LARGE SCALE WAVEFIELD RECONSTRUCTION VIA WEIGHTED MATRIX COMPLETION AND SEISMIC SURVEY DESIGN

A Dissertation
Presented to
The Academic Faculty

By

Yijun Zhang

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering
.

Georgia Institute of Technology

August 2023

LARGE SCALE WAVEFIELD RECONSTRUCTION VIA WEIGHTED MATRIX COMPLETION AND SEISMIC SURVEY DESIGN

Thesis committee:

Dr. Felix J. Herrmann, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Faramarz Fekri
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Ghassan AlRegib
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Oscar Lopez
Ocean Engineering and Applied Technology
Florida Atlantic University

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

Date approved: May 18, 2023

For my father Zhang, Shaolin

ACKNOWLEDGMENTS

Before anything else, I would like to express my sincere gratitude to my PhD advisor, Professor Dr. Felix J. Herrmann, for his invaluable assistance and guidance throughout my research journey. I am truly grateful for the opportunity to complete my PhD dissertation at the Seismic Laboratory for Imaging and Modeling (SLIM) under his supervision. His instruction in analytical problem-solving and collaborative approach with an open mind have had a profound impact on me, not only as a researcher but also as an individual. I feel fortunate to have had his mentorship and will continue to benefit from his teachings in all aspects of my life.

I express my sincere appreciation to Professor Dr. Zhigang Peng, Professor Dr. Ghassan AlRegib, Professor Dr. Faramarz Fekri from Georgia Institute of Technology, and Professor Dr. Oscar Lopez from Florida Atlantic University for their valuable time and guidance as my advisory committee. I am grateful for their support.

I would like to express my gratitude for the friendship and support of all my coworkers at the SLIM lab, and would like to express my gratitude to them. In particular, I would like to thank Henryk Modzelewski for his professional assistance. Moreover, I would like to extend my appreciation to all the students and researchers at SLIM, including Dr. Rajiv Kumar, Dr. Gabrio Rizzuti, Dr. Philipp A. Witte, Dr. Mathias Louboutin, Dr. Mengmeng Yang, Dr. Shashin Sharan, Dr. Ali Siahkoohi, Ziyi Yin, Rafael Orozco, and other members of the SLIM group for their insightful scientific contributions and assistance.

I would like to thanks to Klaas Koster from Occidental for his valuable inputs and assistance. I would like to express my acknowledgement to Professor Dr. Rachel Kuske for her support during a particularly challenging phase of my PhD journey.

I would like to acknowledge the support from Occidental for providing the dataset. I would also like to thanks to the creators of the Julia programming language, as well as the developers of the Devito, JUDI, and JOLI software packages, which were invaluable tools for my research. Furthermore, I would like to thank AWS for their cloud computing support, which greatly facilitated my work. And I appreciate the assistance provided by ChatGPT in editing and improving my writing.

Finally, I would like to express my gratitude for the unwavering support of my wonderful family. I am deeply grateful to my parents, as well as my elder brother and sister, for their love and encouragement. In particular, I want to extend my heartfelt appreciation to my husband, Gang Liu, for his selfless love, patience, and support throughout my academic journey.

This research was partially funded by the US National Science Foundation grant OAC 2203821, the partners of the ML4Seismic Center, the SINBAD Consortium, Georgia Research Alliance and Georgia Institute of Technology.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Seismic data acquisition plays a crucial role in identifying potential oil and gas reservoirs during the early phases of exploration. However, obtaining finely sampled seismic data can be costly and physically impossible. Recent developments in Compressive Sensing have resulted in seismic data being increasingly collected at random along spatial coordinates. Although random sampling improves acquisition efficiency, it shifts the burden from seismic acquisition to data processing. Wavefield recovery is one of the required processes for reconstructing fully seismic data from coarsely subsampled data. Among the various techniques proposed for wavefield reconstruction, matrix completion methods are computationally efficient and straightforward to implement. These methods exploit the low-rank structure of fully seismic data. However, matrix completion performs well at low-to-mid frequencies and degrades at higher frequencies due to the failure of low-rank structure to accurately approximate higher frequencies. To address this issue, this thesis proposed a recursively weighted matrix completion method. Although effective, this method is computationally expensive, and a more efficient method for handling 2D seismic data was also proposed. Compared to 2D seismic data, 3D seismic data can detect reflections outside of the 2D plane but poses a computational challenge due to its large scale. To overcome this challenge, this thesis proposed a parallel weighted reconstruction method to improve the reconstruction of 3D seismic data. Land seismic data presents a greater challenge due to contamination by ground roll, which consists of surface waves with a high spatial frequency content and large amplitude. To address this issue, a practical workflow was proposed in this thesis to improve the recovery of land seismic data. Although matrix completion is an efficient technique for reconstructing fully seismic data, the optimal acquisition design is still being investigated. Recent studies have shown that the spectral gap can be used to predict and characterize the quality of wavefield reconstruction via matrix completion for a given subsampling mask. Based on these findings, a simulation-free seismic survey design

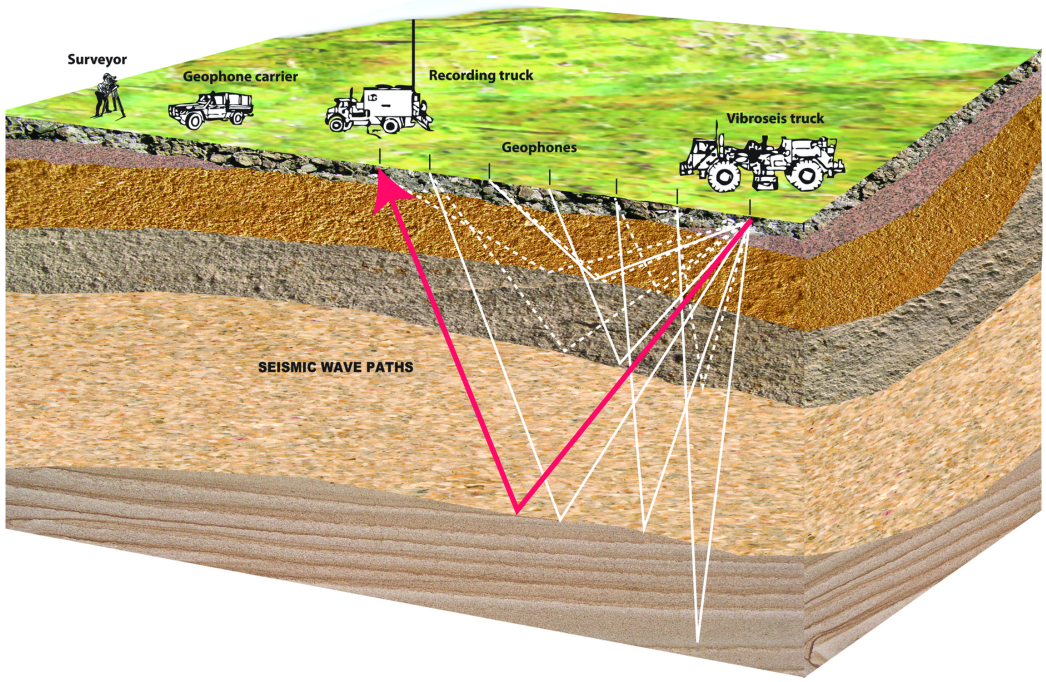
for both 2D and 3D seismic data was proposed in this thesis to obtain an improved subsampling survey by minimizing the spectral gap ratio. Furthermore, this concept was extended to the design of a time-lapse seismic survey, which is essential for reservoir management and monitoring geological carbon storage but is difficult and expensive to acquire. To improve the reconstruction of the time-lapse wavefield, a joint recovery model was proposed that leverages the benefits of the non-replicated baseline and monitor subsampled seismic data. A time-lapse seismic survey design that incorporates the joint recovery model with spectral gap was proposed to generate sparse, non-replicated time-lapse acquisition geometries that favor wavefield recovery.

CHAPTER 1
INTRODUCTION

Seismic data acquisition is essential in the early stages of oil and gas exploration, since it provides a higher degree of precision for predicting the physical properties of the earth's subsurface compared to other geophysical methods. Seismic sources, such as airguns and vibrators, are fired from either the land or sea surface (the land acquisition Figure 1.1a or marine acquisition Figure 1.1b), resulting in the propagation of acoustic waves into the subsurface of the earth. These waves are reflected when they encounter interfaces formed between rocks with distinct physical properties and are recorded by receivers, such as geophones and hydrophones, on the surface of the land or sea. Subsequent processing of the recorded seismic data allows us to obtain subsurface images and estimate the physical properties of the subsurface, which aid in identifying potential oil and gas reservoirs. Moreover, seismic data is also used to help drill wells, which are operationally complex and expensive, and extract oil and gas from the subsurface.

To achieve a higher degree of precision in subsurface images, seismic data must be acquired at fine grids, which can be both costly and time-consuming. However, recent developments in the field of Compressive Sensing (CS) have inspired seismic data collection methods that randomly sample along spatial coordinates to reduce acquisition time and costs (Candès, Romberg, and Tao 2006). While this approach increases acquisition productivity (Mosher, Li, Morley, Ji, Janiszewski, Olson, and Brewer 2014), it also shifts the burden from collecting data in the field to processing the data (Chiu 2019).

Wavefield recovery is one of the key steps in reconstructing seismic data from coarsely subsampled data. Transform-domain-based approaches have been used extensively for wavefield reconstruction. These transforms increase the sparsity of seismic data to varying degrees, which is a critical component of wavefield reconstruction. While these methods



(a)

(b)

Figure 1.1: Seismic acquisition. *(a)* Land seismic acquisition (image courtesy of DMP: *www.dmp.wa.gov.au*). *(b)* Marine seismic acquisition (image courtesy of PGS: *www.pgs.com*).

are advantageous in terms of the quality of reconstructed data, they are also relatively complex and computationally expensive. However, matrix completion techniques based on low-rank matrix factorization are computationally efficient and relatively straightforward. The general idea of these methods is to exploit the low-rank structure of fully sampled frequency slices when organized in a matrix. Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015 and Oropeza and Sacchi 2011 discovered that fully 2D seismic data, which are acquired along seismic lines, exhibit a low-rank structure in the midpoint-offset domain. They exploited the fact that the presence of noise or missing traces increases the rank of these frequency slices. By organizing the data in the appropriate domain, low-rank factorization has been used successfully for low and mid-range frequencies. However, its performance degrades at high frequencies because monochromatic frequency slices can no longer be accurately approximated by low-rank factorization. One of the topics of this thesis is to propose an efficient weighted matrix factorization to improve data reconstruction at higher frequencies.

The sources/receivers used in 2D seismic surveys detect only vertically traveling wavefields between sources and receivers. As a result, reflections outside the 2D source-receiver plane are not captured, which degrades the image quality of the subsurface. To capture three-dimensional effects, the majority of seismic exploration surveys are now conducted in 3D, with sources and receivers distributed across an area rather than along a single line. When permuted in non-canonical form, 3D seismic data reveals a low-rank structure (Da Silva and Herrmann 2015). Although the fully sampled 3D seismic data can be reconstructed via the proposed weighted matrix completion in non-canonical form, the large-scale 3D datasets will bring computational challenges. Therefore, one of the topics of this thesis is to propose an efficient parallel weighted matrix factorization to overcome the computational challenges inherent in large-scale 3D datasets.

Wavefield reconstruction for land seismic data is degraded by promoting structure, such as sparsity or low rank, due to the contamination of ground roll, a surface wave with a strong amplitude and high spatial frequency content (Liu 1999). This deterioration can be attributed to two factors. Firstly, ground roll is a type of Rayleigh-type surface wave that is commonly aliased because it travels more slowly than body waves. Secondly, the high amplitudes of the ground roll require the reconstruction to focus on the ground roll rather than the body waves with low amplitudes. Ground roll is typically dominant at low temporal frequencies due to its spatial aliasing, but separating it from body waves is difficult due to aliasing. This thesis proposes a practical workflow for land seismic wavefield recovery using weighted matrix factorization.

The use of matrix completion as a computationally efficient method to reconstruct fully sampled wavefields from sparsely sampled seismic data has been established (Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015). Random subsampling is increasingly used to reduce acquisition time and costs. However, the design of optimal acquisition geometries is still an ongoing area of research (Manohar, Brunton, Kutz, and Brunton 2018). In matrix completion theory, the spectral gap, a measure of the connectedness of the graph in expander graph theory, has been used to predict, and to some extent quantify, the quality of wavefield reconstruction with a specific subsampling scheme (acquisition mask) (Bhojanapalli and Jain 2014; López, Kumar, Moldoveanu, and Herrmann 2022). Building on these insights, one of the topics of this thesis is to propose an optimization scheme that finds subsampling masks with large spectral gaps to improve the quality of wavefield reconstruction using matrix completion.

While sparse randomized collection of seismic data is an efficient strategy for reducing operational costs, the replication of the baseline and monitor for time-lapse seismic data gathering nullifies the productivity benefits of compressive sensing (Candès, Romberg, and Tao 2006). Collecting time-lapse seismic data is time-consuming and costly, yet it is crucial for reservoir management and monitoring of geological carbon storage (GCS). To address this issue, joint recovery models (JRM) were introduced by Oghenekohwo and Herrmann 2017 and Wason, Oghenekohwo, and Herrmann 2017, inspired by distributed compressive sensing (Baron, Duarte, Sarvotham, Wakin, and Baraniuk 2005), to leverage the benefits of low-cost randomized non-replicated acquisition for time-lapse seismic data. Rather than recovering time-lapse data individually, JRM is designed to invert baseline and monitor surveys for the common component, which contains information shared between the surveys and innovations of the baseline and monitor surveys regarding this common component (Wason, Oghenekohwo, and Herrmann 2017; Kumar, Wason, Sharan, and Herrmann 2017; Oghenekohwo and Herrmann 2017). The quality improvements of the vintages and time-lapse differences reported by Yin, Louboutin, and Herrmann 2021 and Oghenekohwo and Herrmann 2017 can be explained by the fact that the fictitious common component is observed by baseline and monitor surveys, and its recovery improves when baseline and monitor survey acquisition geometries differ (non-replicated). The use of JRM can be extended to seismic denoising (Tian, Wei, Li, Oppert, and Hennenfent 2018; Wei, Tian, Li, Oppert, and Hennenfent 2018), imaging, inversion, monitoring of carbon storage (Oghenekohwo and Herrmann 2017; Oghenekohwo 2017; Yin, Louboutin, and Herrmann 2021), and wavefield recovery (Wason, Oghenekohwo, and Herrmann 2017; Oghenekohwo, Wason, Esser, and Herrmann 2017; Kumar, Wason, Sharan, and Herrmann 2017). As part of this thesis, I investigate the use of JRM and the spectral gap to design time-lapse seismic acquisitions.

1.1 Objectives

To summarize, this thesis aims to achieve the following objectives:

1. To improve the reconstruction of 2D seismic data, particularly at high frequencies where conventional and pair matrix completion methods perform poorly, we propose recursively weighted matrix completion and establish a more computationally efficient formulation by relocating the weight matrices from constraints to the data-misfit term.

2. To design a computationally efficient weighted matrix completion for large-scale 3D seismic datasets, we propose a parallelized alternating optimization approach for parallelizing the weighted low-rank factorization algorithm to expand the application of the new

method to 3D seismic acquisitions.

3. To improve land seismic wavefield recovery with more reflection and diffraction information by using parallel weighted matrix factorization and reduce the noise introduced by ground roll, we propose a practical workflow by reconstructing the body and surface (ground roll) waves separately.

4. To obtain an improved seismic survey, we propose an optimization scheme based on simulated annealing, which finds sub-sampling masks with large spectral gaps that improve the quality of wavefield reconstruction with matrix completion.

5. To design a low-cost time-lapse seismic survey, we propose a simulation-free optimization method that combines the joint-recovery model (JRM) with large spectral gaps to generate improved subsampling surveys for each vintage.

1.2 Thesis outline

This thesis comprises a total of eight chapters, including the present introduction. In **Chapter** 2, we begin by discussing wavefield recovery via matrix completion. We then explain how to incorporate prior information from adjacent lower frequencies into our matrix completion framework on the row and column subspaces. To verify our method, we use a field data from the Gulf of Suez and demonstrate its superior performance compared to conventional matrix completion, particularly at higher frequencies. A version of this chapter was published in *SEG Technical Program Expanded Abstracts* (Zhang, Sharan, and Herrmann 2019).

In **Chapter** 3, we review the recursively weighted matrix factorization wavefield recovery. Following this, we introduce a new formulation where the weight appears in the data misfit term and discuss how to restrict the subspace of our weighted matrix factorizations. We verify our approach using field data from the Gulf of Suez, demonstrating better recovery quality compared to conventional recursively weighted matrix completion. A version of this chapter was published in *SEG Technical Program Expanded Abstracts* (Zhang, Sharan, Lopez, and Herrmann 2020).

In **Chapter** 4, we first discuss the efficient weighted wavefield reconstruction via matrix completion. We then propose approximations that allow us to decouple calculations on a row-by-row and column-by-column basis, parallelizing the alternating optimization upon which our low-rank factorization relies. The combination of weighting and decoupling results in a technique for full-azimuth wavefield reconstruction that is computationally feasible and scalable to industrial-scale problem sizes. We demonstrate the effectiveness of the proposed parallel method using a 3D synthetic dataset with varying subsampling ratios, where our method yields accurate reconstructions of broadband wavefields from severely downsampled data. My main contributions in this work are proposing the efficient weighted wavefield reconstruction and the parallel method, which I have also implemented.

In **Chapter** 5, we discuss the reconstruction of the seismic wavefield by weighted matrix factorization first. Next, we explain the impact of ground roll and introduce our practical workflow in detail. We demonstrate our approach on 3D synthetic data simulated from the Barrett model, showing improved recovery quality in comparison to the conventional workflow. A version of this chapter was published in *SEG Technical Program Expanded Abstracts* (Zhang and Herrmann 2021).

In **Chapter** 6, we present the proposed optimization problem to minimize the spectral gap ratios of subsampling masks. Second, we explain how to approximate acquisition masks using simulated annealing. We conclude by performing numerical experiments on 2D and 3D synthetic Compass datasets (E. Jones, A. Edgar, I. Selvage, and Crook 2012) and demonstrating improvements in recovery quality compared to the reconstruction of data collected via the jittered subsampling technique (Herrmann and Hennenfent 2008). A version of this chapter for the 2D case was published in *International Meeting for Applied Geoscience and Energy Expanded Abstracts* (Zhang, Louboutin, Siahkoohi, Yin, Kumar, and Herrmann 2022). A version of this chapter for the 3D case was submitted to *International Meeting for Applied Geoscience and Energy Expanded Abstracts* (Zhang, Yin, Lopez, Siahkoohi, Louboutin, and Herrmann 2023).

In **Chapter** 7, we first describe the relationship between the connectivity of graphs associated with binary sampling masks and the spectral gap ratio. Then, we explain how to increase the connectedness by decreasing the spectral gap ratios through our proposed optimization, which assists in the wavefield reconstruction process. To achieve this, we propose a new optimization objective that incorporates spectral gap ratios for the common component and baseline/monitoring surveys. This enables the improvement of time-lapse data inversion based on the joint recovery model (JRM). After a brief discussion on how to minimize this objective using simulated annealing, we evaluate the proposed method for automatically generating a binary time-lapse mask numerically using synthetic 2D data. A version of this chapter was accepted in *Geophysics* (Zhang, Yin, Lopez, Siahkoohi, Louboutin, Kumar, and Herrmann 2023).

In the final **Chapter** 8, we present the conclusions of this thesis and discuss future directions for research.

CHAPTER 2

HIGH-FREQUENCY WAVEFIELD RECOVERY WITH WEIGHTED MATRIX FACTORIZATIONS

2.1 Summary

Acquired seismic data is normally not the fully sampled data we would like to work with since traces are missing due to physical constraints or budget limitations. Rank minimization is an effective way to recovering the missing trace data. Unfortunately, this technique's performance may deteriorate at higher frequency because high-frequency data can not necessarily be captured accurately by low-rank matrix factorizations albeit remedies exist such as hierarchical semi-separable matrices. As a result, recovered data often suffers from low signal to noise ratio (SNR) at the higher frequencies. To deal with this situation, we propose a weighted recovery method that improves the performance at the high frequencies by recursively using information from matrix factorizations at neighboring lower frequencies. Essentially, we include prior information from previously reconstructed frequency slices during the wavefield reconstruction. We apply our method to data collected from the Gulf of Suez, which shows that our method performs well compared to the traditional method without weighting.

2.2 Introduction

Seismic data acquisition is one of the key steps in the initial phase of oil & gas exploration. Due to operational complexity and operational costs, acquired seismic data is usually not fully sampled, a prerequisite to subsequent steps such as multiple removal and migration all of which require densely sampled data.

Wavefield recovery is an important tool to solve the problem of poor sampling. In the last decade, wavefield recovery methods based on sparsity promotion in different transform domains, such as the Radon (Bardan 1987), wavelet (Villasenor, Ergas, and Donoho 1996), and curvelet (Herrmann, Wang, Hennenfent, and Moghaddam 2007; Herrmann and Hennenfent 2008) domain have been developed. Although these methods are valuable in terms of the quality of recovered data, they are relatively complex and computationally expensive. Fortunately, matrix completion methods (Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015) based on low-rank matrix factorizations are relatively simple and computationally cheaper. The latter use the property that fully-sampled frequency slices permit accurate low-rank representations when organized in midpoint-offset. In Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015 and Oropeza and Sacchi 2011, authors exploit the fact that presence of noise or missing traces increases the rank of these frequency slices. We use this property to recover frequency slices via factored rank minimization (Kumar, Aravkin, Esser, Mansour, and Herrmann 2014). While this matrix factorization method performs well at the low to mid frequencies, it struggles to recover high-frequency data, which need higher ranks to be accurately represented.

Recent work by Aravkin, Kumar, Mansour, Recht, and Herrmann 2014 and Eftekhari, Yang, and Wakin 2018 has shown that reliable prior information on the row and column subspaces of the underlying low rank matrix can be used to further improve wavefield recovery via matrix completion. For seismic data, we have access to this information when there is a strong similarity between adjacent frequency slices. In that case, the row and column subspaces can serve as prior information. This principle was first demonstrated by Aravkin, Kumar, Mansour, Recht, and Herrmann 2014 and we extend this line of research by recursively invoking prior as we work our way from the relatively low frequencies to the high frequencies where conventional matrix completion methods typically perform poorly.

This chapter is organized as follows. First, we discuss wavefield recovery via matrix completion. Next, we discuss how to incorporate prior information on the row and column subspaces from neighboring lower frequencies in our matrix completion framework. We conclude demonstrating our approach on a field data example from the Gulf of Suez and show its better performance compared to conventional matrix completion especially at the higher frequencies.

2.3 Methodology

2.3.1 Low-rank matrix factorization

In Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015 and Aravkin, Kumar, Mansour, Recht, and Herrmann 2014, authors exploit low rank of fully sampled seismic data by solving for each frequency problem of the type

$$\min_{\mathbf{X}_i} \quad \|\mathbf{X}_i\|_* \quad \text{subject to} \quad \|\mathcal{A}(\mathbf{X}_i) - \mathbf{b}_i\|_2 \leq \epsilon. \tag{2.1}$$

In this expression, the matrices $\mathbf{X}_i$ for $i = 1 \cdots n_f$ with n_f the number of angular frequencies represent fully sampled monochromatic frequency slices in the midpoint-offset domain, $\mathcal{A}$ is the sampling operator collecting the data into a vector, and $\mathbf{b}_i$ represents the observed data at the i^{th} frequency. For each frequency, we recover the fully sampled data by minimizing the nuclear norm $\|\cdot\|_*$ on each $\mathbf{X}_i$ subject we fit the data within ϵ. The nuclear norm itself is defined as the sum of the singular values. We solve Equation 2.1 for all the frequencies to obtain our recovered data $\mathbf{X} \in \mathbb{C}^{n_f \times n_m \times n_h}$, where n_m is the number of midpoints and n_h the number of offsets. As reported by Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015, randomized sampling increases the rank of 2D seismic data in midpoint offset domain, which is a favorable condition for matrix completion.

To avoid computationally expensive singular-value decompositions (SVD) while solving Equation 2.1, we employ a low-rank matrix factorization approach. For this purpose, we factor the matrices (for notational simplicity we drop the subscript $_i$) $\mathbf{X} \in \mathbb{C}^{n_m \times n_h}$ in Equation 2.1 into the low-rank factors $\mathbf{L} \in \mathbb{C}^{n_m \times r}$ and $\mathbf{R} \in \mathbb{C}^{n_h \times r}$ both of rank r. To avoid expensive SVDs, we follow Rennie and Srebro 2005 and replace the nuclear norm in Equation 2.1 by

$$\min_{\mathbf{L},\mathbf{R}} \quad \frac{1}{2}\left\| \begin{bmatrix} \mathbf{L} \\ \mathbf{R} \end{bmatrix} \right\|_F^2 \quad \text{subject to} \quad \|\mathcal{A}(\mathbf{L}\mathbf{R}^H) - \mathbf{b}\|_F \leq \epsilon, \tag{2.2}$$

where ^{H} is the Hermitian transpose and $\|\cdot\|_F$ the Frobenius norm (2-norm of the vectorized matrix). Following Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015 and Aravkin, Kumar, Mansour, Recht, and Herrmann 2014, we solve Equation 2.2 with spectral-projected gradients (Berg and Friedlander 2009).

As we mentioned earlier, the performance of low-rank factorization methods degrade with increasing frequency reflected in increasing poor signal to noise ratios (SNRs) (Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015). To improve the recovered data quality at higher frequencies, we include weighted matrix completion (Aravkin, Kumar, Mansour, Recht, and Herrmann 2014; Eftekhari, Yang, and Wakin 2018).

2.3.2 Weighted low-rank matrix factorization

The key of our methodology is that we approximate fully sampled data in low-rank factored form. When using SVDs, this factored form reads

$$\mathbf{X} \approx \mathbf{U}\Sigma\mathbf{V}^H, \tag{2.3}$$

where $\mathbf{U} \in \mathbb{C}^{n_m \times r}$ and $\mathbf{V} \in \mathbb{C}^{n_h \times r}$ are column and row subspaces of $\mathbf{X}$, respectively. $\Sigma \in \mathbb{C}^{r \times r}$ is a diagonal matrix containing the largest r singular values of the matrix $\mathbf{X}$.

As shown by Aravkin, Kumar, Mansour, Recht, and Herrmann 2014 and Eftekhari, Yang, and Wakin 2018, information on these subspaces can be used to rewrite Equation 2.1

into its weighted form—i.e., we have

$$\min_{\mathbf{X}} \quad \|\mathbf{QXW}\|_* \quad \text{subject to} \quad \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2 \leq \epsilon, \tag{2.4}$$

where

$$\mathbf{Q} = w_1 \mathbf{U}\mathbf{U}^H + \mathbf{U}^{\perp}\mathbf{U}^{\perp H} \tag{2.5}$$

and

$$\mathbf{W} = w_2 \mathbf{V}\mathbf{V}^H + \mathbf{V}^{\perp}\mathbf{V}^{\perp H} \tag{2.6}$$

are projection matrices on subspaces spanned by $\mathbf{U}$, $\mathbf{V}$ and their orthogonal complements $\mathbf{U}^{\perp}$, $\mathbf{V}^{\perp}$. The scalars $w_1 \in (0, 1]$ and $w_2 \in (0, 1]$ are weights that depend on the confidence we have in the priors—i.e., how close the matrix $\mathbf{X}$ is to the actual $\mathbf{X}_i$ we are dealing with at frequency i. Small values for the weights w_1 and w_2 mean that we have confidence in the prior (the matrix $\mathbf{X}$ is close). When $w_1 \uparrow 1$ and $w_2 \uparrow 1$, solving Equation 2.4 becomes equivalent to solving the original Equation 2.1.

As before, we can rewrite Equation 2.4 into a weighted low-rank factored form (Aravkin, Kumar, Mansour, Recht, and Herrmann 2014):

$$\min_{\mathbf{L},\mathbf{R}} \quad \frac{1}{2}\left\| \begin{bmatrix} \mathbf{QL} \\ \mathbf{WR} \end{bmatrix} \right\|_F^2 \quad \text{subject to} \quad \|\mathcal{A}(\mathbf{L}\mathbf{R}^H) - \mathbf{b}\|_2 \leq \epsilon. \tag{2.7}$$

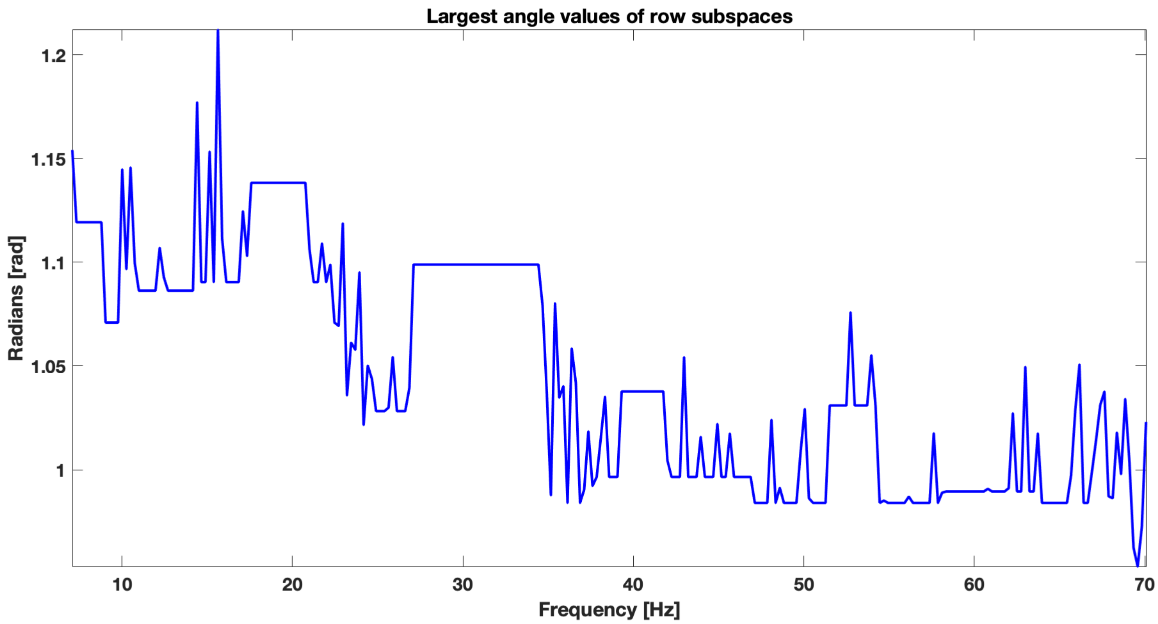
The question now is which subspaces to use for the columns and rows. Because frequency slices have information in common from frequency to frequency, we follow Aravkin, Kumar, Mansour, Recht, and Herrmann 2014, Eftekhari, Yang, and Wakin 2018 and use the $\mathbf{U}$ and $\mathbf{V}$ from the previous lower frequency.

While Aravkin, Kumar, Mansour, Recht, and Herrmann 2014 somewhat successfully applied this approach for a single frequency slice, these authors never justified this approach and neither did they apply the weighting recursively working from the low to the higher frequencies. For this purpose, we quantify similarities (in the form of angles, see Eftekhari, Yang, and Wakin 2018) between subsequent frequency slices for the Gulf of Suez data. This will allow us to predict the performance of our method.

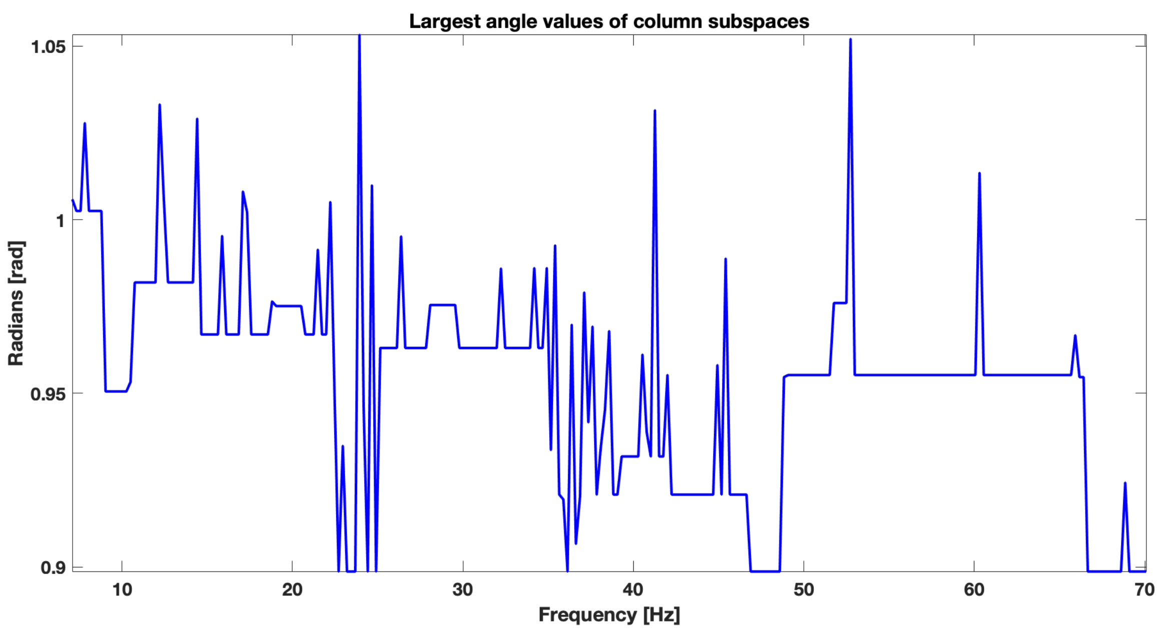
2.3.3 Quantifying similarity

Similarity between subsequent frequency slices depends on the largest principal angles (Eftekhari, Yang, and Wakin 2018) between column subspaces and between row subspaces of subsequent frequency slices. Smaller angles correspond to more similarity between subspaces of subsequent frequency slices and vice versa. Therefore, we can choose smaller weights w_1 and w_2 when the angles are smaller. Smaller weights correspond to larger penalties (Eftekhari, Yang, and Wakin 2018) on matrices that have subspaces orthogonal to $\mathbf{U}$ and $\mathbf{V}$ in Equation 2.4. When weights are small, we have more confidence in $\mathbf{U}$ and $\mathbf{V}$ and less confidence in their orthogonal counterparts. In Figure 2.1, we show angles between column subspaces (Figure 2.1b) and row subspaces (Figure 2.1a) for subsequent frequency slices of the Gulf of Suez data. We observe an overall decreasing trend in angles with increasing frequencies for both row and column subspaces. This trend indicates increasing similarity between subsequent frequency slices with increasing frequency. This trend is consistent with high-frequency approximate behavior of wavefields—i.e., as the frequency increases solutions become more and more like the high-frequency solution, and this gives us a handle how to choose the weights as the frequency increases.

Figure 2.1 shows that as the frequency increase, the largest angles between the subspaces of neighboring frequencies decreases. Smaller the angle, more similar are subspaces. This angle test have demonstrated that the weighted matrix completion will perform better in high frequency band because of smaller angle in comparison to its lower frequency counterpart.



(a)



(b)

Figure 2.1: Largest angle between *(a)* row and *(b)* column subspaces for subsequent frequency slices of Gulf of Suez data

2.4 Numerical experiments

We demonstrate effectiveness of recursive weighted matrix completion method over the conventional matrix completion method and the weighted method just using previous frequency slice (named as single pair weighted method) in terms of data reconstruction quality. In single pair weighted method, we reconstruct previous frequency slice using conventional matrix completion. We use real 2D seismic data with number of sources, $N_s = 354$, and number of receivers, $N_r = 354$ acquired in Gulf of Suez for this comparison. Total number of time samples for this data set is $N_t = 1024$ with sampling interval of $0.004\,\mathrm{s}$. Most of the energy of the seismic line is concentrated in $20\,\mathrm{Hz}$ to $70\,\mathrm{Hz}$ frequency band (Figure 2.3b). To get the subsampled data (Figure 2.2b & Figure 2.4b), we remove 75% of sources using a jittered subsampling mask. Jittered subsampling method not only breaks the inherent properties such as low rank of fully sampled seismic data but also controls the maximum gap size of the incomplete data (Herrmann and Hennenfent 2008). We show the results and comparisons in both frequency domain and time domain. For every frequency slice, we perform 150 iterations of spectral projected gradient algorithm for all these methods. Figure 2.2 shows results on a frequency slice at $30\,\mathrm{Hz}$. Recovery with the unweighted method gives the result with SNR of $11.43\,\mathrm{dB}$. Whereas recovery with the single pair weighted method gives a higher SNR of $15.20\,\mathrm{dB}$, our recursive weighted method gives the highest SNR of $18.48\,\mathrm{dB}$, $7.05\,\mathrm{dB}$ improvement in SNR over the unweighted method. Figure 2.2d, Figure 2.2f and Figure 2.2h show differences between these three different methods and ground truth data. The reconstruction using recursive prior knowledge gives the least residual in Figure 2.2h among these three methods. The residual of reconstruction without using any prior knowledge is Figure 2.2d and the residual of single pair weighted reconstruction is Figure 2.2f.

Figure 2.3a shows recursive weighted method's performance (red color plot) over range of frequencies in terms of signal to noise of completion. Recursive weighted recovery clearly outperforms the conventional recovery without weight (black color plot in Fig-

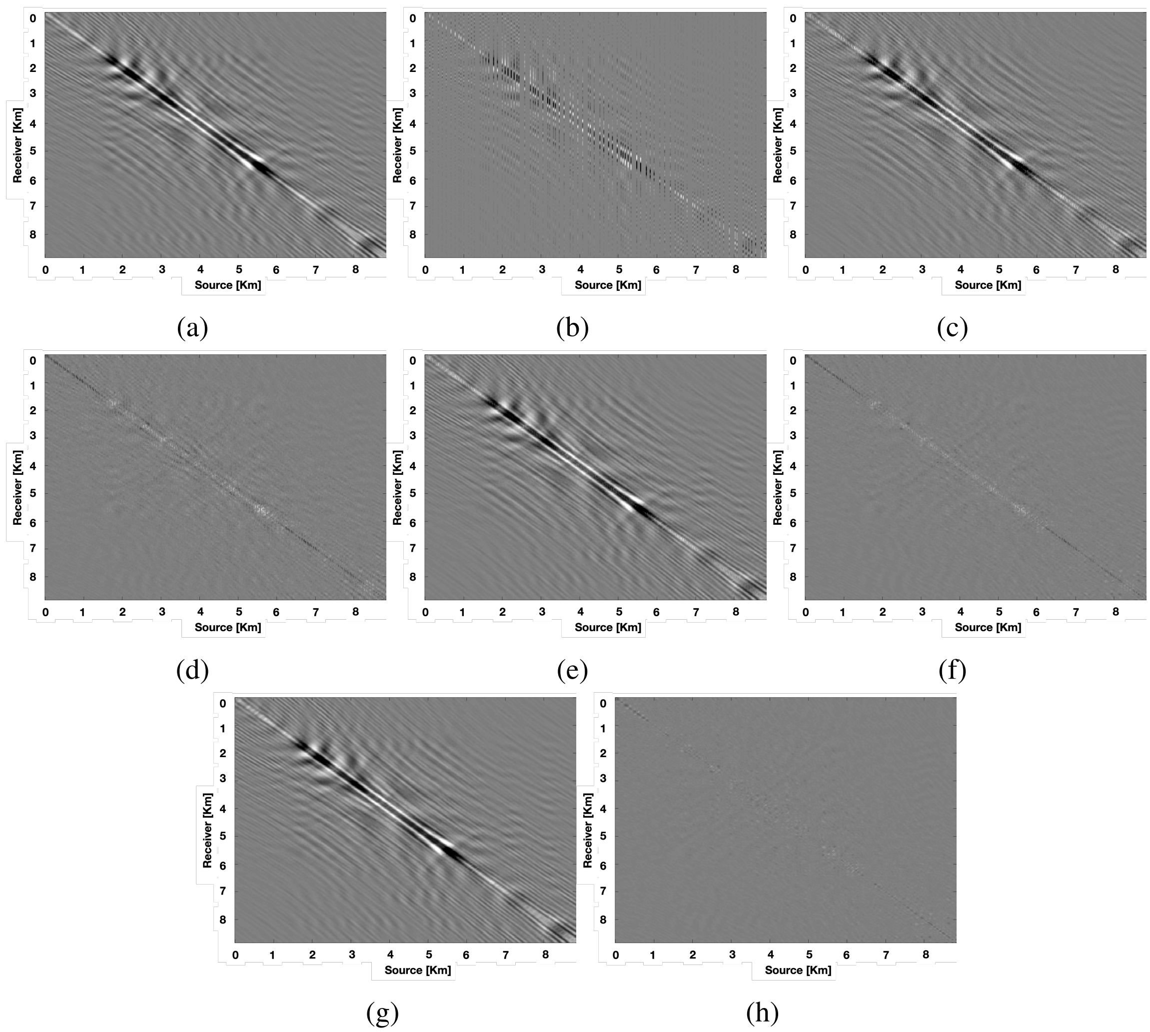


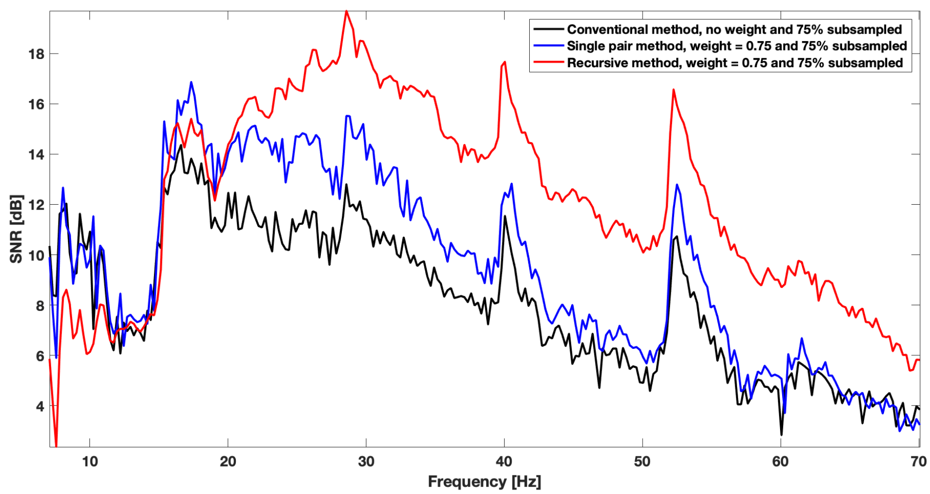
Figure 2.2: Missing trace recovery for a frequency slice at $30\,\mathrm{Hz}$ in source-receiver domain. *(a)* Ground truth, *(b)* 75% subsampled seismic data with jittered subsampling. *(c)* and *(d)* represent recovery ($SNR = 11.43\,\mathrm{dB}$) using conventional method and its difference w.r.t. the ground truth respectively. *(e)* and *(f)* represent recovery ($SNR = 15.20\,\mathrm{dB}$) using single pair weighted method and its difference w.r.t. the ground truth respectively. *(g)* and *(h)* represent recovery ($SNR = 18.48\,\mathrm{dB}$) using recursive weighted method and its difference w.r.t. the ground truth respectively.

ure 2.3a) and the single pair weighted recovery (blue color plot in Figure 2.3a) in the frequency range which contains most of the energy (Figure 2.3b).

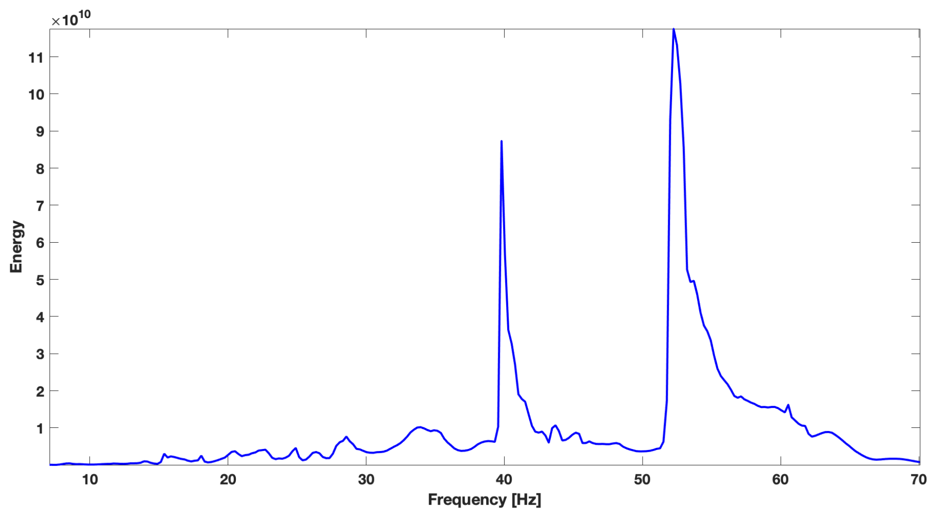
To further compare these three methods for time domain data, we apply them on all frequency slices and transfer results back to time domain. Figure 2.4 shows results and differences for one common receiver gather extracted from complete time domain data. With the conventional recovery we get SNR of $6.87\,\mathrm{dB}$, whereas with single pair weighted recovery we get SNR of $7.85\,\mathrm{dB}$ and with recursive weighted recovery we get SNR of $11.63\,\mathrm{dB}$. With recursive weighted recovery we get almost $5\,\mathrm{dB}$ improvement for complete time domain data over conventional recovery. It is also obvious to see the advantage of the recursive weighted method from three differences in Figure 2.4d, Figure 2.4f and Figure 2.4h. The residual is significantly reduced in Figure 2.4h in contrast to Figure 2.4d and Figure 2.4f.

2.5 Conclusion

In this work, we have proposed recursive weighted matrix completion to improve data reconstruction quality, especially at high frequencies where the conventional matrix completion method performs poorly. In contrast to conventional low-rank matrix factorization without weighting or with non-recurrent pairwise weighting, our recursively weighted method performs better at the high frequencies, especially at frequencies where the data has the most energy. We also demonstrated the effectiveness of our recursive recovery on real data. Future work will be to extend the application of recursive weighted matrix completion to realistic size 3D seismic data reconstruction and also to simultaneous source acquisition.



(a)



(b)

Figure 2.3: *(a)* SNR vs frequency of recovery using recursive weighted method (Red color), single pair weighted method (Blue color) and conventional method (Black color). *(b)* Plot of energy of frequency slices vs frequency.

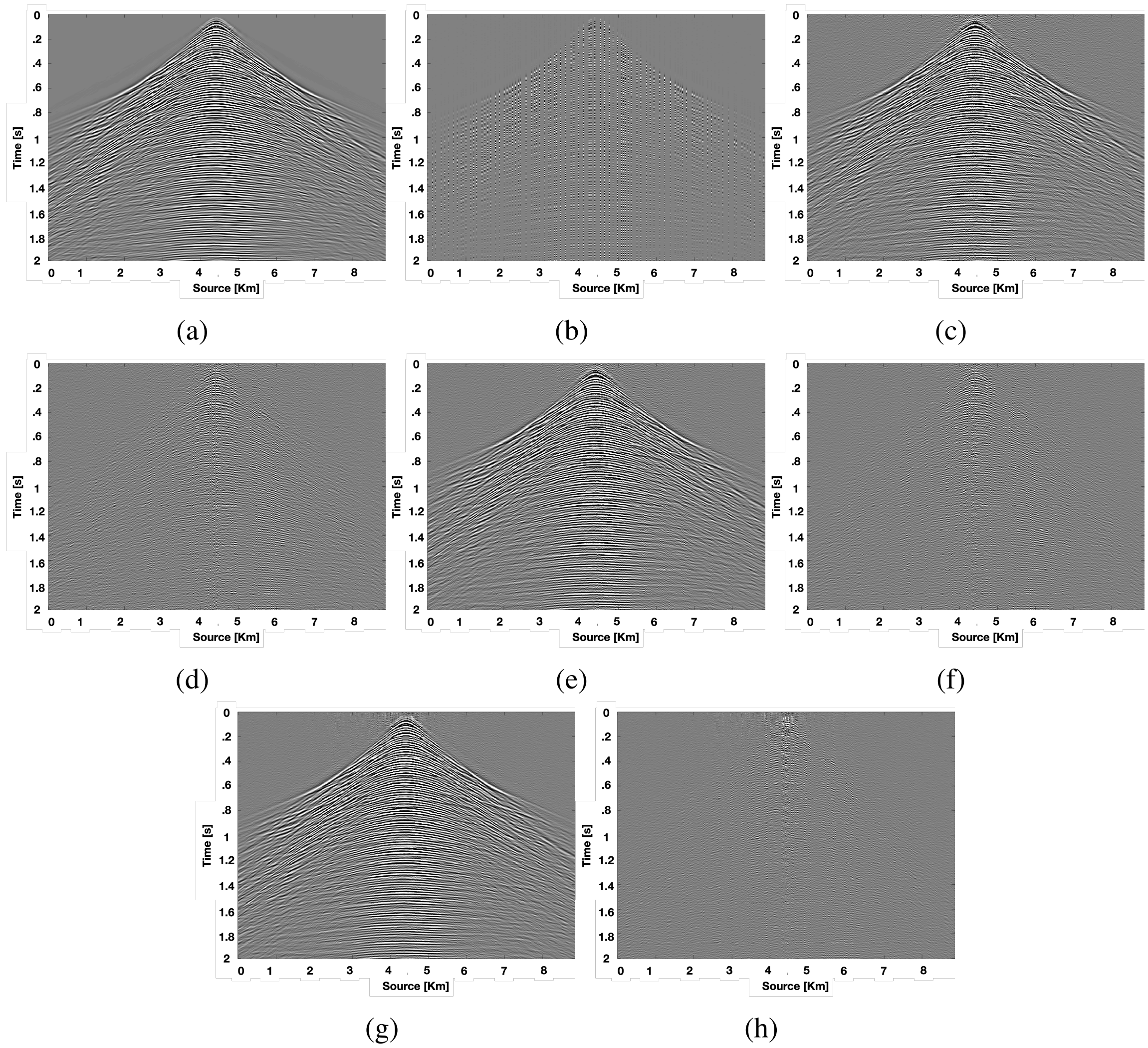


Figure 2.4: Missing trace recovery of time domain data. *(a)* Ground truth. (b) 75% subsampled seismic data with jittered subsampling. *(c)* and *(d)* represent recovery ($SNR = 6.87\,\mathrm{dB}$) using conventional method and its difference w.r.t. the ground truth respectively. *(e)* and *(f)* represent recovery ($SNR = 7.85\,\mathrm{dB}$) using single pair weighted method and its difference w.r.t. the ground truth respectively. *(g)* and *(h)* represent recovery ($SNR = 11.63\,\mathrm{dB}$) using recursive weighted method and its difference w.r.t. the ground truth respectively.

CHAPTER 3
WAVEFIELD RECOVERY WITH LIMITED-SUBSPACE WEIGHTED MATRIX FACTORIZATIONS

3.1 Summary

Modern-day seismic imaging and monitoring technology increasingly rely on dense full-azimuth sampling. Unfortunately, the costs of acquiring densely sampled data rapidly become prohibitive and we need to look for ways to sparsely collect data, e.g. from sparsely distributed ocean bottom nodes, from which we then derive densely sampled surveys through the method of wavefield reconstruction. Because of their relatively cheap and simple calculations, wavefield reconstruction via matrix factorizations has proven to be a viable and scalable alternative to the more generally used transform-based methods. While this method is capable of processing all full azimuth data frequency by frequency slice, its performance degrades at higher frequencies because monochromatic data at these frequencies is not as well approximated by low-rank factorizations. We address this problem by proposing a recursive recovery technique, which involves weighted matrix factorizations where recovered wavefields at the lower frequencies serve as prior information for the recovery of the higher frequencies. To limit the adverse effects of potential overfitting, we propose a limited-subspace recursively weighted matrix factorization approach where the size of the row and column subspaces to construct the weight matrices is constrained. We apply our method to data collected from the Gulf of Suez, and our results show that our limited-subspace weighted recovery method significantly improves the recovery quality.

3.2 Introduction

Seismic data acquisition plays a key role in the initial phase of oil & gas exploration. It also represents a significant budget item for monitoring of carbon sequestration. For these reasons, it is a challenge to come up with new acquisition methodologies that improve acquisition productivity (Mosher, Li, Morley, Ji, Janiszewski, Olson, and Brewer 2014) without sacrificing data quality. Randomized acquisition according to the principles of compressive sensing (Herrmann, Friedlander, and Yilmaz 2012) in combination with large-scale wavefield reconstruction algorithms (Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015) has proven a viable tool to improve the acquisition productivity both in marine and land seismic settings.

So far, many of the employed approached of wavefield reconstruction are based transform-domain sparsity, which is deigned to explore local smoothness typically in small windows in up to five dimensions. While these approaches have been applied successfully on production data, they do not exploit redundancies present in the data over long distances. Recovery techniques based on low-rank matrix factorizations (Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015) do not suffer from this shortcoming because this method works with monochromatic frequency slices that contain data from the complete survey instead of working within small windows limiting the apperture. By organizing the data in the appropriate domain, e.g. midpoint-offset domain for seismic lines, monochromatic frequency slices permit approximations in low-rank form, which can be used to recover fully sample wavefields from subsampled data.

While low-rank factorizations have been employed successfully for low and midrange frequencies, their performance deteriorates at high frequencies because monochromatic frequency slices can no longer be approximated accurately by low-rank factorizations. In this work, we overcome this problem by using the fact that factorizations at neighboring frequencies live in close-by subspaces. As described in early work by Aravkin, Kumar, Mansour, and Recht 2013, Eftekhari, Yang, and Wakin 2018, this property can be exploited by introducing matrix weights defined in terms of factorizations of near-by frequency slices. Recent work by Zhang, Sharan, and Herrmann 2019 took this initial a step further by proposing a recursive approach where factorizations of frequency slices at lower frequencies are used as weight for factorizations at the higher frequencies starting at the low frequencies and working its way up.

While this approach has had some success (see e.g. Zhang, Sharan, and Herrmann 2019), there is challenge related to the fact that high frequencies require higher rank factorizations and this can lead to overfitting when using this higher rank throughout. We avoid this overfitting, by adapting the rank of the weighting matrices such that overfitting is avoided. We do this by actively limiting the row and column subspaces of the weight matrices. Because we avoid overfitting, we are able to further improve the wavefield recovery. We also introduce an alternative formulation where the weight matrices are moved from the constraint, as in Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015, to the data misfit objective, which leads to a significant improvement (20 to 25 times speedup) computational efficiency.

We organize this chapter as follows. First, we review the recursively weighted wavefield recovery via matrix factorization including the new formulation where the weight appear in the data misfit term. Next, we discuss how to limit the subspace of our weighted matrix factorizations. We conclude by demonstrating our approach on a field data example from the Gulf of Suez, which shows improved recovery quality compared to conventional recursively weighted matrix completion.

3.3 Methodology

We start by introducing wavefield reconstruction via weighted matrix factorization. To improve computational efficiency, we move the weight matrices to the data misfit term so we no longer have to carry out numerically expensive weighted projections as in Aravkin, Kumar, Mansour, and Recht 2013. Aside from allowing for a much more computationally efficient implementation, this alternative formulation also forms the basis for our limited-subspace approach designed to prevent overfitting at the low frequencies.

3.3.1 Weighted low-rank matrix factorization

Our proposed extension to wavefield reconstruction via recursively weighted matrix factorization derives from earlier work by Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015, Aravkin, Kumar, Mansour, and Recht 2013, and Zhang, Sharan, and Herrmann 2019, where we solve

$$
\begin{aligned}
\min_{\mathbf{X}_i} \quad & \|\mathbf{Q}\mathbf{X}_i\mathbf{W}\|_* \\
\text{subject to} \quad & \|\mathcal{A}(\mathbf{X}_i) - \mathbf{b}_i\|_2 \leq \tau
\end{aligned}
\tag{3.1}
$$

to within a noise-level dependent data misfit tolerance τ. In this expression, the matrix $\mathbf{X}_i$ corresponds to a monochromatic frequency slice in the midpoint-offset domain (in case of 2D) at the $i\mathrm{th}$ frequency ($i \in [1, \cdots, n_f]$ with n_f the number of frequencies).

During the wavefield recovery, fully sampled frequency slices are represented by the complex valued matrix, $\mathbf{X} \in \mathbb{C}^{n_f \times n_m \times n_h}$ where n_m is the number of midpoints and n_h the number of offsets. The symbol $\mathcal{A}(\cdot)$ stands for the subsampling operator, which collects monochromatic data at the observed source-receiver combinations into the vector $\mathbf{b}_i$. Given these observations, we solve for the fully sampled $\mathbf{X}_i$ for each frequency by minimizing Equation 3.1 with weight matrices $\mathbf{Q}$ and $\mathbf{W}$ given by

$$
\mathbf{Q} = w_1\mathbf{U}\mathbf{U}^H + \mathbf{U}^\perp\mathbf{U}^{\perp H}
\tag{3.2}
$$

and

$$
\mathbf{W} = w_2\mathbf{V}\mathbf{V}^H + \mathbf{V}^\perp\mathbf{V}^{\perp H}.
\tag{3.3}
$$

In these expressions for the weight matrices, the $\mathbf{U} \in \mathbb{C}^{n_m \times r}$ and $\mathbf{V} \in \mathbb{C}^{n_h \times r}$ are the column and row subspaces that derive from the low-rank factorization of the nearby frequency slice. $\mathbf{U}$ and $\mathbf{V}$ have orthonormal columns that span top column and row subspaces of nearby frequency slice. Because these weight matrices include information on the subspaces of the current factorization, they serves as prior information aiding the wavefield recovery via the weighted nuclear norm minimization (denoted by $\|\mathbf{QXW}\|_* = \sum_{j=1}^{r} \sigma_j$ with σ_j the j^{th} singular value). Depending on whether we have confidence in the fact that the neighboring frequency slice has an overlapping subspace, we chose the weights w_1 and w_2 close to 0 if we have confidence and close to 1 if we do not.

While the above weighted formulation has resulted in major improvements in the recovery when reliable information on a neighboring frequency slice is available (Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015; Aravkin, Kumar, Mansour, and Recht 2013; Zhang, Sharan, and Herrmann 2019), the minimization in Equation 3.1 is complicated by the presence of the weighting matrices in the nuclear norm objective. As a result, the minimization becomes computationally expensive. To avoid this complication, we replace the optimization variable by $\bar{\mathbf{X}}_i = \mathbf{Q}\mathbf{X}_i\mathbf{W}$, and rewrite Equation 3.1 as

$$\begin{aligned} \min_{\bar{\mathbf{X}}_i} \quad & \|\bar{\mathbf{X}}_i\|_* \\ \text{subject to} \quad & \|\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{X}}_i\mathbf{W}^{-1}) - \mathbf{b}_i\|_2 \leq \tau \end{aligned} \tag{3.4}$$

where the modified weighting matrices

$$\mathbf{Q}^{-1} = \frac{1}{w_1}\mathbf{U}\mathbf{U}^H + \mathbf{U}^{\perp}\mathbf{U}^{\perp H} \tag{3.5}$$

and

$$\mathbf{W}^{-1} = \frac{1}{w_2}\mathbf{V}\mathbf{V}^H + \mathbf{V}^{\perp}\mathbf{V}^{\perp H} \tag{3.6}$$

are moved from the objective to the data misfit constraint. To reflect that we changed the problem, we introduced barred quantities from which the solution original solution can be readily computed—i.e., we recover the solution $\mathbf{X}_i = \mathbf{Q}^{-1}\bar{\mathbf{X}}_i\mathbf{W}^{-1}$ since $\bar{\mathbf{X}}_i = \mathbf{Q}\mathbf{X}_i\mathbf{W}$ solves the above optimization problem. Compared to Equation 3.1, this new formulation does not require nuclear norm projections onto weighted matrices while its solution is equivalent to Equation 3.1.

Like the original formulation, our new formulation lends also itself to be cast into a low-rank ($r \ll \max(n_m, n_h)$) factorized form so that expensive SVDs are avoided in the nuclear norm. After factorization our wavefield reconstruction involves

$$\begin{aligned} \min_{\bar{\mathbf{L}}_i, \bar{\mathbf{R}}_i} \quad & \frac{1}{2}\left\| \begin{bmatrix} \bar{\mathbf{L}}_i \\ \bar{\mathbf{R}}_i \end{bmatrix} \right\|_F^2 \\ \text{subject to} \quad & \|\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{L}}_i\bar{\mathbf{R}}_i^H\mathbf{W}^{-1}) - \mathbf{b}_i\|_2 \leq \epsilon, \end{aligned} \tag{3.7}$$

where the symbol ^{H} denotes the Hermitian transpose and $\|\cdot\|_F$ is the Frobenius norm (2-norm of the vectorized matrix) (Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015; Aravkin, Kumar, Mansour, and Recht 2013; Zhang, Sharan, and Herrmann 2019). Compared to the original representation for frequency slices, the above factored form is compressed since it entails the low-rank pair $\{\bar{\mathbf{L}}_i, \bar{\mathbf{R}}_i\}$,where $\bar{\mathbf{X}}_i = \bar{\mathbf{L}}_i\bar{\mathbf{R}}_i^H$, and does not rely on storage and manipulation of the original and dense optimization variable $\mathbf{X}_i$ or $\bar{\mathbf{X}}_i$. Despite gains in computation, because of the factored form and redefined data misfit term, challenges remain with recursive weighted matrix factorizations (Zhang, Sharan, and Herrmann 2019) at the high frequencies and as we will show these have to do with overfitting.

3.3.2 Limited subspace weighted implementation

To reduce approximation errors at the high frequencies, we can increase the rank of the factorization throughout. While increasing the rank leads to better approximations at the high frequencies adapting this higher rank at the lower frequencies can lead to overfitting. The resulting poor reconstructions at the lower frequencies can in turn have a detrimental effect on the reconstruction at higher frequencies, which information from the lower frequencies as the recursive algorithm sweeps from the low to the high frequencies.

By choosing the rank for the limited subspace, we reduce the size of the subspaces of the weight matrices to prevent overfitting at the lower frequencies. In Equation 3.2, Equation 3.3, Equation 3.5 and Equation 3.6, we notice that the size of the weight matrices $\mathbf{Q}$ and $\mathbf{W}$ are independent of rank r. Therefore, we can use a limited subspace to remove the influence of overfitting and get better results.

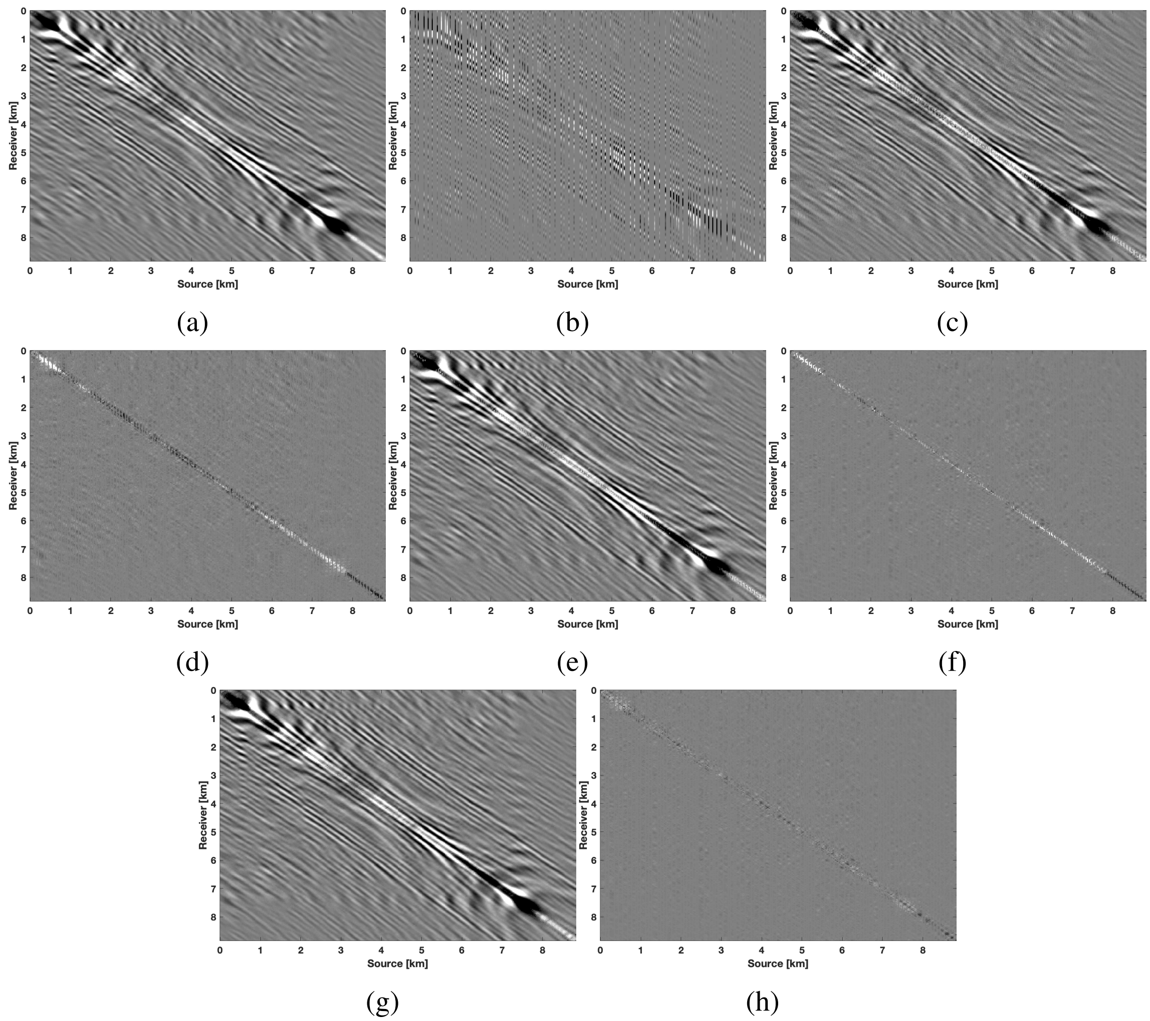
By limited subspace, we mean that at a given frequency slice, instead of using a rank r for row and column subspaces $\mathbf{U}$ and $\mathbf{V}$ respectively, we can use a lower rank r_s. In this way, we can choose higher rank r to reconstruct each frequency but use lower rank r_s to construct the weight matrices ($\mathbf{Q}$ and $\mathbf{W}$). By choosing smaller rank for the subspaces, we mitigate the negative influence of overfitting. Therefore, in the limited-subspace method, we are free to choose smaller values for the r_s for each frequency slice and higher values for the rank r for the factorization itself (not for the weights) for each frequency.

3.4 Numerical Experiments

To demonstrate the advocacy of the proposed method, we use 2D field seismic data acquired in the Gulf of Suez with number of sources, $N_s = 355$, and number of receivers, $N_r = 355$. The total number of time samples in this dataset is $N_t = 1024$ and the sampling interval is $0.004\,\mathrm{s}$. We use a jittered subsampling (Herrmann and Hennenfent 2008) mask to remove 75% of the sources to obtain the subsampled data. When data is organized in the midpoint-offset domain, we know that randomized jittered subsampling method breaks the inherent low-rank property of seismic data while controlling the largest gap size of the subsampled data (Herrmann and Hennenfent 2008). Controlling largest gap is important because very large gaps are not suitable for wavefield reconstruction using sparsity-promotion or low-rank matrix completion. We use the weighted method as described by Zhang, Sharan, and Herrmann 2019 to reconstruct frequency slices starting at $10\,\mathrm{Hz}$ and working our way up to $70\,\mathrm{Hz}$. We use constant rank across all the frequencies for weight matrices and matrix factorization. We base these choices for $r_s < r$ on visual inspection of the recovered frequency slices. To avoid overfitting at lower frequencies we select rank r_s of the limited subspace constant across all the frequencies. And to better approximation of higher frequencies we choose higher rank r across all the frequencies. Combination of higher rank for matrix factorization and smaller rank for limited subspace avoid the risk of overfitting and at the same time improves the data reconstruction quality.

To demonstrate that the limited-subspace recursively weighted method gives improved results compared to conventional recursively weighted method (Zhang, Sharan, and Herrmann 2019), we first show results in the frequency domain. For each frequency slice, we perform 150 iterations for both the methods. For the limited-subspace weighted method, we use rank $r = 85$ and limited subspace rank of $r_s = 25$. For comparison with the conventional weighted method, we perform two experiments with a fixed high rank of $r = 85$ and lower rank of $r = 25$. We choose lower rank for conventional weighted method to show that smaller rank itself is not sufficient for significant improvement in data reconstruction at higher frequencies. On the other hand we choose higher rank of 85 for conventional weighted method to show that higher rank is alone not sufficient to improve the quality of reconstructed data at higher frequencies because of the overfitting at lower frequencies. We show reconstruction results for a frequency slice at $22\,\mathrm{Hz}$ in Figure 3.1. Due to overfiting, the conventional method with rank $r = 85$ gives a reconstruction with a smaller SNR of $13.09\,\mathrm{dB}$ compared to the wavefield reconstruction (Figure 3.1c and Figure 3.1d) obtained with the smaller rank $r = 25$ for which we get SNR of $15.50\,\mathrm{dB}$ (Figure 3.1e and Figure 3.1f). We get SNR of $19.52\,\mathrm{dB}$ for the reconstructed data (Figure 3.1g) using the limited-subspace weighted method. Figure 3.1h shows the data residual with respect to the ground truth (Figure 3.1a). Clearly, our limited-subspace weighted method outperforms the conventional weighted method in terms of improved quality of reconstructed data.

To further compare our limited-subspace method with the original method, we repeat



(a) (b) (c)

(d) (e) (f)

(g) (h)

Figure 3.1: Reconstruction for missing source for a frequency slice at $22\,\mathrm{Hz}$ shown in the source-receiver domain but reconstructed in the midpoint-offset domain. *(a)* Ground truth, *(b)* 75% subsampled seismic data with jittered subsampling. *(c)* and *(d)* recovery by weighted matrix factorization ($SNR = 13.09\,\mathrm{dB}$) using conventional recursively weighted approach with fixed rank $r = 85$ and corresponding residual w.r.t. the ground truth, respectively. *(e)* and *(f)* contain recovery ($SNR = 15.50\,\mathrm{dB}$) for conventional recursively weighted with a rank $r = 25$ and corresponding residual w.r.t. the ground truth respectively. *(g)* and *(h)* represent recovery ($SNR = 19.52\,\mathrm{dB}$) using limited-subspace weighted method with limited-subspace rank $r_s = 25$ and corresponding residual w.r.t. the ground truth respectively.

wavefield reconstructions over a range of frequencies $7-74\,\mathrm{Hz}$. In Figure 3.2, we show the comparison of the SNR's across the whole frequency range. As expected, we observe that limited-subspace weighted method (red line in Figure 3.2) outperforms conventional weighted method for both ranks of 25 (blue line in Figure 3.2) and 85 (black line in Figure 3.2) for most of the frequencies. This is because of using limited subspace we avoid risk of overfitting at lower frequencies and hence get improvement in quality of reconstructed data.

To show the recovery improvement in the time domain, we included Figure 3.3. To make fair comparison, we construct a bandpass filter with pass frequency $7-74\,\mathrm{Hz}$ with a transition width at both ends of $3.66\,\mathrm{Hz}$. We apply this bandpass filter on the true data, the subsampled data, and on recovered data recovered using the three scenarios described above. After applying the filter, we transform the filtered data back to the time domain. As we can see from Figure 3.3e, we observe less leakage of coherent signal in the data residual for results obtained with our limited-subspace weighted method in comparison to the data residual yielded by the conventional weighted method with ranks of $r=85$ (Figure 3.3c) and $r=25$ (Figure 3.3d). With the conventional weighted method for rank equals to $r=85$, we get SNR of $10.69\,\mathrm{dB}$, and for rank $r=25$, we get SNR of $11.49\,\mathrm{dB}$. With the limited-subspace weighted method we get SNR of $13.31\,\mathrm{dB}$, which is a significant improvement.

3.5 Conclusions

In this chapter, we proposed a limited-subspace weighted method to further improve the performance of recursively weighted method in terms of better data reconstruction quality. By exploiting the fact that dimensions of weight matrices are independent of the rank of the subspaces, our method allows us to use higher ranks for data reconstruction while avoiding the risk of overfitting at the lower frequencies. Matrices with higher rank allow for a better approximation of the frequency slices at higher frequencies and hence allow for better quality of reconstructed data if we prevent overfitting by working with limited-subspace weights. Through experiments we performed on a field data acquired in the Gulf of Suez, we demonstrated the advantage of our method in comparison to the recursively weighted method without using limited subspace. We also introduced a computationally more efficient formulation by moving the weight matrices to the data-misfit term. In future work , we would like to extend the application of limited-subspace weighted method to large scale 3D data examples.

3.6 Related materials

In order to facilitate the reproducibility of the results herein discussed, Matlab & Julia implementation of this work are made available on the SLIM GitHub page https://github.com/slimgroup/Software.SEG2020.

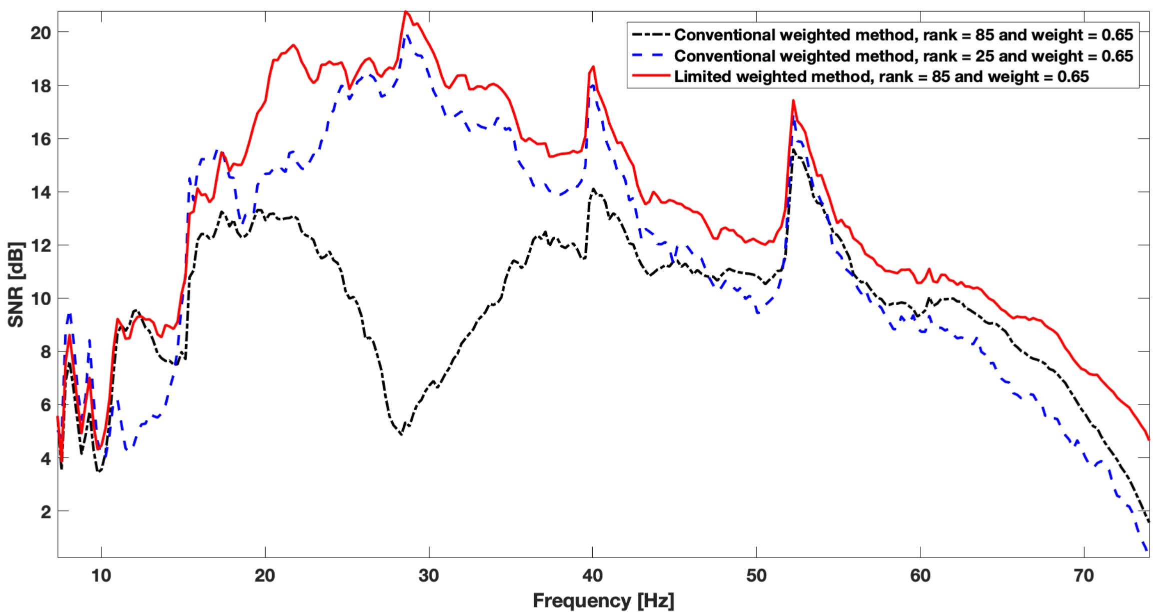


Figure 3.2: SNR of reconstructed data vs frequency based on our limited-subspace weighted method (red color), conventional weighted method with rank equals to 85 (black color) and 25 (blue color).

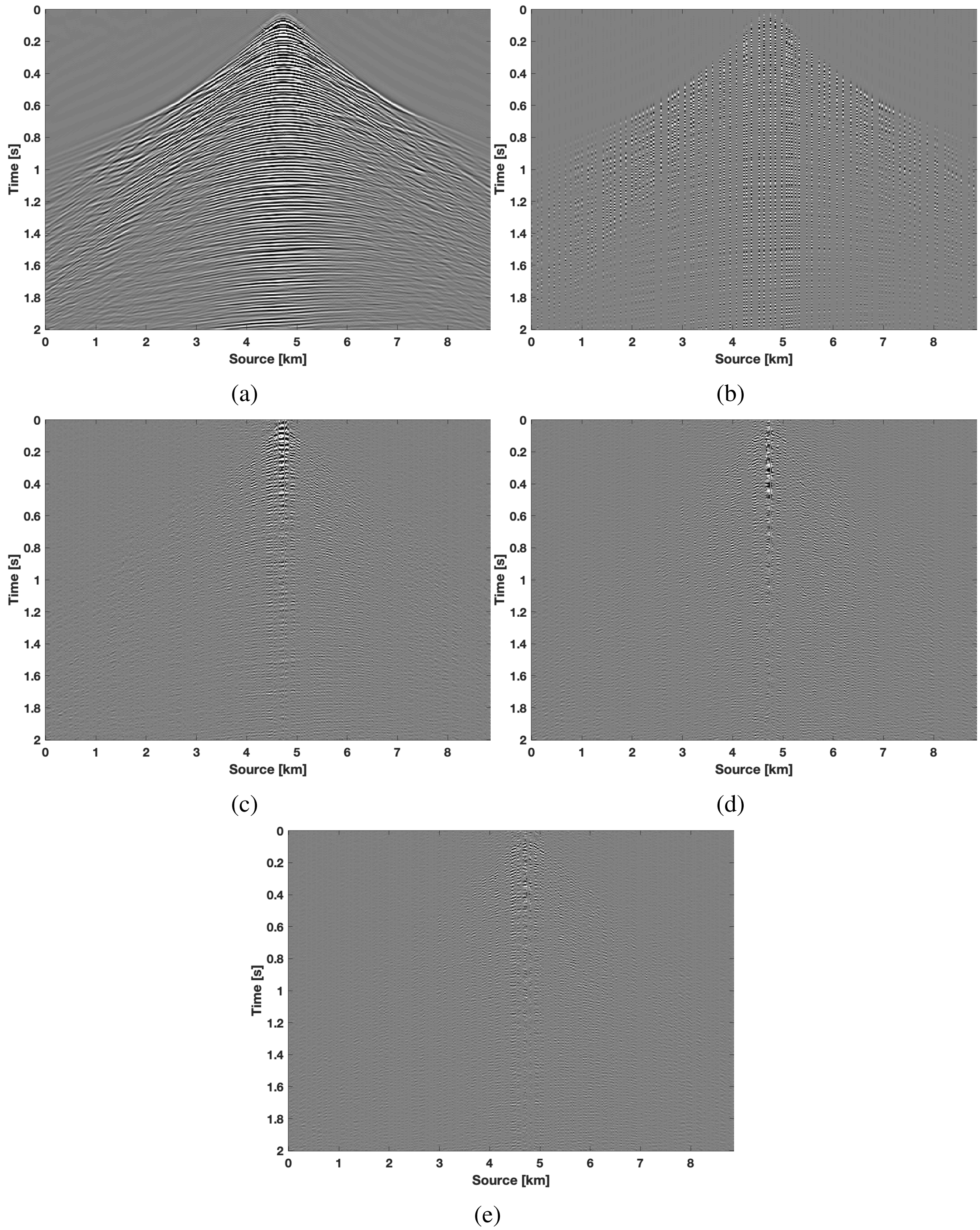


Figure 3.3: Wavefield reconstruction results in the time-domain. *(a)* Ground truth. (b) 75% subsampled seismic data with jittered subsampling. *(c)* using conventional weighted method ($SNR = 10.69\,\mathrm{dB}$) for rank equals to $r = 85$, *(d)* using conventional weighted method ($SNR = 11.49\,\mathrm{dB}$) for rank equals to $r = 25$, *(e)* using limited subspace weighted method ($SNR = 13.31\,\mathrm{dB}$) with limited subspace rank $r_s = 25$.

CHAPTER 4
LARGE SCALE HIGH-FREQUENCY WAVEFIELD RECONSTRUCTION WITH RECURSIVELY WEIGHTED MATRIX FACTORIZATIONS

4.1 Summary

Acquiring seismic data on a regularly spaced fine grid poses a challenge. However, by leveraging the low-rank approximation property of fully sampled seismic data in a specific transform domain, we can employ low-rank matrix completion. This approach offers a scalable solution for reconstructing seismic data on a regularly spaced fine grid from sparsely and randomly sampled data obtained in the field. While wavefield reconstruction has been successfully applied in the lower frequency range, its effectiveness diminishes at higher frequencies where the low-rank assumption no longer holds. This limitation hampers its utility in situations that require high-resolution images. To overcome this drawback, we capitalize on the explicit similarities between adjacent frequency slices. These similarities, manifested during low-rank matrix factorization, result in the alignment of subspaces of the factors. We propose to exploit this notion by recursively reconstructing monochromatic frequency slices, starting from the lower frequencies. Although the core idea is relatively straightforward, transforming this recent insight into a successful scalable wavefield reconstruction scheme for 3D seismic data involves several crucial steps. Firstly, we need to transfer the weighting matrices, which encapsulate prior information from adjacent frequency slices, from the objective to the data misfit constraint. This adjustment significantly enhances the performance of the weighted low-rank matrix factorization that underlies our wavefield reconstructions. Secondly, we introduce approximations that enable us to perform computations on a row-by-row and column-by-column basis, thereby facilitating the parallelization of the alternating optimization process central to our low-rank factorization.

The combination of weighting and decoupling results in a computationally feasible full-azimuth wavefield reconstruction scheme that scales to industry-scale problem sizes. We showcase the performance of the proposed parallel algorithm using both a 2D field dataset and a 3D synthetic dataset. In both cases, our approach produces high-fidelity broadband wavefield reconstructions from severely subsampled data.

4.2 Introduction

To achieve cost-effective extraction of hydrocarbon resources from the subsurface and mitigate hazardous situations, oil and gas companies heavily rely on accurate imaging and estimation of the physical parameters of the Earth's subsurface, such as wavespeed, density, etc. To obtain subsurface images and recover these physical parameters, practitioners employ a series of processing steps on the raw seismic data collected during seismic data acquisition in the field. Some of these processing steps, such as migration, demultiple, etc. necessitate finely sampled seismic data ideally on a regular grid. However, acquiring seismic data on a fine regular grid is often financially prohibitive and operationally complex. Therefore, the common practice in the oil and gas industry is to acquire seismic data on a coarse irregular grid and subsequently perform wavefield reconstruction to obtain a finer grid. In this study, we focus on wavefield reconstruction from randomized samples extracted from a periodic grid. For a more in-depth exploration of the presented wavefield reconstruction methodology, we refer the reader to Lopez, Kumar, Yilmaz, and Herrmann 2016 which discusses an off-the-grid extension.

In recent years, various methods for wavefield reconstruction have been developed. Many of these methods perform wavefield reconstruction in a transformed domain, involving Fourier (Xu, Zhang, Pham, and Lambaré 2005), Radon (Bardan 1987), wavelet (Villasenor, Ergas, and Donoho 1996), or curvelet (Herrmann and Hennenfent 2008) domain. These transformations, to different extents, promote sparsity in seismic data, which is a fundamental aspect of wavefield reconstruction based on compressive sensing (CS) (Candes, Romberg, and Tao 2006; Donoho 2006). Although powerful, sparsity-based wavefield reconstruction encounters scalability challenges in 3D seismic scenarios, where data volumes become excessively large when structure is promoted across more than three dimensions, e.g. along all four source and receiver coordinates. To address these challenges in higher dimensions, exploiting the low-rank properties of matrices and tensors (Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015; Oropeza and Sacchi 2011) has proven effective. This approach builds upon the earlier work of Recht, Fazel, and Parrilo 2010, who extended some of the ideas of CS to matrices. Similar to CS, matrix completion capitalizes on the low-rank approximation permitted by the underlying fully sampled data organized in a matrix. Randomized sampling, like in CS, disrupts the low-rank structure, which guides (convex) optimization techniques to recover wavefields by minimizing the rank. Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015 and Da Silva and Herrmann 2015 leverage this property and formalize matrix- and tensor-based wavefield reconstructions suitable for large-scale seismic datasets (Kumar, Wason, Sharan, and Herrmann 2017).

As demonstrated in the work by Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015, low-rank matrix completion methods perform well when reconstructing seismic data at lower angular frequencies. However, the quality of recovery degrades as we move to higher frequencies ($> 15\,\mathrm{Hz}$). This degradation is due to the fact that high-frequency slices are not accurately approximated by low-rank matrix factorization (Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015; Aravkin, Kumar, Mansour, Recht, and Herrmann 2014). Unfortunately, techniques like multiple elimination and migration require access to high-frequency data to create high-fidelity, artifact-free, high-resolution images. This need becomes even more crucial in areas with complex geology, where understanding the physical properties is of utmost interest.

To address the challenges of recovering seismic data at high frequencies, we build upon the earlier work of Aravkin, Kumar, Mansour, Recht, and Herrmann 2014 and Eftekhari, Yang, and Wakin 2018, who discussed methods to improve the performance of low-rank matrix completion by incorporating prior information in the form of weighting matrices. These weighting matrices are projections spanned by the row and column subspaces, along with their complements, of a low-rank matrix factorization that is close to the target matrix for recovery. Similar to weighted ℓ_1-norm minimization, these weighting matrices enhance wavefield recovery if the principal angle between the subspaces of the weighting matrices and the target matrix is small. Conceptually, this is the matrix equivalent of weighted ℓ_1-norm minimization proposed by Mansour, Herrmann, and Yılmaz 2012, Friedlander, Mansour, Saab, and Yilmaz 2012, Borries, Miosso, and Potes 2007, Vaswani and Lu 2010, Khajehnejad, Xu, Avestimehr, and Hassibi 2009, and Candes, Wakin, and Boyd 2008. Aravkin, Kumar, Mansour, Recht, and Herrmann 2014 and Eftekhari, Yang, and Wakin 2018 demonstrated that wavefield recovery via matrix completion can be improved by using low-rank factorizations from adjacent frequency slices to define these weighting matrices. In their work, Aravkin, Kumar, Mansour, Recht, and Herrmann 2014 employed a modified version of the spectral-projected gradient algorithm (Berg and Friedlander 2009) to assume access to the low-rank factorization of an adjacent frequency slice, while Eftekhari, Yang, and Wakin 2018 showed that for small principal angles, these weighting matrices reduce the sampling requirement for successful data reconstruction by a logarithmic factor compared to conventional matrix completion methods.

Although the initial results on wavefield reconstruction via weighted matrix completion were promising, the approach presented had practical limitations, relying on access to the weights and employing computationally expensive optimization algorithms. To overcome these shortcomings, we propose a parallelizable recursive method that utilizes a recently developed alternating minimization procedure (Xu and Yin 2013; Jain, Netrapalli, and Sanghavi 2013) proposed by Lopez, Kumar, and Herrmann 2015. Through recursive reconstruction, as initially proposed by Zhang, Sharan, and Herrmann 2019, and improved optimization, we demonstrate enhanced performance of our wavefield reconstruction algorithm.

The outline of our chapter is as follows. We first provide a brief overview of the principles of wavefield reconstruction via matrix completion. Following this introduction, we describe the challenges associated with high-frequency wavefield reconstruction and discuss how these challenges can be addressed through weighted matrix completion. We then describe the formulation in factored form, which drastically reduces the problem size, making our approach practical for 3D seismic data. In particular, we explain how our algorithm can be parallelized and applied to large-scale high-frequency seismic wavefield reconstruction problems.

4.3 Wavefield reconstruction via weighted matrix completion

According to the seminal work of Recht, Fazel, and Parrilo 2010, matrices that exhibit a low-rank structure can be recovered from randomly missing entries through a nuclear norm minimization procedure. This procedure minimizes the sum of the singular values. As long as the randomized subsampling reduces the rate of decay of the singular values, this type of minimization enables the recovery of matrices that are well approximated by low-rank matrices when fully sampled. Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015 applied this principle to recover frequency slices from seismic lines in the midpoint-offset domain or from 3D seismic data permuted in a non-canonical form (Da Silva and Herrmann 2015). In both cases, the resulting frequency slice can be accurately approximated by a low-rank matrix factorization.

To illustrate the underlying principle of wavefield reconstruction via matrix completion, let's consider a $12\,\mathrm{Hz}$ monochromatic frequency slice assembled from a 2D line acquired in the Gulf of Suez. Figure 4.1 includes the real part of this frequency slice in the source-receiver and midpoint-offset domain after removing 75

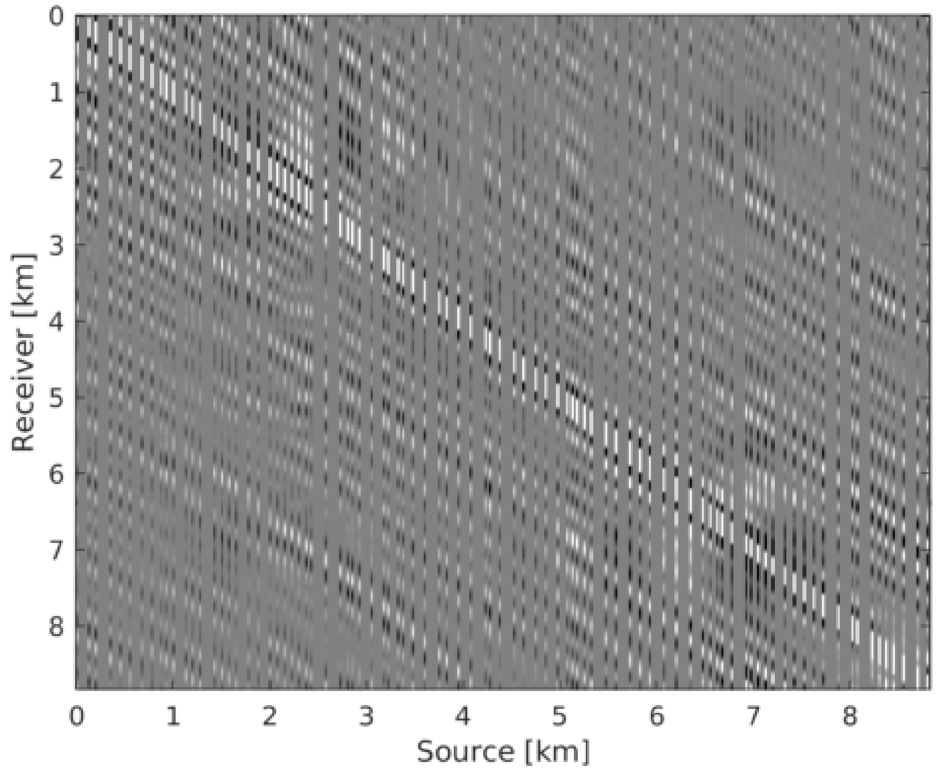
$$\mathbf{X} := \arg\min_{\mathbf{Y}} \|\mathbf{Y}\|_* \quad \text{subject to} \quad \|\mathcal{A}(\mathbf{Y}) - \mathbf{B}\|_F \leq \epsilon, \tag{4.1}$$

which promotes low-rank matrices.

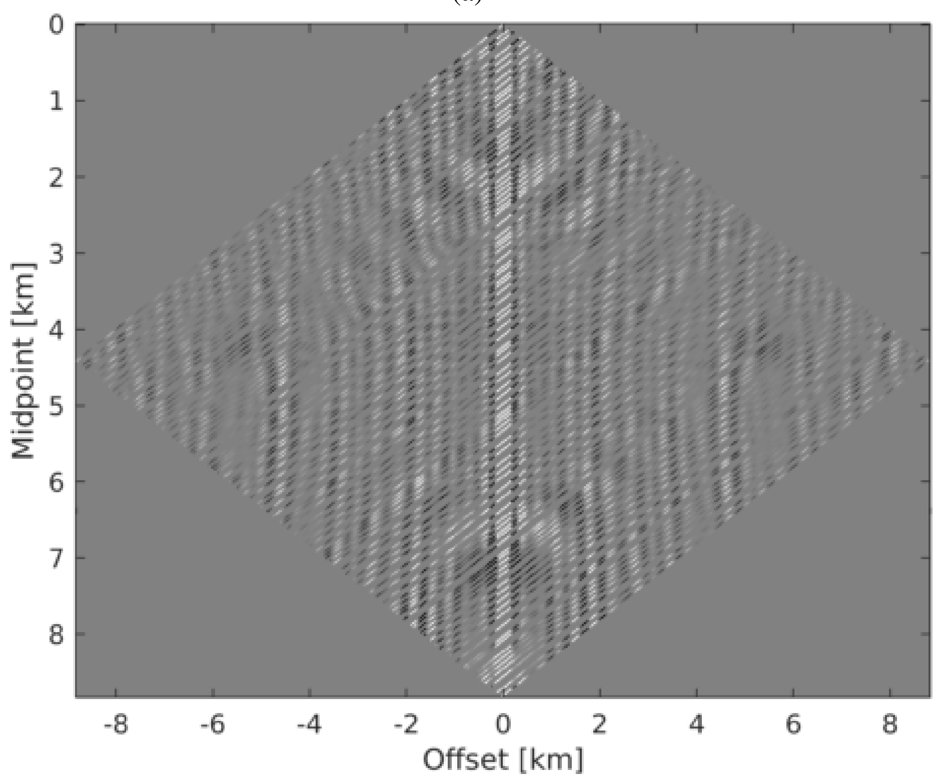
By solving this minimization problem, our goal is to recover the minimum nuclear norm ($\|\mathbf{X}\|_* = \sum \sigma_i$ with the sum running over the singular values of $\mathbf{X}$) of the complex-valued data matrix $\mathbf{X} \in \mathbb{C}^{m \times n}$, where m represents the number of offsets and n represents the number of midpoints. In addition to minimizing the nuclear norm objective, the minimizer fits the observed data $\mathbf{B} \in \mathbb{C}^{m \times n}$ at the sampling locations within a certain tolerance ϵ, measured by the Frobenius norm, i.e., $\|\mathbf{D}\|_F = \sqrt{\sum_j \sum_k D_{jk}^2}$ for a matrix $\mathbf{D}$. The linear operator $\mathcal{A}$ implements the sampling mask by setting zeros at the source (and possibly receiver) locations that were not collected in the field. The optimization variable is represented by the matrix $\mathbf{Y}$. Equation 4.1 resembles the classic Basis Pursuit DeNoising problem (BPDN, Berg and Friedlander 2009) and can be solved using a modified version of the SPGℓ_1 algorithm, adapted for nuclear-norm minimization (Aravkin, Kumar, Mansour, Recht, and Herrmann 2014). To solve problem Equation 4.1, SPGℓ_1 solves a series of constrained subproblems, during which the nuclear-norm constraint is relaxed to fit the observed data.

4.3.1 The challenge of high-frequency wavefield recovery

Wavefield reconstruction via matrix completion (cf. problem Equation 4.1) relies on the assumption that the singular values of monochromatic data organized in matrix decay rapidly. This assumption holds true for the lower frequencies ($< 15.0\,\mathrm{Hz}$), but unfortunately, it no longer holds for higher frequencies. To illustrate this phenomenon, we compare the decay of singular values for the two matricizations of Figure 4.2 at $12.0\,\mathrm{Hz}$ and $60.0\,\mathrm{Hz}$ in Figure 4.3. While the singular values at $12.0\,\mathrm{Hz}$ decay quickly, this is not the case at $60.0\,\mathrm{Hz}$ (compare solid lines in Figure 4.3a and Figure 4.3b), where the singular values for the fully sampled data decay more slowly. The slower decay at the high frequencies is caused by the increased complexity and oscillatory behavior exhibited by data at higher temporal frequencies. Despite the fact that the randomized source subsampling slows down the

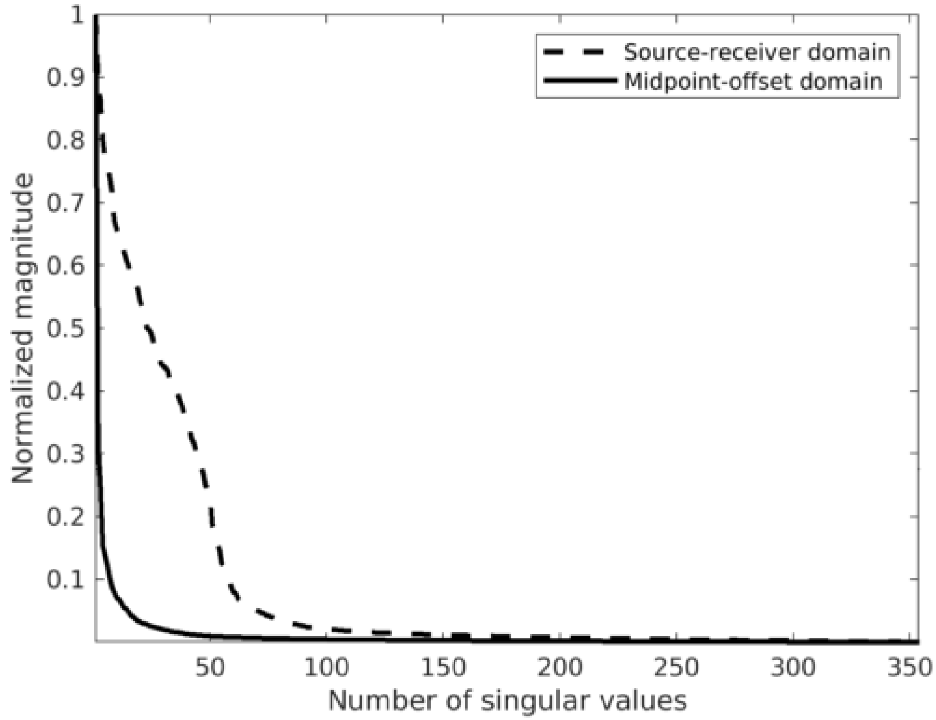


(a)

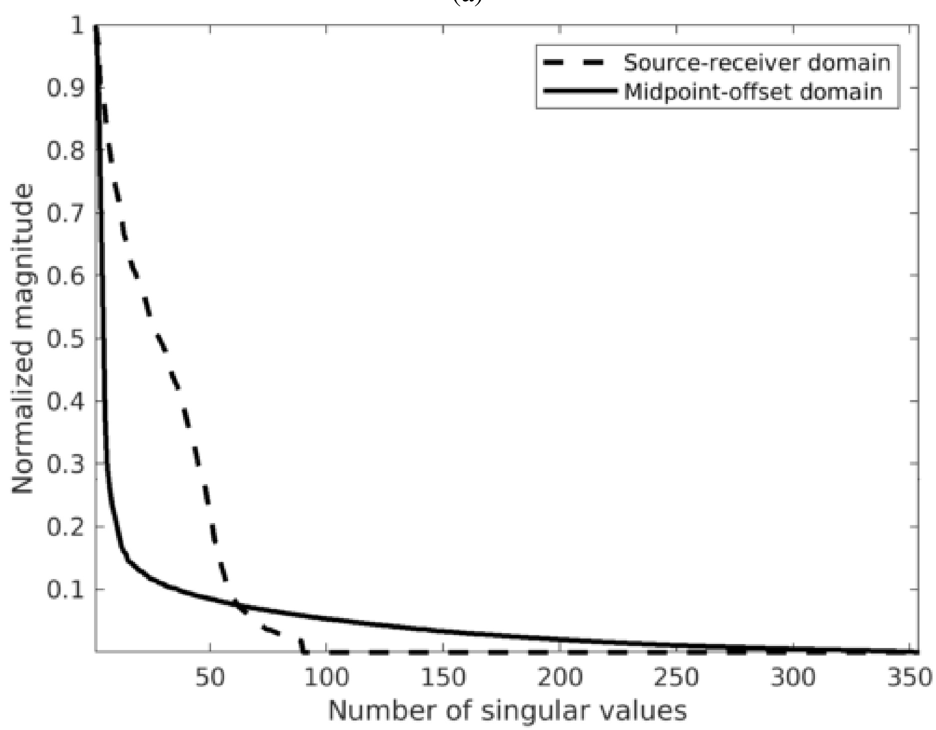


(b)

Figure 4.1: $12.0\,\mathrm{Hz}$ frequency slice extracted from 2D seismic data acquired in Gulf of Suez. Data with 75% missing random jittered sources in *(a)* source-receiver domain and *(b)* in midpoint-offset domain.



(a)



(b)

Figure 4.2: Decay of singular values for $12.0\,\mathrm{Hz}$ frequency slice in source-receiver and midpoint-offset domain for *(a)* full data and for *(b)* subsampled data with 75% missing sources.

decay, the slower decay of the fully sampled data leads to poor wavefield reconstruction (Figure 4.4a) and unacceptable large residuals (Figure 4.4b) at $60.0\,\mathrm{Hz}$.

4.3.2 Weighted matrix completion

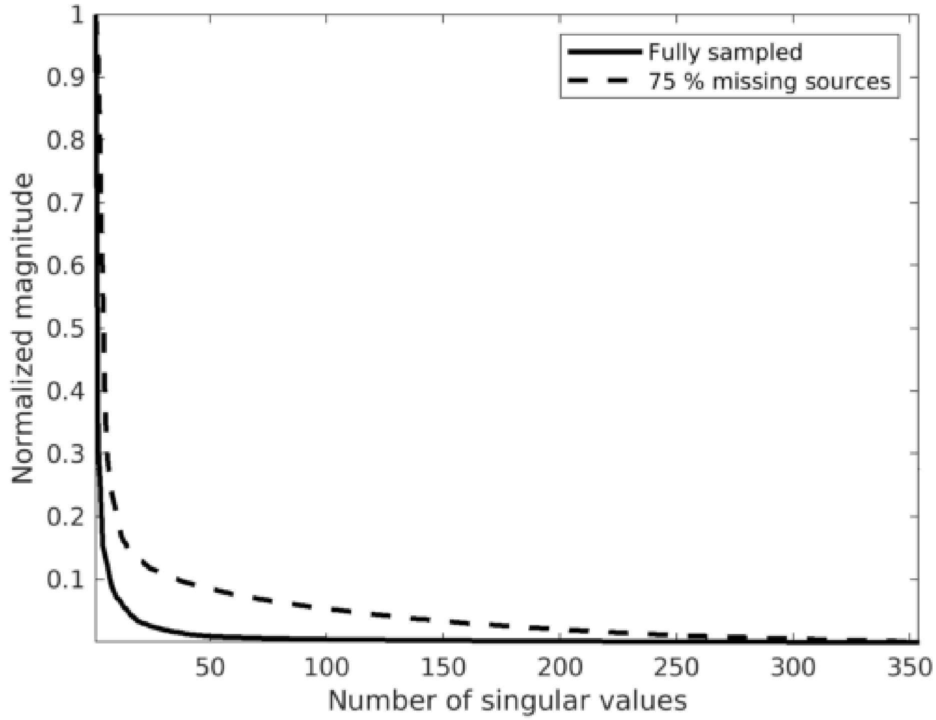
As shown in Figure 4.3, Figure 4.4a, and Figure 4.4b, the success of wavefield reconstruction through the minimization of the nuclear norm (cf. Equation 4.1) hinges on rapid decay of the singular values an assumption violated at the higher frequencies. This shortcoming can, at least in part, be overcome by using prior information from a related problem in the form of weights, an approach initially put forward by Aravkin, Kumar, Mansour, Recht, and Herrmann 2014 and further theoretically analyzed by Eftekhari, Yang, and Wakin 2018. In its original form, the weights were derived from the reconstruction of the wavefield at a neighboring temporal frequency, which leads to a significant improvement for the reconstruction and the residual plotted in Figure 4.4c and Figure 4.4d, respectively. Building upon this approach, Zhang, Sharan, and Herrmann 2019 further enhanced the reconstruction by applying it recursively from low to high frequencies, leading to improved quality as depicted in Figure 4.4e and a reduced residual size as shown in Figure 4.4f. In this work, we extend these results by reformulating the optimization problem and introducing a parallel algorithm that minimizes communication.

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

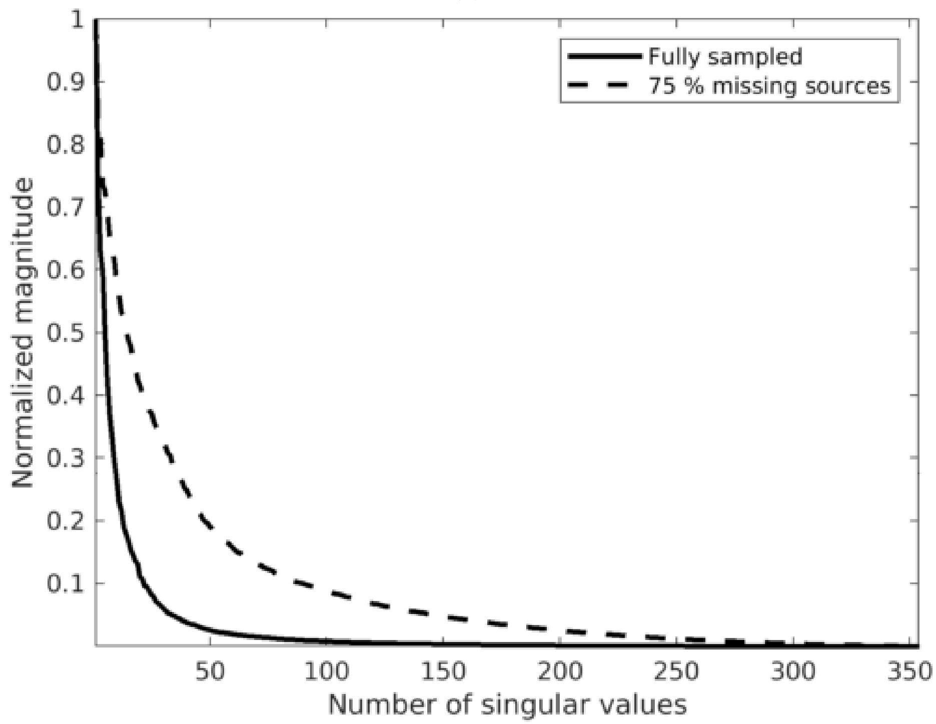
$$\mathbf{X} := \arg\min_{\mathbf{Y}} \|\mathbf{Q}\mathbf{Y}\mathbf{W}\|_* \quad \text{subject to} \quad \|\mathcal{A}(\mathbf{Y}) - \mathbf{B}\|_F \leq \epsilon, \tag{4.2}$$

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

$$\mathbf{Q} = w_1 \mathbf{U}\mathbf{U}^H + \mathbf{U}^{\perp}{\mathbf{U}^{\perp}}^H \tag{4.3}$$



(a)



(b)

Figure 4.3: Singular value decay for fully sampled and subsampled data (75% missing sources) in midpoint-offset domain for *(a)* $12.0\,\mathrm{Hz}$ and *(b)* $60.0\,\mathrm{Hz}$ frequency slice

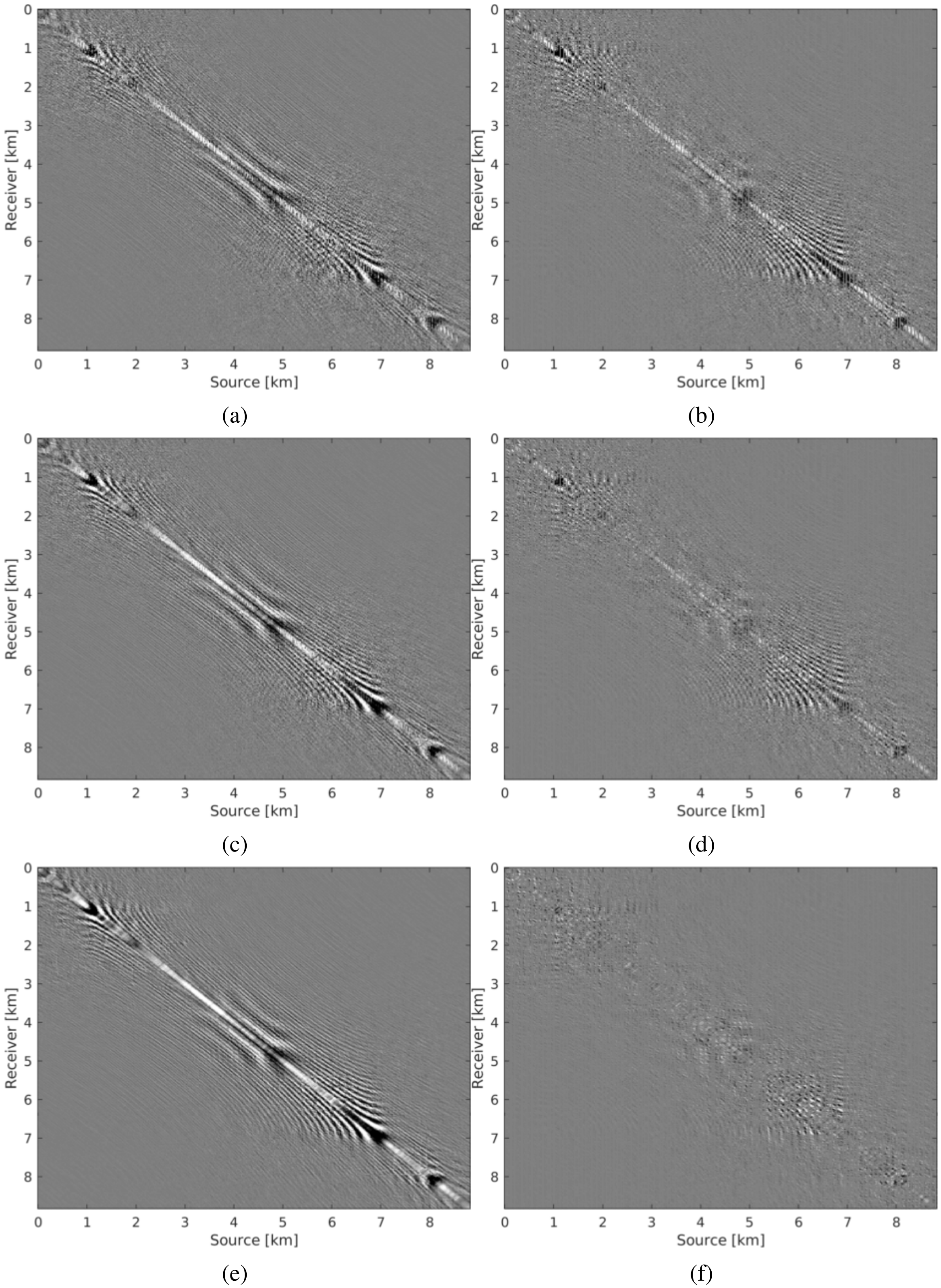


Figure 4.4: Wavefield reconstruction comparison for a 60 Hz frequency slice. *(a)* Reconstructed wavefield from 75% subsampling. *(b)* residual with a poor S/R = 2.83 dB. *(c)* Reconstructed wavefield using the recovery at the adjacent lower frequency as weights and *(d)* improved residual with S/R = 5.08 dB. *(e)* and *(f)* the same but now with the weighting scheme applied recursively with significantly improved S/R = 8.72 dB.

and

$$\mathbf{W} = w_2 \mathbf{V}\mathbf{V}^H + \mathbf{V}^{\perp}\mathbf{V}^{\perp^H}, \tag{4.4}$$

where the symbol ^{H} denotes complex transpose. These projections are spanned by the row and column subspaces $\mathbf{U}$, $\mathbf{V}$ and their orthogonal complements $\mathbf{U}^{\perp}$ and $\mathbf{V}^{\perp}$. Furthermore, the subspaces $\mathbf{U}$, $\mathbf{V}$ have orthonormal columns, resulting in the orthogonal projections $\mathbf{U}\mathbf{U}^H$ and $\mathbf{V}\mathbf{V}^H$. The pair of matrices $\{\mathbf{U}, \mathbf{V}\}$ form a low-rank pair that can be obtained from the factorization of a lower adjacent frequency slice. The choice for the weights w_1 and w_2 in Equation 4.3 and Equation 4.4 depends on the similarity between the corresponding row and column subspaces of the two adjacent frequency slices. Following Eftekhari, Yang, and Wakin 2018, we quantify this similarity using the largest principle angle between these subspaces. A smaller angle indicates a higher similarity between the subspaces from the two adjacent frequency slices. In cases where the adjacent frequency slices are near orthogonal—i.e., have a near 90° angle, we set $w_1 \uparrow 1$ and $w_2 \uparrow 1$ so that the weighting matrices $\mathbf{Q}$ and $\mathbf{W}$ become identity matrices. Here, the symbol $\uparrow 1$ represents an approximation 1 from below. In such situations, the weighting matrices should not introduce additional information—i.e., the solution of problem Equation 4.2 should become equivalent to solving the original problem in Equation 4.1. Conversely, when the subspaces are similar—i.e., they have an angle $\ll 90^\circ$, then the w_1 and w_2 should be chosen small such that we penalize solutions more in the orthogonal complement space. The choice of these weights depends on our confidence in the given factorization: we select weights close to one when we have little confidence and close to zero when we have higher confidence.

While replacing the nuclear-norm objective in Equation 4.1 within its weighted counterpart in Equation 4.2 is a valid approach that has shown improvements as reported in Figure 4.4, solving this weighted problem involves non-trivial weighted projections (see equation 7.3 in Aravkin, Kumar, Mansour, Recht, and Herrmann 2014). These computationally expensive operations can be avoided by reformulating the optimization problem Equation 4.2 in a slightly different manner, where the weights are transferred from the objective to the data constraint—i.e., we have

$$\bar{\mathbf{X}} := \arg\min_{\bar{\mathbf{Y}}} \|\bar{\mathbf{Y}}\|_* \quad \text{subject to} \quad \|\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{Y}}\mathbf{W}^{-1}) - \mathbf{B}\|_F \leq \epsilon. \tag{4.5}$$

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

$$\mathbf{Q}^{-1} = \frac{1}{w_1}\mathbf{U}\mathbf{U}^H + \mathbf{U}^{\perp}\mathbf{U}^{\perp^H} \tag{4.6}$$

and

$$\mathbf{W}^{-1} = \frac{1}{w_2}\mathbf{V}\mathbf{V}^H + \mathbf{V}^{\perp}\mathbf{V}^{\perp^H}. \tag{4.7}$$

By transferring the weighting matrices to the data constraint, we eliminate the need to project onto a more complex constraint, as demonstrated in Aravkin, Kumar, Mansour, Recht, and Herrmann 2014. As a result, the solutions obtained from the modified formulation in Equation 4.5 can be obtained at nearly the same computational cost as the original formulation in Equation 4.1. This formulation serves as the foundation for our wavefield reconstruction approach, designed to handle the substantial data volumes present in 3D seismic applications.

4.4 Scalable multi-frequency seismic wavefield reconstruction

Up until now, our minimization problems have relied on explicitly forming the data matrix and utilizing singular-value decomposition (SVD) techniques, as described in Aravkin, Kumar, Mansour, Recht, and Herrmann 2014. However, these approaches are impractical for industry-scale 3D wavefield reconstruction problems. To tackle this challenge, we explore the transformation of the aforementioned weighted matrix completion approach into a factored form. This alternative formulation offers computational advantages and, as we will demonstrate later, remains amenable to parallelization.

4.4.1 Weighted low-rank matrix factorization

To avoid computing costly SVDs, we first cast the solution of Equation 4.5 into factored form:

$$\bar{\mathbf{L}}, \bar{\mathbf{R}} := \underset{\bar{\mathbf{L}}_\#, \bar{\mathbf{R}}_\#}{\arg\min} \frac{1}{2} \left\| \begin{bmatrix} \bar{\mathbf{L}}_\# \\ \bar{\mathbf{R}}_\# \end{bmatrix} \right\|_F^2 \quad \text{subject to} \quad \|\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{L}}_\#\bar{\mathbf{R}}_\#^H\mathbf{W}^{-H}) - \mathbf{B}\|_F \leq \epsilon, \tag{4.8}$$

where $\bar{\mathbf{L}} = \mathbf{QL}$ and $\bar{\mathbf{R}} = \mathbf{WR}$. Under certain technical conditions (Candes and Recht 2009), which include choosing the proper rank r, the factored solution, $\mathbf{X} = \mathbf{LR}^H$ with $\mathbf{L} = \mathbf{Q}^{-1}\bar{\mathbf{L}}$ and $\mathbf{R} = \mathbf{W}^{-1}\bar{\mathbf{R}}$, corresponds to the solution of the weighted problem included in Equation 4.2. Here, the matrices $\mathbf{L} \in \mathbb{C}^{m\times r}$ and $\mathbf{R} \in \mathbb{C}^{n\times r}$ are the low-rank factors of $\mathbf{X}$. Using the property that the matrices $\mathbf{W}^H = \mathbf{W}$ and $\mathbf{Q}^H = \mathbf{Q}$ in Equation 4.8 are idempotent, we replace $\mathbf{W}^{-H}$ by $\mathbf{W}^{-1}$ to avoid extra computation. Compared to the original convex formulation, solving Equation 4.8 is made possible through the use of block coordinate descent (Xu and Yin 2013), which offers computational efficiency, as demonstrated by the runtimes shown in Figure 4.5 as a function of temporal frequency. The block coordinate descent algorithm efficiently finds the low-rank factors, allowing us to solve the low-rank matrix completion problem. However, it's important to note that this approach is effective only when the monochromatic data matrices can be well approximated by low-rank matrices, meaning that the rank parameter $r \ll \min(m, n)$.

Although the weighted formulation described above enables solving the problem in factored form, it requires access to the subspaces $\{\mathbf{U}, \mathbf{V}\}$, which necessitates computing the full SVD (Eftekhari, Yang, and Wakin 2018). However, since computing the full SVD is not feasible, we adopt a different approach. Instead, we orthogonalize the low-rank factors from adjacent frequency slices themselves by performing computationally inexpensive

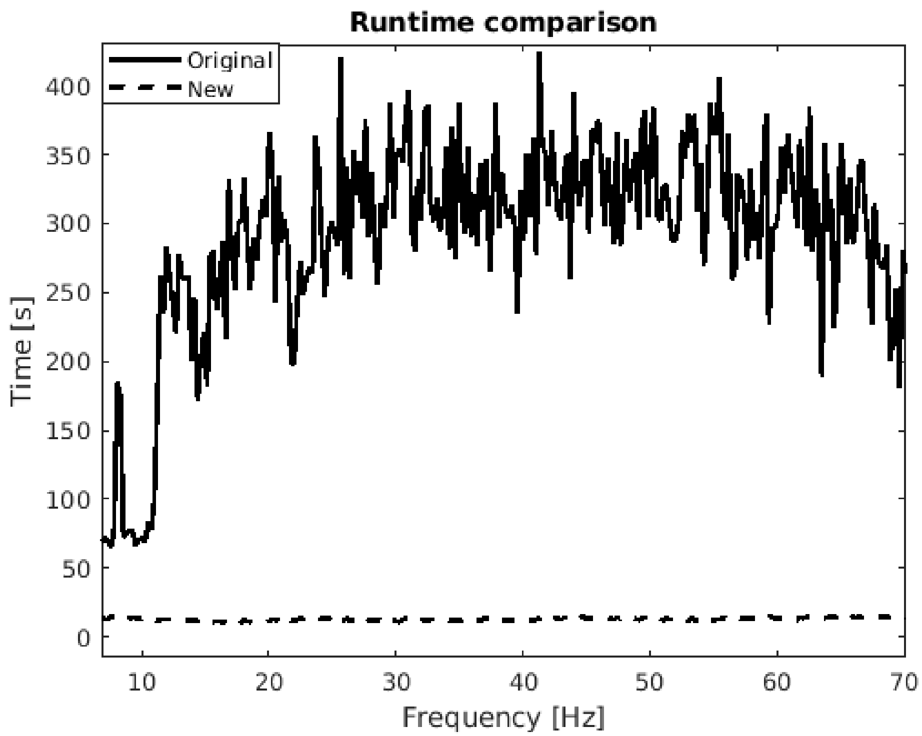


Figure 4.5: Runtime comparison plot: Solid black line shows runtime of the original weighted formulation and dashed black line shows runtime of the new weighted formulation for same number of iterations with same data residual at the end.

SVDs on these factors, rather than on the entire data matrix, and retaining only the top r left singular vectors. This strategy is justified because orthogonalizing the low-rank factors allows us to approximate the orthogonal subspaces spanned by the complete frequency slice.

The results presented in Figure 4.4 were obtained in factored form and clearly demonstrate the benefits of incorporating weight matrices, particularly when these matrices are recursively calculated from low to high frequencies (juxtapose Figure 4.4c, Figure 4.4d and Figure 4.4e, Figure 4.4f). This improvement is attributed to the fact that low-frequency data matrices can be more effectively approximated by low-rank matrices, which enhances the recovery process and, consequently, the quality of the weighted reconstruction.

4.4.2 Weighted parallel recovery

The example in Figure 4.4 clearly demonstrates the enhanced performance of wavefield reconstruction through matrix factorization when weight matrices, containing information on the row and column subspaces, are incorporated. However, the inclusion of these weight matrices complicates the parallelization of the algorithm, as it no longer straightforwardly applies the parallelized alternating optimization approach proposed by Recht and Ré 2013 and Lopez, Kumar, and Herrmann 2015.

This approach relies on independent computations performed on a row-by-row and column-by-column basis (see Figure 4.6) , in which the optimization alternates between minimizing the rows via

$$\mathbf{R}(l_1,:)^H := \arg\min_{\mathbf{v}} \frac{1}{2}\|\mathbf{v}\|^2 \quad \text{subject to} \quad \|\mathcal{A}_{l_1}(\mathbf{L}\mathbf{v}) - \mathbf{B}(:,l_1)\| \leq \gamma \tag{4.9}$$

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

$$\mathbf{L}(l_2,:)^H := \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{u}\|^2 \quad \text{subject to} \quad \|\mathcal{A}_{l_2}((\mathbf{R}\mathbf{u})^H) - \mathbf{B}(l_2,:)\| \leq \gamma \tag{4.10}$$

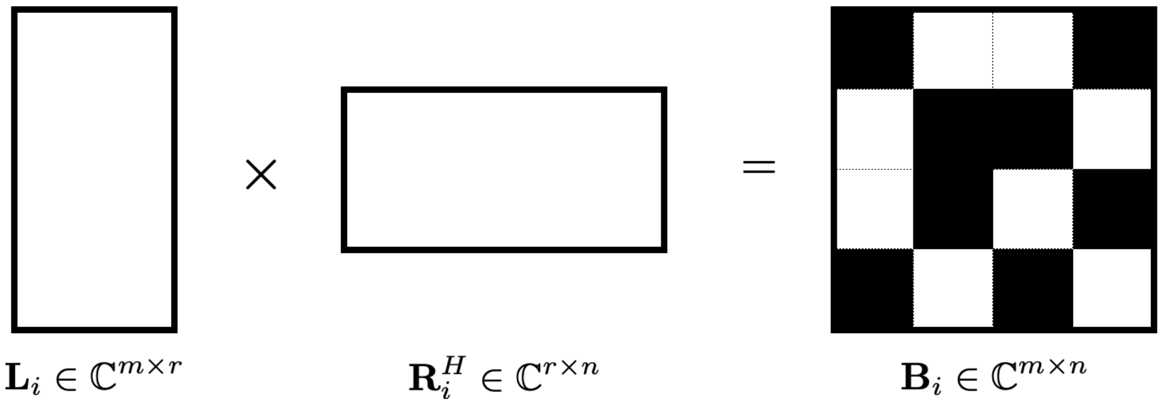
for $l_2 = 1 \cdots m$. With this approach, the rows of the right factor $\mathbf{R}$ are updated first by iterating over the rows via the index $l_1 = 1 \cdots n$. Subsequently, updates are performed on the rows of the left factor $\mathbf{L}$ by iterating over the rows via the index $l_2 = 1 \cdots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

The operations in Equation 4.9 and Equation 4.10 allow for a parallel implementation that scales well for large-scale industrial 3D seismic problems since they decouple the computations. However, this decoupled formulation lacks the inclusion of weighting matrices, which limits its utility for recovery problems at higher frequencies where weighting matrices are required. To address this limitation, we propose a novel approach to ameliorate this problem in which we take Equation 4.9 and Equation 4.10 as a starting point and pre- and post multiply the data misfit terms by $\mathbf{Q}$ and $\mathbf{W}$ after including the weighting matrices as in Equation 4.8. Furthermore, we leverage the property that for large weights, the matrices $\mathbf{Q}$ and $\mathbf{W}$ nearly commute with the measurement operator $\mathcal{A}$—i.e., we have $\mathbf{Q}\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{X}}\mathbf{W}^{-1}) \approx \mathcal{A}(\bar{\mathbf{X}}\mathbf{W}^{-1})$ and $\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{X}}\mathbf{W}^{-1})\mathbf{W} \approx \mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{X}})$ where $\bar{\mathbf{X}}$ represents the fully sampled data matrix or its factored form. With these approximations, we arrive at the following weighted iterations:

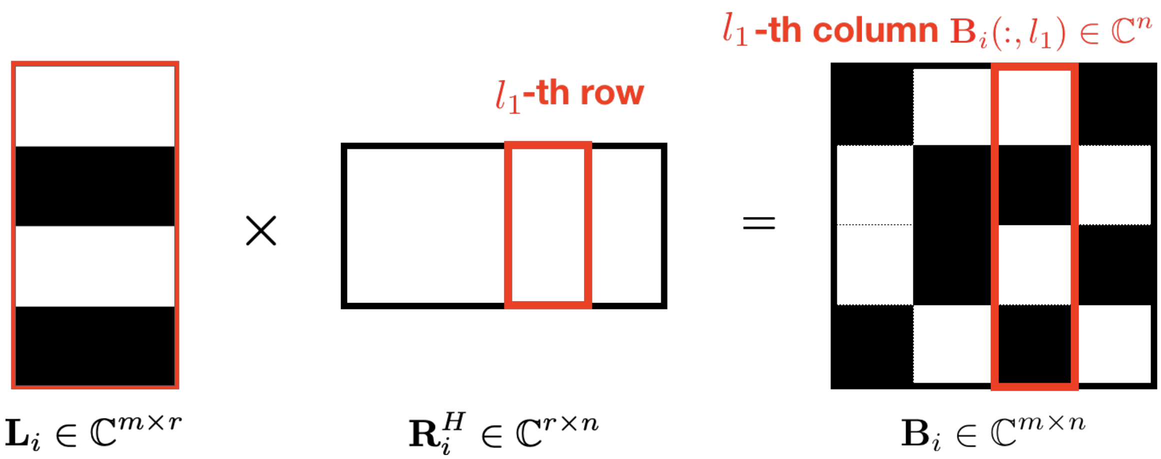
$$\bar{\mathbf{R}}(l_1, :)^H := \arg\min_{\bar{\mathbf{v}}} \frac{1}{2} \|\bar{\mathbf{v}}\|_2^2 \quad \text{subject to} \quad \|\mathcal{A}_{l_1}(\mathbf{Q}^{-1}\bar{\mathbf{L}}\bar{\mathbf{v}}) - \mathbf{B}_R(:, l_1)\| \leq \gamma \tag{4.11}$$

for $l_1 = 1 \cdots n$ and

$$\bar{\mathbf{L}}(l_2, :)^H := \arg\min_{\bar{\mathbf{u}}} \frac{1}{2} \|\bar{\mathbf{u}}\|^2 \quad \text{subject to} \quad \|\mathcal{A}_{l_2}((\bar{\mathbf{R}}\bar{\mathbf{u}})^H\mathbf{W}^{-1}) - \mathbf{B}_L(l_2, :)\| \leq \gamma \tag{4.12}$$



(a)



(b)

Figure 4.6: Alternating minimization and decoupling. *(a)* Solving for the low-rank factor $\mathbf{R}$ by using fixed factor $\mathbf{L}$ and observed data $\mathbf{B}$. *(b)* Solving for the l_1^{th} row of the low-rank factor $\mathbf{R}$ by using rows (in black color) of the fixed factor $\mathbf{L}$ corresponding to the non-zero entries (in black color) of the l_1^{th} column from the observed data $\mathbf{B}$.

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

The validity of the above derivation depends on the accuracy of the approximations involving the commutation of the weight matrices with the sampling operator $\mathcal{A}$. To assess the justification of these approximations, we compare the accuracy in Figure 4.7 by comparing plots of $\mathbf{Q}\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{X}}\mathbf{W}^{-1})$ and $\mathcal{A}(\bar{\mathbf{X}}\mathbf{W}^{-1})$ for two different weight values in Equation 4.6 and Equation 4.7. As expected, for the smaller weight value $w_{1,2} = 0.25$ the weighting matrix $\mathbf{Q}$ does not commute with the sampling matrix (see Figure 4.7a – Figure 4.7c). However, for $w_{1,2} = 0.75$ the approximation is reasonably accurate (see Figure 4.7d – Figure 4.7f). Similarly, in Figure 4.8 we compare plots of $\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{X}}\mathbf{W}^{-1})\mathbf{W}$ and $\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{X}})$ for small and large weights. As before, for smaller weights $w_{1,2} = 0.25$, the weighing matrix $\mathbf{W}$ does not commute with the sampling matrix (see Figure 4.8a – Figure 4.8c). However, for $w_{1,2} = 0.75$ the approximation is again reasonably accurate (see Figure 4.8d – Figure 4.8f). It is important to note that the weights $w_{1,2}$ reflect confidence we have in the weight matrices and are chosen to be small when we have confidence that the weighting matrices $\mathbf{Q}$ and $\mathbf{W}$ contribute valuable information to the recovery process. This means we need to select a value for the weights $w_{1,2}$ that strikes a balance between the amount of prior information we want to incorporate and the desired accuracy of the commutation relations. Choosing small weights results in larger "commutation" errors, while large weights leads to small "commutation" errors but limit the incorporation of prior information through the weights.

Although, the decoupled Equation 4.11 and Equation 4.12 can now be parallelized over the rows of the low-rank factors $\bar{\mathbf{R}}$ and $\bar{\mathbf{L}}$, they come with additional computational cost. Unlike the sparse observed data collected in the matrix $\mathbf{B}$, the data matrices $\mathbf{B}_R$ and $\mathbf{B}_L$ are dense (have all non-zero entries) because of the multiplications by $\mathbf{W}$ and $\mathbf{Q}$. However, when the weights $w_{1,2}$ are relatively large, we observe that both dense matrices $\mathbf{B}_L$,

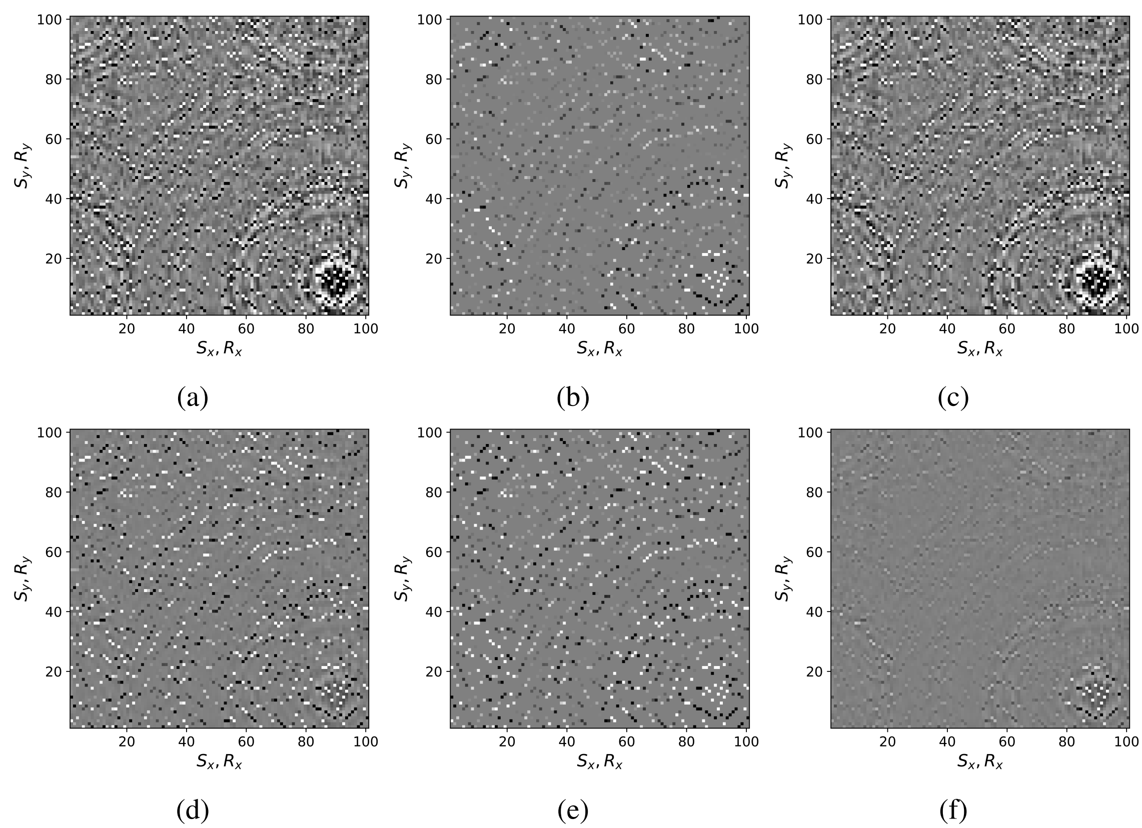


Figure 4.7: Commutation test for small and large weights. *(a)* Subset of 3D frequency slice for $\mathbf{Q}\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{X}}\mathbf{W}^{-1})$ for $w_{1,2} = 0.25$; *(b)* the same but now for $\mathcal{A}(\bar{\mathbf{X}}\mathbf{W}^{-1})$; *(c)* difference plot between *(a)* and *(b)*; *(d)*-*(f)* the same as *(a)*-*(c)* but now for $w_{1,2} = 0.75$.

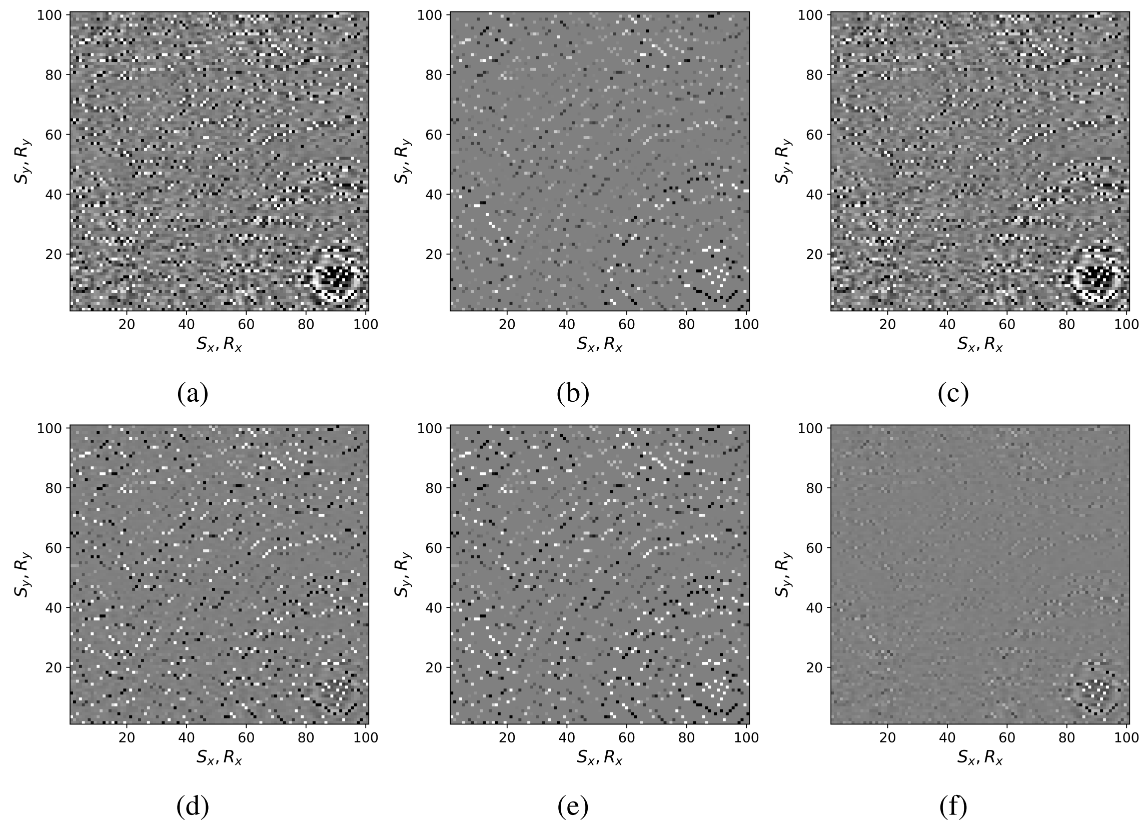


Figure 4.8: Commutation test for small and large weights. *(a)* Subset of 3D frequency slice for $\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{X}}\mathbf{W}^{-1})\mathbf{W}$ for $w_{1,2} = 0.25$; *(b)* the same but now for $\mathcal{A}(\mathbf{Q}^{-1}\bar{\mathbf{X}})$; *(c)* difference plot between *(a)* and *(b)*; *(d)*-*(f)* the same as *(a)*-*(c)* but now for $w_{1,2} = 0.75$.

$\mathbf{B}_R$ (Figure 4.9b and Figure 4.9d) can be well approximated by the sparse observed data matrix $\mathbf{B}$, as indicated by the difference plots in Figure 4.9. With this approximation, Equation 4.11 and Equation 4.12 becomes computationally efficiently.

While the above formulation allows us to perform weighted factored wavefield recovery in parallelized form, we have observed that taking inverses of $\mathbf{Q}$ and $\mathbf{W}$ in the data misfit objective (see Equation 4.8) results in inferior recovery due to the involvement of reciprocals of the weights (see Equation 4.6 and Equation 4.7). The value range of these reciprocals is no longer contained to the interval $(0, 1]$, which can introduce numerical issues during the recovery process. To overcome this problem, we propose an alternative but equivalent form for the weighted formulation, where the weights are defined as

$$\widehat{\mathbf{Q}} = \mathbf{U}\mathbf{U}^H + w_1 \mathbf{U}^\perp {\mathbf{U}^\perp}^H = w_1 \mathbf{Q}^{-1}, \tag{4.13}$$

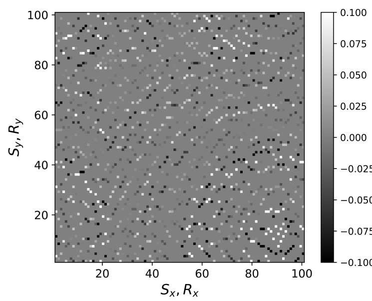
$$\widehat{\mathbf{W}} = \mathbf{V}\mathbf{V}^H + w_2 \mathbf{V}^\perp {\mathbf{V}^\perp}^H = w_2 \mathbf{W}^{-1} \tag{4.14}$$

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

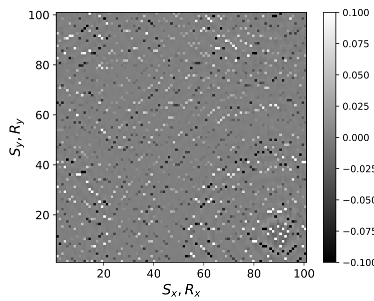
$$\begin{aligned} \bar{\mathbf{R}}(l_1,:)^H := \quad & \arg\min_{\bar{\mathbf{v}}} \frac{1}{2}\|\bar{\mathbf{v}}\|^2 \\ & \text{subject to} \\ & \|\mathcal{A}_{l_1}(\widehat{\mathbf{Q}}\bar{\mathbf{L}}\bar{\mathbf{v}}) - w_1 w_2 \mathbf{B}(:, l_1)\| \leq w_1 w_2 \gamma \end{aligned} \tag{4.15}$$

for $l_1 = 1 \cdots n$ and

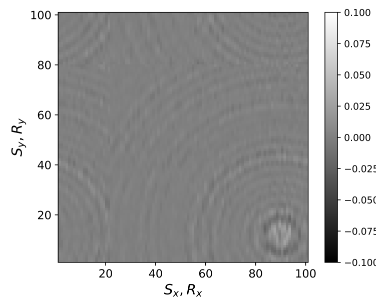
$$\begin{aligned} \bar{\mathbf{L}}(l_2,:)^H := \quad & \arg\min_{\bar{\mathbf{u}}} \frac{1}{2}\|\bar{\mathbf{u}}\|^2 \\ & \text{subject to} \\ & \|\mathcal{A}_{l_2}((\bar{\mathbf{R}}\bar{\mathbf{u}})^H\widehat{\mathbf{W}}) - w_1 w_2 \mathbf{B}(l_2, :)\| \leq w_1 w_2 \gamma \end{aligned} \tag{4.16}$$



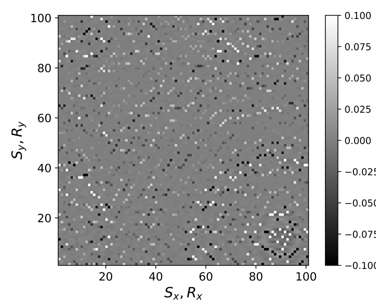
(a)



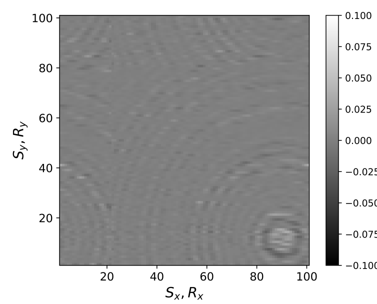
(b)



(c)



(d)



(e)

Figure 4.9: Accuracy of sparse approximation for weights $w_{1,2} = 0.75$, *(a)* Subset of 3D frequency slice for sparse observed data $\mathbf{B}$; *(b)* the same but now for the dense matrix $\mathbf{B}_L$ with $S/R = 17.5\,\mathrm{dB}$; *(c)* difference plot between *(a)* and *(b)*; *(d)* Subset of 3D frequency slice for the dense matrix $\mathbf{B}_R$ with $S/R = 16.5\,\mathrm{dB}$; *(e)* difference plot between *(a)* and *(d)*.

for $l_2 = 1 \cdots m$. Equation 4.15 and Equation 4.16 form the basis for our recovery approach summarized in Algorithm 1 below, which corresponds to

$$\min_{\bar{\mathbf{X}}} \quad \|\bar{\mathbf{X}}\|_* \quad \text{subject to} \quad \|\mathcal{A}(\widehat{\mathbf{Q}}\bar{\mathbf{X}}\widehat{\mathbf{W}}) - w_1 w_2 \mathbf{B}\|_F \leq w_1 w_2 \epsilon, \tag{4.17}$$

which is equivalent to Equation 4.5 as we show in Appendix A.

Algorithm 1 Weighted minimization via Alternating minimization.

---

0. **Input:** Observed data $\mathbf{B}$, rank parameter r, acquisition mask $\mathcal{A}$, priors $\widehat{\mathbf{Q}}$, $\widehat{\mathbf{W}}$ and initial guess $\bar{\mathbf{L}}^{(0)}$
1. **for** $k = 0,\, 1,\, 2, \cdots,\, N-1$ solve for rows of $\bar{\mathbf{R}}$ and $\bar{\mathbf{L}}$ in parallel
2. $\bar{\mathbf{R}}^{(k+1)}(l_1,:)^H := \arg\min_{\bar{\mathbf{v}}} \frac{1}{2}\|\bar{\mathbf{v}}\|^2$ s.t.$\|\mathcal{A}_{l_1}(\widehat{\mathbf{Q}}\bar{\mathbf{L}}^{(k)}\bar{\mathbf{v}}) - w_1 w_2 \mathbf{B}(:,l_1)\| \leq w_1 w_2 \gamma$
3. $\bar{\mathbf{L}}^{(k+1)}(l_2,:)^H := \arg\min_{\bar{\mathbf{u}}} \frac{1}{2}\|\bar{\mathbf{u}}\|^2$ s.t.$\|\mathcal{A}_{l_2}((\bar{\mathbf{R}}^{(k+1)}\bar{\mathbf{u}})^H\widehat{\mathbf{W}}) - w_1 w_2 \mathbf{B}(l_2,:)\| \leq w_1 w_2 \gamma$
4. **end for**
5. $\mathbf{L} = \frac{1}{w_1}\widehat{\mathbf{Q}}\bar{\mathbf{L}}$
6. $\mathbf{R} = \frac{1}{w_2}\widehat{\mathbf{W}}\bar{\mathbf{R}}$

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

In Algorithm 1, **Line 2** corresponds to solving for each row of the low-rank factor $\bar{\mathbf{R}}^{(k+1)}$ at the $(k+1)^{th}$ iteration using the estimate of low-rank factor $\bar{\mathbf{L}}^{(k)}$ from the $(k)^{th}$ iteration. Similarly, **Line 3** corresponds to solving for each row of the low-rank factor $\bar{\mathbf{L}}^{(k+1)}$ at the $(k+1)^{th}$ iteration using the estimated low-rank factor $\bar{\mathbf{R}}^{(k+1)}$. Finally, **Lines 5** and **6** correspond to retrieving the low-rank factors $\mathbf{L}$ and $\mathbf{R}$ from $\bar{\mathbf{L}}$ and $\bar{\mathbf{R}}$, respectively. We compare our modified methodology to the original formulation presented in our reference(Lopez, Kumar, and Herrmann 2015). Unlike the original work, our approach only requires additional matrix multiplications with $\widehat{Q}$ and $\widehat{W}$. This modification introduces a minor additional cost of approximately $\mathcal{O}(r \cdot \max(m,n))$floating-point operations, which is small compared to the numerical complexity of the original approach.

The numbers in Table 4.1 demonstrate the improvement in runtime of the modified weighted formulation (Equation 4.8). By parallelizing our approach (Algorithm 1), we achieve significant further improvement in runtime. For this test, we use a frequency slice of dimension 8241×8241 with 90% missing receivers. By working with eight parallel

Table 4.1: Runtime comparison of the original weighted method (Aravkin, Kumar, Mansour, Recht, and Herrmann 2014), modified weighted (Equation 4.8), and the parallel weighted method (Algorithm 1) terminating at a same data residual.

Method	S/R(dB)	Time(seconds)
Original Weighted	16.28	37251
Modified Weighted	16.56	15128
Parallel Weighted	16.59	954

workers (two threads each) in the Cloud, we are able to achieve a significantly faster ($39\times$) runtime compared to the original weighted method with a signal to noise ratio that is very close to results obtained with the original formulation.

In summary, consistent with earlier work (Eftekhari, Yang, and Wakin 2018) we found that the choice of weights determines the accuracy of the wavefield recovery. In our experience, adding weighting matrices has little to no effect on convergence of the residuals. Depending on the accuracy of the prior subspace information, there exists in principle an optimal choice for the weights that will give the best possible recovery. In practice, we do not have access to the ground truth to calculate these optimal weights. However, we find that sub-optimal weights still provide improvements in the output quality compared to when no prior information in the form of weight matrices is used. In the next section, we will discuss a practical approach how to find sub-optimal weights.

4.5 Case studies

We proceed to perform a series of experiments to assess the accuracy of the proposed weighted wavefield reconstruction methodology. It is important to note that although we have access to the ground truth fully sampled data, we do not utilize this information for selecting the weights. Instead, we solely rely on the ground truth data to evaluate the accuracy through visual inspection and S/R's (Signal to noise ratio). Our experimental examples encompass the seismic line from the Gulf of Suez, which was previously discussed, as well as a complex full-azimuth synthetic 3D dataset.

4.5.1 Gulf of Suez field data: 2D example

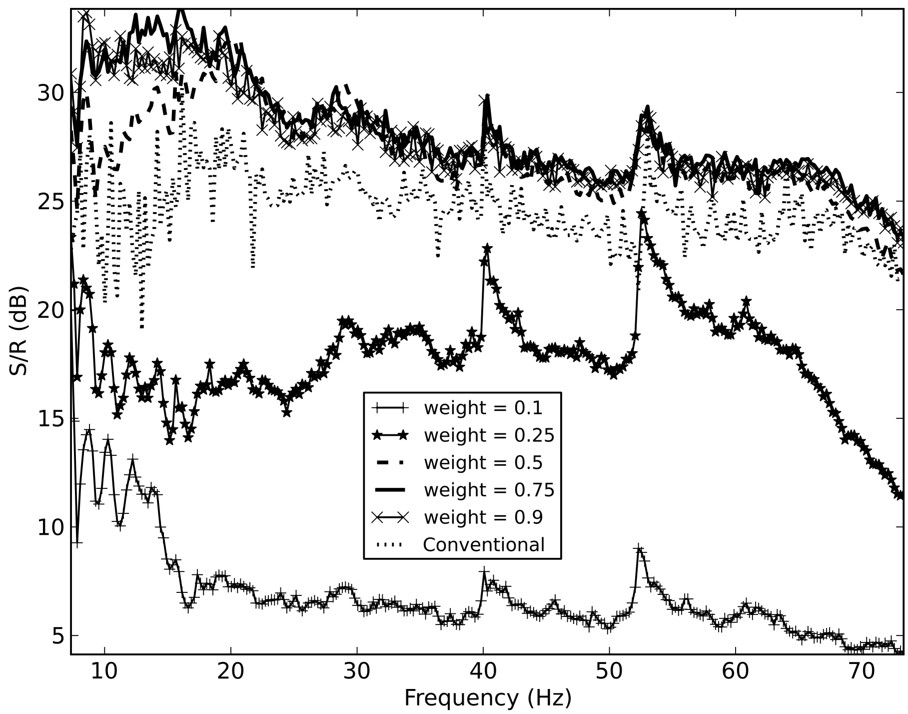
To access the performance of our recursively weighted wavefield recovery method on field data, we revisit the 2D line from the Gulf of Suez. The fully sampled split-spread dataset comprises 1024 time samples, acquired using 355 sources and 355 receivers, with a time sampling interval of $0.004\,\mathrm{s}$. The source-receiver spacing is $25\,\mathrm{m}$. To evaluate our algorithm, we reconstruct this 2D line using randomly subsampled traces obtained by removing 75% of the sources through optimal jittered subsampling (Herrmann and Hennenfent 2008).

We evaluate the performance of recursively weighted matrix factorization by comparing wavefield recovery with and without weighting across different angular frequencies. To mitigate the impact of noise at low frequencies, we start the recovery at $7.0\,\mathrm{Hz}$. Given the relatively small problem, we reconstruct the frequency slices by performing 150 iterations of the SPG-LR algorithm (Berg and Friedlander 2009; Aravkin, Kumar, Mansour, Recht, and Herrmann 2014) for each frequency to reconstruct the frequency slices. The results of wavefield reconstructions with and without weights are compared, and the findings are summarized in Figure 4.11. Based on these results, it is evident that including weights in the reconstruction process yields clear benefits for frequencies above 17Hz compared to reconstructions performed without weighting using the same number of iterations. We chose $w_{1,2} = 0.75$ as values for the weights. This choice for the weights is informed by tests we carried out for different weight parameters across all frequencies. The results of this exercise are summarized in Figure 4.10, which includes plots for the S/R derived from the residuals (Figure 4.10a) and from the misfit between the recovered and fully sampled data (Figure 4.10b). In practice, we only have access to the residuals, which are given by the difference between the incomplete observed data and reconstructed data at the observed locations. The different plots in Figure 4.10b suggest that weights between $w_{1,2} = 0.50$ and $w_{1,2} = 0.75$ give the best S/R across the different frequencies. While the S/R for the residual and misfit with respect to the fully sampled data differ, the residuals provide a practial way to select the weights. Based on the information in Figure 4.10a, the choice of $w_{1,2} = 0.75$ is reasonable since it leads to the best S/R. Notice that for increasing weights, the S/R increase to decrease again for the large values. This behavior can be understood because for small weights, the solution will be dominated by the factorization at the lower frequency and not by the data at the current frequency. As the weights increase, the wavefield recoveries are informed by a combination of observed data and prior information. Since the prior information decreases for increasing weights, we reach, as expected, a point where the results deteriorate again.

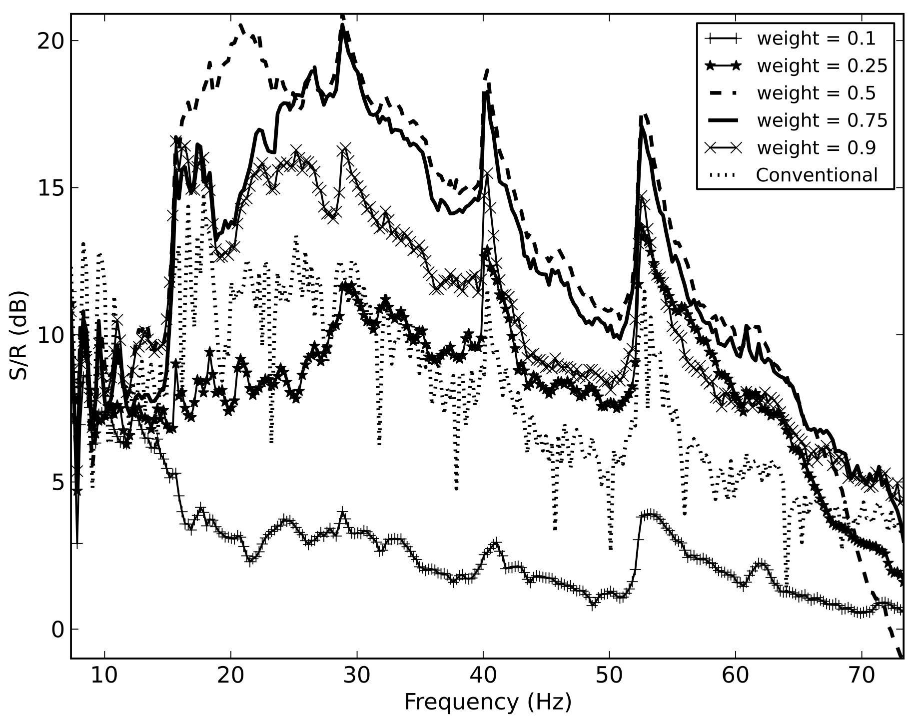
Figure 4.12c and Figure 4.13c display the shallow and deeper parts of a reconstructed common receiver gather obtained from the conventional method, resulting in an $S/R = 6.99\,\mathrm{dB}$. In the data residual plots (Figure 4.12d and Figure 4.13d), we observe signal leakage and noise, due to reconstruction artifact in both shallow and deeper parts. Signal leakage refers to the presence of coherent events in the data residual plot, indicating incomplete data reconstruction. Additionally, there is noticeable signal leakage in the difference plot at far offsets. Far offset data is crucial for FWI (Full Waveform Inversion) purposes since it contains turning waves. In contrast, we observe better reconstructed data in the common receiver gather (Figure 4.12e and Figure 4.13e) extracted from reconstructed data using the recursively weighted approach, exhibits improved quality with an enhanced S/R of $11.91\,\mathrm{dB}$. The corresponding data residual plot (Figure 4.12f and Figure 4.13f) demonstrates reduced signal leakage compared to its conventional counterpart. Even at far offsets, the reconstructed signal is significantly improved.

4.5.2 Synthetic Compass model data: 3D example

In 2D seismic surveys, receivers only measure wavefields within the vertical plane along sources and receivers, limiting the capture of reflections outside of this plane. This omission of out-of-plane scattering ultimately impacts the quality of the subsurface image, particularly in regions with significant lateral heterogeneity. To address this limitation and account for 3D effects, modern seismic exploration surveys employ 3D data acquisition,



(a)



(b)

Figure 4.10: S/R comparison of recursively weighted method with different weights for all frequencies with 75% missing sources. *(a)* S/R of the residuals comparing wavefield recovery and observed data yielded by the recursively weighted method for different weights across all frequencies, *(b)* S/R comparison between recovery and ground truth fully sampled data. As expected other than for the lower frequencies slices, we achieve significant improvements in S/R (Figure 4.11) with our recursively weighted approach when the weights are in a regime where they add prior information without dominating the recovery. The latter occurs for small weights.

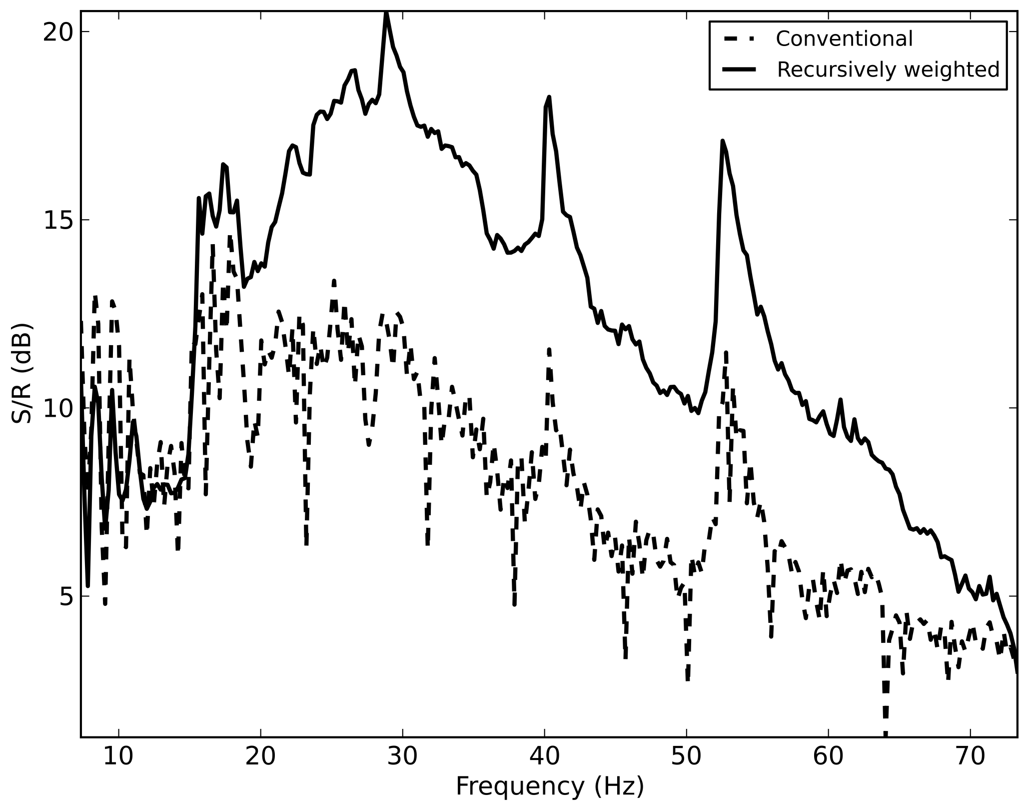


Figure 4.11: *(a)* S/R comparison of conventional (dashed black line) and recursively weighted method (solid black line) with ($w_{1,2} = 0.75$) for all frequencies

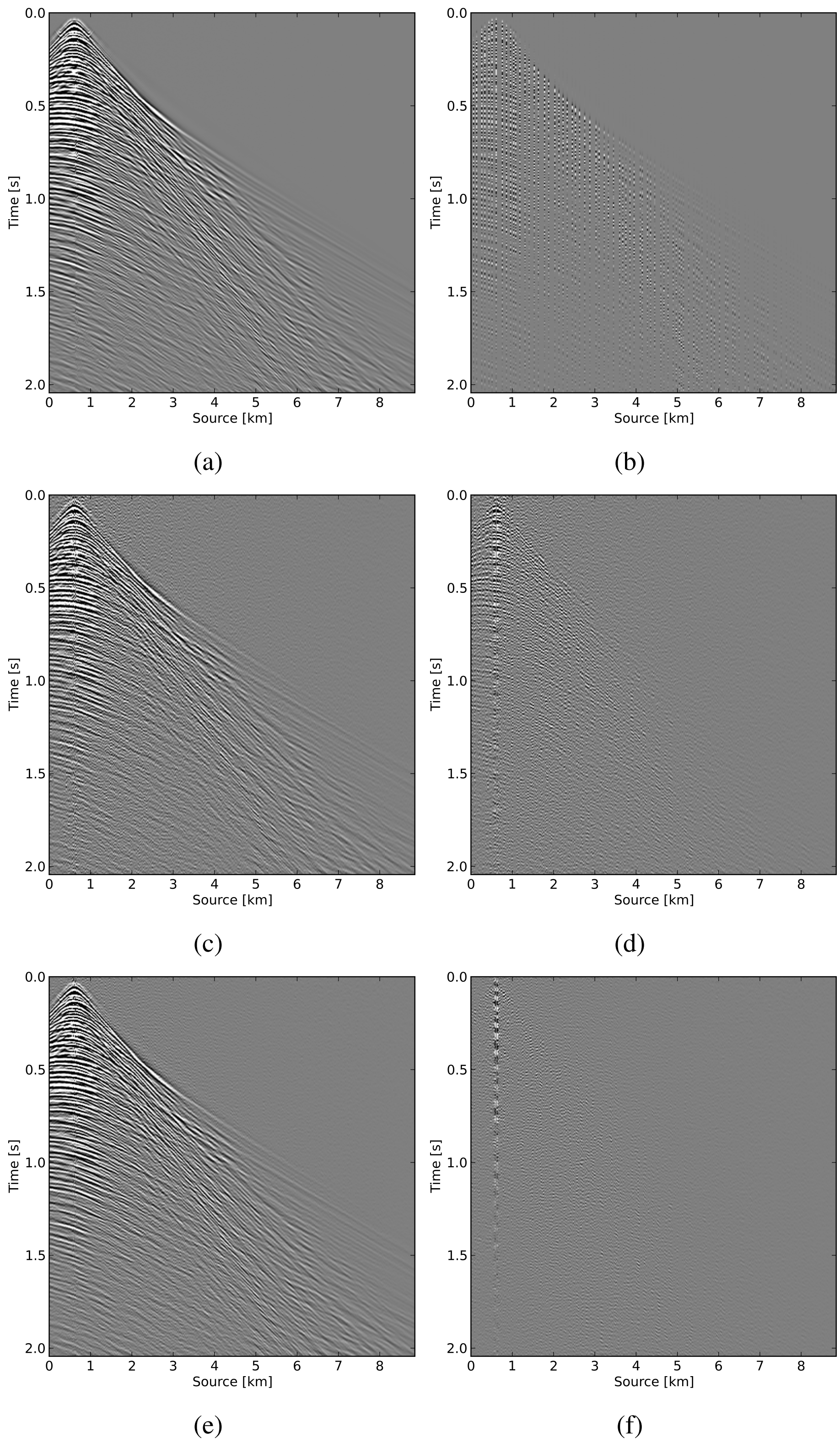


Figure 4.12: Wavefield reconstruction in common receiver gather domain in the shallow part. *(a)* True data, *(b)* Observed data with 75% missing sources. *(c)* Reconstructed data using the conventional method with $S/R = 6.99\,\mathrm{dB}$ and *(d)* corresponding difference with respect to the true data. *(e)* Reconstructed data using the recursively weighted method with $S/R = 11.91\,\mathrm{dB}$ and *(f)* corresponding difference with respect to the true data.

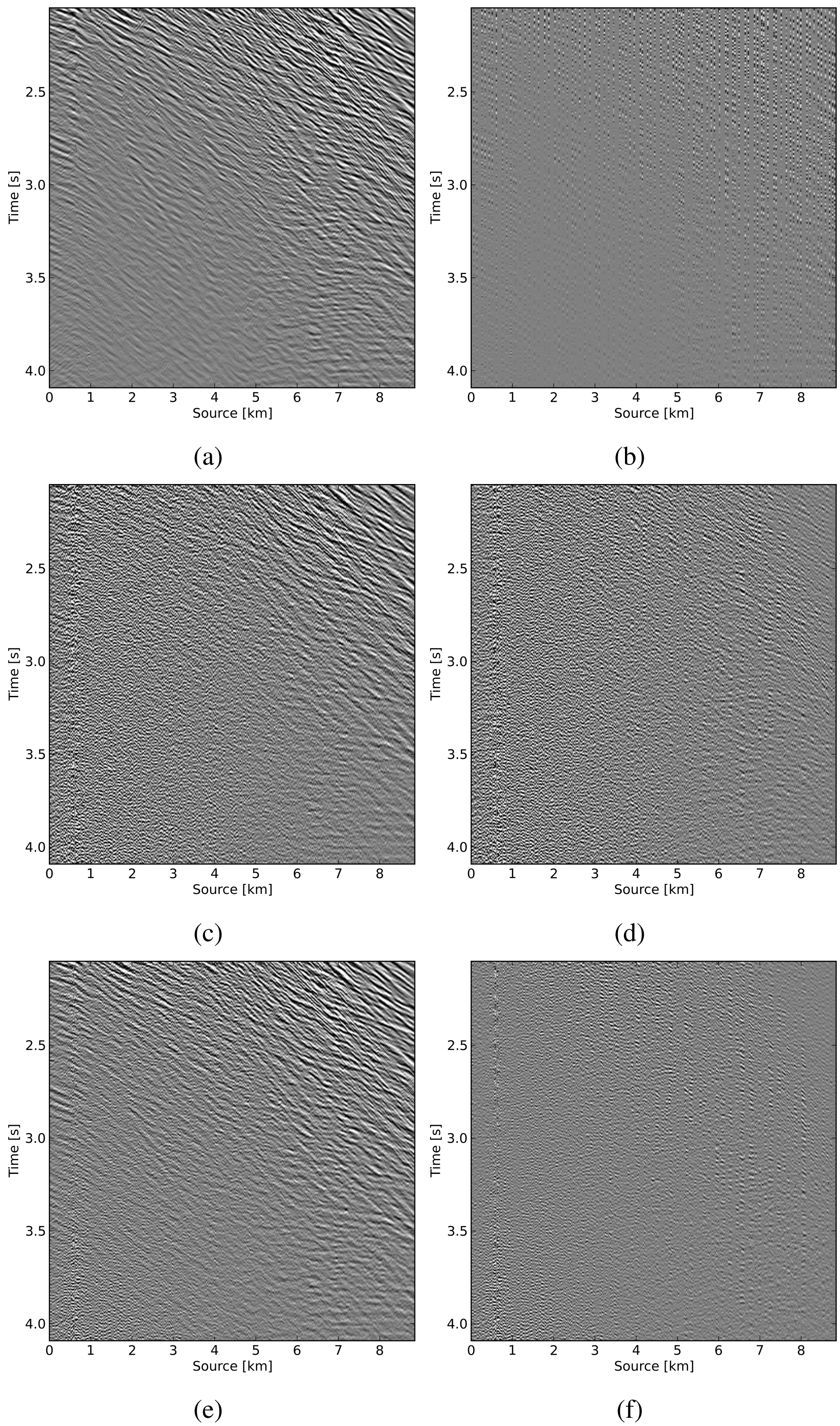


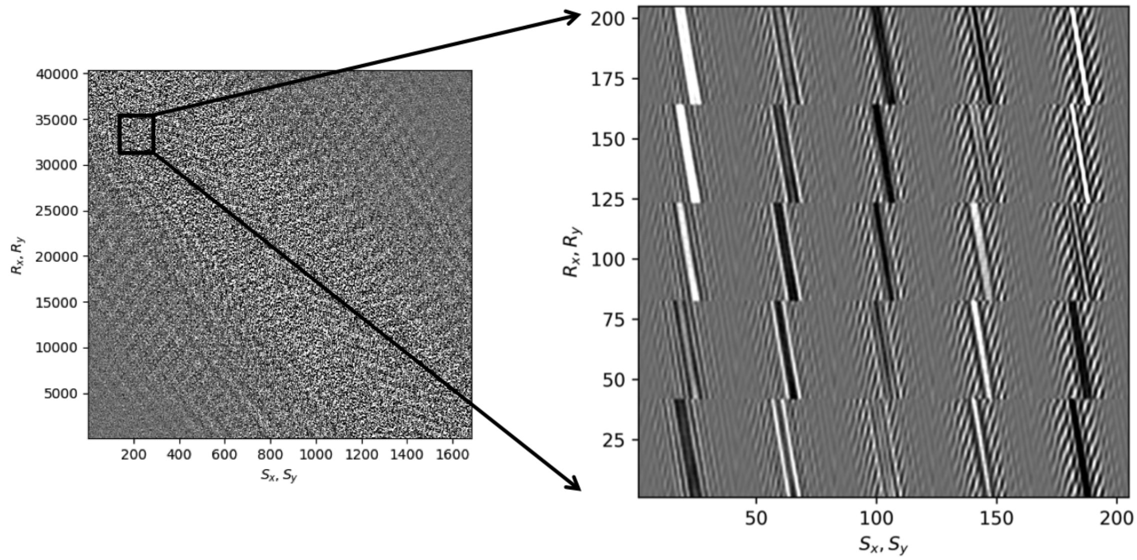
Figure 4.13: Wavefield reconstruction in common receiver gather domain in the deeper part. *(a)* True data, *(b)* Observed data with 75% missing sources. *(c)* Reconstructed data using the conventional method and *(d)* corresponding difference with respect to the true data. *(e)* Reconstructed data using the recursively weighted method and *(f)* corresponding difference with respect to the true data.

where sources and receivers are distributed across the surface rather than confined to a single line. In order to assess the performance of our recursively weighted low-rank matrix factorization methodology in this more challenging 3D setting, we utilize synthetic 3D data generated on the Compass model, which incorporates velocity kickbacks, strong reflectors, and small wavelength details constrained by real well-log data from the North Sea. Due to the complexity of this dataset, resembling marine acquisition with a towed array, we face similar challenges in wavefield reconstruction as we would encounter when dealing with real 3D field data. The authors Da Silva and Herrmann 2015 also employed this 3D dataset to evaluate their tensor-based wavefield reconstruction algorithm based on the Hierarchical Tucker decompositions.

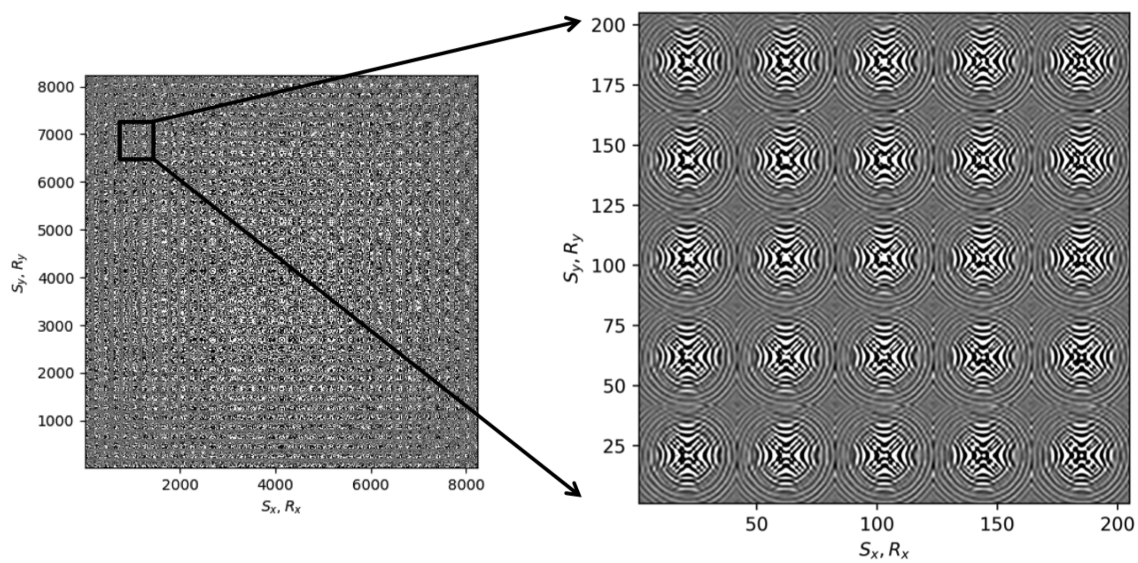
For this experiment, we use a subset of the complete data volume, consisting of $501 \times 201 \times 201 \times 41 \times 41$ gridpoints—i.e., $n_t \times n_{rx} \times n_{ry} \times n_{sx} \times n_{sy}$ along the time, receiver x, receiver y, source x, and source y directions. Here, n_t is the number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

Before proceeding, let us first briefly discuss the organization of the data for wavefield reconstructions. Although we could transform the data into the midpoint-offset domain, as done in the 2D case, we follow Da Silva and Herrmann 2015 and Demanet 2006. We exploit the fact that monochromatic 3D frequency slices rearranged along the x and y-coordinates for sources and receivers can be effectively approximated using a low-rank factorization. In this rearrangement the data is structured as a matrix with S_x, R_x and S_y, R_y coupled along the columns and rows respectively, unfolding along the corresponding coordinate directions. Here $S_{x,y}$ and $R_{x,y}$ are the source and receiver coordinates along the x and y directions. After rearrangement in this non-canonical form, the frequency slices are low-rank while data with randomly missing receivers is not (juxtapose $10.0\,\mathrm{Hz}$ frequency slices in Figure 4.15a and Figure 4.15b and the singular value plots in Figure 4.16a and Figure 4.16b). We choose $10\,\mathrm{Hz}$ frequency slice as the changes in the rate of decay of singular values upon sampling at lower frequencies is more prominent in comparison to changes we observe at higher frequencies. This frequency choice allows us to better demonstrate the reasoning behind choosing S_x, R_x and S_y, R_y domain for reconstruction. In the canonical organization, missing receivers leads to missing rows and this decreases the rank (cf. solid lines in Figure 4.16) in the non-canonical rearrangements the rank increases (cf. dashed lined in Figure 4.16). The sudden decrease in the singular values in the canonical arrangement is a direct result of removing complete rows or columns, which reduces the rank of the data. By examining the behavior of the singular values before and after receiver removal, it becomes evident that the straightforward rearrangement of the data in the non-canonical organization can serve as the transform domain for data recovery through weighted low-rank factorization.

As before, we now proceed with the full-azimuth 3D wavefield reconstruction for each frequency slice using our proposed recursively weighted low-rank matrix factorization approach. Since this is a relatively large problem, we utilize the parallel framework presented in the previous section (Algorithm 1) with 4 alternations and 40 iterations of SPG-ℓ_2 per frequency slice. These parameter values were chosen based on our observation of improved signal continuity and reduced noise in the reconstructed data. In addition to setting the number of alternations, i.e. switching between Equation 4.15 and Equation 4.16, the algorithm needs us to specify the rank of the factorization and the weights. Furthermore, we set the rank of the factorization to $r = 228$ and the weights to $w_{1,2} = 0.75$ after conducting tests with different values. These choices strike a good balance between data quality (in terms of continuity of events, lesser noise) and computational efficiency. The weight

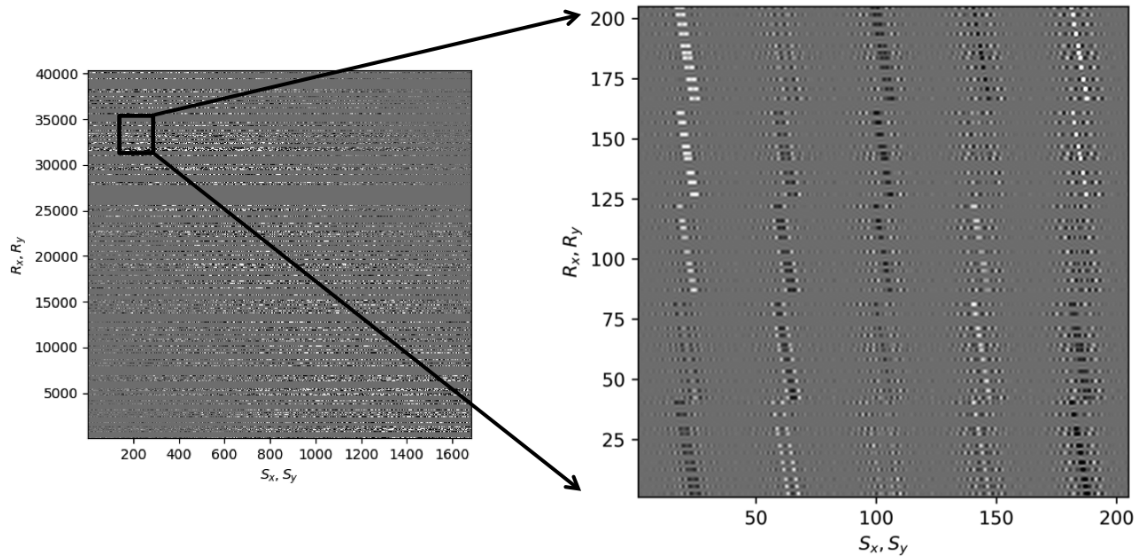


(a)

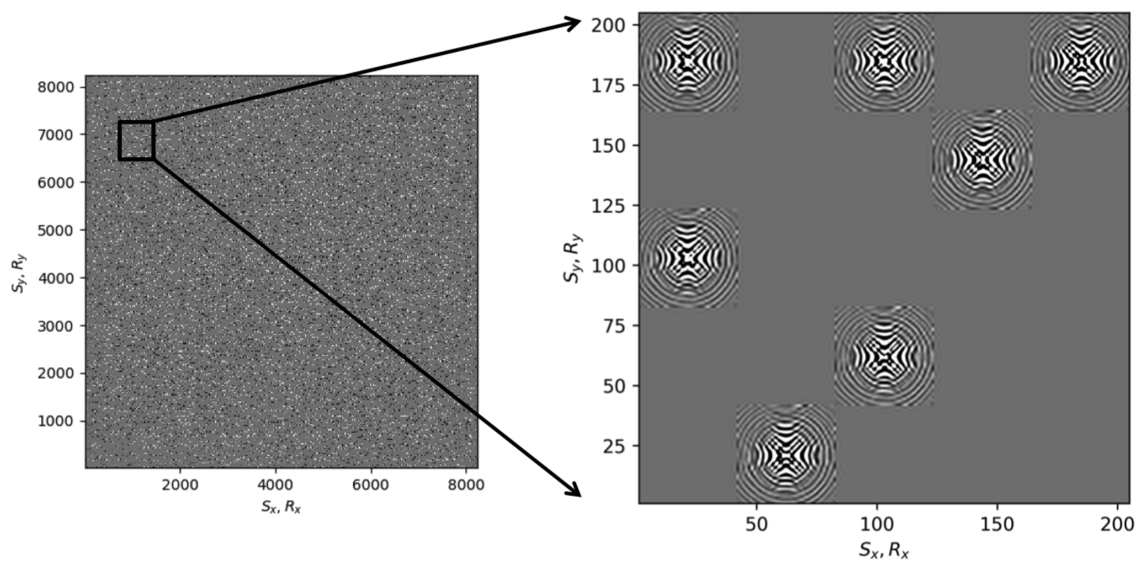


(b)

Figure 4.14: $10.0\,\mathrm{Hz}$ Frequency slice from 3D data: True *(a)* in S_x, S_y domain and *(b)* in S_x, R_x domain. Figures in left column show full data and in right column show data zoomed in the small black box.

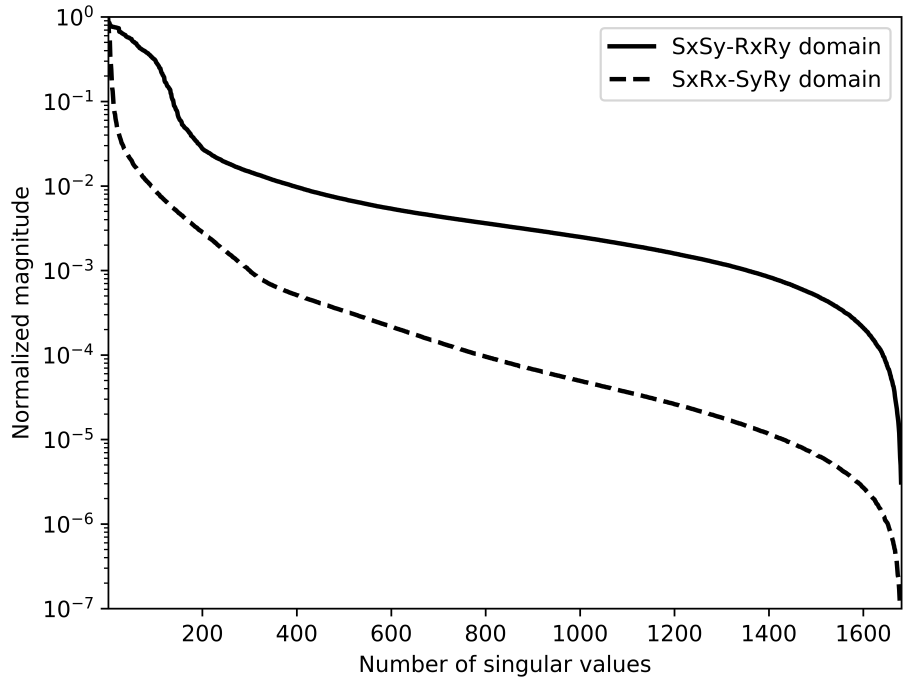


(a)

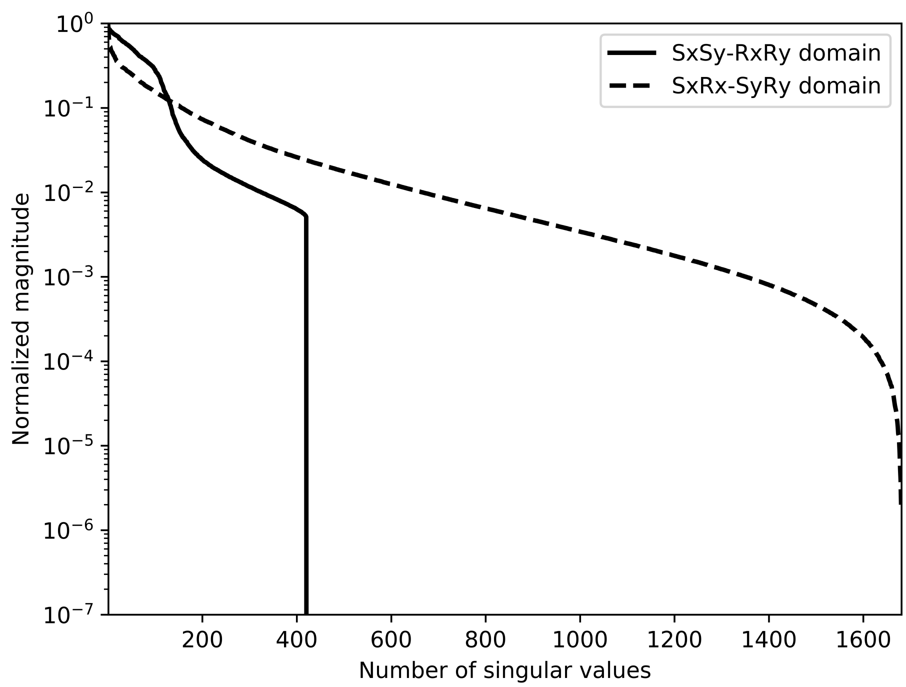


(b)

Figure 4.15: $10.0\,\mathrm{Hz}$ Frequency slice from 3D data: Observed data *(a)* in S_x, S_y domain and *(b)* in S_x, R_x domain with 75% missing receivers. Figures in left column show full data and in right column show data zoomed in the small black box.



(a)



(b)

Figure 4.16: Singular values decay comparison for *(a)* fully sampled and *(b)* subsampled data with 75% missing receivers in S_x, S_y domain (solid black line) and S_x, R_x (dashed black line) domain for $10.0\,\mathrm{Hz}$ frequency slice

selection process for the 2D case is followed to determine the weights.

To mitigate noise resulting from simulation artifacts at very low frequencies, we start our recursively weighted from $4.4\,\mathrm{Hz}$. For comparison, we also use conventional matrix completion for wavefield reconstruction. As a point of comparison, we also employ conventional matrix completion for wavefield reconstruction using the same number of alternations, SPG-ℓ_2 iterations and rank value of 228. For visualization purpose, we present results in a common shot gather (Figure 4.17a) extracted from $15\,\mathrm{Hz}$ frequency slice. Here we choose higher frequency of $15\,\mathrm{Hz}$ instead of $10\,\mathrm{Hz}$ to show how the recursively weighted method is able to give better reconstruction at high frequency in comparison to reconstructed data obtained from the conventional method. In Figure 4.17b we show subsampled shot gather with 75% missing receivers. Using the conventional method we get S/R of $17.7\,\mathrm{dB}$ for the reconstructed data at $15.0\,\mathrm{Hz}$ (Figure 4.17c). Whereas, with the recursively weighted method we get improved S/R of $19.9\,\mathrm{dB}$ (Figure 4.17e). We also observe less leakage of signal and less noise in the residual plots for the data reconstructed using recursively weighted method (Figure 4.17f) in comparison to the data reconstructed using the conventional method (Figure 4.17d). From Figure 4.18a, we also observe improvement in theS/R of reconstructed data for all the frequencies with the recursively weighted method (solid black line in Figure 4.18a) in comparison to its conventional counterpart (dashed black line in Figure 4.18a). In Figure 4.19 and Figure 4.20 we also show comparison of the recursively weighted and conventional method in time domain common shot gather at earlier and later arrivals respectively. In Figure 4.19, we also show comparison of a time slice at $1.6\,\mathrm{s}$ extracted from a 3D common shot gather. Figure 4.19a shows two common shot gathers extracted from the true data along x and y directions along with a time-slice on top left corner. Figure 4.19b shows the corresponding observed data with missing receivers. We observe improved reconstruction of signals in the common shot gather (with $S/R = 17.8\,\mathrm{dB}$) reconstructed from recursively weighted method (Figure 4.19e) in comparison to the reconstructed data from the conventional method (Figure 4.19c) with S/R

Table 4.2: Comparison of S/R (in dB) of reconstructed data from 75% and 90% subsampled data using conventional and recursively weighted method.

Subsampling ratio	Conventional	Recursively weighted
75% (15 Hz)	17.67	19.92
75% (30 Hz)	11.63	14.12
90% (15 Hz)	3.71	12.49
90% (30 Hz)	-0.03	6.64

of $15.3\,\mathrm{dB}$. Even in the residual plots we observe less leakage of signal with the recursively weighted method (Figure 4.19f) in comparison to its conventional counterpart (Figure 4.19d). In Figure 4.20a and Figure 4.20b, we show the same common shot gather at later time along x and y directions extracted from true and observed data respectively. We observe noise in the data and corresponding residual (Figure 4.20c and Figure 4.20d) reconstructed from the conventional method. Whereas, we observe better reconstruction and less noise in the data reconstructed (Figure 4.20e and Figure 4.20f) from the recursively weighted method.

4.5.3 BG synthetic 3D data with 90% missing receivers

Next we test the ability of the recursively weighted method with a reduced number of samples. We subsample the BG synthetic 3D data by 90% using jittered subsampling, i.e. we use only 10% of receivers for wavefield reconstruction. We use 4 alternations and 40 inner iterations of SPG-ℓ_2 in each alternation per frequency slice for both conventional and recursively weighted method. We use rank parameter of 228 for all the frequency slices. Like before we arrive at these values by inspecting the quality of reconstructed data based on the continuity of signal and attenuated noise in the reconstructed data. To find the weights, we follow the same procedure as outlined in the 2D example. To avoid noise at lower frequencies we start recursively weighted method from $4.4\,\mathrm{Hz}$. As evident from the S/R plot (Figure 4.18b), we observe improvement in data reconstruction quality across all the frequency slices using the recursively weighted method (solid black line in Figure 4.18b) in comparison to its conventional counterpart (dashed black line in Figure 4.18b). In a com-

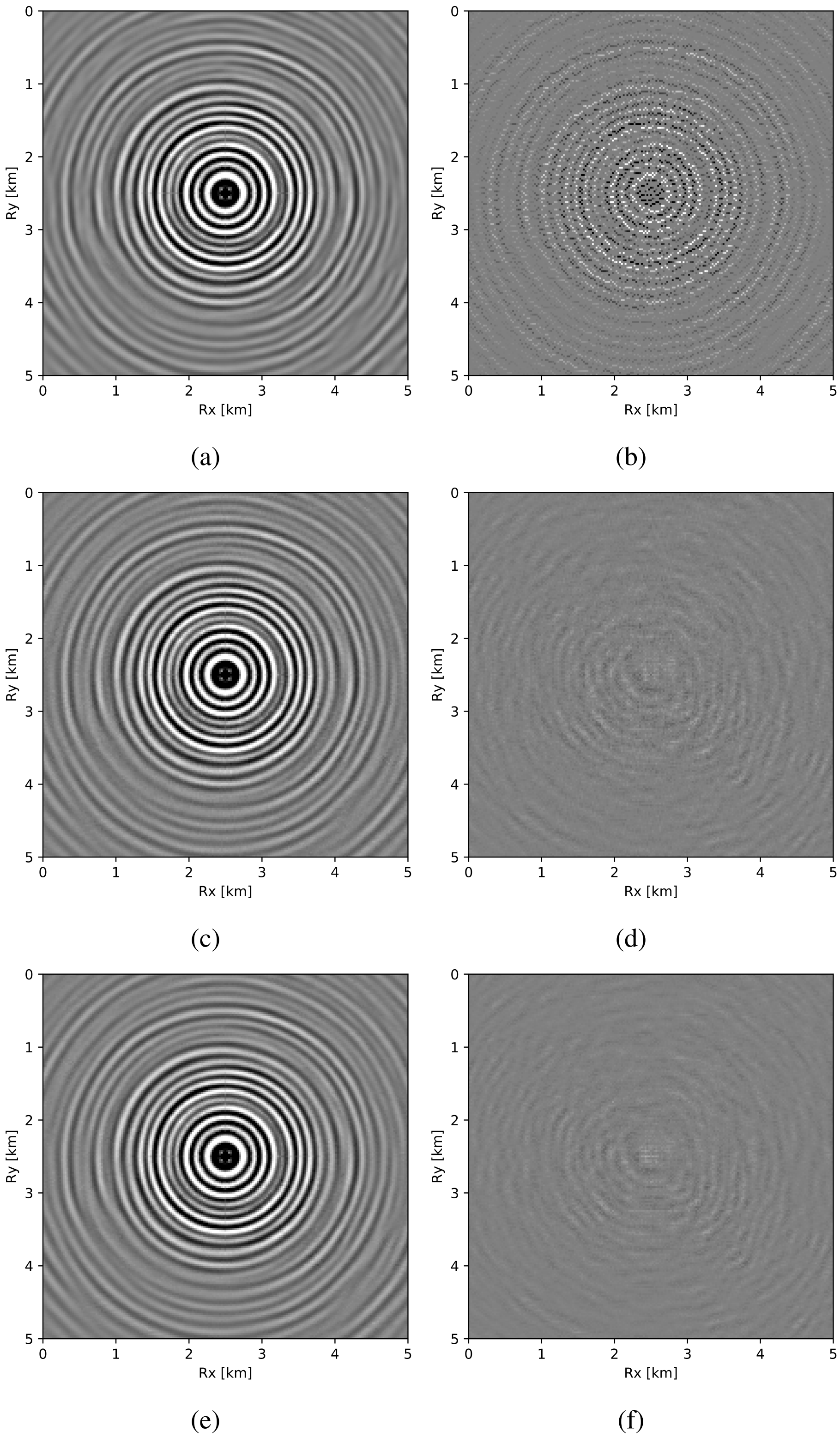
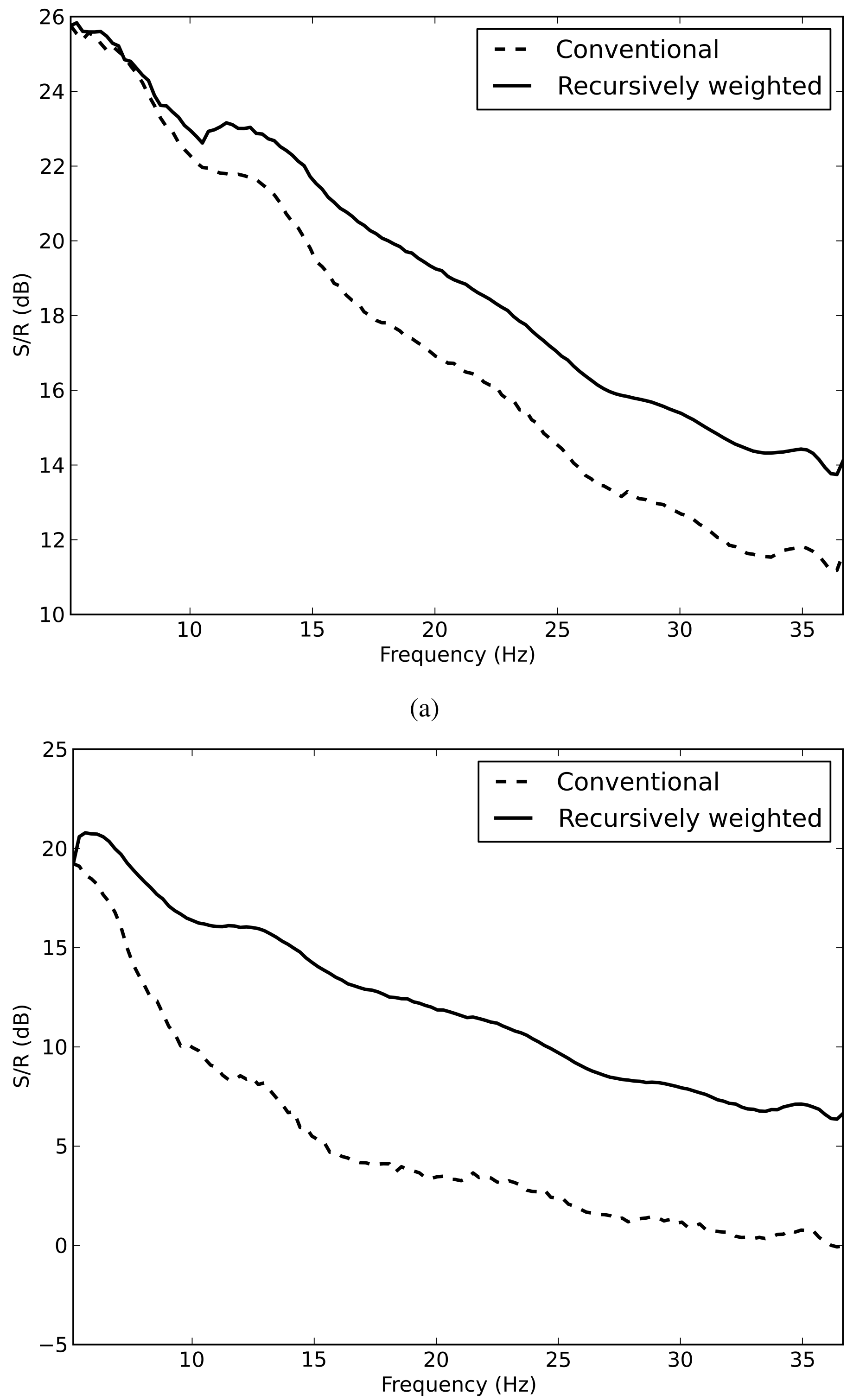


Figure 4.17: Full azimuth wavefield reconstruction comparison for $15.0\,\mathrm{Hz}$ frequency slice in common shot domain. *(a)* True frequency slice. Subsampled frequency slice with *(b)* 75% missing receivers. *(d)* Reconstructed data using conventional method with $S/R = 17.7\,\mathrm{dB}$ and *(e)* corresponding data residual with respect to true data. *(f)* Reconstructed data using recursively weighted method with $S/R = 19.9\,\mathrm{dB}$ and *(g)* corresponding data residual with respect to true data.



(a)

(b)

Figure 4.18: S/R comparison of conventional (dashed black line) and recursively weighted ($w_{1,2} = 0.75$) method (solid black line) for all the frequencies for *(a)* 75% and *(b)* 90% missing receiver scenarios.

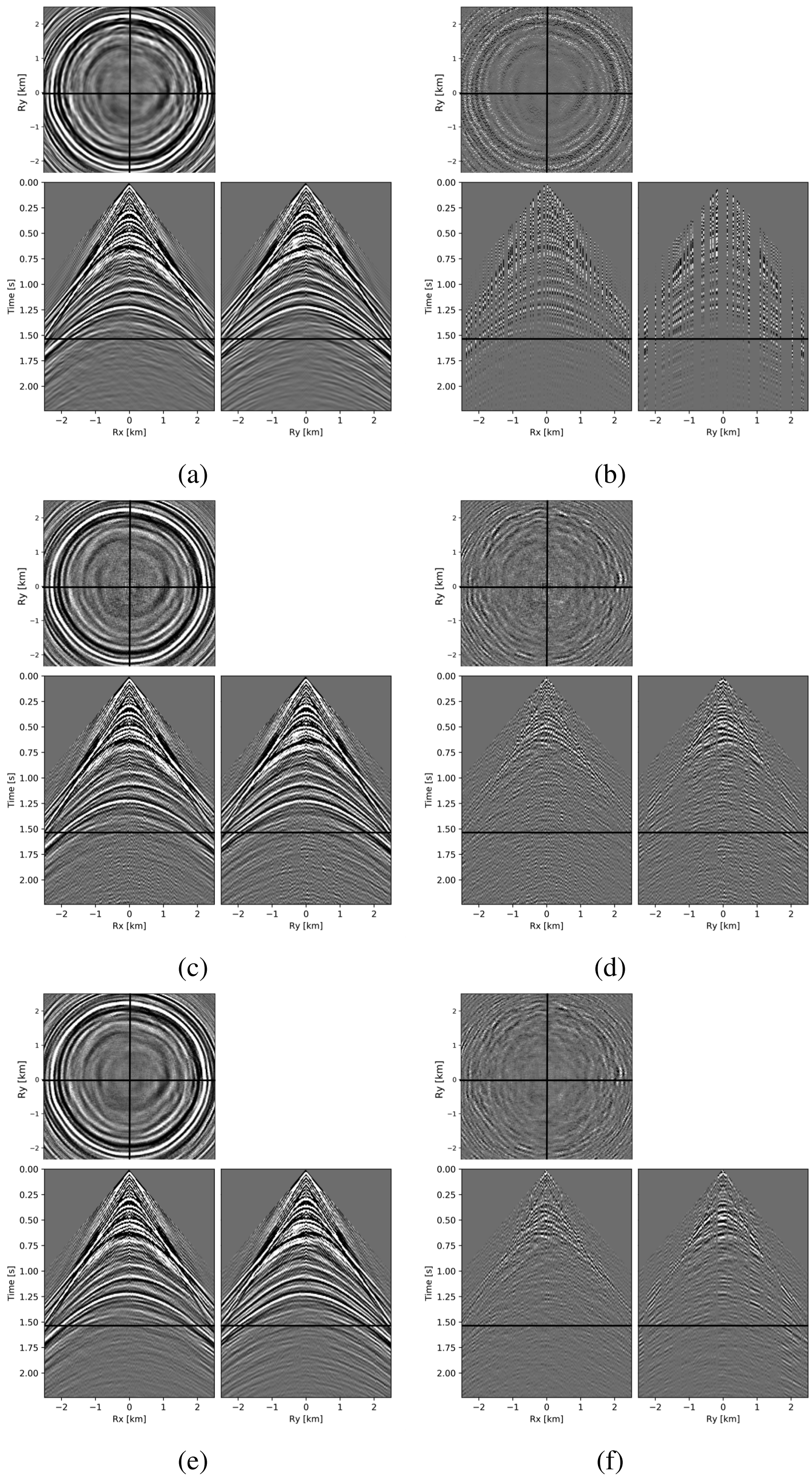


Figure 4.19: Full azimuth wavefield reconstruction in time domain for a common shot gather along with time slice at $1.6\,\mathrm{s}$. *(a)* True data. Subsampled data with *(b)* 75% missing receivers. *(d)* Reconstructed data using the conventional method with $S/R = 15.3\,\mathrm{dB}$ and *(e)* corresponding data residual with respect to the true data. *(f)* Reconstructed data using the recursively weighted method with $S/R = 17.8\,\mathrm{dB}$ and *(g)* corresponding data residual with respect to the true data.

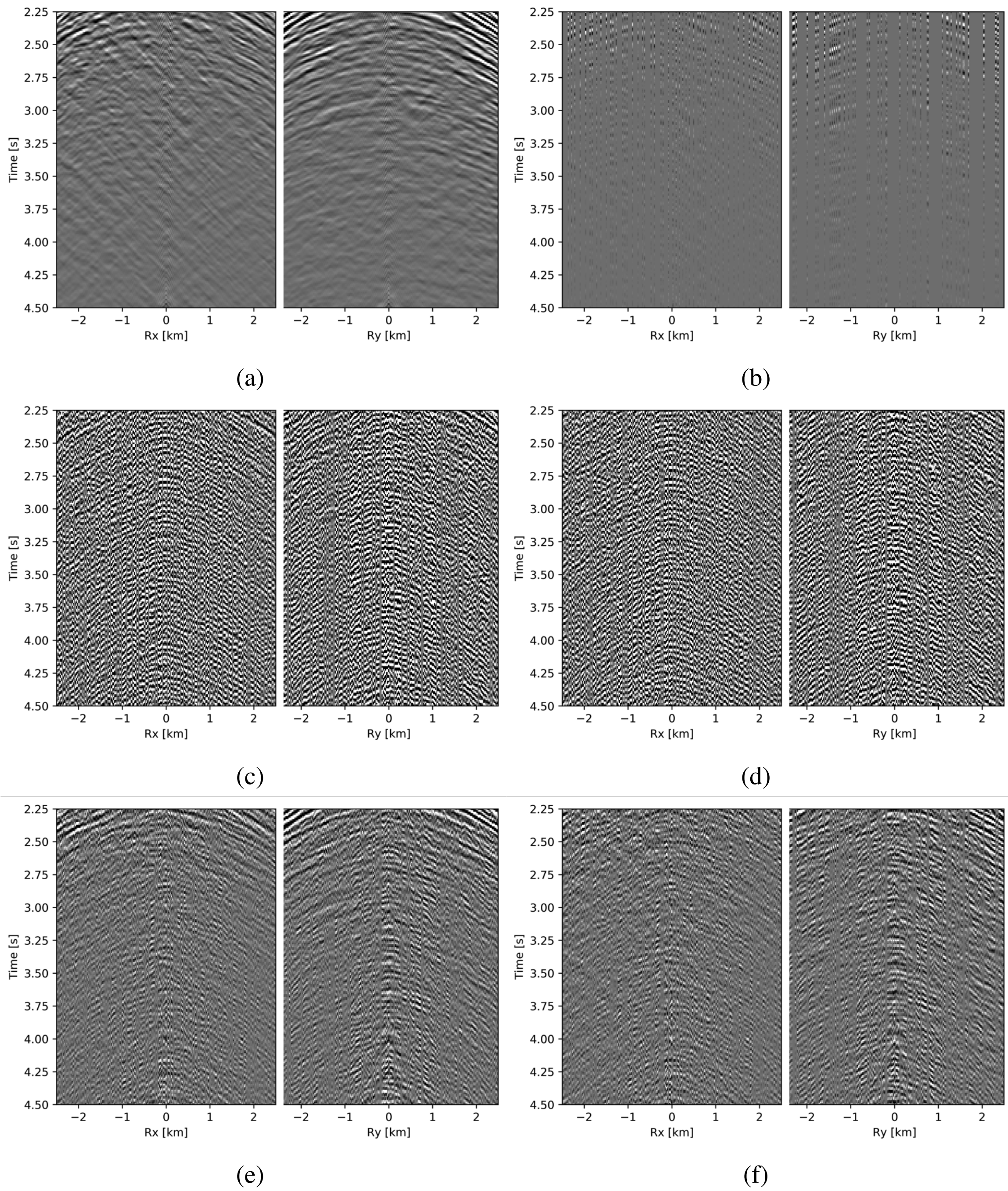


Figure 4.20: Full azimuth wavefield reconstruction in time domain for a common shot gather (deeper section). *(a)* True data. *(b)* Subsampled data with 75% missing receivers. *(c)* Reconstructed data using the conventional method and *(d)* corresponding difference with respect to the true data. *(e)* Reconstructed data using the recursively weighted method and *(f)* corresponding difference with respect to the true data.

mon shot gather extracted from a frequency slice at $15\,\mathrm{Hz}$, we observe better continuity and less noise in the reconstructed wavefield (Figure 4.21d) in comparison to the reconstruction obtained from its conventional counterpart (Figure 4.21b). We observe more leakage of signal in the data residual with the conventional method (Figure 4.21c) in comparison to the data residual obtained from recursively weighted method (Figure 4.21e). In Figure 4.22 we compare data reconstruction in time domain using conventional (Figure 4.22b and Figure 4.22c) and recursively weighted method (Figure 4.22d and Figure 4.22e). We again observe better data reconstruction and reduced data residual with the recursively weighted method in comparison to reconstruction obtained from the conventional method. As an additional measure of comparison, we also plot frequency-wavenumber (f-k) spectrum of shot gathers in Figure 4.23 along both the x and y directions. We observe similarity of f-k spectrum between the data reconstructed from recursively weighted method (Figure 4.23d) and the ground truth (Figure 4.23a). We see improvements compared to the f-k spectrum of data reconstructed with the conventional method (Figure 4.23c).

4.6 Discussion

Our proposed method, which utilizes recursively weighted low-rank matrix completion, surpasses its conventional counterpart in terms of the quality of the reconstructed data, especially at higher frequencies. Conventional low-rank matrix completion exhibits poor performance at higher frequencies due to the increasing complexity of matrices that eventually violate our low-rank assumption. As mentioned earlier, high-quality high-frequency content in the data is crucial for high-resolution imaging of the earth's subsurface and precise inversion of its physical parameters.

Weighted matrix completion was initially introduced by Aravkin, Kumar, Mansour, Recht, and Herrmann 2014 to enhance the seismic data reconstruction quality within the conventional matrix completion framework. In our work, we have harnessed the potential of the weighted method by recursively reconstructing data from low to high frequencies.

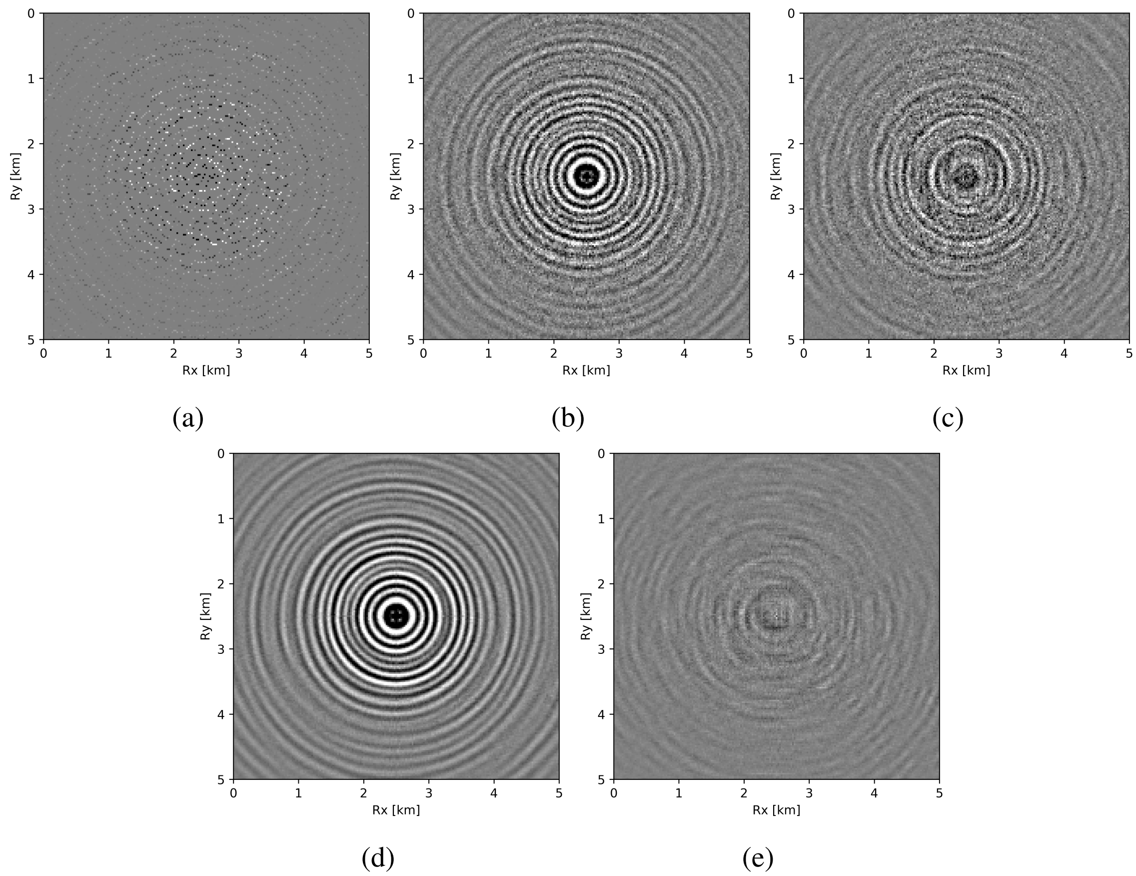


Figure 4.21: Full azimuth wavefield reconstruction comparison for $15.0\,\mathrm{Hz}$ frequency slice in common shot domain. Subsampled frequency slice with *(a)* 90% missing receivers. *(b)* Reconstructed data using conventional method with $S/R = 3.7\,\mathrm{dB}$ and *(b)* corresponding data residual with respect to true data. *(c)* Reconstructed data using recursively weighted method with $S/R = 12.5\,\mathrm{dB}$ and *(d)* corresponding data residual with respect to true data.

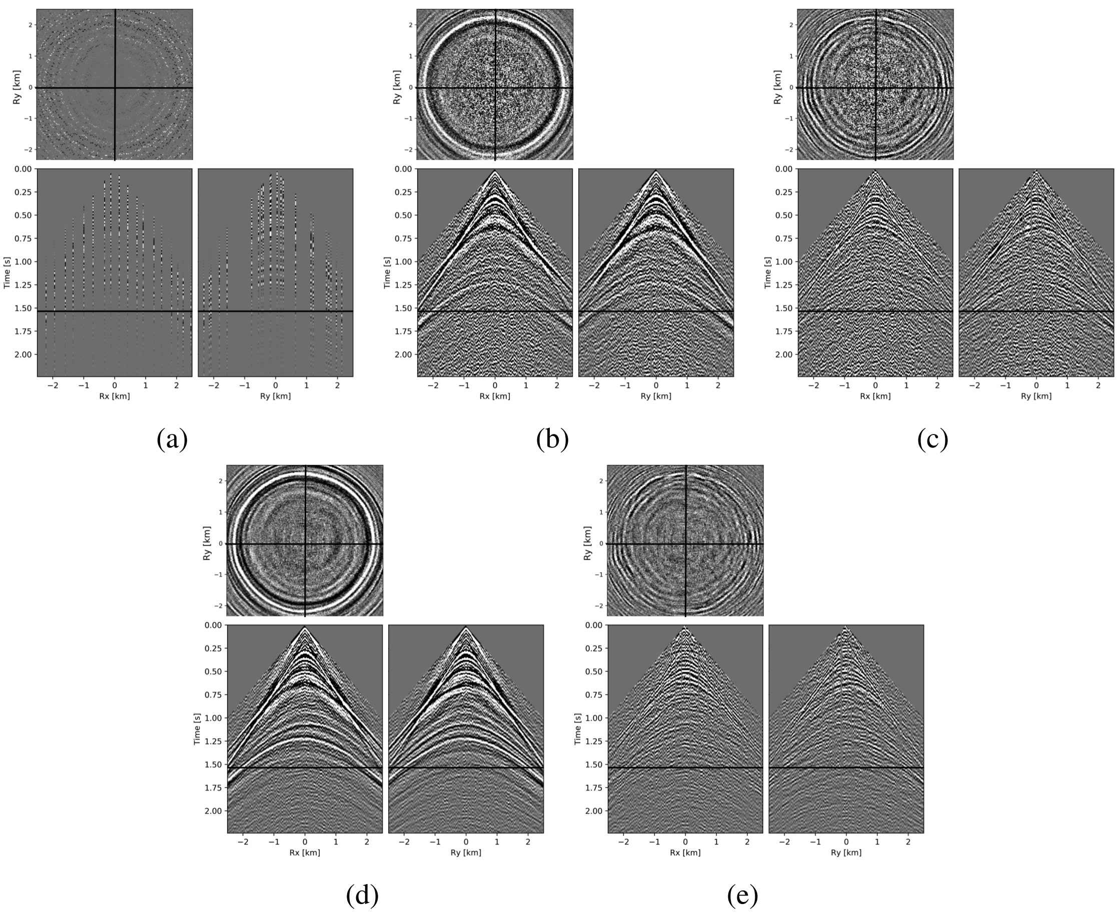


Figure 4.22: Full azimuth wavefield reconstruction in time domain for a common shot gather along with time slice at $1.6\,\mathrm{s}$. Subsampled data with *(a)* 90% missing receivers. *(b)* Reconstructed data using the conventional method with $S/R = 3\,\mathrm{dB}$ and *(c)* corresponding data residual with respect to the true data. *(d)* Reconstructed data using the recursively weighted method with $S/R = 10.2\,\mathrm{dB}$ and *(e)* corresponding data residual with respect to the true data.

Moreover, we have improved the computational efficiency of the original weighted method formulation proposed by Aravkin, Kumar, Mansour, Recht, and Herrmann 2014 by shifting the weights from the objective to the data misfit constraint function.

The success of the recursively weighted method heavily relies on the similarity between adjacent frequency slices and the appropriate choice of weights. The conventional method can be easily parallelized over frequencies, rendering it computationally efficient. However, the interdependence between frequency slices in the recursively weighted method prevents parallelization over frequencies, posing computational challenges, particularly for large-scale 3D datasets. Through the strategies of alternating minimization and decoupling, we have achieved computational efficiency for higher weights in the recursively weighted method. Depending on the availability of computational resources, the recursively weighted method can be efficiently applied to large-scale 3D datasets. Our parallel weighted framework partially leverages the advantages of weighted low-rank matrix factorization as it can only be parallelized for higher weights. Nonetheless, our numerical experiments demonstrate improvements in the quality of the reconstructed data across all frequencies for 3D seismic data generated on a geologically complex velocity model resembling a part of the earth's subsurface. To fully exploit the benefits of the weighted method for large 3D datasets, our future work will focus on extending this methodology to enable parallelism even for smaller weight values.

By directly utilizing the low-rank factors from a preceding frequency slice to calculate the weight matrix, our recursively weighted framework avoids the need for computing SVDs of the complete dataset to determine its row and column subspaces. This SVD-free parallel weighted framework can be applied to industry-scale seismic datasets. With the emergence of cloud computing, abundant computational resources are available. However, the challenge lies in optimizing both the turnaround time and the budget when utilizing these resources. Therefore, our next steps will involve re-engineering the weighted framework to efficiently utilize cloud-based computational resources by incorporating the principles of serverless computing. For example, Witte, Louboutin, Modzelewski, Jones, Selvage, and Herrmann 2019 designed serverless computing architecture to perform large scale 3D seismic imaging.

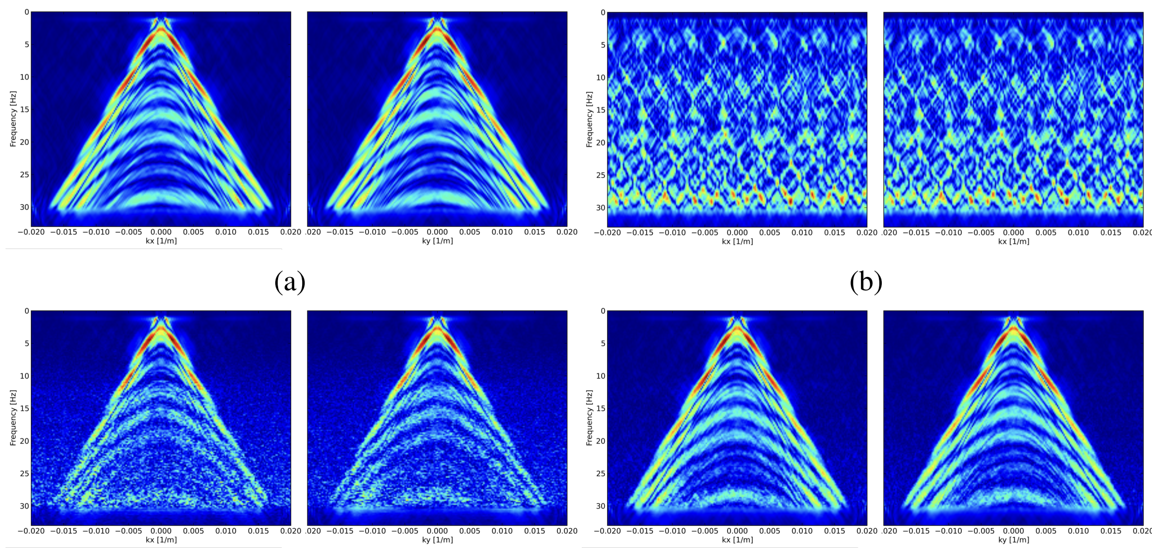
Both datasets used in our experiments feature sources and receivers on a uniform grid, but in reality, this is not always the case. Due to environmental and operational constraints, sources and receivers are often shifted from the uniform grid. If we neglect to account for this shift in our reconstruction framework, the performance of the framework may suffer. By incorporating an additional operator (Lopez, Kumar, Yilmaz, and Herrmann 2016) corresponding to these shifts from the uniform grid, our weighted framework can be applied to field data recorded on a non-uniform grid.

4.7 Conclusions

While wavefield reconstruction based on matrix factorization is successful at low to midrange frequencies, it encounters difficulties at higher frequencies where seismic data is no longer low rank. To address this issue, we leverage the similarities between low-rank factorizations of adjacent monochromatic frequency slices organized in a way that reveals the underlying low-rank structure of fully sampled data, which is often inaccessible for budgetary and physical reasons. These similarities manifest as alignment of the subspaces in which the low-rank factors reside during matrix factorization. By introducing weight matrices that project these factors onto the nearby subspace of the adjacent frequency, we improve the performance of the low-rank matrix factorization, especially when this beneficial feature is recursively applied starting from lower frequencies.

However, transforming this approach into an algorithm capable of scaling to industry-scale wavefield reconstruction problems for full-azimuth data requires several additional crucial steps. Firstly, we need to avoid costly projections onto weighted constraints. We achieve this by transferring the weights to the data misfit, resulting in a computationally faster equivalent formulation. Secondly, although the recursively applied weighting matrices enhance performance at high frequencies, they prevent a row-by-row and column-by-column parallelization of the alternating minimization procedure. This issue is overcome by striking a balance between the importance we assign to information from adjacent frequency slices and our ability to decouple operations, thereby enabling parallelization of the algorithm.

Through carefully selected examples using a 2D field dataset and a full-azimuth 3D dataset, we demonstrate the capability of the proposed algorithm to handle high frequencies. Additionally, we showcase the scalability of the algorithm to 3D problems with a large percentage of missing traces. Based on these results, we argue that the proposed approach could serve as a valuable alternative to transform-based methods, which are constrained to operate on small multidimensional patches.



(a) (b)

(c) (d)

Figure 4.23: f-k spectrum comparison of *(a)* True and *(b)* observed data with 90% missing receivers, reconstructed data using *(c)* conventional method and *(d)* recursively weighted method.

CHAPTER 5

A PRACTICAL WORKFLOW FOR LAND SEISMIC WAVEFIELD RECOVERY WITH WEIGHTED MATRIX FACTORIZATION

5.1 Summary

While wavefield reconstruction through weighted low-rank matrix factorizations has been shown to perform well on marine data, out-of-the-box application of this technology to land data is hampered by ground roll. The presence of these strong surface waves tends to dominate the reconstruction at the expense of the weaker body waves. Because ground roll is slow, it also suffers more from aliasing. To overcome these challenges, we introduce a practical workflow where the ground roll and body wave components are recovered separately and combined. We test the proposed approach blindly on a subset of the 3D SEAM Barrett dataset. With our technique, we recover densely sampled data from 25 percent randomly subsampled receivers. Independent comparisons on a single shot demonstrate significant improvements achievable with the presented workflow.

5.2 Introduction

One of the critical phases in the early stages of oil and gas exploration is seismic data acquisition. Inspired by relatively recent developments encouraged by the field of Compressive Sensing (Candès, Romberg, and Tao 2006), seismic data is increasingly collected randomly along the spatial coordinates to shorten the acquisition time and to reduce cost. While random sampling improves acquisition productivity (Mosher, Li, Morley, Ji, Janiszewski, Olson, and Brewer 2014), it does shift the burden from field acquisition to data processing (Chiu 2019) since fully sampled seismic data is a prerequisite to subsequent steps such as multiple removal and migration.

Wavefield recovery is one of the key steps to reconstruct fully seismic data from subsampled data. Recovery methods based on wavefield reconstruction that exploits data sparsity in different transform domains, such as wavelet (Villasenor, Ergas, and Donoho 1996), Fourier (Sacchi, Ulrych, and Walker 1998), and curvelet (Herrmann and Hennenfent 2008), have been proposed. More recently, several seismic studies have investigated wavefield recovery via low-rank matrix factorizations (Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015), which are relatively simple and computationally cheap. The general idea of these methods is to exploit low-rank structure of fully sampled frequency slices when they are organized in a matrix. Oropeza and Sacchi 2011 and Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015 showed that the presence of noise or missing traces increases the rank of these matrices, and they used this property to recover the fully sampled frequency slices via low-rank matrix factorization.

While the low-rank matrix factorization method has had some success, especially for low to midrange frequencies, it struggles to recover high frequency slices, which require higher ranks because they cannot be accurately approximated by low-rank factorizations. To solve this problem, Aravkin, Kumar, Mansour, Recht, and Herrmann 2014, Eftekhari, Yang, and Wakin 2018, and Zhang, Sharan, and Herrmann 2019 used the wavefield recovery via weighted matrix factorization to reconstruct seismic data by introducing matrix weights defined in terms of factorizations at neighboring frequencies that live in close-by subspaces. By moving the matrix weights from the constraint to the data-misfit term, Zhang, Sharan, Lopez, and Herrmann 2020 proposed a computationally more efficient scheme capable of handling high frequencies.

Even though this weighted approach has had success, there remains the challenge that land seismic data contains ground roll, which because of its strong amplitude and high spatial frequency content is known to (Liu 1999) degrade the wavefield reconstructions based on promoting structure whether this is sparsity or low rank. The reason for this possible degradation is two-fold. First, ground roll corresponds to Rayleigh-type surface waves, which are slow and for this reason often aliased. Second, ground roll has strong amplitudes, which causes the reconstruction to focus on the ground roll at the expense of reconstructing the low-amplitude body waves. While ground roll is typically dominant at the low temporal frequencies, its separation from body waves is complicated by the fact that it is spatially aliased. By reconstructing the wavefield to a fine grid, where the ground roll is no longer aliased, we allow for a separation of ground roll and body waves using $f-k$ filtering (Yilmaz 2001) or Radon domain techniques (Trad, Ulrych, and Sacchi 2003). During this chapter, we present a practical workflow aimed at removing the complications of carrying out wavefield reconstruction on land data dominated by ground roll.

We organize this chapter as follows. First, we discuss the seismic wavefield reconstruction via weighted matrix factorization. Next, we discuss the impact of ground roll. And then, we introduce our proposed practical workflow step by step. We conclude by demonstrating our approach on synthetic 3D data simulated from the Barrett model and show improved recovery quality compared to the conventional workflow.

5.3 Reconstruction with weighted matrix factorizations

In Aravkin, Kumar, Mansour, Recht, and Herrmann 2014, Eftekhari, Yang, and Wakin 2018 and Zhang, Sharan, and Herrmann 2019, the authors proposed a wavefield recovery via weighted matrix factorization. These factorizations are carried out on data organized in monochromatic frequency slices and involve the following optimization problem:

$$\begin{aligned} \min_{\mathbf{X}_i} \quad & \|\mathbf{Q}\mathbf{X}_i\mathbf{W}\|_* \\ \text{subject to} \quad & \|\mathcal{A}(\mathbf{X}_i) - \mathbf{B}_i\|_F \leq \eta. \end{aligned} \tag{5.1}$$

In this expression, the symbol $\|\cdot\|_*$ represents the nuclear norm, given by the sum of the singular values, and $\|\cdot\|_F$ denotes the Frobenius norm, the energy of the matrix entries. The matrix $\mathbf{X}_i$, for $i \in [1, \ldots, N_f]$, represents a fully sampled monochromatic frequency slice at the $i\mathrm{th}$ frequency. N_f corresponds to the number of frequencies, and the matrix

$\mathcal{A}(\cdot)$ represents a mask operator used to subsample the fully sampled frequency slice. The matrix $\mathbf{B}_i$ represents the observed input data with missing traces at the $i\mathrm{th}$ frequency. The misfit tolerance η depends on the noise level in the observed data.

To exploit the fact that seismic data exhibits low-rank behavior in the so-called non-canonical organization (Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015), we matricize the frequency slices in the source-x receiver-x organization—i.e., the to-be-recovered monochromatic data is represented by the matrix $\mathbf{X}_i \in \mathbb{C}^{(N_{sx}\times N_{rx})\times(N_{sy}\times N_{ry})}$ where N_{sx}, N_{sy} are the number of sources along the x and y coordinates, respectively. N_{rx}, N_{ry} are the corresponding numbers of receivers. The $\{\mathbf{Q},\mathbf{W}\} \in \mathbb{C}^{(N_{sx}\times N_{rx})\times(N_{sx}\times N_{rx})}$ are the weighting matrices, which include information on the subspaces of a neighboring factorization as we reconstruct the wavefield from low-to-high frequencies (Zhang, Sharan, and Herrmann 2019). These weighting matrices are given by

$$\mathbf{Q} = w_1\mathbf{U}\mathbf{U}^H + \mathbf{U}^\perp\mathbf{U}^{\perp H} \tag{5.2}$$

and

$$\mathbf{W} = w_2\mathbf{V}\mathbf{V}^H + \mathbf{V}^\perp\mathbf{V}^{\perp H}. \tag{5.3}$$

In these expressions, the symbol ^{H} denotes the Hermitian transpose. The projection matrices $\mathbf{U} \in \mathbb{C}^{(N_{sx}\times N_{rx})\times r}$, $\mathbf{V} \in \mathbb{C}^{(N_{sy}\times N_{ry})\times r}$ contain rank r column and row subspaces that derive from neighboring (lower) frequencies. The matrices $\mathbf{U}^\perp$, $\mathbf{V}^\perp$ are the orthogonal complements of $\mathbf{U}$, $\mathbf{V}$. Because factorizations of neighboring (lower) frequencies share information with the current frequency slice, they can serve as prior information aiding the wavefield recovery. The scalars $w_1 \in (0,1]$ and $w_2 \in (0,1]$ quantify the similarity between prior information and the current to-be-recovered frequency slice. Small values for these scalars indicate that we have more confidence in the prior information.

As shown in Zhang, Sharan, and Herrmann 2019, considerable improvements can be made during the recovery when reliable prior information is available. However, including weighting matrices in the nuclear norm objective function complicates the optimization making the minimization in Equation 5.1 computationally more expensive. To avoid this issue, we follow Zhang, Sharan, Lopez, and Herrmann 2020 and rewrite Equation 5.1 into

$$\begin{aligned} \min_{\tilde{\mathbf{X}}_i} \quad & \|\tilde{\mathbf{X}}_i\|_* \\ \text{subject to} \quad & \|\mathcal{A}(\mathbf{Q}^{-1}\tilde{\mathbf{X}}_i\mathbf{W}^{-1}) - \mathbf{B}_i\|_F \leq \eta. \end{aligned} \tag{5.4}$$

To arrive at this formulation, we replace the optimization variable with $\tilde{\mathbf{X}}_i = \mathbf{Q}\mathbf{X}_i\mathbf{W}$. After solving Equation 5.4, the original solution $\mathbf{X}_i$ can be recovered by $\mathbf{X}_i = \mathbf{Q}^{-1}\tilde{\mathbf{X}}_i\mathbf{W}^{-1}$. Mathematically, Equation 5.1 and Equation 5.4 are equivalent except that the solution of the second formulation is easier to compute by moving the weighting matrices to the data misfit constraint.

To prevent computationally expensive singular value decompositions (SVDs) part of the nuclear norm computations, we write Equation 5.4 in the following factored form:

$$\begin{aligned} \min_{\tilde{\mathbf{L}}_i,\tilde{\mathbf{R}}_i} \quad & \frac{1}{2}\left\| \begin{bmatrix} \tilde{\mathbf{L}}_i \\ \tilde{\mathbf{R}}_i \end{bmatrix} \right\|_F^2 \\ \text{subject to} \quad & \|\mathcal{A}(\mathbf{Q}^{-1}\tilde{\mathbf{L}}_i\tilde{\mathbf{R}}_i^H\mathbf{W}^{-1}) - \mathbf{B}_i\|_F \leq \eta. \end{aligned} \tag{5.5}$$

In this expression, the $\tilde{\mathbf{L}}_i \in \mathbb{C}^{(N_{sx} \times N_{rx}) \times r}$ and $\tilde{\mathbf{R}}_i \in \mathbb{C}^{(N_{sy} \times N_{ry}) \times r}$ represent the low-rank factorization of $\tilde{\mathbf{X}}_i$ with rank $r \ll \min(N_{sx} \times N_{rx}, N_{sy} \times N_{ry})$ (Zhang, Sharan, Lopez, and Herrmann 2020).

While wavefield recovery based on weighted matrix factorization has been applied successfully (see e.g., Zhang, Sharan, and Herrmann 2019 and Zhang, Sharan, Lopez, and Herrmann 2020), its performance is challenged by data that contains strong-amplitude aliased ground roll. Because of its large-amplitude, ground roll dominates the reconstruction at the expense of body waves that are of prime interest. In the next section, we will introduce a practical workflow addressing this challenge.

5.4 Impact of ground roll

Acquisition and processing of land data are often challenged because it is contaminated by strong ground roll. Because ground roll is slow, it is often spatially aliased, complicating subsequent processing efforts to remove this coherent noise component with $f-k$ or Radon filtering (Trad, Ulrych, and Sacchi 2003; Yilmaz 2001). Unfortunately, it is financially unfeasible to decrease the periodic receiver sampling interval to avoid aliasing (Bahia, Papathanasaki, and Sacchi 2020), and we have to resort to alternative randomized acquisition methodologies (Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015; Mosher, Li, Morley, Ji, Janiszewski, Olson, and Brewer 2014) that are in principle conducive to in silico unaliased wavefield reconstruction. While this has proven to work, the presence of strong-amplitude ground roll complicates wavefield reconstruction.

To investigate this issue, we collaborate with Klaas Koster from Occidental to conduct a blind study where we were provided with a subset consisting of 21 source lines, extracted from the synthetic 3D SEAM Barrett dataset (Tan, Li, Jarrah, Lee, Holt, Coevering, and Koster 2019; Van De Coevering, Koster, and Holt 2019). This dataset is designed to benchmark land data processing. As part of this blind study, with the acquisition geometry plotted in Figure 5.1, we receive 3D shot records that are randomly subsampled along the receivers. The 8×8km receiver aperture is moving with the source location, which means that between neighboring shots randomly sampled receivers are mostly shared while some drop-off and others are added (cf. red and blue rectangle in Figure 5.1). Approximately 75 percent of receiver positions are missing from the regular densely sampled periodic grid of 12.5m, yielding an effective average sample interval of 50m, which is well below Nyquist. The data consists of 667 time samples with a sample interval of $0.006\,\mathrm{s}$. The shots are sampled periodically with a sample interval of 25m in the shot-line direction and 100m in the perpendicular direction.

To illustrate, the effects of the strong ground roll on our factorized low-rank wavefield reconstruction scheme, we recover a patch of 4×4 shots with on the dense periodic receiver grid of 641 receivers in each direction sampled at 12.5m. This recovery corresponds to solving a total of 384 monochromatic matrix factorization problems involving data volumes of $667\times641\times641\times4\times4$. These volumes are factored into the product of two $(641\times4)\times340$ matrices where 340 is the rank r. After reconstruction, the results of which are plotted in Figure 5.2 for a single shot record, we recover shot gathers sampled at 12.5m from receivers collected at random at an average receiver spacing of 50m. While we are able to recover this shot record, strong noisy artifacts remain especially at the long offsets. In addition, important reflection and diffraction information is missing and the ground roll is not well recovered making this wavefield recovery unsuitable for subsequent processing.

5.5 Proposed practical workflow

To mitigate the effects of ground roll, we propose the reconstruction of the body and surface (ground roll) waves separately. In this approach, outlined in Figure 5.3, we use the fact that ground roll is slow and relatively easily separable by applying a linear shift to the data. Below, we describe the different steps outlined in the dashed boxes in Figure 5.3.

5.5.1 Ground roll estimation

Because the speed of the ground roll is slower than that of the body wave, it is steep in the 'travel-time' plot and therefore at least in an approximate sense, separable from the body waves. This allows us to devise a separate reconstruction scheme to recover the ground roll before adding it back to the reconstruction of the body waves. We obtain an estimate of the ground roll by carrying out the following steps: (1) after zero-padding the input data (Figure 5.4a), we apply a linear shift aligning the ground roll (Figure 5.4b); (2) we apply a smooth taper with smooth cutoffs around at $t = 0$s and $t = 1$s designed to extract the ground roll (Figure 5.4c), followed by (3) undoing the linear shift, yielding an estimate of

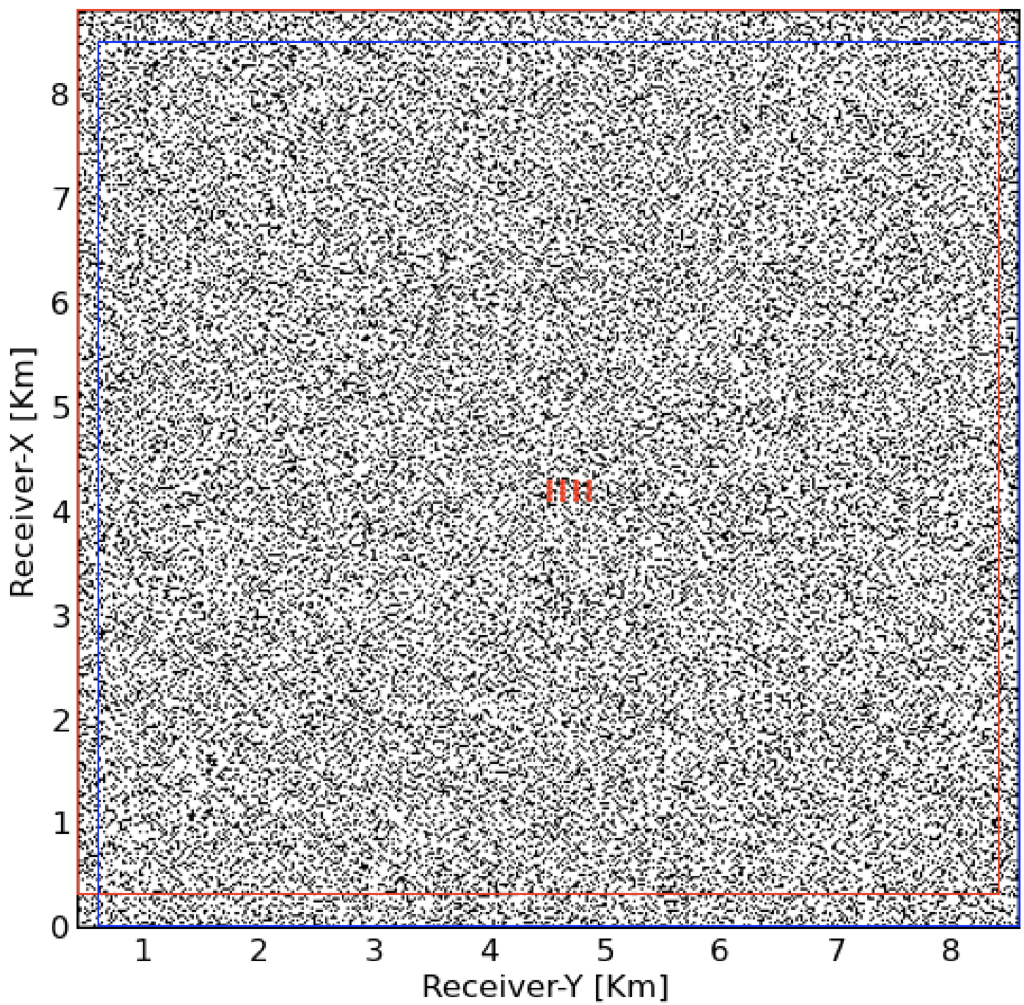
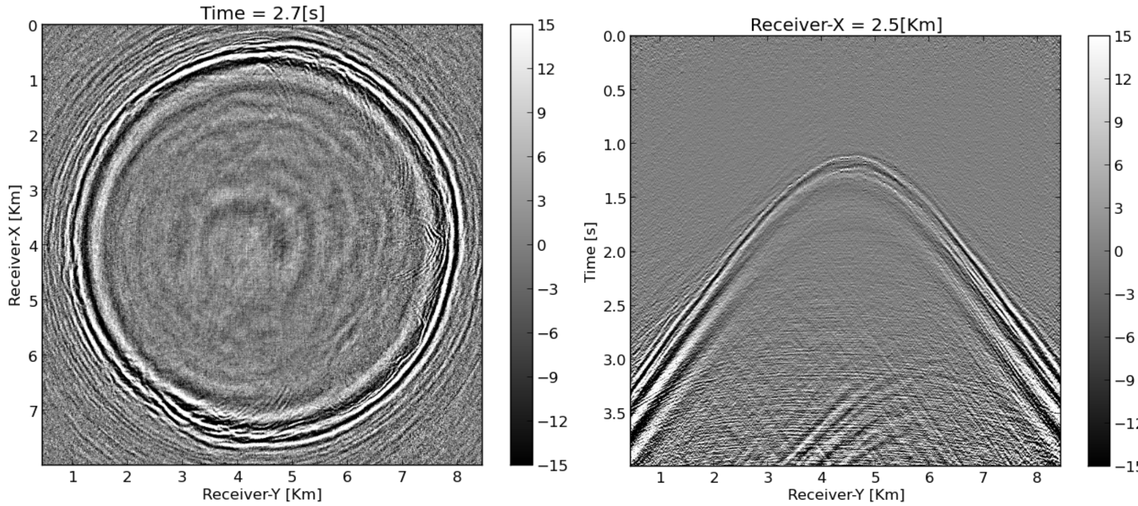


Figure 5.1: Acquisition geometry for the recovered patch. Black ·'s represent receiver locations, and red ·'s in the middle represent the source locations. The red rectangle is the receiver aperture for the top left source and the blue rectangle is the receiver aperture for the bottom right source.



(a)

(b)

Figure 5.2: One common-shot record obtained by factorized wavefield reconstruction. *(a)* Time slice at $2.7\,\mathrm{s}$. *(b)* Shot record along the y direction.

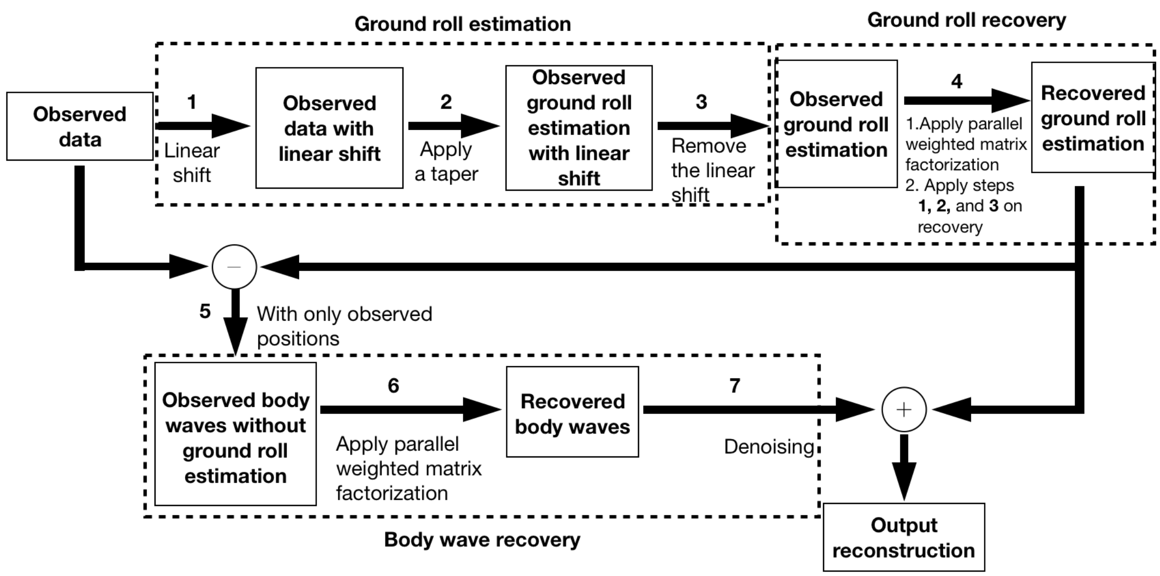


Figure 5.3: Flowchart of proposed method.

the randomly subsampled ground roll plotted in Figure 5.4d. This estimate for the ground roll serves as input for the reconstruction.

5.5.2 Ground roll recovery

We use the estimated ground roll as input to our wavefield reconstruction based on weighted matrix factorizations for a rank $r = 250$, which we find empirically by observing continuity of signals and limited noise in the reconstructed data. We run the reconstruction over all shots simultaneously for 320 iterations of SPG-ℓ_2 (Lopez, Kumar, and Herrmann 2015) per frequency slice. To avoid reconstruction leakage, we apply steps *(1)-(3)* from the previous section again to get the final ground roll recovery.

5.5.3 Body wave recovery

After reconstruction of the ground roll, we apply the mask $\mathcal{A}$ to restrict the ground roll reconstruction to the observed receiver positions again and subtract it from the original subsampled input data. The resulting "ground roll free" estimate for the body waves subsequently serves as input to a second wavefield reconstruction now for the body waves. Since these waves are more complex than ground roll, we choose the rank higher ($r = 340$). As we can observe from Figure 5.5, the reconstructed body waves contain, as expected, some remaining low-amplitude ground roll. Before inverse Fourier transforming the reconstructed body waves, we apply a $f - k$ filter to each shot along both receiver coordinates to remove remnant noise. The resulting recovery for the body waves shown in Figure 5.5b shows reconstruction of high frequency reflected and diffracted energy. To arrive at the final result, we combine the wavefield reconstructions for the ground roll and body waves. The result of this blind study for a single shot is included in Figure 5.6.

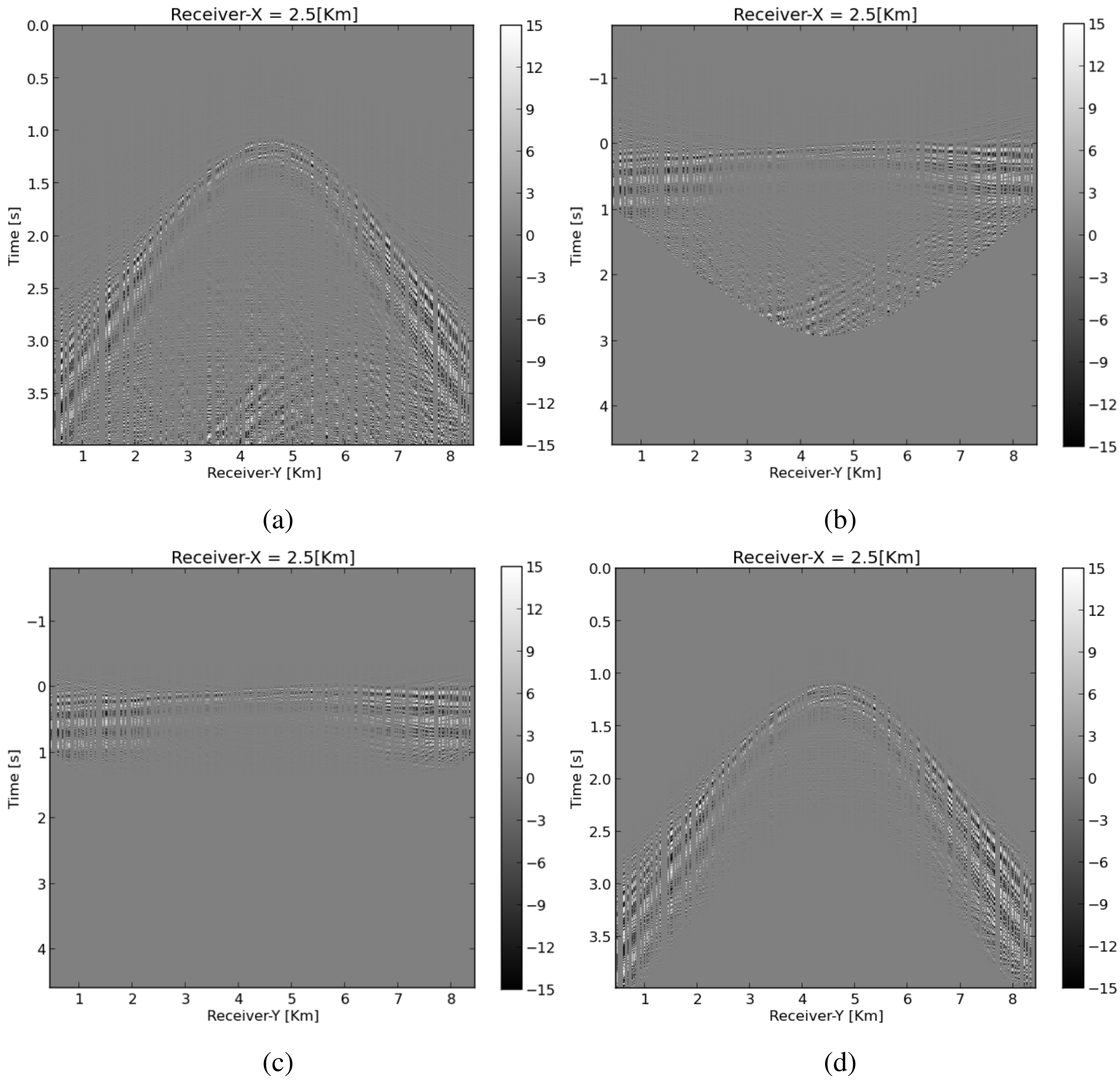
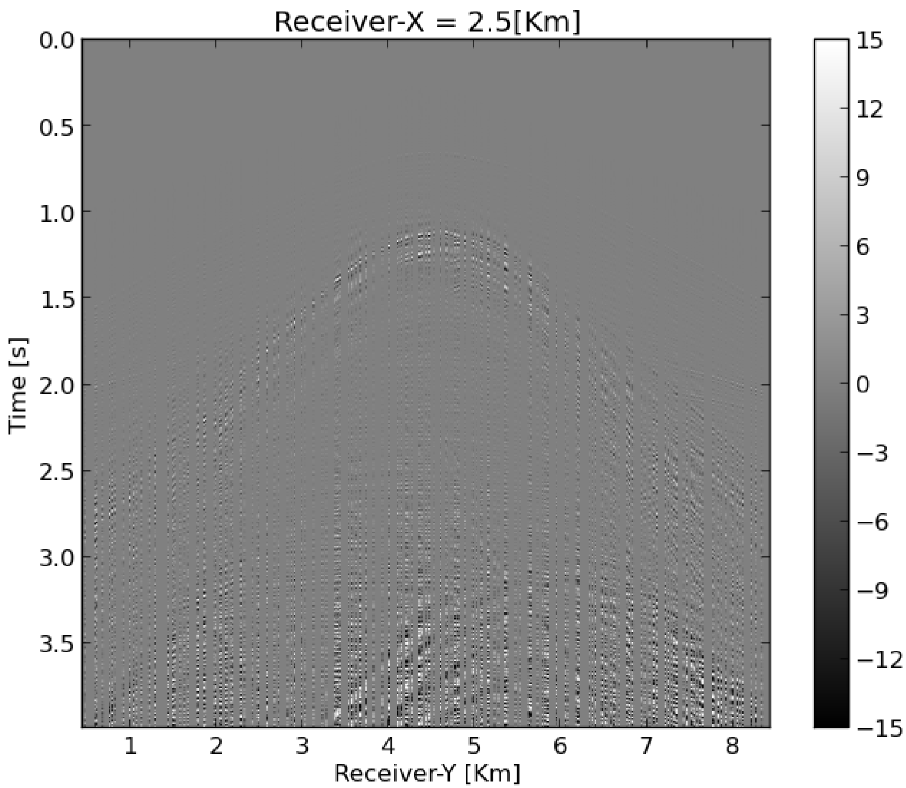
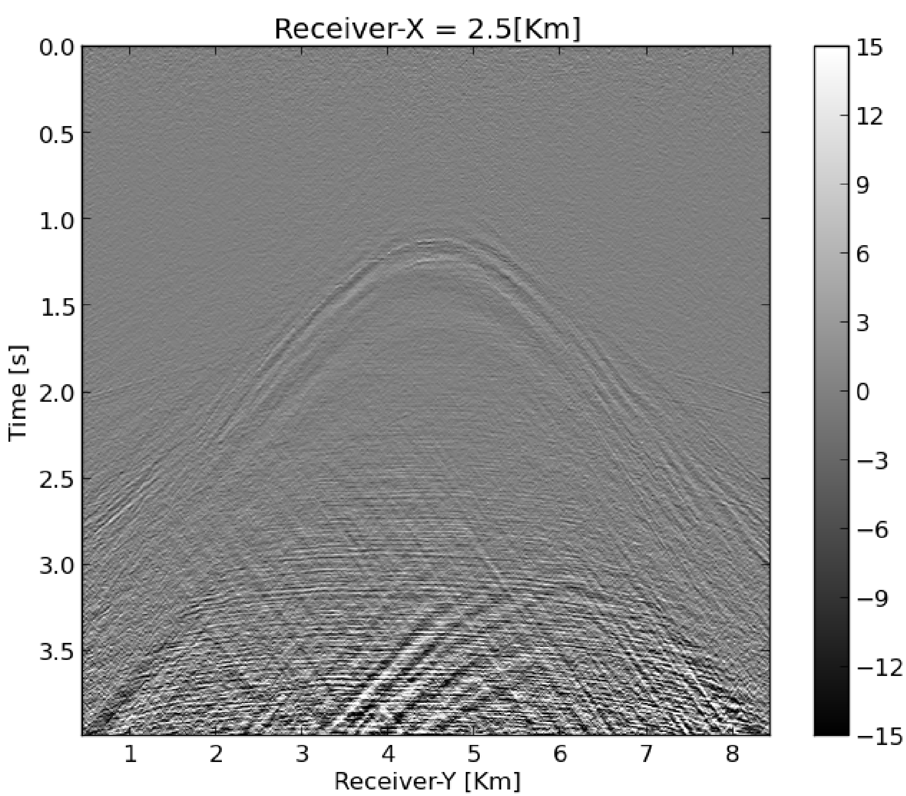


Figure 5.4: Ground roll estimation. *(a)* Input shot record. *(b)* Input after linear shift. *(c)* Tapered data. *(d)* Estimate subsampled ground roll after undoing linear shift.

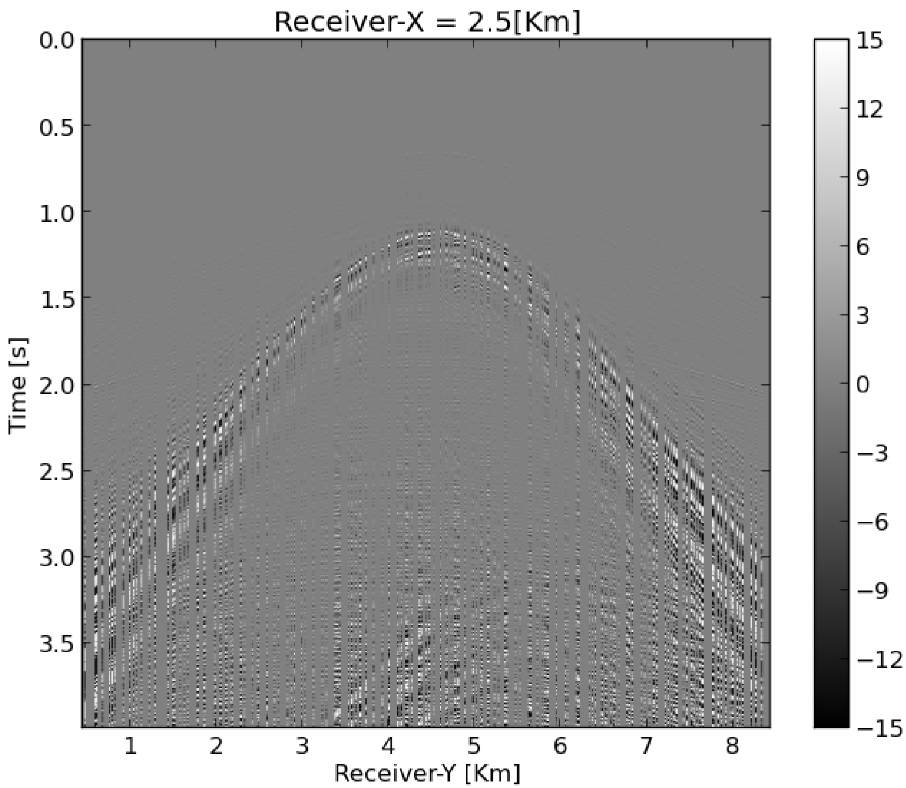


(a)

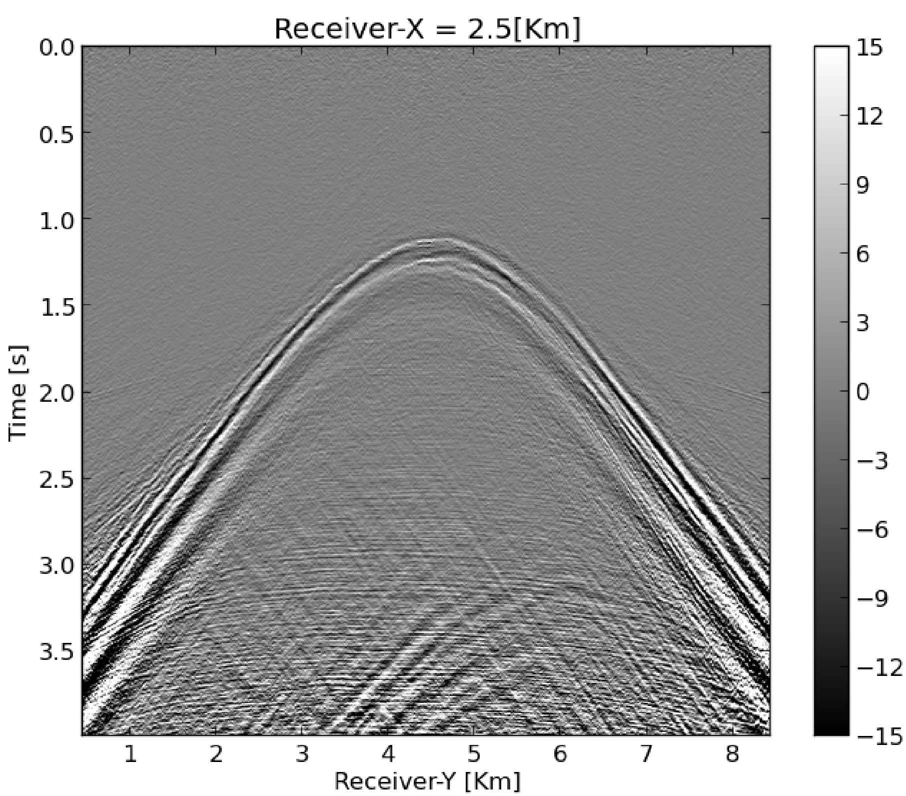


(b)

Figure 5.5: Body waves in the time domain. *(a)* Observed body waves. *(b)* Reconstructed body waves

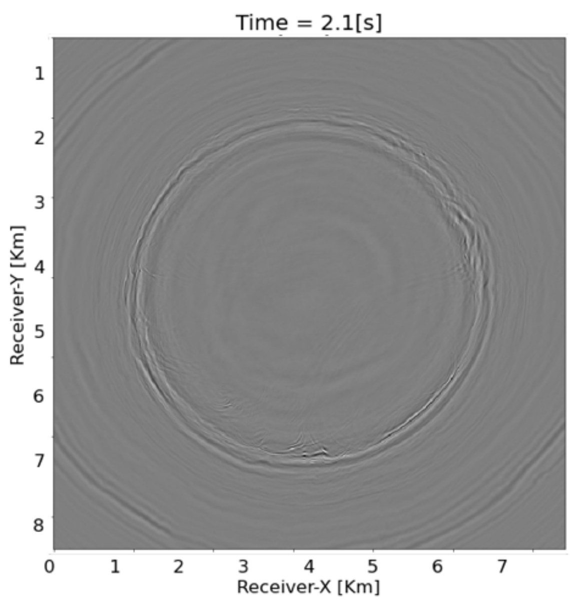


(a)

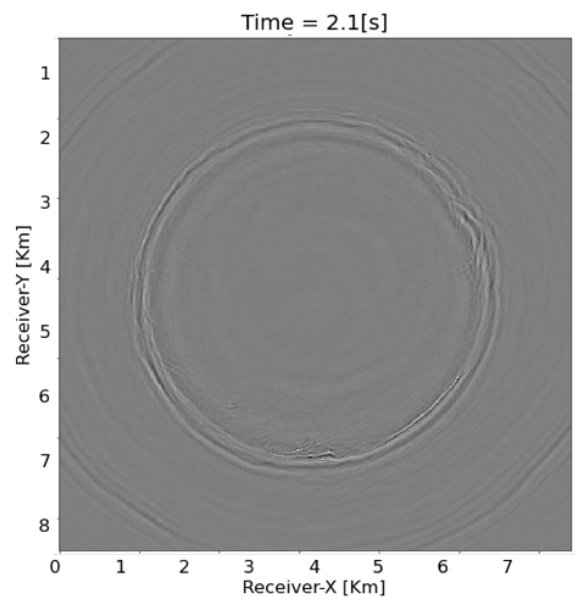


(b)

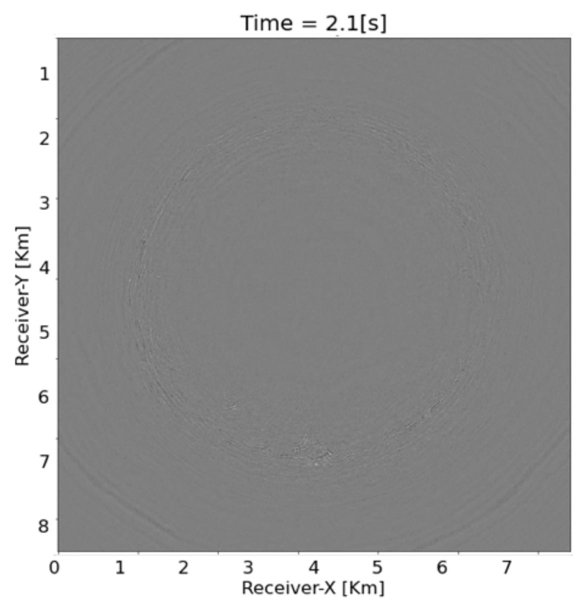
Figure 5.6: Seismic data reconstruction in the time domain. *(a)* Observed shot gather. *(b)* Reconstructed shot gather.



(a)



(b)



(c)

Figure 5.7: Wavefield recovery of one common shot gather. *(a)* Time slice at 2.1s of ground truth. *(b)* Time slice at 2.1s of reconstructed data. *(c)* Time slice at 2.1s of difference between ground truth and recovery. All the subfigures are plotted on the same scale.

5.6 Quality control (QC)

By comparing Figure 5.6b with Figure 5.2b, we observe that the proposed method produces results with less artifacts at the long offsets, and the ground roll is well recovered, especially at the near offsets (see Figure 5.6b for the receiver coordinate y between $4-6$ km for the time interval $1-1.5$ s).

To further verify our proposed practical workflow, we sent one reconstructed shot gather to Occidental and obtained the following plots in return. Figure 5.7 contains time slices at 2.1s with the ground truth (Figure 5.7a), reconstructed wavefield (Figure 5.7b) and difference plot (Figure 5.7c). From these plots, we observe that the proposed method successfully recovers body waves (reflections and diffractions) despite the presence of strong aliased ground roll.

5.7 Conclusions

We presented a practical workflow successfully recovering a subset of the synthetic 3D SEAM Barrett dataset randomly sampled at 25 percent receiver sampling. Our workflow consists of the combination of a weighted matrix factorization scheme and a separation of the subsampled input data into ground roll and body wave components. Thanks to this decomposition, we were able to mitigate the effects induced by the strong aliased ground roll. Initial findings of the blind test we carried out in collaboration with Occidental show that our method is capable of dealing with ground roll while recovering high-frequency body waves.

5.8 Related materials

The Julia code for this work is available on the SLIM GitHub page https://github.com/slimgroup/Software.SEG2021.

CHAPTER 6
A SIMULATION-FREE SEISMIC SURVEY DESIGN BY MAXIMIZING THE SPECTRAL GAP

6.1 Summary

Due to the tremendous cost of seismic data acquisition, methods have been developed to reduce the amount of data acquired by designing optimal missing trace reconstruction algorithms. These technologies are designed to record as little data as possible in the field, while providing accurate wavefield reconstruction in the areas of the survey that are not recorded. This is achieved by designing randomized subsampling masks that allow for accurate wavefield reconstruction via matrix completion methods. Motivated by these recent results, we propose a simulation-free seismic survey design that aims at improving the quality of a given randomized subsampling using a simulated annealing algorithm that iteratively increases the spectral gap of the subsampling mask, a property recently linked to the quality of the reconstruction. We demonstrate that our proposed method improves the data reconstruction quality for a fixed subsampling rate on a 2D and 3D realistic synthetic datasets.

6.2 Introduction

Due to relatively recent breakthroughs in compressive sensing (Candès, Romberg, and Tao 2006), seismic data is increasingly gathered randomly along spatial coordinates to decrease acquisition productivity (Mosher, Li, Morley, Ji, Janiszewski, Olson, and Brewer 2014; Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015; Chiu 2019). Wavefield reconstruction is used to recover fully sampled data from randomly subsampled observed seismic data (Hennenfent and Herrmann 2008; Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015; Zhang, Sharan, Lopez, and Herrmann 2020). To remove the imprint of large gaps in uniform random sampling, Gilles (Gilles 2008) proposed jittered subsampling, which by controlling the maximum gap size of subsampled data creates favorable conditions for seismic wavefield recovery based on sparsity promotion in a transformed domain made of localized atoms including curvelets (Herrmann, Wang, Hennenfent, and Moghaddam 2008). While uniform random (Candes and Recht 2009; Candès and Tao 2010) and random jittered subsampling schemes (Herrmann and Hennenfent 2008; Hennenfent and Herrmann 2008) are relatively straightforward to generate, these sampling strategies are almost certainly suboptimal and have shown to be improvable by solving certain optimization problem (Mosher, Li, Morley, Ji, Janiszewski, Olson, and Brewer 2014; Manohar, Brunton, Kutz, and Brunton 2018; Li, Kaplan, Mosher, Brewer, and Keys 2017). For instance, Mosher, Li, Morley, Ji, Janiszewski, Olson, and Brewer 2014, Li, Kaplan, Mosher, Brewer, and Keys 2017, and later Titova, Wakin, and Tura 2019 improved the reconstruction quality by devising a global optimization scheme that uses the mutual coherence.

In addition to wavefield reconstruction with optimized sampling schemes, Mosher, Li, Morley, Ji, Janiszewski, Olson, and Brewer 2014 also proposed a simulation-based acquisition design to support the use of compressive sensing in seismic data acquisition. For time-lapse seismic, Guo and Sacchi 2020 also used a data-driven approach where the acquisition is optimized by using prior information on the seismic data (Manohar, Brunton, Kutz, and Brunton 2018). While these methods have lead to promising results, they either require significant computational resources to determine the optimal source-receiver layout using combined wavefield simulations and recoveries or require detailed information on the to-be-collected seismic data. Neither is feasible for the design of optimized sampling strategies in 3D. However, the massive cost of 3D acquisition calls for methods to reduce the number of receivers by designing optimal receiver sampling masks.

To overcome this difficulty, this chapter provides a global optimization strategy for determining improved source-receiver layouts suitable for wavefield reconstructions based on matrix completion (Recht, Fazel, and Parrilo 2010; Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015; Kumar 2017) without the need to carry out expensive wavefield simulations. Similar to Li, Kaplan, Mosher, Brewer, and Keys 2017 and Mosher, Li, Morley, Ji, Janiszewski, Olson, and Brewer 2014, who propose a simulation-free optimization method based on the mathematical property of mutual coherence for transform-based wavefield reconstruction, our method involves improving the connectivity of graphs spanned by the binary sampling masks in the midpoint-offset domain for 2D case and the non-canonical Source-X/Receiver-X (columns) Source-Y/Receiver-Y (rows) domain for 3D case. According to Bhojanapalli and Jain 2014, by maximizing the spectral gap—i.e., the distance between the two first singular values—of the binary sampling mask the connectivity of the graph is improved, which favors reconstruction by matrix completion, an observation recently confirmed by López, Kumar, Moldoveanu, and Herrmann 2023 for 2D and 3D seismic wavefield reconstructions. While recent work by López, Kumar, Moldoveanu, and Herrmann 2023 indeed negates the need to run multiple costly wavefield reconstructions for different candidate sampling masks, this work does not yet provide a constructive method to generate sampling masks that maximize the spectral gap.

Unfortunately, the design of acquisition masks that maximize the spectral gap is an NP-hard problem (Guo and Sacchi 2020) whose solution requires a brute-force search through all combinatorial possibilities (Manohar, Brunton, Kutz, and Brunton 2018). When the number of sources or receivers becomes "large" (Li, Petropulu, and Trappe 2016), this precludes its practical use; for example, there are $\binom{n}{m} = \frac{n!}{m!(n-m)!} = 75287520$ possible combinations when selecting $m = 5$ subsampling positions from a pool of $n = 100$ candidate sites. For this reason, we propose to obtain an approximate solution by maximizing the spectral gap using simulated annealing (Kirkpatrick, Gelatt Jr, and Vecchi 1983), a stochastic local search optimization technique that is straightforward to implement, apply, and computationally feasible. Beside studies on 2D seismic showed that maximizing the spectral gap of the subsampling mask leads to better wavefield reconstruction results. We enrich the current study by proposing the simulation-free method to generate optimal 3D acquisition in the non-canonical Source-X/Receiver-X (columns) Source-Y/Receiver-Y (rows) domain by maximizing the spectral gap of the subsampling mask via the simulated annealing algorithm. The proposed method depends only on binary mask optimization, has a minimal computational cost, and should be adaptable to large-scale survey design.

We organize this chapter as follows. First, we present the proposed optimization problem to maximize the spectral gap of subsampling masks. Next, we explain how to obtain the approximate acquisition masks via simulated annealing for 2D case and 3D case. We conclude by demonstrating numerical experiments on the 2D as well as the 3D synthetic Compass datasets (Jones, Edgar, Selvage, and Crook 2012) and show the improvements in recovery quality compared to wavefield reconstruction of data collected with jittered subsampling method (Hennenfent and Herrmann 2008).

6.3 Methodology

Successful matrix-completion based seismic wavefield reconstruction (Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015; Recht, Fazel, and Parrilo 2010; Kumar 2017) hinges on three critical factors, namely: (1) an appropriate randomized subsampling scheme, such as uniform random or jittered subsampling (Hennenfent and Herrmann 2008; Herrmann and Hennenfent 2008); (2) existence of a transform domain in which the fully sampled seismic data organized as a matrix exhibits low-rank structure; (3) a computationally scalable matrix completion technique, which exploits the property that missing source and/or receivers increases the rank of these matrices. In this chapter, we propose a new constructive method to automatically generate improved source-receiver sampling masks, which favor seismic wavefield reconstruction via matrix completion in the midpoint-offset domain for 2D case and the non-canonical domain for 3D case. We begin by describing our approach to acquisition design.

6.3.1 spectral gap ratio based acquisition design for 2D seismic

Following López, Kumar, Moldoveanu, and Herrmann 2023, we constitute the spectral gap by the spectral gap ratio (SGR, the ratio of the first to second singular values), which becomes small for a large spectral gap. While the SGR indeed has been shown to be a valuable quantity to predict the quality of wavefield reconstruction with matrix completion (López, Kumar, Moldoveanu, and Herrmann 2023), ways to automatically generate acquisition masks with small SGRs have so far been lacking. To meet this challenge, we cast the problem of finding optimized acquisition masks with small SGRs as a minimization problem. Given n_s source locations, n_r receiver locations, and the source subsampling ratio r, we propose to solve a non-convex combinatorial optimization problem with respect to the subsampling mask $\mathbf{M} \in \{0,1\}^{n_s \times n_r}$—i.e., we have

$$
\begin{aligned}
\mathcal{L}(\mathbf{M}) = \min_{\mathbf{M}} \quad & \frac{\sigma_2(\mathcal{S}(\mathbf{M}))}{\sigma_1(\mathcal{S}(\mathbf{M}))} \\
& \text{subject to} \\
& \|\mathbf{M}\|_0 = \lfloor n_s \times r \rfloor \times n_r \cap \mathbf{M} \in \mathcal{J} \cap \mathbf{M} \in \{0,1\}^{n_s \times n_r}.
\end{aligned}
\tag{6.1}
$$

In this optimization problem, the objective function consists of the spectral gap ratio (SGR), defined by the ratio of the first, $\sigma_1(\cdot)$, and second, $\sigma_2(\cdot)$, singular values. $\mathcal{S}$ stands for the transformation operator with seismic reciprocity (Fenati and Rocca 1984) from the source-receiver domain to the midpoint-offset domain. We constrain the solution to conserve the subsampling ratio ($\|\mathbf{M}\|_0 = \lfloor n_s \times r \rfloor \times n_r$) and to stay jittered sampled with $\mathcal{J} \subseteq \{0,1\}^{n_s \times n_r}$ being the set of all possible jittered subsampling acquisitions. This constraint guarantees that the spread of the survey will not be modified, but only the local source position will be optimized. The symbol $\lfloor \cdot \rfloor$ denotes the floor operation. As we previously mentioned, in order to solve this combinatorial optimization problem, we implemented a simulated annealing method to obtain a solution in a finite and acceptable time. We now describe this algorithm and link each step to its subsampling mask counterpart.

6.3.2 Simulated annealing

Stochastic local search optimization algorithms are viable approximate methods for solving combinatorial optimization problems (e.g., Equation 6.1). Simulated annealing is a global optimization technique that uses local search to find approximate solutions to combinatorial optimization problems given a computational budget (Şahin, Ertoğral, and Türkbey 2010; Kirkpatrick, Gelatt Jr, and Vecchi 1983).

This optimization method has three main components(Van Laarhoven and Aarts 1987): (1) an initial state, $\mathbf{M}_0$, representing the initial solution to the optimization problem (Equation 6.1); (2) a set of neighboring states for any given state, which will be used to update the current state randomly; and (3) a transition probability that determines the probability of moving from one state to another. During optimization, at each given state, $\mathbf{M}_k$, which represents the current solution to the optimization problem Equation 6.1, a candidate state, $\tilde{\mathbf{M}}_k$, is chosen randomly from the neighboring states. Next, the algorithm transitions from the current state to the candidate state, i.e., from $\mathbf{M}_k$ to $\tilde{\mathbf{M}}_k$, if this transition reduces the objective function, i.e., $\mathcal{L}(\tilde{\mathbf{M}}_k) < \mathcal{L}(\mathbf{M}_k)$. On the other hand, if the objective function evaluated at the candidate state is larger than the current value, the algorithm makes the transition to the candidate state according to a transition probability, defined as follows (Kirkpatrick, Gelatt Jr, and Vecchi 1983):

$$p(\delta\mathcal{L}, k) = \exp\Big(\frac{-\delta\mathcal{L}}{T(k)}\Big), \tag{6.2}$$

where k is the iteration number, $\delta\mathcal{L} = \mathcal{L}(\tilde{\mathbf{M}}_k) - \mathcal{L}(\mathbf{M}_k)$ indicates the change in the objective function (Equation 6.1) by moving to the candidate state, and $T(k) : \mathbb{R} \to \mathbb{R}^+$, typically called temperature function, is a monotonically decreasing function that reduces the uphill transition probability towards the end of optimization while allowing uphill movement early in the optimization. We choose the temperature function as $T(k) = T_0 \times \alpha^k$, following a geometric reduction rule, which is the most commonly used function in the simulated annealing with a start temperature T_0 and the decrease rate α (Abramson, Krishnamoorthy, Dang, *et al.* 1999; Kirkpatrick, Gelatt Jr, and Vecchi 1983). This allows the algorithm to escape from local minima in the initial stages of the optimization while ensuring downhill movement towards the end. Finally, the transition probability is smaller for candidate states that increase the objective function more, i.e., $\delta\mathcal{L} \gg 0$, minimizing the probability of moving to very bad solutions.

To adapt simulated annealing to the acquisition design optimization problem (cf.Equation 6.1), we define the states as arbitrary positioning of sources. The algorithm is summarized in Algorithm 2. This algorithm is initialized with a subsampling mask $\mathbf{M}_0$ that is generated by using jittered subsampling method, known to facilitate seismic wavefield recovery (Hennenfent and Herrmann 2008). After updating the temperature function $T(k)$ (line 1) (Ma 2002), we select a source position $\tilde{\mathbf{M}}_k$ within the neighborhood of the current position (line 2). We then update the source position to this updated state according to the loss decrease and probabilistic update rule (lines $3-8$). After a predetermined number of iterations, the algorithm outputs the source sampling mask $\mathbf{M}_K$ with smaller SGR.

Algorithm 2 spectral gap ratio minimization via simulated annealing.

Inputs:

$\mathbf{M}_0$: Initial source positions using jittered subsampling method.
K: Maximum number of iterations.
$T(k)$: Temperature function.
p: Transition probability (cf. Equation 6.2).

0. **for** $k = 0$ **to** $K-1$ **do**
1. $T(k) = T_0 \times \alpha^k$
2. $\tilde{\mathbf{M}}_k \leftarrow$ randomly pick a neighboring state
3. $\delta\mathcal{L} = \mathcal{L}(\tilde{\mathbf{M}}_k) - \mathcal{L}(\mathbf{M}_k)$
4. **if** $\delta\mathcal{L} \leq 0$
5. $\mathbf{M}_{k+1} = \tilde{\mathbf{M}}_k$
6. **else**
7. $\mathbf{M}_{k+1} = \begin{cases} \tilde{\mathbf{M}}_k \text{ with probability } p(\delta\mathcal{L}, k) \\ \mathbf{M}_k \end{cases}$
8. **end if**
9. **end for**

Output: $\mathbf{M}_K$

In order to satisfy the constraints introduced in Equation 6.1, we carefully define the neighborhood of acceptable state to prevent sources positions to cluster around a few areas of the survey. We now detail its design.

Neighboring states

Given the source sampling at an iteration, we randomly select 20% of the source positions in the current state to balance between exploring the search space and avoiding too large change between adjacent iterations (Assad and Deep 2018; Olorunda and Engelbrecht 2008). Using the subsampling factor $f = \frac{1}{r}$, the fine grid with all possible source locations is divided into f equal regions. Each selected source is allowed to randomly shift within the region in which it is located. For clarity, we summarize the perturbation rule in Figure 6.1 with $n_s = 20$ source positions and a subsampling factor of $f = 5$. We choose the movement range R to ensure that we remain close to the jittered sampling, which has been shown to result in better wavefield recoveries (Hennenfent and Herrmann 2008). We now detailed our synthetic numerical experiment demonstrating the benefits of our method for data reconstruction.

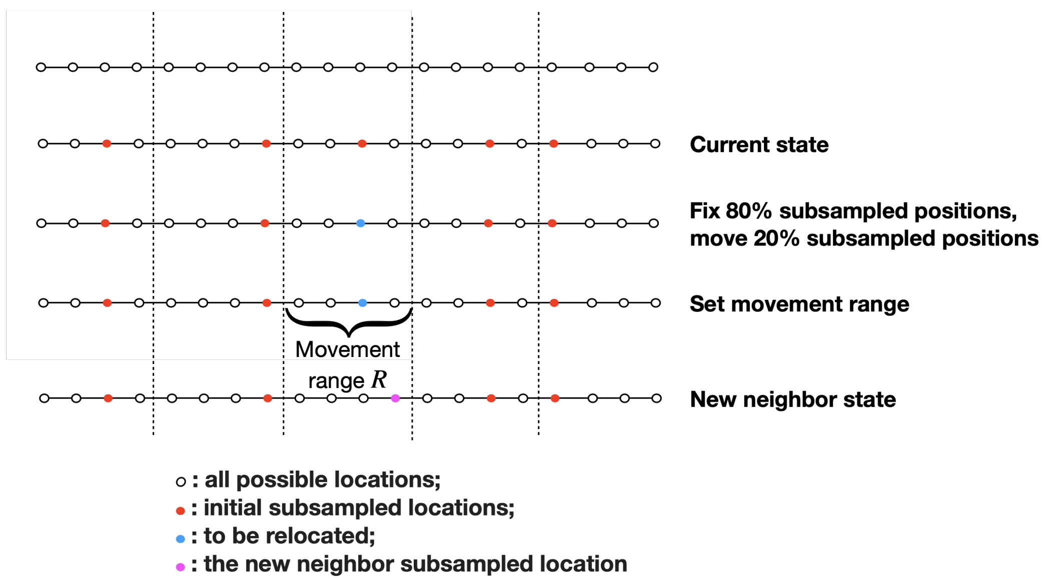


Figure 6.1: Random state perturbation rules to satisfy the constraints. White circles (○) indicate all possible source locations. Five red circles represent an initial state. The blue circle represents the 20% of sources that we will move within movement range R to arrive at a neighboring state (purple circle).

6.3.3 spectral gap ratio based acquisition design for 3D seismic

Motivated by the success on 2D survey design methods driven by SGR minimization, we consider 3D survey design where receivers are missing and sources are fully sampled. Because 3D wavefield reconstruction based on low-rank matrix completion relies on the non-canonical Source-X/Receiver-X (columns) Source-Y/Receiver-Y (rows) organization of the data into a matrix, we aim to minimize the SGR of the subsampling mask in that domain. Fortunately, when sources are fully sampled, each single-receiver block of the global sampling matrix is either fully sampled or empty depending on whether that specific receiver is sampled. Consequently, the block structure of the global matrix leads to the exact same singular values as a single-source receiver sampling mask. We can therefore optimize a single-source mask to obtain the global optimized mask (Figure 6.2). The main computational cost is computing the first two singular values of the receiver sampling mask, which is negligible compared to approaches that require wave simulations. The resulting optimal mask with the lowest SGR indicates the receiver sampling locations that favor 3D wavefield reconstruction via matrix completion in the non-canonical organization domain.

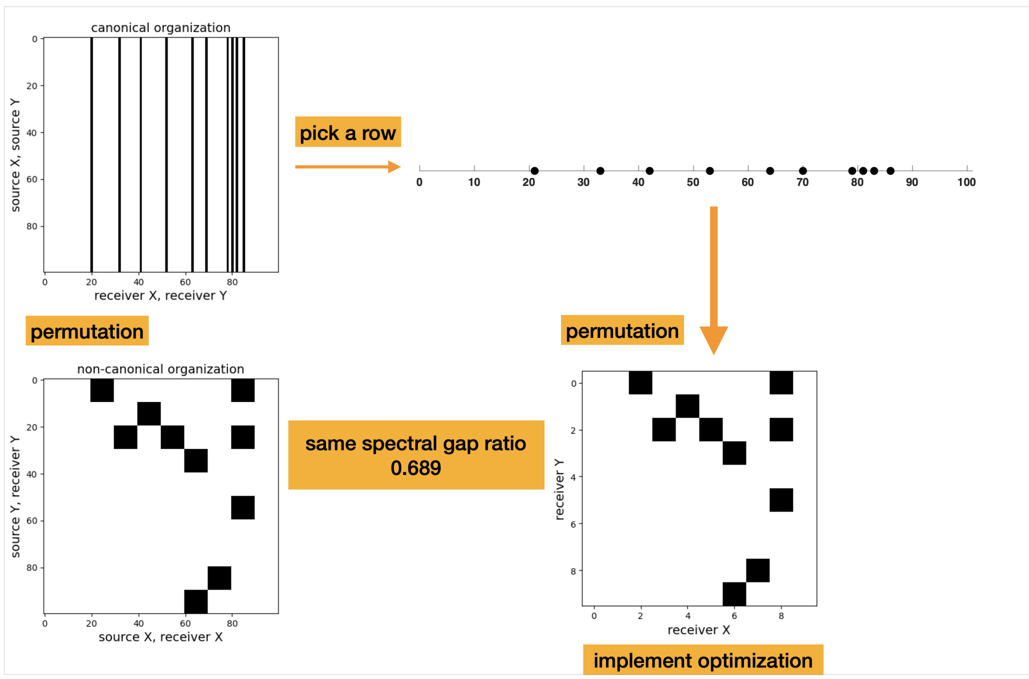


Figure 6.2: SGR of the data matrix in the non-canonical Source-X/Receiver-X (columns) Source-Y/Receiver-Y (rows) domain is the same as the SGR of the single-source receiver sampling matrix.

6.4 Numerical experiments

We consider a 2D and a 3D marine datasets simulated over the realistic Compass model (Jones, Edgar, Selvage, and Crook 2012).

6.4.1 Experiments for 2D dataset

The 2D dataset consists of 300 sources and 150 receivers sampled at $12.5\,\mathrm{m}$. The data is recorded at a $2\,\mathrm{ms}$ sampling rate for $2.046\,\mathrm{s}$ (1024 time samples). Based on this dataset, we proceed in two steps. First, we will compare the subsampling mask we obtain with our proposed method against the standard jittered sampling mask. Second, we will show that the recovered data is of better quality as expected from the optimal SGR that quantifies the expected quality of recovery. The jittered subsampling method used in this abstract allows neighboring subsamples (non-gap subsamples), which creates favorable conditions for recovery and is defined as optimally-jittered subsampling in Hennenfent and Herrmann 2008. We use the weighted matrix completion method (Zhang, Sharan, Lopez, and Herrmann 2020) to recover the observed data and evaluate the quality of the recovered dataset.

We start with picking a subsampling mask using the jittered method (Herrmann and Hennenfent 2008) that includes 20$\%$ of the sources. In Figure 6.3a, we show the subsampling mask in the source-receiver domain. Under the assumption of the source-receiver reciprocity (Fenati and Rocca 1984), we apply this reciprocity on the subsampling mask to implement a realistic seismic survey design. The subsampling mask in the midpoint-offset domain with seismic reciprocity is depicted in Figure 6.3b. With this mask as an initial guess, we perform 4000 iterations of simulated annealing to obtain an optimal subsampling mask. Figure 6.3c and Figure 6.3d show the resulting subsampling mask in the source-receiver and midpoint-offset domains, respectively. We observe that the SGR was reduced by 30$\%$ with a fixed subsampling rate hinting towards well-improved data reconstruction. The proposed method is a simulation-free method that depends only on binary mask optimization.

With this optimized subsampling mask, we now perform data reconstruction via weighted matrix completion (Zhang, Sharan, Lopez, and Herrmann 2020) and compare the result against reconstructing the data sampled with the initial jittered subsampled mask. In both cases, we use the same algorithm and hyperparameters (e.g., number of iterations, rank) for a fair comparison. We summarize the recovery in Figure 6.4.

We first show the ground truth in Figure 6.4a, where the right plot shows the full shot record and the left one depicts the later arrival events between about $1\,\mathrm{s}$ to $2\,\mathrm{s}$. By applying these two masks (jittered mask and proposed mask) individually on the ground truth, we obtain two observed datasets with 80% of sources missing. The proposed subsampled data is illustrated in Figure 6.4b. Figure 6.4c shows the reconstruction from jittered observed data with a signal-to-noise ratio (SNR) of $14.6\,\mathrm{dB}$ for the full shot record and $12.8\,\mathrm{dB}$ for the later arrival events. The reconstruction from the proposed subsampled data is shown in Figure 6.4d, with SNRs of $14.91\,\mathrm{dB}$ for the full shot record and $12.9\,\mathrm{dB}$ for the later arrival events. Figure 6.4e illustrates the difference between Figure 6.4c and Figure 6.4a, whereas Figure 6.4f shows the difference between Figure 6.4d and Figure 6.4a. The wavefield reconstructions demonstrate that the reconstruction from the proposed subsampling mask gives a more accurate data reconstruction for the full shot record and the later arrival events in terms of SNR. The difference is significantly reduced in Figure 6.4f in contrast to Figure 6.4e.

To further validate the performance of our proposed method, we show that our proposed mask outperforms on the average standard jittered acquisition and not just for a single experiment. We randomly generate five independent jittered subsampling masks (removing 80% of sources) and then utilize the proposed approach to minimize the SGR of these five jittered subsampling layouts. These five jittered and proposed masks are then used to perform weighted matrix completion (Zhang, Sharan, Lopez, and Herrmann 2020; Zhang, Sharan, and Herrmann 2019) and reconstruct the full dataset. The results are summarized

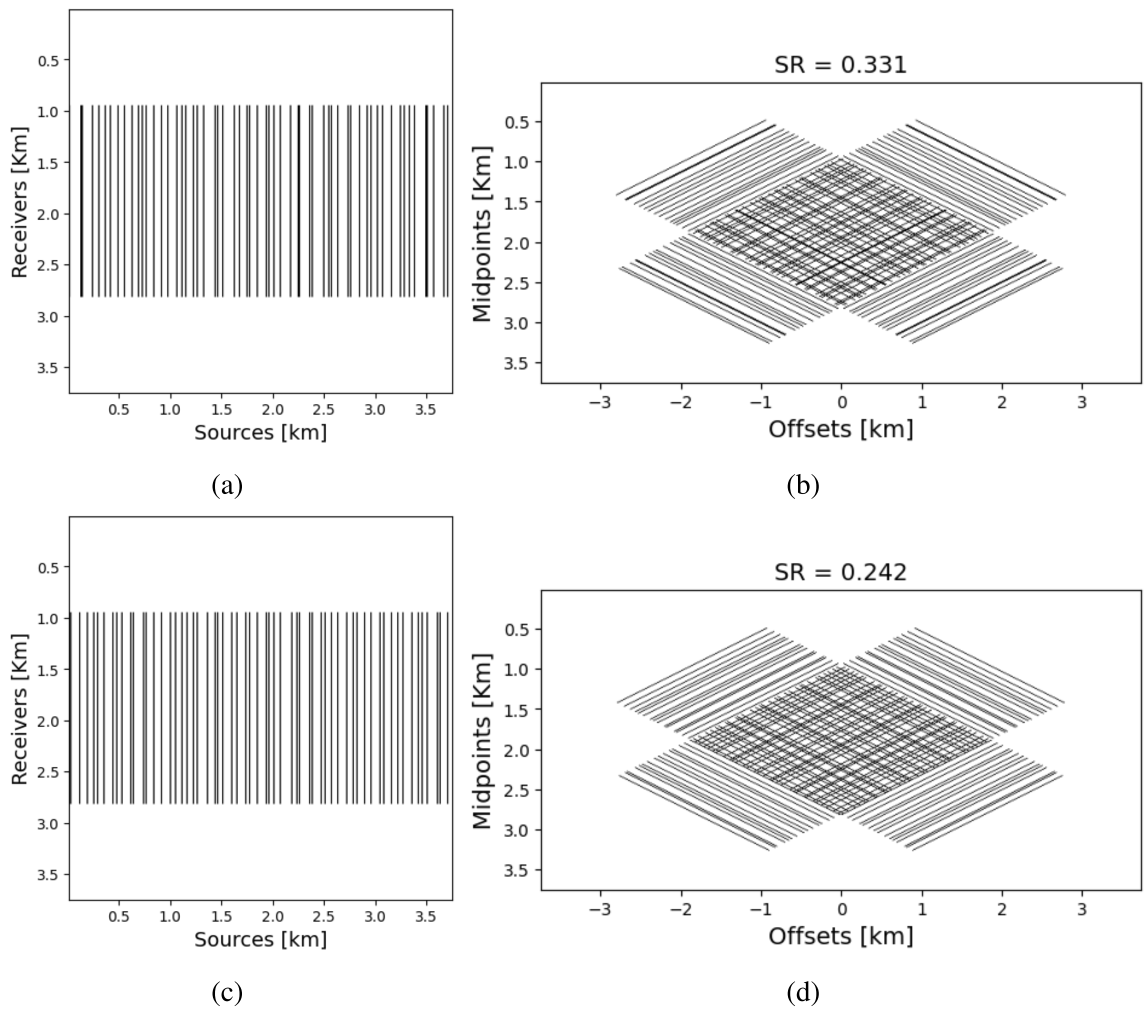


Figure 6.3: Jittered subsampling mask in the *(a)* source-receiver domain and *(b)* midpoint-offset domain (SGR $= 0.331$). Optimized subsampling mask in the *(c)* source-receiver domain and *(d)* midpoint-offset domain (SGR $= 0.242$).

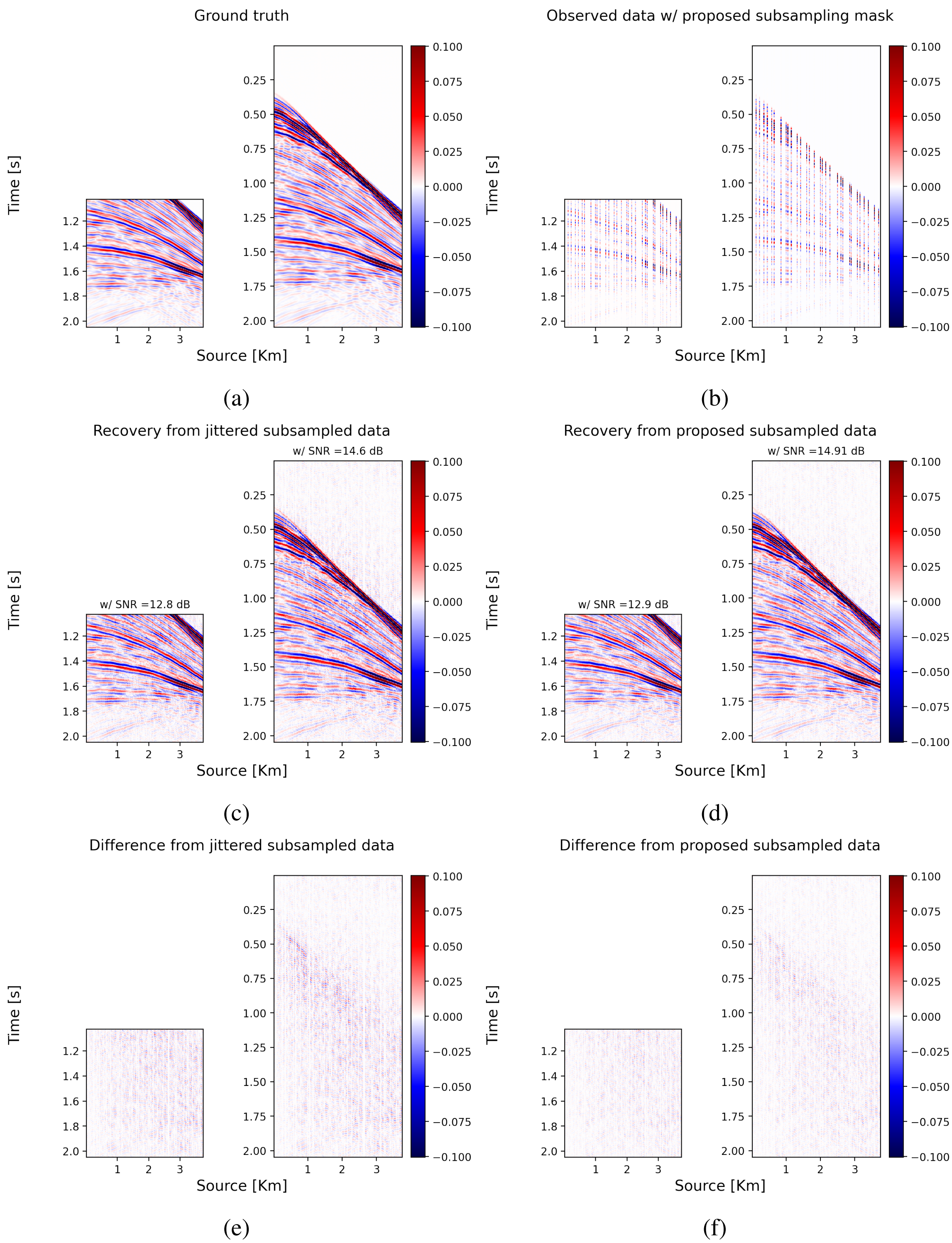
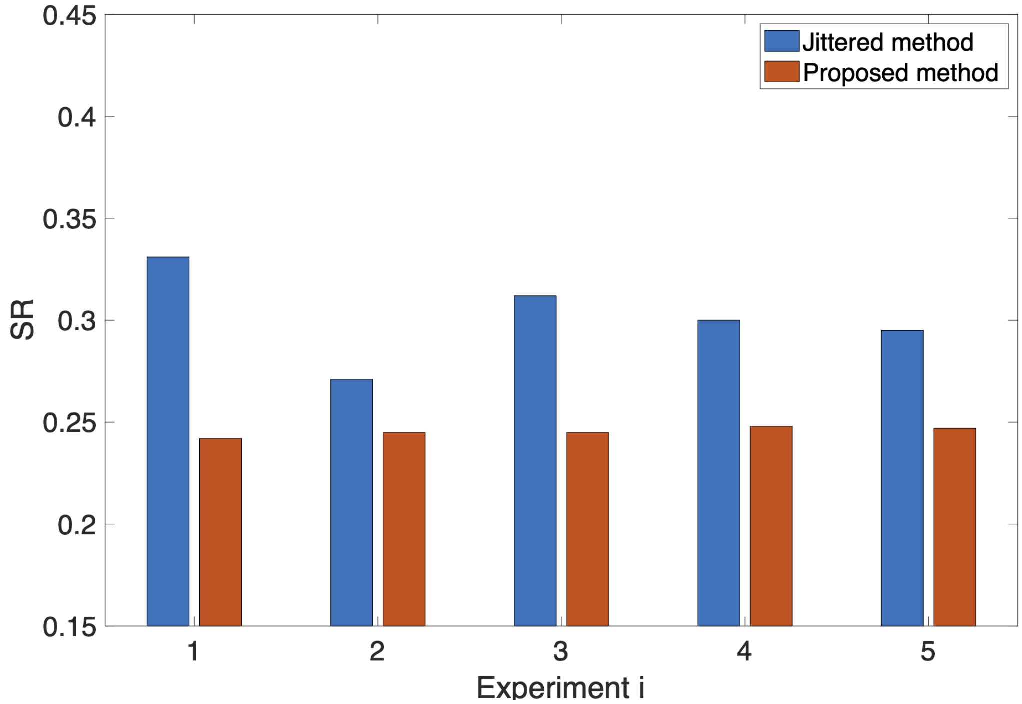
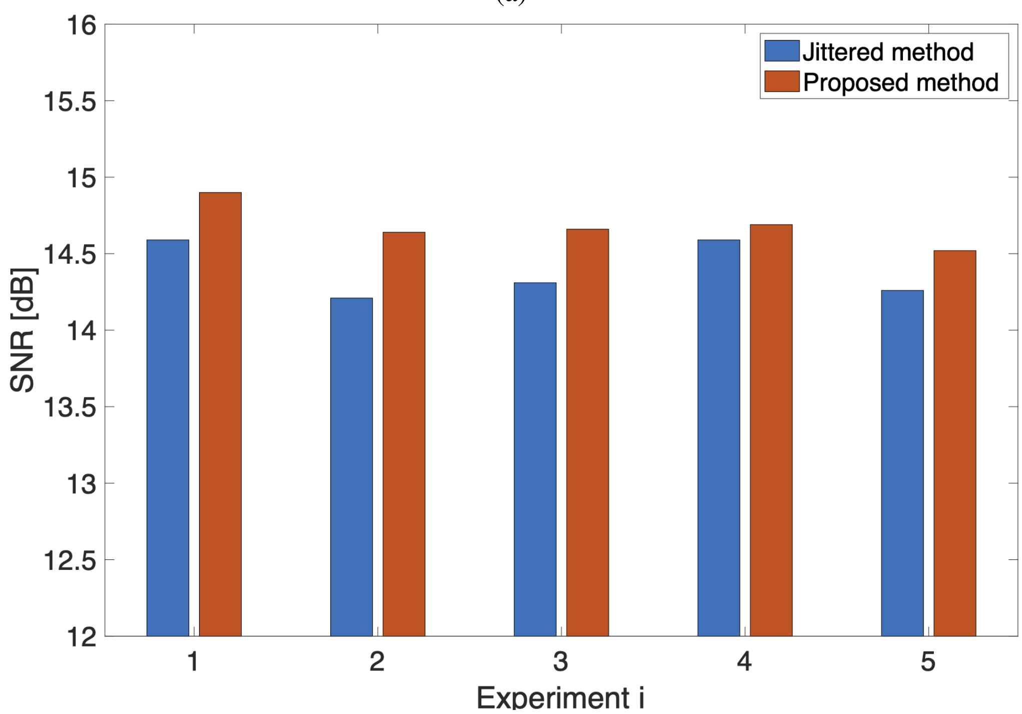


Figure 6.4: Wavefield reconstruction results in the time domain. *(a)* Ground truth. *(b)* 80% subsampled seismic data with proposed subsampling. Reconstructions and differences from 80% missing sources: *(c)* jittered subsampling, SNR = $14.6\,\mathrm{dB}$ and $12.8\,\mathrm{dB}$ for later arrival events, *(d)* improved subsampling with SNR = $14.91\,\mathrm{dB}$ and $12.9\,\mathrm{dB}$ for later arrival events, *(e)* difference of jittered subsampling, *(e)* difference of improved subsampling.

in Figure 6.5. The bar plots in Figure 6.5 lead to the following observations. First, our proposed method consistently reduced the spectral gap ratio by at least 11% leading to a similar optimal SGR for this given subsampling ratio and acquisition. Second, the recovered data always presents a higher SNR representative of a more accurate wavefield reconstruction. These two results show that despite being a potentially aleatory method, our simulated annealing based SGR minimization method consistently provides a subsampling mask best fitted for data reconstruction.



(a)



(b)

Figure 6.5: *(a)* SGR comparison (lower is better) of subsampling masks using jittered method versus proposed method. *(b)* Reconstruction SNR comparison (higher is better) from observed data using jittered method and proposed method. The results are obtained by five independent experiments.

6.4.2 Experiments for 3D dataset

To illustrate the efficacy of our method via a numerical experiment on a simulated 3D marine dataset over the compass model. The data volume consists of 501 time samples, 1681 sources and $10\,\mathrm{k}$ receivers. The distance between the adjacent sources and receivers are $150\,\mathrm{m}$ and $25\,\mathrm{m}$ in each direction, respectively, with a time sampling interval of $0.01\,\mathrm{s}$. By removing 90% of receivers using jittered subsampling, we obtain a binary matrix with the SGR 0.507 in the non-canonical domain. After applying simulated annealing algorithm, the SGR of mask effectively decreases to 0.328. To validate the efficacy of our acquisition design method, we perform data reconstruction on a frequency slice at $16.8\,\mathrm{Hz}$ via weighted matrix completion for the two subsampled datasets with jittered subsampling mask and the proposed mask, with results shown in Figure 6.6. The reconstruction signal-to-noise ratio from the observed data at proposed receiver locations is $12.27\,\mathrm{dB}$, which is about $1.4\,\mathrm{dB}$ higher than the reconstruction signal-to-noise ratio $10.88\,\mathrm{dB}$ achieved from data observed at jittered sampled receiver locations. This confirms that the proposed optimized receiver sampling locations result in a superior seismic survey that leads to better wavefield reconstruction performance.

6.5 Conclusions

We proposed a simulation-free method for seismic survey design in this chapter by minimizing the spectral gap ratio using the simulated annealing algorithm. This are the first numerical case studies that applies spectral gap ratio minimization techniques to seismic acquisition design that favors 2D and 3D wavefield reconstructions. Because the proposed method solely relies on a binary mask optimization rather than being data-driven, the computational cost is minimal and should scale to industry-size survey design. Through analysis and experiments, we conclude that the proposed method generates the improved 2D and 3D seismic surveys better suitable for wavefield reconstruction.

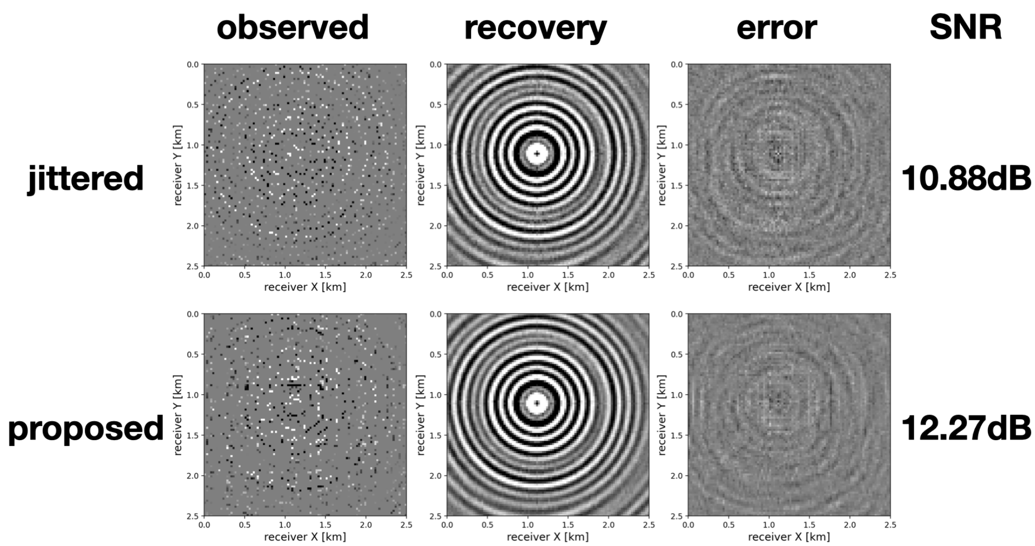


Figure 6.6: Comparison of data reconstruction performance for receiver locations sampled by the jittered method and the proposed method. There is about $1.4\,\mathrm{dB}$ SNR improvement.

CHAPTER 7

OPTIMIZED TIME-LAPSE ACQUISITION DESIGN VIA SPECTRAL GAP RATIO MINIMIZATION

7.1 summary

Modern-day reservoir management and monitoring of geological carbon storage increasingly call for costly time-lapse seismic data collection. In this chapter, we show how techniques from graph theory can be used to optimize acquisition geometries for low-cost sparse 4D seismic. Based on midpoint-offset domain connectivity arguments, the proposed algorithm automatically produces sparse non-replicated time-lapse acquisition geometries that favor wavefield recovery.

7.2 Introduction

Time-lapse seismic data acquisition is a costly but crucial endeavor for reservoir management and monitoring of geological carbon storage. While sparse randomized collection of seismic data can lead to major improvements in acquisition productivity (Herrmann and Hennenfent 2008; Hennenfent and Herrmann 2008; Herrmann 2010; Mosher, Li, Morley, Ji, Janiszewski, Olson, and Brewer 2014), systematic approaches to performance prediction, other than computationally expensive simulation-based studies, are mostly lacking. Besides, acquisition optimization approaches, such as minimizing the mutual coherence (Tang, Ma, and Herrmann 2008; Mosher, Li, Morley, Ji, Janiszewski, Olson, and Brewer 2014; Obermeier and Martinez-Lorenzo 2017) or minimizing the spectral gap ratio (SGR, Zhang, Louboutin, Siahkoohi, Yin, Kumar, and Herrmann 2022; López, Kumar, Moldoveanu, and Herrmann 2023) , do not handle the unique challenges of time-lapse seismic data acquisition.

To meet these challenges, inversion with the joint recovery model (JRM, Oghenekohwo, Wason, Esser, and Herrmann 2017; Wason, Oghenekohwo, and Herrmann 2017) will be combined with automatic binary sampling mask generation driven by SGR minimization (Zhang, Louboutin, Siahkoohi, Yin, Kumar, and Herrmann 2022). We opt for the JRM because it inverts baseline and monitor surveys jointly for the common component, which contains information shared between the surveys, and innovations with respect to the common component. Since the fictitious common component is observed by all surveys, its recovery improves when the time-lapse surveys contain complementary information. This is the case when sparse surveys are not replicated (Oghenekohwo, Wason, Esser, and Herrmann 2017; Wason, Oghenekohwo, and Herrmann 2017) or when the time-lapse datasets contain independent noise terms (Tian, Wei, Li, Oppert, and Hennenfent 2018). In either case, the JRM leads without insisting on replication of the surveys to high degrees of time-lapse repeatability both in the data (Oghenekohwo, Wason, Esser, and Herrmann 2017; Wason, Oghenekohwo, and Herrmann 2017) and image space (Yin, Erdinc, Gahlot, Louboutin, and Herrmann 2023). It also yields better interpretability of time-lapse field data (Wei, Tian, Li, Oppert, and Hennenfent 2018).

As demonstrated by this chapter, including the common component offers additional advantages when optimizing time-lapse acquisition via SGR minimization. To demonstrate this, we first explain the relationship between the SGR and connectivity within graphs associated with binary sampling masks. Next, we describe how this connectivity, which favors wavefield reconstruction, can be improved by minimizing the SGR via optimization. To enhance inversion of time-lapse data with the JRM, a new optimization objective will be introduced that contains SGRs of the common component and of the baseline/monitor surveys. After a brief discussion on minimizing this objective with simulated annealing, the proposed methodology for automatic time-lapse binary mask generation is numerically validated on realistic synthetic 2D data.

7.3 Optimized time-lapse acquisition

While the SGR has been used successfully to predict and improve the performance of wavefield reconstruction, it has not yet been used to optimize time-lapse acquisition. After briefly discussing the SGR and JRM, we introduce our methodology to optimize time-lapse data acquisition.

7.3.1 The spectral gap ratio

As shown by López, Kumar, Moldoveanu, and Herrmann 2023, the success of seismic wavefield reconstruction via universal matrix completion (Bhojanapalli and Jain 2014) can be predicted by the ratio of the first two singular values of binary sampling masks, $\sigma_2(\mathbf{M})/\sigma_1(\mathbf{M}) \in [0,\ 1]$ where $\mathbf{M}$ is a binary matrix with 1's where data is sampled and with 0's otherwise. This ratio is known as the spectral gap ratio (SGR) and provides a cheap-to-compute quantitative measure to predict recovery quality. The smaller the SGR, the better the connectivity within graphs spanned by binary sampling masks. Improved connectivity leads to improved wavefield recovery (López, Kumar, Moldoveanu, and Herrmann 2023). While useful, the SGR itself is not constructive because it does not produce sampling masks with small SGRs. With simulated annealing, Zhang, Louboutin, Siahkoohi, Yin, Kumar, and Herrmann 2022 came up with a practical algorithm to generate acquisition geometries with small SGRs. In this work, we extend this approach by optimizing sparse geometries for time-lapse data acquisition.

7.3.2 Optimized sampling mask generation

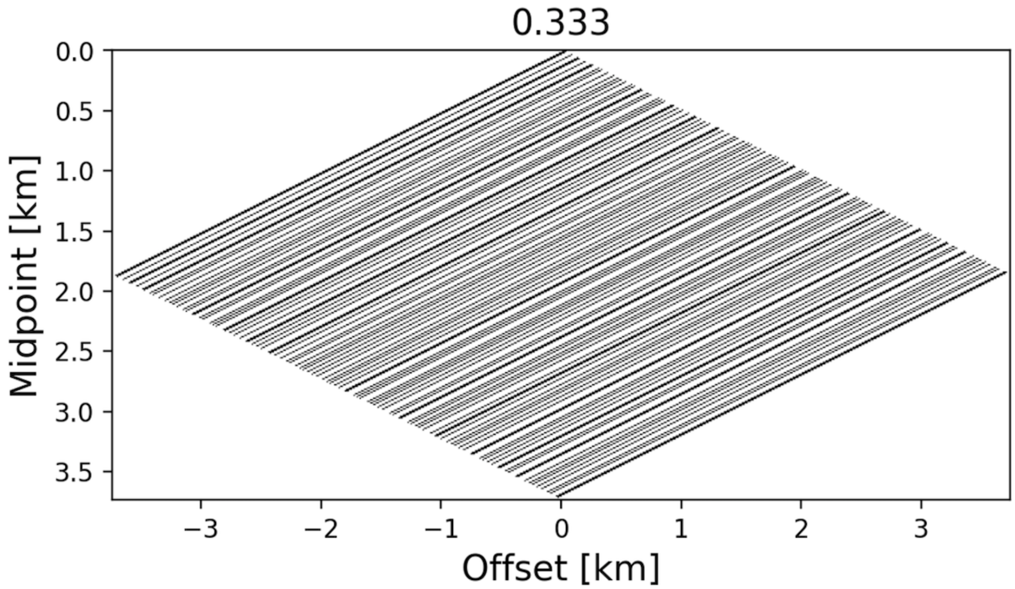
Given an initial binary mask, $\mathbf{M} \in \{0,1\}^{n_s \times n_r}$, with n_s sources and n_r receivers, Zhang, Louboutin, Siahkoohi, Yin, Kumar, and Herrmann 2022 proposed a methodology to minimize the SGR via

$$\underset{\mathbf{M}}{\text{minimize}} \quad \mathcal{L}(\mathbf{M}) \quad \text{subject to} \quad \mathbf{M} \in \mathcal{C}, \tag{7.1}$$

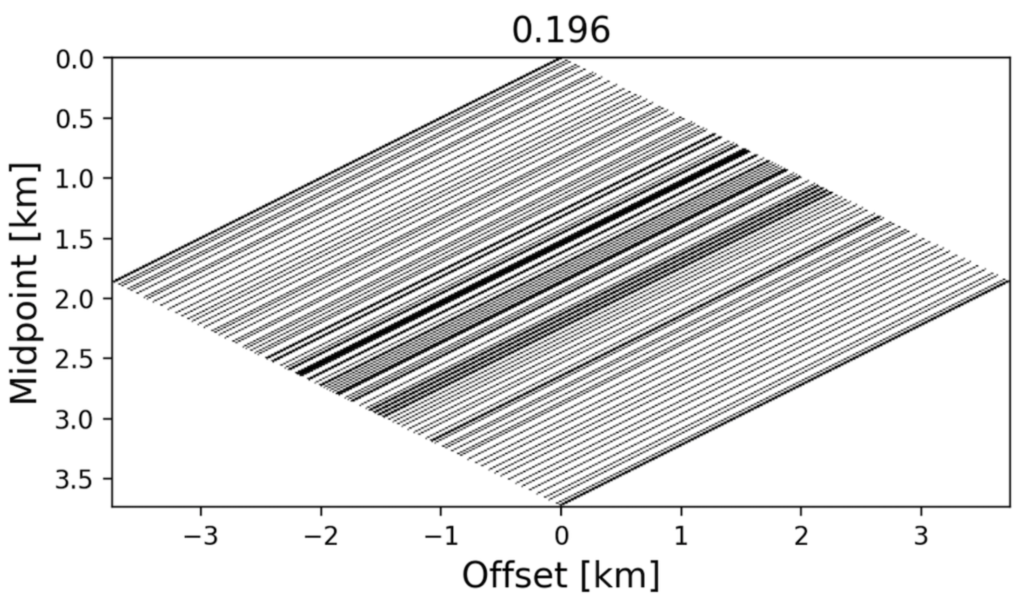
with the objective, $\mathcal{L}(\mathbf{M}) = \sigma_2(\mathbf{M})/\sigma_1(\mathbf{M})$, given by the SGR. To ensure feasibility of the optimized binary masks with source subsampling ratio $\rho \in (0,1)$, the constraint, $\mathcal{C} = \bigcap_{i=1}^{3} \mathcal{C}_i$, is included, which consists of the intersection of the cardinality constraint, $\mathcal{C}_1 = \{\mathbf{M} \mid \#(\mathbf{M}) = \lfloor n_s \times \rho \rfloor \times n_r\}$, the binary mask constraint, $\mathcal{C}_2 = \{\mathbf{M} \mid \mathbf{M} \in \{0,1\}^{n_s \times n_r}\}$, and a constraint on the maximum gap size between consecutive samples, $\mathcal{C}_3 = \{\mathbf{M} \mid \mathrm{maxgap}(\mathbf{M}) \leq \Delta\}$, where Δ is the maximal permitted gap size. By solving Equation 7.1, Zhang, Louboutin, Siahkoohi, Yin, Kumar, and Herrmann 2022 produced binary sampling masks that improved wavefield reconstruction compared to masks generated with randomized jittered sampling (Hennenfent and Herrmann 2008). Figure 7.1 contrasts jittered with optimized sampling in the midpoint-offset domain, reducing the SGR from 0.333 to 0.196. The optimized mask increases the sampling at the near offsets where there are more ways to connect to midpoints, which favors wavefield reconstruction (López, Kumar, Moldoveanu, and Herrmann 2023). This chapter delves deeper into time-lapse survey design and proposes a novel simulation-free method to find near optimal sparse source locations for a baseline survey and one or more monitor survey(s). Our method is guided by the joint recovery model, which has a successful track record in time-lapse wavefield reconstruction (Oghenekohwo, Wason, Esser, and Herrmann 2017; Wason, Oghenekohwo, and Herrmann 2017). Next, we introduce the joint recovery model and present our spectral gap ratio minimization framework, which is tailored to optimize time-lapse survey design in accordance with the joint recovery model.

7.3.3 Joint recovery model

Lowering costs while ensuring time-lapse repeatability are the main challenges in the design of seismic monitoring systems employed to optimize reservoir management and to safeguard geological carbon storage. Both challenges can be met by inverting sparsely sampled baseline and monitor data jointly. For time-lapse acquisition with a single monitor



(a)



(b)

Figure 7.1: (a) Jittered versus (b) optimized sampling mask in the midpoint-offset domain.

survey, this entails inverting

$$\mathbf{b} = \mathcal{A}\left(\mathbf{Z}\right) \quad \text{with} \quad \mathcal{A}\left(\cdot\right) = \begin{bmatrix} \mathcal{A}_1 & \mathcal{A}_1 & 0 \\ \mathcal{A}_2 & 0 & \mathcal{A}_2 \end{bmatrix} \left(\cdot\right). \tag{7.2}$$

In this JRM, the linear operators, $\mathcal{A}_j,\ j = 1, 2$, stand for the combined action of converting monochromatic time-lapse data from the midpoint-offset to the source-receiver domain, followed by trace collection with the acquisition geometries defined by the binary sampling masks, $\mathbf{M}_j,\ j = 1, 2$ with $j = 1$ and $j = 2$ masks for the baseline/monitor surveys. With this model, time-lapse data, $\mathbf{b}$, which contains the baseline, $\mathbf{b}_1$ and monitor data, $\mathbf{b}_2$, are linearly related to $\mathbf{Z}$, which contains matrices for the unknown densely sampled common component, $\mathbf{Z}_0$, and innovations with respect to this common component, $\mathbf{Z}_j,\ j = 1, 2$. Compared to recovering the baseline/monitor surveys separately, the JRM produces repeatable results from non-replicated (Oghenekohwo, Wason, Esser, and Herrmann 2017; Wason, Oghenekohwo, and Herrmann 2017; Kumar, Wason, Sharan, and Herrmann 2017), non-calibrated (Oghenekohwo and Herrmann 2017), and noisy (Tian, Wei, Li, Oppert, and Hennenfent 2018), time-lapse data. These enhanced results are due to the improved recovery of the fictitious common component, and therefore, better resolved vintages and time-lapse differences.

7.3.4 Time-lapse optimized mask generation

Based on the success of the JRM, we carry the argument of minimizing the SGR a step further by optimizing this quantity for the baseline/monitoring surveys. Because $\mathbf{Z}_0$ is observed by both surveys, the set of sampling points, $\{\mathbf{M}_0\}$, equals the union $\{\mathbf{M}_0\} = \{\mathbf{M}_1\} \cup \{\mathbf{M}_2\}$. When surveys are replicated, $\{\mathbf{M}_0\} = \{\mathbf{M}_1\} = \{\mathbf{M}_2\}$. However, $\mathbf{M}_0$ becomes larger when the baseline and monitor surveys are not replicated explaining why the common component is better resolved when the surveys are not replicated.

While Equation 7.1 leads to improved sampling masks for individual surveys, it does not exploit the fact that the common component is observed by all surveys. For this reason, we propose an optimization procedure with respect to $\mathbf{M}_1$ and $\mathbf{M}_2$ with an objective that also includes the SGR for the common component. To avoid generation of poor sampling masks, we follow a mini-max principle where the maximum—i.e., the ℓ_∞-norm—of the SGRs for the common and innovation components is minimized. To compensate for likely smaller SGRs for the common component when the surveys do not overlap ($\#\left\{\mathbf{M}_0\right\} > \#\left\{\mathbf{M}_1\right\}, \#\left\{\mathbf{M}_2\right\}$), we also introduce a scaling. We base this scaling on the property (see Definition 3.1 in Bhojanapalli and Jain 2014; Hoory, Linial, and Wigderson 2006) that the second singular value of d-regular graphs—i.e., seismic sampling masks with d non-zero entries per midpoint or offset— scales with $\sqrt{d}$. Given this scaling, we propose to minimize the following constrained optimization problem, for $j = 1, 2$:

$$\underset{\mathbf{M}_1, \mathbf{M}_2}{\text{minimize}} \quad \mathcal{L}(\mathbf{M}_1, \mathbf{M}_2) \quad \text{subject to} \quad \left\{\mathbf{M}_0\right\} = \left\{\mathbf{M}_1\right\} \cup \left\{\mathbf{M}_2\right\}, \mathbf{M}_j \in \mathcal{C}_j, \tag{7.3}$$

with $\mathcal{L}(\mathbf{M}_1, \mathbf{M}_2) = \left\| \left[\mathcal{L}(\mathbf{M}_0), \sqrt{\frac{\#(\mathbf{M}_1)}{\#(\mathbf{M}_0)}} \mathcal{L}(\mathbf{M}_1), \sqrt{\frac{\#(\mathbf{M}_2)}{\#(\mathbf{M}_0)}} \mathcal{L}(\mathbf{M}_2) \right] \right\|_\infty$. As before, the minimization is subject to constraints, $\mathcal{C}_j,\ j = 1, 2$, with a slight abuse of notation, representing the cardinality, binary mask, and maximum gap constraints for the baseline and the monitor surveys, respectively.

Figure 7.2 demonstrates deployment of the proposed time-lapse acquisition design and how to recover the fully sampled time-lapse data jointly. To calculate the optimized source locations for the baseline and monitor surveys, we solve the optimization problem in Equation 7.3. After collecting seismic traces with the optimized (by minimizing the SGRs) acquisition, we recover fully-sampled time-lapse data by inverting the system of equations included in Equation 7.2 with structure promotion (Kumar, Wason, Sharan, and Herrmann 2017). To produce time-lapse sampling masks, we employ simulated annealing as proposed by Zhang, Louboutin, Siahkoohi, Yin, Kumar, and Herrmann 2022 but with the following differences: *(i)* randomly perturbed masks are drawn for each survey independently; *(ii)* the compound objectives and constraints of Equation 7.3 are used; *(iii)* to be relocated sample points are allowed to move more freely than during jitter sampling, which allows us to better explore candidate sampling masks. Figure 7.3 illustrates how the algorithm progresses in very early iterations when initialized with a replicated jittered subsampled (removing 80% of the sources) acquisition. From Figure 7.2 and Figure 7.3a, we observe that the co-located source positions (denoted by the black dots) are gradually replaced by non-coincident source locations for the baseline (blue dots) and monitor surveys (red dots). Even though the objective of Equation 7.3 decreases non-monotonically (see Figure 7.3b), the reconstruction SNR increases for the baseline and monitor surveys for the selected points.

7.4 Numerical validation

To confirm the benefits of optimized acquisition, we consider time-lapse data, which differs by a complex gas cloud (Wason, Oghenekohwo, and Herrmann 2017; Jones, Edgar, Selvage, and Crook 2012). Using finite-differences (Witte, Louboutin, Kukreja, Luporini, Lange, Gorman, and Herrmann 2019; Louboutin, Witte, Yin, Modzelewski, and Herrmann 2022; Louboutin, Lange, Luporini, Kukreja, Witte, Herrmann, Velesko, and Gorman 2019; Luporini, Louboutin, Lange, Kukreja, Witte, Hückelheim, Yount, Kelly, Herrmann, and Gorman 2020), fully sampled (split spread) 2D baseline and monitor surveys are simulated each consisting of 300 sources/receivers sampled at $12.5\,\mathrm{m}$. By using a single jittered subsampling mask 80% of the sources are removed, yielding an average source sampling rate of $62.5\,\mathrm{m}$ with 100% overlap. After running $40,000$ iterations of the simulated annealing algorithm, the SGRs of the baseline/monitor surveys decreases from 0.346 to 0.268 and 0.262, respectively. The reduction in the overlap ratio (to 22%) leads to improvement in wavefield recovery via matrix completion (Kumar, Da Silva, Akalin, Aravkin, Mansour, Recht, and Herrmann 2015; Kumar, Wason, Sharan, and Herrmann 2017), which results in better SNRs for the baseline from $6.55\,\mathrm{dB}$ to $17.03\,\mathrm{dB}$ and for the monitor from $6.67\,\mathrm{dB}$ to $16.99\,\mathrm{dB}$, shown in Figure 7.4. For reasons explained by Oghenekohwo, Wason, Esser,

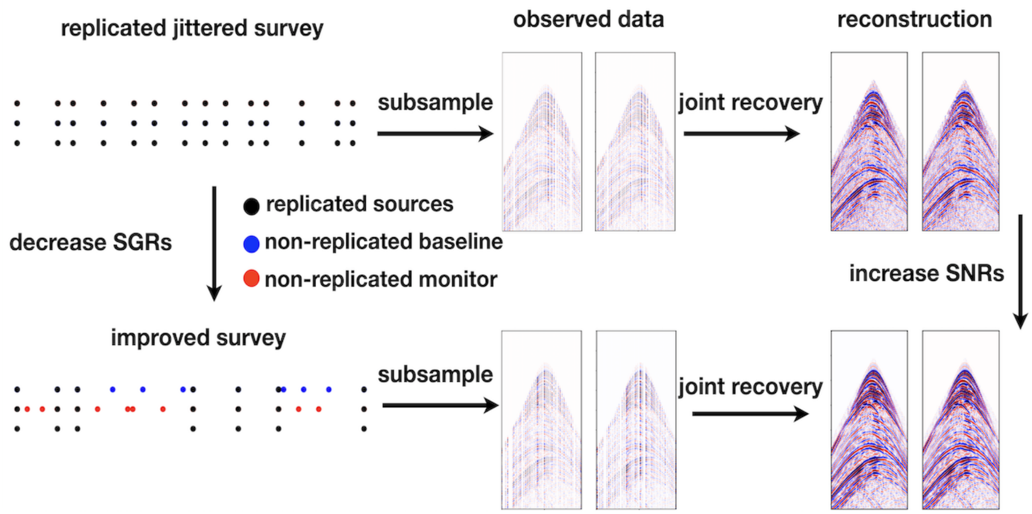
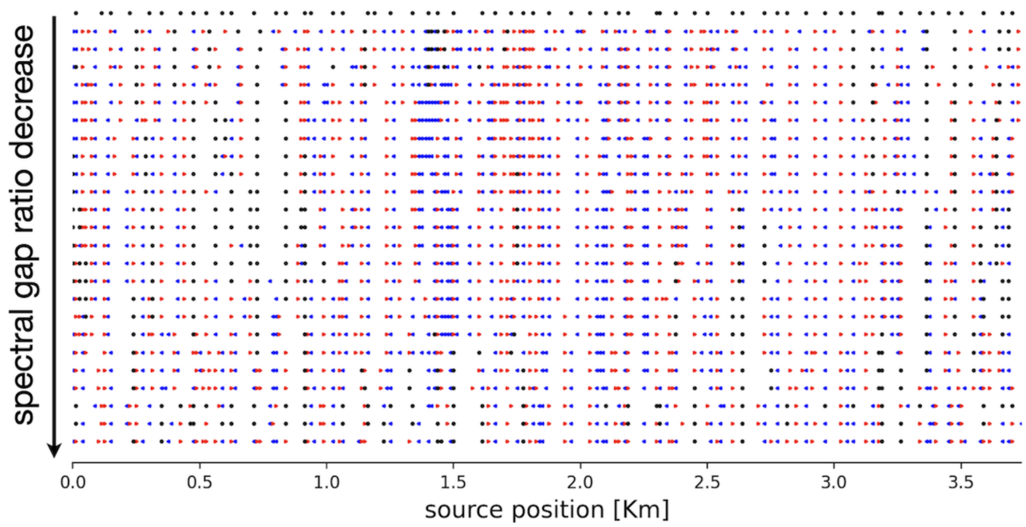
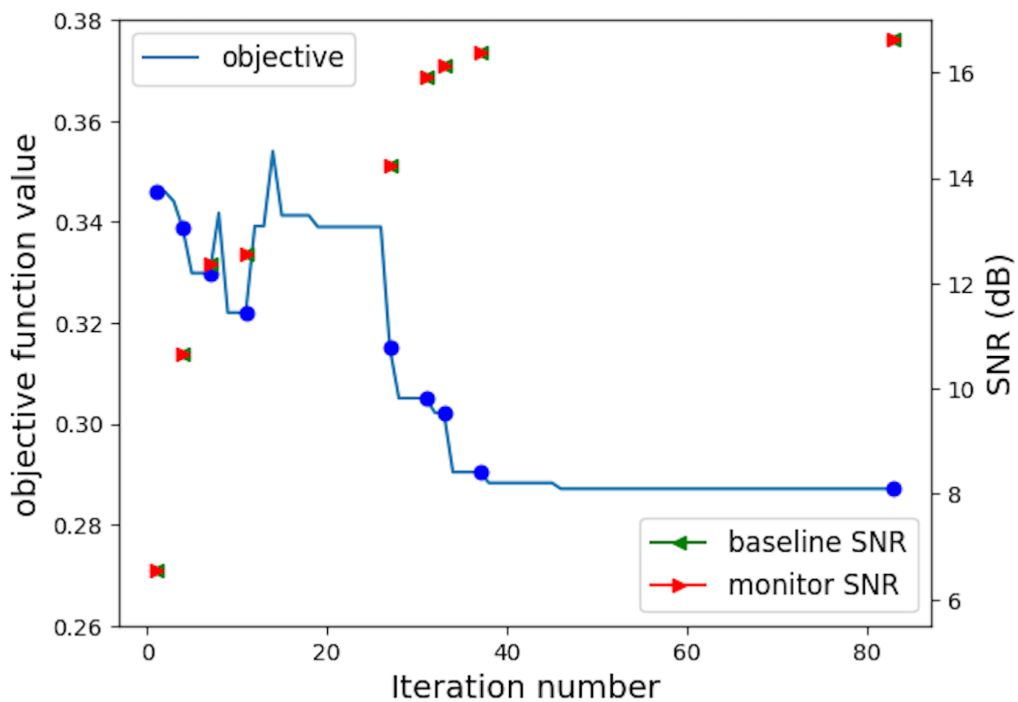


Figure 7.2: Illustration of proposed optimized time-lapse acquisition design. Optimized source locations for the baseline and the monitor surveys are calculated first with Equation 7.3, followed by collecting sparse samples in the field. Fully sampled time-lapse data are obtained by inverting the system of Equations in Equation 7.2 with structure promotion (Kumar, Wason, Sharan, and Herrmann 2017). The black dots of the sampling masks plotted on the left represent replicated source positions. The blue and red dots correspond to non-replicated source locations for the baseline and the monitor surveys. To better highlight replacement of co-located by non-coincident source locations after the SGR minimization, we only display a representative subset of the actual non-optimized and optimized source locations.



(a)



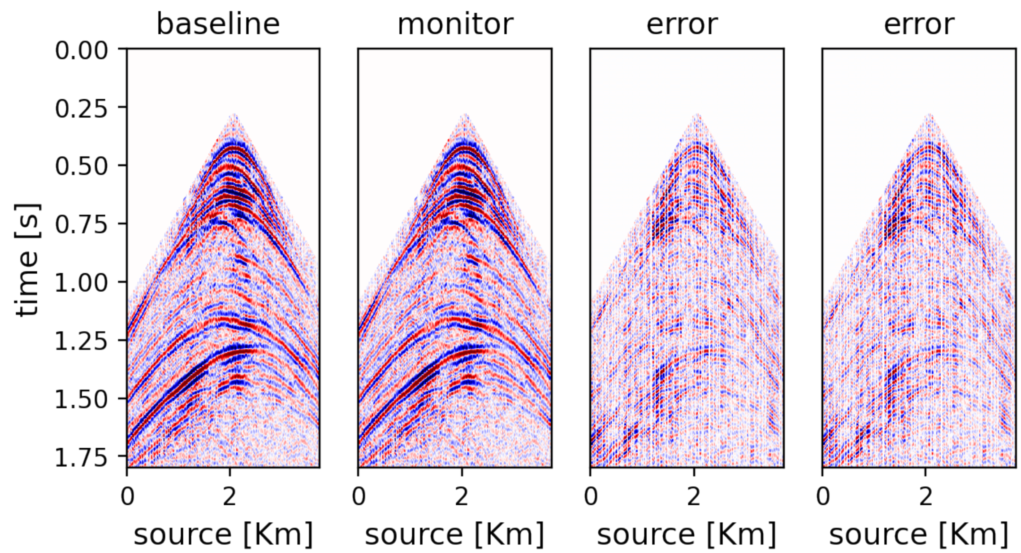
(b)

Figure 7.3: Automatic time-lapse sampling mask generation. (a) Starting from a jittered replicated sampling mask, the algorithm produces masks that have smaller SGRs but are no longer replicated. The overlap ratio decreases from 100% to 39%. The color scheme for markers remains consistent with Figure 7.2. (b) Non-monotonically decaying objective and reconstruction SNR evaluated at points where the objective decreased by more than 0.003.

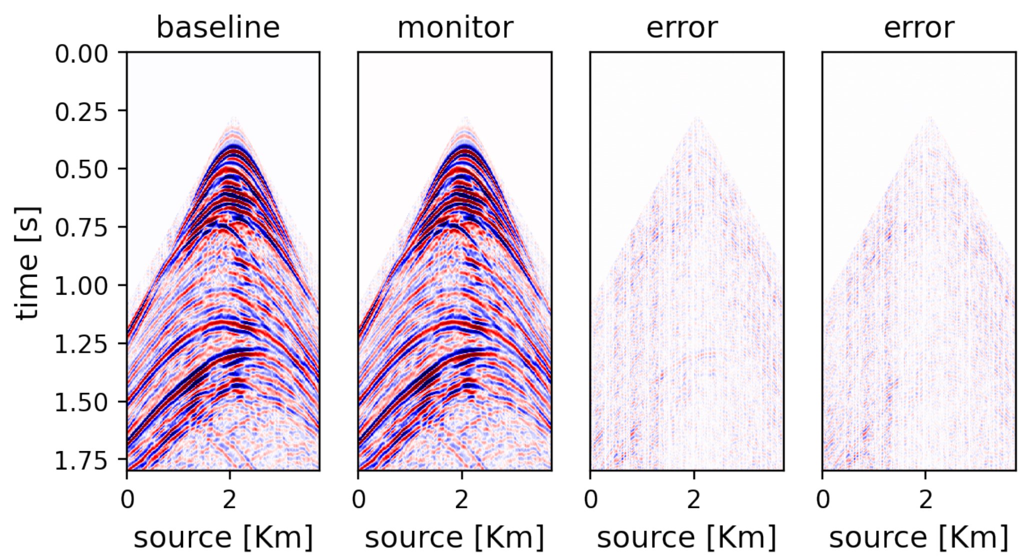
and Herrmann 2017 and Wason, Oghenekohwo, and Herrmann 2017, time-lapse difference plots are not included because the benefits of exact replication vanish when acquisition geometries undergo relatively small ($1-2$m) random shifts.

While these improvements are encouraging, the proposed optimization is approximate and the produced binary masks will be different for different starting masks. To investigate this effect, 30 overlapping jittered masks are generated by removing 75% of the sources. By reducing the overlap to $29\% \pm 8\%$, the optimized masks improve the SGRs as can be observed from the violin plots in Figure 7.5a. As before, the reductions in SGRs translate into improved SNRs as can be seen in Figure 7.5b. Compared to box plots, violin plots display the entire distribution including lines for the median (long dashes), first and third quartile (short dashes). We can make the following observations: *(i)* the SGRs for the baseline/monitor surveys decrease significantly; *(ii)* because of the larger number of sampled sources, the SGR for the common component is smaller and more narrowly distributed; *(iii)* the distribution of the SGRs of the baseline/monitor surveys is also narrow compared to the one of the initial jittered binary sampling masks; *(iv)* the SNRs for the recovered baseline/monitor surveys improve significantly.

Even though the above results are encouraging and consistent with published reports that claim benefits of the JRM (Oghenekohwo, Wason, Esser, and Herrmann 2017; Wason, Oghenekohwo, and Herrmann 2017; Yin, Erdinc, Gahlot, Louboutin, and Herrmann 2023), further scrutiny is in order. To this end, additional experiments were conducted to better understand robustness of the proposed methodology. Aside from predictable behavior for different starting masks (Figure 7.5), we also found that optimized SGRs are relatively insensitive to different runs of the simulated annealing algorithm and to random perturbations in the optimized masks. The first observation implies that while the simulated annealing algorithm may produce different masks, the SGRs remain very close, yielding wavefield reconstructions of near equal quality. The second observation indicates that postplot errors by single gridpoint shifts (12.5m) in the worst scenario offset the gains made by the opti-

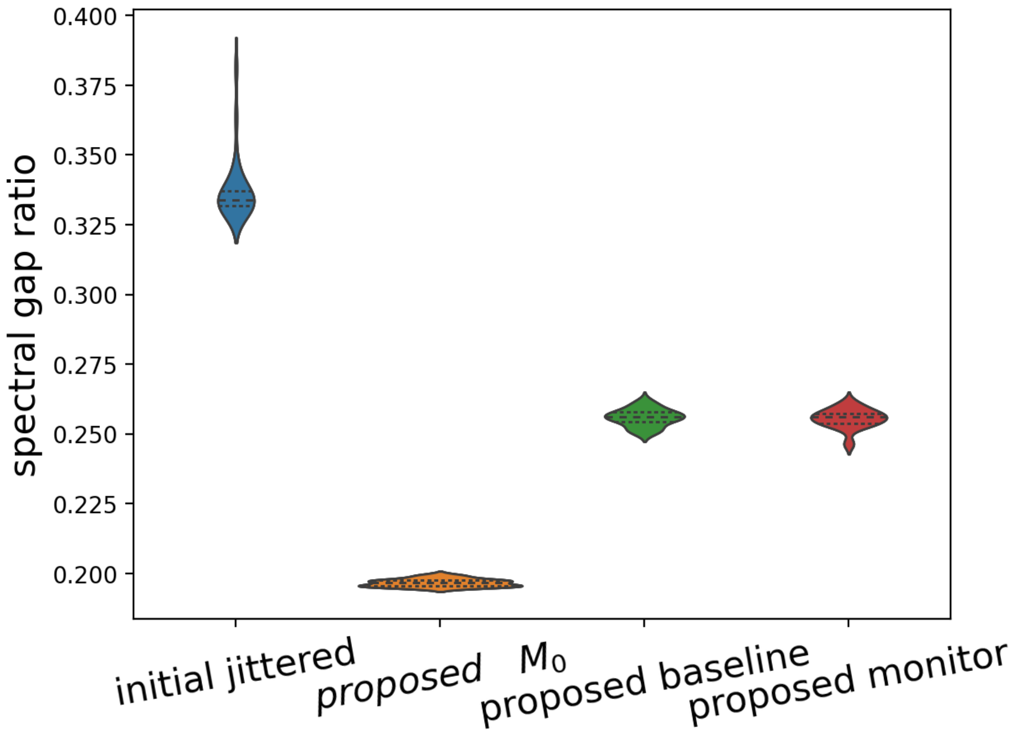


(a)

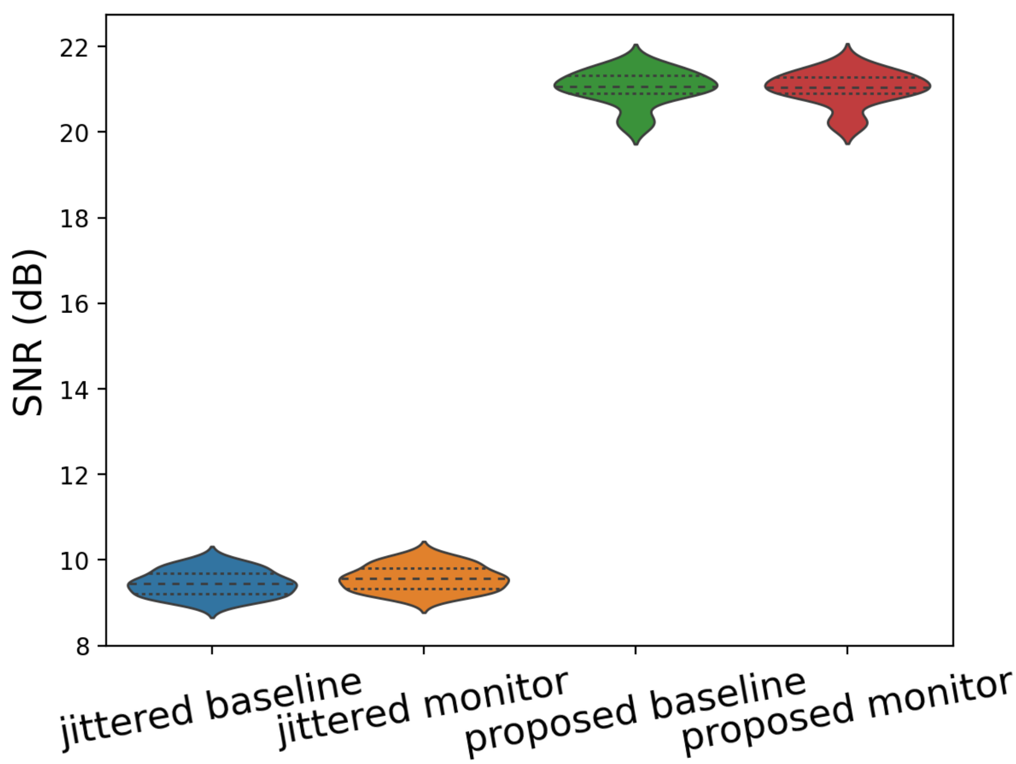


(b)

Figure 7.4: Time-lapse wavefield reconstruction in the time-domain. (a) wavefield reconstruction from 80% jittered subsampling for the baseline SNR = $6.55\,\mathrm{dB}$, monitor SNR = $6.67\,\mathrm{dB}$, and errors between the ground truth and the reconstructed wavefields. (b) the same but with optimized sampling masks, yielding improved recovery baseline/monitor surveys with SNR = $17.03,\ 16.99\mathrm{dB}$, respectively.



(a)



(b)

Figure 7.5: Violin plots for the SGRs (a) and recovery SNRs (b) for 30 independent experiments. These experiments show systematic reductions in SGR and significantly improves reconstruction SNRs for optimized surveys.

mization. However, on average improvements are mostly preserved although with higher variability.

The observed robustness of the presented method is consistent with reported behavior of the JRM. Even though we only considered the on-the-grid case, the argument can be made that improvements will carry over to the off-the-grid situation (Wason, Oghenekohwo, and Herrmann 2017; Oghenekohwo and Herrmann 2017; Lopez, Kumar, Yilmaz, and Herrmann 2016). However, to turn this claim into a more solid argument, we would have to extend the presented approach to the infinite-dimensional case, which is beyond the scope of this chapter.

7.5 Conclusions

Acquisition costs form a major impediment to time-lapse seismic. To reduce these costs while ensuring time-lapse repeatability, a novel acquisition optimization scheme was proposed that produces binary sampling masks that favor wavefield reconstruction with the joint recovery model. Optimized sampling masks were generated automatically by minimizing a new objective function consisting of spectral gap ratios for the baseline/monitor surveys and for the common component shared by the surveys. Aside from allowing for wave-simulation free, and therefore computationally feasible, optimized acquisition design, the proposed method also reaffirms the suggestion that deliberate relaxation of survey replication may lead to improved quality of jointly inverted surveys. This claim is solely based on connectivity arguments for the acquisition geometries associated with the baseline/monitor surveys and the common component. Because the spectral gap ratio is extremely cheap to evaluate, it lends itself very well to be extended to multiple monitoring surveys and to 3D. As long as the time-lapse acquisition geometries are relatively well calibrated—i.e. errors between actual and assumed geometries are small, our simulation-free survey design methodology also eliminates the need for cumbersome 4D processing to a large degree. It enables low-cost surveys, and utilizes the joint recovery model to accurately invert for fully sampled repeatable time-lapse data without insisting on replicating the surveys in the field. The recovered data can subsequently be imaged and inverted to extract changes in the reservoir's elastic properties. The proposed method should also be capable of accommodating pre-existing constraints in the field, including restricted areas where no source/receiver can be placed due to production platforms, to private properties, or due to governmental minimum source/receiver line distance regulations. In principle, these additional constraints can be incorporated to refine the search space of the simulated annealing algorithm. Off-the-grid acquisition geometries are also conducive to being improved by spectral gap ratios, but we will leave these extensions to future work.

CHAPTER 8
CONCLUSIONS

This chapter summarizes the primary contributions of the thesis, discusses its limitations, outlines potential expansions, and proposes future research.

Seismic imaging is crucial in various industries, especially in the exploration and management of hydrocarbon reservoirs. The quality of the acquired seismic data directly affects the image quality, including its spatial and temporal sampling. However, conventional acquisition systems are limited by operational, cost, and environmental considerations, resulting in the collection of sparse and under-sampled data. These under-sampled data cause aliasing artifacts during processing, which degrade the final imaging results. Therefore, wavefield reconstruction is essential for recovering these under-sampled seismic data.

Among the most promising techniques for wavefield reconstruction, matrix completion is computationally efficient and easy to implement by utilizing the low-rank structure of seismic data. However, low-rank factorization cannot accurately approximate monochromatic frequency slices at high frequencies, reducing the efficiency of this method. To address this limitation, this thesis contributes a recursively weighted matrix completion method for reconstructing dense seismic data at higher frequencies. Moreover, a computationally weighted matrix completion method is proposed, which reduces the computational cost of weighted matrix completion. To expand the applicability of this novel method to 3D seismic acquisition, this thesis proposes a parallelized alternating optimization technique for parallelizing the weighted low-rank factorization procedure.

In comparison to marine data, land data are contaminated by ground roll, which is a strong surface wave with a high spatial frequency content. This thesis contributes to a practical workflow that improves the land seismic wavefield recovery via weighted matrix completion. However, designing optimized masks to improve the quality of wavefield reconstruction with matrix completion remains a challenging task. To address this issue, this thesis proposes a simulation-free 2D and 3D seismic survey design that utilizes simulated annealing to optimize the spectral gap of the subsampling masks, thereby improving reconstructions. Additionally, the thesis proposes a method for optimizing the design of a time-lapse seismic survey by combining the spectral gap method and the joint recovery model to generate a sparse, non-replicated time-lapse survey that improves the reconstructions of each survey.

8.1 High frequency seismic data reconstruction using recursively efficient weighted matrix completion

In the first two chapters of this thesis (Chapter 2 and Chapter 3), I proposed a recursively efficient weighted matrix completion method to improve seismic reconstructions at higher frequencies. Aravkin, Kumar, Mansour, Recht, and Herrmann 2014 proposed a weighted matrix completion approach that considers the fact that matrix completion works well at low to mid frequencies and that row and column subspaces of adjacent frequencies exhibit some similarity. However, it should be noted that this method was only implemented to a certain degree of success because they did not justify the method or apply weighting recursively from lower to higher frequencies.

Thanks to the good performance of conventional matrix completion at lower frequencies, we can obtain useful prior information for the weighted matrix completion. The recursively weighted matrix completion is an efficient method that enables the accumulation of this advantage across frequencies. Furthermore, the full potential of the recursively weighted approach requires careful selection of the weight parameters, which indicate the correlation between row and column subspaces of adjacent frequencies.

The original weighted matrix completion method (Aravkin, Kumar, Mansour, Recht, and Herrmann 2014) requires a highly computational projection operation to be performed on each iteration. By relocating the weights to the data constraints of the original weighted matrix completion problem, we can efficiently solve an equivalent version of the original weighted formulation problem. Experiments on 2D field data demonstrate that our proposed method improves reconstructed results over the conventional method and the pair weighted method (Aravkin, Kumar, Mansour, Recht, and Herrmann 2014).

8.2 Large scale high-frequency wavefield reconstruction with recursively weighted matrix factorizations

In Chapter 4 of this thesis, I proposed a large-scale, high-frequency wavefield reconstruction based on parallel, recursively weighted matrix factorizations. The conventional method for reconstructing matrices can be easily parallelized across frequency slices, making it computationally efficient and scalable to large-scale seismic data. However, while our proposed recursively weighted matrix completion performs well with 2D seismic data, the interdependence between frequency slices prevented parallel processing across frequency slices. Furthermore, the introduction of recursively applied weighting matrices improves performance at high frequencies but prevents row-by-row and column-by-column parallelization of the alternating minimization procedure.

To address these challenges, we assume that weight matrices with higher weights are nearly diagonal. Under this assumption, we solved for individual rows of low-rank components in parallel using the recursively weighted technique, combining the alteration and decoupling strategies. As a result, our existing parallel, recursively weighted framework can partially exploit the advantages provided by the recursively weighted method.

Experiments on synthetic 3D seismic data demonstrate that our proposed method improves the reconstructed results compared to the conventional method. Additionally, a comparison of the conventional method, the effective method, and the parallel method demonstrates the effectiveness of the proposed parallel method.

8.3 A practical workflow for land seismic wavefield recovery with weighted matrix factorization

In Chapter 5 of this thesis, I propose a practical workflow for seismic wavefield recovery using weighted matrix factorization. While the weighted technique has been successfully applied to marine data, ground roll in land seismic data is known to degrade wavefield reconstructions by promoting structure. There are two potential explanations for this. Firstly, ground roll corresponds to slow-moving surface waves of the Rayleigh type, which frequently alias. Secondly, the reconstruction places greater emphasis on the high-amplitude ground roll than on the lower-amplitude body wave.

To reduce the effects of ground roll, I propose a practical workflow that incorporates a weighted matrix factorization scheme with the separation of the subsampled input data into ground roll and body wave components.

Tests on a subset of the synthetic 3D SEAM Barrett dataset, randomly sampled at a 25 percent receiver sampling rate, demonstrate that the proposed workflow is capable of producing significant enhancements.

8.4 A simulation-free seismic survey design by maximizing the spectral gap

In Chapter 6 of this thesis, I proposed a simulated annealing-based optimization strategy to improve the quality of wavefield reconstruction with matrix completion. To reduce costs, seismic data is increasingly collected using random subsampling, but optimal acquisition geometries continue to be a topic of ongoing study.

To address this challenge, we used the spectral gap, which is a measure of the connectedness of the graph in expander graph theory and has been used to predict the quality of wavefield reconstruction with a particular subsampling scheme in matrix completion theory. We proposed a simulation-free 2D and 3D seismic survey design that increases the spectral gaps of subsampling masks. Our proposed method effectively increases the spectral gap of the subsampling mask using synthetic 2D and 3D datasets. This increased spectral gap enhances the connectivity between sources and receivers, and as a result, improves wavefield reconstruction.

8.5 Optimized time-lapse acquisition design via spectral gap ratio minimization

In Chapter 7 of this thesis, I proposed a joint recovery model-based simulation-free time-lapse seismic survey design. Although time-lapse seismic has been successfully applied to CO2 sequestration monitoring, it remains a challenging problem due to the extremely high cost of replicated density field surveys. As previously mentioned, wavefield reconstruction based on matrix completion from randomized subsampled data is a cost-effective method for reducing expenses. This technique also enables precise time-lapse reconstruction by using the joint recovery model (JRM), which takes advantage of the fact that various vintages share a common component. However, combining JRM with optimal time-lapse acquisition survey design is an unexplored field of study.

To expand the seismic survey design based on the spectral gap to the time-lapse seismic survey design, we integrated the JRM with the optimal survey design. By designing the algorithm and selecting the parameters carefully, we demonstrate how the spectral gap from graph theory can be used to optimize low-cost sparse 4D seismic acquisition geometries to generate non-replicated time-lapse surveys. Tests on 4D synthetic seismic data confirm that the spectral gap increases from replicated to non-replicated time-lapse acquisition geometries, favoring wavefield recovery with the JRM.

8.6 Future research direction

There are several potential research directions to explore in the future, including the following:

1. Our recursively weighted method is computationally inexpensive and yields excel-

lent results for higher frequency slices. However, the current framework cannot automatically select key parameters such as the rank of the data to be recovered, the rank of the prior information, and the weights across frequencies. Because a constant parameter is used across all frequencies, the advantages of the recursively weighted method are only partially utilized. Thus, a future direction for research is to create an algorithm that autonomously selects these crucial parameters.

2. We developed a parallel, computationally effective version of the recursively weighted technique for large-scale seismic data. However, the present implementation of this parallel framework is weight-dependent and cannot operate with smaller weights, limiting the benefits of the weighted technique. Therefore, a potential topic for future research is the development of a computationally efficient variant of the recursively weighted method for smaller weights, so that the proposed method can realize its full potential.

3. We proposed a practical workflow to handle land seismic data and implemented our algorithm on subsets of the data to reduce computational costs. The number of iterations is a crucial parameter, with smaller subsets requiring fewer iterations and larger subsets requiring more. However, our algorithm cannot automatically select this vital parameter. Therefore, a future direction for research would be to develop an algorithm that automatically selects this crucial parameter.

4. This thesis proposes a seismic survey design free of simulations by optimizing the spectral gap ratio using simulated annealing. Although simulated annealing approximates the global optimal, a large number of iterations can be computationally intensive. Hence, discovering a computationally efficient algorithm to solve this optimization and further improve the outcome could be the subject of future study.

5. We proposed a design for a time-lapse seismic survey that includes the JRM and spectral gap. Our objective function takes into account both the subsampling masks for each vintage and those for the common components. However, we were unable to find a suitable method to describe the subsampling masks for 4D seismic data. Future research could focus on identifying the subsampling mask for 4D signals and incorporating it into our objective function to design a more effective time-lapse seismic survey.

Appendices

APPENDIX A
EXPERIMENTAL EQUIPMENT

In this section, we justify our parallel implementation of the weighted matrix completion problem. Beginning at equation Equation 4.2, our original weighted program, we will arrive at equations Equation 4.15 and Equation 4.16 which specify our implemented parallel counterpart.

Recall equation Equation 4.2

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

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

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

where

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

and

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

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

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

where we have defined

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

and

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

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

$$\begin{aligned} \bar{\mathbf{L}}, \bar{\mathbf{R}} := \quad & \arg\min_{\bar{\mathbf{L}}_\#, \bar{\mathbf{R}}_\#} \frac{1}{2} \left\| \begin{bmatrix} \bar{\mathbf{L}}_\# \\ \bar{\mathbf{R}}_\# \end{bmatrix} \right\|_F^2 \\ & \text{subject to} \\ & \|\mathcal{A}(\widehat{\mathbf{Q}}\bar{\mathbf{L}}_\# \bar{\mathbf{R}}_\#^H \widehat{\mathbf{W}}) - w_1 w_2 \mathbf{B}\|_F \le w_1 w_2 \epsilon, \end{aligned} \tag{A3}$$

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

$$\bar{\mathbf{R}}^k := \arg\min_{\bar{\mathbf{R}}_\#} \|\bar{\mathbf{R}}_\#\|_F^2 \quad \text{subject to} \quad \|\mathcal{A}(\widehat{\mathbf{Q}}\bar{\mathbf{L}}^{k-1}\bar{\mathbf{R}}_\#^H \widehat{\mathbf{W}}) - w_1 w_2 \mathbf{B}\|_F \le w_1 w_2 \epsilon, \tag{A4}$$

and upon output switches to optimize over the left factor

$$\bar{\mathbf{L}}^k := \arg\min_{\bar{\mathbf{L}}_\#} \|\bar{\mathbf{L}}_\#\|_F^2 \quad \text{subject to} \quad \|\mathcal{A}(\widehat{\mathbf{Q}}\bar{\mathbf{L}}_\# (\bar{\mathbf{R}}^k)^H \widehat{\mathbf{W}}) - w_1 w_2 \mathbf{B}\|_F \le w_1 w_2 \epsilon. \tag{A5}$$

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

Our next goal is to approximately solve problems Equation A5 and Equation A4 in a distributed manner, to be implemented in a parallel computing architecture. To this end, we apply our approximate commutative property, i.e., $\mathcal{A}(\widehat{\mathbf{Q}}\mathbf{Y}\widehat{\mathbf{W}}) \approx \mathcal{A}(\widehat{Q}\mathbf{Y})\widehat{\mathbf{W}}$ and $\mathcal{A}(\widehat{\mathbf{Q}}\mathbf{Y}\widehat{\mathbf{W}}) \approx \widehat{\mathbf{Q}}\mathcal{A}(\mathbf{Y}\widehat{\mathbf{W}})$ for large values of w_1 and w_2. Using these approximations, we obtain

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

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

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

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

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

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

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

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

The main advantage in computing iterates $\tilde{\mathbf{R}}^k, \tilde{\mathbf{L}}^k$, rather than $\bar{\mathbf{R}}^k, \bar{\mathbf{L}}^k$, is that these programs allow for a distributed implementation. The data matrices $\widehat{\mathbf{B}}_R$ and $\widehat{\mathbf{B}}_L$ in equations Equation A8 and Equation A7 are dense (have all non-zero entries) making computation expensive. However, when the weights $w_{1,2}$ are relatively large we observe that both dense matrices $\widehat{\mathbf{B}}_R$, $\widehat{\mathbf{B}}_L$ can be well approximated by the sparse observed data matrix $\mathbf{B}$. This leads to subproblems Equation 4.15 and Equation 4.16 and concludes our derivation.

REFERENCES

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

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

[3] E. J. Candès, J. Romberg, and T. Tao, "*Robust uncertainty principles: *Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006.

[4] 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*, vol. 33, no. 4, pp. 386–391, 2014.

[5] S. K. Chiu, "*3D attenuation of aliased ground roll on randomly undersampled data," in *SEG International Exposition and Annual Meeting*, OnePetro, 2019.

[6] R. Kumar, C. Da Silva, O. Akalin, A. Y. Aravkin, H. Mansour, B. Recht, and F. J. Herrmann, "*Efficient matrix completion for seismic data reconstruction," *Geophysics*, vol. 80, no. 5, pp. V97–V114, 2015.

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

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

[9] X. Liu, "*Ground roll supression using the Karhunen-Loeve transform," *Geophysics*, vol. 64, no. 2, pp. 564–566, 1999.

[10] K. Manohar, B. W. Brunton, J. N. Kutz, and S. L. Brunton, "*Data-driven sparse sensor placement for reconstruction: *Demonstrating the benefits of exploiting known patterns," *IEEE Control Systems Magazine*, vol. 38, no. 3, pp. 63–86, 2018.

[11] S. Bhojanapalli and P. Jain, "*Universal matrix completion," in *International Conference on Machine Learning*, PMLR, 2014, pp. 1881–1889.

[12] O. López, R. Kumar, N. Moldoveanu, and F. Herrmann, "*Graph Spectrum Based Seismic Survey Design," *arXiv preprint arXiv:2202.04623*, 2022.

[13] F. Oghenekohwo and F. J. Herrmann, "*Improved time-lapse data repeatability with randomized sampling and distributed compressive sensing," in *EAGE Annual Conference Proceedings*, vol. 6, 2017.

[14] H. Wason, F. Oghenekohwo, and F. J. Herrmann, "*Low-cost time-lapse seismic with distributed compressive sensing—Part 2: *Impact on repeatability," *Geophysics*, vol. 82, no. 3, P15–P30, 2017.

[15] D. Baron, M. F. Duarte, S. Sarvotham, M. B. Wakin, and R. G. Baraniuk, "*An information-theoretic approach to distributed compressed sensing," in *Proc. 45rd Conference on Communication, Control, and Computing*, 2005.

[16] F. Oghenekohwo and F. J. Herrmann, "*Highly repeatable time-lapse seismic with distributed compressive sensing—Mitigating effects of calibration errors," *The Leading Edge*, vol. 36, no. 8, pp. 688–694, 2017.

[17] Z. Yin, M. Louboutin, and F. J. Herrmann, "*Compressive time-lapse seismic monitoring of carbon storage and sequestration with the joint recovery model," in *SEG/AAPG/SEPM First International Meeting for Applied Geoscience & Energy*, OnePetro, 2021.

[18] Y. Tian, L. Wei, C. Li, S. Oppert, and G. Hennenfent, "*Joint sparsity recovery for noise attenuation," in *2018 SEG International Exposition and Annual Meeting*, OnePetro, 2018.

[19] L. Wei, Y. Tian, C. Li, S. Oppert, and G. Hennenfent, "*Improve 4D seismic interpretability with joint sparsity recovery," in *2018 SEG International Exposition and Annual Meeting*, OnePetro, 2018.

[20] F. O. Oghenekohwo, "*Economic time-lapse seismic acquisition and imaging-reaping the benefits of randomized sampling with distributed compressive sensing," Ph.D. dissertation, University of British Columbia, 2017.

[21] F. Oghenekohwo, H. Wason, E. Esser, and F. J. Herrmann, "*Low-cost time-lapse seismic with distributed compressive sensing—Part 1: *Exploiting common information among the vintages," *Geophysics*, vol. 82, no. 3, P1–P13, 2017.

[22] Y. Zhang, S. Sharan, and F. J. Herrmann, "*High-frequency wavefield recovery with weighted matrix factorizations," in *89th Annual International Meeting, SEG, Expanded Abstracts*, 2019, pp. 3959–3963.

[23] Y. Zhang, S. Sharan, O. Lopez, and F. J. Herrmann, "*Wavefield recovery with limited-subspace weighted matrix factorizations," in *SEG Technical Program Expanded Abstracts 2020*, Society of Exploration Geophysicists, 2020, pp. 2858–2862.

[24] Y. Zhang and F. J. Herrmann, "*A practical workflow for land seismic wavefield recovery with weighted matrix factorization," *arXiv preprint arXiv:2104.08155*, 2021.

[25] C. E. Jones, J. A. Edgar, J. I. Selvage, and H. Crook, "*Building Complex Synthetic Models to Evaluate Acquisition Geometries and Velocity Inversion Technologies," in *74th Conference and Exhibition, EAGE*, Jun. 2012.

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

[27] Y. Zhang, M. Louboutin, A. Siahkoohi, Z. Yin, R. Kumar, and F. J. Herrmann, "*A simulation-free seismic survey design by maximizing the spectral gap," in *Second International Meeting for Applied Geoscience & Energy*, Society of Exploration Geophysicists and American Association of Petroleum . . ., 2022, pp. 15–20.

[28] Y. Zhang, Z. Yin, O. Lopez, A. Siahkoohi, M. Louboutin, and F. J. Herrmann, "*3D seismic survey design by maximizing the spectral gap," Submitted to IMAGE 2023, Mar. 2023.

[29] Y. Zhang, Z. Yin, O. Lopez, A. Siahkoohi, M. Louboutin, R. Kumar, and F. J. Herrmann, "*Optimized time-lapse acquisition design via spectral gap ratio minimization," *Geophysics*, Apr. 2023, (Geophysics, just accepted).

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

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

[32] F. J. Herrmann, D. Wang, G. Hennenfent, and P. P. Moghaddam, "*Curvelet-based seismic data processing: *A multiscale and nonlinear approach," *Geophysics*, vol. 73, no. 1, A1–A5, 2007.

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

[34] 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," in *76th EAGE Conference and Exhibition 2014*, 2014, pp. 1–5.

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

[36] J. D. Rennie and N. Srebro, "*Fast maximum margin matrix factorization for collaborative prediction," in *Proceedings of the 22nd international conference on Machine learning*, ACM, 2005, pp. 713–719.

[37] E. van den Berg and M. Friedlander, "*Probing the Pareto Frontier for Basis Pursuit Solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2009.

[38] F. J. Herrmann, M. P. Friedlander, and O. Yilmaz, "*Fighting the curse of dimensionality: *Compressive sensing in exploration seismology," *IEEE Signal Processing Magazine*, vol. 29, no. 3, pp. 88–100, 2012.

[39] A. Y. Aravkin, R. Kumar, H. Mansour, and B. Recht, "*A robust SVD-free approach to matrix completion, with applications to interpolation of large scale data," 2013.

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

[41] O. Lopez, R. Kumar, O. Yilmaz, and F. J. Herrmann, "*Off-the-Grid Low-Rank Matrix Recovery and Seismic Data Reconstruction," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 658–671, 2016.

[42] S. Xu, Y. Zhang, D. Pham, and G. Lambaré, "*Antileakage Fourier transform for seismic data regularization," *GEOPHYSICS*, vol. 70, no. 4, pp. V87–V95, 2005. eprint: https://doi.org/10.1190/1.1993713.

[43] E. J. Candes, J. K. Romberg, and T. Tao, "*Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[44] D. L. Donoho, "*Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[45] B. Recht, M. Fazel, and P. A. Parrilo, "*Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

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

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

[48] M. P. Friedlander, H. Mansour, R. Saab, and Ö. Yilmaz, "*Recovering Compressively Sampled Signals Using Partial Support Information," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1122–1134, 2012.

[49] R. von Borries, C. J. Miosso, and C. Potes, "*Compressed Sensing Using Prior Information," *2007 2nd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, pp. 121–124, 2007.

[50] N. Vaswani and W. Lu, "*Modified-CS: *Modifying Compressive Sensing for Problems With Partially Known Support," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4595–4607, 2010.

[51] M. A. Khajehnejad, W. Xu, A. S. Avestimehr, and B. Hassibi, "*Weighted l1 minimization for sparse recovery with prior information," in *2009 IEEE International Symposium on Information Theory*, 2009, pp. 483–487.

[52] E. Candes, M. Wakin, and S. Boyd, "*Enhancing Sparsity by Reweighted l1 Minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, 2008.

[53] Y. Xu and W. Yin, "*A Block Coordinate Descent Method for Regularized Multiconvex Optimization with Applications to Nonnegative Tensor Factorization and Completion," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1758–1789, 2013.

[54] 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*, 2013, pp. 665–674.

[55] O. Lopez, R. Kumar, and F. J. Herrmann, "*Rank minimization via alternating optimization: *seismic data interpolation," in *77th Conference and Exhibition, EAGE*, Jun. 2015.

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

[57] E. J. Candes and B. Recht, "*Exact Matrix Completion via Convex Optimization," *Foundations of Computational Mathematics*, vol. 9, p. 717, 2009.

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

[59] L. Demanet, "*Curvelets, Wave Atoms, and Wave Equations," Ph.D. dissertation, California Institute of Technology, May 2006.

[60] P. A. Witte, M. Louboutin, H. Modzelewski, C. Jones, J. Selvage, and F. J. Herrmann, "*Event-driven workflows for large-scale seismic imaging in the cloud," in *89th Annual International Meeting, SEG, Expanded Abstracts*, 2019, pp. 3984–3988.

[61] M. D. Sacchi, T. J. Ulrych, and C. J. Walker, "*Interpolation and extrapolation using a high-resolution discrete Fourier transform," *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 31–38, 1998.

[62] Ö. Yilmaz, *Seismic data analysis: Processing, inversion, and interpretation of seismic data*. Society of exploration geophysicists, 2001.

[63] D. Trad, T. Ulrych, and M. Sacchi, "*Latest views of the sparse Radon transform," *Geophysics*, vol. 68, no. 1, pp. 386–399, 2003.

[64] B. Bahia, I. Papathanasaki, and M. D. Sacchi, "*Ground-roll attenuation through quaternionic inversion with sparsity constraints," in *SEG Technical Program Expanded Abstracts 2020*, Society of Exploration Geophysicists, 2020, pp. 3254–3258.

[65] J. Tan, T. Li, F. Jarrah, K. Lee, R. Holt, N. V. D. Coevering, and K. Koster, "*SEAM Phase II Barrett model classic data study: *Processing, imaging, and attributes analysis," in *SEG Technical Program Expanded Abstracts 2019*, Society of Exploration Geophysicists, 2019, pp. 3840–3844.

[66] N. Van De Coevering, K. Koster, and R. Holt, "*A sceptic's view of VVAz and AVAz," in *SEG Technical Program Expanded Abstracts 2019*, Society of Exploration Geophysicists, 2019, pp. 3230–3234.

[67] O. Lopez, R. Kumar, and F. J. Herrmann, "*Rank minimization via alternating optimization-seismic data interpolation," in *77th EAGE Conference and Exhibition 2015*, European Association of Geoscientists & Engineers, vol. 2015, 2015, pp. 1–5.

[68] H. Gilles, "*Sampling and reconstruction of seismic wavefields in the curvelet domain," Ph.D. dissertation, University of British Columbia, 2008.

[69] F. J. Herrmann, D. Wang, G. Hennenfent, and P. P. Moghaddam, "*Curvelet-based seismic data processing: *A multiscale and nonlinear approach," *Geophysics*, vol. 73, no. 1, A1–A5, 2008.

[70] E. J. Candès and T. Tao, "*The power of convex relaxation: *Near-optimal matrix completion," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2053–2080, 2010.

[71] C. Li, S. T. Kaplan, C. C. Mosher, J. D. Brewer, and R. G. Keys, **Compressive sensing*, US Patent 9,632,193, Apr. 2017.

[72] A. Titova, M. B. Wakin, and A. Tura, "*Mutual coherence in compressive sensing seismic acquisition," in *SEG International Exposition and Annual Meeting*, OnePetro, 2019.

[73] Y. Guo and M. D. Sacchi, "*Data-driven time-lapse acquisition design via optimal receiver-source placement and reconstruction," in *SEG Technical Program Expanded Abstracts 2020*, Society of Exploration Geophysicists, 2020, pp. 66–70.

[74] R. Kumar, "*Enabling large-scale seismic data acquisition, processing and waveform-inversion via rank-minimization," Ph.D. dissertation, University of British Columbia, 2017.

[75] O. López, R. Kumar, N. Moldoveanu, and F. J. Herrmann, "*Spectral Gap-Based Seismic Survey Design," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[76] B. Li, A. P. Petropulu, and W. Trappe, "*Optimum co-design for spectrum sharing between matrix completion based MIMO radars and a MIMO communication system," *IEEE Transactions on Signal Processing*, vol. 64, no. 17, pp. 4562–4575, 2016.

[77] S. Kirkpatrick, C. D. Gelatt Jr, and M. P. Vecchi, "*Optimization by simulated annealing," *science*, vol. 220, no. 4598, pp. 671–680, 1983.

[78] C. Jones, J. Edgar, J. Selvage, and H. Crook, "*Building complex synthetic models to evaluate acquisition geometries and velocity inversion technologies," in *74th EAGE Conference and Exhibition incorporating EUROPEC 2012*, European Association of Geoscientists & Engineers, 2012, cp–293.

[79] D. Fenati and F. Rocca, "*Seismic reciprocity field tests from the Italian Peninsula," *Geophysics*, vol. 49, no. 10, pp. 1690–1700, 1984.

[80] R. Şahin, K. Ertoğral, and O. Türkbey, "*A simulated annealing heuristic for the dynamic layout problem with budget constraint," *Computers & Industrial Engineering*, vol. 59, no. 2, pp. 308–313, 2010.

[81] P. J. Van Laarhoven and E. H. Aarts, "*Simulated annealing," in *Simulated annealing: Theory and applications*, Springer, 1987, pp. 7–15.

[82] D. Abramson, M. Krishnamoorthy, H. Dang, *et al.*, "*Simulated annealing cooling schedules for the school timetabling problem," *Asia Pacific Journal of Operational Research*, vol. 16, pp. 1–22, 1999.

[83] X.-Q. Ma, "*Simultaneous inversion of prestack seismic data for rock properties using simulated annealing," *Geophysics*, vol. 67, no. 6, pp. 1877–1885, 2002.

[84] A. Assad and K. Deep, "*A hybrid harmony search and simulated annealing algorithm for continuous optimization," *Information Sciences*, vol. 450, pp. 246–266, 2018.

[85] O. Olorunda and A. P. Engelbrecht, "*Measuring exploration/exploitation in particle swarms using swarm diversity," in *2008 IEEE congress on evolutionary computation (IEEE world congress on computational intelligence)*, IEEE, 2008, pp. 1128–1134.

[86] F. J. Herrmann, "*Randomized sampling and sparsity: *Getting more information from fewer samples," *Geophysics*, vol. 75, no. 6, WB173–WB187, 2010.

[87] W. Tang, J. Ma, and F. J. Herrmann, "*Optimized Compressed Sensing for Curvelet-based Seismic Data Reconstruction," *CiteSeer*, 2008.

[88] R. Obermeier and J. A. Martinez-Lorenzo, "*Sensing matrix design via mutual coherence minimization for electromagnetic compressive imaging applications," *IEEE Transactions on Computational Imaging*, vol. 3, no. 2, pp. 217–229, 2017.

[89] Z. Yin, H. T. Erdinc, A. P. Gahlot, M. Louboutin, and F. J. Herrmann, "*Derisking geologic carbon storage from high-resolution time-lapse seismic to explainable leakage detection," *The Leading Edge*, vol. 42, no. 1, pp. 69–76, 2023.

[90] S. Hoory, N. Linial, and A. Wigderson, "*Expander graphs and their applications," *Bulletin of the American Mathematical Society*, vol. 43, no. 4, pp. 439–561, 2006.

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

[92] M. Louboutin, P. Witte, Z. Yin, H. Modzelewski, and F. Herrmann, "*slimgroup/JUDI.jl: *v3.2.1," version v3.2.1, Dec. 2022.

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

[94] F. Luporini, M. Louboutin, M. Lange, N. Kukreja, P. Witte, J. Hückelheim, C. Yount, P. H. Kelly, F. J. Herrmann, and G. J. Gorman, "*Architecture and performance of Devito, a system for automated stencil computation," *ACM Transactions on Mathematical Software (TOMS)*, vol. 46, no. 1, pp. 1–28, 2020.