without windowing on the entire data set, the data was organized as a matrix using the low-rank promoting organization described previously. We used Jellyfish and SPG-LR implementations to complete the resulting incomplete matrix and compared these methods to the ADMM Tensor method and LMaFit, an alternating least-squares approach to matrix completion detailed in Wen et al. [2012]. LMaFit is a fast matrix completion solver that avoids using nuclear norm penalization but must be given an appropriate rank parameter in order to achieve reasonable results. We use the code available from the author's website. SPG-LR, ADMM, and LMaFit were run on a 2 quad-core 2.6GHz Intel processor with 16 GB memory and implicit multithreading via LAPACK libraries while Jellyfish was run on a dual Xeon X650 CPU



(a) True Data (b) Subsampled data

(c) **SPG-LR** - 34.5 dB (d) **ADMM-50** - 26.6 dB (e) **ADMM-25** - 21.4 dB (f) **ADMM-10** - 16.2 dB

Figure 3.14: ADMM data fit + recovery quality (SNR) for single reflector data, common receiver gather. Middle row: recovered slices, bottom row: residuals corresponding to each method in the middle row. Tensor-based windowing appears to visibly degrade the results, even with overlap.

(6 x 2 cores) with 24 GB of RAM with explicit multithreading. The hardware configurations of both of these environments are very similar, which results in SPG-LR and Jellyfish performing comparably.

For the Jellyfish experiments, the model parameter μ, which plays the same role as the λ parameter above, and the optimization parameters (initial step size and step decay) were selected by validation, which required 120 iterations of the optimization procedure for each (frequency, subsampling ratio) pair. The maximum rank value was set to the rank value used in the SPG-LR results. For the SPG-LR experiments, we interpolate a subsection of the data for various rank values and arrived at 120, 150 and 200 as the best rank parameters for each frequency. We perform the same validation techniques on the rank parameter k of LMaFit. In order to focus solely on comparing computational times, we omit reporting the parameter selection times for the ADMM method.

The results for 75% missing sources in Figure 3.15 demonstrate that, even in the low subsampling

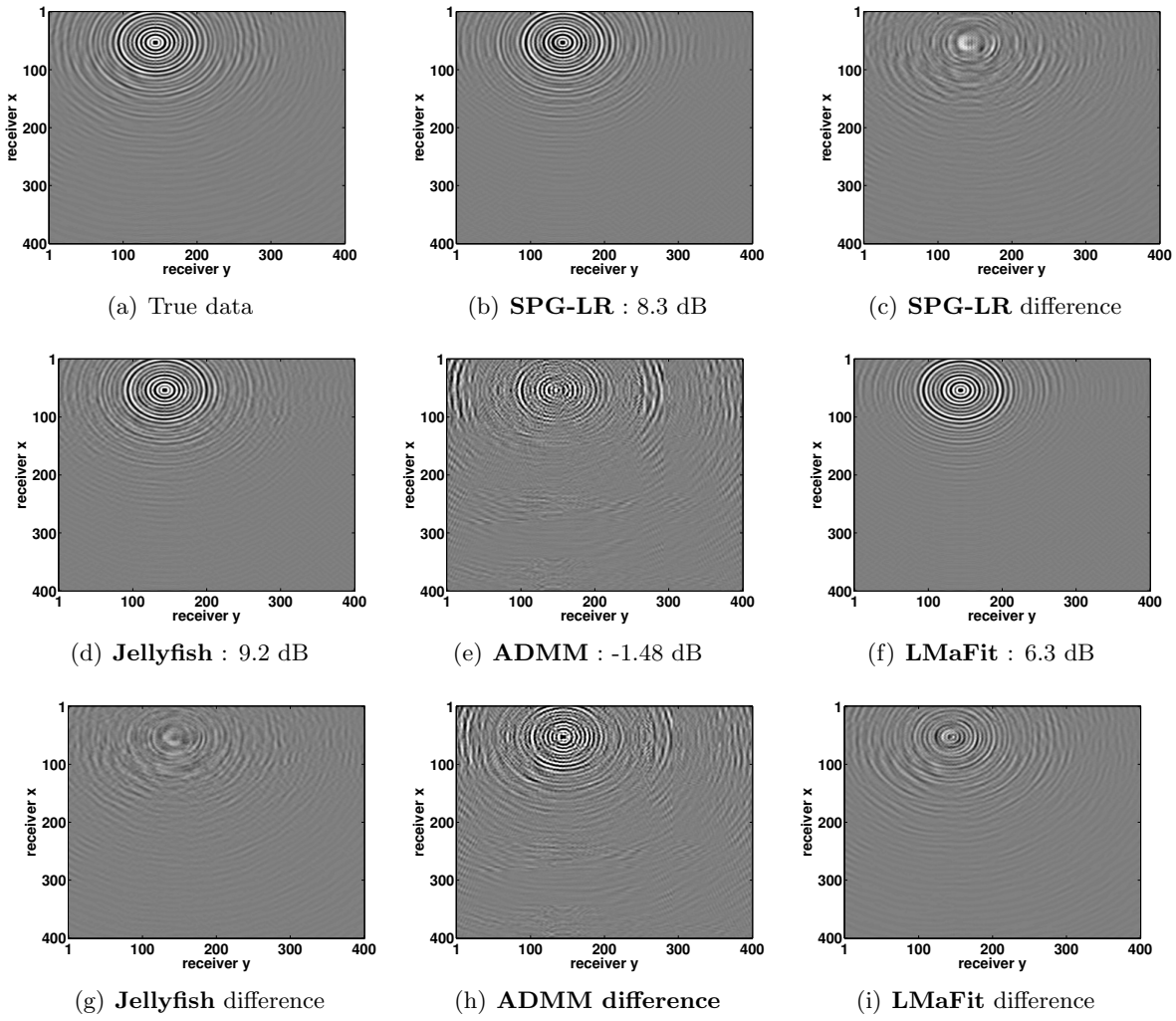
regime, matrix completion methods can successfully recover the missing shot data at these low frequencies. Table 3.3 gives an extensive summary of our results for different subsampling ratios and frequencies. The comparative results between Jellyfish and SPG-LR agree the with the theoretical results that establish the equivalence of BPDN_σ and QP_λ formulations. The runtime values include the parameter estimation procedure, which was carried out individually in each case. As we have seen previously, the ADMM approach does not perform well both in terms of computational time and in terms of recovery.

In our experiments, we noticed that successful parameter combinations work well for other problems too. Hence we can argue that in a real-world problem, once a parameter combination is selected, it can be used for different instances or it can be used as an initial point for a local parameter search.

Matrix completion with windowing

When windowing the data, we use the same matricizations of the data as discussed previously, but now split the volume in to nonoverlapping windows. We now use matrix completion on the resulting windows of data individually. We used Jellyfish for matrix completion on individual windows. Again, we use cross validation to select our parameters. We performed the experiments with two different window sizes. For the small window case, the matricization was partitioned into 4 segments along rows and columns, totalling 16 windows. For the large window case, the matricization was split into 16 segments along rows and columns, yielding 256 windows. This windowing is distinctly different from the windowing explored for the single-layer model, since here we are windowing the matricized form of the tensor, in the $(x_{\text{src}}, x_{\text{rec}})$ unfolding, as opposed to the per-dimension windowing in the previous section. The resulting windows created in this way contain much more sampled data than in the tensor-windowing case yet are still small enough in size to be processed efficiently.

The results in Figure 3.16 suggest that for this particular form of windowing, the matrix completion results are particularly degraded by only using small windows of data at a time. As mentioned previously, since we are relying on a high redundancy (with respect to the SVD) in the underlying and sampled data to perform matrix completion, we are reducing the overall redundancy of the input data by partitioning it. On the other hand, the real-world benefits of windowing in this context become apparent when the data cannot be fit into the memory at the cost of reconstruction quality. In this case, windowing allows us to partition the problem, offsetting the I/O cost that would result from memory paging. Based on these results, whenever possible, we strive to include as much data as possible in a given problem in order to recover the original matrix/tensor adequately.



(a) True data (b) **SPG-LR** : 8.3 dB (c) **SPG-LR** difference

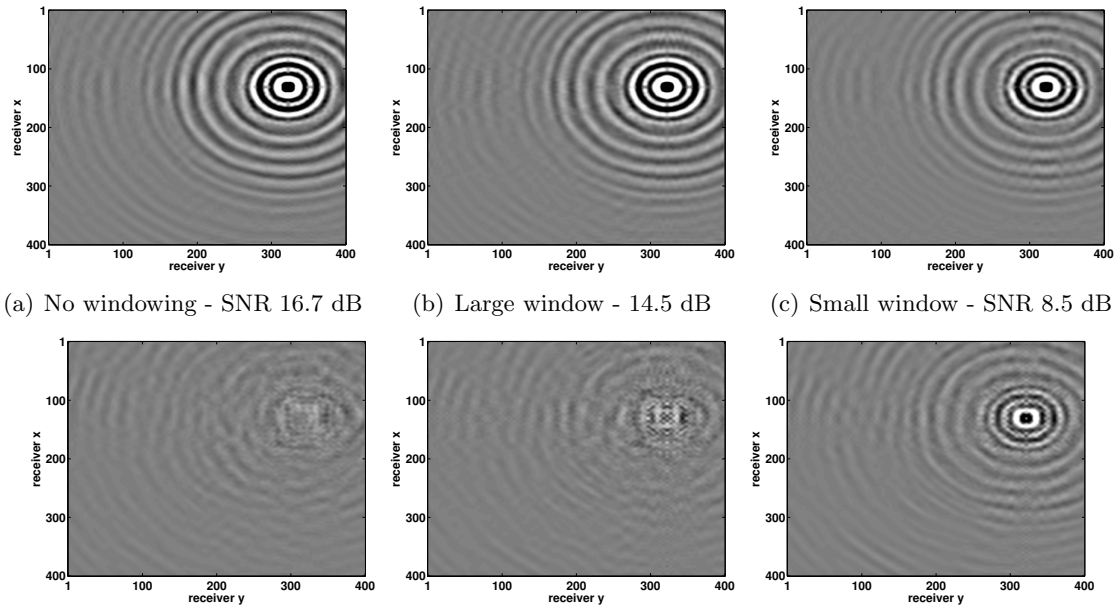
(d) **Jellyfish** : 9.2 dB (e) **ADMM** : -1.48 dB (f) **LMaFit** : 6.3 dB

(g) **Jellyfish** difference (h) **ADMM difference** (i) **LMaFit** difference

Figure 3.15: BG 5-D seismic data, 12.3 Hz, 75% missing sources. Middle row: interpolation results, bottom row: residuals.

3.7 Discussion

As the above results demonstrate, the $\mathbf{L}, \mathbf{R}$ matrix completion approach significantly outperforms the ADMM tensor-based approach due to the need to avoid the computation of SVDs as well as the minimal duplication of variables compared to the latter method. For the simple synthetic data, the ADMM method is able to achieve a similar recovery SNR to matrix completion, albeit at a much larger computational cost. For realistically sized data sets, the difference between the two methods can mean the difference between hours and days to produce an adequate result. In terms of the difference between SPG-LR and Jellyfish matrix completion, both return results that are similar in quality, which agrees with the fact that they are both based off of $\mathbf{L}, \mathbf{R}$ factorizations and the ranks used in these experiments are identical. Compared to these two methods, LMaFit converges



(a) No windowing - SNR 16.7 dB (b) Large window - 14.5 dB (c) Small window - SNR 8.5 dB

Figure 3.16: BG 5D seismic data, 4.68 Hz, Comparison of interpolation results with and without windowing using Jellyfish for 75% missing sources. Top row: interpolation results for differing window sizes, bottom row: residuals.

Frequency	Missing sources	SPG-LR		Jellyfish		ADMM		LmaFit	
		SNR	Time	SNR	Time	SNR	Time	SNR	Time
4.68 Hz	75%	15.9	84	**16.34**	**36**	0.86	1510	14.7	204
	50%	**20.75**	96	19.81	**82**	3.95	1510	17.5	91
	25%	**21.47**	114	19.64	124	9.17	1510	18.9	**66**
7.34 Hz	75%	11.2	84	**11.99**	**52**	0.39	1512	10.7	183
	50%	**15.2**	126	15.05	146	1.71	1512	14.1	**37**
	25%	**16.3**	138	15.31	195	4.66	1512	14.3	**21**
12.3 Hz	75%	7.3	324	**9.34**	**223**	0.06	2840	8.1	814
	50%	**12.6**	438	12.12	706	0.21	2840	11.1	**72**
	25%	**14.02**	450	12.90	1295	0.42	2840	11.3	**58**

Table 3.3: 3D seismic data results. The recovery quality (in dB) and the computational time (in minutes) is reported for each method.

much faster for regimes when there is more data available, while producing a lower quality result. When there is very little data available, as is typical in realistic seismic acquisition scenarios, the algorithm has issues converging. We note that, since it is a sparse linear algebra method, Jellyfish tends to outperform SPG-LR when the number of missing traces is high. This sparse linear algebra approach can conceivably be employed with the SPG-LR machinery. In these examples, we have not made any attempts to explicitly parallelize the SPG-LR or ADMM methods, instead relying on the efficient dense linear algebra routines used in Matlab, whereas Jellyfish is an inherently parallelized method.

Without automatic mechanisms for parameter selection as in SPG-LR, the Jellyfish, ADMM, and LMaFit algorithms rely on cross-validation techniques that involve solving many problem instances at different parameters. The inherently parallel nature of Jellyfish allows it to solve each problem instance very quickly and thus achieves very similar performance to SPG-LR. LMaFit has very fast convergence when there is sufficient data, but slows down significantly in scenarios with very little data. The ADMM method, on the other hand, scales much more poorly for large data volumes and spends much more time on parameter selection than the other methods. However, in practice, we can assume that across frequency slices, say, optimally chosen parameters for one frequency slice will likely work well for neighbouring frequency slices and thus the parameter selection time can be amortized over the whole volume.

In our experiments, aside from the simple 3D layer model and the Nelson dataset, the geological models used were not low rank. That is to say, the models had complex geology and were not simply horizontally layered media. Instead, through the use of these low rank techniques, we are exploiting the low rank structure of the data volumes achieved from the data acquisition process, not merely any low rank structure present in the models themselves. As the temporal frequency increases, the inherent rank of the resulting frequency slices increases, which makes low rank interpolation more challenging. Despite this observation, we still achieve reasonable results for higher frequencies using our methods.

As predicted by our theoretical considerations, the choice of windowing in this case has a negative effect on the generated results in the situation where the earth model permits a low-rank representation that is reflected in the midpoint-offset domain. In case of earth models that are not inherently low-rank, such as those with salt bodies, we can still recover the low-frequency slices as shown by the examples without performing the windowing on the data sets. As a general rule of thumb, we advise to incorporate as much of the input data is possible in to a given matrix-completion problem but clearly there is a tradeoff between the size of the data windows, the amount of memory available to process such volumes, and the inherent complexity of the model. Additionally, one should avoid methods that needlessly create extraneous copies of the data when working with large scale volumes.

Here we have also demonstrated the importance of theoretical components for signal recovery using matrix and tensor completion methods. By ignoring these principles of matrix completion, a

practitioner can unintentionally find herself in a disadvantageous scenario and produce sub-optimal results without a guiding theory to remedy the situation. However, by choosing an appropriate transform domain in which to complete the matrix or tensor, we can successfully employ this rank-minimizing machinery to interpolate a signal with missing traces in a computationally efficient manner.

From a practitioner's point of view, the purpose of this interpolation machinery is to remove the acquisition footprint from missing-trace data that is used in further downstream seismic processes such as migration and full waveform inversion. These techniques can help mitigate the lack of data coverage in certain areas that would otherwise have created artifacts or non-physical regions in a seismic image.

3.8 Conclusion

Building upon existing knowledge of compressive sensing as a successful signal recovery paradigm, this work has outlined the necessary components of using matrix and tensor completion methods for interpolating large-scale seismic data volumes. As we have demonstrated numerically, without the necessary components of a low-rank domain, a rank-increasing sampling scheme, and a rank-minimizing optimization scheme, matrix completion-based techniques cannot successfully recover subsampled seismic signals. Once all of these ingredients are in place, however, we can use existing convex solvers to recover the fully sampled data volume. Since such solvers invariably involve computing singular-value decomposition of large matrices, we have presented two alternative factorized-based formulations that scale much more efficiently than their strictly convex counterparts when the data volumes are large. We have shown that our factorization-based matrix completion approach is very competitive compared to existing curvelet-based methods for 2D seismic data and alternating direction method of multipliers tensor-based methods for 3D seismic data.

From a practical point of view, this theoretical framework is exceedingly flexible. We have shown the effectiveness of midpoint-offset organization for 2D data and $(x_{\text{source}}, x_{\text{receiver}})$ matrix organization for 3D data for promoting low-rank structure in the data volumes but it is conceivable that other seismic data organizations could also be useful in this regard, e.g., midpoint-offset-azimuth. Our optimization framework also allows us to operate on both large-scale data without having to select a large number of parameters and we do not need to recourse to using small windows of data, which may degrade the recovery results. In the seismic context, reusing the interpolated results from lower frequencies as a warm-start for interpolating data at higher frequencies can further reduce the overall computational costs. The proposed approach to matrix completion and the Jellyfish method are very promising for large scale data sets and can conceivably be applied to interpolate wide azimuth data sets as well.

Chapter 4

Source separation for simultaneous towed-streamer marine acquisition—a compressed sensing approach

4.1 Summary

Apart from performing seismic data interpolation as shown in previous chapters, rank-minimization based techniques have shown great potential in dealing with seismic data acquired in simultaneous fashion. In marine environment, simultaneous acquisition is an economic way to sample seismic data and speedup acquisition, wherein single and/or multiple source vessels fire sources at near-simultaneous or slightly random times, resulting in overlapping shot records. The current paradigm for simultaneous towed-streamer marine acquisition incorporates "low-variability" in source firing times—i.e., $0 \leq 1$ or 2 seconds, since both the sources and receivers are moving. This results in low degree of randomness in simultaneous data, which is challenging to separate (into its constituent sources) using compressed sensing based separation techniques since randomization is the key to successful recovery via compressed sensing. In this chapter, we address the challenge of source separation for simultaneous towed-streamer acquisitions via two compressed sensing based approaches—i.e., sparsity-promotion and rank-minimization. We illustrate the performance of both the sparsity-promotion and rank-minimization based techniques by simulating two simultaneous towed-streamer acquisition scenarios—i.e., over/under and simultaneous long offset. A field data example from the Gulf of Suez for the over/under acquisition scenario is also included. We observe that the proposed approaches give good and comparable recovery qualities of the separated sources, but the rank-minimization technique outperforms the sparsity-promoting technique in terms of the computational time and memory. We also compare these two techniques with the NMO-based

A version of this chapter has been published in Geophysics, 2016, vol 80, pages WD73-WD88.

median filtering type approach.

4.2 Introduction

The benefits of simultaneous source marine acquisition are manifold—it allows the acquisition of improved-quality seismic data at standard (conventional) acquisition turnaround, or a reduced turnaround time while maintaining similar quality, or a combination of both advantages. In simultaneous marine acquisition, a single or multiple source vessels fire sources at near-simultaneous or slightly random times resulting in overlapping shot records [de Kok and Gillespie, 2002, Beasley, 2008, Berkhout, 2008b, Hampson et al., 2008, Moldoveanu and Quigley, 2011, Abma et al., 2013], as opposed to non-overlapping shot records in conventional marine acquisition. A variety of simultaneous source survey designs have been proposed for towed-streamer and ocean bottom acquisitions, where small-to-large random time delays between multiple sources have been used [Beasley, 2008, Moldoveanu and Fealy, 2010, Mansour et al., 2012c, Abma et al., 2013, Wason and Herrmann, 2013b, Mosher et al., 2014].

An instance of low-variability in source firing times—e.g., $0 \leq 1$ (or 2) second, is the over/under (or multi-level) source acquisition [Hill et al., 2006, Moldoveanu et al., 2007, Lansley et al., 2007, Long, 2009, Hegna and Parkes, 2012, Torben Hoy, 2013]. The benefits of acquiring and processing over/under data are clear, the recorded bandwidth is extended at both low and high ends of the spectrum since the depths of the sources produce complementary ghost functions, avoiding deep notches in the spectrum. The over/under acquisition allows separation of the up- and down-going wavefields at the source (or receiver) using a vertical pair of sources (or receivers) to determine wave direction. Simultaneous long offset acquisition (SLO) is another variation of simultaneous towed-streamer acquisition, where an extra source vessel is deployed, sailing one spread-length ahead of the main seismic vessel [Long et al., 2013]. The SLO technique is better in comparison to conventional acquisition since it provides longer coverage in offsets, less equipment downtime (doubling the vessel count inherently reduces the streamer length by half), easier maneuvering, and shorter line turns.

Simultaneous acquisition (e.g., over/under and SLO) results in seismic interferences or source crosstalk that degrades quality of the migrated images. Therefore, an effective (simultaneous) source separation technique is required, which aims to recover unblended interference-free data—as acquired during conventional acquisition—from simultaneous data. The challenge of source separation (or deblending) has been addressed by many researchers [Stefani et al., 2007, Moore et al., 2008, Akerberg et al., 2008, Huo et al., 2009], wherein the key observation has been that as long as the sources are fired at suitably randomly dithered times, the resulting interferences (or source crosstalk) will appear noise-like in specific gather domains such as common-offset and common-receiver, turning the separation problem into a (random) noise removal procedure. Inversion-type algorithms [Moore, 2010, Abma et al., 2010, Mahdad et al., 2011, Doulgeris et al., 2012, Baardman

and van Borselen, 2013] take advantage of sparse representations of coherent seismic signals. Wason and Herrmann [2013a]; Wason and Herrmann [2013b] proposed an alternate sampling strategy for simultaneous acquisition (*time-jittered marine*) that leverages ideas from compressed sensing (CS), addressing the deblending problem through a combination of tailored (blended) acquisition design and sparsity-promoting recovery via convex optimization using one-norm constraints. This represents a scenario of high-variability in source firing times—e.g., > 1 second, resulting in irregular shot locations.

One of the source separation techniques is the normal moveout based median filtering, where the key idea is as follows: *i*) transform the blended data into the midpoint-offset domain, *ii*) perform semblance analysis on common-midpoint gathers to pick the normal moveout (NMO) velocities followed by NMO corrections, *iii*) perform median filtering along the offset directions and then apply inverse NMO corrections. One of the major assumptions in the described workflow is that the seismic events become flat after NMO corrections, however, this can be challenging when the geology is complex and/or with the presence of noise in the data. Therefore, the above process along with the velocity analysis is repeated a couple of times to get a good velocity model to eventually separate simultaneous data.

Recently, rank-minimization based techniques have been used for source separation by Maraschini et al. [2012] and Cheng and Sacchi [2013]. The general idea is to exploit the low-rank structure of seismic data when it is organized in a matrix. Low-rank structure refers to the small number of nonzero singular values, or quickly decaying singular values. Maraschini et al. [2012] followed the rank-minimization based approach proposed by Oropeza and Sacchi [2011], who identified that seismic temporal frequency slices organized into a block Hankel matrix, in ideal conditions, is a matrix of rank k, where k is the number of different plane waves in the window of analysis. Oropeza and Sacchi [2011] showed that additive random noise increase the rank of the block Hankel matrix and presented an iterative algorithm that resembles seismic data reconstruction with the method of projection onto convex sets, where they use a low-rank approximation of the Hankel matrix via the randomized singular value decomposition [Liberty et al., 2007, Halko et al., 2011a] to interpolate seismic temporal frequency slices. While this technique may be effective the approach requires embedding the data into an even larger space where each dimension of size n is mapped to a matrix of size $n \times n$. Consequently, these approaches are applied on small data windows, where one has to choose the size of these windows. Although mathematically desirable due to the seismic signal being stationary in sufficiently small windows, Kumar et al. [2015a] showed that the act of windowing from a matrix-rank point of view degrades the quality of reconstruction in the case of missing-trace interpolation. Choosing window sizes apriori is also a difficult task, as it is not altogether obvious how to ensure that the resulting sub-volume is approximately a plane-wave.

4.2.1 Motivation

The success of CS hinges on randomization of the acquisition, as presented in our previous work on simultaneous source acquisition [Mansour et al., 2012c, Wason and Herrmann, 2013b], which represents a case of high-variability in source firing times—e.g., within a range of 1-20 seconds, resulting in overlapping shot records that lie on irregular spatial grids. Consequently, this made our method applicable to marine acquisition with ocean bottom cables/nodes. Successful separation of simultaneous data by sparse inversion via one-norm minimization, in this high-variability scenario, motivated us to analyze the performance of our separation algorithm for the low-variability, simultaneous towed-streamer acquisitions. In this chapter, we address the challenge of source separation for two types of simultaneous towed-streamer marine acquisition—over/under and simultaneous long offset. We also compare the sparsity-promoting separation technique with separation via rank-minimization based technique, since the latter is relatively computationally faster and memory efficient, as shown by Kumar et al. [2015a] for missing-trace interpolation.

4.2.2 Contributions

Our contributions in this work are the following: first, we propose a practical framework for source separation based upon the compressed sensing (CS) theory, where we outline the necessary conditions for separating the simultaneous towed-streamer data using sparsity-promoting and rank-minimization techniques. Second, we show that source separation using the rank-minimization based framework includes a "transform domain" where we exploit the low-rank structure of seismic data. We further establish that in simultaneous towed-streamer acquisition each monochromatic frequency slice of the fully sampled blended data matrix with periodic firing times has low-rank structure in the proposed transform domain. However, uniformly random firing-time delays increase the rank of the resulting frequency slice in this transform domain, which is a necessary condition for successful recovery via rank-minimization based techniques.

Third, we show that seismic frequency slices in the proposed transform domain exhibit low-rank structure at low frequencies, but not at high frequencies. Therefore, in order to exploit the low-rank structure at higher frequencies we adopt the Hierarchical Semi-Separable matrix representation (HSS) method proposed by Chandrasekaran et al. [2006] to represent frequency slices. Finally, we combine the (SVD-free) matrix factorization approach recently developed by Lee et al. [2010a] with the Pareto curve approach proposed by Berg and Friedlander [2008]. This renders the framework suitable for large-scale seismic data since it avoids the computation of the singular value decomposition (SVD), a necessary step in traditional rank-minimization based methods, which is prohibitively expensive for large matrices.

We simulate two simultaneous towed-streamer acquisitions—over/under and simultaneous long offset, and also use a field data example for over/under acquisition. We compare the recovery in terms of the separation quality, computational time and memory usage. In addition, we also make

comparisons with the NMO-based median filtering type technique proposed by Chen et al. [2014].

4.3 Theory

Compressed sensing is a signal processing technique that allows a signal to be sampled at sub-Nyquist rate and offers three fundamental principles for successful reconstruction of the original signal from relatively few measurements. The first principle utilizes the prior knowledge that the underlying signal of interest is sparse or compressible in some transform domain—i.e., if only a small number k of the transform coefficients are nonzero or if the signal can be well approximated by the k largest-in-magnitude transform coefficients. The second principle is based upon a sampling scheme that breaks the underlying structure—i.e., decreases the sparsity of the original signal in the transform domain. Once the above two principles hold, a sparsity-promoting optimization problem can be solved in order to recover the fully sampled signal. It is well known that seismic data admit sparse representations by curvelets that capture "wavefront sets" efficiently (see e.g., Smith [1998], Candès and Demanet [2005], Hennenfent and Herrmann [2006a] and the references therein).

For high resolution data represented by the N-dimensional vector $\mathbf{f}_0 \in \mathbb{R}^N$, which admits a sparse representation $\mathbf{x}_0 \in \mathbb{C}^P$ in some transform domain characterized by the operator $\mathbf{S} \in \mathbb{C}^{P\times N}$ with $P \geq N$, the sparse recovery problem involves solving an underdetermined system of equations:

$$\mathbf{b} = \mathbf{A}\mathbf{x}_0, \tag{4.1}$$

where $\mathbf{b} \in \mathbb{C}^n$, $n \ll N \leq P$, represents the compressively sampled data of n measurements, and $\mathbf{A} \in \mathbb{C}^{n\times P}$ represents the measurement matrix. We denote by $\mathbf{x}_0$ a sparse synthesis coefficient vector of $\mathbf{f}_0$. When $\mathbf{x}_0$ is strictly sparse—i.e., only $k < n$ nonzero entries in $\mathbf{x}_0$, sparsity-promoting recovery can be achieved by solving the ℓ_0 minimization problem, which is a combinatorial problem and quickly becomes intractable as the dimension increases. Instead, the basis pursuit denoise (BPDN$_\epsilon$) convex optimization problem:

$$\underset{\mathbf{x}\in\mathbb{C}^P}{\text{minimize}} \quad \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{b}-\mathbf{A}\mathbf{x}\|_2 \leq \epsilon, \tag{BPDN$_\epsilon$}$$

can be used to recover $\tilde{\mathbf{x}}$, which is an estimate of $\mathbf{x}_0$. Here, ϵ represents the error-bound in the least-squares misfit and the ℓ_1 norm $\|\mathbf{x}\|_1$ is the sum of absolute values of the elements of a vector $\mathbf{x}$. The matrix $\mathbf{A}$ can be composed of the product of an $n \times N$ sampling (or acquisition) matrix $\mathbf{M}$ and the sparsifying operator $\mathbf{S}$ such that $\mathbf{A} := \mathbf{M}\mathbf{S}^H$, here ^{H} denotes the Hermitian transpose. Consequently, the measurements are given by $\mathbf{b} = \mathbf{A}\mathbf{x}_0 = \mathbf{M}\mathbf{f}_0$. A seismic line with N_s sources, N_r receivers, and N_t time samples can be reshaped into an N dimensional vector $\mathbf{f}$, where $N = N_s \times N_r \times N_t$. For simultaneous towed-streamer acquisition, given two unblended data vectors

$\mathbf{x}_1$ and $\mathbf{x}_2$ and (blended) measurements $\mathbf{b}$, we can redefine Equation 4.1 as

$$\overbrace{\begin{bmatrix}\mathbf{M}\mathbf{T_1}\mathbf{S}^H & \mathbf{M}\mathbf{T_2}\mathbf{S}^H\end{bmatrix}}^{\mathbf{A}} \overbrace{\begin{bmatrix}\mathbf{x}_1 \\ \mathbf{x}_2\end{bmatrix}}^{\mathbf{x}} = \mathbf{b}, \tag{4.2}$$

where $\mathbf{T}_1$ and $\mathbf{T}_2$ are defined as the firing-time delay operators which apply uniformly random time delays to the first and second source, respectively. Note that accurate knowledge of the firing times is essential for successful recovery by the proposed source separation techniques. We wish to recover a sparse approximation $\tilde{\mathbf{f}}$ of the discretized wavefield $\mathbf{f}$ (corresponding to each source) from the measurements $\mathbf{b}$. This is done by solving the BPDN$_\epsilon$ sparsity-promoting program, using the SPGℓ_1 solver [see Berg and Friedlander, 2008, Hennenfent et al., 2008, for details], yielding $\tilde{\mathbf{f}} = \mathbf{S}^H\tilde{\mathbf{x}}$ for each source.

Sparsity is not the only structure seismic data exhibits where three- or five-dimensional seismic data is organized as a vector. High-dimensional seismic data volumes can also be represented as matrices or tensors, where the low-rank structure of seismic data can be exploited [Trickett and Burroughs, 2009, Oropeza and Sacchi, 2011, Kreimer and Sacchi, 2012a, Silva and Herrmann, 2013b, Aravkin et al., 2014b]. This low-rank property of seismic data leads to the notion of matrix completion theory which offers a reconstruction strategy for an unknown matrix $\mathbf{X}$ from its known subsets of entries [Candès and Recht, 2009, Recht et al., 2010a]. The success of matrix completion framework hinges on the fact that regularly sampled target dataset should exhibit a low-rank structure in the rank-revealing "transform domain" while subsampling should destroy the low-rank structure of seismic data in the transform domain.

4.3.1 Rank-revealing "transform domain"

Following the same analogy of CS, the main challenge in applying matrix completion techniques to the source separation problem is to find a "transform domain" wherein: *i*) fully sampled conventional (or unblended) seismic data have low-rank structure—i.e., quickly decaying singular values; *ii*) blended seismic data have high-rank structure—i.e., slowly decaying singular values. When these properties hold, rank-minimization techniques (used in matrix completion) can be used to recover the deblended signal. Kumar et al. [2013a] showed that the frequency slices of unblended seismic data do not exhibit low-rank structure in the source-receiver (s-r) domain since strong wavefronts extend diagonally across the s-r plane. However, transforming the data into the midpoint-offset (m-h) domain results in a vertical alignment of the wavefronts, thereby reducing the rank of the

frequency slice matrix. The midpoint-offset domain is a coordinate transformation defined as:

$$x_{midpoint} = \frac{1}{2}(x_{source} + x_{receiver}),$$
$$x_{offset} = \frac{1}{2}(x_{source} - x_{receiver}).$$

These observations motivate us to exploit the low-rank structure of seismic data in the midpoint-offset domain for simultaneous towed-streamer acquisition. Figures 4.1a and 4.1c show a monochromatic frequency slice (at 5 Hz) for simultaneous acquisition with periodic firing times in the source-receiver (s-r) and midpoint-offset (m-h) domains, while Figures 4.1b and 4.1d show the same for simultaneous acquisition with random firing-time delays. Note that we use the source-receiver reciprocity to convert each monochromatic frequency slice of the towed-streamer acquisition to split-spread type acquisition, which is required by our current implementation of rank-minimization based techniques for 2-D seismic acquisition. For 3-D seismic data acquisition, however, where seismic data exhibit 5-D structure, we can follow the strategy proposed by Kumar et al. [2015a], where different matricization is used as a transformation domain to exploit the low-rank structure of seismic data. Therefore, in 3-D seismic data acquisition we do not have to work in the midpoint-offset domain which removes the requirement of source-receiver reciprocity.

As illustrated in Figure 4.1, simultaneously acquired data with periodic firing times preserves continuity of the waveforms in the s-r and m-h domains, which inherently do not change the rank of blended data compared to unblended data. Introducing random time delays destroys continuity of the waveforms in the s-r and m-h domains, thus increasing the rank of the blended data matrix drastically, which is a necessary condition for rank-minimization based algorithms to work effectively. To illustrate this behaviour, we plot the decay of the singular values of a 5 Hz monochromatic frequency slice extracted from the periodically and randomized simultaneous acquisition in the s-r and m-h domains, respectively in Figure 4.2a. Note that uniformly random firing-time delays do not noticeably change the decay of the singular values in the source-receiver (s-r) domain, as expected, but significantly slow down the decay rate in the m-h domain.

Similar trends are observed for a monochromatic frequency slice at 40 Hz in Figure 4.2b. Following the same analogy, Figures 4.2c and 4.2d show how randomization in acquisition destroys the sparse structure of seismic data in the source-channel (or source-offset) domain—i.e., slow decay of the curvelet coefficients, hence, favouring recovery via sparsity-promotion in this domain. Similarly, for simultaneous long offset acquisition, we exploit the low-rank structure of seismic data in the m-h domain, and the sparse structure in the source-channel domain.

Seismic frequency slices exhibit low-rank structure in the m-h domain at low frequencies, but the same is not true for data at high frequencies. This is because in the low-frequency slices, the vertical alignment of the wavefronts can be accurately approximated by a low-rank representation. On the other hand, high-frequency slices include a variety of wave oscillations that increase the rank,

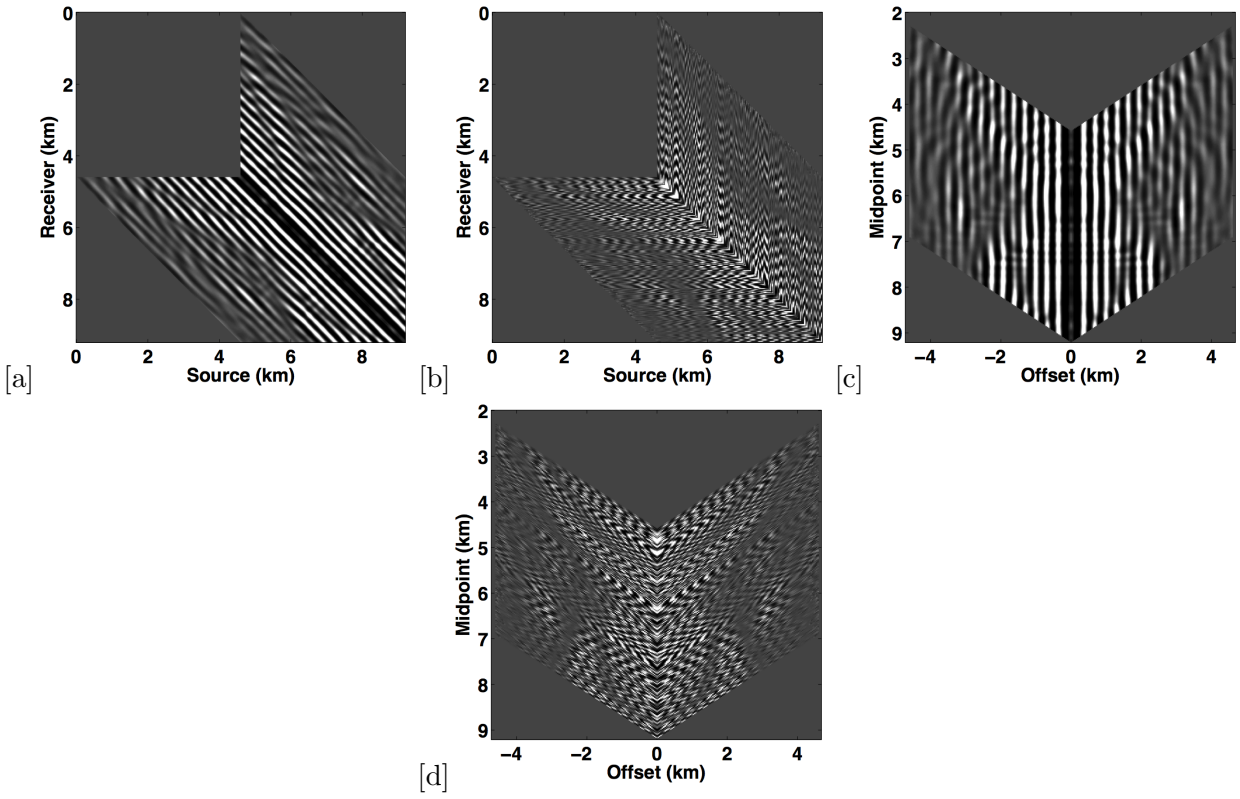


Figure 4.1: Monochromatic frequency slice at 5 Hz in the source-receiver (s-r) and midpoint-offset (m-h) domain for blended data (a,c) with periodic firing times and (b,d) with uniformly random firing times for both sources.

even though the energy remains focused around the diagonal [Kumar et al., 2013a]. To illustrate this phenomenon, we plot a monochromatic frequency slice at 40 Hz in the s-r domain and the m-h domain for over/under acquisition in Figure 4.3. When analyzing the decay of the singular values for high-frequency slices in the s-r domain and the m-h domain (Figure 4.2b), we observe that the singular value decay is slower for the high-frequency slice than for the low-frequency slice. Therefore, rank-minimization in the high-frequency range requires extended formulations that incorporate the low-rank structure.

To exploit the low-rank structure of high-frequency data, we rely on the Hierarchical Semi-Separable matrix representation (HSS) method proposed by Chandrasekaran et al. [2006] to represent frequency slices. The key idea in the HSS representation is that certain full-rank matrices, e.g., matrices that are diagonally dominant with energy decaying along the off-diagonals, can be represented by a collection of low-rank sub-matrices. Kumar et al. [2013a] showed the possibility of finding accurate low-rank approximations of sub-matrices of the high-frequency slices by partitioning the data into the HSS structure for missing-trace interpolation. Jumah and Herrmann [2014] showed that HSS representations can be used to reduce the storage and computational cost

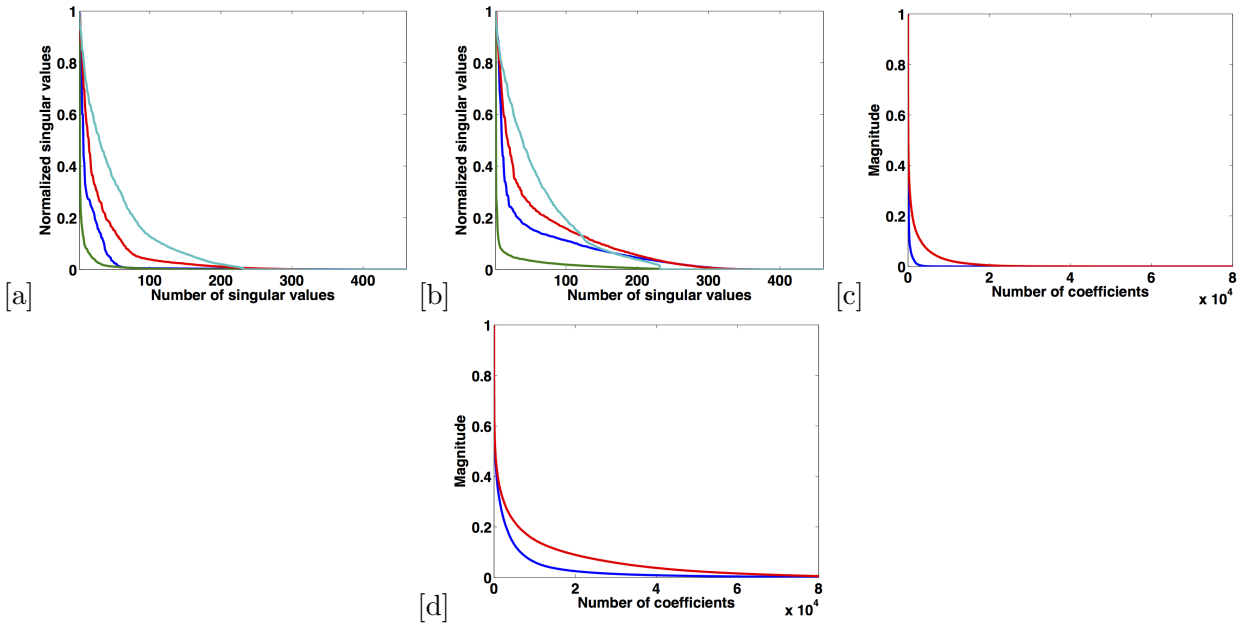


Figure 4.2: Decay of singular values for a frequency slice at (a) 5 Hz and (b) 40 Hz of blended data. Source-receiver domain: blue—periodic, red—random delays. Midpoint-offset domain: green—periodic, cyan—random delays. Corresponding decay of the normalized curvelet coefficients for a frequency slice at (c) 5 Hz and (d) 40 Hz of blended data, in the source-channel domain.

for the estimation of primaries by sparse inversions. They combined the HSS representation with the randomized SVD proposed by Halko et al. [2011c] to accelerate matrix-vector multiplications that are required for sparse inversion.

4.3.2 Hierarchical semi-separable matrix representation (HSS)

The HSS structure first partitions a matrix into diagonal and off-diagonal sub-matrices. The same partitioning structure is then applied recursively to the diagonal sub-matrices only. To illustrate the HSS partitioning, we consider a 2-D monochromatic high-frequency data matrix at 40 Hz in the s-r domain. We show the first-level of partitioning in Figure 4.4a and the second-level partitioning in Figure 4.4b in their corresponding source-receiver domains. Figures 4.5a and 4.5b display the first-level off-diagonal sub-blocks, Figure 4.5c is the diagonal sub-block, and the corresponding decay of the singular values is displayed in Figure 4.6. We can clearly see that the off-diagonal sub-matrices have low-rank structure, while the diagonal sub-matrices have higher rank. Further partitioning of the diagonal sub-blocks (Figure 4.4b) allows us to find better low-rank approximations. The same argument holds for the simultaneous long offset acquisition. Therefore, for low-variability acquisition scenarios, each frequency slice is first partitioned using HSS and then deblended in its respective m-h domain, as showed for missing-trace interpolation by Kumar et al. [2013a].

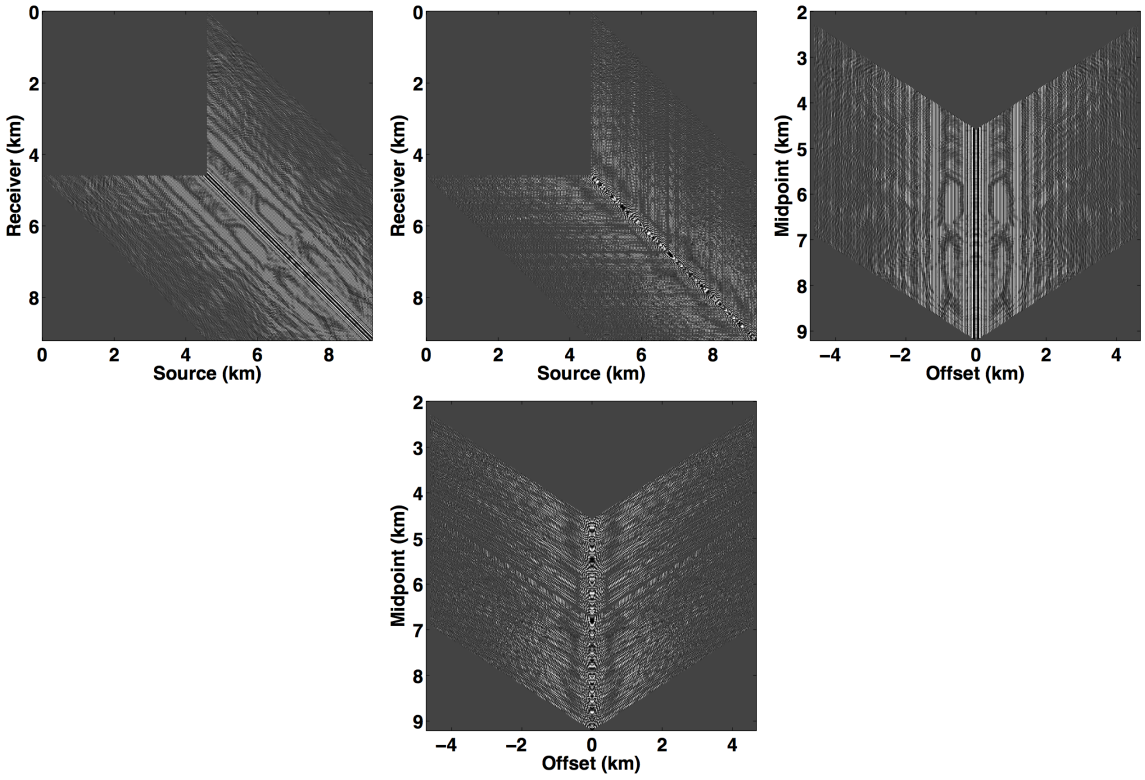
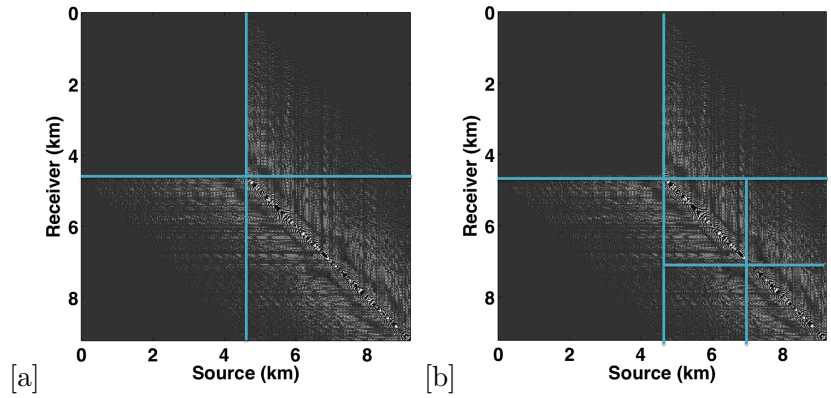


Figure 4.3: Monochromatic frequency slice at 40 Hz in the s-r and m-h domain for blended data (a,c) with periodic firing times and (b,d) with uniformly random firing times for both sources.



[a] [b]

Figure 4.4: HSS partitioning of a high-frequency slice at 40 Hz in the s-r domain: (a) first-level, (b) second-level, for randomized blended acquisition.

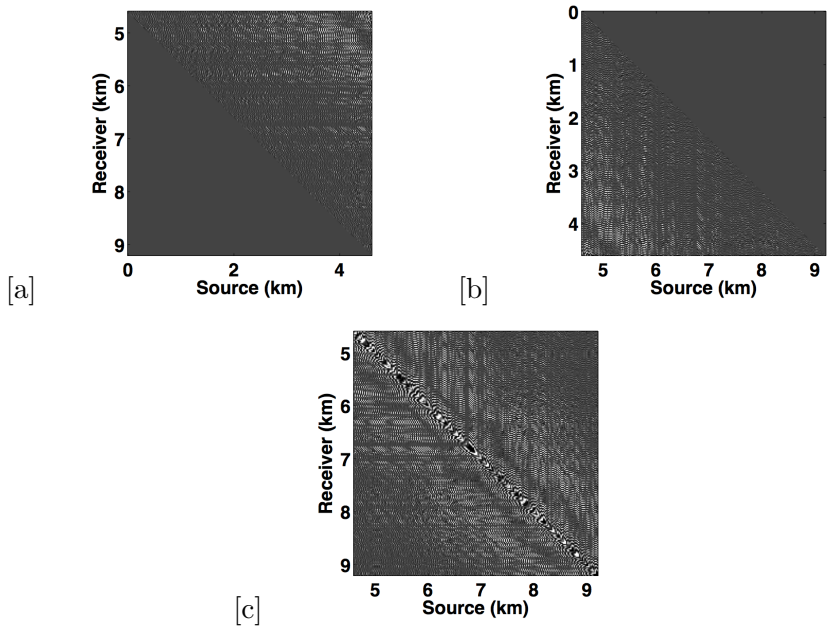


Figure 4.5: (a,b,c) First-level sub-block matrices (from Figure 4.4a).

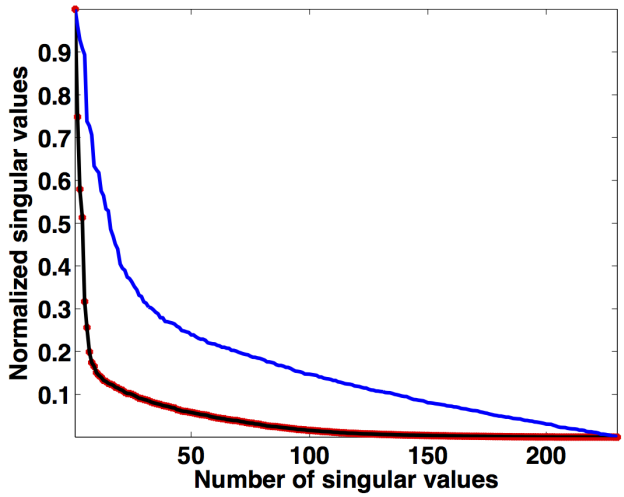


Figure 4.6: Decay of singular values of the HSS sub-blocks in s-r domain: red—Figure 4.5a, black—Figure 4.5b, blue—Figure 4.5c.

One of the limitations of matrix completion type approaches for large-scale seismic data is the nuclear-norm projection, which inherently involves the computation of singular value decompositions (SVD). Aravkin et al. [2014b] showed that the computation of SVD is prohibitively expensive for large-scale data such as seismic, therefore, we propose a matrix-factorization based approach to avoid the need for expensive computation of SVDs [see Aravkin et al., 2014b, for details]. In the next section, we introduce the matrix completion framework and explore its necessary extension to separate large-scale simultaneous seismic data.

4.3.3 Large-scale seismic data: SPG-LR framework

Let $\mathbf{X}_0$ be a low-rank matrix in $\mathbb{C}^{n\times m}$ and $\mathcal{A}$ be a linear measurement operator that maps from $\mathbb{C}^{n\times m} \rightarrow \mathbb{C}^p$ with $p \ll n \times m$. Under the assumption that the blending process increases the rank of the matrix $\mathbf{X}_0$, the source separation problem is to find the matrix of lowest possible rank that agrees with the above observations. The rank-minimization problem involves solving the following problem for $\mathcal{A}$, up to a given tolerance ϵ:

$$\underset{\mathbf{X}}{\text{minimize}} \quad \text{rank}(\mathbf{X}) \quad \text{subject to} \quad \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2 \leq \epsilon,$$

where rank is defined as the maximum number of linearly independent rows or column of a matrix, $\mathbf{b}$ is a set of blended measurements. For simultaneous towed-streamer acquisition, we follow equation 4.2 and redefine our system of equations as

$$\overbrace{\begin{bmatrix}\mathbf{M}\mathbf{T}_1\mathcal{S}^H & \mathbf{M}\mathbf{T}_2\mathcal{S}^H\end{bmatrix}}^{\mathcal{A}} \overbrace{\begin{bmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{bmatrix}}^{\mathbf{X}} = \mathbf{b},$$

where $\mathcal{S}$ is the transformation operator from the s-r domain to the m-h domain. Recht et al. [2010b] showed that under certain general conditions on the operator $\mathcal{A}$, the solution to the rank-minimization problem can be found by solving the following nuclear-norm minimization problem:

$$\underset{\mathbf{X}}{\text{minimize}} \quad \|\mathbf{X}\|_* \quad \text{subject to} \quad \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2 \leq \epsilon, \qquad (\text{BPDN}_\epsilon)$$

where $\|\mathbf{X}\|_* = \|\sigma\|_1$, and σ is a vector of singular values. Unfortunately, for large-scale data, solving the BPDN_ϵ problem is difficult since it requires repeated projections onto the set $\|\mathbf{X}\|_* \leq \tau$, which means repeated SVD or partial SVD computations. Therefore, in this paper, we avoid computing singular value decompositions (SVD) of the matrices and use an extension of the SPGℓ_1 solver [Berg and Friedlander, 2008] developed for the BPDN_ϵ problem in Aravkin et al. [2013b]. We refer to this extension as SPG-LR in the rest of the paper. The SPG-LR algorithm finds the solution to the BPDN_ϵ problem by solving a sequence of LASSO (least absolute shrinkage and selection

operator) subproblems:

$$\underset{\mathbf{X}}{\text{minimize}} \quad \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2 \quad \text{subject to} \quad ||\mathbf{X}||_* \leq \tau, \tag{LASSO$_\tau$}$$

where τ is updated by traversing the Pareto curve. The Pareto curve defines the optimal trade-off between the two-norm of the residual and the one-norm of the solution [Berg and Friedlander, 2008]. Solving each LASSO subproblem requires a projection onto the nuclear-norm ball $\|\mathbf{X}\|_* \leq \tau$ in every iteration by performing a singular value decomposition and then thresholding the singular values. For large-scale seismic problems, it becomes prohibitively expensive to carry out such a large number of SVDs. Instead, we adopt a recent factorization-based approach to nuclear-norm minimization [Rennie and Srebro, 2005a, Lee et al., 2010a, Recht and Ré, 2011]. The factorization approach parametrizes the matrix ($\mathbf{X}_1$, $\mathbf{X}_2$) $\in \mathbb{C}^{n\times m}$ as the product of two low-rank factors ($\mathbf{L}_1$, $\mathbf{L}_2$) $\in \mathbb{C}^{n\times k}$ and ($\mathbf{R}_1$, $\mathbf{R}_2$) $\in \mathbb{C}^{m\times k}$ such that,

$$\mathbf{X} = \begin{bmatrix} \mathbf{L}_1\mathbf{R}_1^H \\ \mathbf{L}_2\mathbf{R}_2^H \end{bmatrix}. \tag{4.3}$$

Here, k represents the rank of the $\mathbf{L}$ and $\mathbf{R}$ factors. The optimization scheme can then be carried out using the factors $(\mathbf{L}_1, \mathbf{L}_2)$ and $(\mathbf{R}_1, \mathbf{R}_2)$ instead of $(\mathbf{X}_1, \mathbf{X}_2)$, thereby significantly reducing the size of the decision variable from $2nm$ to $2k(n+m)$ when $k \ll m, n$. Rennie and Srebro [2005a] showed that the nuclear-norm obeys the relationship:

$$\|\mathbf{X}\|_* \quad \leq \quad \frac{1}{2}\left\|\begin{bmatrix}\mathbf{L}_1 \\ \mathbf{R}_1\end{bmatrix}\right\|_F^2 + \frac{1}{2}\left\|\begin{bmatrix}\mathbf{L}_2 \\ \mathbf{R}_2\end{bmatrix}\right\|_F^2 \quad =: \quad \Phi(\mathbf{L}_1, \mathbf{R}_1, \mathbf{L}_2, \mathbf{R}_2), \tag{4.4}$$

where $\|\cdot\|_F^2$ is the Frobenius norm of the matrix—i.e., sum of the squared entires. Consequently, the LASSO subproblem can be replaced by

$$\underset{\mathbf{L}_1,\mathbf{R}_1,\mathbf{L}_2,\mathbf{R}_2}{\text{minimize}} \quad \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2 \quad \text{subject to} \quad \Phi(\mathbf{L}_1, \mathbf{R}_1, \mathbf{L}_2, \mathbf{R}_2) \leq \tau\ , \tag{4.5}$$

where the projection onto $\Phi(\mathbf{L}_1, \mathbf{R}_1, \mathbf{L}_2, \mathbf{R}_2) \leq \tau$ is easily achieved by multiplying each factor $(\mathbf{L}_1, \mathbf{L}_2)$ and $(\mathbf{R}_1, \mathbf{R}_2)$ by the scalar $\sqrt{2\tau/\Phi(\mathbf{L}_1, \mathbf{R}_1, \mathbf{L}_2, \mathbf{R}_2)}$. Equation 4.4, for each HSS sub-matrix in the m-h domain, guarantees that $\|\mathbf{X}\|_* \leq \tau$ for any solution of 4.5. Once the optimization problem is solved, each sub-matrix in the m-h domain is transformed back into the s-r domain, where we concatenate all the sub-matrices to get the deblended monochromatic frequency data matrices. One of the advantages of the HSS representation is that it works with recursive partitioning of a matrix and sub-matrices can be solved in parallel, speeding up the optimization formulation.

4.4 Experiments

We perform source separation for two simultaneous towed-streamer acquisition scenarios—over/under and simultaneous long offset, by generating synthetic datasets on complex geological models using the IWAVE [Symes et al., 2011a] time-stepping acoustic simulation software, and also use a field dataset from the Gulf of Suez. Source separation for over/under acquisition is tested on two different datasets. The first dataset is simulated on the Marmousi model [Bourgeois et al., 1991], which represents a complex-layer model with steeply dipping reflectors that make the data challenging. With a source (and channel/receiver) sampling of 20.0 m, one dataset is generated with a source-depth of 8.0 m (Figures 4.7a and 4.7d), while the other dataset has the source at 12.0 m depth (Figures 4.7b and 4.7e), resulting in 231 sources and 231 channels. The temporal length of each dataset is 4.0 s with a sampling interval of 0.004 s. The second dataset is a field data example from the Gulf of Suez. In this case, the first source is placed at 5.0 m depth (Figures 4.8a and 4.8d) and the second source is placed at 10.0 m depth (Figures 4.8b and 4.8e). The source (and channel) sampling is 12.5 m, resulting in 178 sources and 178 channels with a time sampling interval of 0.004 s.

The simultaneous long offset acquisition is simulated on the BP salt model [Billette and Brandsberg-Dahl, 2004], where the presence of salt-bodies make the data challenging. The two source vessels are 6.0 km apart and the streamer length is 6.0 km. Both the datasets (for source 1 and source 2) contain 361 sources and 361 channels with a spatial interval of 12.5 m, where the source and streamer depth is 6.25 m. The temporal length of each dataset is 6.0 s with a sampling interval of 0.006 s. A single shot gather from each dataset is shown in Figures 4.9a and 4.9b and the corresponding channel gathers are shown in Figures 4.9d and 4.9e. The datasets for each source in both the acquisition scenarios are (simply) summed for simultaneous acquisition with periodic firing times, while uniformly random time delays between 0-1 second are applied to each source for the randomized simultaneous acquisition. Figures 4.7c, 4.8c and 4.9c show the randomized blended shot gathers for the Marmousi, the Gulf of Suez and the BP datasets, respectively. As illustrated in the figures, both the sources fire at random times (independent of each other) within the interval of 0-1 second, hence, the difference between the firing times of the sources is always less than 1 second. The corresponding randomized blended channel gathers are shown in Figures 4.7f, 4.8f and 4.9f. Note that the speed of the vessels in both the acquisition scenarios is no different than the current practical speed of the vessels in the field.

For deblending via rank-minimization, second-level of HSS partitioning, on each frequency slice in the s-r domain, was sufficient for successful recovery in both the acquisition scenarios. After transforming each sub-block into the m-h domain, deblending is then performed by solving the nuclear-norm minimization formulation (BPDN_ϵ) on each sub-block, using 350 iterations of SPG-LR. In order to choose an appropriate rank value, we first perform deblending for frequency slices at 0.2 Hz and 125 Hz. For the over/under acquisition simulated on the Marmousi model, the best

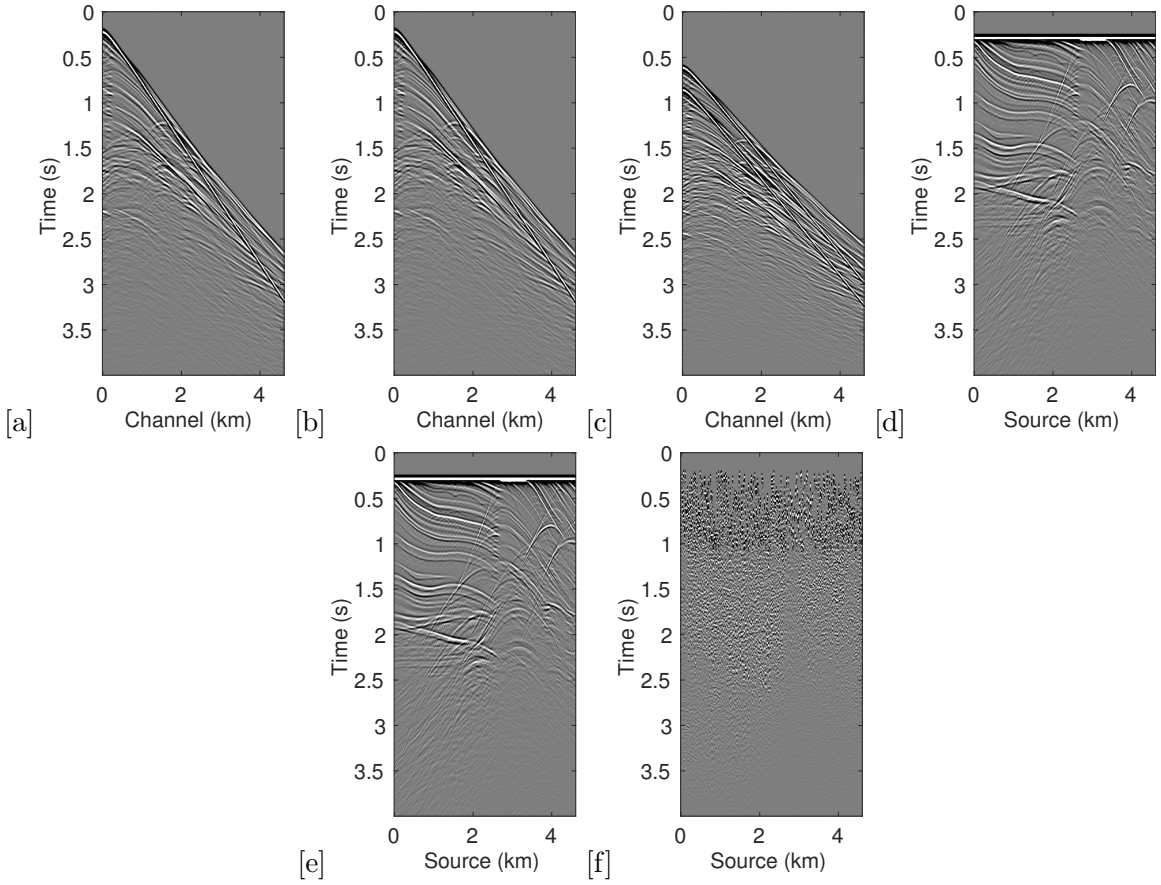


Figure 4.7: Original shot gather of (a) source 1, (b) source 2, and (c) the corresponding blended shot gather for simultaneous over/under acquisition simulated on the Marmousi model. (d, e) Corresponding common-channel gathers for each source and (f) the blended common-channel gather.

rank value is 30 and 80 for each frequency slice, respectively. The best rank values for the Gulf of Suez dataset are 20 and 100, respectively. For simultaneous long offset acquisition, the best rank value is 10 and 90 for frequency slices at 0.15 Hz and 80 Hz, respectively. Hence, we adjust the rank linearly within these ranges when moving from low to high frequencies, for each acquisition scenario. For deblending via sparsity-promotion, we use the BPDN$_\epsilon$ formulation to minimize the ℓ_1 norm (instead of the nuclear-norm) where the transformation operator $\mathcal{S}$ is the 2-D curvelet operator. Here, we run 350 iterations of SPGℓ_1.

For the over/under acquisition scenario simulated on the Marmousi model, Figures 4.10a and 4.10c show the deblended shot gathers via rank-minimization and Figures 4.10e and 4.10g show the deblended shot gathers via sparsity-promotion, respectively. The deblended common-channel gathers via rank-minimization and sparsity-promotion are shown in Figures 4.11a, 4.11c and Figures 4.11e, 4.11g, respectively. For the Gulf of Suez field dataset, Figures 4.12 and 4.13 show

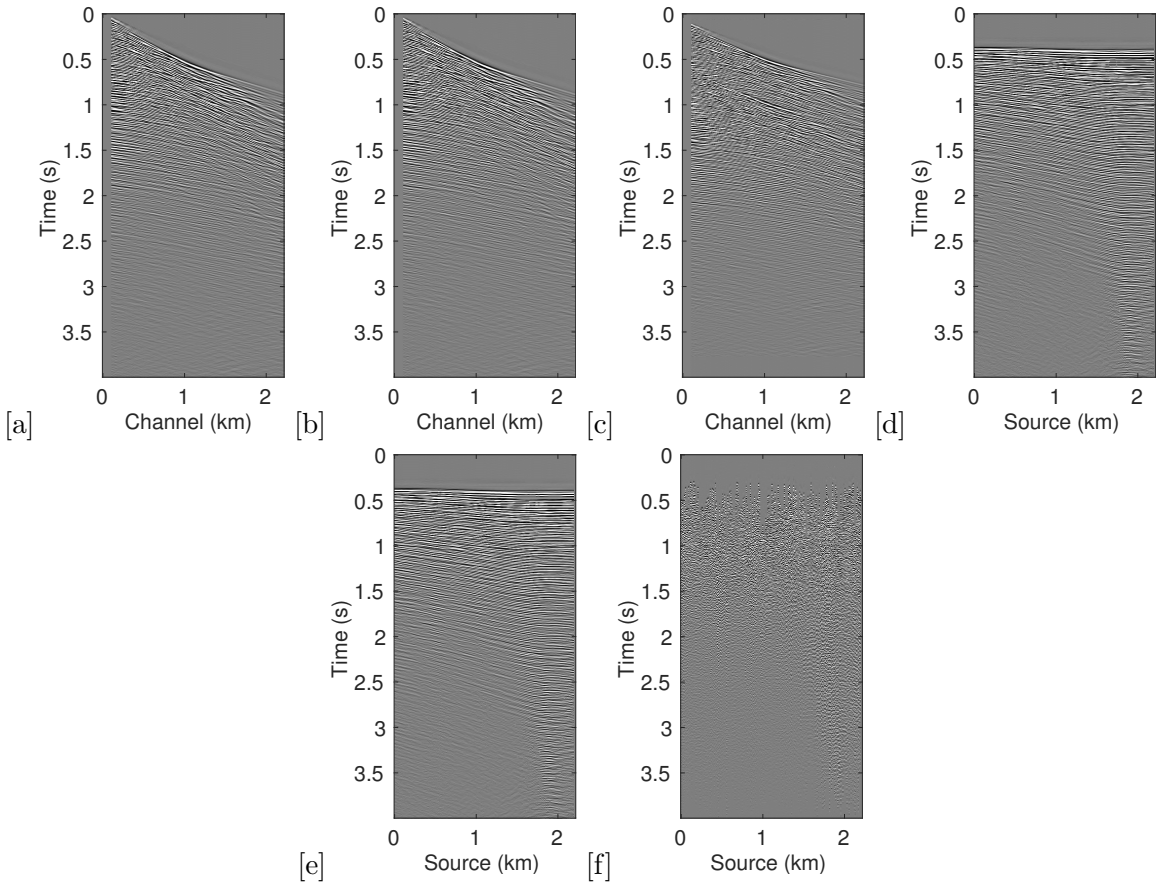


Figure 4.8: Original shot gather of (a) source 1, (b) source 2, and (c) the corresponding blended shot gather for simultaneous over/under acquisition from the Gulf of Suez dataset. (d, e) Corresponding common-channel gathers for each source and (f) the blended common-channel gather.

the deblended gathers and difference plots in the common-shot and common-channel domain, respectively. The corresponding deblended gathers and difference plots in the common-shot and common-channel domain for the simultaneous long offset acquisition scenario are shown in Figures 4.14 and 4.15.

As illustrated by the results and their corresponding difference plots, both the CS-based approaches of rank-minimization and sparsity-promotion are able to deblend the data for the low-variability acquisition scenarios fairly well. In all the three different datasets, the average SNRs for separation via sparsity-promotion is slightly better than rank-minimization, but the difference plots show that the recovery via rank-minimization is equivalent to the sparsity-promoting based recovery where it is able to recover most of the coherent energy. Also, rank-minimization outperforms the sparsity-promoting technique in terms of the computational time and memory usage as represented in Tables 4.1, 4.2 and 4.3. Both the CS-based recoveries are better for the simultane-

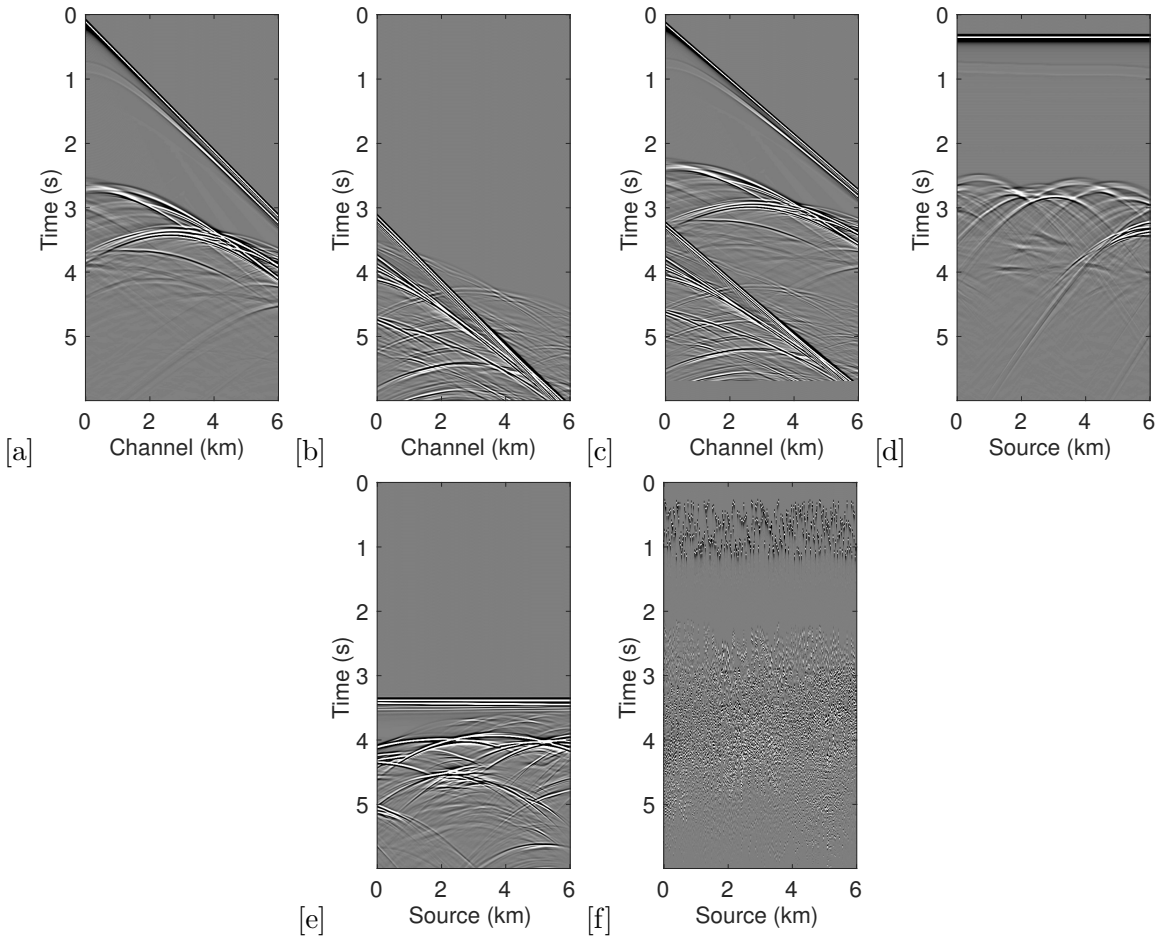


Figure 4.9: Original shot gather of (a) source 1, (b) source 2, and (c) the corresponding blended shot gather for simultaneous long offset acquisition simulated on the BP salt model. (d, e) Corresponding common-channel gathers for each source and (f) the blended common-channel gather.

ous long offset acquisition than the recoveries from the over/under acquisition scenario. A possible explanation for this improvement is the long offset distance that increases randomization in the simultaneous acquisition, which is a more favourable scenario for recovery by CS-based approaches. Figure 4.16 demonstrates the advantage of the HSS partitioning, where the SNRs of the deblended data are significantly improved.

4.4.1 Comparison with nmo-based median filtering

We also compare the performance of our CS-based deblending techniques with deblending using the NMO-based median filtering technique proposed by Chen et al. [2014], where we work on a common-midpoint gather from each acquisition scenario. For the over/under acquisition simulated on the Marmousi model, Figures 4.17a and 4.17e show the blended common-midpoint gathers and

	Marmousi model		
	Time	Memory	SNR
Sparsity	167	7.0	16.7, 16.7
Rank	12	2.8	15.0, 14.8

Table 4.1: Comparison of computational time (in hours), memory usage (in GB) and average SNR (in dB) using sparsity-promoting and rank-minimization based techniques for the Marmousi model.

	Gulf of Suez		
	Time	Memory	SNR
Sparsity	118	6.6	14.6
Rank	8	2.6	12.8

Table 4.2: Comparison of computational time (in hours), memory usage (in GB) and average SNR (in dB) using sparsity-promoting and rank-minimization based techniques for the Gulf of Suez dataset.

deblending using the median filtering technique is shown in Figures 4.17b and 4.17f. The corresponding deblended common-midpoint gathers from the two CS-based techniques are shown in Figures 4.17(c,d,g,h). Figure 4.18 shows the blended and deblended common-midpoint gathers for the field data from the Gulf of Suez. We observe that recoveries via the proposed CS-based approaches are comparable to the recovery from the median filtering technique. Similarly, Figure 4.19 shows the results for the simultaneous long offset acquisition simulated on the BP salt model. Here, the CS-based techniques result in slightly improved recoveries.

4.4.2 Remark

It is important to note here that we perform the CS-based source separation algorithms only once, however, we can always perform a few more runs of the algorithms where we can first subtract the deblended source 1 and source 2 from the acquired blended data and then re-run the algorithms to deblend the energy in the residual data. Hence, the recovery can be further improved until necessary. Since separation via rank-minimization is computationally faster than the sparsity based technique, multiple passes through the data is a computationally viable option for the former deblending technique.

4.5 Discussion

The above experiments demonstrate the successful implementation of the proposed CS-based approaches of rank-minimization and sparsity-promotion for source separation in the low-variability,

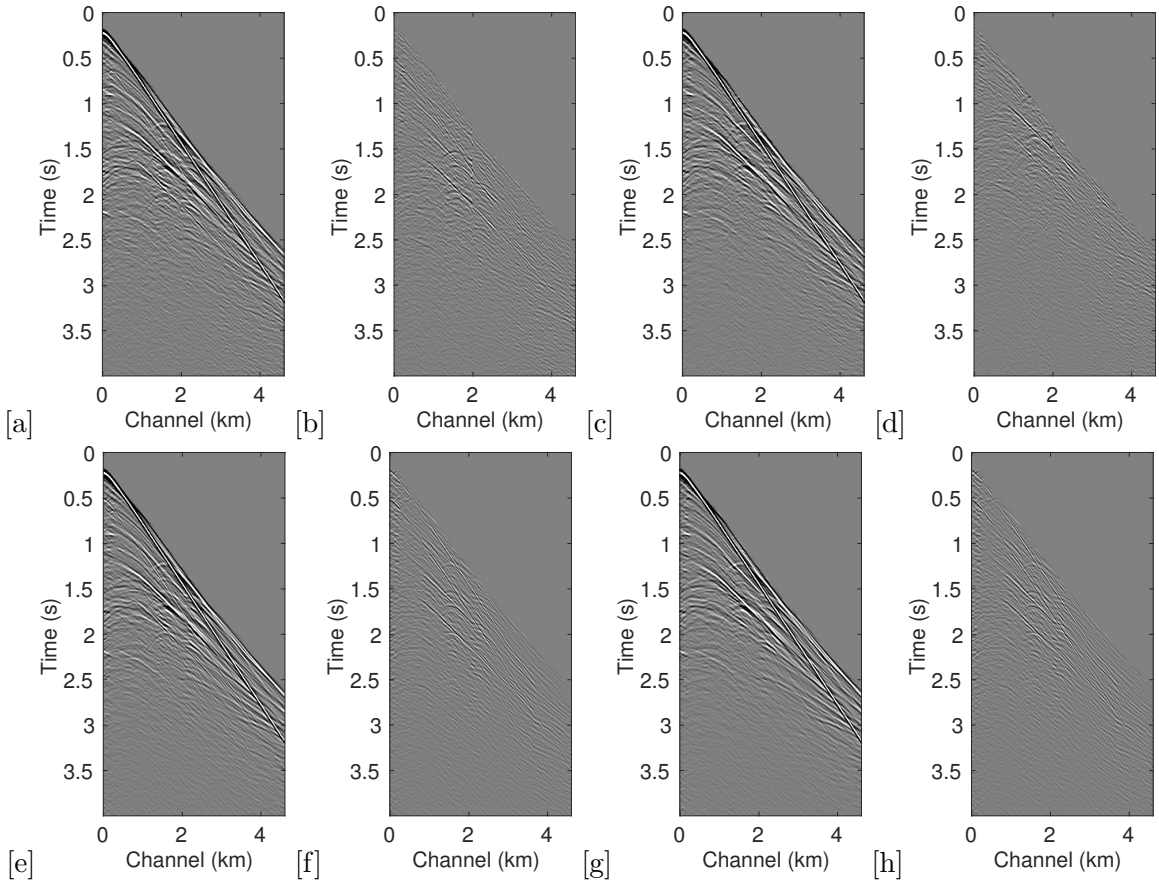


Figure 4.10: Deblended shot gathers and difference plots (from the Marmousi model) of source 1 and source 2: (a,c) deblending using HSS based rank-minimization and (b,d) the corresponding difference plots; (e,g) deblending using curvelet-based sparsity-promotion and (f,h) the corresponding difference plots.

	BP model		
	time	memory	SNR
Sparsity	325	7.0	32.0, 29.4
Rank	20	2.8	29.4, 29.0

Table 4.3: Comparison of computational time (in hours), memory usage (in GB) and average SNR (in dB) using sparsity-promoting and rank-minimization based techniques for the BP model.

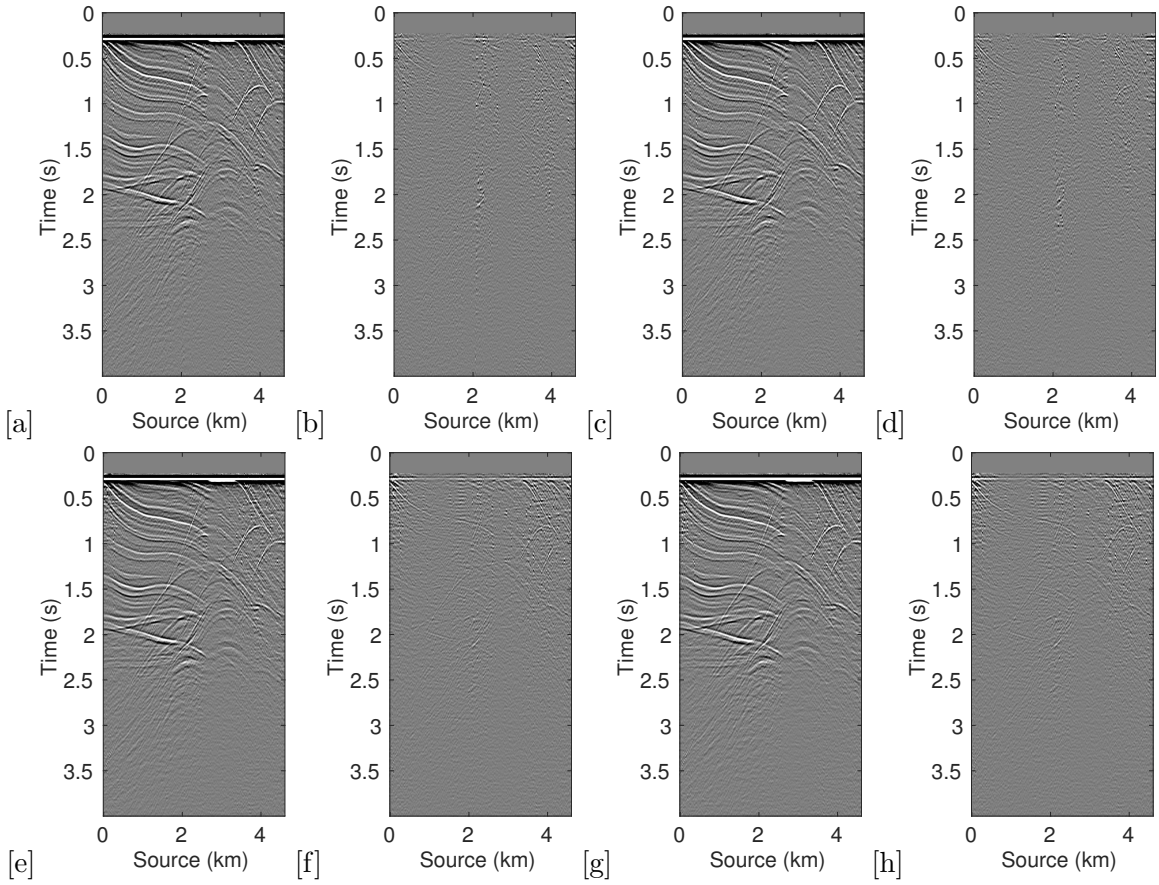


Figure 4.11: Deblended common-channel gathers and difference plots (from the Marmousi model) of source 1 and source 2: (a,c) deblending using HSS based rank-minimization and (b,d) the corresponding difference plots; (e,g) deblending using curvelet-based sparsity-promotion and (f,h) the corresponding difference plots.

simultaneous towed-streamer acquisitions. The recovery is comparable for both approaches, however, separation via rank-minimization is significantly faster and memory efficient. This is further enhanced by incorporating the HSS partitioning since it allows the exploitation of the low-rank structure in the high-frequency regime, and renders its extension to large-scale data feasible. Note that in the current implementation, we work with each temporal frequency slice and perform the source separation individually. The separation results can further be enhanced by incorporating the information from the previously recovered frequency slice to the next frequency slice, as shown by Mansour et al. [2013] for seismic data interpolation.

The success of CS hinges on randomization of the acquisition. Although, the low degree of randomization (e.g., $0 \leq 1$ second) in simultaneous towed-streamer acquisitions seems favourable for source separation via CS-based techniques, however, high-variability in the firing times enhances the recovery quality of separated seismic data volumes, as shown in Wason and Herrmann [2013a];

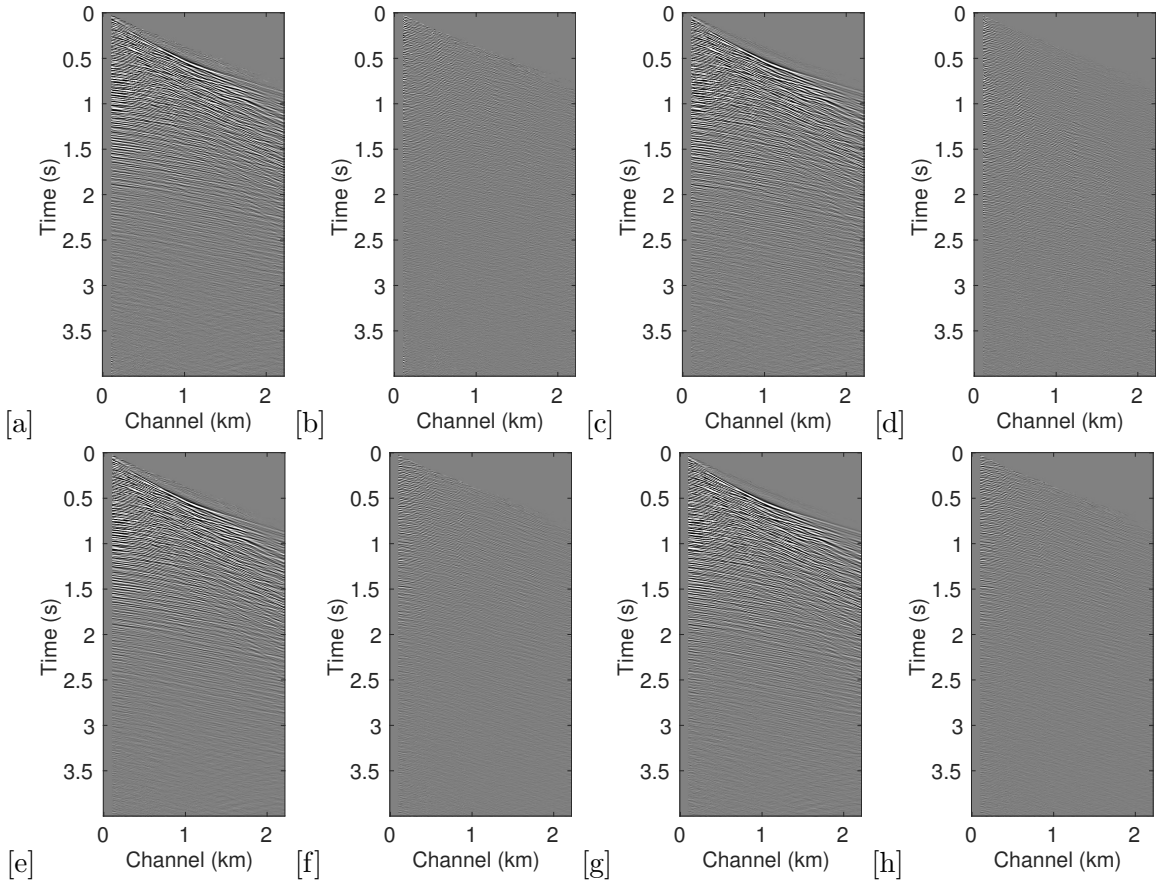


Figure 4.12: Deblended shot gathers and difference plots (from the Gulf of Suez dataset) of source 1 and source 2: (a,c) deblending using HSS based rank-minimization and (b,d) the corresponding difference plots; (e,g) deblending using curvelet-based sparsity-promotion and (f,h) the corresponding difference plots.

Wason and Herrmann [2013b] for ocean-bottom cable/node acquisition with continuous recording. One of the advantages of the proposed CS-based techniques is that it does not require velocity estimation, which can be a challenge for data with complex geologies. However, the proposed techniques require accurate knowledge of the random firing times.

So far, we have not considered the case of missing traces (sources and/or receivers), however, incorporating this scenario in the current framework is straightforward. This makes the problem a joint deblending and interpolation problem. In reality, seismic data are typically irregularly sampled along spatial axes, therefore, future work includes working with non-uniform sampling grids. Finally, we envisage that our methods can, in principle, be extended to separate 3-D blended seismic data volumes.

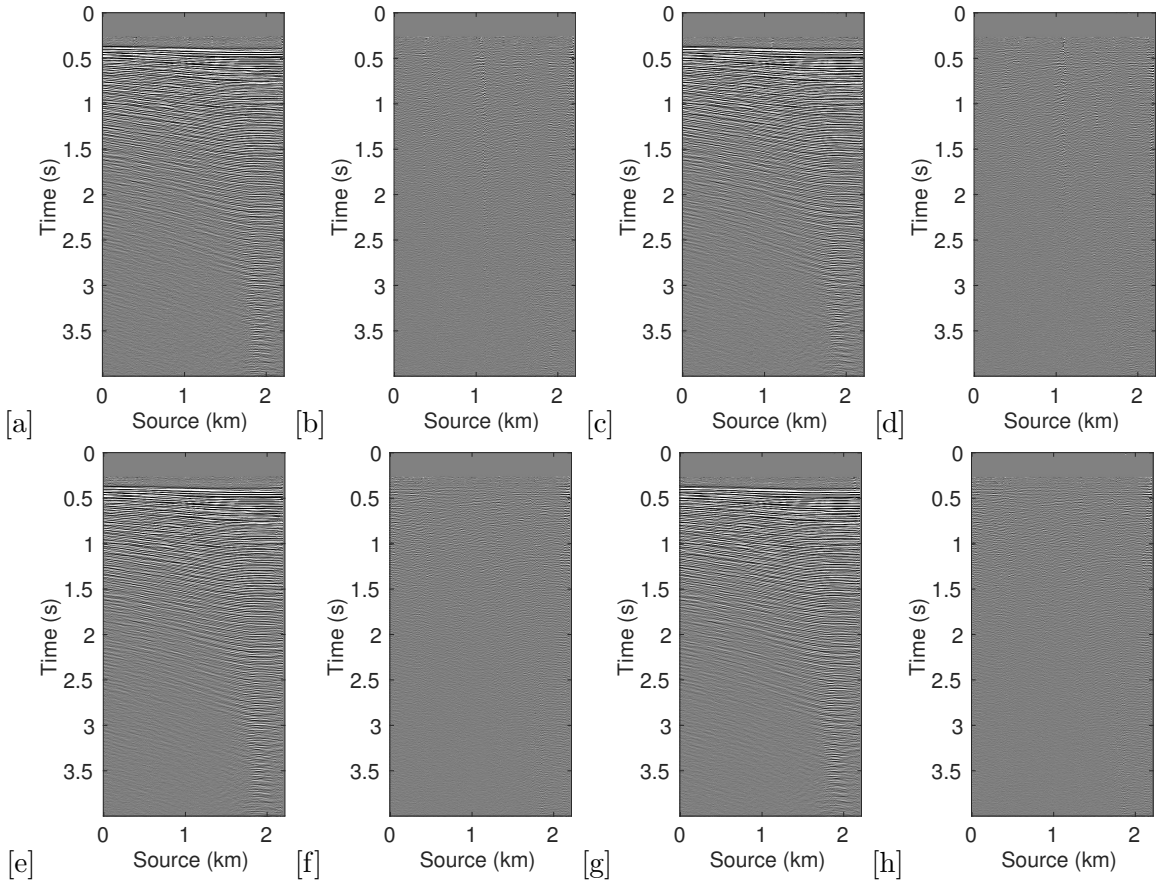


Figure 4.13: Deblended common-channel gathers and difference plots (from the Gulf of Suez dataset) of source 1 and source 2: (a,c) deblending using HSS based rank-minimization and (b,d) the corresponding difference plots; (e,g) deblending using curvelet-based sparsity-promotion and (f,h) the corresponding difference plots.

4.6 Conclusions

We have presented two compressed sensing based methods for source separation for simultaneous towed-streamer type acquisitions, such as the over/under and the simultaneous long offset acquisition. Both the compressed sensing based approaches of rank-minimization and sparsity-promotion give comparable deblending results, however, the former approach is readily scalable to large-scale blended seismic data volumes and is computationally faster. This can be further enhanced by incorporating the HSS structure with factorization-based rank-regularized optimization formulations, along with improved recovery quality of the separated seismic data. We have combined the Pareto curve approach for optimizing BPDN$_\epsilon$ formulations with the SVD-free matrix factorization methods to solve the nuclear-norm optimization formulation, which avoids the expensive computation of the singular value decomposition (SVD), a necessary step in traditional rank-minimization based methods. We find that our proposed techniques are comparable to the commonly used NMO-based

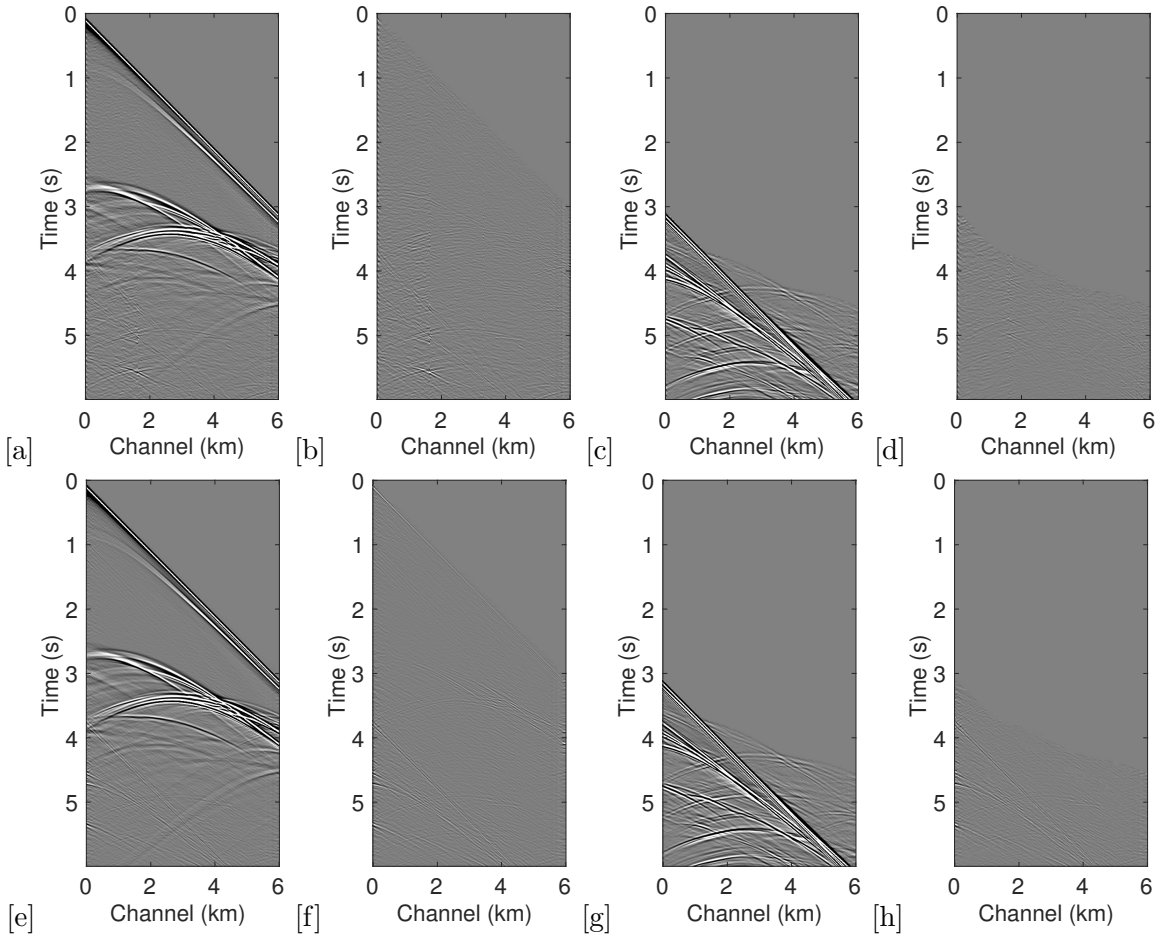


Figure 4.14: Deblended shot gathers and difference plots (from the BP salt model) of source 1 and source 2: (a,c) deblending using HSS based rank-minimization and (b,d) the corresponding difference plots; (e,g) deblending using curvelet-based sparsity-promotion and (f,h) the corresponding difference plots.

median filtering techniques.

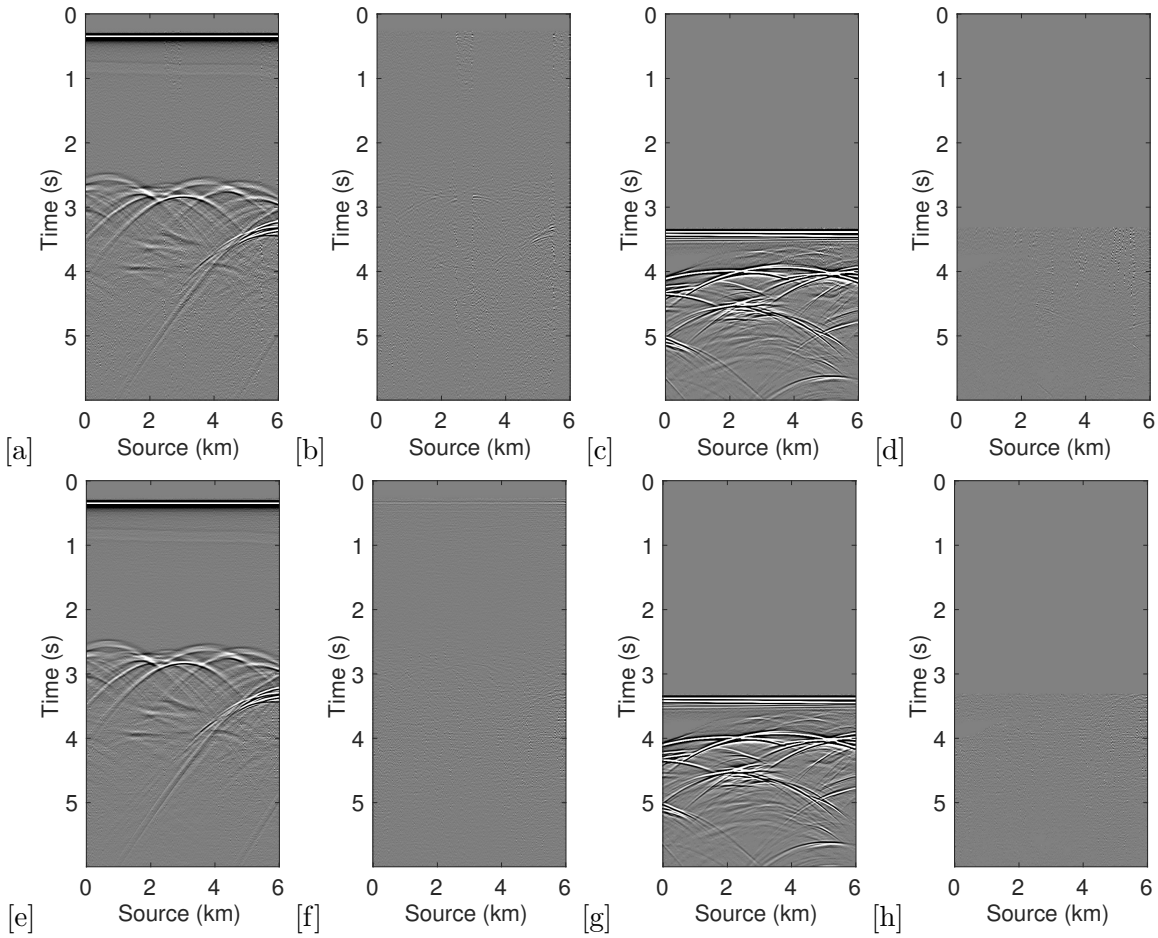


Figure 4.15: Deblended common-channel gathers and difference plots (from the BP salt model) of source 1 and source 2: (a,c) deblending using HSS based rank-minimization and (b,d) the corresponding difference plots; (e,g) deblending using curvelet-based sparsity-promotion and (f,h) the corresponding difference plots.

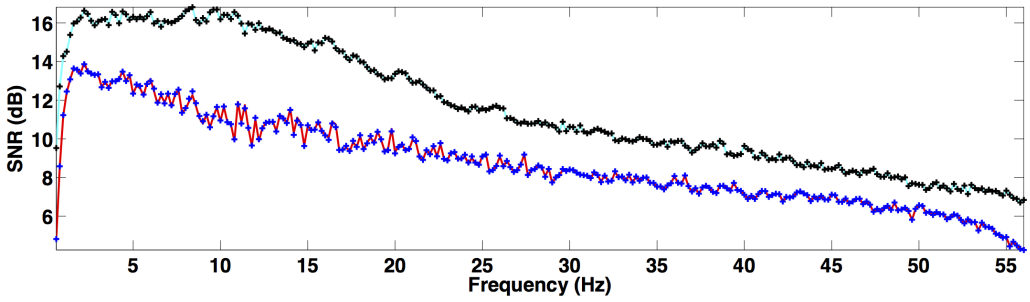


Figure 4.16: Signal-to-noise ratio (dB) over the frequency spectrum for the deblended data from the Marmousi model. Red, blue curves—deblending without HSS; cyan, black curves—deblending using second-level HSS partitioning. Solid lines—separated source 1, + marker—separated source 2.

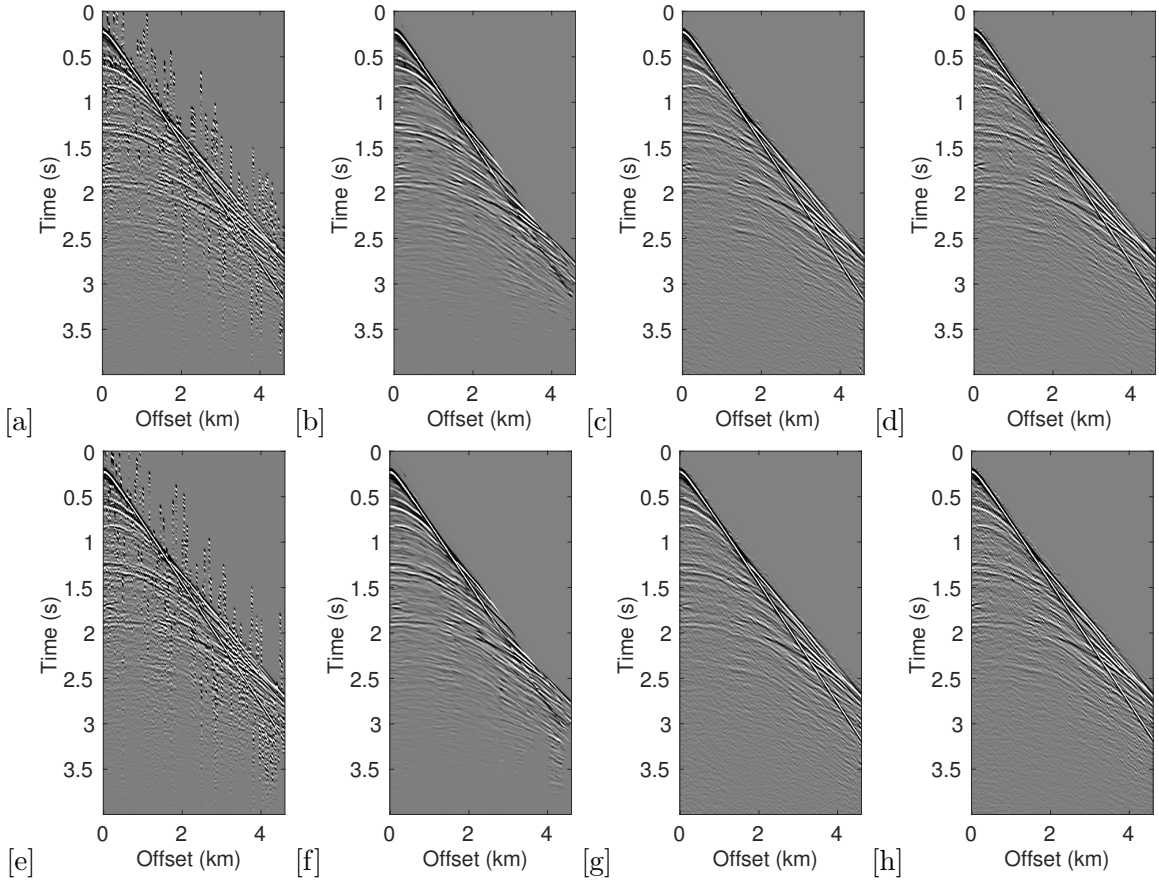


Figure 4.17: Blended common-midpoint gathers of (a) source 1 and (e) source 2 for the Marmousi model. Deblending using (b,f) NMO-based median filtering, (c,g) rank-minimization and (d,h) sparsity-promotion.

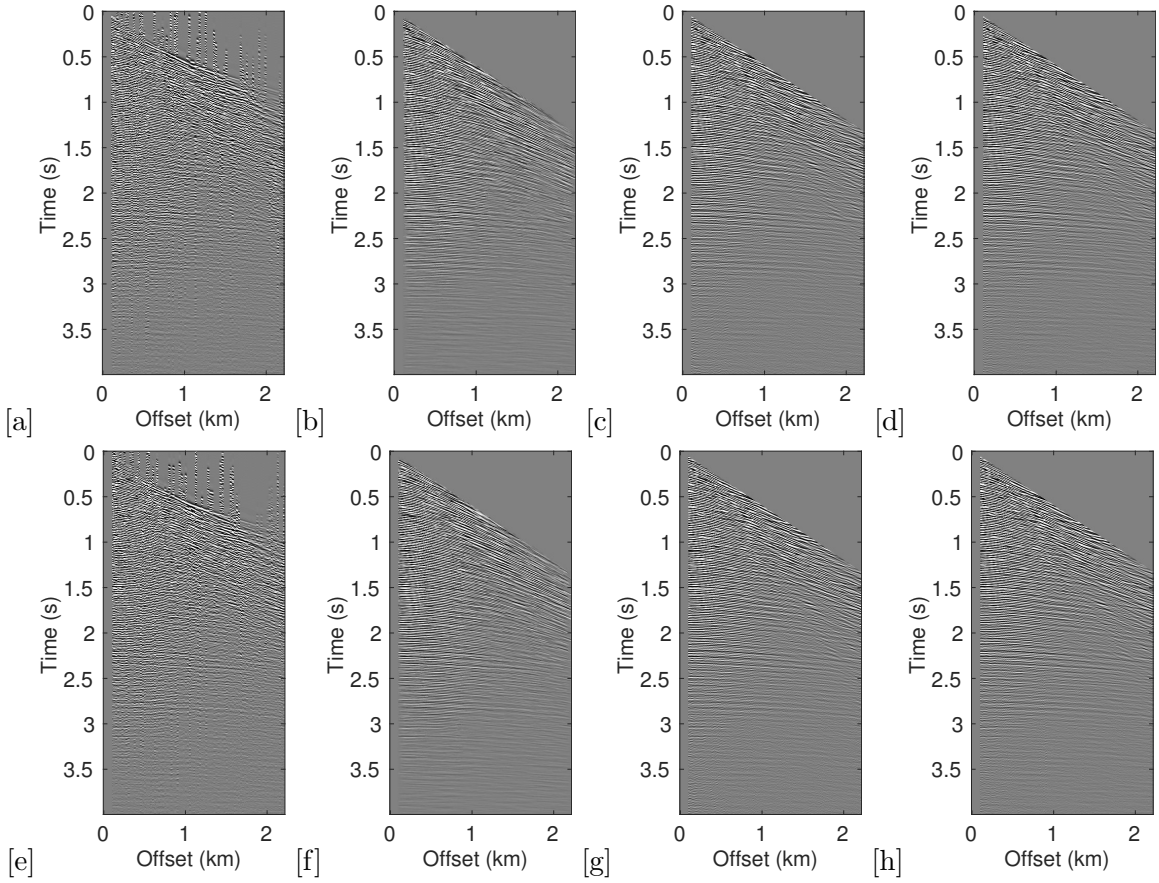


Figure 4.18: Blended common-midpoint gathers of (a) source 1, (e) source 2 for the Gulf of Suez dataset. Deblending using (b,f) NMO-based median filtering, (c,g) rank-minimization and (d,h) sparsity-promotion.

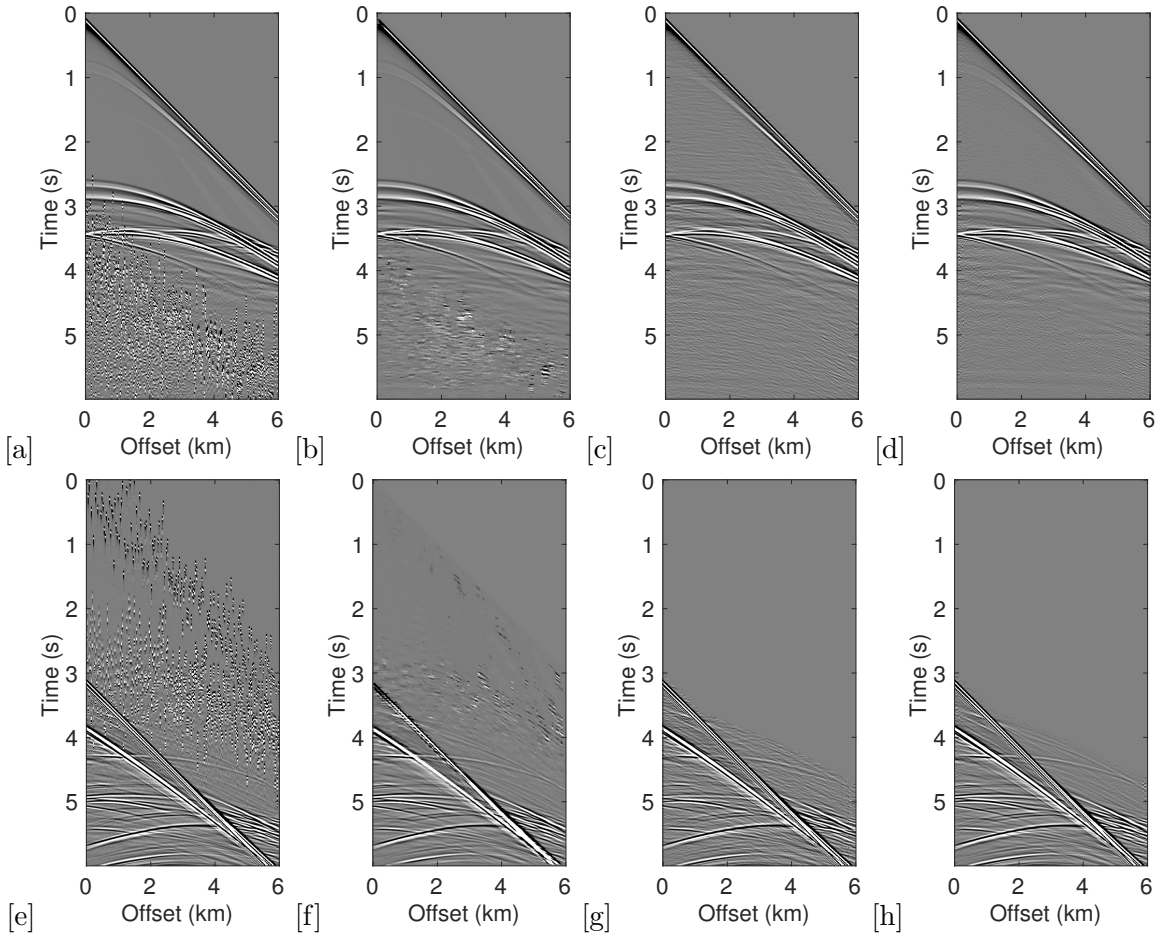


Figure 4.19: Blended common-midpoint gathers of (a) source 1, (e) source 2 for the BP salt model. Deblending using (b,f) NMO-based median filtering, (c,g) rank-minimization and (d,h) sparsity-promotion.

Chapter 5

Large-scale time-jittered simultaneous marine acquisition: rank-minimization approach

5.1 Summary

In this chapter, we present a computationally tractable rank-minimization algorithm to separate simultaneous time-jittered continuous recording for a 3D ocean-bottom cable survey. We will show experimentally that the proposed algorithm is computationally tractable for 3D time-jittered marine acquisition.

5.2 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 et al. [1998], de Kok and Gillespie [2002], Berkhout [2008a]. The final objective of source separation is to get interference-free shot records. Wason and Herrmann [2013b] 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

A version of this chapter has been published in the proceedings of SEG Annual Meeting, 2016, Denver, USA

seismic data volumes.

Recently, nuclear-norm minimization based methods have shown the potential to overcome the computational bottleneck [Kumar et al., 2015a], hence, these methods are successfully used for source separation [Maraschini et al., 2012, Cheng and Sacchi, 2013, Kumar et al., 2015b]. 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 chapter 4, 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 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 et al. [2010a]. The experimental results on a synthetic 5D data set demonstrate successful implementation of the proposed methodology.

5.3 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 [2013b], 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 5.1). 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 coordinates 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 $\mathcal{A} = \mathcal{M}\mathcal{R}\mathcal{S}$, where the operator $\mathcal{S}$ permutes the tensor coordinates from $(n_{rx}, n_{sx}, n_{ry}, n_{sy})$ (rank-revealing domain, i.e., Figure 5.2a) to $(n_{rx}, n_{ry}, n_{sx}, n_{sy})$ (standard acquisition ordering, i.e., Figure 5.2b) and its adjoint reverses this permutation. The restriction operator $\mathcal{R}$ subsamples the conventional data volume at jittered source

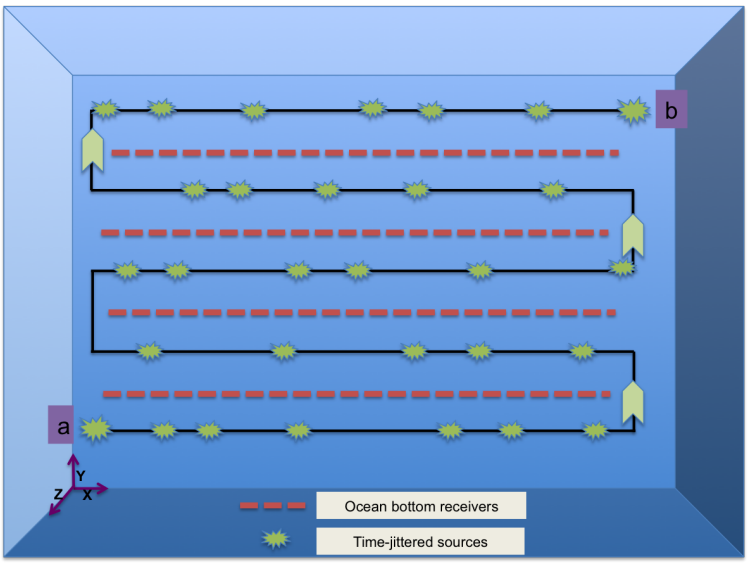


Figure 5.1: Aerial view of the 3D time-jittered marine acquisition. Here, we consider one source vessel with two airgun arrays firing at jittered times and locations. Starting from point a, the source vessel follows the acquisition path shown by black lines and ends at point b. The receivers are placed at the ocean bottom (red dashed lines).

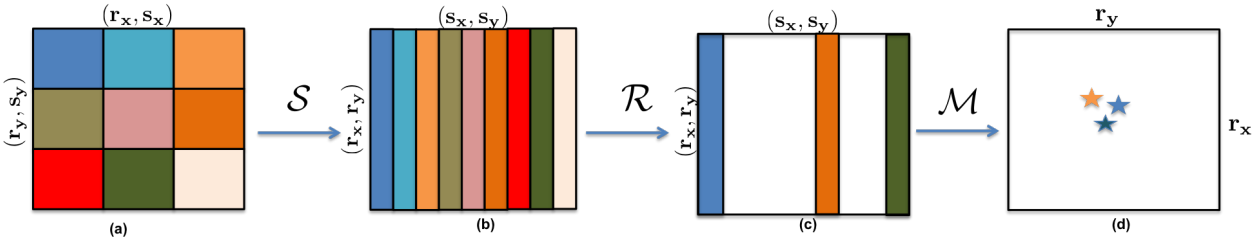


Figure 5.2: Schematic representation of the sampling-transformation operator $\mathcal{A}$ during the forward operation. The adjoint of the operator $\mathcal{A}$ follows accordingly. (a, b, c) represent a monochromatic data slice from conventional data volume and (d) represents a time slice from the continuous data volume.

locations (Figure 5.2c), the sampling operator $\mathcal{M}$ maps the conventional subsampled temporal-frequency domain data to the simultaneous time-domain data (Figure 5.2d). Note that Figure 5.2d 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 [2012b] 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

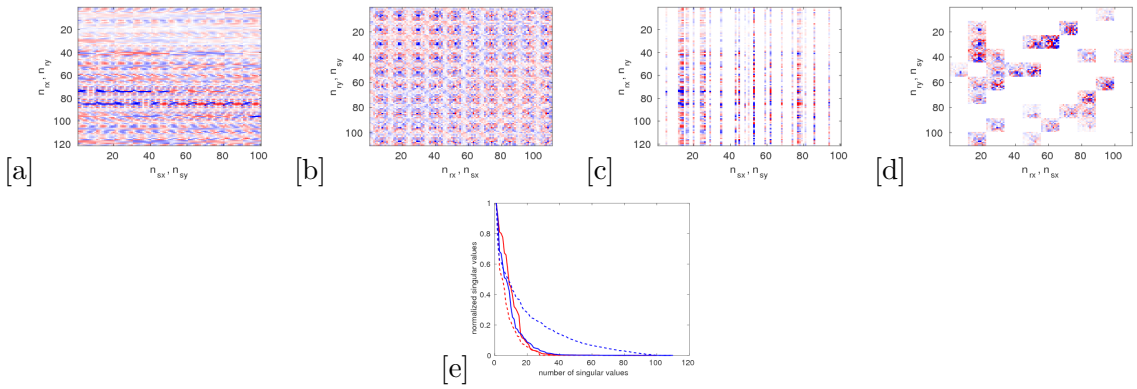


Figure 5.3: Monochromatic slice at 10.0 Hz. Fully sampled data volume and simultaneous data volume matricized as (a, c) $i = (n_{sx}, n_{sy})$, and (b, d) $i = (n_{rx}, n_{sx})$. (e) Decay of singular values. Notice that fully sampled data organized as $i = (n_{sx}, n_{sy})$ has slow decay of the singular 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.

the columns. In this work, we consider the matricization proposed by Silva and Herrmann [2013a], where $i = (n_{sx}, n_{sy})$—i.e., placing both source coordinates along the columns (Figure 5.3a), or $i = (n_{rx}, n_{sx})$—i.e., placing receiver-x and source-x coordinates along the columns (Figure 5.3b). As we see in Figure 5.3e, 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 5.3c), 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 5.3d), hence, slowing down the decay of the singular values (dotted blue curve). This scenario is much closer to the matrix-completion problem (Recht et al. [2010b]), 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 et al. [2010b] showed that solutions to rank-minimization problems can be found by solving a nuclear-norm minimization problem. Silva and Herrmann [2013a] 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 temporal-frequency domain as follows:

$$\min_{\mathbf{D}} \sum_{j}^{n_f} \|\mathbf{D}_j^{(i)}\|_* \quad \text{subject to} \quad \|\mathcal{A}(\mathbf{D}) - \mathbf{b}\|_2 \leq \epsilon, \tag{5.1}$$

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 [2005a], Lee et al. [2010a], 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 [2005a], 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).

5.4 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 [2008a], 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 $4\,\text{km} \times 4\,\text{km}$. Conventional time-sampling interval is 4.0 ms, source- and receiver-sampling interval is 6.25 m. We use a Ricker wavelet with central frequency of 15.0 Hz

as source function. Figure 5.4a 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 5.4b. 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 5.4c. 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 5.4d shows the recovered shot gather, with an SNR of 20.8 dB, and the corresponding residual is shown in Figure 5.4e. As illustrated, we are able to separate the shots along with interpolating the data to the finer grid of 6.25 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 5.4e). 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.

5.5 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.

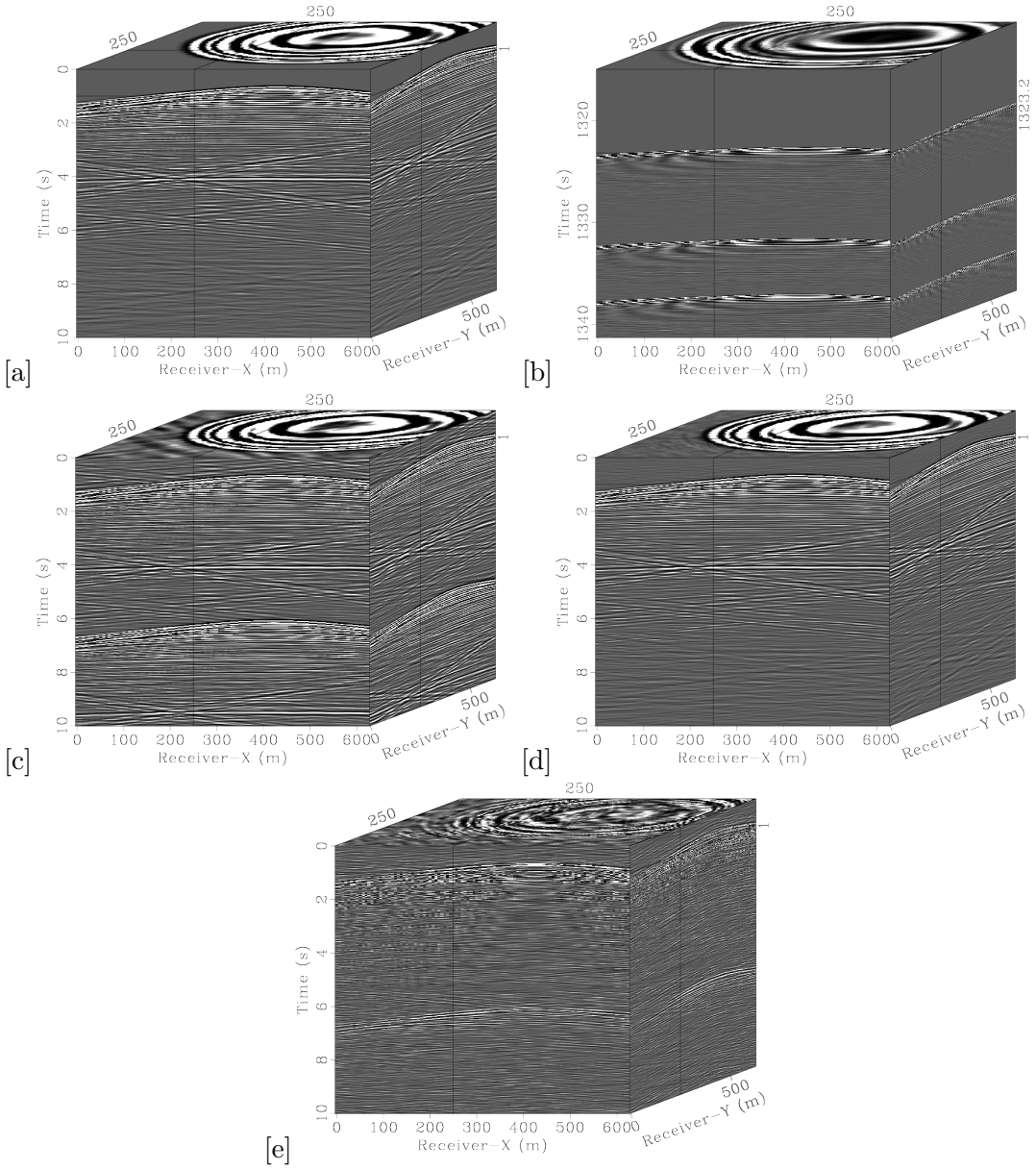


Figure 5.4: Source separation recovery. A shot gather from the (a) conventional data; (b) a section of 30 seconds from the continuous time-domain simultaneous data (**b**); (c) recovered data by applying the adjoint of the sampling operator $\mathcal{M}$; (d) data recovered via the proposed formulation (SNR = 20.8 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 6

Enabling affordable omnidirectional subsurface extended image volumes via probing

6.1 Summary

Image gathers as a function of subsurface offset are an important tool for the inference of rock properties and velocity analysis in areas of complex geology. Traditionally, these gathers are thought of as multidimensional correlations of the source and receiver wavefields. The bottleneck in computing these gathers lies in the fact that one needs to store, compute, and correlate these wavefields for all shots in order to obtain the desired image gathers. Therefore, the image gathers are typically only computed for a limited number of subsurface points and for a limited range of subsurface offsets, which may cause problems in complex geological areas with large geologic dips. We overcome increasing computational and storage costs of extended image volumes by introducing a formulation that avoids explicit storage and removes the customary and expensive loop over shots, found in conventional extended imaging. As a result, we end up with a matrix-vector formulation from which different image gathers can be formed and with which amplitude-versus-angle and wave-equation migration velocity analyses can be performed without requiring prior information on the geologic dips. Aside from demonstrating the formation of two-way extended image gathers for different purposes and at greatly reduced costs, we also present a new approach to conduct automatic wave-equation based migration-velocity analysis. Instead of focussing in particular offset directions and preselected subsets of subsurface points, our method focuses every subsurface point for all subsurface offset directions using a randomized probing technique. As a consequence, we obtain good velocity models at low cost for complex models without the need to provide information on the

A version of this chapter has been published in Geophysical Prospecting, 2016, issn 1365-2478.

geologic dips.

6.2 Introduction

Seismic reflection data are a rich source of information about the subsurface and by studying both dynamic and kinematic properties of the data we can infer both large-scale velocity variations as well as local rock properties. While seismic data volumes – as a function of time, source and receiver positions – contains all reflected and refracted events, it is often more convenient to map the relevant events present in pre-stack data to their respective positions in the subsurface. This stripping away of the propagation effects leads to the definition of an *image volume* or *extended image* – as a function of depth, lateral position and some redundant spatial coordinate(s) – that has the same (or higher) dimensionality as the original data volume. This mapping can be thought of as a coordinate transform, depending on the large-scale background velocity model features, that maps reflection events observed in data collected at the surface to focussed secondary point sources tracing out the respective reflectors. The radiation pattern of these point sources reveals the angle-dependent reflection coefficient and can be used to infer the local rock properties. Errors in the large-scale background velocity model features are revealed through the failure of the events to fully focus. This principle forms the basis of many velocity-model analysis procedures.

Over the past decades various methods have been proposed for computing and exploiting image volumes or slices thereof – the so-called image gathers [Claerbout, 1970, Doherty and Claerbout, 1974, de Bruin et al., 1990, Symes and Carazzone, 1991, ten Kroode et al., 1994, Chauris et al., 2002, Biondi and Symes, 2004, Sava and Vasconcelos, 2011, Koren and Ravve, 2011]. These approaches differ in the way the redundant coordinate is introduced and the method used to transform data volumes into image volumes. Perhaps the most well-known example are normal moveout corrected midpoint gathers, where simple move-out corrections transform observed data into volumes that can be migrated into extended images that are a function of time, midpoint position and surface offset [Claerbout, 1985b].

While this type of surface-offset pre-stack images have been used for migration-velocity analysis and reservoir characterization, these extended images may suffer from unphysical artifacts. For this reason extended images are nowadays formed as functions of the subsurface offset, rather than surface offset as justified by Stolk et al. [2009], who showed that this approach produces artifact-free image gathers in complex geological settings. So far, wave-equation based constructions of these extended images are based on the double-square-root equation [Claerbout, 1970, Doherty and Claerbout, 1974, Biondi et al., 1999, Prucha et al., 1999, Duchkov and Maarten, 2009], which produces image volumes as a function of depth and virtual subsurface source and receiver positions.

With the advent of reverse-time migration [Whitmore et al., 1983, Baysal et al., 1983, Levin, 1984], these one-way, and therefore angle-limited one-way approaches, are gradually being replaced by methods based on the two-way wave-equation, which is able to produce gathers as a function

of depth and horizontal or vertical offset, or both [Biondi and Symes, 2004, Sava and Vasconcelos, 2011]. This is achieved by forward propagating the source and backward propagating the data and subsequently correlating these source and receiver wavefields at non-zero offset/time lag. Depending on the extended imaging conditions [Sava and Biondi, 2004, Sava and Fomel, 2006], various types of subsurface extended image gathers can be formed by restricting the multi-dimensional cross-correlations between the propagated source and receiver wavefields to certain specified coordinate directions. The motivation to depart from horizontal offset only is that steeply dipping events do not optimally focus in the horizontal direction, even for a kinematically correct velocity model. A temporal shift instead of a subsurface offset is sometimes also used as it is more computationally efficient MacKay and Abma [1992], Sava and Fomel [2006].

Up to this point, most advancements in image-gather based velocity-model building and reservoir characterization are due to improved sampling of the extended images and/or a more accurate wave propagators. However, these improvements carry a heavy price tag since the computations typically involve the solution of the forward and adjoint wave equation for each shot, followed by subsequent cross-correlations. As a result, computational and memory requirements grow uncontrollably and many practical implementations of subsurface offset images are therefore restricted, (e.g., by allowing only for lateral interaction over a short distance), and the gathers are computed for a subset of image points only. Unfortunately, forming full subsurface extended image volumes rapidly becomes prohibitively expensive even for small two-dimensional problems.

Extended images play an important role in wave-equation migration-velocity analysis (WEMVA), where velocity model updates are calculated by minimizing an objective function that measures the coherency of image gathers [Symes and Carazzone, 1991, Shen and Symes, 2008]. Regrettably, the computational and storage costs associated with this approach easily becomes unmanageable unless we restrict ourselves to a few judiciously chosen subsurface points [Yang and Sava, 2015]. As a result, we end up with an expensive method, which needs to loop over all shot records and allowing for the computation of subsurface offset gathers for a limited number of subsurface points and directions. This is problematic because the effectiveness of WEMVA is impeded for dipping reflectors that do not focus as well along horizontal offsets.

Extended images also serve as input to amplitude-versus-angle (AVA) or amplitude-versus-offset (AVO) analysis methods that derive from the (linearized) Zoeppritz equations [Aki and Richards, 1980]. In this situation, the challenge is to produce reliable amplitude preserving angle/offset gathers in complex geological environments (see e.g., de Bruin et al. [1990], van Wijngaarden [1998]) from primaries only or, more recently from surface-related multiples, as demonstrated by Lu et al. [2014]. In the latter case, good angular illumination can be obtained from conventional marine towed-streamer acquisitions with dense receiver sampling. As WEMVA, these amplitude analysis are biased when reflectors are dipping, calling for corrections dependent on geological dips that are generally unknown [Brandsberg-Dahl et al., 2003].

We present a new wave-equation based factorization principle that removes computational bot-

tlenecks and gives us access to the kinematics and amplitudes of full subsurface offset extended images without carrying out explicit cross-correlations between source and receiver wavefields for each shot. We accomplish this by carrying out the correlations implicitly among all wavefields via actions of full extended image volumes on certain probing/test vectors. Depending on the choice of these probing vectors, these actions allow us to compute common-image-point (CIPs), common-image- (CIGs) and dip-angle gathers at certain predefined subsurface points in 2- and 3-D as well as objective functions for WEMVA. None of these require storage of wavefields and loops over shots.

The paper proceeds as follows. After introducing the governing equations of continuous-space monochromatic extended image volumes, their relation to migrated images, CIPs and CIGs, we derive the two-way equivalent of the *double square-root equation* and move to a discrete setting that reveals a wave-equation based factorization of discrete full extended image volumes. Next, we show how this factorization can be used to extract information on local geological dips and radiation patterns for AVA purposes and how full offset extended image volumes can be used to carry out automatic WEMVA. Each application is illustrated by carefully selected stylized numerical examples in 2- and 3-D.

6.2.1 Notation

In this paper, we use lower/upper case letters to represents scalars, e.g., x_m, R, and lower case boldface letters to represents vectors (i.e. one-dimensional quantities), e.g., $\mathbf{x}, \mathbf{x}'$. We denote matrices and tensors using upper case boldface letters, e.g., $\mathbf{E}, \widetilde{\mathbf{E}}, \mathbf{R}$. Subscript i represents frequencies, j represents the number of sources and/or receivers and k represents the subsurface grid points. 2-D represents seismic volumes with one source, one receiver, and time dimensions. 3-D represents seismic volumes with two sources, two receivers, and time dimensions.

6.3 Anatomy & physics

Before describing the proposed methodology of extracting information from full subsurface-offset extended image volumes using the proposed probing technique, we first review the governing equations. We denote the full-extended image volumes as

$$e(\omega, \mathbf{x}, \mathbf{x}') = \int_{\mathcal{D}_s} \mathrm{d}\mathbf{x}_s \; u(\omega, \mathbf{x}, \mathbf{x}_s)\overline{v(\omega, \mathbf{x}', \mathbf{x}_s)}, \tag{6.1}$$

where the overline denotes complex conjugation, u and v are source and receiver wavefields as a function of the frequency $\omega \in \Omega \subset \mathbb{R}$, subsurface positions $\mathbf{x}, \mathbf{x}' \in \mathcal{D} \subset \mathbb{R}^n$ ($n = 2$ or 3) and source positions $\mathbf{x}_s \in \mathcal{D}_s \subset \mathbb{R}^{n-1}$. These monochromatic wavefields are obtained by solving

$$H(m)u(\omega, \mathbf{x}, \mathbf{x}_s) = q(\omega, \mathbf{x}, \mathbf{x}_s), \tag{6.2}$$

$$H(m)^* v(\omega, \mathbf{x}, \mathbf{x}_s) = \int_{\mathcal{D}_r} \mathrm{d}\mathbf{x}_r \; d(\omega, \mathbf{x}_r, \mathbf{x}_s)\delta(\mathbf{x} - \mathbf{x}_r), \tag{6.3}$$

where $H(m) = \omega^2 m(\mathbf{x})^2 + \nabla^2$ is the Helmholtz operator with Sommerfeld boundary conditions, the symbol * represents the conjugate-transpose (adjoint), m is the squared slowness, q is the source function and d represents the reflection-data at receiver positions $\mathbf{x}_r \in \mathcal{D}_r \subset \mathbb{R}^{n-1}$.

Equations (6.1-6.3) define a linear mapping from a $2n-1$ dimensional data volume $d(\omega, \mathbf{x}_r, \mathbf{x}_s)$ to a $2n+1$ dimensional image volume $e(\omega, \mathbf{x}, \mathbf{x}')$. Conventional migrated images are obtained by applying the zero-time/offset imaging condition [Sava and Vasconcelos, 2011]

$$r(\mathbf{x}) = \int_\Omega \mathrm{d}\omega \; e(\omega, \mathbf{x}, \mathbf{x}). \tag{6.4}$$

Once we have the full extended image, we can extract all conceivable image and angle gathers by applying appropriate imaging conditions [Sava and Vasconcelos, 2011]. For instance, conventional 2-D common-image-gathers (CIGs), as a function of lateral positions x_m, depth z and horizontal offset h_x are defined as

$$I_{\mathrm{CIG}}(z, h; x_m) = \int_\Omega \mathrm{d}\omega \; e\left(\omega, (x_m - h_x, z)^T, (x_m + h_x, z)^T\right), \tag{6.5}$$

where ^{T} denotes the transpose.

Similarly, time-shifted common-image-point (CIP) gathers at subsurface points $\mathbf{x}_k$, as a function of all spatial offset vectors $\mathbf{h}$ and temporal shifts Δt becomes

$$I^{\mathrm{ext}}_{\mathrm{CIP}}(\mathbf{h}, \Delta t; \mathbf{x}_k) = \int_\Omega \mathrm{d}\omega \; e(\omega, \mathbf{x}_k, \mathbf{x}_k + \mathbf{h}) e^{\imath\omega\Delta t}, \tag{6.6}$$

where $\imath = \sqrt{-1}$. Note that we departed in this expression from the usual symmetric definition $e(\omega, \mathbf{x}_k - \mathbf{h}, \mathbf{x}_k + \mathbf{h})$ because it turns out to be more natural for our shot-based computations. Because we consider these gathers of full-subsurface offset only, we apply a zero-time imaging condition to the CIPs–i.e., $I_{\mathrm{CIP}}(\mathbf{h}; \mathbf{x}_k) = I^{\mathrm{ext}}_{\mathrm{CIP}}(\mathbf{h}, \Delta t = 0; \mathbf{x}_k)$. These gathers form the theoretical basis for AVA and WEMVA.

Moving to a discrete setting, we have N_s sources $\{\mathbf{x}_{s,j}\}_{j=1}^{N_s} \in \mathcal{D}_s$, N_r receivers $\{\mathbf{x}_{r,j}\}_{j=1}^{N_r} \in \mathcal{D}_r$, N_f frequencies $\{\omega_i\}_{i=1}^{N_f} \in \Omega$ and discretize the domain $\mathcal{D}$ using a rectangular grid with a total of N_x grid points $\{\mathbf{x}_k\}_{k=1}^{N_x}$. We organize the source and receiver wavefields in tensors of size $N_f \times N_x \times N_s$ with elements $u_{ikj} \equiv u(\omega_i, \mathbf{x}_k, \mathbf{x}_{s,j})$ and $v_{ikj} \equiv v(\omega_i, \mathbf{x}_k, \mathbf{x}_{s,j})$. For the i^{th} frequency, we can represent the wavefields for all sources as complex valued $N_x \times N_s$ matrices $\mathbf{U}_i$ and $\mathbf{V}_i$, where each column of the matrix represents a monochromatic source experiment.

Full $2n+1$ dimensional image volumes can now be represented as a 3-D tensor with elements $e_{ikk'} \equiv e(\omega_i, \mathbf{x}_k, \mathbf{x}_{k'})$. A slice through this tensor at frequency i is an $N_x \times N_x$ matrix, which can be expressed as an outer product of the matrices $\mathbf{U}_i$ and $\mathbf{V}_i$

$$\mathbf{E}_i = \mathbf{U}_i \mathbf{V}_i^*, \tag{6.7}$$

where * denotes the complex conjugate transpose. The matrix $\mathbf{E}_i$ is akin to Berkhout's *reflectivity matrices* [Berkhout, 1993], except that $\mathbf{E}_i$ captures vertical interactions as well since it is derived from the two-way wave-equation. These source and receiver wavefields obey the following discretized Helmholtz equations:

$$\mathbf{H}_i(\mathbf{m})\mathbf{U}_i = \mathbf{P}_s^*\mathbf{Q}_i, \tag{6.8}$$

$$\mathbf{H}_i(\mathbf{m})^*\mathbf{V}_i = \mathbf{P}_r^*\mathbf{D}_i, \tag{6.9}$$

where $\mathbf{H}_i(\mathbf{m})$ represents the discretized Helmholtz operator for frequency ω_i with absorbing boundary conditions and for a gridded squared slowness $\mathbf{m}$. The $N_s \times N_s$ matrix $\mathbf{Q}_i$ represents the sources (i.e., each column is a source function), the $N_r \times N_s$ matrix $\mathbf{D}_i$ contains reflection-data (i.e., each column is a monochromatic shot gather after subtraction of the direct arrival) and the matrices $\mathbf{P}_s, \mathbf{P}_r$ sample the wavefields at the source and receiver positions (and hence, their transpose injects the sources and receivers into the grid). Remark that these are the discretized versions of equations (6.2, 6.3).

Substituting relations (6.8, 6.9) into the definition of the extended image (6.7) yields

$$\mathbf{H}_i\mathbf{E}_i\mathbf{H}_i = \mathbf{P}_s^*\mathbf{Q}_i\mathbf{D}_i^*\mathbf{P}_r. \tag{6.10}$$

This defines a natural 2-way analogue of the well-known double-square-root equation. From equation (6.10), we derive the following expression for monochromatic full extended image volumes

$$\mathbf{E}_i = \mathbf{H}_i^{-1}\mathbf{P}_s^*\mathbf{Q}_i\mathbf{D}_i^*\mathbf{P}_r\mathbf{H}_i^{-1}, \tag{6.11}$$

which is a discrete analogue of the linear mapping from data to image volumes defined in equations (6.1-6.3). Note that for co-located sources and receivers ($\mathbf{P}_r = \mathbf{P}_s$) and ideal discrete point sources ($\mathbf{Q}$ is the identity matrix) we find that $\mathbf{E}$ is complex symmetric (i.e., $\Re(\mathbf{E}^*) = \Re(\mathbf{E})$ and $\Im(\mathbf{E}^*) = -\Im(\mathbf{E})$) because of source-receiver reciprocity.

The usual zero-time/offset imaging conditions translate to

$$\mathbf{r} = \sum_{i=1}^{N_f} \mathrm{diag}(\mathbf{E}_i), \tag{6.12}$$

where $\mathbf{r}$ is the discretized reflectivity and $\mathsf{diag}(\mathbf{A})$ denotes the diagonal elements of $\mathbf{A}$ organized in a vector. Various image gathers are embedded in our extended image volumes as illustrated in Figure 6.1.

As we will switch between continuous and discrete notation throughout the paper, the correspondence between the image volume $e(\omega, \mathbf{x}, \mathbf{x}')$ and the matrices $\mathbf{E}_i$ is listed in Table 6.1. For simplicity, we will drop the frequency-dependence from our notation for the remainder of the pa-

per and implicitly assume that all quantities are monochromatic with the understanding that all computations can be repeated as needed for multiple frequencies. We will also assume that the zero-time imaging condition is applied by summing over the frequencies, followed by taking the real part. We further note that full extended image volumes can be severely aliased in case of insufficient source-receiver sampling. Therefore, one would in practice only extract gathers from image volumes in well-sampled directions.

6.4 Computational aspects

Of course, we can never hope to explicitly form the complete image volumes owing to the enormous computational and storage costs associated with these volumes that are quadratic in the number of grid points N_x. In particular, we will discuss the computation of monochromatic image volumes $\mathbf{E}_i$ and drop the subscript i in the remainder of the section.

To avoid forming extended images $\mathbf{E}$ explicitly, we instead propose to probe these volumes by right-multiplying them with $N_x \times K$ sampling matrices $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_K]$, where K denotes the number of samples and $\mathbf{w}_k$ denotes a single probing or sampling vector. After sampling, the reduced image volume $\widetilde{\mathbf{E}}$ now reads

$$\widetilde{\mathbf{E}} = \mathbf{E}\mathbf{W} = \mathbf{H}^{-1}\mathbf{P}_s^*\mathbf{Q}\mathbf{D}^*\mathbf{P}_r\mathbf{H}^{-1}\mathbf{W}. \tag{6.13}$$

Our main contribution is that we can compute these compressed volumes efficiently with algorithm 1 or 2. As one can see, these computations derive from wave-equation based factorizations that avoid storage and loops over all shots.

Algorithm 1 Compute matrix-vector multiplication of image volume matrix with given vectors $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_K]$. The computational cost is $2N_s$ wave-equation solves plus the cost of correlating the wavefields.

---

compute all the source wavefields $\mathbf{U} = \mathbf{H}^{-1}\mathbf{P}_s^*\mathbf{Q}$,
compute all the receiver wavefields $\mathbf{V} = \mathbf{H}^{-*}\mathbf{P}_r^*\mathbf{D}$
compute weights $\widetilde{\mathbf{Y}} = \mathbf{V}^*\mathbf{W}$
compute the product $\widetilde{\mathbf{E}} = \mathbf{U}\widetilde{\mathbf{Y}}$.

Algorithm 2 Compute matrix-vector multiplication of image volume matrix with given vectors $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_K]$. The computational cost is $2K$ wave-equation solves plus the cost of correlating the data matrices.

---

compute $\widetilde{\mathbf{U}} = \mathbf{H}^{-1}\mathbf{W}$ and sample this wavefield at the receiver locations $\widetilde{\mathbf{D}} = \mathbf{P}_r\widetilde{\mathbf{U}}$;
correlate the result with the data $\widetilde{\mathbf{W}} = \mathbf{D}^*\widetilde{\mathbf{D}}$ to get the source weights;
use the source weights to generate the simultaneous sources $\widetilde{\mathbf{Q}} = \mathbf{Q}\widetilde{\mathbf{W}}$;
compute the resulting wavefields $\widetilde{\mathbf{E}} = \mathbf{H}^{-1}\mathbf{P}_s^*\widetilde{\mathbf{Q}}$.

While both algorithms produce the same compressed image volume, Algorithm 1 corresponds

to the traditional way of computing image volumes where *all* source and receiver wavefields are computed first and subsequently cross-correlated. Algorithm 2, on the other hand, produces *exactly* the same result without looping over all shots and without carrying out explicit cross-correlations. It arrives at this result by cross-correlating the data and source wavefields and by solving only two wave-equations per probing vector instead of solving the forward and adjoint wave-equations for each shot. This means that our method (Algorithm 2) gains computationally as long as the number of probings is small compared to the number of sources ($K < N_s$). The choice of the probing vectors depends on the application.

For completeness, we summarized the computational complexities of the two schemes in Table 6.2 in terms of the number of sources N_{s}, receivers N_{r}, subsurface sample points N_x and desired number of subsurface offsets in each direction $N_{h_{\{x,y,z\}}}$.

6.5 Case study 1: computing gathers

We now describe how common-image-point gathers (CIPs) and common-image-gathers (CIGs) can be formed with Algorithm 2.

CIPs: According to Algorithm 2, a CIP at midpoint $\mathbf{x}_k$ can be computed efficiently by extracting the corresponding column from the monochromatic matrix $\mathbf{E}$ representing the full extended image. We achieve this at the cost of only two wave-equation solves by setting $\mathbf{w}_k$ to a cardinal basis vector with its non-zero entry corresponding to the location of a single point source at $\mathbf{x}_k$. Because of the proposed factorization, the number of required wave-equation solves is reduced from twice the number of shots to only two per subsurface point, representing an order-of-magnitude improvement. As long as the number of subsurface points are not too large, this reduction allows for targeted quality control with omnidirectional extended image gathers.

For reasonable background velocity models, we can even compute these image gathers simultaneously as long as the corresponding subsurface points are spatially separated. In this case, the probing vectors $\mathbf{w}_k$ correspond to simultaneous subsurface sources without encoding. As a result, we may introduce cross-talk in the gathers, but expect that there will be very little interference when the locations being probed are sufficiently far away from each other. Even though this cross-talk may interfere with the ability to visually inspect the CIPs, we will see that these interferences can be rendered into incoherent noise via randomized source encoding as shown by van Leeuwen et al. [2011] for full waveform inversion (FWI), a property we will later use in automated migration-velocity analyses.

CIGs We can also extract common-image-gather at a lateral position x_k by (densely) sampling the image volumes at $\mathbf{x}_k = (x_k, z_k)^T$ at all depth levels. In this configuration, the sampling matrix takes the form $\mathbf{W} = \mathbf{W}_x \otimes \mathbf{W}_z$, where $\mathbf{W}_x = (0, 0, \ldots, 1, \ldots, 0)^T$ is a cardinal basis vector as

before with a single non-zero entry located at the index corresponding to the midpoint position x_k and $\mathbf{W}_z = \mathbf{I}$ is the identity matrix, sampling all grid points in the vertical direction. The resulting gathers $\widetilde{\mathbf{E}} = \mathbf{EW}$ are now a function of $(z, \Delta x, \Delta z)$ and contain both vertical and lateral offsets. We can form conventional image gathers by extracting a slice of the volume at $\Delta z = 0$. The computational cost per CIG in 2-D are roughly the same as the conventionally computed CIGs; $2N_z$ vs. $2N_s$ wave-equation solves, where N_z denotes the number of samples in depth. In 3-D however, the proposed way of computing the gathers via algorithm 2 is a order of magnitude faster because the number of sources in a 3-D seismic acquisition is an order of magnitude bigger while the number of samples in depth stays the same.

As with the CIPs, we can generate these gathers simultaneously, by sampling various lateral positions at the same time for each depth level.

6.5.1 Numerical results in 2-D

To illustrate the discussed methodology, we compute various gathers on a central part of Marmousi model. We use a grid spacing of 10 m, 81 co-located sources and receivers with 50 m spacing and frequencies between 5 and 25 Hz with 0.5 Hz spacing. The source wavelet is a Ricker wavelet with a peak frequency of 15 Hz. The wavefields are modeled using a 9-point finite difference discretization of the Helmholtz operator with a sponge boundary condition. The direct wave is removed prior to computing the image gathers.

Figure 6.2 shows the migrated image for wrong (a) and correct background velocity models (b). At three locations indicated by the $*$ symbol we extract the CIP gathers, shown in (c) for a background velocity model that is too low and in (d) for the correct velocity. As a result of the cross-correlations between various events in propagated wavefields several spurious events are present so the recovered CIP gathers are only meaningful for interpretation close to the image point. However, as expected most of the energy concentrates along the normal to the reflector. We can generate these three CIPs *simultaneously* at the cost of generating one CIP by defining the sampling vector $\mathbf{w}_k$ to represent the three point sources simultaneously. As we mentioned before, this may result in cross-talk between the CIPs, however, in this case the events do not significantly interfere because the events are separated laterally, as shown in Figure 6.2 (c,d).

To illustrate the benefits of the proposed scheme, we also report the computational time (in sec) and memory (in MB) required to compute a single common-image point gather using Algorithm 1 and 2. The results are shown in Table 6.3. We can see that even for a small toy model, the probing technique reduces the computational time and memory requirement by a factor of 20 and 30, respectively.

Finally, Figures 6.2 (e-f) contain CIGs at two lateral positions for a background velocity that is too low (e) and correct (f). In this example, we generated the image volume for each depth level simultaneously for all lateral positions. When creating many gathers, such a simultaneous probing

can substantially reduce the required computational cost.

6.5.2 Numerical results in 3-D

Notwithstanding the achieved speedup and memory reduction in 2-D, the proposed probing method outlined in Algorithm 2 is a true enabler in 3-D where there is no realistic hope to store full extended image volumes whose size is quadratic in the number of grid points N_x. Besides, the number of sources also becomes quadratic for full-azimuth acquisitions. As in the 2-D case, our probing technique is a key enabler allowing us to compute CIPs without allocating exorbitant amounts of memory and computational resources. To illustrate our claim, we compute a single CIP for the Compass velocity model provided to us by BG Group. Figures 6.3, 6.4 contains both vertical and lateral cross-sections of this complicated 3D velocity model, which contains $131 \times 101 \times 101$ grid points. The model is 780 m deep and 2.5 km across in both lateral directions.

We generated data from this velocity model following an ocean-bottom node configuration where sources are placed at the water surface along the x and y directions with a sampling interval of 75 m. The receivers are placed at the sea bed with a sampling of 50 m resulting in a data volume with 1156 sources and 2601 receivers. We generated this marine data volume with a 3-D time-harmonic Helmholtz solver, based on a 27 point discretization, perfectly-matching boundary conditions, and a Ricker wavelet with a central frequency of 15 Hz. During the simulations and imaging, we used 15 frequencies ranging from 5 to 12 Hz with a sampling interval of 0.5 Hz. For further details on the employed wave-equation solver, we refer to Lago et al. [2014], van Leeuwen and Herrmann [2014].

Figures 6.3b and 6.5 show an example of a full CIP gather extracted at $z = 390$, $x = 1250$ and $y = 1250$ m. Aside from behaving as expected, the observed running time with Algorithm 2 is 1500 times faster compared to the corresponding time needed by Algorithm 1. While Algorithm 1 and 2 lend themselves well for parallelization over frequencies, shots, and probing vectors, our method avoids a loop over 1156 shots while avoiding explicit formation of extended image that would require the allocation of a matrix with 101^6 entries.

6.6 Case study 2: dip-angle gathers

Aside from their use for kinematical quality control during velocity model building, common-image gathers (CIGs) also contain dynamic amplitude-versus-angle (AVA) information on reflecting interfaces that can serve to invert associated rock properties [Mahmoudian and Margrave, 2009]. For this purpose, various definitions of angle-domain CIGs have been proposed [de Bruin et al., 1990, van Wijngaarden, 1998, Rickett and Sava, 2002, Sava and Fomel, 2003, Kuhel and Sacchi, 2003, Biondi and Symes, 2004, Sava and Vasconcelos, 2011]. Extending the work of de Bruin et al. [1990] to include corrections for the geological dip θ, we extract angle-dependent reflection

coefficients from a subsurface point $\mathbf{x}_0$ by evaluating the following integral:

$$R(\mathbf{x}_0, \alpha; \theta) \propto \int_\Omega \mathrm{d}\omega \int_{-h_{\max}}^{h_{\max}} \mathrm{d}h \;\; e(\omega, \mathbf{x}_0, \mathbf{x}_0 + h\mathbf{n}(\theta)^\perp) e^{i\omega \sin(\alpha) h / v(\mathbf{x})} \tag{6.14}$$

over frequencies and offsets. We use the $\propto$ symbol to indicate that this expression holds up to proportionality constant.

In this expression, α is the angle of incidence with respect to the normal (see Figure 6.6), $v(\mathbf{x})$ is the local background velocity used to convert subsurface offsets to angles, and $\mathbf{n}(\theta)^\perp$ denotes the tangent vector to the reflector defining the offset vector in this direction ($h\mathbf{n}(\theta)^\perp$). The integral is carried out over the effective offset range denoted by $h_{\max}$, which decreases for deeper parts of the model. The integral over frequencies correspond to the zero-time imaging condition. As illustrated in Figure 6.6, the normal $\mathbf{n}(\theta) = (\sin\theta, \cos\theta)^T$ and tangent vectors to a reflecting interface depend on the geological dip τ, which is unknown in practice. Unfortunately, ignoring this factor may lead to erroneous amplitudes in cases where this dip is steep. This means that we need to extract CIGs, following the procedure outlined above, as well as estimates on the geological dip from the extended image volumes.

Following Brandsberg-Dahl et al. [2003], we use the stack power to estimate the geological dip at subsurface point $\mathbf{x}_k$ via

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmax}} \int_{-h_{\max}}^{h_{\max}} \mathrm{d}h \;\; \left| \int_\Omega \mathrm{d}\omega \; e(\omega, \mathbf{x}, \mathbf{x}_k + h\mathbf{n}(\theta)) \right|^2 . \tag{6.15}$$

This maximization is based on the assumption that the above integral attains a maximum value when we collect energy from the image volume at time zero and along a direction that corresponds to the true geologic dip. Both integrals in equation 6.14 and 6.15 require information on monochromatic extended image volumes at subsurface position $\mathbf{x}_k$ only–i.e., $e(\omega, \mathbf{x}_k, \mathbf{x}')$ for $\omega \in \Omega$, to which we have readily access via probing as described in Algorithm 2. The resulting gathers $e(\omega, \mathbf{x}_k, \mathbf{x}')$ contain the required information to estimate local geologic-dip corrected angle-dependent reflection coefficient $R(\alpha; \mathbf{x}_k, \widehat{\theta})$ where the correction is carried with dip estimate $\widehat{\theta}$ that maximizes the stack power. Since we have access to all subsurface offsets at no additional computational costs, there is no need to select a maximum subsurface offsets $h_{\max}$ priori even though the effective subsurface offset decreases with depth due to the finite aperture in the seismic data acquisition.

6.6.1 Numerical results

To illustrate the proposed method of computing angle-domain common-image gathers (CIGs), we compare the modulus for plotting reasons of the estimated reflection coefficients $|R(\alpha; \mathbf{x}_k, \theta_k)|$ with the theoretical PP-reflection coefficients predicted by the Zoeppritz equations [Koefoed, 1955,

Shuey, 1985].

To make this comparison, we used a finite-difference time-domain acoustic modelling code [Symes et al., 2011b] to generate three synthetic data sets for increasingly complex models, namely a two-layer velocity and constant density model (Figures 6.7 (a, b)); a four layer model with properties taken from de Bruin et al. [1990] (Figures 6.8 (a,b)), and a two layer lateral varying velocity and density model–i.e., one-horizontal reflector and one-dipping reflector (Figures 6.9(a, b)). The purpose of these three experiments is to (i) verify the velocity-change-only angle dependent reflection coefficients; (ii) study the effects of density and decreasing effective horizontal offset with depth; and (iii) illustrate the effect of the geologic dip on the reflection coefficients. We used a Ricker wavelet with a peak frequency of 15 Hz as a source signature. In all three examples, seismic data is simulated using split-spread acquisition. The gathers used in the AVA analyses are obtained using Algorithm 2 and discrete version of equations 6.14 and 6.15 for a smoothed version of the layered velocity models.

The estimated angle-dependent reflection coefficients for the first model is displayed in Figure 6.7 (c). We can see that these estimates, extracted from our extended image volumes, match the theoretical reflection coefficients according to the Zoeppritz equations after a single amplitude scaling fairly well up to angles of 50° that are reasonably close to effective aperture angle (depicted by the black line in the Figures of the AVA curves). Beyond these angles, the Zoeppritz equations are no longer accurate.

The results for the deeper four layer model depicted in Figures 6.8 (c, d, e, f) clearly show the imprint of smaller horizontal offsets with depth.

From the AVA plots in Figure 6.8, we can observe that the reflection coefficients for the first and second reflector are well matched up to 50° and 40° and, for the third and fourth reflectors, to only 20°. We also confirm that that the angle-dependent reflection coefficient associated with a changes in density is approximately flat (see Figure 6.8 (e)). The finite aperture of the data accounts for the discrepancy beyond these angles.

To illustrate our method's ability to correct for the geological dip, we consider a common-image-point (CIP) gather at $x = 2250$ m and $z = 960$ m. As we can see from Figure 6.9 (c), this CIP is well focused and with the local geologic dip of the reflector. The corresponding stack power is plotted in Figure 6.9 (d), which attains maximum power at $\widehat{\theta} = 10.8°$ an estimate close to the actual dip of 11°. To demonstrate the effect of ignoring the geological dip when computing angle gathers, we evaluate equation 6.14 for $\theta = 0$. The results included in Figure 6.10 clearly illustrate the benefit of incorporating the dip-information in angle-domain image gathers as it allows for a more accurate estimation of the angle-dependent reflection coefficients. This experimental result supports our claim that proposed extended image volume framework lends itself well to estimate geologic-dip corrected angle gathers. As we mentioned before, the proposed method also reaps the computational benefits of probing extended image volumes as outlined in Algorithm 2.

6.7 Case study 3: wave-equation migration-velocity analysis (WEMVA)

Aside from providing localized information on the kinematics and dynamics, full subsurface offset extended image volumes also lend themselves well for automatic velocity analyses that minimize some global focusing objective [Symes and Carazzone, 1991, Biondi and Symes, 2004, Shen and Symes, 2008, Sava and Vasconcelos, 2011, Mulder, 2014, Yang and Sava, 2015]. For this reason, wave-equation migration-velocity analysis (WEMVA) can be considered as another important application where subsurface image gathers are being used extensively. In this case, the aim is to build kinematically correct background velocity models that either promote similarity amongst different surface offset gathers, as in the original work by Symes and Carazzone [1991] on Differential Semblance, or that aim to focus at zero subsurface offset, an approach promoted by recent work on WEMVA [Brandsberg-Dahl et al., 2003, Shen and Symes, 2008, Symes, 2008b]. While these wave-equation based methods are less prone to imaging artifacts Stolk and Symes [2003], they are computationally expensive restricting the number of directions in which the subsurface offsets can be calculated. This practical limitation may affect our ability to handle unknown geological dips [Sava and Vasconcelos, 2011, Yang and Sava, 2015].

Before presenting an alternative approach that overcomes these practical limitations, let us first formulate an instance of WEMVA, based on the probing technique outlined in Algorithm 2. In this discrete case, WEMVA corresponds to minimizing

$$\min_{\mathbf{m}} \sum_{k=1}^{N_x^2} \|\mathbf{S}_k \overline{\mathbf{E}}(\mathbf{m}) \mathbf{w}_k\|_2^2, \tag{6.16}$$

with $\overline{\mathbf{E}}(\mathbf{m}) = \sum_{i \in \Omega} \mathbf{E}_i(\mathbf{m})$. As before the $\mathbf{E}_i(\mathbf{m})$'s denote monochromatic extended image volumes for the background velocity model $\mathbf{m}$. With this definition, we absorb both the zero-time imaging condition, by summing over frequencies. The vector $\mathbf{w}_n$ represents a point-source at a subsurface point corresponding to the k^{th} entry of $\mathbf{w}$. We compute this sum efficiently via first probing the each monochromatic extended image volumes with the vector $\mathbf{w}_k$ followed by sum over frequency instead of summing over the monochromatic full subsurface image volumes followed by probing with the vector $\mathbf{w}_k$. Finally, the diagonal matrix $\mathbf{S}_k$ penalizes defocussed energy by applying a weighting function. Often, this weight is chosen proportional to the lateral subsurface offset [Shen and Symes, 2008]. The main costs of this approach (cf. equation 6.16) are formed by the number of gathers, which equals the number of grid points. Of course, WEMVA is conducted in practice with a subset of grid points N_x, the number of which depends on the complexity of the subsurface.

To arrive at an alternative more cost effective formulation that offers flexibility to focus in all offset directions, we take a different tack by using the fact that focussed extended image volumes commute with diagonal weighting matrices [Kumar et al., 2013b, Symes, 2014] that penalize off-

diagonal energy–i.e., we have

$$\overline{\mathbf{E}}\,\mathrm{diag}(\mathbf{s}) \approx \mathrm{diag}(\mathbf{s})\,\overline{\mathbf{E}} \tag{6.17}$$

for a given weighting vector $\mathbf{s}$. For the specific case where the entries in $\mathbf{s}$ correspond to lateral positions for each grid point in the model, forcing the above commutation relation corresponds to the objective of equation 6.16 with weights proportional to the horizontal subsurface offset. However, other options are also available including focusing in all offset directions. The optimization problem can now be written as

$$\min_{\mathbf{m}} \left\{ \phi(\mathbf{m}) = ||\overline{\mathbf{E}}(\mathbf{m})\mathrm{diag}(\mathbf{s}) - \mathrm{diag}(\mathbf{s})\overline{\mathbf{E}}(\mathbf{m})||_F^2 \right\}, \tag{6.18}$$

where $||\mathbf{A}||_F^2 = \sum_{i,j} a_{i,j}^2$ denotes the Frobenius norm. Minimization of this norm forces $\overline{\mathbf{E}}$ to focus as a function of the velocity model $\mathbf{m}$.

For obvious reasons, this formulation is impractical because even in 2-D we can not hope to store the extended image volumes $\overline{\mathbf{E}}$, whose size is quadratic in the number of grid points – $\overline{\mathbf{E}}$ is a $N_x^2 \times N_x^2$ matrix. However, we can still minimize the above objective function via random-trace estimation [Avron and Toledo, 2011], a technique that also underlies phase encoding techniques in full-waveform inversion van Leeuwen et al. [2011]. With this technique, equation 6.18 can be evaluated to arbitrary accuracy via actions of $\overline{\mathbf{E}}$ on random vectors $\mathbf{w}$. With this approximation, the WEMVA objective becomes

$$\phi(\mathbf{m}) \approx \widetilde{\phi}(\mathbf{m}) = \frac{1}{K}\sum_{k=1}^{K} \|\mathbf{R}(\mathbf{m})\mathbf{w}_k\|_2^2, \tag{6.19}$$

where $\mathbf{R}(\mathbf{m}) = \overline{\mathbf{E}}\,(\mathbf{m})\mathrm{diag}(\mathbf{s}) - \mathrm{diag}(\mathbf{s})\,\overline{\mathbf{E}}(\mathbf{m})$. While other choices are possible, we select the $\mathbf{w}_k$ as Gaussian vectors with independent, identically distributed random entries with zero mean and unit variance. For this choice, $\widetilde{\phi}(\mathbf{m})$ and $\phi(\mathbf{m})$ are equal in expectation, which means that the above sample average is unbiased [van Leeuwen et al., 2011, Haber et al., 2012]. As we know from source-encoding in full-waveform inversion [Krebs et al., 2009, van Leeuwen et al., 2011, Haber et al., 2012], good approximations can be obtained for small sample size $K \ll N_x$. We will study the quality of this approximation below.

The gradient of the approximate objective is given by

$$\nabla\widetilde{\phi}(\mathbf{m}) = \frac{1}{K}\sum_{k=1}^{K}\big(\nabla\overline{\mathbf{E}}(\mathbf{m}, \mathrm{diag}(\mathbf{s})\mathbf{w}_k) - \mathrm{diag}(\mathbf{s})\nabla\overline{\mathbf{E}}(\mathbf{m}, \mathbf{w}_k)\big)^*\mathbf{R}(\mathbf{m})\mathbf{w}_k, \tag{6.20}$$

where

$$\nabla\overline{\mathbf{E}}(\mathbf{m}, \mathbf{y}) = \frac{\partial\overline{\mathbf{E}}(\mathbf{m})\mathbf{y}}{\partial\mathbf{m}}$$

is the Jacobian of $\overline{\mathbf{E}}(\mathbf{m})\mathbf{y}$. We do not form this Jacobian matrix explicitly, but instead compute its action on a vector as follows:

$$\nabla\overline{\mathbf{E}}(\mathbf{m},\mathbf{y})\delta\mathbf{m} \quad = \quad -\omega^2\left(\overline{\mathbf{E}}(\mathbf{m})\text{diag}(\widetilde{\mathbf{y}}) + \mathbf{H}(\mathbf{m})^{-1}\text{diag}(\widetilde{\mathbf{e}})\right)\delta\mathbf{m} \tag{6.21}$$

where $\widetilde{\mathbf{w}} = \mathbf{H}(\mathbf{m})^{-1}\mathbf{y}$, $\widetilde{\mathbf{e}} = \overline{\mathbf{E}}(\mathbf{m})\mathbf{y}$. The computation of the action of the adjoint of the Jacobian follows naturally.

We can now employ an iterative gradient-based method to find a minimizer of $\phi(\mathbf{m})$ by using a K-term approximation of this objective and its gradient [van Leeuwen et al., 2011, Haber et al., 2012]. To remove possible bias from using a fixed set of random probing vectors, we redraw these vectors after each gradient update. Remaining errors can be controlled by increasing K [Haber et al., 2012, Friedlander and Schmidt, 2012, van Leeuwen and Herrmann, 2014].

6.7.1 Numerical results

We test the proposed wave-equation migration-velocity analysis (WEMVA) formulation on synthetic examples, illustrating the efficacy of the probing techniques. In all experiments, we use a 9-point finite-difference discretization of the 2-D Helmholtz equation to simulate the wavefields. The direct wave is removed from the data prior to performing the velocity analysis. To regularize the inversion, we parameterize the model using cubic B-splines [Symes, 2008a]. We solve the resulting optimization problem with the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method [Nocedal and Wright, 2000].

Need for full subsurface-offset image volumes

Our first experiment is designed to illustrate the main benefit of working with full subsurface offsets in all directions in situations where both horizontal and vertical reflectors are present (Figure 6.11 (a) adapted from [Yang and Sava, 2015]). For this purpose, we juxtapose the focussing, for a velocity model with the correct kinematics, of common-image gathers (CIGS) against focussing with common-image point gathers (CIPs) for a model with vertical and horizontal horizontal. While both CIG and CIP gathers can be used to form WEMVA objectives, their performance can be quite different. Ideally, CIGs measure focusing along offsets in the direction of the geologic dip for subsurface points sampled along spatial coordinated perpendicular to this direction, denoted by the yellow lines in Figure 6.11 (a). The green dots denote the location of the selected CIPs.

Since prior knowledge of geologic dips is typically not available, why not device a WEMVA scheme that focusses in all spatial directions? In that way, we are guaranteed to focus as illustrated in Figure 6.11 (b-d), for the horizontal reflector, and in Figure 6.11 (e-g) for the vertical reflector. From these figures it is clear that CIGs do not focus (Figure 6.11 (c,e)), despite the fact that the velocity model is correct. This lack of focusing may lead to erroneous biases during CIG-based WEMVA, a problem altogether avoided when we work with CIPs (Figure 6.11 (d,f)). Because CIPs

are sensitive to focusing in all directions, there is no need to focus in time [Sava and Vasconcelos, 2011]. While the advantages of CIP-based WEMVA are clear, we cheated by selecting CIPs on top of the reflecting interfaces, whose position is also generally not known in advance. This is where the randomized sampling comes to our rescue. By treating all subsurface points as random amplitude encoded sources, we will be able to form a CIP-based WEMVA objective, which does not require explicit knowledge on the location of the reflecting interfaces.

Quality of the stochastic approximation

As with FWI, randomly encoded sources—e.g. via random Gaussian weights, provide a vehicle to approximate (prohibitively) expensive to evaluate wave-equation based objectives (cf. equations 6.18 and 6.19) to controllable accuracy by increasing K. Following Haber et al. [2012], we evaluate for a small subset (Figure 6.12 (a)) of the Marmousi model [Bourgeois et al., 1991] the true $\phi(\mathbf{m}_0 + \alpha\delta\mathbf{m})$ (solid blue line) and approximate $\tilde{\phi}(\mathbf{m}_0 + \alpha\delta\mathbf{m})$ (denoted by the error bars) objectives as a function of α in the direction of the gradient—i.e., $\delta\mathbf{m} = -\nabla\phi$, evaluated at the starting model $\mathbf{m}_0$ depicted in Figure 6.12 (b). The results for $K = 10$ and $K = 80$ are included in Figures 6.12 (c,d). We calculated the error bars from 5 independent realizations. As we increase K, the true objective is better approximated reflected in tighter and better centred (compared to the true objective) error bars. These results also show that we can substantially reduce the computational costs while approximating the true objective function accurately.

WEMVA on the marmousi model

To validate our approach to WEMVA based on random probing and full-subsurface offset common-image point gathers (CIPs), we minimize the approximate objective in equation with a quasi-Newton method using approximate evaluations for the objective (equation 6.18) and gradients (equation 6.20). We choose the Marmousi model plotted in Figure 6.13 (a), because of its complexity and relatively steep reflectors. The model is 3.0 km deep and 9.2 km wide, sampled at 12 m. We acquired synthetic data for this model using a split-spread acquisition geometry resulting in 767 sources and receivers sources sampled at a 12 m interval. The data simulation and inversion are carried out over 201 frequencies, sampled at 0.1 Hz ranging from 5 to 25 Hz and scaled by a Ricker wavelet with a central frequency of 10 Hz.

We use a highly smoothed starting model with small (or no) lateral variations to start the inversions for different numbers of probing vectors, $K = 10$ and $K = 100$ respectively. To regularize the inversion, we use B-splines sampled at $x = 48$ m and $z = 48$ m in the lateral and vertical directions. Figures 6.13 (c, d) show the inverted models after 25 L-BFGS iterations. We can clearly see that even 10 probing vectors are good enough to reveal the structural information. Compared to computing the full image gathers (which would need $2 \cdot 767$ PDE solves per evaluation), this reduces the computational cost and memory use by roughly a factor 60.

Encouraged by this result, we conduct a second experiment for a poor starting model (Figure 6.14 (b)) and for data with fewer low frequencies ($8 - 25\,\mathrm{Hz}$). In this case, we use a slightly higher number of probing vectors ($K = 100$) to reduce the error in approximating the true objective function. Figures 6.12 (c, d) show that the errors for small K are relatively large when the starting model is poor. This is the case for small α's in Figure 6.12. To regularize the inversion, we use B-splines sampled at $x = 96\,\mathrm{m}$ and $z = 96\,\mathrm{m}$ in the lateral and vertical directions in order to recover the smooth (low-frequency) component of the true velocity model. We can clearly see in Figure 6.14 (c) that we build the low-frequency component of the velocity model. Figure 6.14 (d) shows our velocity estimate overlaid with a contour plot of the true velocity perturbation. We can see that despite missing low frequencies the inverted model captures the shallow complexity of the model reasonably well.

6.8 Discussion

To our knowledge this work represents the first instance of deriving a discrete two-way equivalent of Claerbout's double square-root equation [Claerbout, 1970, 1985a] that enables us to compute full-subsurface offset extended image volumes. Contrary to approaches based on the one-way wave equation, which are dip-limited and march along depth, our method provides access to extended image volumes via actions of a wave-equation based factorization on certain probing vectors. This factorization enables matrix-vector products with matrices that encode wavefield interactions between arbitrary pairs of points in the subsurface that can not formed explicitly. As a result, we arrive at a formulation that is computationally feasible and that offers new perspectives on the design and implementation of workflows that exploit information embedded in various types of subsurface extended images.

Depending on the choice of the probing vectors, we either obtain local information, tied to individual subsurface points that can serve as quality control for velocity analyses or as input to localized amplitude-versus-offset analysis, or global information that can as we have shown be used to drive automatic velocity analyses. Aside from guiding kinematical inversions through focusing, we expect that our randomized probings of extended image volumes also provide information on the rock properties. Further extension of the methodology include forming fully elastic image volumes, which turns our matrix representation into a tensor representation. We leave such generalizations to a future paper.

6.9 Conclusions

Extended subsurface image volumes carry information on interactions between pairs of subsurface points encoding essential information on the kinematics—to be used during velocity analysis—and the dynamics, which serve as input to inversions of the rock properties. While conceptually beautiful, full-subsurface offset image volumes have not yet been considered in practice because

these objects are too large to be formed explicitly. Through a wave-equation based factorization, we avoid explicit computations by forming matrix-vector products instead that only require two wave-equation solves each, and thereby removing the customary and expensive loop over shots, found in conventional extended imaging. This approach leads to significant computational gains in certain situations since we are no longer constrained by costs that scale with the number of shots and the number of subsurface points and offsets visited during the cross-correlation calculations. Instead, we circumvent these expensive explicit computations by carrying out these correlations implicitly through the wave equation solves. As a result, we end up with a matrix-vector formulation from which different image gathers can be formed and with which amplitude-versus-angle and wave-equation migration-velocity analyses can be performed without requiring prior information on the geologic dip. We showed that these operations can be accomplished at affordable computational costs.

By means of concrete examples, we demonstrate how localized information on focussing and scattering amplitudes can be revealed by forming different extended image volumes in both 2-D and 3-D. Because full-subsurface extended image volumes are quadratic in the number of gridded subsurface parameters, it would be difficult if not impossible to obtain these results by conventional methods. We also verify that our matrix-vector formulation lends itself well for automatic migration-velocity analysis if Gaussian random probing vectors are used. These vectors act as simultaneous sources and allow for significant computational gains in the evaluation of global focusing objectives that are key to migration-velocity analysis. Instead of focussing in a particular offset direction, as in most current approaches, our objective and its gradient force full-subsurface extended image volumes to focus in all offset directions at all subsurface points. It accomplishes this by forcing a commutation relation between the extended image volume and a matrix that expresses the Euclidean distance between points in the subsurface. The examples show that the computational costs can be controlled by probing. Application of this new automatic migration-velocity analysis technique to a complex synthetic shows encouraging results in particular in regions with steep geological dips.

Table 6.1: Correspondence between continuous and discrete representations of the image volume. Here, ω represents frequency, $\mathbf{x}$ represents subsurface positions, and (i, j) represents the subsurface grid points. The colon (:) notation extracts a vector from e at the grid point i, j for all subsurface offsets.

	Continuous	Discrete
full image volume	$e(\omega_i, \mathbf{x}, \mathbf{x}')$	$\mathbf{E}_i$
migrated image	$\int_\Omega \mathrm{d}\omega\ e(\omega, \mathbf{x}, \mathbf{x})$	$\sum_{i=1}^{N_f} \mathsf{diag}(\mathbf{E}_i)$
CIP	$e(\omega_i, \mathbf{x}_k, \mathbf{x}_{k'})$	$e_{ikk'}$

	# of PDE solves	flops
conventional	$2N_s$	$N_s N_{h_x} N_{h_y} N_{h_z}$
this paper	$2N_x$	$N_r N_s$

Table 6.2: Computational complexity of the two schemes in terms of the number of sources N_s, receivers N_r sample points N_x and desired number of subsurface offsets in each direction $N_{h_{\{x,y,z\}}}$.

	time (s)	memory (MB)
conventional	456	152
this paper	23	5.1

Table 6.3: Comparison of the computational time (in sec) and memory (in megabytes) for computing CIP's gather on a central part of Marmousi model. We can see the significant difference in time and memory using the probing techniques compared to the conventional method and we expect this difference to be greatly exacerbated for realistically sized models.

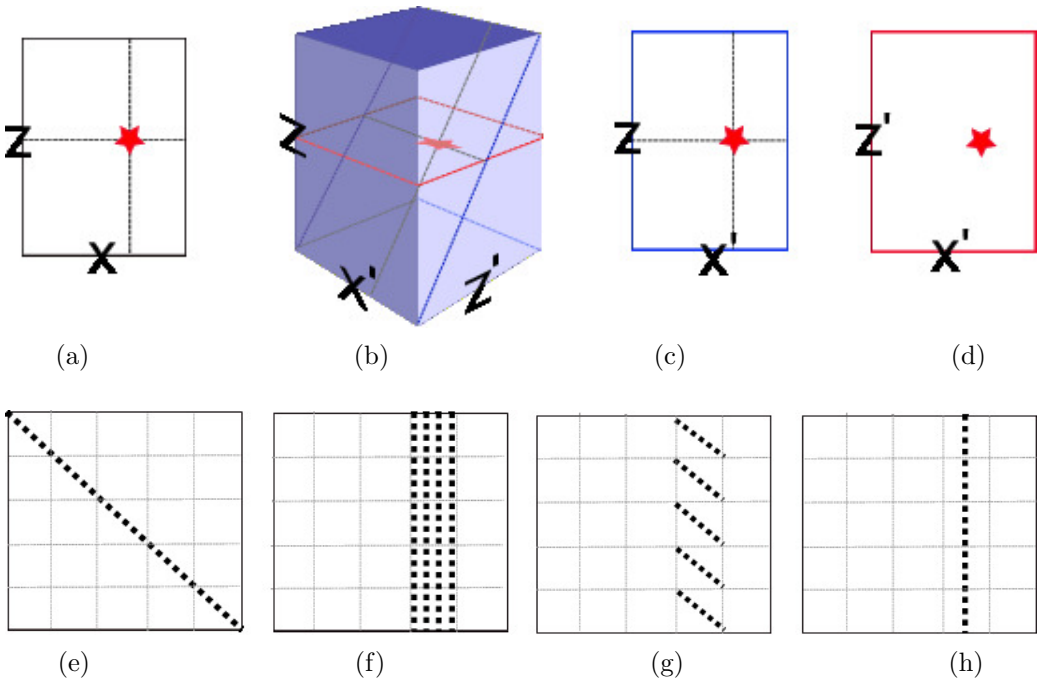


Figure 6.1: Different slices through the 4-dimensional image volume $e(z, z', x, x')$ around $z = z_k$ and $x = x_k$. (a) Conventional image $e(z, z, x, x)$, (b) Image gather for horizontal and vertical offset $e(z, z', x_k, x')$, (c) Image gather for horizontal offset $e(z, z, x_k, x')$ and (d) Image gather for a single scattering point $e(z_k, z', x_k, x')$. (e-g) shows how these slices are organized in the matrix representation of e.

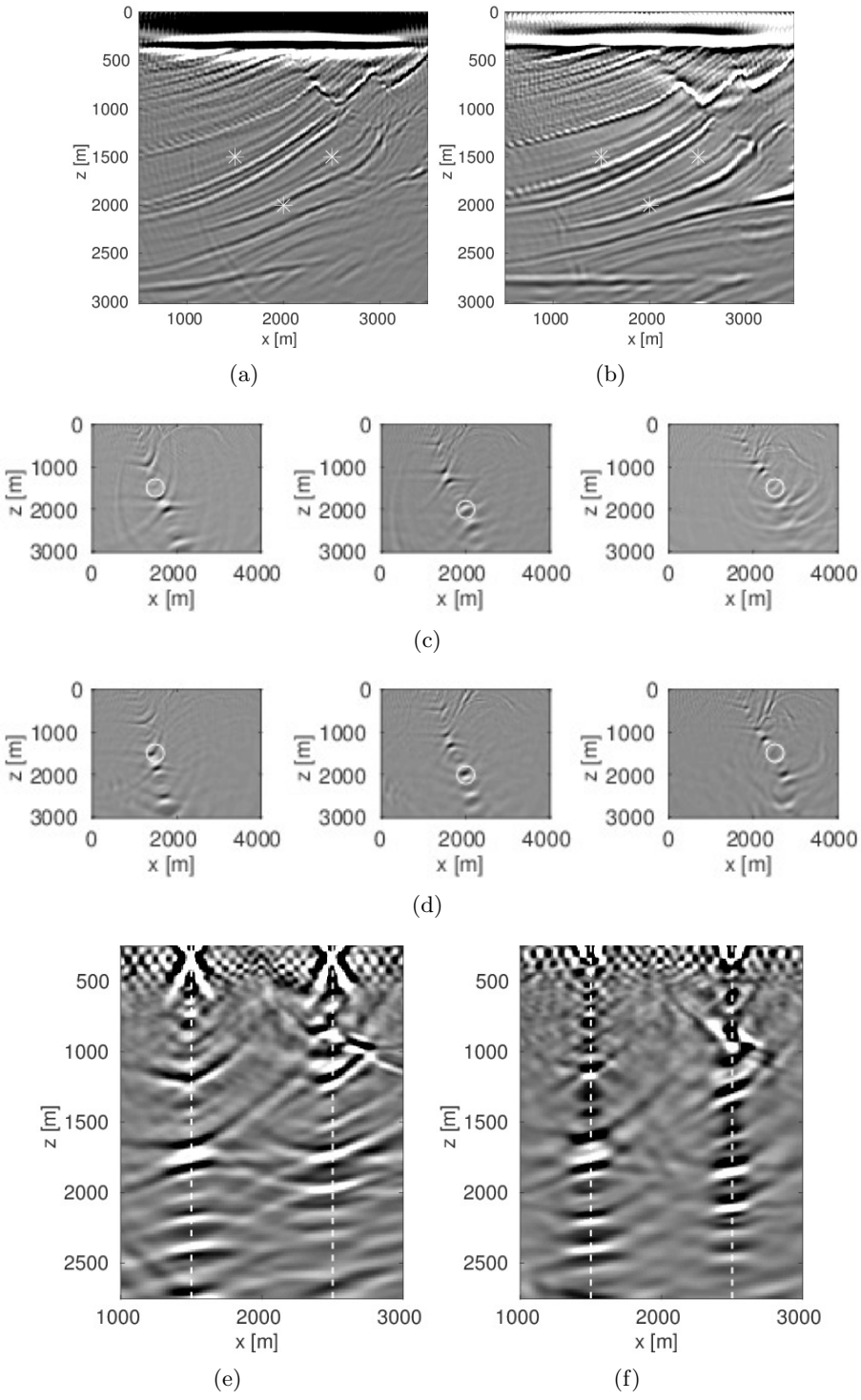
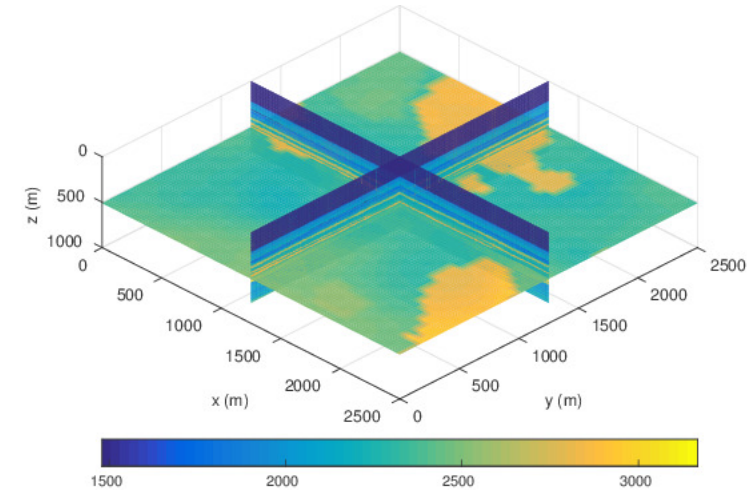
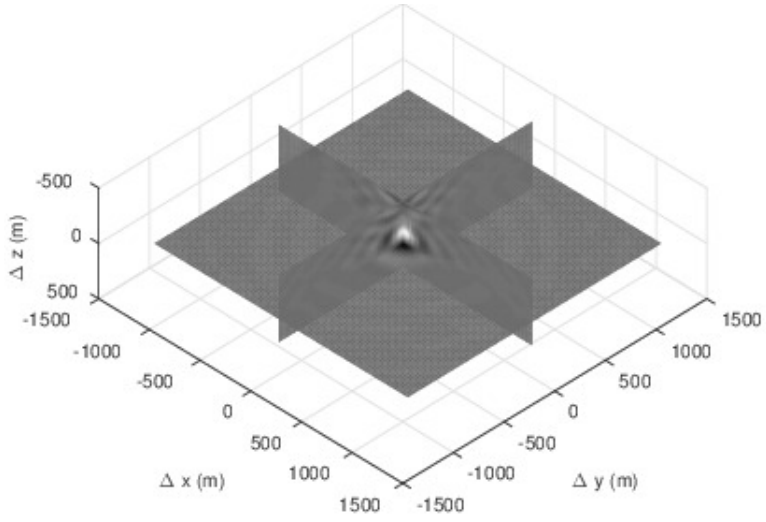


Figure 6.2: Migrated images for a wrong (a) and the correct (b) background velocity are shown with 3 locations at which we extract CIPs for a wrong (c) and the correct (d) velocity. The CIPs contain many events that do not necessarily focus. However, these events are located along the line normal to the reflectors. Therefore, it seems feasible to generate multiple CIPs simultaneously as long as they are well-separated laterally. A possible application of this is the extraction of CIGs at various lateral positions. CIGs at $x = 1500$ m and $x = 2500$ m for a wrong (e) and the correct (f) velocity indeed show little evidence of crosstalk, allowing us to compute several CIGs at the cost of a single CIG.

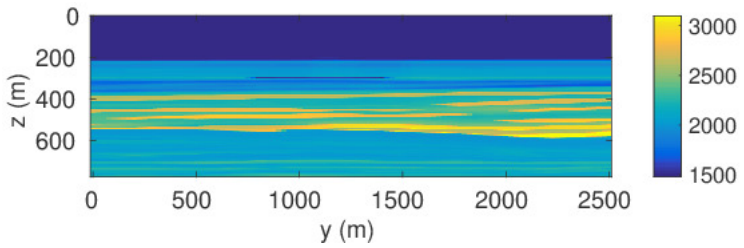


(a)

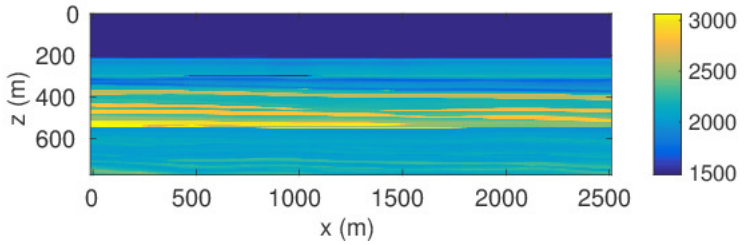


(b)

Figure 6.3: (a) Compass 3D synthetic velocity model provided to us by BG group. (b) A CIP gather at $(x, y, z) = (1250, 1250, 390)$m. The proposed method (Algorithm 2) is 1500 times faster then the classical method (Algorithm 1) to generate CIP gather.

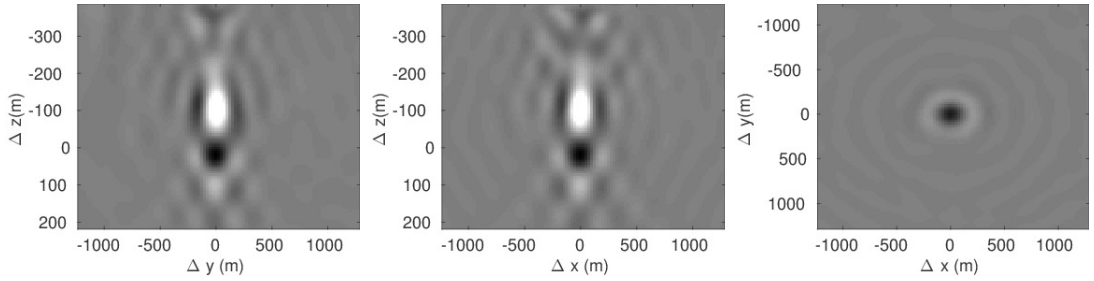


(a)



(b)

Figure 6.4: Cross-section of Compass 3D velocity model (Figure 6.3 (a)) along (a) x, and (b) y direction.



(a) (b) (c)

Figure 6.5: Slices extracted along the horizontal (a,b) and vertical (c) offset directions from the CIP gather shown in Figure 6.3 (b).

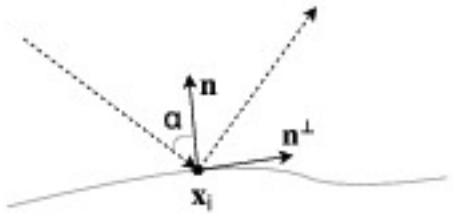


Figure 6.6: Schematic depiction of the scattering point and related positioning of the reflector.

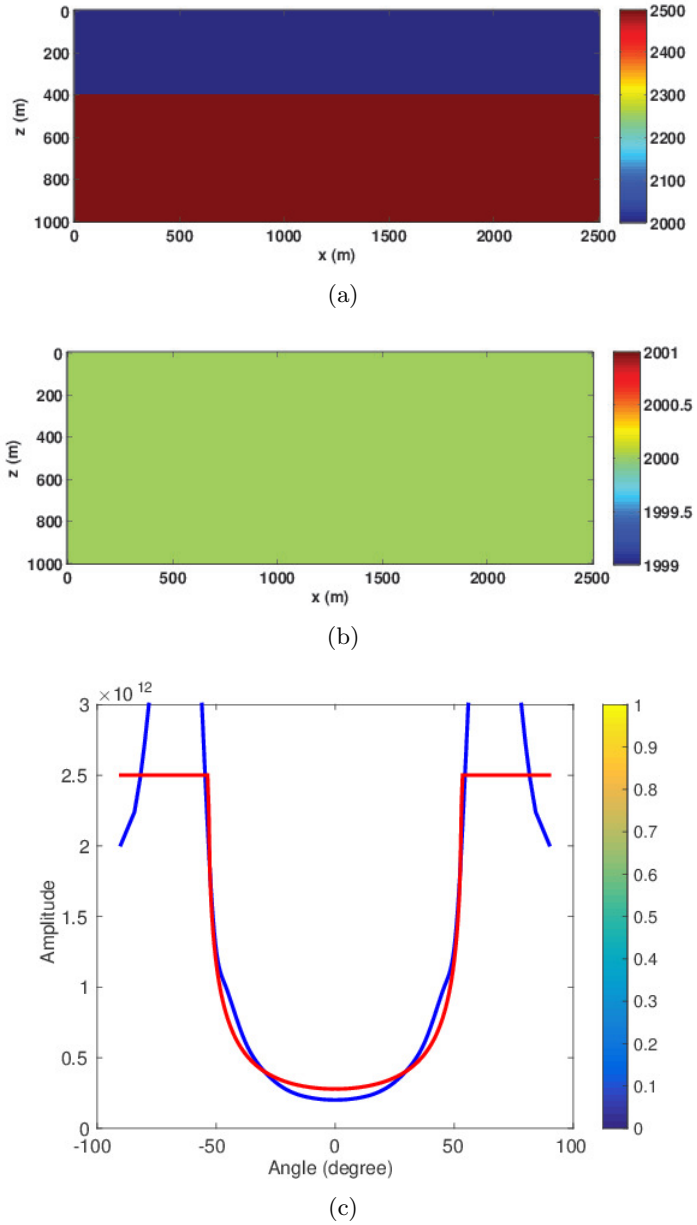


Figure 6.7: (a) Horizontal one-layer velocity model and (b) constant density model. CIP location is $x = 1250$ m and $z = 400$ m. (c) Modulus of angle-dependent reflectivity coefficients at CIP. The black lines are included to indicate the effective aperture at depth. The red lines are the theoretical reflectivity coefficients and the blue lines are the wave-equation based reflectivity coefficients.

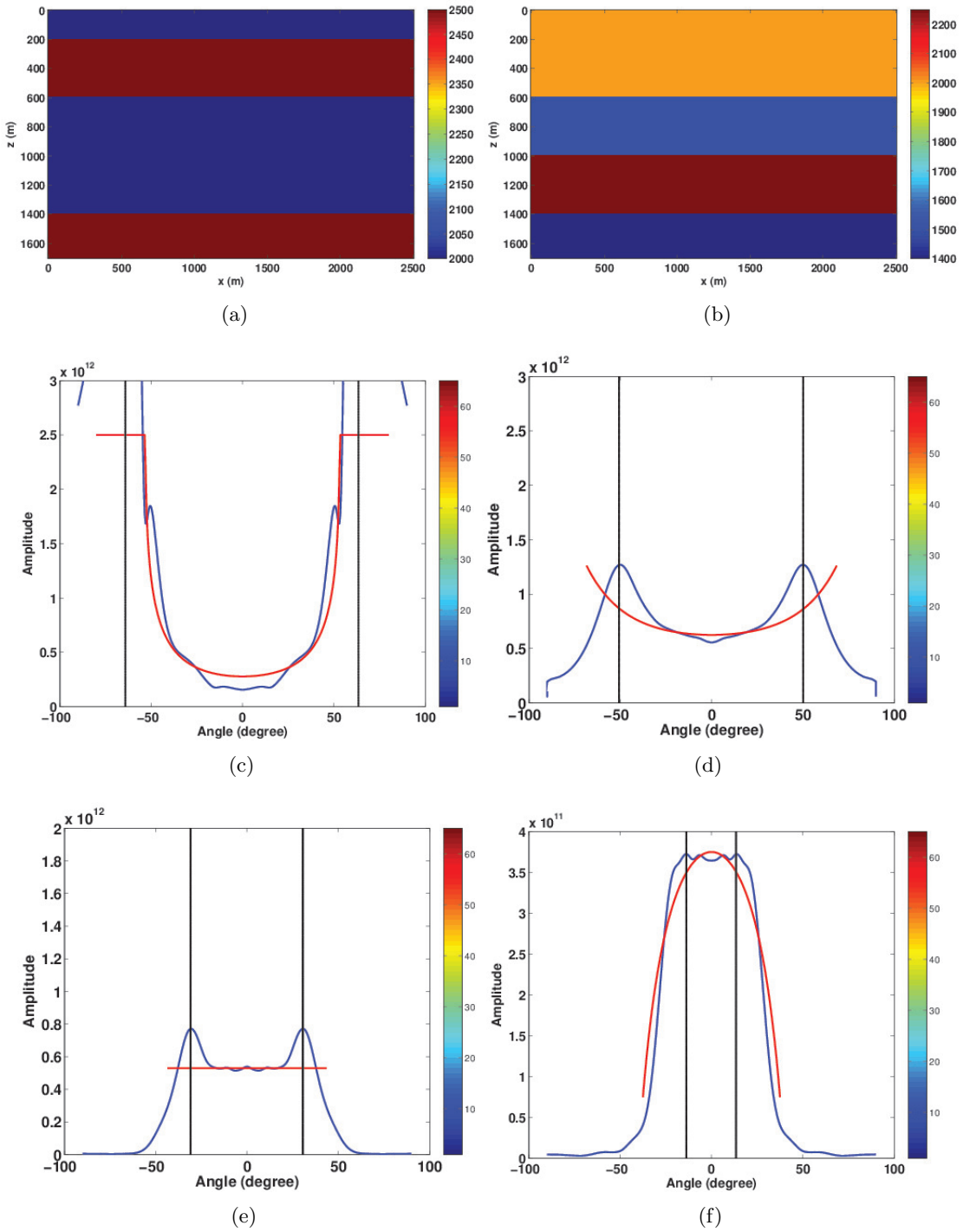


Figure 6.8: Angle dependent reflectivity coefficients in case of horizontal four-layer (a) velocity and (b) density model at $x = 1250$ m. Modulus of angle-dependent reflectivity coefficients at (c) $z = 200$ m, (d) $z = 600$ m, (e) $z = 1000$ m, (f) $z = 1400$ m.

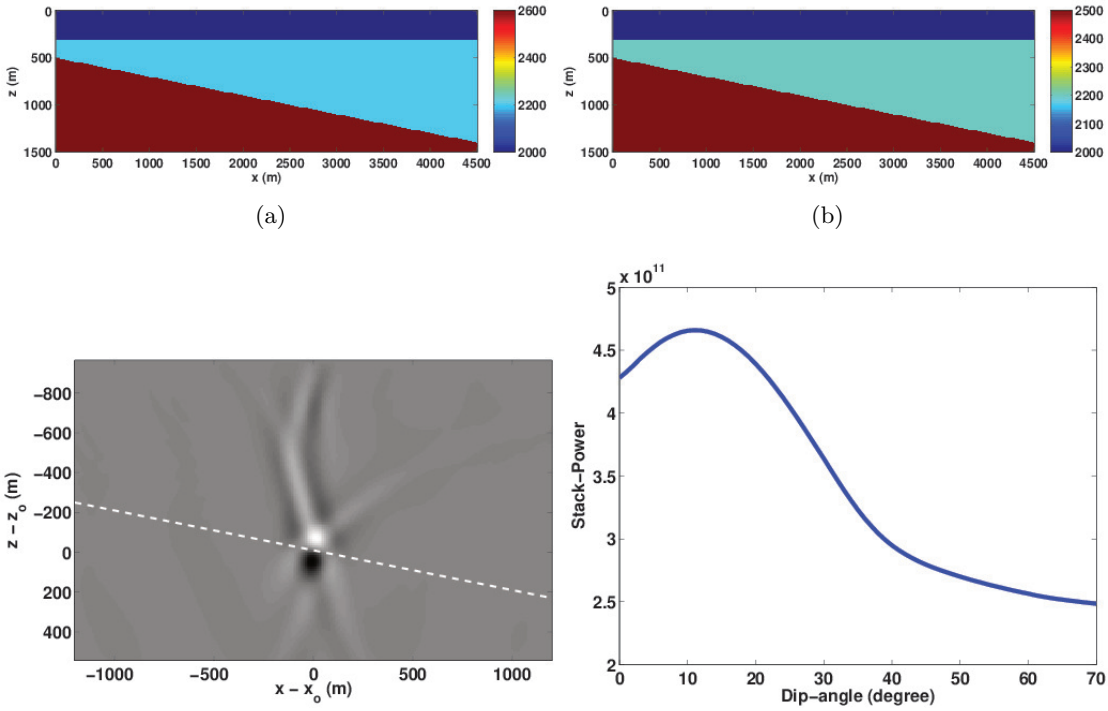


Figure 6.9: Estimation of local geological dip. (a,b) Two-layer model. (c) CIP gather at $x = 2250$ m and $z = 960$ m overlaid on dipping model. (d) Stack-power versus dip-angle. We can see that the maximum stack-power corresponds to the dip value of 10.8°, which is close to the true dip value of 11°.

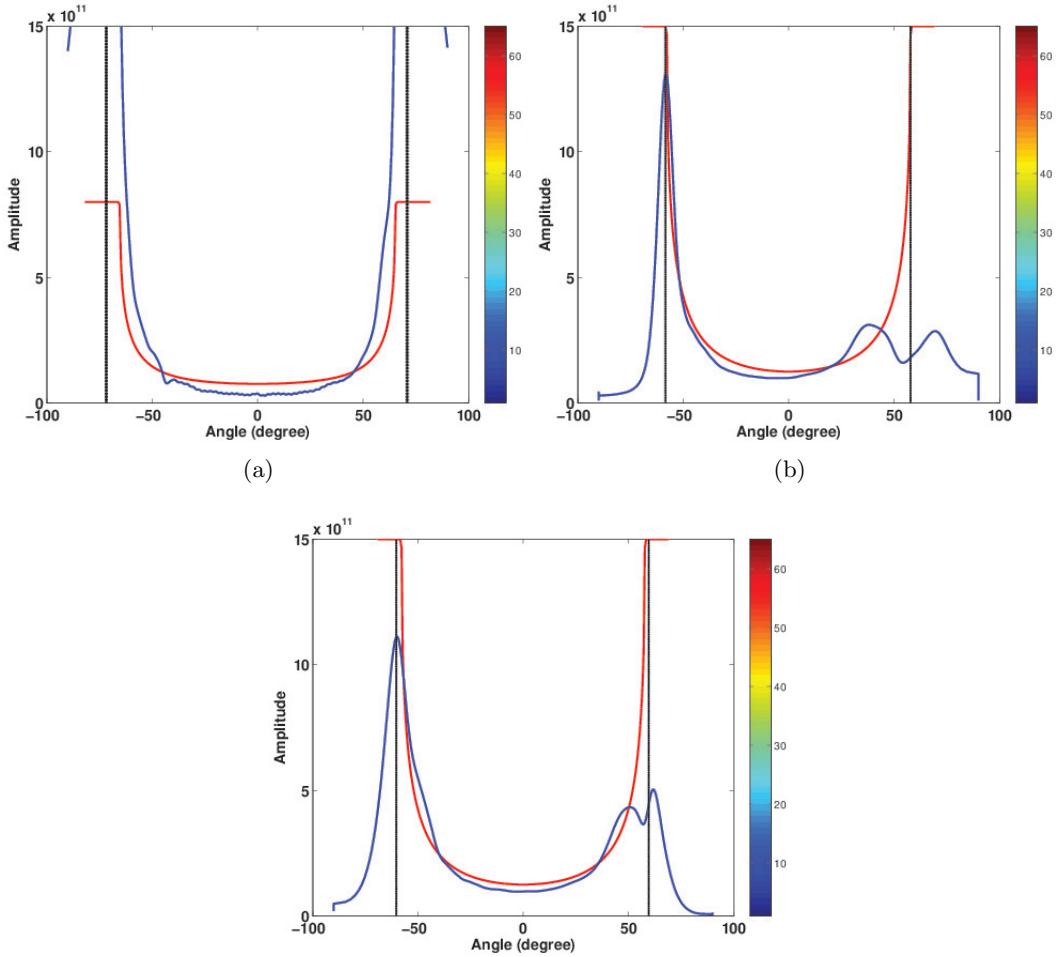


Figure 6.10: Modulus of angle-dependent reflectivity coefficients in two-layer model at $z = 300$ and 960 m and $x = 2250$ m. (a) Reflectivity coefficients at $z = 300$ m and $x = 2250$ m. Reflectivity coefficients at $z = 900$ m (b) with no dip $\theta = 0°$ and (c) with the dip obtained via the method described above ($\theta = 10.8°$).

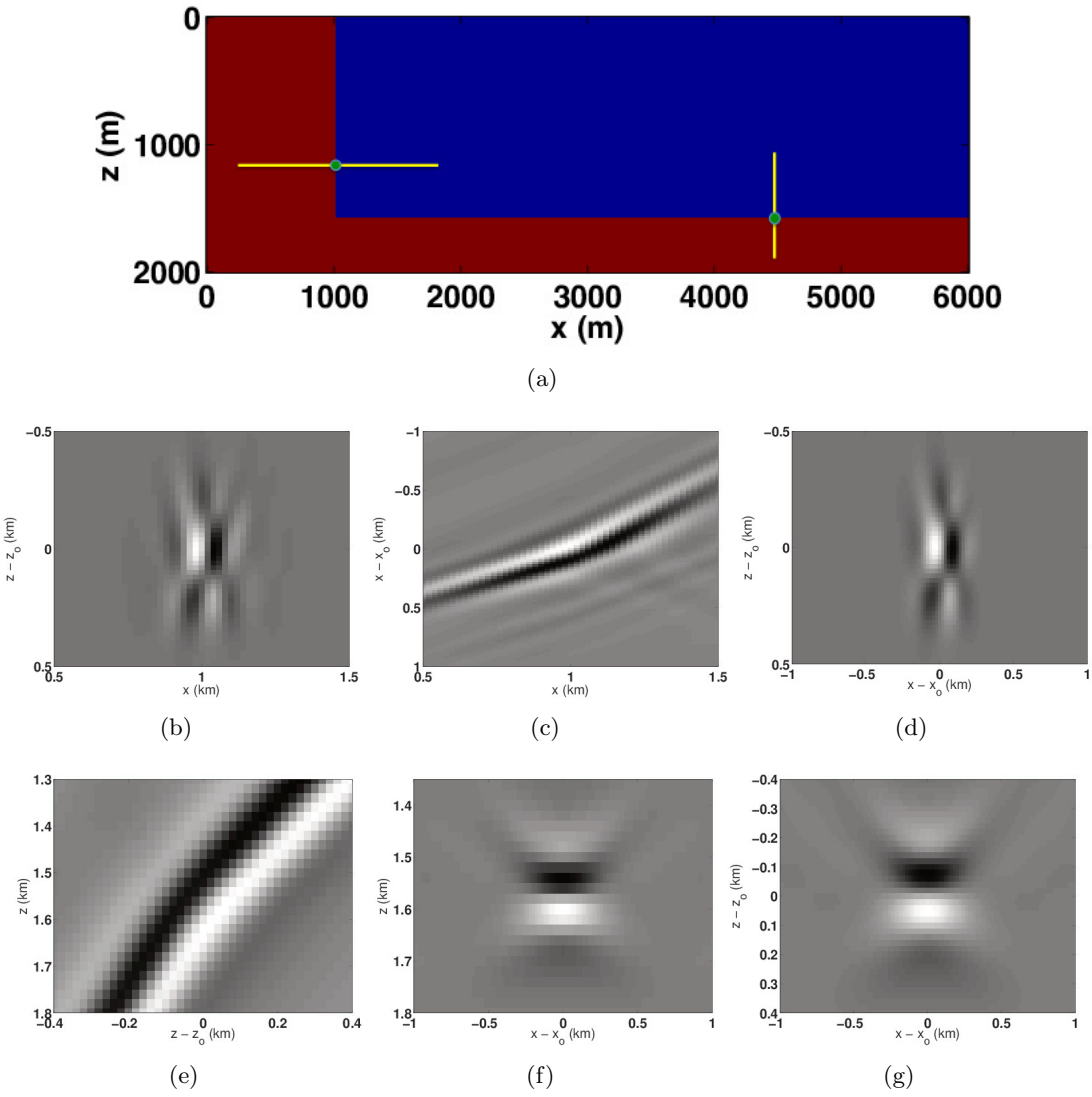


Figure 6.11: Comparison of working with CIGs versus CIPs. (a) True velocity model. The yellow line indicates the location along which we computed the CIGs and the green dot is the location where we extracted the CIPs. (b,c) CIGs extracted along vertical and horizontal offsets directions in case of vertical reflector. (d) CIPs extracted along vertical ($z = 1.2$ km, $x = 1$ km) reflector. (e,f) CIGs extracted along vertical and horizontal offsets directions in case of horizontal reflector. (g) CIPs extracted along horizontal ($z = 1.5$ km, $x = 4.48$ km) reflector.

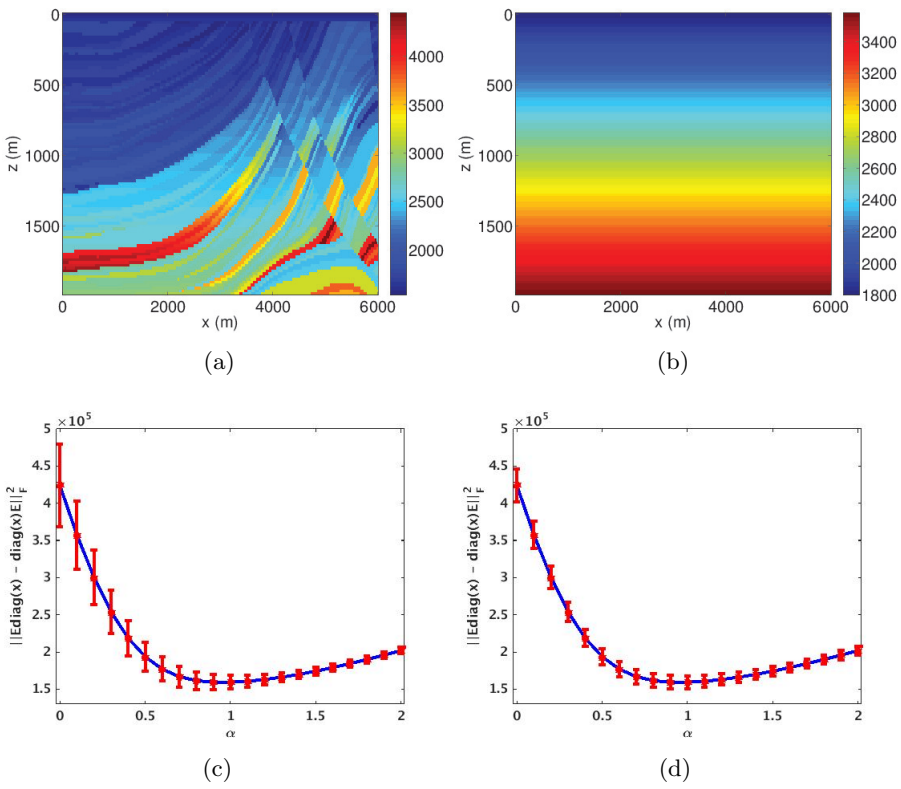


Figure 6.12: Randomized trace estimation. (a,b) True and initial velocity model. Objective functions for WEMVA based on the Frobenius norm, as a function of velocity perturbation using the complete matrix (blue line) and error bars of approximated objective function evaluated via 5 different random probing with (c) K=10 and (d) K = 80 for the Marmousi model.

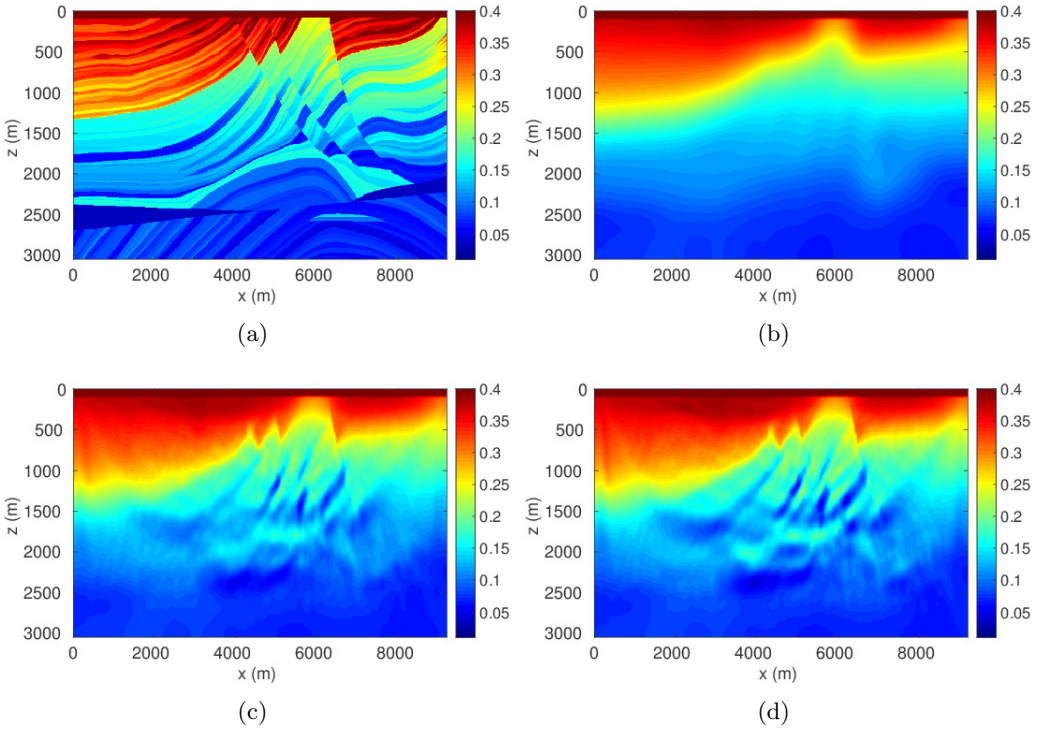


Figure 6.13: WEMVA on Marmousi model with probing technique for a good starting model. (a,b) True and initial velocity models. Inverted model using (c) $K = 10$ and (b) $K = 100$ respectively. We can clearly see that even 10 probing vectors are good enough to start revealing the structural information.

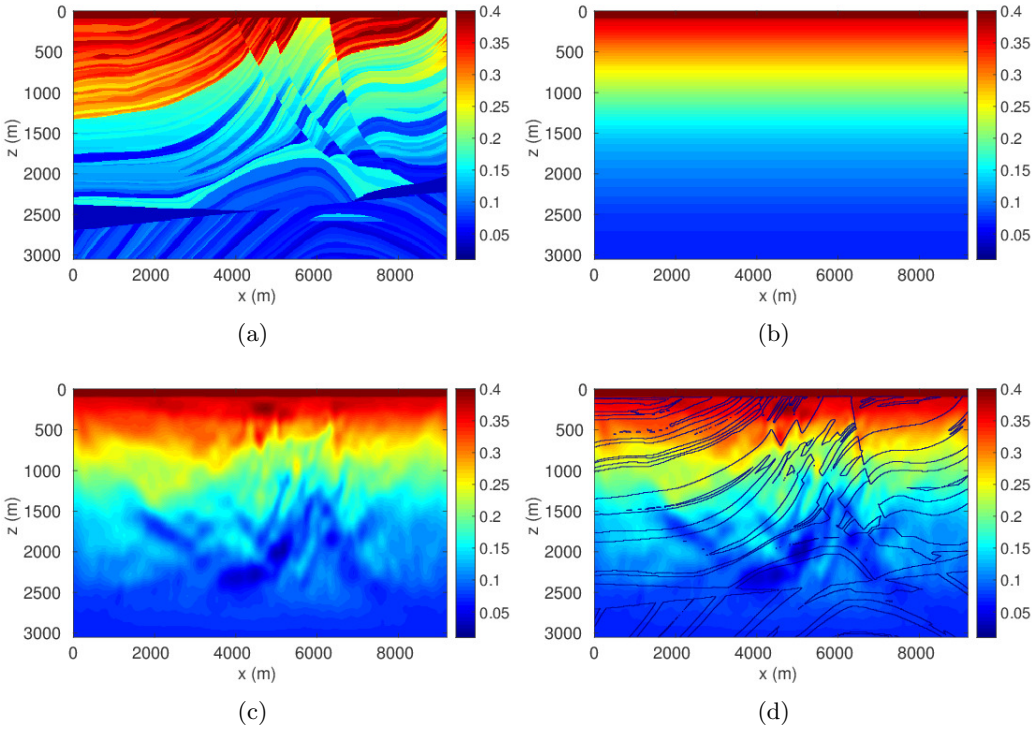


Figure 6.14: WEMVA on Marmousi model with probing technique for a poor starting velocity model and 8-25 Hz frequency band. (a,b) True and initial velocity models. Inverted model using (c) $K = 100$. (d) Inverted velocity model overlaid with a contour plot of the true model perturbation. We can see that we captures the shallow complexity of the model reasonably well when working with a realistic seismic acquisition and inversion scenario.

Chapter 7

Conclusions

In this chapter I summarize the main contributions of this thesis, propose follow-up work, discuss some limitations, and outline possible extensions.

7.1 Main contributions

In this thesis, I have developed fast computational techniques for large-scale seismic applications such as missing-trace interpolation, simultaneous source separation, and wave-equation based migration velocity analysis. In Chapters 2 and 3, I designed a large-scale singular value decomposition (SVD)-free factorization based rank-minimization approach, which is built upon the existing knowledge of compressed sensing as a successful signal recovery paradigm, and outlined the necessary components of a low-rank domain, a rank-increasing sampling scheme, and a SVD-free rank-minimizing optimization scheme for successful missing-trace interpolation. Note that for the matrix completion approach, the limiting component for large scale data is that of the nuclear norm projection, which requires the computation of SVD. For large scale 5D seismic data, it is prohibitively expensive to compute singular value decompositions because the underlying matrix has tens of thousands or even millions of rows and columns. Hence, we design a SVD-free factorization based approach for missing-trace interpolation and source separation. In Chapter 4 and 5, I extended the SVD-free rank-minimization framework to remove source cross-talk during simultaneous source acquisition because subsequent seismic data processing and imaging assumes well separated data. In Chapter 6, I proposed a matrix-vector formulation, which overcomes computational and storage costs of full-subsurface offset extended image volumes that are prohibitively expensive to form for large-scale seismic data imaging problems. This matrix-vector formulation avoids explicit storage of full-subsurface offset extended image volumes and removes the expensive loop over shots found in the conventional extended imaging [Sava and Vasconcelos, 2011].

The remainder of this section provides more details about the aforementioned contributions.

7.1.1 SVD-free factorization based matrix completion

Current efforts towards imaging subsalt structures under complex overburdens have led to a move towards wide-azimuth towed streamer (WATS) acquisition. Due to budgetary and/or physical constraints, WATS is typically coarsely sampled (sub-sampled) either along sources and/or receivers. However, most of the processing and imaging techniques require densely sampled seismic data on a regular periodic grid, thus call for large-scale seismic data interpolation techniques. Following ideas from the field of compressed sensing [Donoho, 2006b, Candès et al., 2006], a variety of methodologies, each based on various mathematical techniques, have been proposed to interpolate seismic data where the underlying principle is to exploit the sparse or low-rank structure of seismic data in a transform domain [Hennenfent et al., 2010, Kreimer, 2013]. However, these previous compressed sensing (CS)-based approaches, using sparsity or rank-minimization, incur computational difficulties when applied to large scale seismic data volumes. For instance, methods that involve redundant transforms, such as curvelets [Hennenfent et al., 2010], or that add additional dimensions, such as taking outer products of tensors [Kreimer, 2013], are computationally no longer tractable for large data volumes with four or more dimensions.

From a theoretical point of view, the success of rank-minimization based techniques depends upon two main principles, namely a low-rankifying transform and a sub-Nyquist sampling strategy that subdues coherent aliases in the rank-revealing transform domain. As identified in Chapter 2 (Figures 2.2 and 2.3), for 2D seismic data acquisition, seismic frequency slices exhibit low-rank structure (fast decay of the singular values) in the midpoint-offset domain, whereas, subsampling increases the rank in the midpoint-offset domain, i.e., singular values decay slowly. Hence, we can use the rank-minimization based techniques to reconstruct the low-rank representation of seismic data in the midpoint-offset domain. Similarly as shown in Chapter 3, 4D monochromatic frequency slices (extracted from a 3D seismic data acquisition) exhibit a low-rank structure in the (source-x, receiver-x) matricization compared to other possible matricization, whereas, subsampling destroys the low-rank structure in the (source-x, receiver-x) matricization (Figures 3.4 and 3.5). Here, matricization refers to an operation which reshapes a tensor into a matrix along specific dimensions. When these two principles hold, i.e., a low-rank domain and a rank-increasing sampling scheme, rank-minimization formulations allow the recovery of missing traces.

From a practical standpoint, we need a large-scale seismic data interpolation framework, which can handle seismic data measurements in the order of 10^{10} to 10^{12}. Unfortunately, one of the limitations of rank-minimization based techniques for large-scale seismic problems is the nuclear-norm projection, which inherently involves prohibitively expensive computations of singular value decompositions (SVD). For this reason, I proposed, in Chapter 2 and 3, a practical framework for recovering missing-traces in large-scale seismic data volumes using a matrix-factorization based approach [Lee et al., 2010b, Recht and Ré, 2011], where I avoided the need for expensive computations of SVDs. I showed that the proposed SVD-free factorization based framework can interpolate

large-scale seismic data volumes with high fidelity from significantly subsampled, and therefore economic and environmental friendly, seismic data volumes. In Chapter 3, I also demonstrated that the proposed factorization framework allows me to work with fully sampled seismic data volume without relying on a windowing operation that divides 5D data up into small cubes followed by normal-moveout corrections designed to remove curvature of seismic reflection events—one of the more recent approaches in missing-trace interpolation literature [Kreimer, 2013]. The reported results in Chapter 2 and 3 on realistic 3D and synthetic 5D seismic data interpolation demonstrate that while the SVD-free factorization based approaches are comparable (in reconstruction quality) to the existing matrix completion [Becker et al., 2011, Wen et al., 2012] and sparsity-promotion based techniques that use sparsity in transform domains (such as wavelets and curvelets) [Herrmann and Hennenfent, 2008b], it significantly outperforms these techniques in terms of the computational time and memory requirements.

Large-scale simultaneous source separation

Practitioners also proposed to acquire simultaneous seismic data acquisition to mitigate the sampling related issues and improve the quality of seismic data, wherein a single or multiple source vessels fire sources at near-simultaneous or slightly random times. This results in overlapping and missing shot records, as opposed to non-overlapping shot records in conventional marine acquisition. Although simultaneous source surveys result in improved quality recorded seismic data volumes at a reduced acquisition turnaround time, the costs of reduced acquisition are coherent artifacts from seismic cross-talk. This may degrade the quality of final migrated images because subsequent seismic data processing and imaging assumes well separated data acquired in the conventional (non-overlapping) way. Therefore, a practical (simultaneous) source separation and interpolation technique is required, which separates overlapping and interpolated missing shots. Following the compressed sensing principles outlined in Chapter 1 and 2 for matrix completion framework, I found that fully sampled conventional seismic data has low-rank structure—i.e., quickly decaying singular values in the midpoint-offset domain, whereas, overlapping and missing seismic data results in high-rank structure—i.e., slowly decaying singular values. For this reason, in Chapter 4 and 5, I developed a highly parallelizable singular-value decomposition (SVD)-free matrix factorization approach for source separation and missing-trace interpolation. I showed that the proposed framework regains the low-rank structure structure of the separated and interpolated seismic data volumes. I tested the SVD-free framework on two different seismic data acquisition scenarios, namely static (receivers are moving with sources) and dynamic (receivers are fixed at ocean-bottom) geometries. For dynamic geometry, in Chapter 4, I investigated two instances of low-variability in source firing times—e.g., $0 \leq 1$ (or 2) second, namely over/under and simultaneous long offset (Figure 4.7 and 4.9). In Chapter 5, I investigated an instance of static geometry in source firing times—e.g., $>$ 1 second, 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 5.1). For both the cases, I showed that the proposed SVD-free rank-minimization framework is able to separate and interpolate the large-scale seismic data to a desired grid with negligible loss of the coherent energy where the reconstruction quality is comparable to sparsity-promotion [Wason and Herrmann, 2013b] and NMO-based median filtering type techniques [Chen et al., 2014]. I also demonstrated that the proposed SVD-free factorization based rank-minimization approach for source separation outperforms sparsity-promotion based techniques by an order of magnitude in computational speed while using only 1/20th of the memory.

Missing-trace interpolation without windowing

Seismic data is often acquired in large volumes, which makes the processing of seismic data computationally expensive. For this reason, seismic practitioners follow workflows that involve: i) normal-moveout (NMO) correction to remove the curvature of seismic reflection events such that the events tend to become linear in small enough windows, ii) dividing data into overlapping spatial-temporal windows, iii) performing Fourier transform along the time coordinate within each window, iv) using a matrix or tensor based technique to interpolate individual monochromatic slices, v) doing the inverse Fourier transform along the frequency coordinate within each window, vi) combining all the windows to get reconstructed seismic data volumes. While this approach is computationally feasible because it can be readily parallelized and utilized with success, there are several issues during the windowing process: i) one needs a reasonably accurate root-mean square velocity model to perform the NMO correction that may be difficult to compute when data is missing, ii) proper window size selection, which depends upon the complexity of seismic data, i.e., large window size for linear reflection events and smaller window size for complex seismic reflection events with curvature, iii) averaging operations along the overlapping windows (see [Kreimer, 2013] for details). These issues are very important because they may lead to underperformance.

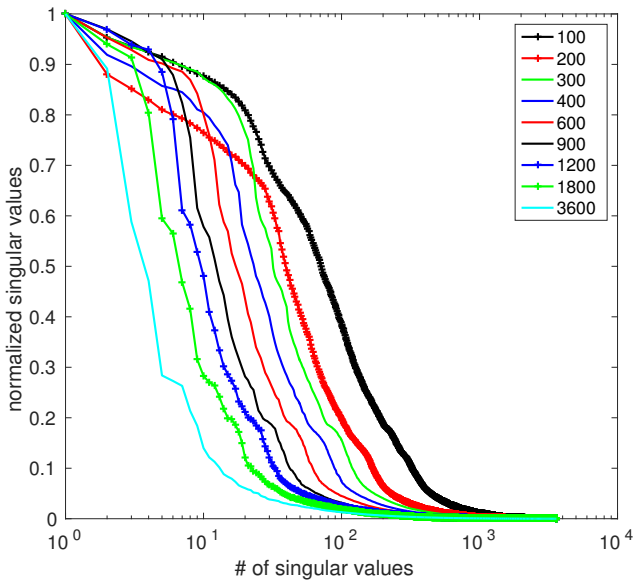
One of the main contributions of this thesis was the design and implementation of a SVD-free matrix-factorization approach, which exploits the inherent redundancy in seismic data while avoiding the customary preprocessing steps of windowing and NMO corrections. Seismic data is inherently redundant because we are collecting data on the same subsurface at different angles. In this thesis, I used this inherent redundancy of seismic data to exploit the low-rank structure, where the idea is to represent complete seismic data volume using only a few singular vectors corresponding to the largest singular values. Note that this redundancy can only be exploited when the data is organized in a particular manner, namely exploiting the low-rank structure of seismic data in the midpoint-offset domain for 2D seismic data surveys or (source-x, receiver-x) matricizations for 3D seismic data surveys.

When data volumes are organized in this way, the singular values decay rapidly, which is a reflection of the intrinsic redundancy exhibited by seismic data. To illustrate this inherent redundancy,

I considered a split-spread seismic acquisition over a complex geological model provided by the BG Group, where each source is recorded by all the receivers. This simulation resulted in a seismic data with 1024 time samples, 60×60 sources and 60×60 receivers. From this seismic volume, I extracted a monochromatic 4D tensor at 10 Hz, where the size of the tensor is $60 \times 60 \times 60 \times 60$, followed by an analysis of the decay of the singular values as a function of window sizes. To compare the recovery within windows versus carrying out the interpolation over the complete non-windowed survey, I computed the singular values as follows: i) I perform x_{src}, x_{rec} matricization (as explained in Chapter 3) on a 10 Hz monochromatic slice, resulting in a matrix of 3600 rows and 3600 columns, ii) I define window sizes of $100, 200, 300, 400, 600, 900, 1200, 1800$ and 3600, iii) for each window size, I extract all possible sub-matrices and compute their corresponding singular values, iv) I concatenate all the singular values of sub-windowed matrices, for a given window size, and sort them in the descending order. In Figure 7.1 (a), I compared the decay of singular values of monochromatic slices at 10 Hz for different window sizes, where we can see that i) fully sampled seismic data volumes have the fastest decay of singular values, and ii) smaller window sizes result in a slower decay rate of the singular values, i.e., we need relatively more singular values to approximate the underlying fully sampled seismic data. This simple experiment demonstrated that fully sampled non-windowed monochromatic slices exhibit low-rank structure because we can approximate them by using a few singular vectors that correspond to the first few largest singular values with minimal loss of coherent seismic energy. Therefore, the non-windowed seismic data should be used during missing-trace interpolation and source separation using the rank-minimization based techniques and avoiding the customary windowed- and NMO-based preprocessing steps.

7.1.2 Enabling computation of omnidirectional subsurface extended image volumes

Image gathers contain kinematics and dynamics information on interactions between pairs of subsurface points. We can use this information to design wave-equation based velocity analysis, amplitude-versus-angle inversion to estimate the rock properties, and target-oriented imaging using redatuming techniques in geological challenging environments to overcome the effects of complex overburden. Unfortunately, forming full-subsurface offset extended image volumes is prohibitively expensive because they are quadratic in the image size and therefore impossible to store. Apart from storage impediments, the costs associated with computing the extended image volumes scale with the number of shots, the number of subsurface points, and the number of subsurface offsets visited during the cross-correlation calculations. In Chapter 6, I proposed a computationally and memory efficient way of gleaning information from the full-subsurface offset extended image volumes, where I first organized the full-subsurface offset extended image volumes as a matrix. Then, I computed the action of the matrix on a given vector without explicitly constructing the full-subsurface offset extended image volumes. The proposed matrix-vector product removes the expensive loop over shots found in the conventional methods, thus leading to significant computational and memory



(a)

Figure 7.1: To understand the inherent redundancy of seismic data, we analyze the decay of singular values of windowed versus non-windowed cases. We see that fully sampled seismic data volumes have fastest decay of singular values, whereas, smaller window sizes result in the slower decay rate of the singular values.

gains in certain situations. Using this matrix-vector formulation, I formed the image-gathers along all offset directions to perform amplitude-versus-angle and automatic wave-equation migration-velocity analyses without requiring prior information on the geologic dip. By means of concrete examples, I demonstrated that we can extract the localized scattering amplitudes information from the image volumes computed on 2D and 3D velocity models. I further validated the potential of probing techniques to perform automatic wave-equation migration-velocity analyses on a complex synthetic model with steep geological dips.

7.2 Follow-up work

Although I tested the proposed SVD-free factorization-based rank-minimization framework on realistic 3D and complex 5D synthetic data volumes and compared it to the existing matrix / tensor completions and sparsity-promotion based interpolation techniques (such as curvelets), it would be interesting to see its application to a realistic 5D data set observed on complex geological structures such as salt. Since our SVD-free rank-minimization framework works with the full seismic data volumes without windowing, it will be computationally infeasible to interpolate the complete data on a single computing node, i.e., performing the interpolation in a serial mode, hence, we need

to devise a parallel interpolation framework that can exploit the inherent redundancy of seismic data without performing the windowing operation and overcome the computational bottleneck of handling the large-scale seismic data volumes where unknowns are of the order of 10^{10} to 10^{12}. Finally, it would be good to see the impact of 5D seismic interpolation on various seismic post processing workflows such as surface-related multiple estimation, wavefield-decomposition, migration, waveform inversion, target-imaging and uncertainty analysis.

7.3 Current limitations

One of the main requirements of the current rank-minimization approach is to find a transform domain where the target recovered data exhibit a low-rank structure. In this thesis, I showed that 2D seismic monochromatic slices exhibit low-rank structure in the midpoint-offset domain at the lower frequencies, but not at the higher frequencies. This behaviour is due to the increase in wave oscillations as we move from low to high-frequency slices in the midpoint-offset domain, even though the energy of the wavefronts remains focused around the diagonal (see Figures 2.2 and 2.3 in Chapter 2). Therefore, interpolation via rank minimization in the high-frequency regime requires extended formulations that incorporate low-rank structure. [Kumar et al., 2013a] proposed to addressed this issue for 2D seismic data acquisition using Hierarchically Semi-Separable matrix representation (HSS) [Chandrasekaran et al., 2006]. Although the initial results of HSS based techniques for 3D seismic data interpolation and source separation are encouraging, it would be interesting to see its benefits to interpolate higher frequencies in large-scale realistic 5D seismic data volumes generated using 3D seismic data acquisitions. Moreover one of the limitations in proposed SVD-free matrix-factorization approach is to find the rank parameter k associated with each low-rank factor. I estimated it using the cross-validation techniques (see chapter 3 for more details) in all the examples presented in this thesis, however, it is still not a practical approach when dealing with large-scale subsampled seismic data volumes, since it will involve finding the appropriate small volume of seismic data to run the cross-validation techniques—a computationally expensive process.

7.4 Future extensions

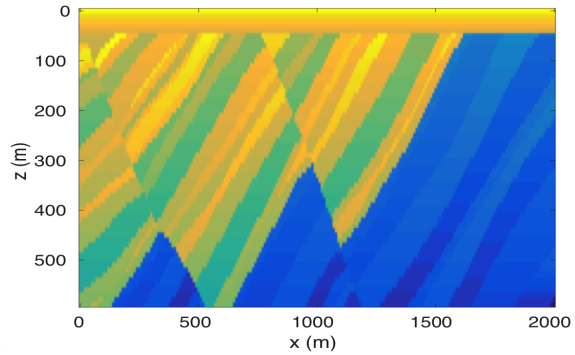
7.4.1 Extracting on-the-fly information

While the traditional interpolation and/or source separation approaches can deal with missing information and/or source cross-talk during seismic data acquisition, these processing methods result in massive seismic data volumes in the orders of terabytes to petabytes. This overwhelming amount of data makes subsequent imaging and inversion workflows daunting for realistic data sets, since extracting various data gathers, such as common-source and receiver gathers can be very challenging (Input/Output costs), apart from storing the massive volumes of interpolated and/or separated seismic data on disks. To overcome this computational and memory burden, we can

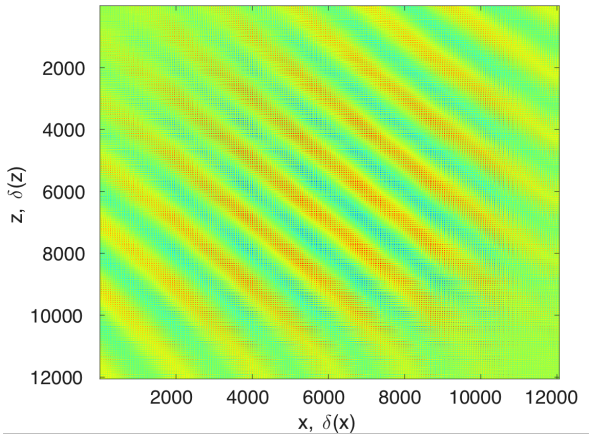
design a fast, resilient, and scalable workflow, where we first compress the seismic data volumes using a SVD-free rank-minimizing optimization scheme. Then, we can access the information from the compressed volumes on-the fly, i.e., extracting the common-source and/or receiver gathers from the low-rank factors without forming the fully sampled seismic data volumes during the objective and gradient calculations of inversion framework.

7.4.2 Compressing full-subsurface offset extended image volumes

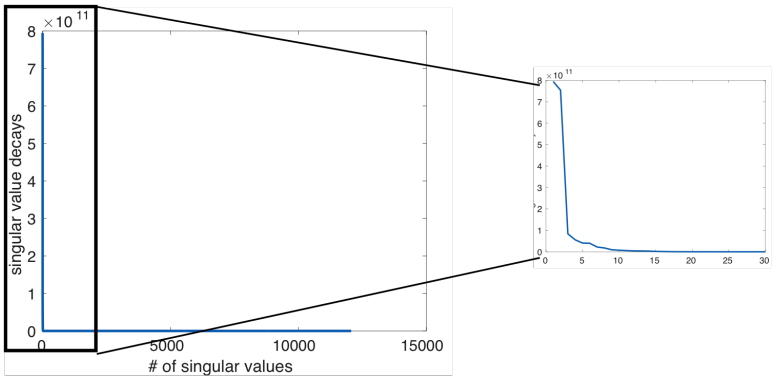
Even though the probing techniques circumvent the computational and storage requirements of forming the full-subsurface offset extended image volumes, we have only limited access to the information from the full-subsurface offset image volumes at the subsurface locations defined by the probing vectors. As a consequence, the proposed framework of probing vectors can only be beneficial when the number of probing vectors is very small compared to the number of sources. To circumvent the computational requirement of forming the full-subsurface offset extended image volumes at every point in the subsurface, I proposed to exploit the low-rank structure of image volumes. To understand the low-rank behaviour of image volumes, I analyzed its singular value decay on a small section of Marmousi model (Figure 7.2 (a)). I chose this particular part of the model because it consists of highly dipping reflectors with strong lateral variations in the velocity. For this 2D model, the monochromatic full-subsurface offset extended image volume is a four dimensional tensor with dimensions, depth z, offset x, horizontal lag δx, and vertical lag δz. As outlined in Chapter 3, there is no unique generalization of the SVD to tensors and as a result, there is no unique notion of rank for tensors. However, rank can be computed on the different matricizations of tensors where matricization reshapes a tensor to a matrix along specific dimensions. To analyze the decay of singular values of each monochromatic full-subsurface offset extended image volume, I matricize this 4D tensor where the depth z and vertical lag coordinates $(z, \delta z)$ are grouped along the rows, and the offset x and horizontal lag δx coordinates are grouped along the columns. Figure 7.2 (b) shows the matricized tensor and Figure 7.2 (c) shows its associated singular values decay. I also tested the decay of singular values for all possible combination of matricizations, i.e., $(z, \delta x), (z, x), (z, \delta z), (\delta x, \delta z)$, and find that $(z, \delta z)$ matricization gives the fastest decay of the singular values. I further approximated the full-subsurface offset extended image volume with its first 10 singular vectors computed using $(z, \delta z)$ matricization, which results in a reconstruction error of 10^{-4}. This example demonstrates that even for highly complex geological structures, the full-subsurface offset extended image volumes for all subsurface points exhibits low-rank structure, which can be exploited to compress full-subsurface offset extended image volumes while gleaning information from the full-subsurface offset extended image volumes during wave-equation migration velocity analyses.



(a)



(b)



(c)

Figure 7.2: To visualize the low-rank nature of image volumes, I form a full-subsurface offset extended image volume using a subsection of the Marmousi model and analyzed the decay of singular values. (a) Complex subsection of the Marmousi model with highly dipping reflectors with strong lateral variations in the velocity, and (b) corresponding full-subsurface offset extended image volume at 5 Hz. (c) To demonstrate the low-rank nature of image volumes, I plot the decay of singular values, where I observed that we only required the first 10 singular vectors to get a reconstruction error of 10^{-4}.

7.4.3 Comparison of MVA and FWI

Apart from performing velocity analysis using extended image volumes, which is known as migration velocity analysis (MVA) in seismic literature, full-waveform inversion (FWI) is another useful tool to invert for the velocity model of the subsurface. FWI is a nonlinear data-fitting procedure, where the aim is to get the velocity updates via minimizing the mismatch between the observed and predicted seismic data. Here, the predicted data is generated using an initial guess of the subsurface by solving a wave-equation [Virieux and Operto, 2009]. Both MVA and FWI have its own pitfall, which can lead to spurious artifacts in the velocity inversion. For example, FWI is mainly driven by the turning waves and MVA is mainly driven by the reflected waves. FWI has deeper depth of penetration for lower frequencies but not for higher frequencies, whereas MVA can provide the velocity updates in the deeper part of the model for both lower and higher frequencies. Another well-known drawback of FWI is the occurrence of local-minima in the misfit functional, which leads to erroneous artifacts in the velocity updates. This can be circumvented by either recording the low-frequencies in the observed data [Virieux and Operto, 2009], which are generally absents from the seismic data, or starting with a good initial velocity model that is sufficiently close to the true model. In future work, I would like to come up with an inversion framework, which has flavors of both FWI and MVA and help us to overcome some of these pitfalls.

Bibliography

AAPG. Seismic data display. *American Association of Petroleum Geologists Wiki*, January 24 2017. URL http://wiki.aapg.org/Seismic_data_display. → pages 2

R. Abma, T. Manning, M. Tanis, J. Yu, and M. Foster. High quality separation of simultaneous sources by sparse inversion. In *72nd EAGE Conference and Exhibition*, 2010. → pages 78

R. Abma, A. Ford, N. Rose-Innes, H. Mannaerts-Drew, and J. Kommedal. Continued development of simultaneous source acquisition for ocean bottom surveys. In *75th EAGE Conference and Exhibition*, 2013. doi: 10.3997/2214-4609.20130081. URL http://earthdoc.eage.org/publication/publicationdetails/?publication=68897. → pages 78

P. Akerberg, G. Hampson, J. Rickett, H. Martin, and J. Cole. Simultaneous source separation by sparse radon transform. *SEG Technical Program Expanded Abstracts*, 27(1):2801–2805, 2008. doi: 10.1190/1.3063927. URL http://link.aip.org/link/?SGA/27/2801/1. → pages 78

K. Aki and P. G. Richards. *Quantitative seismology*. Freeman and Co. New York, 1980. → pages 114

A. Aravkin, M. Friedlander, F. Herrmann, and T. van Leeuwen. Robust inversion, dimensionality reduction, and randomized sampling. *Mathematical Programming*, 134(1):101–125, 2012. → pages 14, 24

A. Aravkin, J. Burke, and M. Friedlander. Variational properties of value functions. *SIAM Journal on Optimization*, 23(3):1689–1717, 2013a. doi: 10.1137/120899157. URL http://dx.doi.org/10.1137/120899157. → pages 14, 15, 16, 23, 24, 25

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*, 36(5):S237–S266, 2014a. doi: 10.1137/130919210. → pages 50, 57, 58, 60, 69

A. Y. Aravkin, J. V. Burke, and M. P. Friedlander. Variational properties of value functions. *SIAM Journal on optimization*, 23(3):1689–1717, 2013b. → pages 88

A. Y. 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*, 36(5):S237–S266, 10 2014b. doi: 10.1137/130919210. URL http://epubs.siam.org/doi/abs/10.1137/130919210. → pages 82, 88

H. Avron and S. Toledo. Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix. *Journal of the Association for Computing Machinery*, 58(2):P1–P16, Apr. 2011. ISSN 00045411. doi: 10.1145/1944345.1944349. URL http://portal.acm.org/citation.cfm?doid=1944345.1944349http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.163.6194&rep=rep1&type=pdf. → pages 125

R. H. Baardman and R. G. van Borselen. Method and system for separating seismic sources in marine simultaneous shooting acquisition. *Patent Application*, EP 2592439 A2, 2013. → pages 78

V. Bardan. Trace interpolation in seismic data processing. *Geophysical Prospecting*, 35(4): 343–358, 1987. ISSN 1365-2478. doi: 10.1111/j.1365-2478.1987.tb00822.x. URL http://dx.doi.org/10.1111/j.1365-2478.1987.tb00822.x. → pages 49

E. Baysal, D. D. Kosloff, and J. W. Sherwood. Reverse time migration. *Geophysics*, 48(11): 1514–1524, 1983. → pages 113

C. J. Beasley. A new look at marine simultaneous sources. *The Leading Edge*, 27(7):914–917, 2008. doi: 10.1190/1.2954033. URL http://tle.geoscienceworld.org/cgi/content/abstract/27/7/914. → pages 78

C. J. Beasley, R. E. Chambers, and Z. Jiang. A new look at simultaneous sources. *SEG Technical Program Expanded Abstracts*, 17(1):133–135, 1998. doi: 10.1190/1.1820149. URL http://link.aip.org/link/?SGA/17/133/1. → pages 105

S. R. Becker, E. J. Candes, and M. C. Grant. Templates for convex cone problems with applications to sparse signal recovery. *Mathematical Programming Computation*, 3(3):165–218, 2011. ISSN 1867-2949. doi: 10.1007/s12532-011-0029-5. URL http://dx.doi.org/10.1007/s12532-011-0029-5. → pages 23, 28, 35, 145

E. v. Berg and M. P. Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2):890–912, 2008. → pages 13, 14, 15, 16, 20, 24, 29, 51, 80, 82, 88, 89

E. v. Berg and M. P. Friedlander. Sparse optimization with least-squares constraints. *SIAM J. Optimization*, 21(4):1201–1229, 2011. → pages 13, 14, 15, 28, 35

A. Berkhout. Changing the mindset in seismic data acquisition. *The Leading Edge*, 27(7): 924–938, 2008a. → pages 105, 109

A. Berkhout and D. Verschuur. Imaging of multiple reflections. *Geophysics*, 71, no. 4(4): SI209–SI220, 2006. doi: 10.1190/1.2215359. URL http://dx.doi.org/10.1190/1.2215359. → pages 61

A. J. Berkhout. A unified approach to acoustical reflection imaging. I: The forward model. *The Journal of the Acoustical Society of America*, 93(4):2005–2016, 1993. ISSN 00014966. doi: 10.1121/1.406714. URL http://link.aip.org/link/JASMAN/v93/i4/p2005/s1&Agg=doi. → pages 117

A. J. Berkhout. Changing the mindset in seismic data acquisition. *The Leading Edge*, 27(7): 924–938, 2008b. doi: 10.1190/1.2954035. URL http://tle.geoscienceworld.org/cgi/content/abstract/27/7/924. → pages 78

F. J. Billette and S. Brandsberg-Dahl. The 2004 bp velocity benchmark. In *67th EAGE Conference & Exhibition*, 2004. → pages 90

B. Biondi and W. W. Symes. Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging. *Geophysics*, (5), 69(5):1283, 2004. ISSN 00168033. doi: 10.1190/1.1801945. URL http://link.aip.org/link/GPYSA7/v69/i5/p1283/s1&Agg=doi. → pages 113, 114, 121, 124

B. Biondi, P. Sava, et al. Wave-equation migration velocity analysis. *69th Ann. Internat. Mtg Soc. of Expl. Geophys*, pages 1723–1726, 1999. → pages 113

A. Bourgeois, M. Bourget, P. Lailly, M. Poulet, P. Ricarte, and R. Versteeg. Marmousi, model and data. In *1990 workshop on Practical Aspects of Seismic Data Inversion, European Association of Exploration Geophysicists*, volume 5-16. EAGE, 1991. → pages 90, 127

S. Brandsberg-Dahl, M. de Hoop, and B. Ursin. Focusing in dip and ava compensation on scattering angle/azimuth common image gathers. *Geophysics*, (1), 68(1):232–254, 2003. → pages 114, 122, 124

S. Burer and R. D. Monteiro. Local minima and convergence in low-rank semidefinite programming. *Mathematical Programming*, 103(3):427–444, 2005. → pages 60

S. Burer and R. D. C. Monteiro. Local minima and convergence in low-rank semidefinite programming. *Mathematical Programming*, 103:2005, 2003. → pages 15, 18, 41

J. Caldwell and C. Walker. An overview of marine seismic operations. *The international association of oil and gas producers*, January 24 2017. URL http://www.ogp.org.uk/pubs/448.pdf. → pages xiii, 1, 2, 3

E. J. Candès and L. Demanet. The curvelet representation of wave propagators is optimally sparse. *Communications on Pure and Applied Mathematics*, 58(11):1472–1528, 2005. ISSN 1097-0312. doi: 10.1002/cpa.20078. URL http://dx.doi.org/10.1002/cpa.20078. → pages 81

E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009. → pages 54, 82

E. J. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies. *Information Theory, IEEE Transactions on*, 52(12):5406–5425, December 2006. ISSN 0018-9448. doi: 10.1109/TIT.2006.885507. → pages 13

E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *CPAM*, 59(8):1207–1223, 2006. → pages 144

E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3), May 2011. → pages 13

E. J. Candès, T. Strohmer, and V. Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013. ISSN 1097-0312. doi: 10.1002/cpa.21432. URL http://dx.doi.org/10.1002/cpa.21432. → pages 21

A. Canning and G. Gardner. Reducing 3-d acquisition footprint for 3-d dmo and 3-d prestack migration. *Geophysics*, 63(4):1177–1183, 1998. doi: 10.1190/1.1444417. → pages 49

S. Chandrasekaran, P. Dewilde, M. Gu, W. Lyons, and T. Pals. A fast solver for hss representations via sparse matrices. *SIAM J. Matrix Analysis Applications*, 29(1):67–81, 2006. doi: http://dx.doi.org/10.1137/050639028. → pages 80, 84, 149

C.-H. Chang and C.-W. Ha. Sharp inequalities of singular values of smooth kernels. *Integral Equations and Operator Theory*, 35(1):20–27, 1999. → pages 65

H. Chauris, M. S. Noble, G. Lambaré, and P. Podvin. Migration velocity analysis from locally coherent events in 2-d laterally heterogeneous media, part i: Theoretical aspects. *Geophysics*, 67(4):1202–1212, 2002. → pages 113

Y. Chen, J. Yuan, Z. Jin, K. Chen, and L. Zhang. Deblending using normal moveout and median filtering in common-midpoint gathers. *Journal of Geophysics and Engineering*, 11(4):045012, 2014. → pages 81, 93, 146

J. Cheng and M. D. Sacchi. Separation of simultaneous source data via iterative rank reduction. In *SEG Technical Program Expanded Abstracts*, pages 88–93, 2013. URL http://dx.doi.org/10.1190/segam2013-1313.1. → pages 79, 106

J. Claerbout. *Imaging the earth's interior*. Blackwell Scientific Publishers, 1985a. → pages 128

J. F. Claerbout. Coarse grid calculations of waves in inhomogeneous media with application to delineation of complicated seismic structure. *Geophysics*, 35(3):407–418, 1970. doi: 10.1190/1.1440103. URL http://geophysics.geoscienceworld.org/content/35/3/407.abstract. → pages 113, 128

J. F. Claerbout. *Fundamentals of geophysical data processing*. Pennwell Books, Tulsa, OK, 1985b. → pages 113

W. J. Curry. Interpolation with fourier radial adaptive thresholding. *79th Annual International Meeting, SEG, Expanded Abstracts*, pages 3259–3263, 2009. doi: 10.1190/1.3255536. URL http://library.seg.org/doi/abs/10.1190/1.3255536. → pages 49

C. de Bruin, C. Wapenaar, and A. Berkhout. Angle-dependent reflectivity by means of prestack migration. *Geophysics*, (9), 55(9):1223, Sept. 1990. ISSN 1070485X. doi: 10.1190/1.1442938. URL http://library.seg.org/getabs/servlet/GetabsServlet?prog=normal&id=GPYSA7000055000009001223000001&idtype=cvips&gifs=yes. → pages 113, 114, 121, 123

R. de Kok and D. Gillespie. A universal simultaneous shooting technique. In *64th EAGE Conference and Exhibition*, 2002. → pages 78, 105

L. Demanet. *Curvelets, Wave Atoms, and Wave Equations*. PhD thesis, California Institute of Technology, 2006a. → pages 35

L. Demanet. *Curvelets, Wave Atoms, and Wave Equations*. PhD thesis, California Institute of Technology, 2006b. → pages 56

S. M. Doherty and J. F. Claerbout. Velocity analysis based on the wave equation. Technical Report 1, Stanford Exploration Project, 1974. URL http://sepwww.stanford.edu/oldreports/sep01/01_12.pdf. → pages 113

D. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006a. → pages 13, 49

D. L. Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006b. → pages 6, 144

P. Doulgeris, K. Bube, G. Hampson, and G. Blacquiere. Convergence analysis of a coherency-constrained inversion for the separation of blended data. *Geophysical Prospecting*, 60 (4):769–781, 2012. doi: 10.1111/j.1365-2478.2012.01088.x. → pages 78

A. A. Duchkov and V. Maarten. Velocity continuation in the downward continuation approach to seismic imaging. *Geophysical Journal International*, 176(3):909–924, 2009. → pages 113

A. Duijndam, M. Schonewille, and C. Hindriks. Reconstruction of band-limited signals, irregularly sampled along one spatial direction. *Geophysics*, 64(2):524–538, 1999. doi: 10.1190/1.1444559. → pages 49

B. Engquist and L. Ying. Fast directional multilevel algorithms for oscillatory kernels. *SIAM Journal on Scientific Computing*, 29(4):1710–1737, 2007. → pages 62

J.-M. Enjolras. Total pioneers cable-less 3d seismic surveys in uganda. *Oil in Uganda*, January 24 2017. URL http://www.oilinuganda.org/features/environment/uganda-pioneers-3d-seismic-surveys.html. → pages xiii, 2

M. Fazel. *Matrix rank minimization with applications*. PhD thesis, Stanford University, 2002. → pages 13

M. Figueiredo, R. Nowak, and S. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):586 –597, dec. 2007. → pages 15

M. Friedlander, H. Mansour, R. Saab, and O. Yilmaz. Recovering compressively sampled signals using partial support information. *IEEE Transactions on Information Theory*, 58(1), January 2011. → pages 14, 26

M. P. Friedlander and M. Schmidt. Hybrid Deterministic-Stochastic Methods for Data Fitting. *SIAM Journal on Scientific Computing*, 34(3):A1380–A1405, May 2012. ISSN 1064-8275. doi: 10.1137/110830629. URL http://epubs.siam.org/doi/abs/10.1137/110830629. → pages 126

S. Funk. Netflix update: Try this at home, December 2006. URL http://sifter.org/~simon/journal/20061211.html. → pages 28

S. Gandy, B. Recht, and I. Yamada. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Problems*, 27(2):025010, Jan. 2011. → pages 60

R. Giryes, M. Elad, and Y. C. Eldar. The projected gsure for automatic parameter tuning in iterative shrinkage methods. *Applied and Computational Harmonic Analysis*, 30(3):407–422, 2011. → pages 15, 20

D. Gross. Recovering Low-Rank Matrices From Few Coefficients in Any Basis. *IEEE Transactions on Information Theory*, 57:1548–1566, 2011. → pages 28

E. Haber, M. Chung, and F. Herrmann. An effective method for parameter estimation with pde constraints with multiple right-hand sides. *SIAM Journal on Optimization*, 22(3):739–757, 2012. doi: 10.1137/11081126X. URL http://epubs.siam.org/doi/abs/10.1137/11081126X. → pages 125, 126, 127

N. Halko, P.-G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011a. → pages 79

N. Halko, P.-G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011b. → pages 49

N. Halko, P.-G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2):217–288, 2011c. → pages 85

G. Hampson, J. Stefani, and F. Herkenhoff. Acquisition using simultaneous sources. *The Leading Edge*, 27(7):918–923, 2008. doi: 10.1190/1.2954034. URL http://tle.geoscienceworld.org/cgi/content/abstract/27/7/918. → pages 78

S. Hegna and G. E. Parkes. Method for acquiring and processing marine seismic data to extract and constructively use the up-going and down-going wave-fields emitted by the source (s), november 2012. US Patent Application 13/686,408. → pages 5, 78

G. Hennenfent and F. J. Herrmann. Seismic denoising with nonuniformly sampled curvelets. *Computing in Science & Engineering*, 8(3):16–25, 05 2006a. doi: 10.1109/MCSE.2006.49. URL https://www.slim.eos.ubc.ca/Publications/Public/Journals/CiSE/2006/hennenfent06CiSEsdn/hennenfent06CiSEsdn.pdf. → pages 81

G. Hennenfent and F. J. Herrmann. Application of stable signal recovery to seismic data interpolation. In *SEG Technical Program Expanded Abstracts*, volume 25, pages 2797–2801. SEG, 2006b. → pages 49

G. Hennenfent, E. van den Berg, M. P. Friedlander, and F. J. Herrmann. New insights into one-norm solvers from the Pareto curve. *Geophysics*, 73(4):A23–A26, 07 2008. doi: 10.1190/1.2944169. URL https://www.slim.eos.ubc.ca/Publications/Public/Journals/Geophysics/2008/hennenfent08GEOnii/hennenfent08GEOnii.pdf. → pages 82

G. Hennenfent, L. Fenelon, and F. J. Herrmann. Nonequispaced curvelet transform for seismic data reconstruction: A sparsity-promoting approach. *Geophysics*, 75(6):WB203–WB210, 12 2010. URL https://www.slim.eos.ubc.ca/Publications/Public/Journals/Geophysics/2010/hennenfent2010GEOPnct/hennenfent2010GEOPnct.pdf. → pages 144

F. J. Herrmann and G. Hennenfent. Non-parametric seismic data recovery with curvelet frames. *Geophysical Journal International*, 173(1):233–248, 2008a. ISSN 1365-246X. → pages 13, 30, 32, 49, 62

F. J. Herrmann and G. Hennenfent. Non-parametric seismic data recovery with curvelet frames. *Geophysical Journal International*, 173:233–248, April 2008b. URL https://www.slim.eos.ubc.ca/Publications/Public/Journals/GeophysicalJournalInternational/2008/herrmann08nps/herrmann08nps.pdf. → pages 145

F. J. Herrmann, M. P. Friedlander, and O. Yilmaz. Fighting the curse of dimensionality: Compressive sensing in exploration seismology. *Signal Processing Magazine, IEEE*, 29(3): 88–100, 2012a. ISSN 1053-5888. doi: 10.1109/MSP.2012.2185859. URL https://www.slim.eos.ubc.ca/Publications/Public/Journals/IEEESignalProcessingMagazine/2012/Herrmann11TRfcd/Herrmann11TRfcd.pdf. → pages 13

F. J. Herrmann, M. P. Friedlander, and O. Yilmaz. Fighting the curse of dimensionality: compressive sensing in exploration seismology. *IEEE Signal Processing Magazine*, 29:88–100, May 2012b. doi: 10.1109/MSP.2012.2185859. → pages 20

D. Hill, C. Combee, and J. Bacon. Over/under acquisition and data processing: the next quantum leap in seismic technology? *First Break*, 24(6):81–95, 2006. → pages 5, 78

P. Huber. *Robust Statistics*. Wiley, 1981. → pages 14, 23

S. Huo, Y. Luo, and P. Kelamis. Simultaneous sources separation via multi-directional vector-median filter. *SEG Technical Program Expanded Abstracts*, 28(1):31–35, 2009. doi: 10.1190/1.3255522. URL http://link.aip.org/link/?SGA/28/31/1. → pages 78

P. Jain, R. Meka, and I. Dhillon. Guaranteed rank minimization via singular value projection. In *In NIPS 2010*, 2010. → pages 23

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*, STOC '13, pages 665–674, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2029-0. doi: 10.1145/2488608.2488693. URL http://doi.acm.org/10.1145/2488608.2488693. → pages x, 21, 22

B. Jumah and F. J. Herrmann. Dimensionality-reduced estimation of primaries by sparse inversion. *Geophysical Prospecting*, 62(5):972–993, 09 2014. doi: 10.1111/1365-2478.12113. URL http://onlinelibrary.wiley.com/doi/10.1111/1365-2478.12113/abstract. → pages 84

M. N. Kabir and D. Verschuur. Restoration of missing offsets by parabolic radon transform1. *Geophysical Prospecting*, 43(3):347–368, 1995. ISSN 1365-2478. doi: 10.1111/j.1365-2478.1995.tb00257.x. URL http://dx.doi.org/10.1111/j.1365-2478.1995.tb00257.x. → pages 49

B. Kanagal and V. Sindhwani. Rank selection in low-rank matrix approximations: A study of cross-validation for nmfs. *Advances in Neural Information Processing Systems*, 1:10, 2010. → pages 51

O. Koefoed. On the effect of poisson's ratios of rock strata on the reflection coefficients of plane waves. *Geophysical Prospecting*, 3(4):381–387, 1955. ISSN 1365-2478. doi: 10.1111/j.1365-2478.1955.tb01383.x. URL http://dx.doi.org/10.1111/j.1365-2478.1955.tb01383.x. → pages 122

Z. Koren and I. Ravve. Full-azimuth subsurface angle domain wavefield decomposition and imaging part i: Directional and reflection image gathers. *Geophysics*, 76(1):S1–S13, 2011. → pages 113

J. Krebs, J. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M. Lacasse. Fast full-wavefield seismic inversion using encoded sources. *Geophysics*, (6), 74(6):P177–P188, 2009. doi: 10.1190/1.3230502. URL http://dx.doi.org/10.1190/1.3230502. → pages 125

N. Kreimer. *Multidimensional seismic data reconstruction using tensor analysis*. 2013. → pages 144, 145, 146

N. Kreimer and M. Sacchi. A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation. *Geophysics*, 77, no. 3(3):V113–V122, 2012a. doi: 10.1190/geo2011-0399.1. → pages 49, 50, 82

N. Kreimer and M. D. Sacchi. A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation. *GEOPHYSICS*, 77(3):V113–V122, 2012b. doi: 10.1190/geo2011-0399.1. → pages 107

N. Kreimer and M. D. Sacchi. Tensor completion via nuclear norm minimization for 5d seismic data reconstruction. In *83rd Annual International Meeting, SEG, Expanded Abstracts*, pages 1–5, 2012c. doi: 10.1190/segam2012-0529.1. URL http://library.seg.org/doi/abs/10.1190/segam2012-0529.1. → pages 49

N. Kreimer, A. Stanton, and M. D. Sacchi. Tensor completion based on nuclear norm minimization for 5d seismic data reconstruction. *Geophysics*, 78, no. 6(6):V273–V284, 2013. → pages 51, 60, 69

H. Kuhel and M. Sacchi. Least-squares wave-equation migration for avp/ava inversion. *Geophysics*, (1), 68(1):262–273, 2003. doi: 10.1190/1.1543212. URL http://library.seg.org/doi/abs/10.1190/1.1543212. → pages 121

R. Kumar, H. Mansour, A. Y. Aravkin, and F. J. Herrmann. Reconstruction of seismic wavefields via low-rank matrix factorization in the hierarchical-separable matrix representation. *84th Annual International Meeting, SEG, Expanded Abstracts*, pages 3628–3633, 9 2013a. doi: 10.1190/segam2013-1165.1. URL https://www.slim.eos.ubc.ca/Publications/Public/Conferences/SEG/2013/kumar2013SEGHSS/kumar2013SEGHSS.pdf. → pages 62, 82, 84, 85, 149

R. Kumar, T. van Leeuwen, and F. J. Herrmann. Efficient WEMVA using extended images. In *SEG Workshop on Advances in Model Building, Imaging, and FWI; Houston*, 9 2013b. → pages 124

R. Kumar, A. Y. Aravkin, E. Esser, H. Mansour, and F. J. Herrmann. SVD-free low-rank matrix factorization : wavefield reconstruction via jittered subsampling and reciprocity. *76st Conference and Exhibition, EAGE, Extended Abstracts*, 06 2014. doi:

10.3997/2214-4609.20141394. URL https://www.slim.eos.ubc.ca/Publications/Public/Conferences/EAGE/2014/kumar2014EAGErank/kumar2014EAGErank.pdf. → pages 63

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*, 80(05):V97–V114, 09 2015a. doi: 10.1190/geo2014-0369.1. URL https://www.slim.eos.ubc.ca/Publications/Public/Journals/Geophysics/2015/kumar2014GEOPemc/kumar2014GEOPemc.pdf. (Geophysics). → pages 79, 80, 83, 106

R. Kumar, H. Wason, and F. J. Herrmann. Source separation for simultaneous towed-streamer marine acquisition –- a compressed sensing approach. *Geophysics*, 80(06):WD73–WD88, 11 2015b. doi: 10.1190/geo2015-0108.1. URL https://www.slim.eos.ubc.ca/Publications/Public/Journals/Geophysics/2015/kumar2015sss/kumar2015sss_revised.pdf. (Geophysics). → pages 106

R. Lago, A. Petrenko, Z. Fang, and F. J. Herrmann. Fast solution of time-harmonic wave-equation for full-waveform inversion. In *EAGE Annual Conference Proceedings*, 06 2014. doi: 10.3997/2214-4609.20140812. URL https://www.slim.eos.ubc.ca/Publications/Public/Conferences/EAGE/2014/lago2014EAGEfst/lago2014EAGEfst.pdf. → pages 121

K. L. Lange, R. J. A. Little, and J. M. G. Taylor. Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 84(408):881–896, 1989. → pages 14, 24

R. M. Lansley, M. Berraki, and M. M. M. Gros. Seismic array with spaced sources having variable pressure, september 2007. US Patent Application 12/998,723. → pages 5, 78

A. Lee, B. Recht, R. Salakhutdinov, Nathan Srebro, and J. A. Tropp. Practical Large-Scale Optimization for Max-Norm Regularization. In *Advances in Neural Information Processing Systems*, 2010a. → pages 17, 51, 80, 89, 106, 109

J. Lee, B. Recht, R. Salakhutdinov, N. Srebro, and J. Tropp. Practical large-scale optimization for max-norm regularization. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23*, pages 1297–1305. 2010b. → pages 14, 17, 20, 144

S. A. Levin. Principle of reverse-time migration. *Geophysics*, 49(5):581–583, 1984. → pages 113

C. Li, C. C. Mosher, L. C. Morley, Y. Ji, and J. D. Brewer. Joint source deblending and reconstruction for seismic data. *SEG Technical Program Expanded Abstracts 2013*, pages 82–87, 2013. → pages 5

E. Liberty, F. Woolfe, P.-G. Martinsson, V. Rokhlin, and M. Tygert. Randomized algorithms for the low-rank approximation of matrices. *National Academy of Sciences*, 104(51):20167–20172, 2007. doi: 10.1073/pnas.0709640104. → pages 49, 79

Z. Lin and S. Wei. A block lanczos with warm start technique for accelerating nuclear norm minimization algorithms. *CoRR*, abs/1012.0365, 2010. → pages 23

A. Long. A new seismic method to significantly improve deeper data character and interpretability. 2009. → pages 5, 78

A. S. Long, E. von Abendorff, M. Purves, J. Norris, and A. Moritz. Simultaneous long offset (SLO) towed streamer seismic acquisition. In *75th EAGE Conference & Exhibition*, 2013. → pages 5, 78

S. Lu, N. Whitmore, A. Valenciano, and N. Chemingui. Illumination from 3d imaging of multiples: An analysis in the angle domain. *84th SEG, Denver, USA, Expended Abstracts*, 2014. → pages 114

S. MacKay and R. Abma. Imaging and velocity analysis with depth-focusing analysis. *Geophysics*, (12), 57(12):1608–1622, 1992. URL http://dx.doi.org/10.1190/1.1443228. → pages 114

A. Mahdad, P. Doulgeris, and G. Blacquiere. Separation of blended data by iterative estimation and subtraction of blending interference noise. *Geophysics*, 76(3):Q9–Q17, 2011. doi: 10.1190/1.3556597. URL http://link.aip.org/link/?GPY/76/Q9/1. → pages 78

F. Mahmoudian and G. F. Margrave. A review of angle domain common image gathers. *Technical Report, University of Calgary*, 2009. URL http://www.crewes.org/ForOurSponsors/ResearchReports/2009/CRR200953.pdf. → pages 121

M. W. Mahoney. Randomized algorithms for matrices and data. *Foundations and Trends in Machine Learning*, 3(2):123–224, Feb. 2011. ISSN 1935-8237. doi: 10.1561/2200000035. URL http://dx.doi.org/10.1561/2200000035. → pages 49

J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing*, 17(1):53–69, Jan. 2008. → pages 13

H. Mansour, R. Saab, P. Nasiopoulos, and R. Ward. Color image desaturation using sparse reconstruction. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 778–781, March 2010. → pages 13

H. Mansour, F. J. Herrmann, and O. Yilmaz. Improved wavefield reconstruction from randomized sampling via weighted one-norm minimization. *submitted to Geophysics*, 2012a. → pages 37

H. Mansour, H. Wason, T. T. Lin, and F. J. Herrmann. Randomized marine acquisition with compressive sampling matrices. *Geophysical Prospecting*, 60(4):648–662, July 2012b. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2478.2012.01075.x/abstract. → pages 13

H. Mansour, H. Wason, T. T. Lin, and F. J. Herrmann. Randomized marine acquisition with compressive sampling matrices. *Geophysical Prospecting*, 60(4):648–662, 2012c. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2478.2012.01075.x/abstract. → pages 78, 80

H. Mansour, F. J. Herrmann, and O. Yilmaz. Improved wavefield reconstruction from randomized sampling via weighted one-norm minimization. *Geophysics*, 78, no. 5(5):V193–V206, 08 2013. doi: 10.1190/geo2012-0383.1. URL https://www.slim.eos.ubc.ca/Publications/Public/Journals/Geophysics/2013/mansour2013GEOPiwr/mansour2013GEOPiwr.pdf. → pages 62, 96

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. → pages 79, 106

R. A. Maronna, D. Martin, and Yohai. *Robust Statistics*. Wiley Series in Probability and Statistics. Wiley, 2006. → pages 23, 24

B. Mishra, G. Meyer, F. Bach, and R. Sepulchre. Low-rank optimization with trace norm penalty. *SIAM Journal on Optimization*, 23(4):2124–2149, 2013. → pages 50

N. Moldoveanu and S. Fealy. Multi-vessel coil shooting acquisition. *Patent Application*, US 20100142317 A1, 2010. → pages 78

N. Moldoveanu and J. Quigley. Random sampling for seismic acquisition. In *73rd EAGE Conference & Exhibition*, 2011. → pages 78

N. Moldoveanu, L. Combee, M. Egan, G. Hampson, L. Sydora, and W. Abriel. Over/under towed-streamer acquisition: A method to extend seismic bandwidth to both higher and lower frequencies. *The Leading Edge*, 26:41–58, 01 2007. URL http://www.slb.com/~/media/Files/westerngeco/resources/articles/2007/jan07_tle_overunder.pdf. → pages 5, 78

I. Moore. Simultaneous sources - processing and applications. In *72nd EAGE Conference and Exhibition*, 2010. → pages 78

I. Moore, B. Dragoset, T. Ommundsen, D. Wilson, C. Ward, and D. Eke. Simultaneous source separation using dithered sources. *SEG Technical Program Expanded Abstracts*, 27(1): 2806–2810, 2008. doi: 10.1190/1.3063928. URL http://link.aip.org/link/?SGA/27/2806/1. → pages 78

C. Mosher, C. Li, L. Morley, Y. Ji, F. Janiszewski, R. Olson, and J. Brewer. Increasing the efficiency of seismic data acquisition via compressive sensing. *The Leading Edge*, 33(4):386–391, 2014. → pages 78

W. Mulder. Subsurface offset behaviour in velocity analysis with extended reflectivity images. *Geophysical Prospecting*, 62(1):17–33, 2014. ISSN 1365-2478. doi: 10.1111/1365-2478.12073. URL http://dx.doi.org/10.1111/1365-2478.12073. → pages 124

R. N. Neelamani, C. E. Krohn, J. R. Krebs, J. K. Romberg, M. Deffenbaugh, and J. E. Anderson. Efficient seismic forward modeling using simultaneous random sources and sparsity. *Geophysics*, 75(6):WB15–WB27, 2010. doi: 10.1190/1.3509470. URL http://geophysics.geoscienceworld.org/content/75/6/WB15.abstract. → pages 13

J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, Aug. 2000. ISBN 0387987932. URL http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0387987932. → pages 126

V. Oropeza and M. Sacchi. Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis. *Geophysics*, 76(3):V25–V32, 2011. → pages 13, 28, 49, 79, 82

A. B. Owen and P. O. Perry. Bi-cross-validation of the svd and the nonnegative matrix factorization. *The Annals of Applied Statistics*, 3(2):564–594, 06 2009. doi: 10.1214/08-AOAS227. → pages 51

S. Oymak, A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi. Simultaneously structured models with application to sparse and low-rank matrices. *IEEE Transactions on Information Theory*, pages 1–1, 2012. → pages 61

M. Prucha, B. Biondi, W. Symes, et al. Angle-domain common image gathers by wave-equation migration. *69th Ann. Internat. Mtg: Soc. of Expl. Geophys*, pages 824–827, 1999. → pages 113

B. Recht. A simpler approach to matrix completion. *The Journal of Machine Learning Research*, 7:3413–3430, 2011. → pages 54

B. Recht and C. Ré. Parallel stochastic gradient algorithms for large-scale matrix completion. In *Optimization Online*, 2011. → pages 14, 17, 20, 28, 89, 109, 144

B. Recht and C. Ré. Parallel stochastic gradient algorithms for large-scale matrix completion. *Mathematical Programming Computation*, 5(2):201–226, 2013. ISSN 1867-2949. doi: 10.1007/s12532-013-0053-8. URL http://dx.doi.org/10.1007/s12532-013-0053-8. → pages 51, 60

B. Recht, M. Fazel, and P. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010a. doi: 10.1137/070697835. → pages 49, 82

B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010b. → pages 13, 18, 88, 108

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*, ICML '05, pages 713–719, New York, NY, USA, 2005a. ACM. ISBN 1-59593-180-5. → pages 89, 109

J. D. M. Rennie and N. Srebro. Fast maximum margin matrix factorization for collaborative prediction. In *ICML '05 Proceedings of the 22nd international conference on Machine learning*, pages 713 – 719, 2005b. → pages 14, 17

J. Rickett and P. Sava. Offset and angle domain common imagepoint gathers for shot profile migration. *Geophysics*, (3), 67(3):883–889, 2002. doi: 10.1190/1.1484531. URL http://library.seg.org/doi/abs/10.1190/1.1484531. → pages 121

RigZone. How does marine seismic work. *RigZone*, January 24 2017. URL http://www.rigzone.com/training/insight.asp-insight_id=303. → pages xiii, 2

R. Rockafellar and R.-B. Wets. *Variational Analysis*, volume 317. Springer, 1998. → pages 16

M. Sacchi, T. Ulrych, and C. Walker. Interpolation and extrapolation using a high-resolution discrete fourier transform. *Signal Processing, IEEE Transactions on*, 46(1):31 –38, jan 1998. → pages 28, 32, 49

M. Sacchi, S. Kaplan, and M. Naghizadeh. F-x gabor seismic data reconstruction. *71st Conference and Exhibition, EAGE, Extended Abstracts*, 2009. → pages 49

M. D. Sacchi and B. Liu. Minimum weighted norm wavefield reconstruction for ava imaging. *Geophysical Prospecting*, 53(6):787–801, 2005. ISSN 1365-2478. doi: 10.1111/j.1365-2478.2005.00503.x. URL http://dx.doi.org/10.1111/j.1365-2478.2005.00503.x. → pages 49

P. Sava and S. Fomel. Angle-domain common-image gathers by wavefield continuation methods. *Geophysics*, (3), 68(3):1065–1074, 2003. doi: 10.1190/1.1581078. URL http://library.seg.org/doi/abs/10.1190/1.1581078. → pages 121

P. Sava and I. Vasconcelos. Extended imaging conditions for wave-equation migration. *Geophysical Prospecting*, 59(1):35–55, Jan. 2011. ISSN 00168025. doi: 10.1111/j.1365-2478.2010.00888.x. URL http://doi.wiley.com/10.1111/j.1365-2478.2010.00888.x. → pages 113, 114, 116, 121, 124, 127, 143

P. C. Sava and B. Biondi. Wave-equation migration velocity analysis. I. Theory. *Geophysical Prospecting*, 52:593–606, 2004. URL http://dx.doi.org/10.1111/j.1365-2478.2004.00447.x. → pages 114

P. C. Sava and S. Fomel. Time-shift imaging condition in seismic migration. *Geophysics*, (6), 71 (6):S209—-S217, 2006. URL http://dx.doi.org/10.1190/1.2338824. → pages 114

H. Schaeffer and S. Osher. A low patch-rank interpretation of texture. *SIAM Journal on Imaging Sciences*, 6(1):226–262, 2013. → pages 50

P. Shen and W. W. Symes. Automatic velocity analysis via shot profile migration. *Geophysics*, 73 (5):VE49–VE59, 2008. URL http://dx.doi.org/10.1190/1.2972021. → pages 114, 124

R. Shuey. A simplification of the zoeppritz equations. *Geophysics*, (4), 50(4):609–614, 1985. doi: 10.1190/1.1441936. URL http://dx.doi.org/10.1190/1.1441936. → pages 123

M. Signoretto, R. Van de Plas, B. De Moor, and J. A. Suykens. Tensor versus matrix completion: a comparison with application to spectral data. *Signal Processing Letters, IEEE Transactions on Information Theory*, 18(7):403–406, 2011. → pages 61

C. D. Silva and F. J. Herrmann. Hierarchical Tucker tensor optimization - applications to 4D seismic data interpolation. In *EAGE Annual Conference Proceedings*, 06 2013a. doi: 10.3997/2214-4609.20130390. URL https://www.slim.eos.ubc.ca/Publications/Public/Conferences/EAGE/2013/dasilva2013EAGEhtucktensor/dasilva2013EAGEhtucktensor.pdf. → pages 50, 108

C. D. Silva and F. J. Herrmann. Structured tensor missing-trace interpolation in the Hierarchical Tucker format. In *SEG Technical Program Expanded Abstracts*, volume 32, pages 3623–3627, 9 2013b. doi: 10.1190/segam2013-0709.1. URL https://www.slim.eos.ubc.ca/Publications/Public/Conferences/SEG/2013/dasilva2013SEGhtuck/dasilva2013SEGhtuck.pdf. → pages 82

C. D. Silva and F. J. Herrmann. Hierarchical Tucker tensor optimization - applications to 4d seismic data interpolation. In *EAGE*, 06 2013c. URL https://www.slim.eos.ubc.ca/Publications/Public/Conferences/EAGE/2013/dasilva2013EAGEhtucktensor/dasilva2013EAGEhtucktensor.pdf. → pages 35

S. Sinha, P. S. Routh, P. D. Anno, and J. P. Castagna. Spectral decomposition of seismic data with continuous-wavelet transform. *Geophysics*, 70, no. 6(6):P19–P25, 2005. → pages 50

H. F. Smith. A Hardy space for Fourier integral operators. *Journal of Geometric Analysis*, 8(4): 629–653, 1998. → pages 81

N. Srebro. *Learning with matrix factorizations, PhD Thesis*. PhD thesis, Massachusetts Institute of Technology, 2004. URL http://hdl.handle.net/1721.1/28743. → pages 58

J.-L. Starck, M. Elad, and D. Donoho. Image decomposition via the combination of sparse representation and a variational approach. *IEEE Transaction on Image Processing*, 14(10), 2005. → pages 13

J. Stefani, G. Hampson, and F. Herkenhoff. Acquisition using simultaneous sources. In *69th EAGE Conference and Exhibition*, 2007. → pages 78

C. C. Stolk and W. W. Symes. Smooth objective functionals for seismic velocity inversion. *Inverse Problems*, 19:73–89, 2003. → pages 124

C. C. Stolk, M. V. de Hoop, and W. W. Symes. Kinematics of shot-geophone migration. *Geophysics*, (6), 74(6):WCA19—-WCA34, 2009. ISSN 00168033. doi: 10.1190/1.3256285. URL http://library.seg.org/getabs/servlet/GetabsServlet?prog=normal&id=GPYSA70000740000060WCA19000001&idtype=cvips&gifs=yeshttp://dx.doi.org/10.1190/1.3256285. → pages 113

M. Stoll. A krylov-schur approach to the truncated svd. *Linear Algebra and its Applications*, 436 (8):2795 – 2806, 2012. ISSN 0024-3795. doi: http://dx.doi.org/10.1016/j.laa.2011.07.022. URL http://www.sciencedirect.com/science/article/pii/S0024379511005349. → pages 57

R. Sun and Z.-Q. Luo. Guaranteed matrix completion via non-convex factorization. *http://arxiv.org/abs/1411.8003, accessed 28 November 2014*, 2014. → pages 60

W. Symes. Approximate linearized inversion by optimal scaling of prestack depth migration. *Geophysics*, (2), 73(2):R23–R35, 2008a. doi: 10.1190/1.2836323. URL http://dx.doi.org/10.1190/1.2836323. → pages 126

W. Symes. Seismic inverse poblems : recent developments in theory and practice. In A. Louis, S. Arridge, and B. Rundell, editors, *Proceedings of the Inverse Problems from Theory to Applications Conference*, pages 2–6. IOP Publishing, 2014. → pages 124

W. W. Symes. Migration velocity analysis and waveform inversion. *Geophysical Prospecting*, 56 (6):765–790, 2008b. → pages 124

W. W. Symes and J. J. Carazzone. Velocity inversion by differential semblance optimization. *Geophysics*, (5), 56(5):654–663, 1991. → pages 113, 114, 124

W. W. Symes, D. Sun, and M. Enriquez. From modelling to inversion: designing a well-adapted simulator. *Geophysical Prospecting*, 59(5):814–833, 2011a. ISSN 1365-2478. doi: 10.1111/j.1365-2478.2011.00977.x. URL http://dx.doi.org/10.1111/j.1365-2478.2011.00977.x. → pages 61, 90

W. W. Symes, D. Sun, and M. Enriquez. From modelling to inversion: designing a well-adapted simulator. *Geophysical Prospecting*, 59(5):814–833, 2011b. ISSN 1365-2478. doi: 10.1111/j.1365-2478.2011.00977.x. URL http://dx.doi.org/10.1111/j.1365-2478.2011.00977.x. → pages 123

A. ten Kroode, D.-J. Smit, and A. Verdel. Linearized inversed scattering in the presence of caustics. In *SPIE's 1994 International Symposium on Optics, Imaging, and Instrumentation*, pages 28–42. International Society for Optics and Photonics, 1994. → pages 113

P. Torben Hoy. A step change in seismic imaging–using a unique ghost free source and receiver system. *CSEG, Geoconvetion*, 2013. → pages 5, 78

D. Trad. Five-dimensional interpolation: Recovering from acquisition constraints. *Geophysics*, 74, no. 6(6):V123–V132, 2009. → pages 49

S. Trickett and L. Burroughs. Prestack rank-reducing noise suppression: Theory. *Society of Exploration Geophysicists*, 2009. → pages 82

S. Trickett, L. Burroughs, A. Milton, L. Walton, and R. Dack. Rank reduction based trace interpolation. *80th Annual International Meeting, SEG, Expanded Abstracts*, pages 3829–3833, 2010. doi: 10.1190/1.3513645. → pages 49

S. Trickett, L. Burroughs, and A. Milton. Interpolation using hankel tensor completion. *83rd Annual International Meeting, SEG, Expanded Abstracts*, pages 3634–3638, 2013. doi: 10.1190/segam2013-0416.1. URL http://library.seg.org/doi/abs/10.1190/segam2013-0416.1. → pages 50

T. van Leeuwen and F. J. Herrmann. 3D frequency-domain seismic inversion with controlled sloppiness. *SIAM Journal on Scientific Computing*, 36(5):S192–S217, 10 2014. doi: 10.1137/130918629. URL http://epubs.siam.org/doi/abs/10.1137/130918629. (SISC). → pages 121, 126

T. van Leeuwen, A. Y. Aravkin, and F. J. Herrmann. Seismic waveform inversion by stochastic optimization. *International Journal of Geophysics*, 2011, 12 2011. doi: 10.1155/2011/689041. URL https://www.slim.eos.ubc.ca/Publications/Public/Journals/InternationJournalOfGeophysics/2011/vanLeeuwen10IJGswi/vanLeeuwen10IJGswi.pdf. → pages 119, 125, 126

A. van Wijngaarden. *Imaging and Characterization of Angle-dependent Seismic Reflection Data*. PhD thesis, Delft University of Technology, 1998. URL http://books.google.ca/books?id=x5NaAAAACAAJ. → pages 114, 121

B. Vandereycken. Low-rank matrix completion by Riemannian optimization. *SIAM Journal on Optimization*, 23(2):1214—1236, 2013. → pages 29

J. Virieux and S. Operto. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26, 2009. doi: 10.1190/1.3238367. URL http://library.seg.org/doi/abs/10.1190/1.3238367. → pages 152

J. Wang, M. Ng, and M. Perz. Seismic data interpolation by greedy local radon transform. *Geophysics*, 75, no. 6(6):WB225–WB234, 2010. doi: 10.1190/1.3484195. → pages 49

L. Wang, J. Gao, W. Zhao, and X. Jiang. Nonstationary seismic deconvolution by adaptive molecular decomposition. In *Geoscience and Remote Sensing Symposium (IGARSS)*, pages 2189–2192. IEEE Transactions on Information Theory, 2011. → pages 50

H. Wason and F. J. Herrmann. Ocean bottom seismic acquisition via jittered sampling. In *EAGE*, 06 2013a. doi: 10.3997/2214-4609.20130379. URL https://www.slim.eos.ubc.ca/Publications/Public/Conferences/EAGE/2013/wason2013EAGEobs/wason2013EAGEobs.pdf. → pages 79, 96

H. Wason and F. J. Herrmann. Time-jittered ocean bottom seismic acquisition. 32:1–6, 9 2013b. doi: 10.1190/segam2013-1391.1. URL https://www.slim.eos.ubc.ca/Publications/Public/Conferences/SEG/2013/wason2013SEGtjo/wason2013SEGtjo.pdf. → pages 5, 78, 79, 80, 97, 105, 106, 146

Z. Wen, W. Yin, and Y. Zhang. Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm. *Mathematical Programming Computation*, 4 (4):333–361, 2012. → pages 70, 145

N. Whitmore et al. Iterative depth migration by backward time propagation. In *1983 SEG Annual Meeting*. Society of Exploration Geophysicists, 1983. → pages 113

T. Yang and P. Sava. Image-domain wavefield tomography with extended common-image-point gathers. *Geophysical Prospecting*, 63(5):1086–1096, 2015. ISSN 1365-2478. doi: 10.1111/1365-2478.12204. URL http://dx.doi.org/10.1111/1365-2478.12204. → pages 114, 124, 126

Y. Yang, J. Ma, and S. Osher. Seismic data reconstruction via matrix completion. *Inverse problem and imaging*, 7(4):1379–1392, 2013. → pages 49, 50