

**SEISMIC IMAGING WITH EXTENDED IMAGE VOLUMES AND SOURCE
ESTIMATION**

A Dissertation
Presented to
The Academic Faculty

By

Mengmeng Yang

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Geophysics

Georgia Institute of Technology

May 2020

Copyright © Mengmeng Yang 2020

**SEISMIC IMAGING WITH EXTENDED IMAGE VOLUMES AND SOURCE
ESTIMATION**

Approved by:

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

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

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

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

Dr. Tristan van Leeuwen
Department of Mathematics
Utrecht University

Date Approved: April 22, 2020

To my family, for your endless love and support.

ACKNOWLEDGEMENTS

First and foremost, I would like to give my sincere gratitude to my supervisor Professor Dr. Felix J. Herrmann. I am greatly appreciative of the precious opportunity to continue as one PhD student in such a world-class geophysical lab and. His gorgeous supports and guidance in motivation and knowledge during my research are essential for me to finish my PhD. I get benefits from his training in critical thinking and open-minded cooperation, which I believe will continue to influence me profoundly in all aspects of my life. I also appreciate the opportunities he provided for me to present my research and communicate with the international geophysical communities via conferences.

My gratitude also goes to Professor Dr. Zhigang Peng, Professor Dr. Andrew Newman from Georgia Institute of Technology, Professor Dr. Micheal Bostock, Professor Dr. Ozgur Yilmaz from University of British Columbia for serving as my supervisory committee with their gorgeous time and advice. Thanks also go to Dr. Tristan Van Leeuwen and Professor Dr. James McClellan for their valuable inputs and attendance in my defense as the external examiners.

I would like to express my thanks to all my colleagues at the SLIM lab for their support and friendship. I would like to thank Henryk Modzelewski, Miranda Joyce and Diana Ruiz for their professional assistance. I would also like to thank all students and researchers in SLIM, especially my coauthors, Dr. Zhilong Fang, Dr. Rajiv Kumar, Dr. Rongrong Wang, Dr. Marie Graff, Philipp Witte, and Dr. Emmanouil Daskalakis for their valuable scientific insights as collaborators. Also, I appreciate the help of Dr. Ning Tu, Dr. Xiang Li, Dr. Ernie Esser (1980-2015), Mathias Louboutin, and other members of the SLIM group.

My grateful appreciation goes to Dr. Xin Cheng for offering me the opportunity of Internship with Schlumberger. I also would like to thank Dr. James Xu for mentorship at Schlumberger and everybody in the group of Migrations and FWI in Schlumberger for such an enjoyable experience.

I would also like to thank the authors of Devito, JOLI and Julia based on which the inversion algorithms are built up for all my thesis, and the authors of Marmousi model and Sigsbee2A model which I used in chapter 2, 3, 4, 5. Thanks to the author of software iWave which I used in chapter 5. Appreciation goes to Delphi for releasing the SRME package to us, which I used in chapter 5.

Many thanks to my family: my parents and my sister for their selfless and endless love and support. My parents always act as my shield no matter how hard the situation is. My sister was also pursuing her PhD degree and acting as my spiritual comrade via sharing her experiences with me. I would also like to thank the love and support from my fiance who spent the hardest time with me. I express my appreciation and love to Mrs. Marnie Tamaki and Mr. David Winter who spent more than four years with me and love me as one family member of them in Vancouver. All praise, honour and glory to my Lord Jesus Christ for His richest grace and mercy for the accomplishment of this thesis.

This work was in part of financially supported by several institutions and grants that I would like to acknowledge, they are the Natural Sciences and Engineering Research Council of Canada Collaborative Research and Development Grant (DNOISE II), members of the SINBAD consortium, as well as the Georgia Research Alliance and the Georgia Institute of Technology.

TABLE OF CONTENTS

Acknowledgments	v
List of Tables	xi
List of Figures	xii
Chapter 1: Introduction	1
1.1 Objectives	8
1.2 Outline	8
Chapter 2: Low-rank recovery of subsurface extended image volumes based on time-stepping propagator and power scheme	11
2.1 Summary	11
2.2 Introduction	12
2.3 Full subsurface monochromatic extended image volumes	15
2.3.1 Extended image volumes with Helmholtz	15
2.3.2 Low-rank factorization of time-harmonic EIVs	17
2.4 Full subsurface extended image volumes based on time-stepping	22
2.4.1 Time-domain EIVs	22
2.4.2 Time-domain probing	24
2.4.3 Low-rank representation via rSVD	26

2.5	Low-rank factorization with the power method	26
2.5.1	Power schemes	27
2.5.2	Complexity analysis for power scheme based rSVDs	29
2.5.3	Numerical experiments for power schemes	31
2.6	Slicing and dicing	36
2.7	Numerical experiments	40
2.8	Discussion	48
2.9	Conclusion	51
2.10	Appendix	52
Chapter 3: Low-rank recovery for extended image volumes via invariance relationship		55
3.1	summary	55
3.2	Introduction	55
3.3	Invariance relationship of EIV	58
3.3.1	Monochromatic invariance relationship	58
3.3.2	Velocity continuation in the time domain	60
3.3.3	Power scheme based rSVD plus invariance relationship together	61
3.4	Numerical experiments	61
3.5	Discussion	72
3.6	Conclusion	73
Chapter 4: Time-domain sparsity promoting least-squares reverse time migration with source estimation		74
4.1	Summary	74

4.2	Introduction	75
4.3	From RTM to LS-RTM	77
4.4	Stochastic optimization with sparsity promotion	79
4.5	On-the-fly source estimation	82
4.6	Numerical experiments	84
4.6.1	Stylized example 1	85
4.6.2	Stylized example 2	85
4.6.3	Experiments on the modified Marmousi model	87
4.6.4	Experiments on Sigsbee model	96
4.7	Discussion	105
4.8	Conclusion	106
Chapter 5: Sparsity-promoting least-squares reverse time migration with multiples		108
5.1	Summary	108
5.2	Introduction	109
5.3	SRME and areal source	111
5.3.1	EPSI	114
5.3.2	Areal source and its adaption into linearized modeling	116
5.4	Joint inversion of multiples and primaries based on LB	117
5.5	Numerical experiments	118
5.5.1	Test on linear data set in time-harmonic domain	119
5.5.2	Test on nonlinear data set in the time domain	120
5.6	Discussion	130

Chapter 6: Removing density effects in least square reverse time migration with a low-rank matched filter	131
6.1 Summary	131
6.2 Introduction	131
6.3 Methodology	133
6.4 Numerical experiments	137
6.5 Conclusion	144
Chapter 7: Conclusion	145
7.1 Low-rank recovery of subsurface extended image volumes based on power-schemed rSVD and mapping via the invariance relationship	145
7.2 Time-domain sparsity promoting least-squares reverse time migration with source estimation	147
7.3 Sparsity-promoting least-squares reverse time migration with multiples and removing density effects with a low-rank matched filter	147
7.4 Current limitations and future work	148
References	159

LIST OF TABLES

2.1	Storage and computational cost for rSVD with simultaneous iterations . . .	29
2.2	Storage and computational cost for rSVD with block Krylov iteration . . .	30

LIST OF FIGURES

1.1	Marine surveys, (a) using hydrophones,(b) using OBN.	2
1.2	Seismic workflow for marine data	4
1.3	Illustration of illuminations. Red stars represent the source and yellow triangles represent hydrophones. The illuminated areas are marked in yellow. The first blue layer denotes the water column, and the arrowhead lines indicate the incident and reflected rays.	6
1.4	Illustration the phantoms from multiples in RTM for a simple two-layer model with only one interface. (a) RTM with only primaries, (b) RTM with both primaries and multiples	7
2.1	Extended image volume computed for the small Marmousi model for all frequencies [5, 50]Hz with step 0.5Hz: (a) full EIV, (b) RTM image from the reshaped diagonal of the EIV, (c) common image-point gather at (7110m, 480m) from the 3987 th column of the EIV, (d) common image-point gather at (7360m, 300m) from the 6494 th column of the EIV.	20
2.2	Extended image volume computed for the small Marmousi model (Figure 2.1) at 5Hz: (a) singular value decay of the data matrix, (d) singular value decay of the corresponding EIV matrix.	21
2.3	Estimated rank for the extended image volume computed for the small Marmousi model with respect to the frequency [5, 50] Hz with step 0.5 Hz. This figure shows the rank of the data matrix D and the EIV E when truncated at 1%, 5%, 10% of the highest singular values.	22

2.4	(a) The singular value bars obtained by SVD, rSVD, SI and BKI with $n_p = 8$. (b) The corresponding relative errors of the singular values obtained by rSVD, SI and BKI. (c) The singular value bars obtained by SVD, rSVD, SI and BKI with $n_p = 16$, (d) the corresponding relative errors of the singular values obtained by rSVD, SI and BKI. (e) The singular value bars obtained by SVD, rSVD, SI and BKI with $n_p = 30$. (f) The corresponding relative errors of the singular values obtained by rSVD, SI and BKI.	34
2.5	Comparison between RTMs obtained via carrying out the conventional SVD(a), the rSVD(b),and BKI for $q = 1$. We only use the first eight singular values – i.e. $n_p = 8 \ll n_s$ where $n_s = 100$. (d) The difference between (a) and (b), (e) the difference between (a) and (c).	35
2.6	(a) shows the low-rank representation \mathbf{L} and \mathbf{R} and organize the 4D EIV into 2D matrix. The diagonal dash line indicates where the traditional RTM images extracted from. (b) presents the \mathbf{L} and \mathbf{R} by 3D cubes whose third dimension is the probing size n_p , and the 3D sub-cube of EIV is $\mathbf{E}[:, i_x; :, :]$ by fixing x dimension with i_x , where the green cross section $\mathbf{E}[i_z, i_x; :, :]$ indicates one CIP gather at i_z, i_x , and the red slice $\mathbf{E}[i_z, i_x; j_z = i_z, :], i_z = 1 \cdots n_z$ indicates one CIG gather at i_x along all the depth.	38
2.7	schematic diagram for stacking power on the slice $\mathbf{E}[i_z, i_x; :, :]$ within d offset.	40
2.8	Selected rank to capture 95% of the data’s energy as a function of frequency. 42	42
2.9	Comparison RTM images. (a) The true conventional RTM image obtained by migration 650 shot records and (b) the RTM image computed with a factorization with the BKI method for $q = 1$ and $n_p = 130$	43
2.10	Comparison CIP images at ($z = 870m, x = 5250m$). (a) The true CIP image via probing with a bandwidth-limited point source and (b) the CIP image extracted with Algorithm 3.	44
2.11	The exact and approximate CIGs at $x = 1.8, 3.6, 5.4\text{km}$, where the offset range is from -150 to 150m . (a) The merged exact CIGs, (b) the recovered CIGs from the low-rank representation via Algorithm 4.	45
2.12	(a) The original CIG image at $x = 5.4\text{km}$ with offset range 150m , (b) the stack power curves at depth $z = 0.432, 1.44, 2.07\text{km}$, (d) the corrected CIG image with the dips from stack power.	47

3.1	The background models used as different velocity scenarios for velocity continuation test. (a) The initial guess of the background model where the top of the salt is not correct, (b) the kinematically correct background model.	64
3.2	Selected rank to capture 95% of the data's energy as a function of frequency.	65
3.3	The RTMs extracted from the recovered low-rank representations of the EIV from (a) the initial guess model, (b) the kinematical correct model, (c) the RTM extracted from the recovered EIV that mapped from the initial guess model to the correct model.	67
3.4	The CIP images extracted from the recovered EIVs at $x = 2640\text{m}$, $z = 1590\text{m}$. (a) the CIP image of the EIVs from the initial guessed model, (b) the CIP image of the EIVs from the correct model, (c) the CIP image of the EIVs via mapping from the initial guess to the correct model.	69
3.5	The CIGs extracted from the low-rank representations with (a) the initial guess of the background model (as in Figure 3.1a), (b) the kinematically correct background model (as in Figure 3.1b), (c) via invariance relationship mapping from the initial background model to the correct background model.	71
4.1	Comparison between the solutions of randomized SPGL1 and LB with different block sizes: figure (a) shows the results of SPGL1 with different block sizes. For small block size, e.g. 20% or 10% of the whole size of \mathbf{A} , randomized SPGL1 obtains results with more noisy elements at wrong locations whereas less amplitudes at the correct locations, indicating the failure in recovering the accurate solutions; Figure (b) shows that LB converges to the correct solution for all the selections of the block size.	86
4.2	Comparison of solutions obtained with the LB iterations (see Algorithm 8) for a fixed true source (denoted by the blue line) and for on-the-fly source estimations with and without penalties. We obtained results with five passes through the data. Our method is well capable of estimating the "reflectivity" (a) and "source function" (b) after normalizing the ℓ_2 norm. The proposed method is also robust with respect to additive noise as we can see in (c) and (d). We added 10% Gaussian noise.	88
4.3	Comparison true, initial, and estimated source time functions (\mathbf{q}_0 , $\hat{\mathbf{q}} = \mathbf{w}_{k+1} * \mathbf{q}_0$) and their associated amplitude spectra. (a) the time signatures and (b) the frequency spectra. The estimated source time functions and spectra were obtained from noise-free data and from data to which zero-centered Gaussian noise was added with energy ranging from 50% and 200% of the simulated linear data.	90

4.4	(a) and (b) are the RTM images of Marmousi with true source and initial guess of source respectively.	91
4.5	Inverted images with (a) the true source and (b) initial source, generated by the Marmousi model.	94
4.6	Inverted images with source estimation (Algorithm 8) for (a) noise-free data, (b) data with 50% noise and (c) data with 200% noise, generated from the Marmousi model.	96
4.7	Convergence plots for the relative residual error (a)—i.e., the ℓ_2 -norm error between the observed data and the demigrated data for estimated reflectivity $\hat{\delta\mathbf{m}}$ and the relative model error (b)—i.e., the ℓ_2 -norm error between the true reflectivity and the recovered reflectivity.	97
4.8	Comparison true, initial, and estimated source time functions (\mathbf{q}_0 , $\hat{\mathbf{q}} = \mathbf{w}_{k+1}\mathbf{q}_0$) and their associated amplitude spectra. (a) the time signatures and (b) the frequency spectra. The estimated source time functions and spectra were obtained during the first five iterations with the conventional imaging condition.	100
4.9	Comparison between SPLS-RTM with the true source wavelet and SPLS-RTM with on-the-fly source estimation. In both cases, we did total 20 iterations amounting to two data passes.	103
4.10	Trace by trace comparisons between the true model perturbation, the images from SPLS-RTM with the true source and with on-the-fly source estimation. The traces in (a) and (b) are extracted from lateral positions $x = 4.5\text{km}$ and $x = 11.3\text{km}$, respectively	104
5.1	The generation of different ordered reflections. The star and the triangles floating in the water volume stand for the source and receivers respectively. From left to right, the lines with arrowhead in different colors indicate the reflected paths of primaries and multiples with higher and higher order.	112
5.2	One shot gather that contains multiples, obtained by the marine survey shown in Figure 5.1. From top to bottom, different colored arrows mark out the reflection events of different orders.	112
5.3	Imaging workflow with SRME demultiples. The primaries \mathcal{P}_0 , as the input in the looping migration-demigration sub-workflow, are predicted by SRME	113

5.4	Example for SRME's failure in shallow water. (a) The designed total up-going data that generated by summing the designed primaries in (b) with its one-order higher data. (c) The predicted primaries by applying the ideal multiples to the adaptive subtraction.	115
5.5	Primaries and the Green's function sketched by the green lines	116
5.6	RTM images for linear data sets. (a) The RTM image for only primaries; (b) The RTM image for the total deghosted up-going wavefield with areal source.	121
5.7	SPLS-RTM images for linear data sets. (a) is the SPLS-RTM image for only primaries; (b) is the SPLS-RTM image for the total deghosted up-going wavefield with areal source.	122
5.8	The mask applied on the acquisition geometry and the masked data.	123
5.9	SPLS-RTM images for the masked linear data sets. (a) is the SPLS-RTM image for the masked primaries; (b) is the SPLS-RTM image for the masked total deghosted up-going wavefield with areal source.	124
5.10	Slowness square background model modified from Sigsbee2A.	125
5.11	One shot record of areal source.	125
5.12	(a) and (b) Images by the total up-going data without and with areal source injection, respectively. (c) Differences between (a) and (b).	127
5.13	Comparison between shot gathers. (a) The deghosted total up-going data, (b) the ideal primaries with absorbing surface and dipole sources mirrored about the surface. (c) The recovered primaries by the inverted δm with absorbing surface and dipole sources mirrored about the surface. (d) The differences between (b) and (c). The red and block arrows indicate that the reflections in the recovered primaries are weaker or stronger than those in the ideal primaries, respectively. (e) The ideal multiples that got by subtracting the synthetic primaries with dipole sources from the deghosted total up-going data. (f) The recovered multiples that generated by areal source.	129
6.1	Quasi layered model with faults. (a, b, c) background velocity, the corresponding perturbation and the true velocity. (c, d, e) Same for density model.	138

- 6.2 Data comparison. (a),(b) and (c) are the nonlinear data slices at 5, 12, 25Hz, respectively. (d),(e) and (f) are the synthetic linear data slices at 5, 12, 25Hz, respectively. (g),(h) and (i) are the predicted linear data slices with the inverted filters at 5, 12, 25Hz respectively. 140
- 6.3 Inverted filters at 5, 12, 25Hz. 141
- 6.4 In the last iteration of out-loop, ranks of the matched-filters along the inner iterations at 5, 12, 25Hz. 141
- 6.5 In the later iteration of out-loop, the Frobenius residuals of the objective function along inner iterations at 5, 12, 25Hz. 142
- 6.6 LS-RTM results using velocity-only Born modeling operator. (a) Idealized linearized data, nonlinear data (b) without, and (c) with matched-filter approach. 143

SUMMARY

Seismic imaging is an important tool for the exploration and production of oil & gas, carbon sequestration, and the mitigation of geohazards. Through the process of seismic migration, images of subsurface geological structures are created from data collected at the surface. These images reflect changes in the physical rock properties such as wave speed and density. While significant progress has been made in the development of 3D imaging technology for complex geological areas, several challenges remain, some of which are addressed in this thesis. The first main contribution of this thesis is in the area of creating so-called subsurface-offset gathers, which play an increasingly important role in seismic imaging because they provide a multitude of information ranging from the reflection mechanism itself to information of the dips of specific reflectors and the accuracy of the background velocity model. Unfortunately, the formation and manipulation of these gathers come with exceedingly high computational and storage costs because extended image volumes are quadratic in the image size. These high costs are avoided by using techniques from modern randomized linear algebra that allow for compression of extended image volumes into low-rank factorized form—i.e., the image volume is approximately written as an outer product of a tall and a wide matrix. It is demonstrated that this factorization provides access to different types of sub-surface offset gathers, including common-image (point) gathers, without the need to explicitly form this outer product. As a result, challenging steep dip imaging situations, where conventional horizontal offset gathers no longer focus, can be handled. Moreover, extended image volumes for one background velocity model can directly be mapped to those of another background velocity model. As a result, factorization costs are incurred only once when examining imaging scenarios for different background velocity models. The second main contribution of this thesis is on the development of computationally efficient sparsity-promoting imaging techniques and on-the-fly source estimation. In this work, an adaptive technique is proposed where the unknown

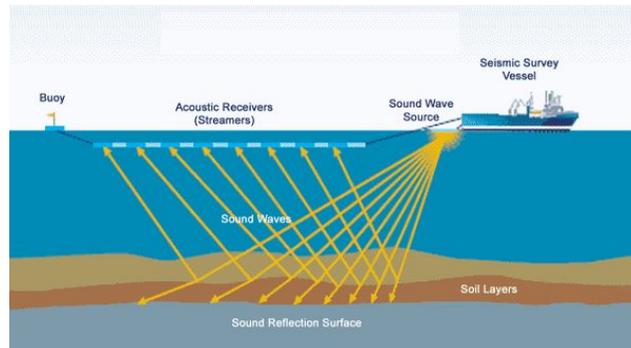
time signature of the sources is estimated during imaging. Without accurate knowledge of these source signatures, seismic images can be wrongly positioned and can have the wrong amplitudes hampering subsequent geophysical and geological interpretations. With the presented technique, this problem is mitigated. Finally, a contribution is made to address the detrimental effects of surface-related multiples. If not handled correctly, these multiples give rise to unwanted artifacts in the image. A new technique is introduced to address this issue in realistic settings where there is a strong density contrast at the ocean bottom. As a result, the surface-related multiples are mapped to the reflectors. Because bounce points at the surface can be considered as sources, this mapping of the multiples rather than removal increases the subsurface illumination.

CHAPTER 1

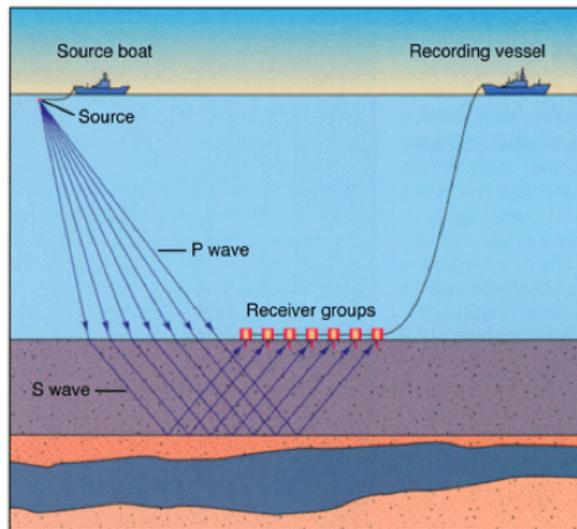
INTRODUCTION

The petroleum industry conducts seismic exploration to search for subsurface deposits of crude oil and gas. This exploration is based on the geophysical principle that seismic waves reflect and refract at geologic interfaces where the impedances (i.e., the products of wave velocities and rock densities) change. We can use the dispersion features of waves in the earth to detect the Earth's interior properties within 10km (Sheriff and Geldart 1983) at the human-generated waves bandwidths up to 100Hz. Such detection can be carried out by either land or marine surveys, the latter of which is mainly conducted in two ways. The first (illustrated in Figure 1.1a) uses specially-equipped vessels that tow several cables, known as streamers, which contain a set of hydrophones that record pressure at fixed intervals (Vaage 2004). Streamers are deployed below the surface of the water, away from the vessel. The seismic sources generated by airguns are set in the water between the vessel and the first receiver as in Figure 1.1a (figure from <https://www.tes.com/lessons/TJUnGRN16pfVg/copy-of-marine-seismic-survey>). Marine surveys can also be conducted by recording seismic waves by ocean bottom cables (i.e., ocean bottom nodes, OBN)(Vaage 2004), illustrated in Figure 1.1b (figure from <http://www.peakseismic.com/content/ocean-bottom-seismic.asp>). In general, marine surveys are capable of acquiring millions of recording traces covering an exploration area of 10^2km^2 level with the subsurface discretized by $10^6 - 10^9$ gridpoints.

After the pro-processing of data quality controls such as denoise, demultiple, and deghost (as shown in the workflow in Figure 1.2), the seismic workflow typically requires the construction of a background velocity model that describes the long-wavelength characteristics of the subsurface and predicts the kinematics of wave propagation in the true subsurface. Then the short-wavelength features, including reflectivity or model perturbations with re-



(a) Marine survey using hydrophones



(b) Marine survey using OBN

Figure 1.1: Marine surveys, (a) using hydrophones,(b) using OBN.

spect to the background model need to be imaged to facilitate locating oil and gas deposits. Among these imaging methods, the most popular method is reverse-time migration (RTM) (Baysal, Kosloff, and Sherwood 1983; Chang and McMechan 1986) because of its ability to generate images with all possible arrivals in lateral velocity variation scenarios without any dip limitations. RTM (Figure 1.2) achieves the zero-offset images by applying the zero-lag cross-correlation imaging condition to the forward and backward wavefields of the source and recording data (or taking the real part of the element-wise multiplies between the harmonic slices of the two wavefields) and depicting the reflectors as the locations in which the likelihood that these two wavefields encounter one another is high. The extended version (Figure 1.2) of this imaging condition uses the multi-dimensional correlation between these two wavefields (matrix-matrix multiplies between the harmonic slices of the two wavefields). The extended image volumes (EIVs) resulted from this imaging condition contain not only RTM image as their diagonal elements, but also the common image point gathers (CIPs) (Rickett and Sava 2002; Leeuwen and Herrmann 2012; Leeuwen, Kumar, and Herrmann 2017; Kumar, Graff-Kray, Leeuwen, and Herrmann 2018b) as their columns and common image gathers (CIGs) (Symes 2008; Stolk, Hoop, and Symes 2009; Rickett and Sava 2002), which help determine how accurate the background model is. The EIVs generally contain information used for not only creating images, but also interpreting rock properties and analyzing velocity in complex geological settings. Obtaining EIVs in a traditional way, however, may be impossible, especially in the case of an extensive number of sources and receivers, which leads to a multitude of wave equations to solve and large gridpoints in the subsurface requiring an enormous amount of memory for storage.

The challenges of exploiting reservoirs with complex geology such as faults or salt bodies lead to increasing demand for high-resolution images of areas with complicated structures. Such demand has given rise to inversion methods based on RTM, namely least-squares RTM (Schuster 1993; Nemeth, Wu, and Schuster 1999; Guitton, Kaelin, and Biondi 2006), the purpose of which is to iteratively fit the observed data to synthetic data

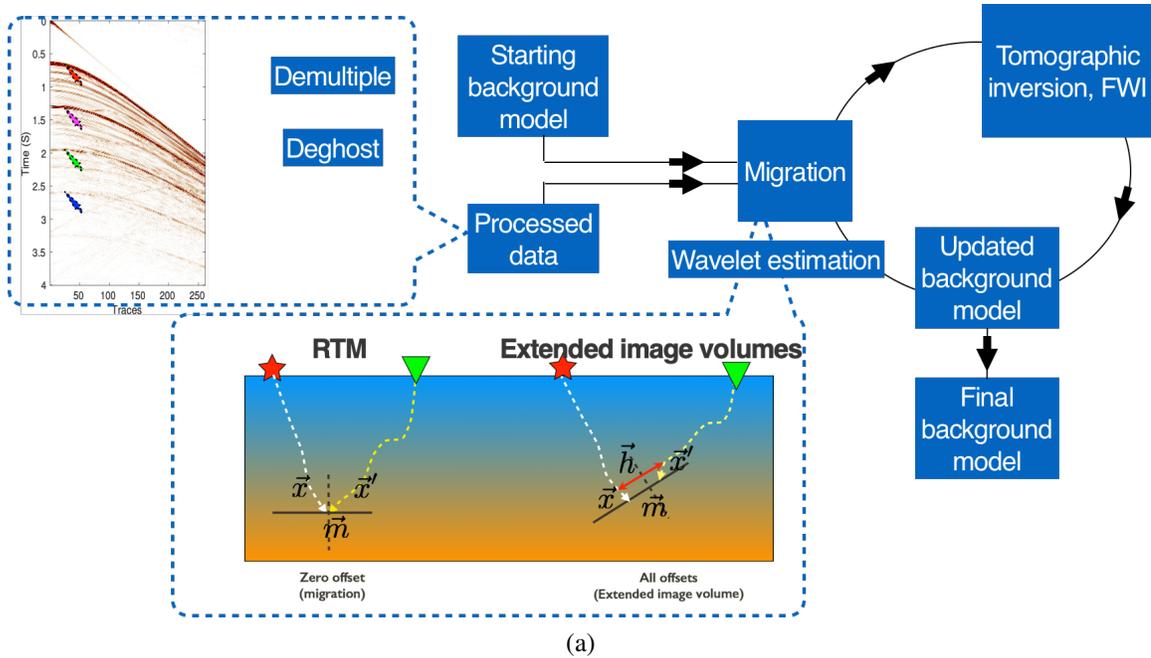


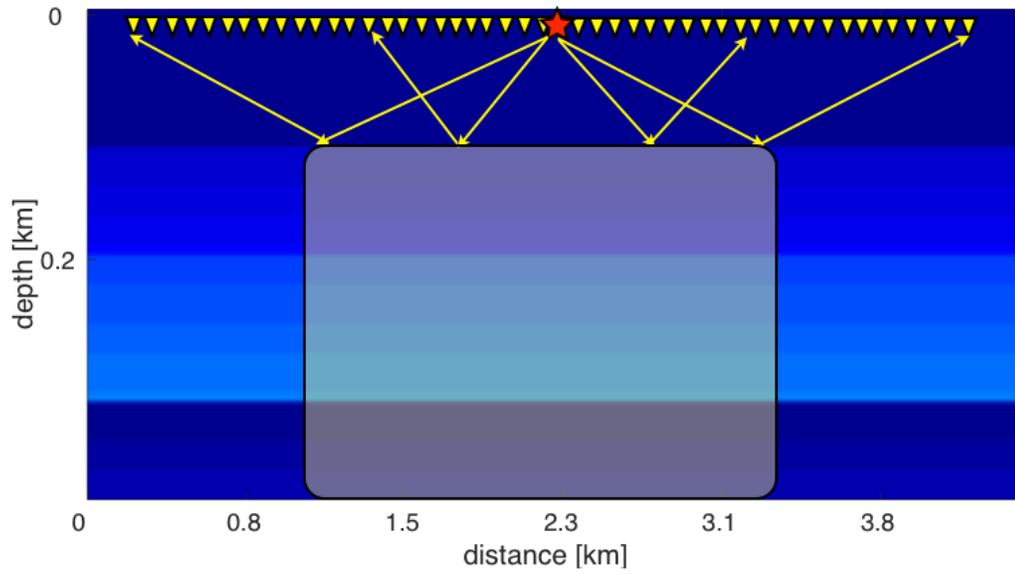
Figure 1.2: Seismic workflow for marine data

in a least-squares sense to remove the source and receiver imprints and the limited wavelet bandwidth influence from RTM. As each iteration of LS-RTM requires one RTM, further application of this method to realistic problems is compromised because of the exorbitant potential cost incurred by solving wave-equations. This situation is further complicated by the missing source function in real problems.

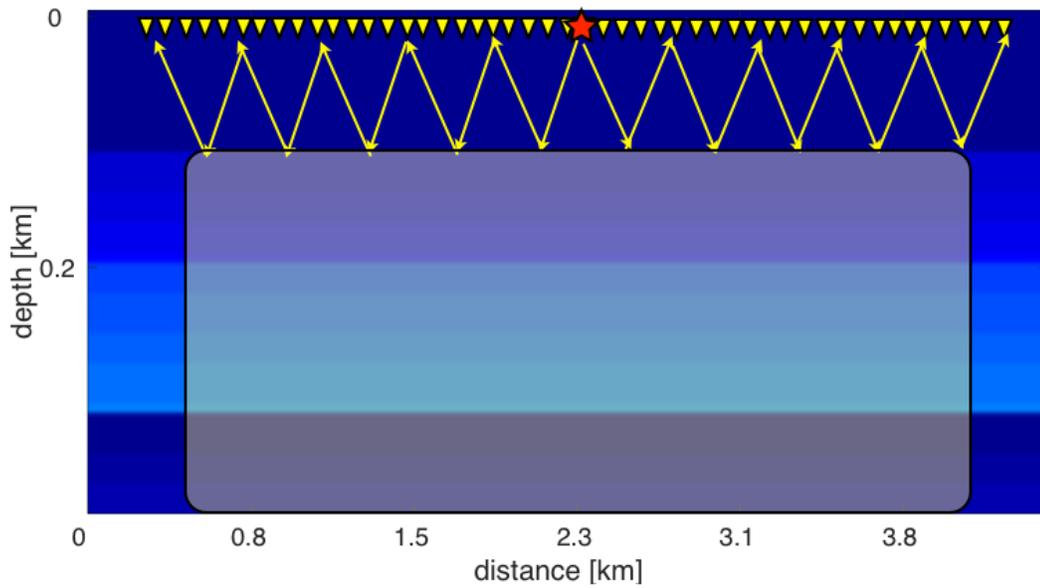
As the above imaging methods are based on single-scattering approximation (Keho and Beydoun 1988; Claerbout, Green, and Green 1985), they are limited to only primaries in marine data. Such data, however, contain strong surface-related multiples because waves are strongly reflected multi times between the water-air interface and strong impedance interface at the ocean bottom (Berkhout 1986). With migrations directly applied to untreated marine data, the multiples will not be correctly mapped to subsurface reflectors locations, resulting in artificial interfaces. A great deal of effort has been devoted to removing multiples in imaging, including the most popular method—surface-related multiples elimination (SRME) (Verschuur, Berkhout, and Wapenaar 1992).

Although the elimination of multiples from all reflection data avoids the artifacts from

multiples, the abandonment of multiples wastes a large amount of energy in the total reflection data (Verschuur 2006; Guitton, Valenciano, Bevc, and Claerbout 2007; Whitmore, Valenciano, Sollner, and Lu 2010; Lu, Whitmore, Valenciano, and Chemingui 2011). Whitmore, Valenciano, Sollner, and Lu 2010 found that the multiples supply extra illumination to the migrated images with more small incident angle energy because each receiver acts as a secondary source, illustrated in Figure 1.3. The figure displays illuminations with and without multiples. If traditional migration is applied, multiples will form phantom reflectors shown in Figure 1.4. Migrating multiples into high-resolution images efficiently, however, remains an issue. Also, the traditional inversion for imaging only inverts the velocity perturbations with a corresponding Born modeling kernel; that is, it converts the data components generated by strong density variations at the ocean bottom into artificial velocity perturbations, which is also problematic.

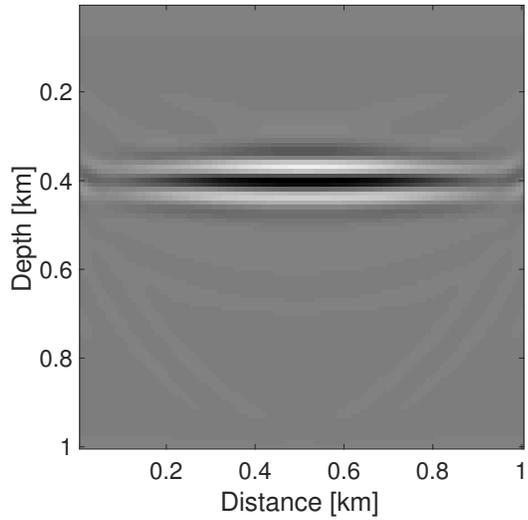


(a) Illumination of only primaries

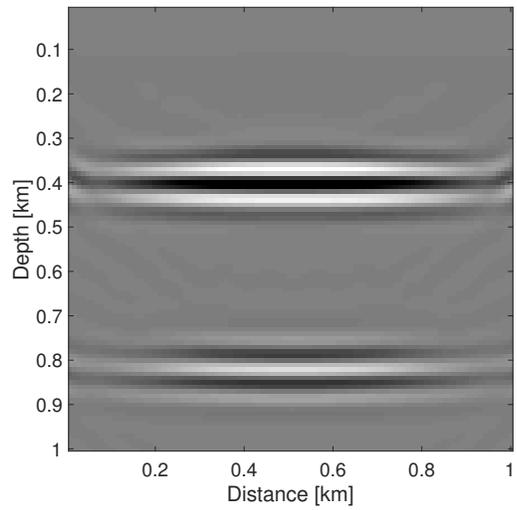


(b) Illumination of multiples and primaries

Figure 1.3: Illustration of illuminations. Red stars represent the source and yellow triangles represent hydrophones. The illuminated areas are marked in yellow. The first blue layer denotes the water column, and the arrowhead lines indicate the incident and reflected rays.



(a) RTM with only primaries



(b) RTM with primaries and multiples

Figure 1.4: Illustration the phantoms from multiples in RTM for a simple two-layer model with only one interface. (a) RTM with only primaries, (b) RTM with both primaries and multiples

1.1 Objectives

To summarize, the aim of this thesis is to achieve the following objectives:

1. To develop a computationally feasible two-way wave-equation-based factorization framework in time domain that provides us access to the amplitudes of full subsurface extended image volumes without explicitly forming the source and receiver wavefields for each shots; and to adapt this framework with velocity variation scenarios.

2. To develop a robust time-domain sparsity-promoting LS-RTM with on-the-fly source estimation. Leveraging insights from both stochastic optimization and compressive sensing to reduce the substantial computational cost of LS-RTM without compromising image quality; and to develop an on-the-fly source estimation approach to enhance applications to realistic problems.

3. To jointly invert the primaries and multiples in the time domain with only the Born modeling operator with respect to velocity; and to propose a method of removing the strong artificial velocity from the density perturbation at the ocean bottom without developing the Born modelling with respect to density.

1.2 Outline

Chapter 2 begins by introducing the underlying theory of monochromatic EIVs, their low-rank representations, and the probing technique used in low-rank recovery method—randomized singular value decomposition (rSVD). To overcome computational bottlenecks and extend the technique to large-scale application in the future, we derive a time-domain version of EIV and corresponding probing techniques and rSVD based on a fast time-stepping propagator and fast Fourier transforms. To alleviate the problem of the increasing ranks of monochromatic EIV along frequency, we propose to combine power schemes to rSVD to accelerate the decay of EIVs' singular values at high frequency. We compare the performance of the basic rSVD and two power schemes: block Krylov iterations (BKI) and

simultaneous iterations (SI) with difference probing sizes and powers on a small EIV. We also introduce the methods of extracting different image gathers from the low-rank factors, e.g. RTM, CIPs, CIGs and the geological dip-corrected CIGs, without any further wave-equation solves; then we conduct the test on the Marmousi model.

In Chapter 3, we introduce the invariance relationship, which could facilitate the mapping of current low-rank factors to another pair of factors for velocity variation scenarios. We begin by proposing the theory of low-rank factor mapping on monochromatic EIVs and then extend it to the time-domain version. The mapping, which entails only wave-equation solves, is an SVD-free approach. We test this method for velocity continuation scenarios on a challenging part of the Sigsbee model, whose initial guess has the wrong salt dome, which will make the reflectors of the RTM (i.e. diagonals of the EIV) under the salt dome distorted, and the energy of the CIPs (i.e. columns of the EIV) or CIGs unfocused around the image point. The update from the initial low-rank factors results in favorable image volumes from which we extract and compare different image gathers.

In Chapter 4, we propose an on-the-fly source estimation method for time-domain sparsity promoting least-squares RTM (SPLS-RTM), in which the image condition is now the zero-lag cross-correlation. We begin by introducing the time-domain SPLS-RTM as a Basis Pursuit Denoise (BPDN) problem and replace the ℓ_1 -norm by an elastic net consisting of a strongly convex combination of ℓ_1 - and ℓ_2 -norm to relax the objective function. Then we introduce an easily-implemented algorithm, the linearized Bregman method, which solves the optimization problem. Following that, we propose on-the-fly source estimation via variable projections that solve the least-squares sub-problem with penalties that prevents overfitting. Finally, we demonstrate the effectiveness of the proposed method with a stylized example on how it eliminates overfitting in the estimated source and the sparsity solutions. We also test the robustness of our method to noise on the Marmousi model and design a hybrid framework for the challenging Sigsbee model that contains salt body.

In Chapter 5, we exploit the capability of our LB-based framework in the joint inversion

of primaries and multiples. We first introduce the fundamental theory of the surface-related multiples elimination relationship (SRME), widely used in the pre-processing of demultiples and primaries prediction. We discuss the drawbacks of SRME in demultiples, that is, the assumption of least energy for the predicted primaries, which will lead to failure of this method in a shallow water scenario, where the multiples strongly interrupt the primaries. Then we continue with the proposed joint inversion of primaries and multiples, which incorporates the SRME relationship into the wave equation by areal source injection. By extending the work in Chapter 4, we implement this joint inversion in the time domain, and to generate strong multiples in shallow water, we introduce strong density perturbation at the ocean bottom. We initially test our method on a linear dataset based on a portion of the Sigsbee model in the time-harmonic domain, and then we test the joint inversion for nonlinear data in the time domain. Since we are inverting with only Born modeling with respect to the velocity perturbation, we obtain some strong artificial velocity perturbations converted from the true density perturbations at the ocean bottom.

In Chapter 6, we continue to discuss the problems encountered by sparsity-promoting least-squares reverse-time migration when it inverts the strong density perturbation related data components with the velocity-only Born modeling operator. As observed in chapter 5, the strong density perturbation will be inverted as strong dummy velocity perturbation. Instead of developing the Born operator with respect to both density and velocity or modifying the image condition, we propose a matched-filter-based LS-RTM for the velocity-only Born modeling operator to remove the artifacts created by the strong density variations. This method does not necessitate extra work on finite difference stencils. We conduct a preliminary test based on a discontinuous layered model with strong density variations at the ocean bottom to demonstrate the efficacy of the proposed formulation.

In Chapter 7, we summarize the work in this thesis and propose the possible research directions in the future.

CHAPTER 2

LOW-RANK RECOVERY OF SUBSURFACE EXTENDED IMAGE VOLUMES BASED ON TIME-STEPPING PROPAGATOR AND POWER SCHEME

2.1 Summary

Extended image volumes (EIVs) contain rich subsurface information and is low rank due to the rank limitation of the data term contained inside its formulation. Low-rank recovery is proposed to explore the EIV instead of forming it explicitly which will take amounts of memory storage and computation. Randomized SVD method could help to recovery the low-rank factors of EIV based on randomized probing technology. To extend this work to the possible industry scale in the future, we implement the low-rank recovery based on rSVD in the time domain in this chapter. To make the randomized probing feasible and numerical stable in time domain, we combine the source term in the formulation of EIV with the Gaussian random noise to form the bandwidth limited source wavefield. The multi-dimensional convolution is implemented by matrix-matrix multiplications monochromatically along the discredited frequencies after fast Fourier transform to the corresponding terms. Another problem that might impede the realistic application of low-rank recovery of EIV is the fact that the rank of the data or EIV will increase along frequency. We have to prob the EIV with larger probing size at higher frequencies or accelerate the decay of the singular values of EIV. In this chapter we introduce the power scheme based rSVD methods that help to narrow the gap between the neighbor singular values. We compare the errors of the recovered singular values and the diagonal RTM by different probing size and power settings for a small part of the Marmousi model to demonstrate the advantage of the power iterations, and point out the block Krylov method (BKI) with power 1 would be the best choice. Also we explain the ways of extracting different gathers out from the

low-rank factors, e.g. reverse time migration(RTM), common image point gather(CIP), common image gather(CIG) and dip corrected CIG. Finally we test our proposed method on the Marmousi model to demonstrate its effectiveness by supplying good approximations of the RTM, CIPs and CIGs with the cost smaller than one RTM,i.e. the probing size n_p smaller than one quarter of the number of source $n_s/4$.

2.2 Introduction

The formation of subsurface-offset gathers, such as common-image gathers (CIGs, (Symes 2008; Stolk, Hoop, and Symes 2009; Rickett and Sava 2002)), angle-domain common-image gathers (ADCIGs, (De Bruin, Wapenaar, and Berkhout 1990; Sava and Fomel 2003; Kroode 2012; Kühn and Sacchi 2003; Mahmoudian and Margrave 2009; Dafni and Symes 2016b; Dafni and Symes 2016a)), and more recently common-image point gathers (CIPs, (Leeuwen and Herrmann 2012; Leeuwen, Kumar, and Herrmann 2017; Kumar, Graff-Kray, Leeuwen, and Herrmann 2018b)), has become an essential component of modern seismic imaging workflows. Each of these gathers provides information on the quality of the velocity model and the scattering mechanism, which is dependent on the subsurface itself as well as the acquisition geometry. Contrary to CIGs, CIPs provide information on the complete scattering mechanism since they are a function of the full omni-directional subsurface offset.

Usage of these gathers includes quality control during velocity model building (Yang and Sava 2015; Biondi and Symes 2004a); automatic model updates during migration velocity analysis (Symes and Carazzone 1991; Shen and Symes 2008); and inferences made on rock properties from amplitude versus offset analysis (De Bruin, Wapenaar, and Berkhout 1990). All of these usages rely on having access to high quality subsurface image volumes. While access to fast hardware and memory has made imaging modalities, such as reverse-time migration (RTM, (Baysal, Kosloff, and Sherwood 1983)) computationally feasible in 3D (Kukreja, Louboutin, Vieira, Luporini, Lange, and Gorman 2016),

the formation of subsurface-offset image volumes remains a major challenge because it involves a loop over the sources and multi-dimensional cross-correlations between (shifted) spatial-temporal forward and adjoint wavefields. Aside from the extra computational burden, subsurface-offset or angle gathers add one or more dimensions making image volumes more challenging and costly to store and manipulate.

By relying on the wave-equation itself in combination with a (randomized) probing technique, Leeuwen, Kumar, and Herrmann 2017 was able to get access to full subsurface-offset images via actions of the double two-way wave equation on probing vectors. This double wave-equation is the two-way wave-equation counterpart of Claerbout's double square-root equation (Claerbout 1970; Symes and Carazzone 1991; Biondi and Symes 2004b; Sava and Vasconcelos 2011a), which is based on the one-way wave equation limiting its accuracy in media with steeply dipping reflectors. The two-way wave equation remedies this shortcoming.

By choosing probing vectors consisting of a single point scatter, Leeuwen, Kumar, and Herrmann 2017 was able to extract CIPs, which are the size of the original image but now as a function of the omni-directional subsurface offset. Compared to conventional CIGs that are generally computed as a function of the horizontal offset alone, CIPs contain all offsets in all directions and this offers important advantages in situations where we are dealing with steeply dipping reflectors in which case CIGs no longer focus (see Figure 11 of Leeuwen, Kumar, and Herrmann 2017).

While this probing method provides access to an object that can not be formed explicitly—i.e., image volumes are quadratic in the image size, the cost of this access scales with the number of probing vectors limiting its use. Despite this shortcoming, the formulation presented by Leeuwen, Kumar, and Herrmann 2017 provided new insights to migration velocity analysis, localized amplitude versus offset analysis including correction for the geologic dip, and the derivation of completely novel approaches to velocity continuation (Leeuwen and Herrmann 2012; Kumar, Graff-Kray, Leeuwen, and Herrmann 2018b), which derive

from an invariance relation of the double two-way wave equation, and redatuming (Kumar, Graff-Kray, Vasconcelos, and Herrmann 2019).

Even though the derivation of the double two-way wave-equation resulted in fundamentally new insights how to form and manipulate certain aspects of omni-directional full-subsurface extended image volumes (EIVs), the proposed technique relied on frequency-domain propagators and access via probing. This reliance limits its potential application to more realistic imaging scenarios where time-domain propagators are needed for the wave simulations and where access to many (geologic-dip) corrected CIGs is desired. We overcome these shortcomings by proposing a low-rank matrix factorization technique, based on the randomized singular-value decomposition (rSVD, (Halko, Martinsson, and Tropp 2011)). To allow low-rank approximations at higher frequencies, where the singular values decay slower, we propose a Block-Krylov method (Musco and Musco 2015). This method requires more costly probings but leads to more accurate low-rank factorizations of EIVs.

Aside from achieving a massive compression of EIVs, we will show that low-rank factorizations also give us readily access to CIPs and (geologic dip-corrected) CIGs without the need to form EIVs explicitly or to solve additional wave equations—an observation also made by Da Silva, Zhang, Kumar, and Herrmann 2019 where subsurface-offset gathers were formed from a tensor factorization based on the hierarchical Tucker format (Silva and Herrmann 2015).

Our contributions are organized as follows. We first briefly review the definition of monochromatic extended image volumes, their relation to the double two-way wave-equation, and a low-rank factorization based on the rSVD. We also show how to migrate and derive CIPs from this low-rank factorization. To accommodate more realistic imaging scenarios, we introduce representations for time-domain EIVs including time-domain probing. Since we are now able to image at high frequencies, we present and compare more elaborated probing techniques that involve powers of the double-wave equation. After showing that these method lead to more accurate factorizations, we show how CIPs and (dip-corrected)

CIGs can be formed from these factors directly and without the need of forming the EIVs explicitly. Via carefully selected experiments, we validate the presented approach by comparing true CIPs and CIGs with their approximate counterpart calculated from the proposed factorization.

2.3 Full subsurface monochromatic extended image volumes

Before discussing our novel approach to factorize image volumes, we first briefly summarize the formation and probing of image volumes in the frequency and time domain.

2.3.1 Extended image volumes with Helmholtz

According to Leeuwen, Kumar, and Herrmann 2017, monochromatic extended image volumes (EIVs), with subsurface offsets in all directions, can be formed by an outer product. This product is calculated between the forward wavefields, collected for n_s different sources and $N = n_x \times n_z$ (with n_x, n_y number of gridpoints in the $x - z$ directions) gridpoints in the tall matrix $\mathbf{U}_i \in C^{N \times n_s}$ at the i^{th} frequency, and the corresponding matrix for the adjoint wavefields $\mathbf{V}_i \in C^{N \times n_s}$ —i.e., we have

$$\mathbf{E}_i = -\omega_i^2 \mathbf{V}_i \mathbf{U}_i^*, \quad (2.1)$$

with ω_i the i^{th} angular frequency. In this expression, $*$ represents the complex conjugate transpose. This monochromatic image volume represents a discretized version of $E(\vec{x}; \vec{x}')$ where in 2D $\vec{x} = (x, z)$ refers to the spatial coordinates and $\vec{x}' = (x', z')$ to a second set of coordinates from which we derive

$$\vec{m} = \frac{\vec{x} + \vec{x}'}{2} \quad \text{and} \quad \vec{h} = \frac{\vec{x} - \vec{x}'}{2} \quad (2.2)$$

with $\vec{m} = (m_x, m_z)$ the midpoint coordinates and $\vec{h} = (h_x, h_z)$ the subsurface offset coordinates along the two spatial coordinate directions. We use semicolons ; to separate

the coordinate directions so that can be represented by a matrix.

The above forward and adjoint wavefields satisfy the following forward and adjoint wave equations:

$$\mathbf{H}_i(\mathbf{m})\mathbf{U}_i = \mathbf{P}_s^\top \mathbf{Q}_i, \quad (2.3)$$

$$\mathbf{H}_i^*(\mathbf{m})\mathbf{V}_i = \mathbf{P}_r^\top \mathbf{D}_i, \quad (2.4)$$

where $\mathbf{H}_i(\mathbf{m})$ represents the discretized Helmholtz operator at the i^{th} frequency. The Helmholtz operator itself is parameterized by the discretized squared slowness collected in the vector $\mathbf{m} \in R^N$. The $n_s \times n_s$ matrix \mathbf{Q}_i denotes the source matrix, where n_s is the number of sources. The observed data itself is collected in the monochromatic $n_r \times n_s$ data matrix \mathbf{D}_i , where each column represents a single monochromatic source experiment with n_r receivers. The matrices \mathbf{P}_s and \mathbf{P}_r are projections that restrict the full wavefields to the source and receiver positions, respectively. The symbol $^\top$ denotes matrix transpose. Finally, by substituting equations 2.3 and 2.4 into Equation 2.1, we can express \mathbf{E}_i as a function of \mathbf{Q}_i and \mathbf{D}_i as follows:

$$\mathbf{E}_i = -\omega_i^2 \mathbf{H}_i^{-*} \mathbf{P}_r^\top \mathbf{D}_i \mathbf{Q}_i^* \mathbf{P}_s \mathbf{H}_i^{-*} = \mathbf{H}_i^{-*} \mathbf{P}_r^\top \dot{\mathbf{D}}_i \dot{\mathbf{Q}}_i^* \mathbf{P}_s \mathbf{H}_i^{-*}, \quad (2.5)$$

where to simplify our notations as far as possible, we introduce the symbol $\dot{\cdot}$ to monochromatic matrix to involve the $j\omega_i$ implicitly. Equation 2.5 corresponds the solution of the double two-way wave-equation given by

$$\mathbf{H}_i^* \mathbf{E}_i \mathbf{H}_i^* = \mathbf{P}_r^\top \dot{\mathbf{D}}_i \dot{\mathbf{Q}}_i^* \mathbf{P}_s. \quad (2.6)$$

As during migration, EIVs are computed using a background velocity model, which defines the squared slowness in the above discretized Helmholtz operators. For now, we

assume this background velocity model to be known and we are interested in finding ways to form and manipulate image volumes in realistic imaging scenarios. Because N easily becomes too large, it becomes unfeasible to form, store, or even manipulate EIVs in explicit form. We address this issue by exploiting reported (Leeuwen, Kumar, and Herrmann 2017; Yang, Graff, Kumar, and Herrmann 2019; Kumar, Graff-Kray, Leeuwen, and Herrmann 2018a) low-rank properties of EIVs. Without loss of generality, we will work exclusively on 2D imaging problems that can feasibly be extended to 3D. We will focus on accuracy and develop techniques to cast EIVs into factored form, which allows for computationally feasible manipulation and extraction of useful gathers for migration velocity, amplitude versus offset analyses, and redatuming (Leeuwen, Kumar, and Herrmann 2017; Kumar, Graff, Vasconcelos, and Herrmann 2019).

As in earlier work by Leeuwen, Kumar, and Herrmann 2017, our approach relies on probing EIVs—i.e. computing the action of EIVs on certain probing vectors. Aside from giving us access to Common Image Point gathers(CIPs)—i.e. full omni-directional subsurface-offset gathers, probings provide information necessary to factor EIVs using randomized Singular Value Decompositions (Halko, Martinsson, and Tropp 2011). To enabling scale up, we extend earlier work by using wave propagators based on time-stepping, in combination with a more sophisticated randomized probing methodology. Before introducing probing with times-stepping, let us first briefly review probing of monochromatic EIVs and show how this technique leads to an alternative formation of subsurface zero-offset reverse-time migration (RTM).

2.3.2 Low-rank factorization of time-harmonic EIVs

To form our EIVs in a computationally feasible manner, we compute the action of these EIVs on a limited number (n_p) of monochromatic probing vectors collected in the tall matrix $\mathbf{W}_i \in C^{N \times n_p}$ with $n_p < n_s \ll N$. Following Leeuwen, Kumar, and Herrmann

2017, the probing entails

$$\mathbf{Y}_i = \mathbf{E}_i \mathbf{W}_i = \mathbf{H}_i^{-*} \mathbf{P}_r^\top \dot{\mathbf{D}}_i \dot{\mathbf{Q}}_i^* \mathbf{P}_s \mathbf{H}_i^{-*} \mathbf{W}_i, \quad (2.7)$$

which involves $2n_p$ wave-equation solves. There are several different choices possible for \mathbf{W}_i . For now¹, we chose the entries of \mathbf{W}_i to be drawn from zero-centered Gaussian noise with unit standard deviation to span the range of \mathbf{E}_i (Kumar, Graff-Kray, Leeuwen, and Herrmann 2018a).

In situation where the EIVs can be approximated accurately by a low-rank matrix, the result of the probing \mathbf{Y}_i contains all information on the range of \mathbf{E}_i as long as n_p is slightly larger than the rank k (Halko, Martinsson, and Tropp 2011). It also will allow to represent EIVs via a low-rank factorization

$$\mathbf{E}_i \approx \mathbf{L}_i \mathbf{R}_i^*, \mathbf{L}_i \text{ and } \mathbf{R}_i \in \mathbf{C}^{N \times n_p}, \quad (2.8)$$

where the factors \mathbf{L}_i and \mathbf{R}_i are computed with the randomized singular-value decomposition (Halko, Martinsson, and Tropp 2011) as described in Algorithm 6 included in the Appendix. This algorithm takes the above probing as input. Compared to the more expensive standard SVD method, which involves $2n_s$ PDE solves cost in the order of $\mathcal{O}(N^3)$ the rSVD only costs $4n_p$ PDE solves and $\mathcal{O}(2Nn_p^2)$.

To illustrate the concept of factorizing EIVs with rSVDs, we consider a small ($N = 100 \times 100$) EIV computed from the Marmousi model and study the behavior of its singular values and the frequency dependence of its low-rank factored approximation. In addition to giving us access to full subsurface-offset image gathers, low-rank factorization gives us access to migrated images via

$$\delta \hat{\mathbf{m}} = \sum_i \text{diag}(\mathbf{E}_i) \approx \sum_i (\mathbf{L}_i \odot \bar{\mathbf{R}}_i) \mathbf{1}. \quad (2.9)$$

¹We can relax this assumption by using fast Fourier-based probing methods (Leeuwen, Kumar, and Herrmann 2017).

In this expression, $\text{diag}(\cdot)$ extracts the diagonal from the EIVs for each frequency and the symbol \odot represents element-wise multiplication also known as the Hadamard product. The symbol $\bar{\cdot}$ represents complex conjugation and $\mathbf{1}$ represents a column vector with n_p 1's. The second part of the above expression corresponds to taking the Hadamard product of the factors for each frequency, followed by summing over the columns. The sum over frequencies, which range between 5 and 50 Hz with a step of 0.5 Hz, corresponds to the zero time-lag imaging condition (Berkhout 1986; Claerbout, Green, and Green 1985) while extraction of the diagonal corresponds to imposing the zero subsurface-offset imaging condition.

Results of this procedure are summarized in Figure 2.1, where we show how to extract a zero-subsurface offset migrated image (Figure 2.1b) from the diagonal of the EIV plotting in Figure 2.1a. In addition to containing information to form a migrated image, EIVs also contain CIPs, which correspond to extracting columns from the EIVs, followed by summing over frequency. As with migration (cf. Equation 2.9), this information is accessible from low-rank factored form given in Equation 2.8. As long as we increase the rank from $n_p = 10$ to 40 for increasing frequencies, the low-rank approximation in Equation 2.9, yielding images and CIPs close to the ones obtained with regular RTM, looping over all n_s sources, or CIP computation via probing. Compared to conventional Common Image Gathers (CIGs), CIPs contain full-subsurface offsets in all directions (Leeuwen, Kumar, and Herrmann 2017). As a result, they nicely show the directivity pattern and geologic dip of the different reflectors as we can see from the overlays in Figures 2.1c and 2.1d.

We were able to obtain the results in Figure 2.1 by making use of the relative fast decay for the singular values of the EIVs compared the decay for the singular values of the data matrix as illustrated in Figures 2.2 and 2.3. This suggests we should aim to factorize EIVs rather than the data. However, this fast decay for the singular values slows down for increasing frequency. This effect is illustrated in Figure 2.3 where we plot the rank we would need to choose for each frequency if we want to capture the singular values to within

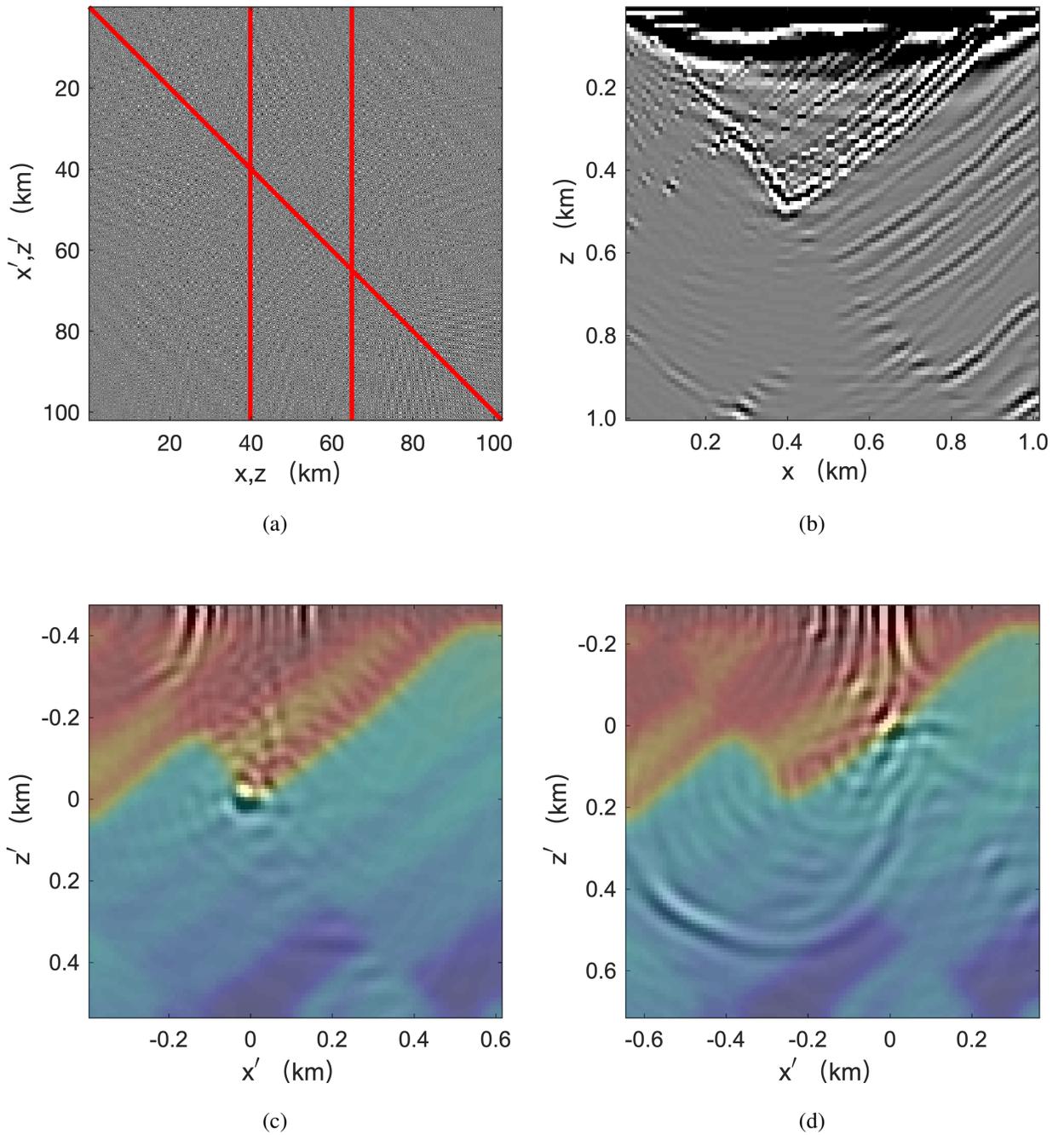


Figure 2.1: Extended image volume computed for the small Marmousi model for all frequencies $[5, 50]$ Hz with step 0.5 Hz: (a) full EIV, (b) RTM image from the reshaped diagonal of the EIV, (c) common image-point gather at $(7110\text{m}, 480\text{m})$ from the 3987^{th} column of the EIV, (d) common image-point gather at $(7360\text{m}, 300\text{m})$ from the 6494^{th} column of the EIV.

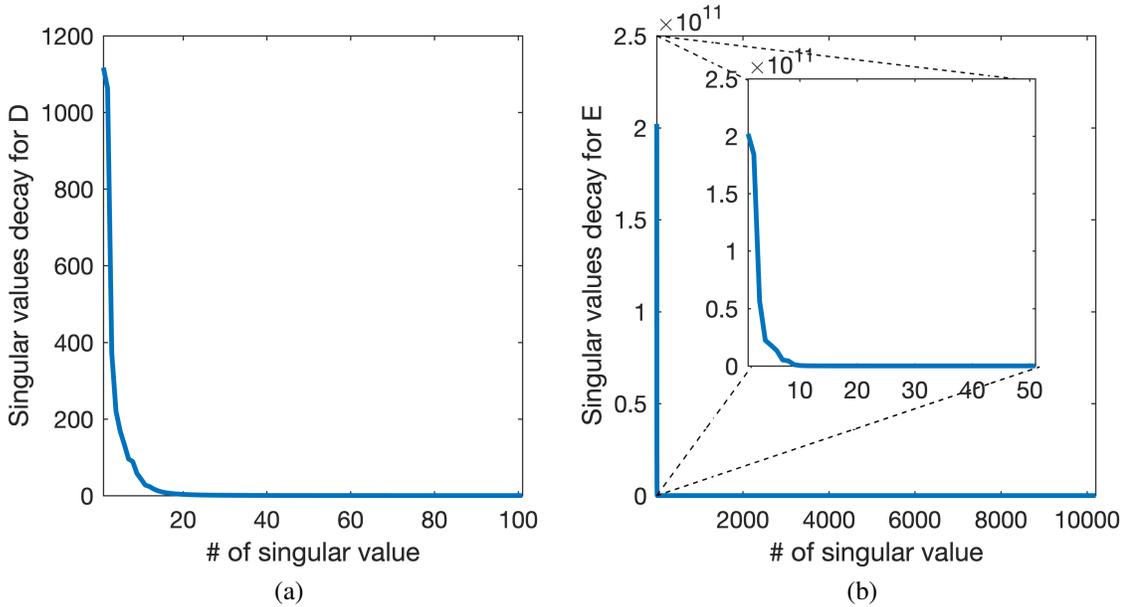


Figure 2.2: Extended image volume computed for the small Marmousi model (Figure 2.1) at 5Hz: (a) singular value decay of the data matrix, (d) singular value decay of the corresponding EIV matrix.

a 1, 5 or 10% of the largest singular value. Since the decays of the singular value decrease with frequency, we observe that the minimal rank we can select increases with frequency. Fortunately, this effect is smaller for the EIVs compared to the data and this explains why $n_p = 9-40$ was sufficient in example included in Figure 2.1.

While the above approach allows us to form and manipulate EIVs in low-rank factorized form, without ever forming the EIV matrix explicitly, several challenges remain to scale this approach to more realistic settings, which include larger models and higher frequencies. Both of these call for computationally more efficient wave propagators and randomized SVDs able to factor matrices that can not be approximated accurately by low rank factorizations. Before demonstrating our approach on a realistic example, we discuss how to probe with times-stepping propagators and how to handle factorizations of high-frequency EIVs.

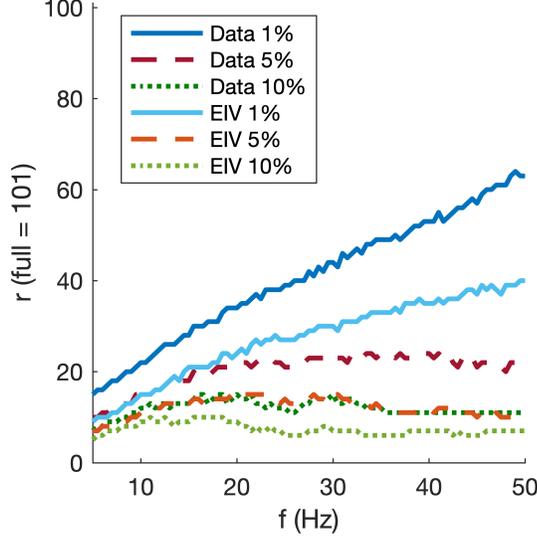


Figure 2.3: Estimated rank for the extended image volume computed for the small Marmousi model with respect to the frequency $[5, 50]$ Hz with step 0.5 Hz. This figure shows the rank of the data matrix \mathbf{D} and the EIV \mathbf{E} when truncated at 1%, 5%, 10% of the highest singular values.

2.4 Full subsurface extended image volumes based on time-stepping

2.4.1 Time-domain EIVs

To substitute time-harmonic wave-equation solvers in equation 2.5 with salable time-stepping, we introduce discrete temporal forward, \mathbf{U} , and adjoint wavefields \mathbf{V} as the solutions of

$$\mathcal{A}(\mathbf{m})[\mathbf{U}] = \mathcal{P}_s^\top[\mathbf{Q}] \quad (2.10)$$

and

$$\mathcal{A}^\top(\mathbf{m})[\mathbf{V}] = \mathcal{P}_r^\top[\mathbf{D}]. \quad (2.11)$$

In these expressions, the symbols $\mathbf{U} \in R^{n_t \times N \times n_s}$ and $\mathbf{V} \in R^{n_t \times N \times n_r}$ are tensors representing the forward and adjoint wavefields, respectively, with n_t the number of time

subsamples. The linear operator \mathcal{A} denotes the discretized wave-equation, which we solve via time-stepping using Devito¹. Similarly, we solve for the adjoint wavefield via backward time-stepping with the adjoint \mathcal{A}^\top . The square brackets $[\]$ are used to indicate application of linear time-domain operators to the respective tensors. As before, the source terms are given by impulsive sources and data collected in the tensors \mathbb{Q} and \mathbb{D} that are injected into the computational grid by the linear operators \mathcal{P}_s and \mathcal{P}_r .

With these definitions for the time-domain wavefields, we can write the time-domain EIV as follows:

$$\begin{aligned} \mathbb{E} &= \dot{\mathbb{V}} *_t \dot{\mathbb{U}}^\top \\ &= \mathcal{F}^\top [\dot{\mathbb{V}} \cdot \dot{\mathbb{U}}^*], \end{aligned} \tag{2.12}$$

where the symbol $\dot{\ }$ applied to the wavefield tensors introduces the first order time derivative to the corresponding term in time domain. And the symbol $*_t$ stands for multi-dimensional convolution² between the two derivative wavefields $\dot{\mathbb{V}}$ and $\dot{\mathbb{U}}^\top$. As before, we implement these convolutions via matrix-matrix multiplies of the monochromatic frequency slices $\dot{\mathbb{V}}_i$ and $\dot{\mathbb{U}}_i^*$, $i = 1 \cdots n_f$, which we shortly handle using the \cdot operator between the respective frequency tensors $\dot{\mathbb{V}}$ and $\dot{\mathbb{U}}^*$. We obtain the time-domain EIV by applying the inverse Fourier transform \mathcal{F}^\top along time to the respective frequency tensor.

To set the stage for probing of EIVs formed with the above time-domain propagators,

¹In our implementation, we used Devito (<https://www.devitoproject.org>) for our time-domain finite difference simulations and gradient computations (Luporini, Lange, Louboutin, Kukreja, Hüchelheim, Yount, Witte, Kelly, Herrmann, and Gorman 2018) and JUDI (<https://github.com/slimgroup/JUDI.jl>) as an abstract linear algebra interface to our Algorithms (Witte, Louboutin, Kukreja, Luporini, Lange, Gorman, and Herrmann 2019).

²With some abuse of notation, we assume that the wavefields collected in the tensors are multiplied as in an outer product.

we write

$$\begin{aligned}
\mathbf{E} &= \overbrace{(\mathcal{F} \circ \mathcal{A}^{-\top} \circ \mathcal{P}_r^\top [\dot{\mathbf{D}}])}^{\dot{\mathbf{V}}} \cdot \overbrace{(\mathcal{F} \circ \mathcal{A}^{-1} \circ \mathcal{P}_s^\top [\dot{\mathbf{Q}}])^*}^{\dot{\mathbf{U}}^*} \\
&= \mathbf{F} \circ \mathcal{A}^{-\top} \circ \mathcal{P}_r^\top \circ \mathcal{F}^\top [\dot{\mathbf{D}} \cdot \dot{\mathbf{Q}}^* \cdot (\mathcal{F} \circ \mathcal{P}_s \circ \mathcal{A}^{-\top} \circ \mathcal{F}^\top [\mathbf{I}])].
\end{aligned} \tag{2.13}$$

The symbol \circ refers to the composition operator between the time-domain operators.

As we can see, the above expression represents the double two-way wave-equation (cf. Equation 2.6) as proposed by Leeuwen, Kumar, and Herrmann 2016 but now based on wave propagation via time-stepping. The temporal convolutions are carried out by complex-valued matrix-matrix products in the temporal Fourier domain. Here the frequency tensor \mathbf{I} contains a set of monochromatic identity matrices $\mathbf{I}_i \in C^{N \times N}$, where $i = 1 \cdots n_f$.

2.4.2 Time-domain probing

While equations 2.12 and 2.13 in principle allow us to form EIVs in the time or Fourier domain using time-domain propagators, these expressions do not readily lend themselves to probing. Moreover, time-stepping propagators impose additional conditions on the wavefields they propagate—e.g. the source wavefield has to be bandwidth limited in time to ensure stability of our numerical scheme³. To ensure this requirement, we assume a single temporal source signature for all sources that are assumed to be delta distributions in space—i.e., $\dot{\mathbf{Q}}_i = j\omega_i \alpha_i \mathbf{I}_{n_s}$ where α_i is the i^{th} Fourier coefficient of the source and \mathbf{I}_{n_s} the identity matrix of size $n_s \times n_s$. Because of this particular choice, the action of the source commutes with the other operators so we can probe our EIVs with independent realizations

³All our time-domain wave simulations are carried out with the open-source package Devito (Kukreja, Louboutin, Vieira, Luporini, Lange, and Gorman 2016).

of bandwidth-limited Gaussian random noise—i.e. we have

$$\begin{aligned} \mathbf{Y} &= \mathcal{F} \circ \mathcal{A}^{-\top} \circ \mathcal{P}_r^\top [\dot{\mathbf{D}} *_t (\mathcal{P}_s \circ \mathcal{A}^{-\top} [\mathbf{W}])] \\ &= \mathcal{F} \circ \mathcal{A}^{-\top} \circ \mathcal{P}_r^\top \circ \mathcal{F}^\top [\dot{\mathbf{D}} \cdot \dot{\mathbf{Q}}^* \cdot (\mathcal{F} \circ \mathcal{P}_s \circ \mathcal{A}^{-\top} \circ \mathcal{F}^\top [\mathbf{I}])]. \end{aligned} \quad (2.14)$$

In this expression, the probing is carried out by the tensor $\mathbf{W} \in R^{n_t \times N \times n_p}$, which contains zero-centered Gaussian noise that is filtered by the time signature of the source-time function. As we will show below, the tall matrices \mathbf{Y} contain the necessary information to factor EIVs from which subsurface-offset gathers can be computed. Contrary to subsurface-offset gathers computed via image-domain cross-correlations of the forward and adjoint wavefields, each of which are of size $N \times n_t$, the above probing involves for each probing vector a single matrix-vector multiply with the $n_s \times n_r \times n_f$ data matrices. Since $n_s \times n_r \ll N$ and $n_f \ll n_t$, the probings are relative cheap.

The above expression for time-domain probing forms the basis for the remainder of this paper where the randomized SVD and other manipulations are carried out for each frequency, indexed by $i = 1 \cdots n_f$, separately. To simplify notation, we will tacitly assume loops over the frequency whenever we refer to monochromatic entities, e.g $\mathbf{Y} = f(\mathbf{X})$ corresponds to $\mathbf{Y}_i = f(\mathbf{X}_i)$ for $i = 1 \cdots n_f$ and $f(\cdot)$ arbitrary function. Note that the extraction of RTM image or CIGs in time-domain are similar as we implement in frequency domain. Here we avoid to show the duplicate extracted images as in the above subsection.

Even though the use of time-domain propagators allows us to computationally feasibly probe EIVs, the singular values for high-frequency EIVs decay slowly (as shown in Figure 2.3), which prevents us from forming low-rank factorizations at these frequencies. Unless we have a solution for this problem, lack of low-rank representations for the EIVs prohibit manipulations such as extracting subsurface offset images and CIPs. In addition, the slow decay of the singular values calls for a larger number of probings, which may render our approach computationally infeasible.

2.4.3 Low-rank representation via rSVD

To efficiently recover the low-rank representation, the forward and adjoint operations in line 2 and 4 will be substituted by the corresponding forward (Equation 2.14) and adjoint based on time-stepping. Also, the QR and SVD factorizations will be overloaded, which implicitly include the operations that loop over all frequencies.

Note that all the QR factorization and SVD decomposition are implemented on the related single monochromatic matrix in looping over all the frequencies. And the extractions of RTM image, CIPs or CIGs based one time-stepping are similar as the way in frequency domain. So far we achieve the computational efficient expression of the low-rank recovery of the full EIVs in time-domain.

2.5 Low-rank factorization with the power method

To address the problem of forming and manipulating EIVs at high frequencies, we propose an alternative approach where we increase the decay of the singular values through linear algebra manipulations. More specifically, we follow recent work by Musco and Musco 2015, which provably offers guarantees on the accuracy of low-rank factorizations in both the Frobenius and spectral norms, and on the accuracy of the factors themselves compared to k -term factorization based on an unattainable singular value decomposition of the original matrix, the EIV in our case. Their core idea to improve the accuracy is to use the fact that the decay of singular values of a matrix increases when we raise this matrix to some $q \geq 1$ power. Due to this property, the accuracy of low-rank factorizations improves since the truncation error decreases because of the increased decay for the singular values. However, as we can see in Algorithm 1, this improvement comes at the cost of having to solve more wave equations. The increase in computational cost depends on the selected power q in line 2, which involve multiple applications of \mathbf{E} and its adjoint.

2.5.1 Power schemes

According to Musco and Musco 2015, the use of the power method for rSVD amounts to the slightly modified Algorithm below.

Algorithm 1 Monochromatic rSVD with simultaneous power iterations (SI)

- 1: given power q and generate n_p random Gaussian vectors $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_{n_p}]$
 - 2: $\mathbf{K} := (\mathbf{E}\mathbf{E}^*)^q \mathbf{E}\mathbf{W}$, $\mathbf{K} \in C^{N \times n_p}$ with probing according to Equation 2.14
 - 3: $[\mathbf{Q}, \sim] = \text{qr}(\mathbf{K})$, $\mathbf{Q} \in C^{N \times n_p}$
 - 4: $\mathbf{Z} = \mathbf{E}^* \mathbf{Q}$, $\mathbf{Z} \in C^{N \times n_p}$
 - 5: $[\Phi, \Sigma, \Psi] = \text{svd}(\mathbf{Z}^*)$, svd computes the top n_p singular vectors
 - 6: set $\Phi \leftarrow \mathbf{Q}\Phi$
 - 7: $\mathbf{L} = \Phi\sqrt{\Sigma}$, $\mathbf{R} = \Psi\sqrt{\Sigma}$
 - 8: Output: factors \mathbf{L}, \mathbf{R} from which actions can be formed via $\mathbf{E} \approx \mathbf{L}\mathbf{R}^*$
-

The above algorithm computes for each frequency, a rank k factorization using $n_p > k$ probings (actions of the double wave equation on random probing factors, see Equation 2.14), a ‘qr’-factorization on a tall matrix and a ‘svd’ on a wide matrix of size $n_p \times N$. After the ‘qr’ factorization, we capture the range of the EIVs in the matrix \mathbf{Q} not to be confused with the source matrix we introduced earlier. After applying the ‘svd’, we obtain the left and right singular vectors collected in Φ and Ψ and Σ , a $k \times k$ matrix with the singular values on its diagonal. As before, the output of Algorithm 1 are the left and right factors \mathbf{L} and \mathbf{R} for each frequency.

Compared to the original rSVD (see Algorithm 6 in the Appendix), Algorithm 1 includes more involved probing (line 2), which now includes the action of \mathbf{E} and $(\mathbf{E}\mathbf{E}^*)^q$. The latter requires q iterations of $\mathbf{K} := (\mathbf{E}\mathbf{E}^*)\mathbf{K}$ where \mathbf{K} is initialized by $\mathbf{K} = \mathbf{E}\mathbf{W}$. For increasing powers of q , the accuracy improves as $\epsilon = \mathcal{O}(\frac{\log N}{q})$, which in practice means that the low-rank factorizations at the higher frequencies become more accurate but this comes at the price of having to carry out an extra $2qn_p$ probings. However, the memory

imprint of Algorithm 1 is roughly the same as Algorithm 6.

While the simultaneous power iterations(SI) in Algorithm 1 allow for an improvement in accuracy by increasing the largest singular values in comparison to the small singular values in the tail, the error only decays linearly in q . To overcome this problem, we follow Musco and Musco 2015 and introduce Algorithm 2, which involves more intricate Block Krylov iterations (BKI) that are better capable of capturing the tail of the singular values. The use of these iterations results in an improvement for the error ($\epsilon = \mathcal{O}(\frac{\log N}{q^2})$), which now decreases quadratically with q . As a consequence, algorithms based on BKI iterations allow for smaller q to attain the same accuracy. However, as we can see in line 2 of Algorithm 2, this improvement goes at the expense of extra memory use because the algorithm works now with multiple vectors defined in terms of the intermediate iterations we used to compute $(\mathbf{E}\mathbf{E}^*)^q\mathbf{K}$. Aside from extra memory use, these additional vectors lead to additional computational costs during the subsequent ‘qr’ and ‘svd’ factorizations, which now involve $(q + 1)n_p$ vectors rather than n_p as before. The number of probings, however, remains the same.

Algorithm 2 Monochromatic rSVD with block Krylov iteration (BKI)

- 1: given power q and generate n_p random Gaussian vectors $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_{n_p}]$
 - 2: $\mathbf{K} := [\mathbf{E}\mathbf{W}, (\mathbf{E}\mathbf{E}^*)\mathbf{E}\mathbf{W}, \dots, (\mathbf{E}\mathbf{E}^*)^q\mathbf{E}\mathbf{W}]$, $\mathbf{K} \in \mathbb{C}^{N \times (q+1)n_p}$
 - 3: $[\mathbf{Q}, \sim] = \text{qr}(\mathbf{K})$, $\mathbf{Q} \in \mathbb{C}^{N \times (q+1)n_p}$
 - 4: $\mathbf{Z} = \mathbf{E}^*\mathbf{Q}$, $\mathbf{Z} \in \mathbb{C}^{N \times (q+1)n_p}$
 - 5: $[\Phi, \Sigma, \Psi] = \text{svd}(\mathbf{Z}^*)$, svd computes the top n_p singular vectors
 - 6: set $\Phi \leftarrow \mathbf{Q}\Phi$, \mathbf{Q} choose the first n_p singular vectors
 - 7: $\mathbf{L} = \Phi\sqrt{\Sigma}$, $\mathbf{R} = \Psi\sqrt{\Sigma}$
 - 8: Output: factors \mathbf{L}, \mathbf{R} from which actions can be formed via $\mathbf{E} \approx \mathbf{L}\mathbf{R}^*$
-

2.5.2 Complexity analysis for power scheme based rSVDs

Although with the same probing size, the recovery accuracy of EIV based on SI and BKI increase with power, the investment in computation and storage also increase with the power. It is necessary to draw complexity analysis and comparison to guide in investing strategy when we have limited resources.

Step	Size	Cost
1). Generate an $N \times n_p$ Gaussian random matrix \mathbf{W}	$N \times n_p$	-
2). Form $\mathbf{K} = (\mathbf{E}\mathbf{E}^*)^q \mathbf{E}\mathbf{W}$	$N \times n_p$	$2n_p + 4qn_p$ wave equations
3). Construct $[\mathbf{Q}, \sim] = qr(\mathbf{K})$	$N \times n_p$	$\mathcal{O}(Nn_p^2)$ flops
4). Form $\mathbf{Z} = \mathbf{E}^* \mathbf{Q}$	$N \times n_p$	$2n_p$ wave equations
5). $[\Phi, \Sigma, \Psi] = \text{svd}(\mathbf{Z}^*)$	$n_p \times n_p, n_p, N \times n_p$	$\mathcal{O}(Nn_p^2)$ flops
6). Update $\Phi \leftarrow \mathbf{Q}\Phi$	$N \times n_p, n_p, N \times n_p$	

Table 2.1: Storage and computational cost for rSVD with simultaneous iterations

Step	Size	Cost
1). Generate an $N \times n_p$ Gaussian random matrix \mathbf{W}	$N \times n_p$	-
2). Form $\mathbf{K} = [\mathbf{E}\mathbf{W}, \dots, (\mathbf{E}\mathbf{E}^*)^q \mathbf{E}\mathbf{W}]$	$N \times (q+1)n_p$	$2n_p + 4qn_p$ wave equations
3). Construct $[\mathbf{Q}, \sim] = qr(\mathbf{K})$	$N \times (q+1)n_p$	$\mathcal{O}(N((q+1)n_p)^2)$ flops
4). Form $\mathbf{Z} = \mathbf{E}^* \mathbf{Q}$	$N \times (q+1)n_p$	$2(q+1)n_p$ wave equations
5). $[\Phi, \Sigma, \Psi] = \text{svd}(\mathbf{Z}^*)$ with only n_p singular vectors	$n_p \times n_p, n_p, N \times n_p$	$\mathcal{O}(N(q+1)n_p^2)$ flops
6). Update $\Phi \leftarrow \mathbf{Q}\Phi$	$N \times n_p, n_p, N \times n_p$	-

Table 2.2: Storage and computational cost for rSVD with block Krylov iteration

In an effort to deal with the challenge of factorizing large-scale EIVs at high frequencies, we introduce an algorithm based on probing alone, rSVD (see Algorithm 6 in the Appendix), and more involved algorithms based on SI (Algorithm 1) and BKI (Algorithm 2) iterations, designed to handle situations where the singular values decay more slowly. These three algorithms differ in attainable accuracy as a function of the number of probings, memory use, and computational expense to carry out the ‘qr’ and ‘svd’ factorizations. With these different approaches, we have freedom to select the algorithm that best fits our needs. To help with this selection process, we include Tables 2.1 and 2.2. Based on these tables, we can make the following observations: *(i)* the accuracy of the factorizations based on SI and BKI increases with the power q ; *(ii)* as we increase q , the computational cost increases for both SI and BKI iterations; and *(iii)* the memory use and computational cost increase for BKI with increasing q with an error that decreases quadratically. Remember that errors in our context refer to inaccuracies related feasible SVDs based on random-probing compared to the inaccessible ground truth given by the k -term SVD derived from the full EIV. This means that we assumed the number of probing vectors to be fixed and equal to k .

Our main goal is to get the most accurate k -term factorization of EIVs through $n_p = k$ random probings with the double wave equation (cf. Equation 2.14). Because all subsequent manipulations on the factored form of these EIVs scale with k , whether we extract CIPs, derive subsurface offset images, redatum or carry out velocity continuation

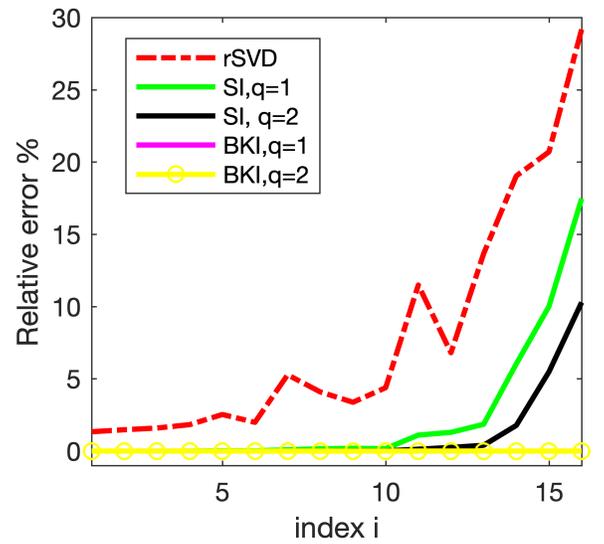
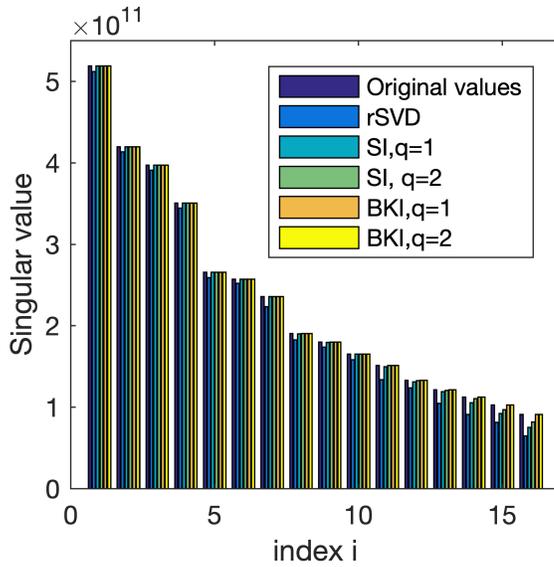
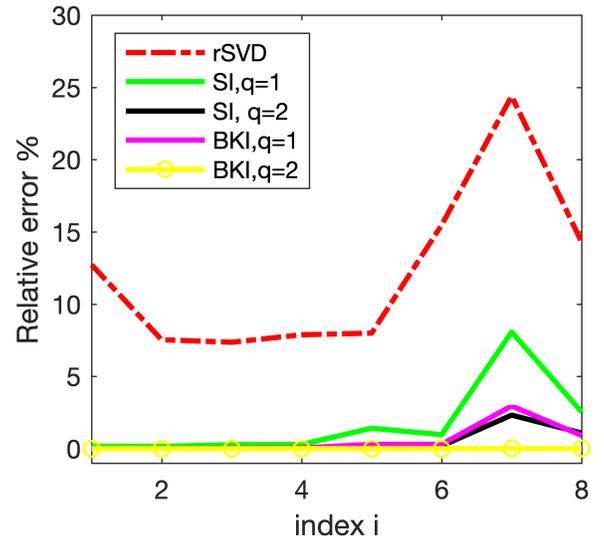
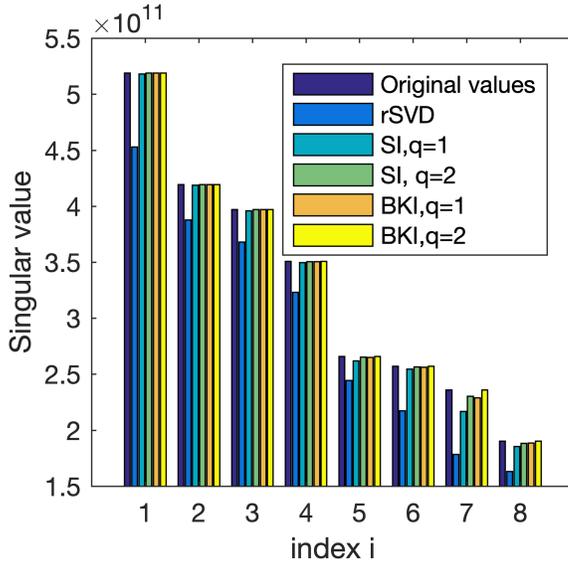
(Leeuwen, Kumar, and Herrmann 2017), we want to control the error. Contrary to truncation errors in conventional svd-based factorizations, factorizations based on randomized probing have additional inaccuracies related to the tail of the singular values. When moving to higher frequencies, it is important that we control these additional errors because the singular values decay slower at these frequencies.

2.5.3 Numerical experiments for power schemes

To illustrate the performance of SI and BKI iterations compared to conventional rSVD, we carry out experiments on the small 25Hz monochromatic frequency slice of the small EIV included in Figure 2.1, for which we can form the EIV itself and its factorization explicitly via a conventional SVD. Our results are summarized in figure 2.4, where we plot the first n_p singular values for factorization based on probings with $n_p = 8, 16, 30$. Comparing the bar plots in Figure 2.4a, 2.4c, and 2.4e, leads to the following observations. First, inaccuracies in the estimates for the singular values are large when there is a large truncation error—i.e., when there is still a lot of energy left in the tail. In that case, there is a large difference between the actual singular value (depicted in dark blue) and the singular values obtained by the standard rSVD method. These errors are much smaller when using factorizations based on SI and BKI iterations for either $q = 1$ or $q = 2$. Smaller errors in the singular values lead to smaller errors in the factorization. Second, the plots for the relative errors in Figure 2.4b, 2.4d, and 2.4f show a rapid increase towards the smaller singular values. Even for a probing size n_p as high as 30 where we leave only 0.4% of the total energy in the tail, the relative error for 30th singular value calculated by rSVD exceeds 35%. Both power methods SI and BKI help to accurately recover the singular values and decrease the relative errors. Even for a very small probing size of $n_p = 8$, where we leave 11.4% energy in the tail, the BKI with power $q = 2$ recovers the first n_p singular values very well. SI, on the other hand, still leaves some errors in the recovery for $q = 1, 2$. For $n_p = 16$, the remaining energy in the tail decreases to 5% and the BKI recovers the n_p singular values

very well with $q = 1, 2$. We can also see that BKI with $q = 1$ works even better than SI for $q = 2$. In summary, both the SI and BKI methods help in decreasing the relative errors in the singular values. With the same probing size n_p and power q , BKI always outperforms the other methods.

Usually BKI with $q = 1$ can satisfy our requirement of a good recovery of the first n_p singular values. As we can observe from Figure 2.5, the errors in the RTM recovered by the rSVD method (cf. Figure 2.5a and b) for $n_p = 8$ are more obvious than those from the BKI with $q = 1$ (cf. Figure 2.5a and c). For the rSVD, coherent energy is lost (cf. Figure 2.5d and e).



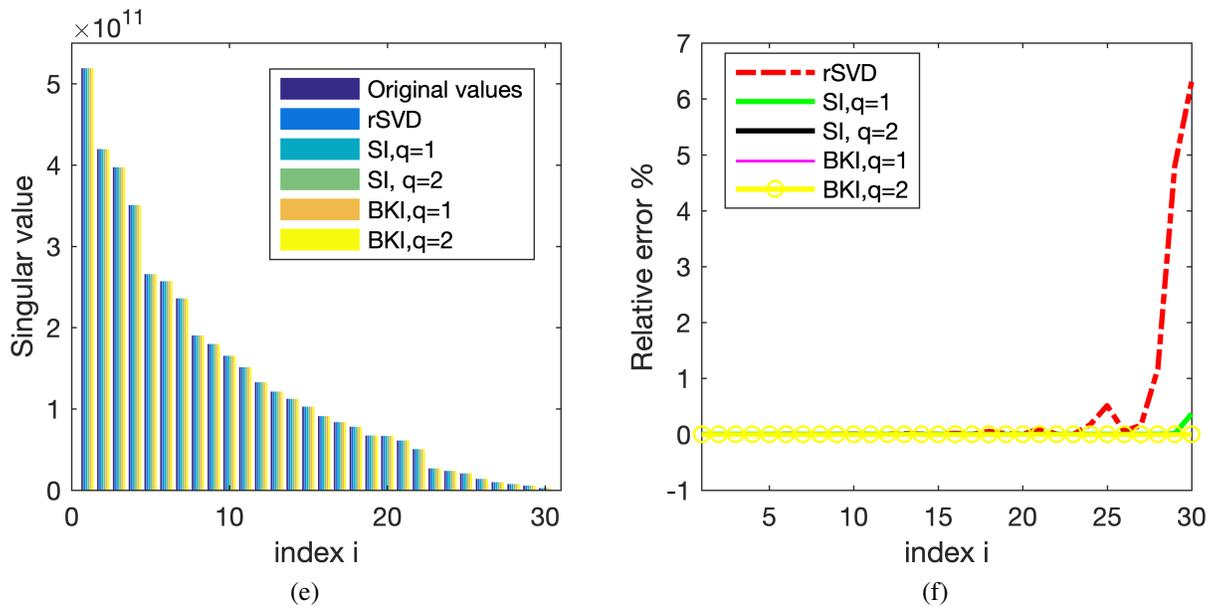


Figure 2.4: (a) The singular value bars obtained by SVD, rSVD, SI and BKI with $n_p = 8$. (b) The corresponding relative errors of the singular values obtained by rSVD, SI and BKI. (c) The singular value bars obtained by SVD, rSVD, SI and BKI with $n_p = 16$, (d) the corresponding relative errors of the singular values obtained by rSVD, SI and BKI. (e) The singular value bars obtained by SVD, rSVD, SI and BKI with $n_p = 30$. (f) The corresponding relative errors of the singular values obtained by rSVD, SI and BKI.

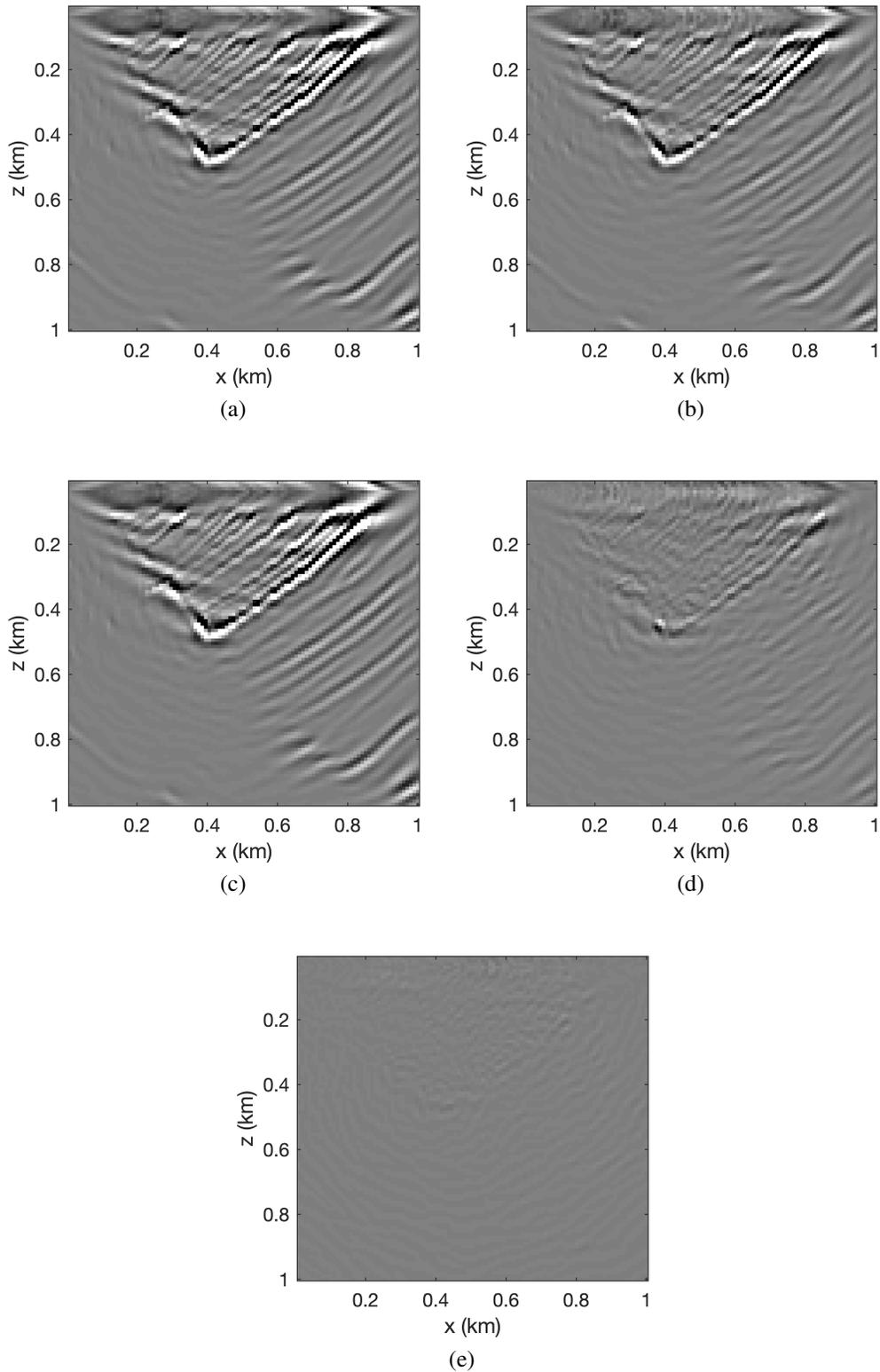


Figure 2.5: Comparison between RTMs obtained via carrying out the conventional SVD(a), the rSVD(b),and BKI for $q = 1$. We only use the first eight singular values – i.e. $n_p = 8 \ll n_s$ where $n_s = 100$. (d) The difference between (a) and (b), (e) the difference between (a) and (c).

2.6 Slicing and dicing

Now that we have established a method to factorize EIVs based on time-domain propagators, we would like to discuss how to extract different gathers without having to form the EIVs themselves explicitly (see also Da Silva, Zhang, Kumar, and Herrmann 2019 who accomplished the same when using a tensor factorization in the hierarchical Tucker format). Aside from having major advantages regarding memory use and storage, all described operations scale with the rank of the factorization and as long as this rank $k < \frac{n_s}{4}$ we gain computationally compared to methods that loop over shots. In addition, after the EIVs are factorized, no additional wave-equation solves are needed to extract the gathers. We are interested in three different gathers, namely Common Image Point gathers (CIPs), Common Image Gathers (CIGs), and geological dip corrected CIGs. The latter correspond to CIGs where we compute the subsurface offset in the direction perpendicular to the geological dip. As outlined by Leeuwen, Kumar, and Herrmann 2017, including this rotation has advantages for amplitude versus offset analyses and as we will show that it leads to improved focusing.

To set the stage, we use the following notation for the discretized image volume $\mathbf{E}[i_z, i_x; j_z, j_x]$ with $i_z = 1 \cdots n_z$, $i_x = 1 \cdots n_x$ the indices along the spatial coordinates (we use the "FORTRAN" convention where first dimension runs over the rows) and with $j_z = 1 \cdots n_z$, $j_x = 1 \cdots n_x$ the indices that run over the second set of coordinates. For notational simplicity we drop the frequency index. We impose the time imaging condition by summing over this index after extracting the different gathers.

While formally a matrix of size $N \times N$, we consider the discretized image volume as a four dimensional array. Similarly, we can regard the factors as multi-dimensional arrays—i.e., we have $\mathbf{L}[i_z, i_x; i_p]$ with $i_p = 1 \cdots n_p$ and $\mathbf{R}[j_z, j_x; i_p]$ as depicted in Figure 2.6a). We adopt the Matlab-like $:$ notation to extract vectors or matrices from these multi-dimensional arrays.

According to Leeuwen, Kumar, and Herrmann 2017, a CIP gather indexed by a single point (i_z, i_x) is given by the following 2D slice $\mathbf{E}[i_z, i_x; :, :]$, indicated by the green plane in Figure 2.6b. To avoid having to form the EIV explicitly, we implement the extraction of CIPs directly on the factors as outlined in Algorithm 3. In this algorithm, we extract the vector $\mathbf{l} = \mathbf{L}[i_z, i_x; :] \in C^{1 \times n_p}$ once followed by a loop over depth during which this vector is applied to the transpose of the matrix $\mathbf{R}[j, :, :] \in C^{n_x \times n_p}$ for $j = 1 \dots n_z$.

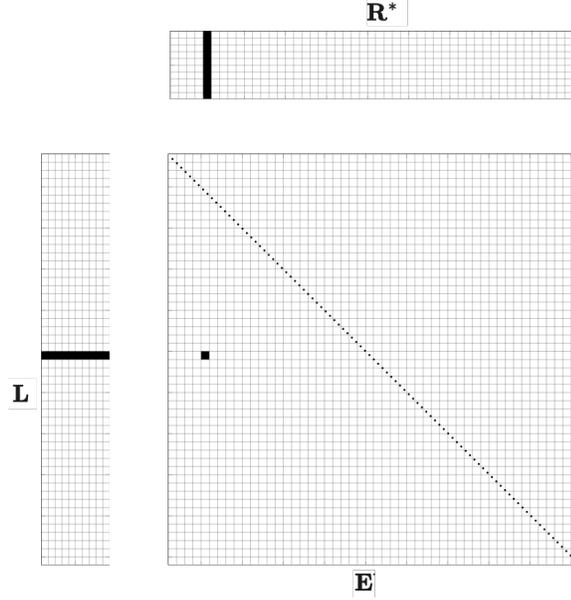
Algorithm 3 Pseudo code for CIP gather extraction

- 1: **Input:** location common image point (i_z, i_x) and low rank factors $\{\mathbf{L}, \mathbf{R}\}$
 - 2: extract the vector $\mathbf{l} = \mathbf{L}[i_z, i_x; :] \in C^{1 \times n_p}$
 - 3: **for** $j=1:n_z$
 - 4: $\mathbf{E}[i_z, i_x; j, :] = \mathbf{lR}^*[j, :, :]$
 - 5: **end**
 - 6: **output:** Real part of $\mathbf{E}[i_z, i_x; :, :]$
-

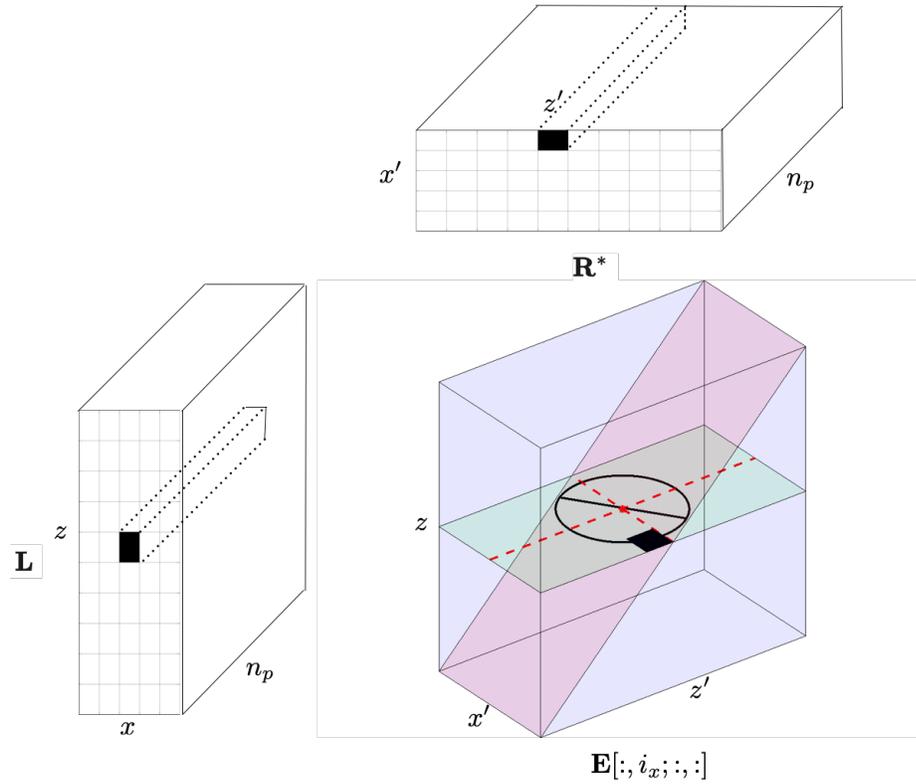
CIG gathers for horizontal subsurface offset correspond to extracting $\mathbf{E}[i_z, i_x; j_z = i_z, :]$, $i_z = 1 \dots n_z$ as depicted by the red plane in Figure 2.6b. Algorithm 4 extracts CIGs along all depth and at a single lateral index i_x by extracting the vector $\mathbf{l} = \mathbf{L}[j, i_x; :] \in C^{1 \times n_p}$ now within the loop over the vertical coordinate followed by a multiplication with the matrix $\mathbf{R}^*[j, :, :]$. The resulting CIG corresponds to real part of $\mathbf{E}[i_z, i_x; j_z = i_z, :]$ for $i_z = 1 \dots n_z$.

Algorithm 4 Pseudo code for CIG gather extraction

- 1: given the lateral index of the CIG i_x and low rank factors $\{\mathbf{L}, \mathbf{R}\}$
 - 2: **for** $j=1:n_z$
 - 3: extract the vector $\mathbf{l} = \mathbf{L}[j, i_x; :] \in C^{1 \times n_p}$
 - 4: $\mathbf{E}[j, i_x; j, :] = \mathbf{lR}^*[j, :, :]$
 - 5: **end**
 - 6: **output:** Real part of $\mathbf{E}[i_z, i_x; j_z = i_z, :]$ for $i_z = 1 \dots n_z$
-



(a) Element-wise extraction schematic diagram



(b) Element-wise extraction and different slices through the 3D sub-cube of EIV

Figure 2.6: (a) shows the low-rank representation **L** and **R** and organize the 4D EIV into 2D matrix. The diagonal dash line indicates where the traditional RTM images extracted from. (b) presents the **L** and **R** by 3D cubes whose third dimension is the probing size n_p , and the 3D sub-cube of EIV is $\mathbf{E}[:, i_x; :, :]$ by fixing x dimension with i_x , where the green cross section $\mathbf{E}[i_z, i_x; :, :]$ indicates one CIP gather at i_z, i_x , and the red slice $\mathbf{E}[i_z, i_x; j_z = i_z, :], i_z = 1 \cdots n_z$ indicates one CIG gather at i_x along all the depth.

In addition to the CIPs and CIGs, we introduce geologic-dip corrected CIGs that combine information from both gathers by estimating the geologic dip as a function of depth along the CIG. We find the geologic dip by maximizing the stack power in CIPs as illustrated in Figure 2.7. As shown by Leeuwen, Kumar, and Herrmann 2017, CIGs computed with horizontal offsets poorly focus when reflectors are steeply dipping. These authors showed that there is a complete lack of focusing for vertical reflectors using horizontal offsets and for horizontal reflectors using vertical offsets. Following their work, we compute the geologic dip using stack power and subsequently correct for it such that the offset direction is always taken perpendicular to the geological dip. The procedure that we follow to correct CIGs is summarized in Algorithm 5.

Algorithm 5 Pseudo code for dip corrected CIG gather extraction

```

1: Input: lateral index  $i_x$ , offset range  $d$ , number of angles  $n_\theta$ , and  $\{\mathbf{L}, \mathbf{R}\}$ 
2:  $\theta = [0, \pi]$ 
3: for  $j=1:n_z$ 
4:   extract the vector  $\mathbf{l} = \mathbf{L}[j, i_x; :] \in C^{1 \times n_p}$ 
5:   for  $i=1:n_\theta$ 
6:     find indices  $I_z, I_x = \{(i_z, i_x) \mid \text{along lines with angle } \theta_i\}$ 
7:     extract  $\widehat{\mathbf{R}} = \mathbf{R}[I_z, I_x; :]$ 
8:      $\mathbf{e} = \widehat{\mathbf{R}}^*$ 
9:     Stack Power  $\Gamma[i, j] = \|\mathbf{e}\|^2$ 
10:   end
11:   find the maximum of  $\Gamma[:, j]$  and the corresponding  $\theta_{max}$  and  $\mathbf{e}$  perpendicular to
       $\theta_{max}$ 
12:    $\text{CIG}_{dip}[:, j] = \mathbf{e}$ 
13: end
14: output: Real part of  $\text{CIG}_{dip}$  and the stack power image  $\Gamma$ 

```

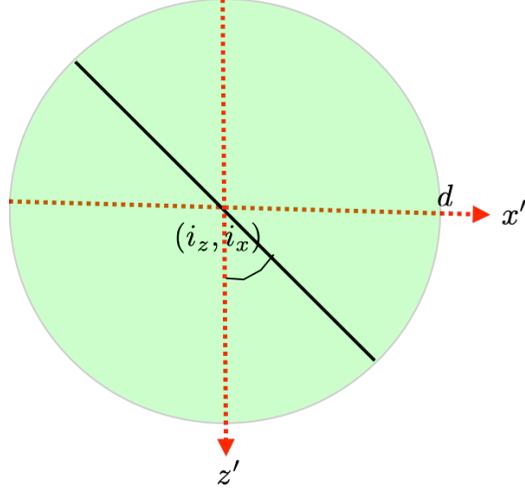


Figure 2.7: schematic diagram for stacking power on the slice $\mathbf{E}[i_z, i_x; :, :]$ within d offset.

2.7 Numerical experiments

We presented a numerical framework to represent and manipulate EIVs by low-rank factorizations obtained via randomized probing and BKI iterations. We argue that these factorizations are a natural parameterization for full-subsurface offset EIVs. We will now show how these factorizations can be applied to a series of imaging problems with an emphasis on how to make informed choices on the rank and the order of the BKI method given computational constraints. As we observed from the example with the explicit EIV calculated from a small subset of the Marmousi model, the BKI method with power $q = 1$ outperforms the rSVD and SI. For this reason, we will employ BKI iterations for $q = 1$ for the remainder of the paper.

Our imaging experiments will be conducted on the Marmousi model and are designed to demonstrate our ability to compute RTM images, CIPs, CIGs, and dip corrected CIGs, from factorization obtained with randomized probing and BKI iterations. To establish accuracy of the proposed method, we compare exact CIPs and CIGs with their approximate counterparts derived from the low-rank factorization without the need to form the full EIV. To handle imaging problems with steep dips, we show how CIGs can be calculated that

correct for the local geologic dip.

To verify the validity of the proposed factorization, we conduct a series of imaging experiments that involve the Marmousi model, which is 8km wide and 3.2km deep and discretized on a 6×6 m grid. We acquire data from 650 co-located sources and receivers positioned at a depth of 18m and sampled with a 12m interval horizontally. We simulate data with the acoustic constant density wave equation with an absorbing boundaries all around and a Ricker wavelet centered at 23Hz. Before imaging, we remove the direct wave from the simulated "observed" data.

Given this factorization, we compute a migrated image via Equation 2.9 and we compare this result with a regular computed over all $n_s = 650 \gg 130$ sources. Aside from some noise, the migrated image obtained from the factorization and conventional RTM compare well (cf. Figure 2.9b and 2.9a). As we can see in Figure 2.9, the extracted RTM from the recovered low-rank factorization of EIV has good quality, the image is not perfect due to the lose of the energy presented by the singular values after 130th singular. The fact that these images are not the same is not surprising because our factorization is approximate, which makes the image a bit noisy and the amplitudes are slightly less well resolved. However, we argue that this is a relatively small price to pay since the factorization gives us access to much more information such as CIP gathers without the need to compute additional wave-equation solves. To demonstrate that this is indeed the case, we compare in Figure 2.10a true CIP, obtained by time-domain probing of the EIV with a bandwidth-limited point source located at ($z = 870$ m, $x = 5250$ m) (Leeuwen, Kumar, and Herrmann 2017), with a CIP computed from the factors using Algorithm 3. As with the RTM image itself, the CIP derived from the factors while noisy captures most of the energy. As with the true CIP, the approximate CIP shows a nice directivity pattern with a rotation that is consistent with the geologic dip. Remember that the approximate CIP did not require additional wave-equation solves. Finally, we also computed three CIGs at $x = 1.8, 3.6, 5.4$ km for an offset range between -150 to 150 m. The results are included in Figure 2.11. Again,

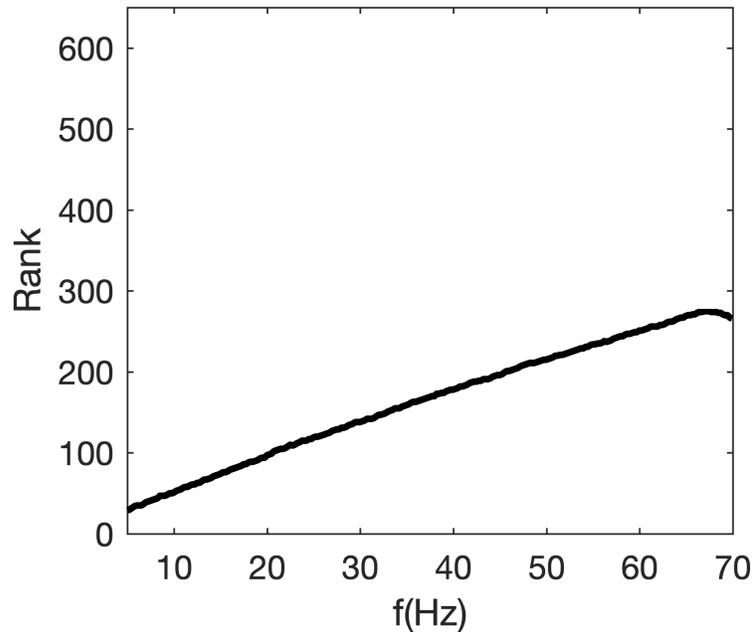
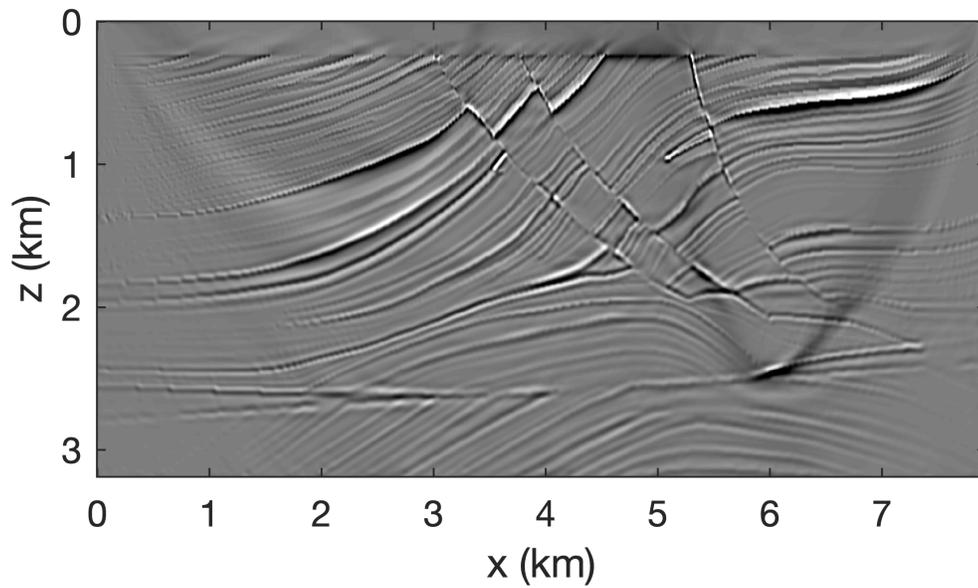


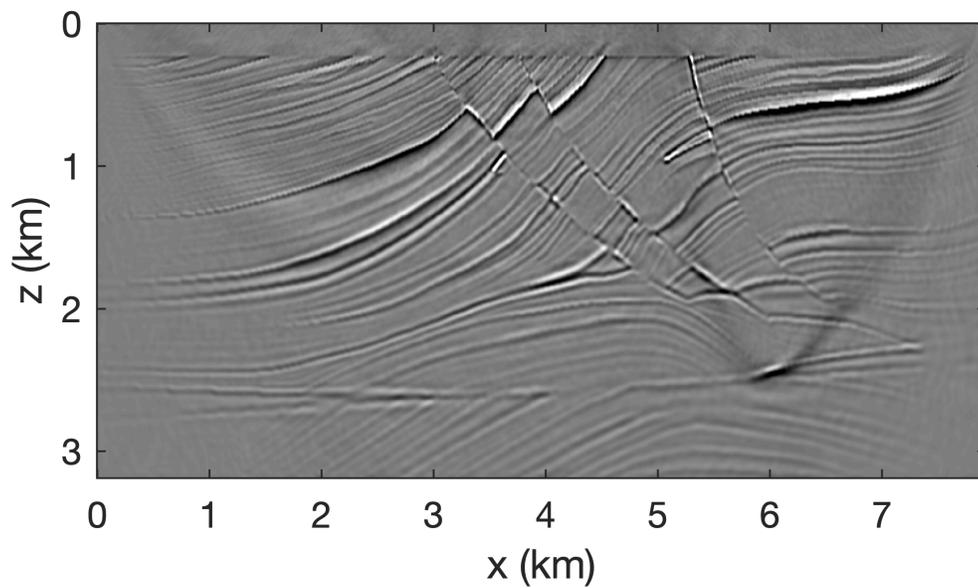
Figure 2.8: Selected rank to capture 95% of the data's energy as a function of frequency.

the results computed from the factorization with Algorithm 4 compare well with the true CIGs. Except for the presence of some noise, the approximate CIGs (Figure 2.11b) capture the behavior of the true CIGs (Figure 2.11b). As expected, the energy is well focused for flat reflectors because the background velocity model is kinematically correct.

To improve the focusing of CIGs for steeply dipping reflectors, we ran Algorithm 5 at $x = 5.4\text{km}$. This algorithm is designed to correct for the geologic dip so that the subsurface offset is always taken in the direction perpendicular to the reflector. As we can see from Figure 2.12a, the CIG is not well focused at locations where the geologic dip is steep. We can correct for this geologic dip by computing the stack power of CIPs for each depth level along lines with different angles (see Figure 2.7). The stack power is maximum when the angle is close to the geologic dip as can be seen in Figure 2.12, where the stack power is plotted as a function of the angle for three different depth levels. By using the angles where the stack power is maximum, we are able to correct for the geologic dip by rotating the direction of the subsurface offset by 90 degrees compared to this angle. The dashed white



(a)



(b)

Figure 2.9: Comparison RTM images. (a) The true conventional RTM image obtained by migration 650 shot records and (b) the RTM image computed with a factorization with the BKI method for $q = 1$ and $n_p = 130$.

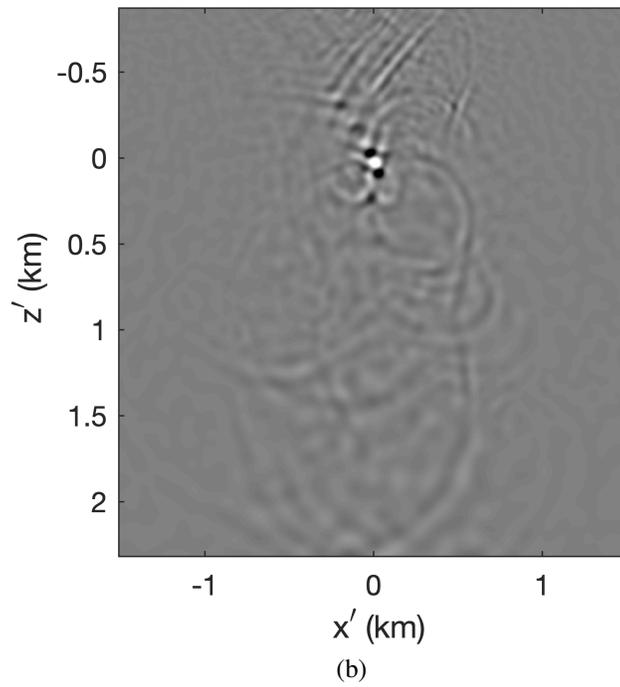
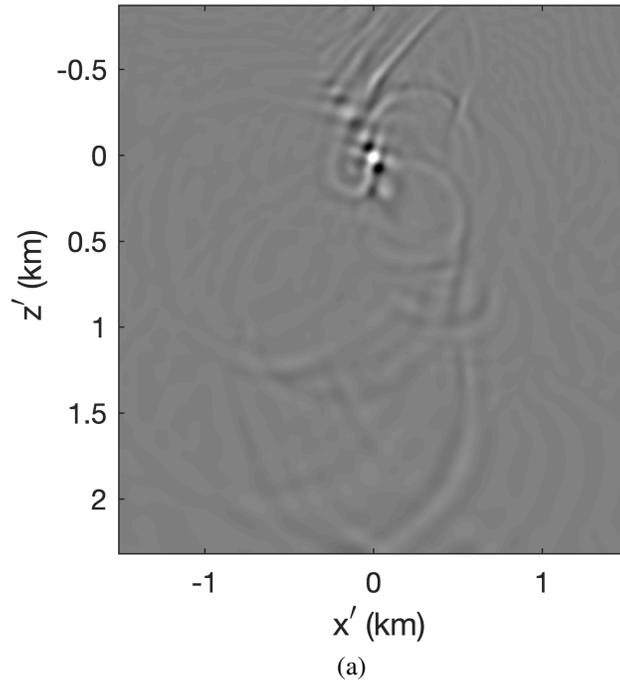
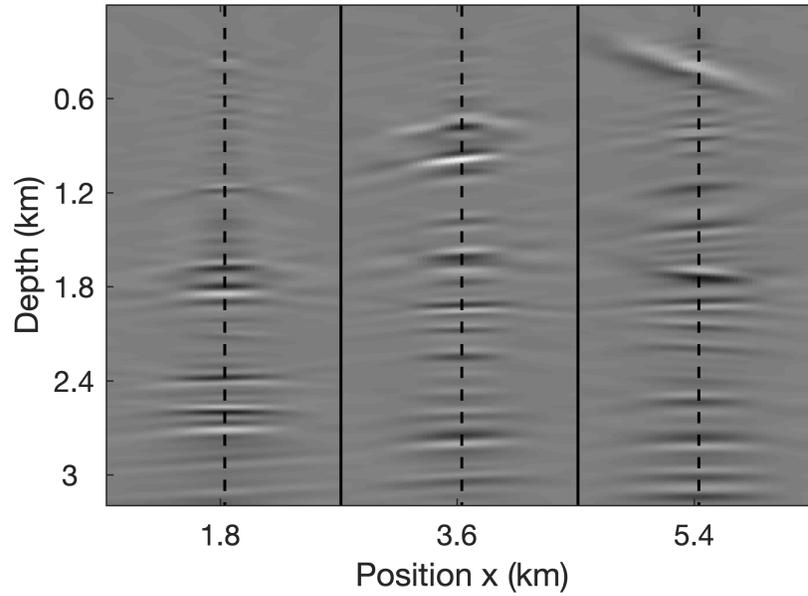
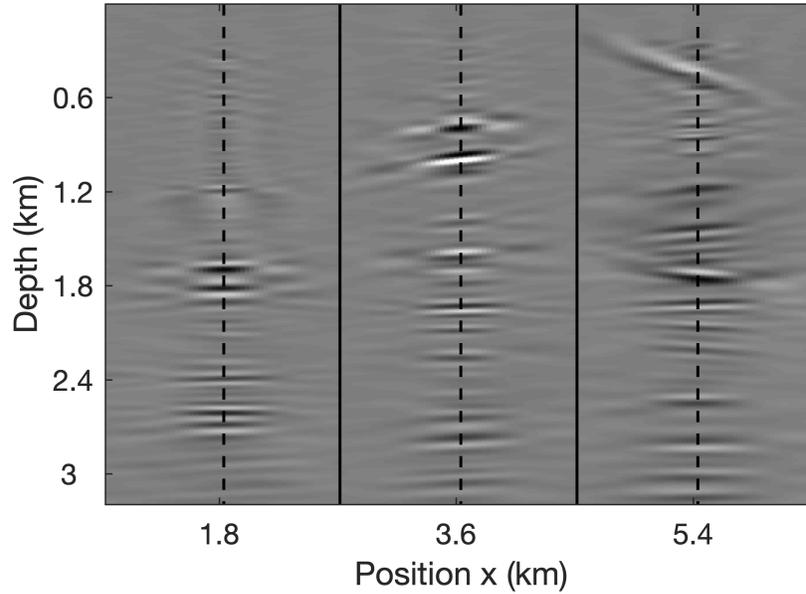


Figure 2.10: Comparison CIP images at ($z = 870m, x = 5250m$). (a) The true CIP image via probing with a bandwidth-limited point source and (b) the CIP image extracted with Algorithm 3.



(a)



(b)

Figure 2.11: The exact and approximate CIGs at $x = 1.8, 3.6, 5.4$ km, where the offset range is from -150 to 150 m. (a) The merged exact CIGs, (b) the recovered CIGs from the low-rank representation via Algorithm 4.

lines in Figure 2.12a correspond to the estimated geologic dips, which is close to the true but unknown geologic dip. The CIGs computed with this correction using Algorithm 5 are included in Figure 2.12c. Compared to the CIG included in Figure 2.12a, the corrected CIG is much better focused in areas where the geologic dip is large.

These examples nicely demonstrate that accurate imaging results can be obtained using our factored formulation. Aside from being able to approximate RTM images well, the factored form also provides rapid access to CIPs and (dip-corrected) CIGs at no additional wave-equation solves. This is made possible by working with the low-rank factored form without the need to form full EIVs.

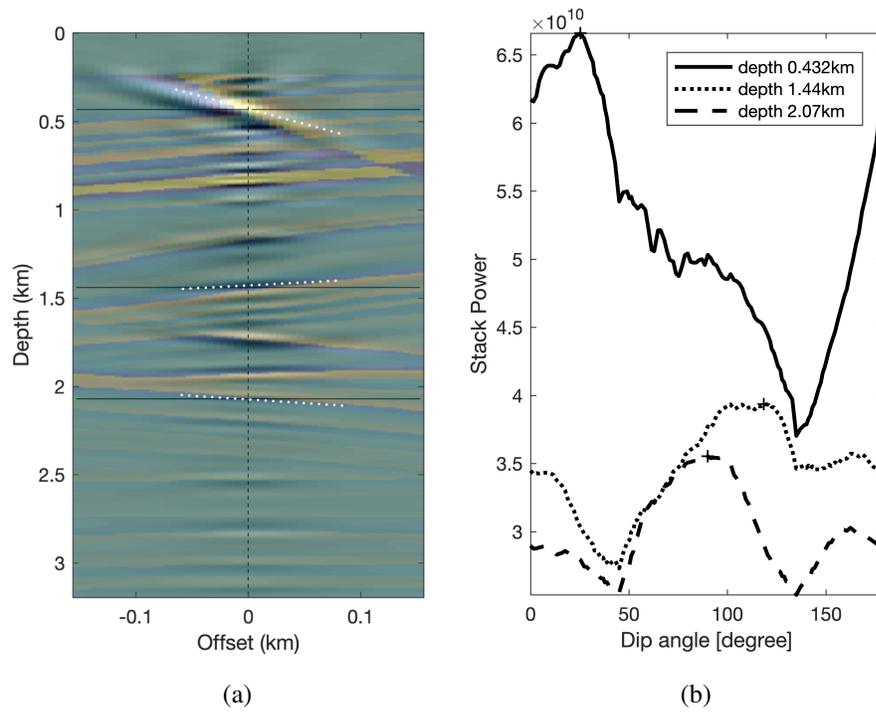


Figure 2.12: (a) The original CIG image at $x = 5.4\text{km}$ with offset range 150m, (b) the stack power curves at depth $z = 0.432, 1.44, 2.07\text{km}$, (d) the corrected CIG image with the dipoles from stack power.

2.8 Discussion

Aside from incurring the cost of solving two wave equations per source, conventional RTM imaging runs substantial costs in computing subsurface offset gathers such as CIGs (Symes 2008; Stolk, Hoop, and Symes 2009; Rickett and Sava 2002) or angle-domain CIGs (De Bruin, Wapenaar, and Berkhout 1990; Sava and Fomel 2003; Kroode 2012; Köhl and Sacchi 2003; Mahmoudian and Margrave 2009; Dafni and Symes 2016b; Dafni and Symes 2016a). CIGs are routinely used during migration-velocity and amplitude-versus offset analyses (De Bruin, Wapenaar, and Berkhout 1990) and as part of quality control during (automatic) velocity-model building (Symes and Carazzone 1991; Shen and Symes 2008). Calculation of these gathers often occurs via brute force cross-correlations between space-, or sometimes time- (Sava and Fomel 2006), shifted versions of the forward and adjoint wavefields. Depending on the number of offsets and the number of CIGs, the costs of these multidimensional cross-correlations (Sava and Vasconcelos 2011b) can become comparable to calculating the wave equation solves themselves.

Probing techniques based on the double two-way wave-equation (cf. Equation 2.6) avoid some of these costs by computing CIPs for all subsurface offsets at the price of only two wave equation solves and a multi-dimensional convolution with the data matrix per probing vector. While this probing technique, introduced by Leeuwen, Kumar, and Herrmann 2017, gives us access to objects (e.g. CIPs) to which we normally would not have access, its complexity scales linearly with the number of CIPs, which rapidly becomes computationally infeasible. By using randomized probing techniques in combination with Block Krylov iterations, we overcome this shortcoming by casting EIVs in an approximate low-rank factored form. As we have shown, this factored form gives us access to conventional RTM images (cf. Equation 2.9), various subsurface-offset gathers (Algorithms 3 – 5), and multi-scenario imaging with costs that scale with the number of factors n_p . This number is typically much smaller than the number of source experiments, $n_s \ll n_p$.

While the examples were all in 2-D, our formulation is suitable to scale to 3-D for the following reasons: *(i)* our use of highly optimized time-domain finite-difference propagators from Devito; *(ii)* our Fourier-domain implementation of the multi-dimensional convolution with the data matrix (see Equations 2.7 or 2.14); and *(iii)* the factorizations themselves. Because we work with subsets of frequencies, we are able to limit memory use and the compute needed to factorize. For now, we implemented the probing with the regular Fourier transform, followed by subsampling, which requires storage of the full wavefield. As shown recently by Witte, Louboutin, Luporini, Gorman, and Herrmann 2019, we can remove the need to store the full wavefield by using the on-the-fly Fourier transform. Since the factors are in the Fourier domain, it is trivial to implement the zero-time imaging condition via a simple stack.

In addition to having a computational feasible and manipulatable representation for EIVs, our factorization allows for the establishment of a completely new iterative seismic imaging workflow during which

1. we follow the heuristic explained in the experiment section and select n_p , followed by probing with random Gaussian vectors to calculate $\mathbf{K} := [\mathbf{E}\mathbf{W}, (\mathbf{E}\mathbf{E}^*)\mathbf{E}\mathbf{W}, \dots, (\mathbf{E}\mathbf{E}^*)^q\mathbf{E}\mathbf{W}]$ (line 1 Algorithm 2) requiring $6n_p$ wave-equation solves when we set $q = 1$. From \mathbf{K} , we compute its QR-factorization, followed by another probing with \mathbf{E}^* at $2n_p$ wave-equation solves in turn followed by an SVD producing our factorization $\{\mathbf{L}, \mathbf{R}\}$ with $\mathbf{E} \approx \mathbf{L}\mathbf{R}^*$ for each frequency.

2. we have access to migrated images via Equation 2.9, to CIPs (via Algorithm 3), and (dip-corrected) CIGs (via Algorithms 4 or 5) at costs that scale with n_p and which do not require additional wave-equation solves. We compute these gather for each frequency, followed by summing to impose the zero-time imaging condition.

Aside from having access to different kinds of subsurface offset CIGs or angle-domain ADCIGs (Dafni and Symes 2016b), this new imaging scheme has the advantage that it can relatively cheaply recompute these CIGs for a different velocity model via velocity

continuation. We consider this as a highly desirable feature. For instance, this feature would allow us to recompute CIGs for quality control during velocity model building. It would also allow for the evaluation of different models during redatuming (Kumar, Graff-Kray, Vasconcelos, and Herrmann 2019).

Subsurface-offset image gathers exist in various forms and are parameterized by subsurface offset as in CIGs or by angle as in angle-domain CIGs (ADCIGs). In either case, the parameterization of these gathers, and their recent extensions including dip-angle decomposition (Dafni and Symes 2016b; Dafni and Symes 2016a) or micro-local parameterization (see e.g. Kroode 2012), does not make use of the underlying low-rank structure of EIVs. By explicitly using this low-rank structure, our ability to probe, factorize, and velocity-continue, we offer an alternative formulation where the underlying linear algebra offers a natural and scalable parameterization. Informed by the singular-value decay of the data and tolerance for errors, we make an informed decision on the underlying rank n_p . This number determines the overall computational complexity. As long as n_p is sufficiently small, our formulation can arguably compete computationally while offering unique features such as access to arbitrary subsurface-offset or angle gathers, to geologic dips, and to the option to recompute these gathers for different background velocity models at significantly reduced costs.

The above workflow during which we produced geologic-dip corrected CIGs is one example of what our factored approach has to offer. Other imaging schemes are possible. Since we have access to omni-directional subsurface offset gathers, we have flexibility to derive filters designed to remove certain imaging artifacts as recently proposed by Dafni and Symes 2016a. Since CIPs contain the full scattering information for each point in the image, we have access to the local geologic dip. The latter corresponds to the specular dip angle of reflection discussed in recent work by Dafni and Symes 2016a.

In addition to allowing for manipulations of full subsurface-offset EIVs, the proposed formulation essentially boils down to an imaging algorithm with a computational com-

plexity that scales with the number of probing factors n_p instead of with the number of shots n_s . We found empirically, that the singular values of EIVs decay faster than those of monochromatic data matrices. This allows us to choose the probing size and an imaging paradigm with determined by n_p , which is smaller than n_s and arguably also smaller than a low row-rank approximation of the data matrix as proposed by Hu, Abubakar, and Habashy 2009. In future work, we plan to select the rank adaptively per frequency, which should increase the performance of or low-rank factorization even further.

2.9 Conclusion

Wave-equation based imaging techniques, such as reverse-time migration including the formation and manipulation of subsurface-offset gathers, is becoming more and more common place in modern-day seismic imaging workflows. While subsurface-offset image gathers carry important information on the velocity model and the local scattering mechanism, they are because of their high dimensionality difficult to form and manipulate. By combining probing of full-subsurface offset extended image volumes via the double two-way wave equation with techniques from randomized linear algebra, we were able to cast these extended image volumes into a highly compressed and manipulatable factored form. To meet the demands of high-resolution imaging, we based our factorization on probing with the time-domain wave equation and an advanced Block-Krylov randomized singular-value decomposition technique. The latter is designed to increase the accuracy of the factorization for high frequencies where the singular values decay more slowly. Given this factorization, we demonstrated how various subsurface image gathers can be computed without having to form the extended image volume explicitly.

While the initial cost of the factorization may exceed the cost of regular reverse-time migration, the factors give us access to gathers as a function of the omni-directional subsurface offset. These gathers allows for the computations of geologic-dip corrected common image gathers that remain focused in situations where the reflectors are strongly dipping.

Common image gathers based on horizontal offset alone do not focus in this situation even when the background velocity model is correct.

2.10 Appendix

Due to the low-rank property of the EIV, we expect to express the monochromatic \mathbf{E}_i as the products of small or narrow matrices. For instance, we can approximate \mathbf{E}_i with the singular value decomposition truncated for the n_p larger singular values of \mathbf{E}_i :

$$\mathbf{E}_i \approx \mathbf{\Phi}_i \mathbf{\Sigma}_i \mathbf{\Psi}_i^*, \quad (2.15)$$

where $\mathbf{\Phi}_i$ and $\mathbf{\Psi}_i$ are now the $N \times n_p$ matrix containing the n_p left singular vectors and right singular vectors respectively, associated to the n_p larger singular values listed in the $n_p \times n_p$ diagonal matrix $\mathbf{\Sigma}_i$. As $n_p \ll N$, we expect to be able to store the matrices $\mathbf{\Phi}_i$, $\mathbf{\Psi}_i$ and the diagonal of matrix $\mathbf{\Sigma}_i$, and extract information, e.g, RTM or CIPs by matrix-vector multiplication successively.

In the same spirit, we may write \mathbf{E} as the product

$$\mathbf{E}_i \approx \mathbf{L}_i \mathbf{R}_i^*, \quad (2.16)$$

with \mathbf{L}_i and \mathbf{R}_i being two $N \times n_p$ matrices, obtained from the SVD (2.15):

$$\begin{aligned} \mathbf{L}_i &= \mathbf{\Phi}_i \sqrt{\mathbf{\Sigma}_i}, \\ \mathbf{R}_i &= \mathbf{\Psi}_i \sqrt{\mathbf{\Sigma}_i}. \end{aligned} \quad (2.17)$$

Note that for this monochromatic \mathbf{E}_i , matrix \mathbf{S}_i is real, positive and diagonal, so the computation of its square root is implemented element-wisely on the diagonal. Then the construction of \mathbf{L}_i and \mathbf{R}_i is really cheap once we have the SVD decomposition of \mathbf{E}_i . However the basic SVD method has limitations in computation which involves (1) $2n_s$ PDE solves that are extraordinarily expensive for large-scale model, (2) the cost of SVD on

the full EIV which is of the order of $\mathcal{O}(N^3)$ according to Holmes, Gray, and Isbell 2007.

To circumvent the computational cost of the basic SVD, we propose to use the randomized SVD based approach (Halko, Martinsson, and Tropp 2011) to obtain the low-rank representation of the full EIV. Note that we have already wrapped up the monochromatic EIV as a linear operator \mathbf{E}_i based on Helmholtz solves, and EIVs along frequency as a linear operator \mathbf{E} based on time-stepping solves. To keep it simple and concise, we adopt the monochromatic notations in all the low-rank recovery algorithms below, following with explanations on the implementation in both frequency harmonic domain and time-domain. The randomized SVD algorithm with the subscript neglected, is listed as Algorithm 6:

Algorithm 6 Monochromatic randomized SVD algorithm from Halko, Martinsson, and Tropp 2011.

- 1: Generate n_p random Gaussian vectors $\mathbf{W}_i = [\mathbf{w}_1, \dots, \mathbf{w}_{n_p}]$
 - 2: $\mathbf{Y} = \mathbf{E}\mathbf{W}$, $\mathbf{Y} \in C^{N \times n_p}$
 - 3: $[\mathbf{Q}, \mathbf{T}] = \text{qr}(\mathbf{Y})$, $\mathbf{Q} \in C^{N \times n_p}$
 - 4: $\mathbf{Z} = \mathbf{E}^* \mathbf{Q}$, $\mathbf{Z} \in C^{N \times n_p}$
 - 5: $[\mathbf{\Phi}, \mathbf{\Sigma}, \mathbf{\Psi}] = \text{svd}(\mathbf{Z}^*)$, svd computes the top n_p singular vectors
 - 6: $\mathbf{\Phi} \leftarrow \mathbf{Q}\mathbf{\Phi}$
 - 7: $\mathbf{L} = \mathbf{\Phi}\sqrt{\mathbf{\Sigma}}$, $\mathbf{R} = \mathbf{\Psi}\sqrt{\mathbf{\Sigma}}$
 - 8: output: factors \mathbf{L}, \mathbf{R} from which actions can be formed via $\mathbf{E} \approx \mathbf{L}\mathbf{R}^*$
-

Here in line 0 the vector \mathbf{w}_{n_p} is one Gaussian vector. And the following steps in Algorithm 6 are implemented monochromatically. The corresponding step 0 in time-domain version generates the tensor W which is the band-limited noisy simultaneous shots located at every subsurface grid point, used in time-domain probing method. Analogously, the following steps in time-domain version get the corresponding monochromatic tensors: \mathbf{Y} , \mathbf{Q} , \mathbf{Z} , $\mathbf{\Phi}$, $\mathbf{\Psi}$, \mathbf{L} and \mathbf{R} . Also the ‘qr’ and ‘svd’ factorizations are overloaded and implemented over all frequency slices of the corresponding tensors \mathbf{Y} and \mathbf{Z}^* . Finally this algorithm combined with time-domain wave-equation solver can get the recovered EIVs along all

frequencies.

CHAPTER 3

LOW-RANK RECOVERY FOR EXTENDED IMAGE VOLUMES VIA INVARIANCE RELATIONSHIP

3.1 summary

Continued with the work in chapter 2, the proposed factorization provides a mechanism to use the invariance relation of extended image volumes in factored form. This invariance relation states that extended image volumes obtained for one background velocity model can directly be mapped to those of another background velocity model without the need to re-factorize. Our low-rank factorization inherits this invariance property so we only incur the relatively high factorization costs once for common imaging workflows during which different velocity model scenarios are examined. All subsequent imaging experiments only involve the factors and are therefore computationally cheap compared to conventional imaging where the cost scale with the number of source experiments.

3.2 Introduction

In chapter 2 we introduce the extended image volumes (EIVs) that contain rich subsurface information, including RTM, CIPs and GIGs used for not only creating images, but also for the interpretation of rock properties and velocity analysis in complex geological settings (Shen and Symes 2008; Sava and Vasconcelos 2011c). And we also commend that the full EIVs are too expensive in computation and storage, which impedes the widely usage in industrial large-scale problem. Leeuwen, Kumar, and Herrmann 2017 proposed the probing technology that extracts the desired CIPs (column of the EIVs) or angle gathers with limited computation that only proportional to the number of columns (usually proportional to the grid points $2n_z$ in depth dimension). However if the RTM image (diagonal elements of

EIVs) is desired, this probing technology still needs to compute as high as N , the number of the subsurface grid points.

Thanks to the redundancy of the full EIVs that we explained in chapter 2, we are able to present the EIVs with low-rank representations (Kumar, Graff-Kray, Leeuwen, and Herrmann 2018a). As we have shown in chapter 2, the rank of the EIV at low frequency is much lower than that of the data. We can imagine that the upper limitation of EIV's rank is the rank of data because the definition of monochromatic EIV involves full-rank operators of wave-equation solver and the low-rank data matrix. Finally we show in chapter 2 that we can recover EIVs based on rSVD methods, where the range of the EIV is first obtained by probing the implicit function of EIV's forward with n_p simultaneous random shots, then extracted out by QR factorization and probed back to EIV's implicit adjoint function. The computation for any image gather could be limited to $4n_p$ wave-equation solves, where n_p slightly larger than the rank of EIV leads to accurate recovery (Kumar, Graff-Kray, Leeuwen, and Herrmann 2018a).

However the implementation in the early stage, which based on frequency Helmholtz solvers, scales badly to the size of the model due to the Helmholtz solver involved here is harder to solve for larger model and higher frequencies. In order to extend this idea to the industrial scale 2- or 3-D problem in the future, we implement the low-rank recovery of EIV based on time-stepping propagator Devito (Kukreja, Louboutin, Vieira, Luporini, Lange, and Gorman 2016) instead of Helmholtz solver. However there still are open issues to solve. For example, the rank of the EIV will increase along frequency, which makes it more expensive to recover sharper EIV because higher rank for higher frequencies is demanded, rising again the PDE-related (or wave-equation-related) cost and storage.

To limit the probing size, in chapter 2, we propose to use power iterations (Halko, Martinsson, and Tropp 2011)—i.e. simultaneous iterations and block Krylov iterations, which could promote the decay of the singular values of the aimed frequency components, together with time-stepping probing framework. We compare the performance of these two

power schemes for a small model from Marmousi with limited power $q = 1, 2$ by analyzing the errors in the recovered singular values and the RTMs, and give the computation and storage complexity analysis. The experiment demonstrates that we could improve the recovery accuracy without increasing the probing size when the rank goes higher along frequency. So the following slicing and dicing do not cost any extra PDE solves. However, according to the computation and storage complexity analysis, the power scheme based rSVD still cost more wave-equation-related computations. And the BKI method even takes more temporary memory to form the Krylov space that gathers different orders together. Yet the superiority of the power scheme based rSVD over the basic rSVD is not very clear in computation and storage. Since EIVs are solutions to the double two-way wave equation, which itself adheres to an invariance relation, they exhibit this invariance as well and we will show this property is inherited by our factorization. We will demonstrate that this invariance leads to imaging workflows where we incur the relatively expensive computational costs of the factorization only once. All subsequent costs scale with the rank of the factorization and this includes imaging in different background velocity models. This property is unique and can be seen as an image-domain extension of early work by Hu, Abubakar, and Habashy 2009.

In this chapter, based on the work of chapter 2, we discuss time-harmonic and time-domain versions of the invariance relationship for EIVs. This relationship allows us to map EIVs for one background velocity model to another without the need to re-factorize. We demonstrate that a multi-scenario imaging workflow that works on the factors alone is feasible by virtue of velocity continuation that derives from the invariance relationship of EIVs by carrying out a realistic imaging scenario involving salt. During that experiment, we demonstrate that a multi background velocity model imaging scenario that work on the factors alone are computationally feasible by virtue of our velocity-continuation approach, which derives from the invariance relationship of EIVs (Leeuwen and Herrmann 2012; Kumar, Graff-Kray, Leeuwen, and Herrmann 2018b).

3.3 Invariance relationship of EIV

So far, we concentrated on finding representations for EIVs using factorizations informed by randomized probings. Our factorizations incurred an upfront cost dominated by the number of randomized probings that determined the number of wave-equation solves. All subsequent manipulations, such as forming CIPs and (angle corrected) CIGs, did not involve wave-equation solves and are therefore relatively cheap. Moreover, the number of probing vectors is often smaller than the number of source experiments, i.e., $n_p \ll n_s$.

To allow for more realistic imaging workflows, where different imaging scenarios involving different velocity models are conducted, we propose to leverage an important invariance property of the double wave equation, which model EIVs. This property allows us to do velocity continuation (Kumar, Graff-Kray, Leeuwen, and Herrmann 2018b)—i.e., to directly map an image volume obtained with one velocity model to an image volume yielded by another velocity model without the need to remigrate involving a loop over n_s shots. We will adapt this invariance property of the double wave equation to the low-rank factorizations for the EIVs as introduced earlier. Since $n_p \ll n_s$, this formulation will allow us to test different imaging scenarios with varying background velocity models—e.g., different picks of top salt, at a greatly reduced cost. To firmly establish this opportunity where we derive invariance relations for the factors themselves, we first introduce the invariance relationship in factored form in the Fourier domain, followed by its time-domain equivalent.

3.3.1 Monochromatic invariance relationship

Because the right-hand-side of the two-way wave-equation does not depend on the background velocity model, Leeuwen, Kumar, and Herrmann 2017 derived an invariance relationship directly linking image volumes \mathbf{E}_1 and \mathbf{E}_2 pertaining to background velocity model \mathbf{m}_i , $i = 1, 2$ (for simplicity, we abandoned the frequency subscript and the subscript

now specifies extended image volumes pertaining to velocity models for scenario 1 or 2). According to the double wave equation, these image volumes are related via

$$\mathbf{H}_1^* \cdot \mathbf{E}_1 \cdot \mathbf{H}_1^* = \mathbf{H}_2^* \cdot \mathbf{E}_2 \cdot \mathbf{H}_2^*, \quad (3.1)$$

where the Helmholtz operators $\mathbf{H}_i = \mathbf{H}(\mathbf{m}_i)$, $i = 1, 2$ depend on the background velocity models for the two imaging scenarios. This relationship allows us to directly calculate EIV, \mathbf{E}_2 , for imaging scenario 2 from the EIV yielded by imaging scenario 1 via

$$\mathbf{E}_2 = \mathbf{H}_2^{-*} \cdot \mathbf{H}_1^* \cdot \mathbf{E}_1 \cdot \mathbf{H}_1^* \cdot \mathbf{H}_2^{-*}. \quad (3.2)$$

Since EIVs can be represented in factored form, the factors for both imaging scenarios are related via

$$\begin{aligned} \mathbf{L}_2 &= \mathbf{H}_2^{-*} \cdot \mathbf{H}_1^* \cdot \mathbf{L}_1, \\ \mathbf{R}_2 &= \mathbf{H}_2^{-1} \cdot \mathbf{H}_1 \cdot \mathbf{R}_1. \end{aligned} \quad (3.3)$$

With this relationship, we only need to factor an EIV once, say for the velocity model of scenario 1. All subsequent factors for different imaging scenarios with different velocity models can be derived with Equation 3.3, avoiding the computational expensive step of randomized probing, followed by the relative expensive BKI iterations. Consequently, we arrived at a formulation where the expensive computational costs of factorizing EIVs are incurred only once up front. After the initial factorization, we only need to spend $2n_p$ wave-equation solves per factor, which easily negates the computational overhead associated with the initial factorization. Obviously this is a powerful result in situations where there is uncertainty in the background velocity model. In the next section, we discuss how to implement these invariance relations using time-domain wave-equation solvers.

3.3.2 Velocity continuation in the time domain

The combination of low-rank factorizations with the above invariance relationship gives us the possibility to directly form factored EIVs for different velocity models without the need redo the factorization including the probing. According to Equation 3.3, we only need access to the factors in the current velocity model, the action of the forward and adjoint wave equation itself in the current and the new velocity model, and the solution of these wave equations for the two velocity models.

While the monochromatic invariance relations allows us to map factors from one velocity model to another, the formulation hinges on having access to the action of the discrete wave-equation operators and their inverse. The need for the latter can become problematic since Helmholtz solvers do not scale very well to high frequencies and 3D models. To address this issue, we employ wave-equation solvers, including the action of the wave-equation operators, based on time-stepping and finite differences implemented with Devito (Louboutin, Lange, Luporini, Kukreja, Witte, Herrmann, Velesko, and Gorman 2018). Based on this time-domain implementation, the new factors, which now become tensors, can be written as

$$\begin{aligned} \mathbf{L}_2 &= \mathcal{F} \circ \mathcal{A}_2^{-\top} \circ \mathcal{A}_1^\top [\mathbf{L}_1] & \text{with } \mathbf{L}_1 &= \mathcal{F}^\top [\mathbf{L}_1], \\ \mathbf{R}_2 &= \mathcal{F} \circ \mathcal{A}_2^{-1} \circ \mathcal{A}_1 [\mathbf{R}_1] & \text{with } \mathbf{R}_1 &= \mathcal{F}^\top [\mathbf{R}_1]. \end{aligned} \tag{3.4}$$

Here \mathcal{A}_2^{-1} represents forward modeling in the velocity model \mathbf{m}_2 for scenario 2, and $\mathcal{A}_2^{-\top}$ is the corresponding adjoint operator. The linear operator \mathcal{A}_1 is the inverse forward modeling operator for imaging scenario 1 with the velocity model \mathbf{m}_1 , and \mathcal{A}_1^\top is the corresponding adjoint operator. As in the monochromatic case, the direct mapping of the factored form of an EIV from one to another velocity model only involves $2n_p$ actions of the forward/adjoint wave-equation and their inverses. Note that compared to the frequency domain formulation, the action of time-domain operators and their inverses is roughly the same while the cost of applying the Helmholtz operator is cheap compared to applying its

inverse.

3.3.3 Power scheme based rSVD plus invariance relationship together

Aside from reducing the cost (from $\mathcal{O}(n_s)$ to $\mathcal{O}(n_p)$ actions with wave operators and their inverse), the main advantage of working with factored EIVs is that we only incur the costs of the initial factorization once, which involves randomized probing and a SVD based on BKI iterations. After this initial cost, factorizations of EIVs for different imaging scenarios with different velocity models can be formed and this process can be repeated at will because the mapping in Equation 3.4 preserves accuracy of the original factorization. We base this claim on the fact that the action of the wave-equation operators on their inverse is the identity by definition. This means that if we apply the wave-operator to a factor, we will undo possible wave simulation errors, such as numerical dispersion. In view of these properties, we argue that it is beneficial to develop a strategy where we work with as few as possible factors (n_p) calculated with an as high as possible accuracy.

3.4 Numerical experiments

To test the approach we proposed above, we consider a large-scale complex imaging problem with Salt. To mimic a realistic imaging scenario, we examine an imaging scenario where the background velocity model for the top salt is wrong. We demonstrate that our velocity continuation technique is capable of mapping the low-rank representation for the wrong velocity model to corrected factors that lead to an image that is well focused without the need to recompute the factorization.

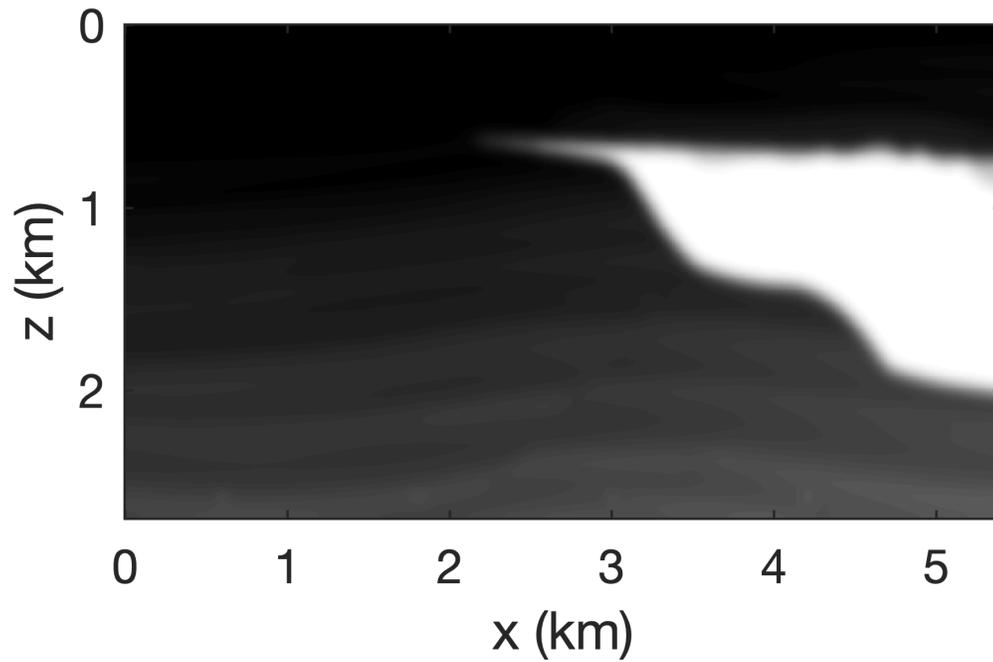
While working with EIVs in factored form gave us access to accurate RTM images and subsurface offset gathers at limited costs, imaging in complex areas remain challenging because of inaccuracies in the background velocity model. For instance, errors in top salt can lead to a rapid deterioration of the imaging quality beneath salt. In practice, this means that imaging teams go through many cycles of updating the velocity model, followed by imag-

ing. By using the invariance relation for EIVs, we propose to use a velocity-continuation technique based on our factorization and captured by Equation 3.3. According to those equations, the left and right factors can be mapped directly from one velocity model to another without the need to completely recompute the factorization. Instead, we incur costs equalling $4n_p$, which is relatively cheap when n_p is small.

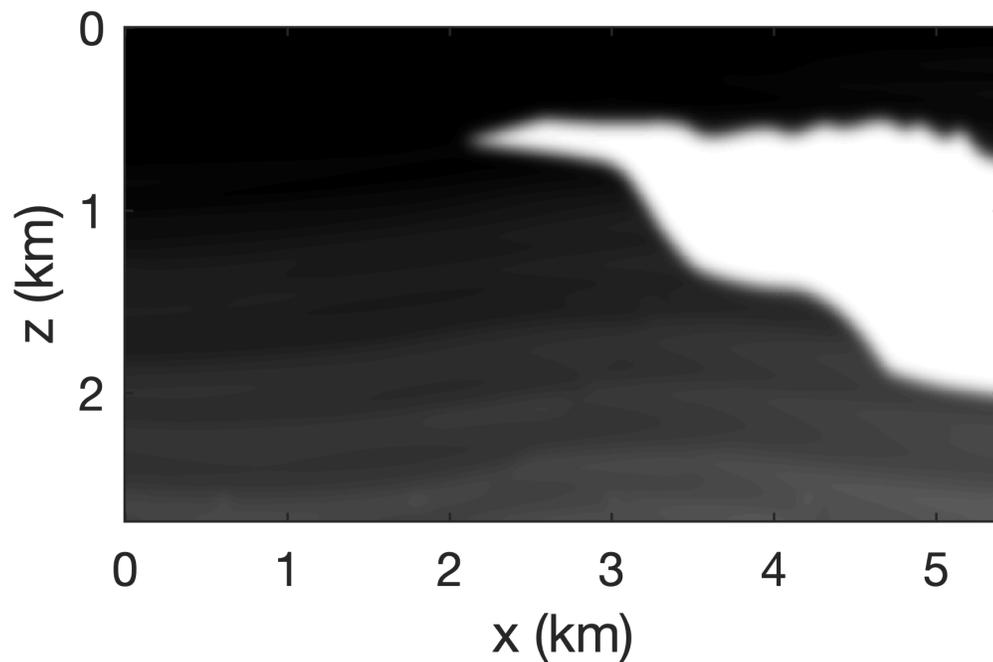
To mimic a realistic subsalt imaging scenario, we compare three scenarios that derive from a subset of the Sigsbee model. In the first scenario, we compute an image for a background velocity model where top salt is wrong. We compare this result to the second scenario where the background velocity model is corrected but where we remigrate the data by recomputing the factorization. In the third scenario, we compute the image by mapping the EIV in factors form with Equation 3.3. To avoid salt-related imaging artifacts, we use the inverse-scattering imaging condition [add references] on linearized data simulated with the correct background velocity model depicted in Figure 3.1b and the true perturbation given by the difference between the 2.7×5.4 km true velocity model, sampled on a 6×6 m grid, and the correct background velocity model. We simulate the data for 450 co-located sources and receivers spread over the top of the model and at a depth of 18m sampled with a 12m interval. The source signature is a Ricker wavelet centered at 23Hz.

As before, we choose n_p according to the rank needed to capture 95% of the energy in the data. We plot this rank in Figure 3.2. Based on our empirical finding that the singular values of EIVs decay more rapidly, we choose $n_p = 100$ roughly half of the maximum rank needed to accurately represent the data at 70Hz. Figure 3.3a contains the image obtained with a background velocity model that contains errors in the definition of top salt (cf. Figures 3.1a and 3.1b). Figure 3.1a misses key details on the salt sediment boundary, which has a detrimental effect of the image beneath the salt (cf. Figures 3.3a and 3.3b). Not only the bottom salt is out of focus but so are the sediments and fault beneath the salt. The image for the correct background velocity model is obtained via a completely new probing, factorization, and application of diagonal extraction. This shows that our

factorization is capable of handling imaging complex salt areas, and observation confirmed by the CIPs (cf. Figures 3.4a and 3.4b) and CIGs (cf. Figures 3.5a and 3.5b). While all slightly noisy, the images and subsurface offset gathers behave as expected with energy focused onto the reflectors. What is more important, is that exactly the same image quality is attained for the RTM and subsurface-offset gathers when we directly map the original factorization, obtained for the wrong velocities of scenario one, to the factorization for the correct velocity model using Equation 3.3 instead of recomputing the factorization after probing. As a result, we are with scenario three able to obtain the RTM image (Figure 3.3c), CIP (Figure 3.4) and CIG (Figure 3.5) at only $4n_p$ wave-equation solves. For comparison, conventional RTM without having access to EIVs would have cost $2n_s$ wave-equation solves while scenario two would, according to Table 2.2 for $q = 1$, have cost $2n_p + 4n_p + 4n_p = 10n_p$, while the direct map only costs $4n_p$. Remember we choose $n_p = 100 \ll n_s = 450$, which means that we incur slightly more cost when conducting a single migration (1000 wave-equation solves versus 900 for conventional RTM). However, after this factorization each additional RTM only costs 400 wave-equation solves. Also, remember that factorizing EIVs gives us easy access to subsurface offset gathers.



(a)



(b)

Figure 3.1: The background models used as different velocity scenarios for velocity continuation test. (a) The initial guess of the background model where the top of the salt is not correct, (b) the kinematically correct background model.

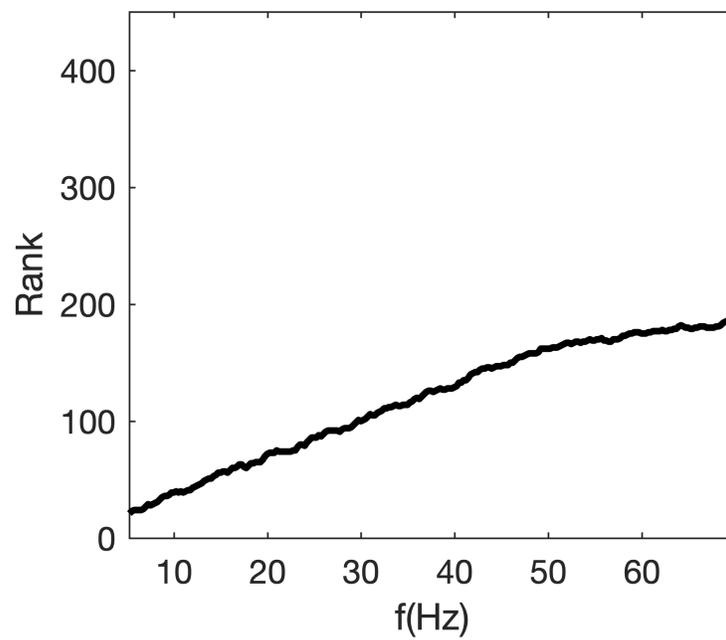
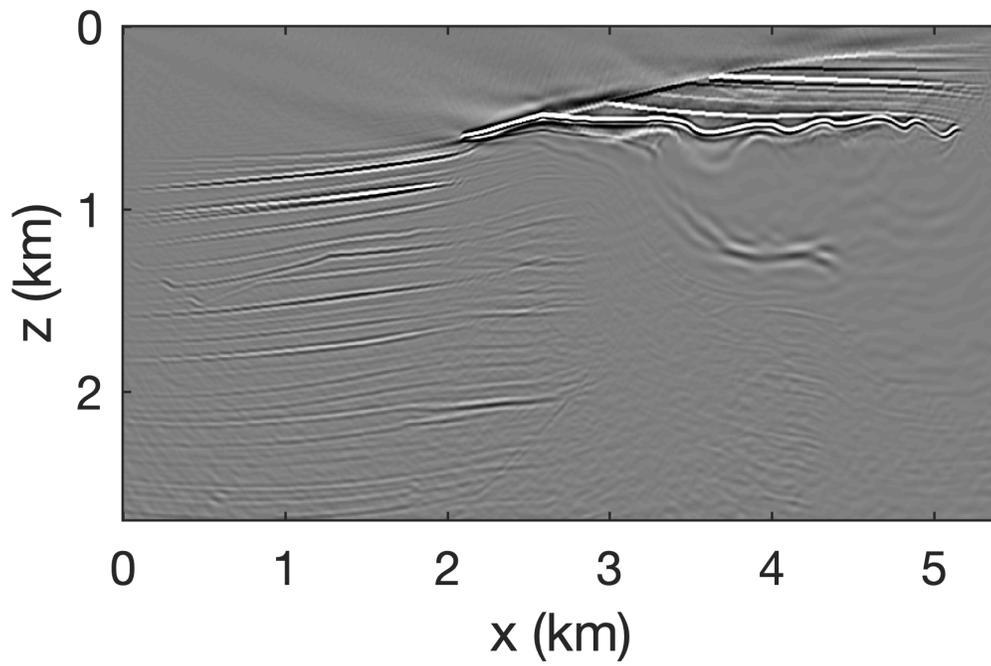
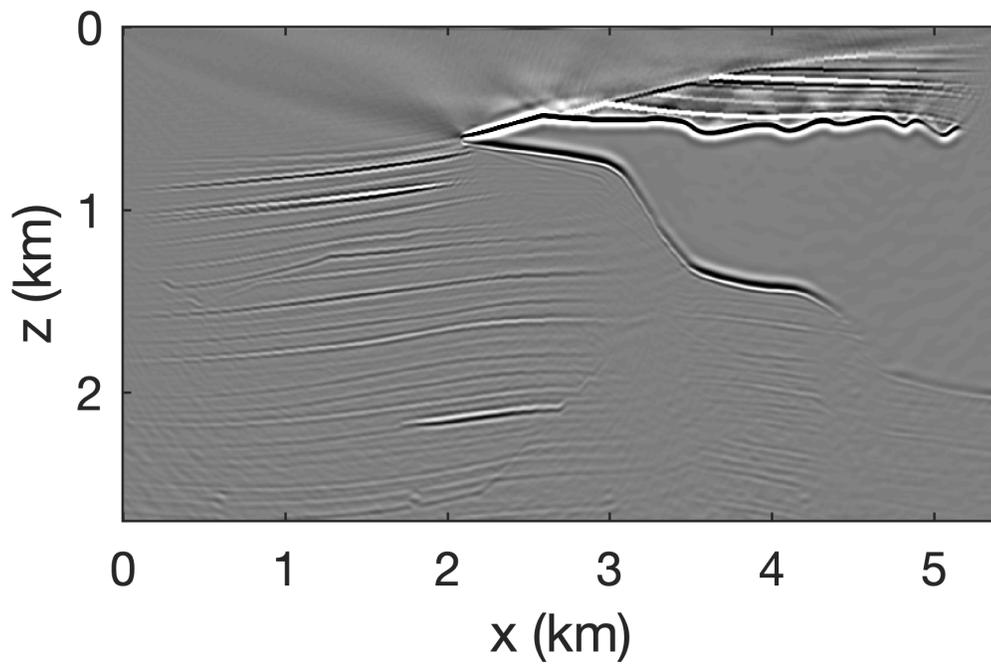


Figure 3.2: Selected rank to capture 95% of the data's energy as a function of frequency.



(a)



(b)

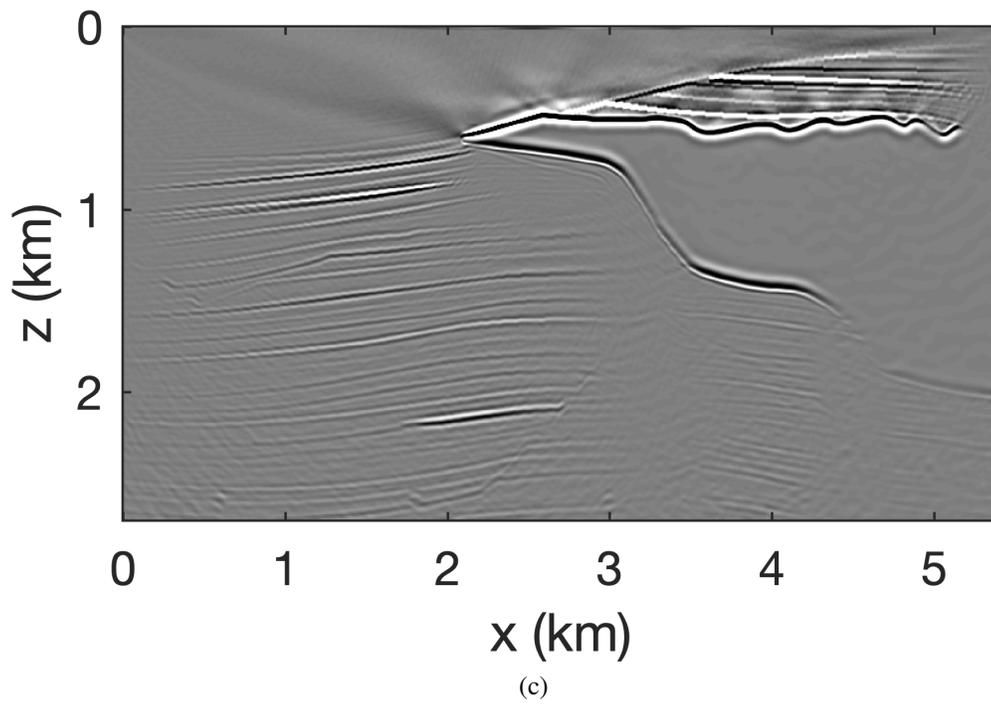
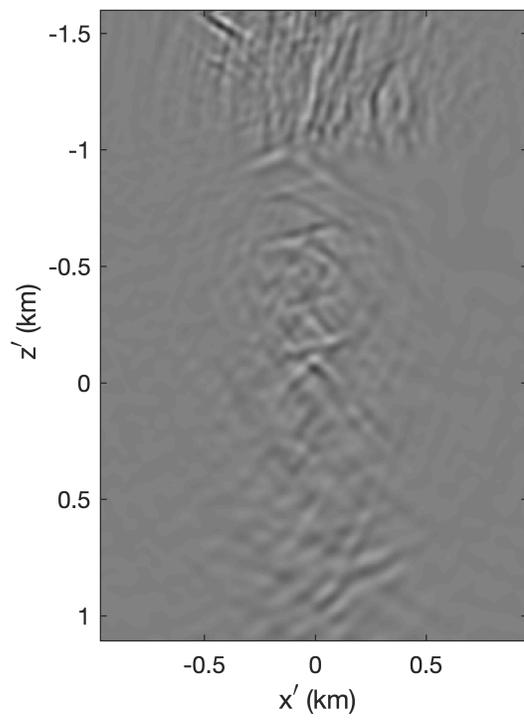
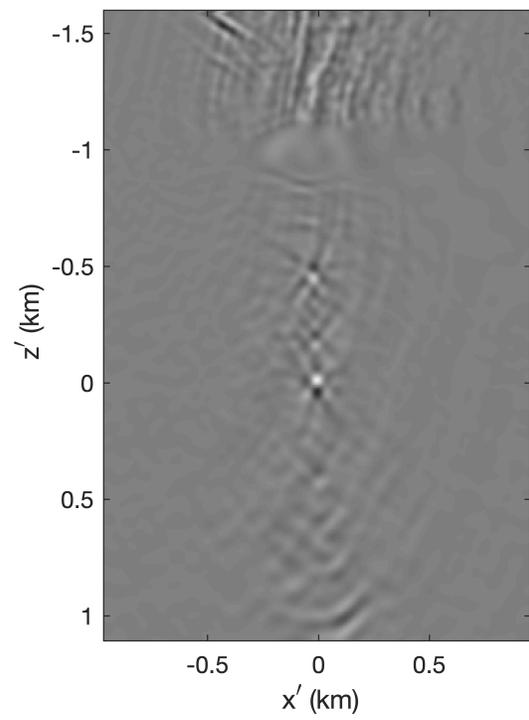


Figure 3.3: The RTMs extracted from the recovered low-rank representations of the EIV from (a) the initial guess model, (b) the kinematical correct model, (c) the RTM extracted from the recovered EIV that mapped from the initial guess model to the correct model.



(a)



(b)

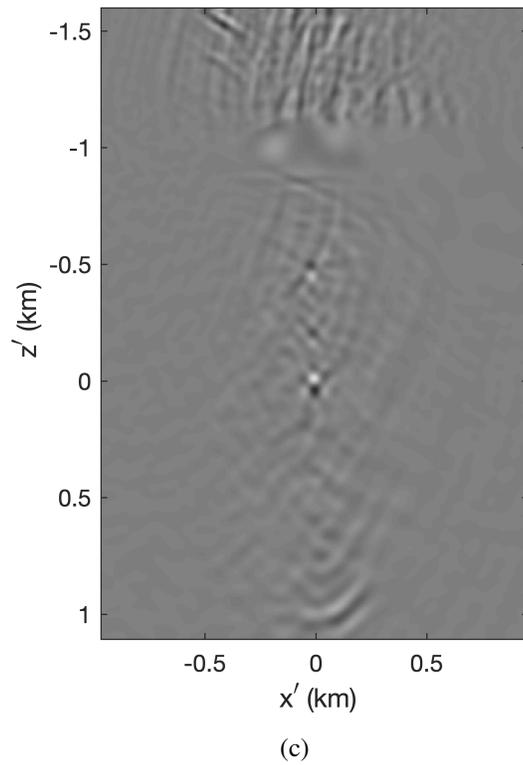
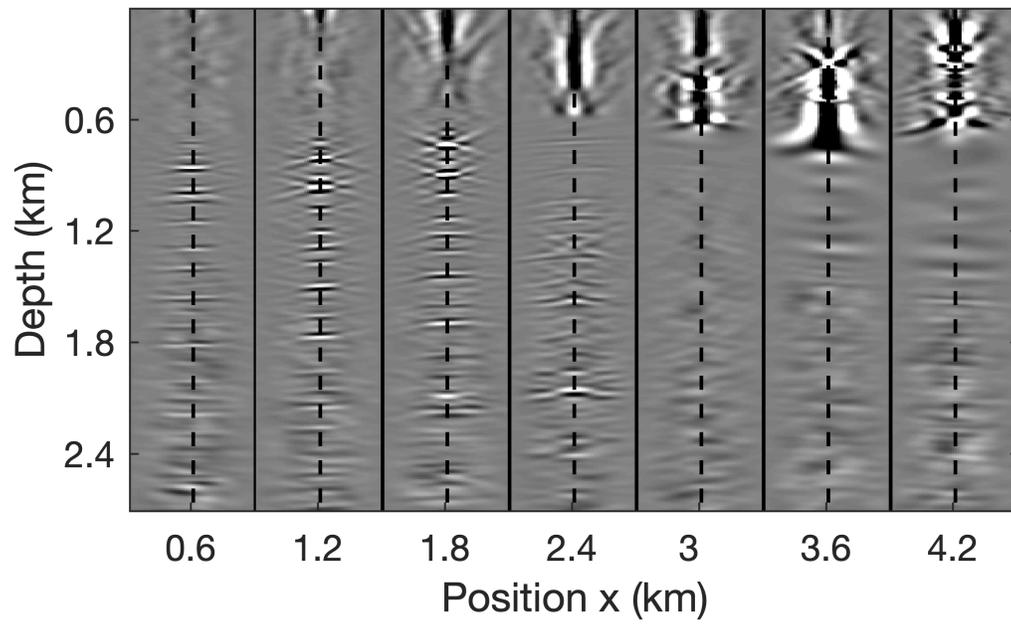
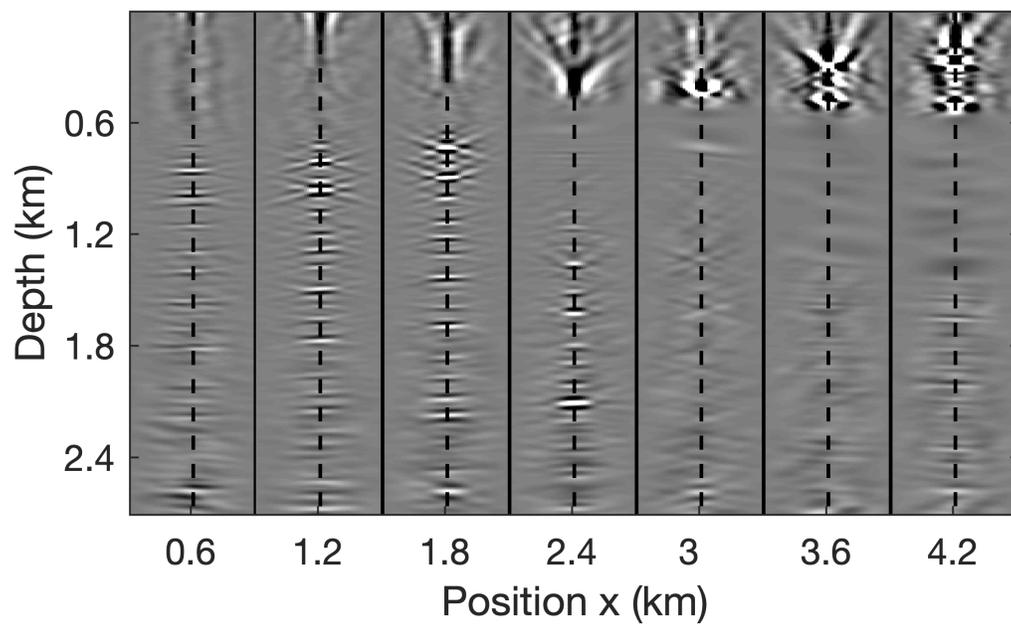


Figure 3.4: The CIP images extracted from the recovered EIVs at $x = 2640\text{m}$, $z = 1590\text{m}$. (a) the CIP image of the EIVs from the initial guessed model, (b) the CIP image of the EIVs from the correct model, (c) the CIP image of the EIVs via mapping from the initial guess to the correct model.



(a)



(b)

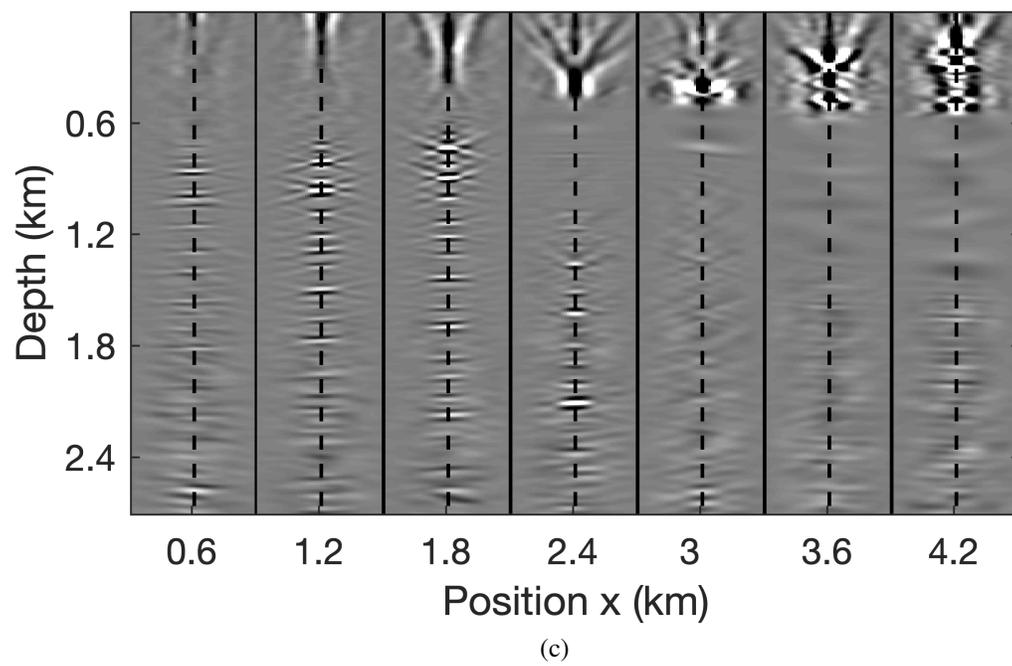


Figure 3.5: The CIGs extracted from the low-rank representations with (a) the initial guess of the background model (as in Figure 3.1a), (b) the kinematically correct background model (as in Figure 3.1b), (c) via invariance relationship mapping from the initial background model to the correct background model.

3.5 Discussion

In this chapter, we introduce the invariance relationship in both frequency domain and time domain. And we use the invariance relationship to map between different background models. In the numerical experiments, we test with part of the Sigsbee2A model. The initial guessed smooth model has wrong salt dome, so the corresponding RTM has wrong located reflectors, and the CIPs are not focused to the image points, so neither do the CIGs. After mapping the low-rank factors we obtained with BKI method from this initial guess to the kinematically correct background model, we get the updated low-rank factors, where the reflectors of RTM are located correctly, and the CIPs are focused correctly to the image point. When we use power method to obtain the low-rank factors for the initial model, we spend more than $4n_p + 4qn_p$ and $2n_p + 4qn_p + 2(q + 1)n_p$ wave-equation solves for SI and BKI respectively, and the corresponding ‘qr’ and ‘SVD’ factorizations. But the mapping to other model only costs $4n_p$ wave-equation solves and no factorizations anymore. The superiority of the power-scheme-based rSVD combined with the invariance relationship indicates that when the computational resource is limited, we could consider to achieve more accurate low-rank representations with the probing size as low as possible via powered rSVDs.

Based on the invariance relationship, it is also possible to update the background model according to the diagonals of the gradient of the EIV by minimizing the objective functions that focus the EIV commute with diagonal weighting matrices that penalize off-diagonal energy (Leeuwen, Kumar, and Herrmann 2017). And as we mentioned in the discussion in chapter 2, when there is salt, the EIV would be ill-conditioned, which means the singular values are dominated by the salt. We need to design some preconditioners in the future to improve the images extracted from the low-rank factors of EIV.

3.6 Conclusion

Because we are able to directly map the factors from one background velocity model to another, without the need to factorize again, we are justified to incur the relatively high initial factorization costs. This direct mapping of the factors is known as velocity continuation. We argue that our approach is one of the first concrete examples where this technique results in a viable workflow for imaging involving salt. We demonstrate that we are capable of carrying out a completely new imaging experiment without the need to refactor. We accomplish this by using the invariance relation of extended image volumes whose applicability extends to its factored form. Since our factorization is low rank, the costs of repeated imaging experiments is small since the rank is typically much smaller than the number of shots in an imaging experiment.

CHAPTER 4

TIME-DOMAIN SPARSITY PROMOTING LEAST-SQUARES REVERSE TIME MIGRATION WITH SOURCE ESTIMATION

4.1 Summary

In chapter 2, we mainly discuss the low-rank recovery on the extended image volume (EIV), where the imaging condition is multi-dimensional convolution in time domain and matrix-matrix multiplication in monochromatic frequency domain. The future work on EIV could be focused on developing the preconditioners which could alleviate the ill-conditioning of the EIV and compensate the wavelet imprint from the source and receivers, so further improve the resolutions. From this chapter on, we focus on imaging problems limited to only the diagonals of the EIV, namely the reverse time migration image where the imaging condition is the cross-correlation between the forward and backward wavefields, instead of multi-dimensional convolutions in time domain, and element-wise multiplication instead of matrix-matrix multiplication in monochromatic frequency domain. Reverse time migration also suffers from the wavelet related imprints, and the amplitudes are distorted.

Least-squares reverse time migration is well-known for its capability of generating true-amplitude subsurface images through fitting observed data in the least-squares sense. However, when applied to realistic problems, this approach is faced with issues related to overfitting and excessive computational costs induced by many wave-equation solves. The fact that the source function is unknown complicates this situation further. Motivated by compressive sensing, recently developed sparsity-promoting approaches are capable of substantially reducing computational costs while avoiding imaging artifacts and restoring amplitudes. Nevertheless, these approaches still raise issues of lack of convergence, algorithmic complexity of the solver, and the need to do source estimation possibly on-the-fly. We

address these problems by presenting an algorithm that allows us to work with randomly drawn subsets of sources during each iteration. Moreover, we adapt this algorithm with an on-the-fly source estimation through variable projection, which involves only inexpensive penalized least-squares sub-problems instead of expensive PDE solvers. Applications of our algorithms to the Marmousi model and the Sigsbee model illustrate that the proposed method generates high-resolution images along with accurate estimates for the source signature using only one to two data passes with a computational cost that roughly equals that of two conventional reverse-time migrations. Our numerical results also demonstrate robustness of the proposed method against noise.

4.2 Introduction

Reverse-time migration (RTM) is a popular wave-equation-based seismic imaging method where the inverse of the Born scattering operator is approximated by applying its adjoint directly to the observed reflection data (Baysal, Kosloff, and Sherwood 1983; Whitmore 1983). Because the adjoint does not equal the pseudo inverse conventional RTM produces images with incorrect amplitudes. Amongst the factors that contribute to low fidelity amplitudes, the imprint of the temporal bandwidth limitation of the typically unknown source wavelet features prominently and so does the fact that the Born scattering operator is not inverted. To overcome these issues, we formulate our imaging problem as a linear least-squares inversion problem where the difference between observed and predicted data is minimized in an ℓ_2 -norm (Schuster 1993; Nemeth, Wu, and Schuster 1999; Dong, Cai, Guo, Suh, Zhang, Wang, and Li 2012; Zeng, Dong, and Wang 2014). While least-squares migration is a powerful technique, its successful application to industry-scale problems is hampered by three key issues. First, iterative demigrations (= Born modeling) and migration become computationally prohibitively expensive when carried out over all shots. Second, we run the risk of overfitting the data when minimizing the ℓ_2 -norm of the data residual. This overfitting may introduce noise-related artifacts in inverted images. Third,

while the source location is generally well known, the temporal source function is often not known accurately. Because imaging relies on knowing the source function, this may have a detrimental effect on the image and makes it necessary to come up with source estimation methodology. Since we carry out our imaging iteratively, we propose to estimate the wavelet on-the-fly as we build up the image.

We address the issue of computational feasibility by combining techniques from stochastic optimization (Leeuwen, Aravkin, and Herrmann 2011; Haber, Chung, and Herrmann 2012; Powell 2014), curvelet-domain sparsity-promotion herrmann2012efficient, and online convex optimization (Lorenz, Schopfer, and Wenger 2014) with linearized Bregman. Stochastic optimization allows us to work with small random subsets of shots, which limits the number of passes through the data. Convergence is guaranteed (Herrmann, Tu, and Esser 2015b; Yang, Witte, Fang, and Herrmann 2016; Witte, Louboutin, Luporini, Gorman, and Herrmann 2019) by replacing the ℓ_1 -norm, by an elastic net consisting of a strongly convex combination of ℓ_1 - and ℓ_2 -norm objectives. Inclusion of the ℓ_2 -norm result in a greatly simplified algorithm involving linearized Bregman iterations, which corresponds to gradient descent on the dual variable supplemented by a simple soft thresholding operation (Yin 2010; Cai, Osher, and Shen 2009) with a threshold that is fixed. We refer to this method as sparsity-promoting least-squares reverse-time migration (SPLS-RTM).

In addition to the high computational cost, the lack of accurate knowledge on the unknown temporal source signature may also adversely affect the performance of the inversion. Errors in the source signature lead to erroneous residuals, which in turn result in inaccurately imaged reflectors, which now may be positioned wrongly or may have the wrong amplitude or phase. To mitigate these errors, we need an embedded procedure where the source signature is updated along with the image during the inversion (Pratt 1999; Aravkin, Leeuwen, and Tu 2013; Fang, Wang, and Herrmann 2018; Aravkin, Leeuwen, Calandra, and Herrmann 2012) using a technique known as variable projection (Leeuwen, Aravkin, Herrmann, Li, Rickett, and Abubakar 2014; Rickett 2013). For time-harmonic imaging,

variable projection involves estimation of the source function by solving a least-squares problems for each frequency separately. Since the unknown for each frequency is single complex-valued variable this process is simple and has resulted in accurate estimation and compensation for the source-time function (see e.g. Tu, Aravkin, Leeuwen, and Herrmann 2013 and Fang, Wang, and Herrmann 2018). Unfortunately, the situation is more complicated during imaging in the time-domain where we have to estimate the complete source signature during each iteration. For this purpose, we build on early work by Yang, Witte, Fang, and Herrmann 2016 by making it suitable realistic imaging scenarios that may include salt and inverse-scattering introduced by Witte, Louboutin, Luporini, Gorman, and Herrmann 2019 .

Our work is outlined as follows. First, we introduce the basic equations for time-domain reverse time migration and least-squares reverse time migration. To overcome the computational cost associated with the latter, we introduce a stochastic optimization method with sparsity promotion. This method is designed to provide an image at a fraction of the cost. Next, we extend this approach so it includes on-the-fly source estimation. This allows us to remove the requirement of the source function. We conclude by presenting a number of synthetic case studies designed to demonstrate robustness with respect to noisy data and to complex imaging scenarios that include salt.

4.3 From RTM to LS-RTM

Since our approach hinges on cost-effective least-squares imaging, we first introduce our formulation of sparsity-promoting least-squares migration with stochastic optimization followed by our approach to on-the-fly source estimation during the iterations.

Reverse time migration derives from a linearization (see e.g. Mulder and Plessix 2004 with respect to the squared background slowness. For the i^{th} source this linearization reads

$$\delta \mathbf{d}_i = \mathbf{F}_i(\mathbf{m}_0 + \delta \mathbf{m}, \mathbf{q}) - \mathbf{F}_i(\mathbf{m}_0, \mathbf{q}) \approx \nabla \mathbf{F}_i(\mathbf{m}_0, \mathbf{q}) \delta \mathbf{m}, \quad (4.1)$$

where the vectors $\delta\mathbf{m}$, \mathbf{q} , and $\delta\mathbf{d}$ denote the model perturbation, the source-time function, and the corresponding perturbation in the data, respectively. We model the data for n_t time samples of a time period of T s. The number of receivers is n_r so a single shot record is of size $n_t \times n_r$. The nonlinear forward modeling operator $\mathbf{F}_i(\mathbf{m}, \mathbf{q})$ for the i th source location involves the solution of the discretized acoustic wave equation

$$\begin{aligned} \left(\Delta - \mathbf{m} \odot \frac{\partial^2}{\partial t^2} \right) \mathbf{u}_i &= \mathbf{P}_{s,i}^\top \mathbf{q}, \\ \mathbf{P}_{r,i} \mathbf{u}_i &= \mathbf{d}_i, \end{aligned} \tag{4.2}$$

parameterized by the squared slowness collected in the vector \mathbf{m} (for simplicity, we kept the density constant and we used the symbol \odot to denote element wise multiplication.) The symbol Δ represents the discretized Laplacian and the linear operators $\mathbf{P}_{r,i}$ restrict the wavefield for the i th source to the corresponding receiver locations, while the linear operator $\mathbf{P}_{s,i}^\top$ injects the source time function at the location of the i th source in the computational grid. The Jacobian $\nabla \mathbf{F}_i(\mathbf{m}_0, \mathbf{q})$ is known as the Born modeling operator and is given by the derivative of $\mathbf{F}_i(\mathbf{m}, \mathbf{q})$ at the point of \mathbf{m}_0 . Applying the Jacobian $\nabla \mathbf{F}_i(\mathbf{m}_0, \mathbf{q})$ to the model perturbation $\delta\mathbf{m}$ requires the solution of the following linearized equation::

$$\begin{aligned} \left(\mathbf{m}_0 \odot \frac{\partial^2}{\partial t^2} - \Delta \right) \delta\mathbf{u}_i &= -\frac{\partial^2}{\partial t^2} (\delta\mathbf{m} \odot \mathbf{u}_i), \\ \mathbf{P}_{r,i} \delta\mathbf{u}_i &= \delta\mathbf{d}_i, \end{aligned} \tag{4.3}$$

where the vector $\delta\mathbf{u}_i$ corresponds to the wavefield perturbation for the i^{th} source.

The goal of seismic imaging is to estimate model perturbations from observed data. We can expect this reconstruction process to be successful in situations where the above linear approximation is accurate—i.e., the background velocity model needs to be sufficiently accurate, which we assume it is. We also need accurate knowledge on the source function, an important aspect we will address below.

While the above linearization allows us to create an image via

$$\delta \mathbf{m}_{\text{RTM}} = \sum_{i=1}^{n_s} \nabla \mathbf{F}_i^\top \delta \mathbf{d}_i, \quad (4.4)$$

with n_s the number of shots, the adjoint (denoted by the symbol $^\top$) of the Jacobian does not correspond to its inverse and $\delta \mathbf{m}_{\text{RTM}}$ will suffer from wavelet side lobes and inaccurate and unbalanced amplitude (mulder2004comparison, bednar2006two, hou2016accelerating). Unlike RTM (Equation 4.4), LS-RTM (Aoki and Schuster 2009; Herrmann and Li 2012; Tu and Herrmann 2015a) reconstructs the model perturbation by computing the pseudo-inverse of the Born modeling operator, which can significantly mitigate these defects. LS-RTM iteratively solves the following least squares data-fitting problem:

$$\min_{\delta \mathbf{m}} \frac{1}{2} \sum_{i=1}^{n_s} \|\nabla \mathbf{F}_i(\mathbf{m}_0, \mathbf{q}) \delta \mathbf{m} - \delta \mathbf{d}_i\|^2. \quad (4.5)$$

Compared to Equation 4.4, the above minimization requires multiple evaluations of the Jacobian and its adjoint, which becomes rapidly computationally prohibitive for large 2D, 3D imaging problems with the number of sources n_s large. This in part explains the relatively slow adaptation of least-squares reverse time migration (cf. Equation 4.5) by industry. As we show below, we overcome this problem by combining ideas from stochastic optimization and sparsity promotion (Herrmann, Tu, and Esser 2015b; Yang, Witte, Fang, and Herrmann 2016; Witte, Louboutin, Luporini, Gorman, and Herrmann 2019), which allow us to obtain artifact-free images at the cost of two to three passes through the data.

4.4 Stochastic optimization with sparsity promotion

As we mentioned above, minimization of Equation 4.5 over all n_s shots is computationally prohibitively expensive. In addition, the minimization is unconstrained and misses regularization to battle the adverse effects of noise and the null space (missing frequencies and finite aperture) associated with solving the least-squares imaging problems of

the type listed in Equation 4.5. To address these two problems, we combine ideas from stochastic optimization, during which we only work on randomized subsets of shots during each iteration, and ideas from sparsity-promoting optimization designed to remove the imprint of the null space and source subsampling related artifacts. As we have learned from the field of Compressive Sensing (Candès 2006; Donoho 2006; Candès and Wakin 2008), transform-domain sparsity promotion is a viable technique to remove subsample related noise in imaging via

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{x}\|_1, \\ \text{subject to} \quad & \sum_{i=1}^{n_s} \|\nabla \mathbf{F}_i(\mathbf{m}_0, \mathbf{q}) \mathbf{C}^\top \mathbf{x} - \delta \mathbf{d}_i\|_2 \leq \sigma. \end{aligned} \tag{4.6}$$

In this formulation, known as the Basis Pursuit Denoise (BPDN, Chen, Donoho, and Saunders 2001) problem, we included the sparsity-promoting ℓ_1 -norm as the objective on the curvelet coefficients \mathbf{x} of the image. These coefficients are related to the linearized data via the adjoint of the curvelet transform (\mathbf{C}^\top) and the above program seeks to find the sparsest curvelet coefficient vector that matches the data within the noise level σ . While the above problem is known to produce high-fidelity results, its solution relies on iterations that involves a loop over all n_s shots.

Stochastic gradient descent (Haber, Chung, and Herrmann 2012) is a widely used tool to make unconstrained optimization problems of the type included in Equation 4.5 computationally feasible by computing the gradient over randomized subsets of shots with a batch size (= number of shots ≥ 1) used for each gradient calculation of Equation 4.5 of $n'_s \ll n_s$. This popular algorithm solves Equation 4.5 in a few epochs (= passes through data consisting of n_s shot records) as long as the step lengths adhere to certain conditions to guarantee convergence. Unfortunately, this complicates the solution of BPDN. To avoid this complication, we reformulate, following Cai, Osher, and Shen 2009, Equation 4.6 by

replacing its convex ℓ_1 -norm objective by the strictly convex objective involving

$$\begin{aligned} & \min_{\mathbf{x}} \lambda_1 \|\mathbf{x}\|_1 + \frac{1}{2} \|\mathbf{x}\|_2^2 \\ & \text{subject to } \sum_{i=1}^{n_s} \|\nabla \mathbf{F}_i(\mathbf{m}_0, \mathbf{q}) \mathbf{C}^\top \mathbf{x} - \delta \mathbf{d}_i\|_2 \leq \sigma \end{aligned} \quad (4.7)$$

with the estimate for the image given by $\delta \hat{\mathbf{m}} = \mathbf{C}^\top \hat{\mathbf{x}}$ where $\hat{\mathbf{x}}$ is the minimizer of the above optimization problem. The mixed objective in this problem is known as an elastic net in machine learning, which offers convergence guarantees (see Lorenz, Schopfer, and Wenger 2014) in situations where during each iteration we work with different randomized subsets of shots indexed by $\mathcal{I}_k \subset [1 \cdots n_s]$ with cardinality $|\mathcal{I}| = n'_s \ll n_s$. We chose this subsets without replacement.

For $\lambda \rightarrow \infty$, which in practice means λ large enough, iterative solutions of Equation 4.7 as summarized in Algorithm 7 converge to the solution of Equation 4.6 even in situations where we work with randomized subsets of shots. Compared to iterative solutions of Equation 4.6, the iterations (lines 7–8 in Algorithm 7) correspond to iterative thresholding with a fixed threshold λ on the dual variable (\mathbf{z}_k) with a dynamic step length given by $t_k = \|\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k\|_2^2 / \|\mathbf{A}_k^\top (\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k)\|_2^2$ (Lorenz, Schopfer, and Wenger 2014). During each iteration, known as linearized Bregman iterations, the residual is projected onto an ℓ_2 -norm ball of σ by \mathcal{P}_σ . To avoid too many iterations, we set the threshold λ , related to the tradeoff between the ℓ_1 and ℓ_2 -norm objectives in Equation 4.7, to a value that is not too large—i.e., typically proportional to the maximum of $|\mathbf{z}_k|$ at the first iteration ($k = 1$). As reported by Yang, Witte, Fang, and Herrmann 2016; Witte, Louboutin, Luporini, Gorman, and Herrmann 2019, high quality images can be obtained running Algorithm 7 for a few epochs as long as the source time function \mathbf{q} and background velocity model are sufficiently accurate. As we will show below, the background velocity model also needs to be smooth so tomography-related imaging are avoided.

Algorithm 7 Linearized Bregman for SPLS-RTM

- 1: Initialize $\mathbf{x}_0 = \mathbf{0}, \mathbf{z}_0 = \mathbf{0}, \mathbf{q}, \lambda_1, \text{batchsize} n'_s \ll n_s$
 - 2: **for** $k = 0, 1, \dots$ **do**
 - 3: Randomly choose shot subsets $\mathcal{I} \subset [1 \dots n_s], |\mathcal{I}| = n'_s$
 - 4: $\mathbf{A}_k = \{\nabla \mathbf{F}_i(\mathbf{m}_0, \mathbf{q}) \mathbf{C}^\top\}_{i \in \mathcal{I}}$
 - 5: $\mathbf{b}_k = \{\delta \mathbf{d}_i\}_{i \in \mathcal{I}}$
 - 6: $t_k = \|\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k\|_2^2 / \|\mathbf{A}_k^\top (\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k)\|_2^2$
 - 7: $\mathbf{z}_{k+1} = \mathbf{z}_k - t_k \mathbf{A}_k^\top \mathcal{P}_\sigma(\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k)$
 - 8: $\mathbf{x}_{k+1} = \mathcal{S}_{\lambda_1}(\mathbf{z}_{k+1})$
 - 9: **end for**
 - 10: **Output:** $\hat{\delta \mathbf{m}} = \mathbf{C}^\top \mathbf{x}_{k+1}$
 - 11: note: $\mathcal{S}_{\lambda_1}(\mathbf{z}_{k+1}) = \text{sign}(\mathbf{z}_{k+1}) \max\{0, \|\mathbf{z}_{k+1}\| - \lambda_1\}$
 - 12: $\mathcal{P}_\sigma(\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k) = \max\{0, 1 - \frac{\sigma}{\|\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k\|}\} \cdot (\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k)$
-

4.5 On-the-fly source estimation

In practice, we unfortunately do not have access to the source time function \mathbf{q} required by Algorithm 7. Following our earlier work on source estimation in time-harmonic imaging and full-waveform inversion (Tu and Herrmann 2015b; Leeuwen, Aravkin, and Herrmann 2011), we propose an approach during which we estimate the source-time signature after each model update by solving a least-squares problem that matches predicted and observed data via a time-domain filter.

To keep our time-domain wave-equation solvers with finite differences⁴ numerically stable, we introduce an initial guess for the source time function \mathbf{q}_0 with a bandwidth limited spectrum that is flat over the frequency range of interest. Under some assumptions

⁴In our implementation, we used Devito (<https://www.devitoproject.org>) for our time-domain finite difference simulations and gradient computations (Luporini, Lange, Louboutin, Kukreja, Hückelheim, Yount, Witte, Kelly, Herrmann, and Gorman 2018), and JUDI (<https://github.com/slimgroup/JUDI.jl>) as an abstract linear algebra interface to our Algorithms (Witte, Louboutin, Kukreja, Luporini, Lange, Gorman, and Herrmann 2019)

on the source time function, we can write the true source time function as the convolution between the initial guess and the unknown filter \mathbf{w} —i.e., we have $\mathbf{q} = \mathbf{w} * \mathbf{q}_0$ where the symbol $*$ denotes temporal convolution. Because we assume one and the same source time function for all shots, we can write

$$\nabla \mathbf{F}_i(\mathbf{m}_0, \mathbf{w} * \mathbf{q}_0) = \mathbf{w} * \nabla \mathbf{F}_i(\mathbf{m}_0, \mathbf{q}_0) \quad (4.8)$$

for all sources $i = 1 \cdots n_s$. In this expression, we made use of linearity of the wave equation with respect to its source. To simplify notation, we also overloaded the temporal convolution (denoted by the symbol $*$) to apply to all data—i.e. all traces in the shot records.

Based on the above relationship, we propose to solve for \mathbf{w} after each linearized Bregman iteration (line 10 of Algorithm 8 via

$$\min_{\mathbf{w}} \sum_{i \in \mathcal{I}_k} \|\mathbf{w} * \nabla \mathbf{F}_i(\mathbf{m}_0, \mathbf{q}_0) \mathbf{C}^T \mathbf{x} - \delta \mathbf{d}_i\|_2^2 + \|\mathbf{r} \odot (\mathbf{w} * \mathbf{q}_0)\|_2^2 \quad (4.9)$$

To prevent overfitting while fitting the generated data $\tilde{\mathbf{b}}_k$ at the k th iteration to the observed data \mathbf{b}_k , we included penalties exponential weighting vector \mathbf{r} given by discretizing

$$\mathbf{r}(t) = \nu + \log(1 + e^{\alpha(t-t_0)}). \quad (4.10)$$

In this expression, the scalar α determines the rate of growth after $t = t_0$. We chose t_0 such that oscillations related to overfitting are suppressed after this time. This prevents overfitting and ensures the filters \mathbf{w}_k to be short such that the estimated source time function $\mathbf{q} = \mathbf{w}_k * \mathbf{q}_0$ remains short as well. The weight parameter ν penalizes the energy of the estimated source \mathbf{q} , which also help to relief the ill-conditionness of this sub-problem.

We summarize the different steps of our approach in Algorithm 8 below. As earlier, we solve the sparsity-promoting optimization problem via linearized Bregman iterations, which now includes in line 8 a correlation (correlation denoted by the symbol \star is the

adjoint of convolution) with the current estimate for source time correction (\mathbf{w}_k), which we initialize with the discrete Delta distribution ($\mathbf{w}_0 = \delta$). We refer to this method with on-the-fly source estimation as sparsity-promoting LS-RTM with source estimation (SPLS-RTM-SE).

Algorithm 8 LB for LS-RTM with source estimation

- 1: Initialize $\mathbf{x}_0 = \mathbf{0}, \mathbf{z}_0 = \mathbf{0}, \mathbf{q}_0, \lambda_1, \mathbf{w}_0 = \delta, \nu$, batch size $n'_s \ll n_s, \mathbf{r}$
 - 2: **for** $k = 0, 1, \dots$ **do**
 - 3: Randomly choose shot subsets $\mathcal{I} \subset [1 \dots n_s], |\mathcal{I}| = n'_s$
 - 4: $\mathbf{A}_k = \{\nabla \mathbf{F}_i(\mathbf{m}_0, \mathbf{q}_0) \mathbf{C}^\top\}_{i \in \mathcal{I}}$
 - 5: $\mathbf{b}_k = \{\delta \mathbf{d}_i\}_{i \in \mathcal{I}}$
 - 6: $\tilde{\mathbf{b}}_k = \mathbf{A}_k \mathbf{x}_k$
 - 7: $t_k = \|\tilde{\mathbf{b}}_k - \mathbf{b}_k\|_2^2 / \|\mathbf{A}_k^\top (\tilde{\mathbf{b}}_k - \mathbf{b}_k)\|_2^2$
 - 8: $\mathbf{z}_{k+1} = \mathbf{z}_k - t_k \mathbf{A}_k^\top \left(\mathbf{w}_k \star \mathcal{P}_\sigma(\mathbf{w}_k \star \tilde{\mathbf{b}}_k - \mathbf{b}_k) \right)$
 - 9: $\mathbf{x}_{k+1} = S_{\lambda_1}(\mathbf{z}_{k+1})$
 - 10: $\mathbf{w}_{k+1} = \arg \min_{\mathbf{w}} \|\mathbf{w} \star \tilde{\mathbf{d}}_k - \mathbf{b}_k\|_2^2 + \|\text{diag}(\mathbf{r})(\mathbf{w} \star \mathbf{q}_0)\|_2^2$
 - 11: **end for**
 - 12: **Output:** $\mathbf{q} = \hat{\mathbf{w}}_{k+1} \star \mathbf{q}_0$, and $\delta \hat{\mathbf{m}} = \mathbf{C}^\top \mathbf{x}_{k+1}$
-

In Algorithm 8, the symbol \star stands for the correlation, which is the adjoint operation of the convolution. Since the initial guess of \mathbf{x} is zero, we initialize the filter with one Dirac function. The sub-problem in line 10 can be solved by formulating the optimal condition and solving for \mathbf{w}_{k+1} directly.

4.6 Numerical experiments

In this experiment section, we demonstrate the viability of our approach by means of carefully designed synthetic examples. First we demonstrate the effectiveness of LB over randomized SPGL1 used in Herrmann and Li 2012 by conduct one stylized example that aims

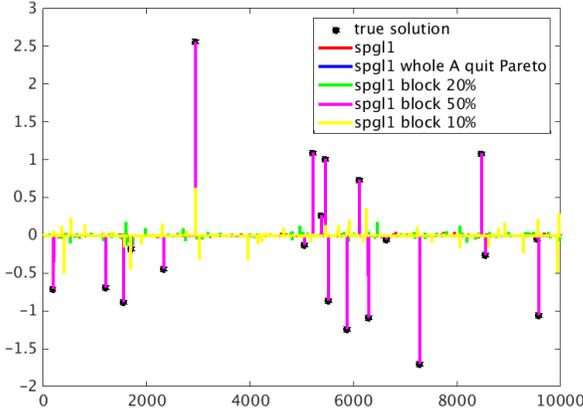
to recover one sparse vector by solving one overdetermined problem. Based on this test, we continue to design the second example to show that linearized Bregman iterations with on-the-fly source estimation is indeed able to jointly estimate the source and the sparse vector. Next, we consider the imaging experiments on the Marmousi model emphasizing the importance of including the source function and the influence of noise. We conclude by introducing a practical workflow that is capable of handling salt-related imaging problem.

4.6.1 Stylized example 1

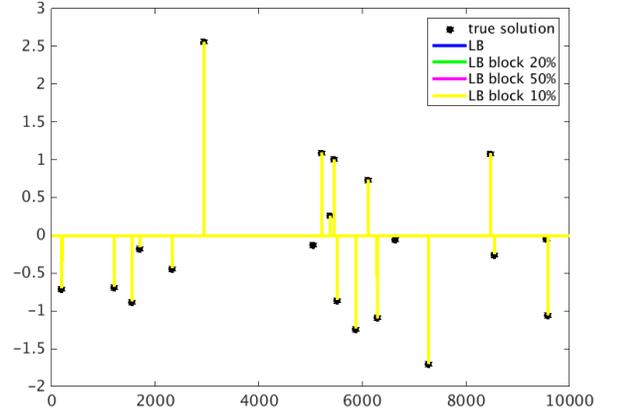
In the first example, we design a simplified experiment to show the advantages of LB over SPGL1 when applying to randomized subsets of data. Considering the enormous computational cost of the migration operator, we substitute it by a tall ill-conditioned matrix $\mathbf{A} \in \mathcal{R}^{20000 \times 10000}$ with $\text{rank}(\mathbf{A}) = 500$. The sparse vector $\mathbf{x} \in \mathcal{R}^{10000 \times 1}$ has only 20 random non-zero elements. In every iteration, a block of several rows \mathbf{A}_k are redrawn randomly. Both LB and randomize SPGL1 pass the full data set 5 times. Figure 4.1 shows the comparison of the recovery results obtained by SPGL1 and LB with different block size. Clearly, the split LB guarantees the convergence despite the block size. On the contrary, randomized SPGL1 fails to recover the correct solution. As is shown in Figure 4.1(a), due to the failure of the warming-up strategy between difference sub-problems, the recovery quality of the randomized SPGL1 becomes worse along with the decrease of the block size.

4.6.2 Stylized example 2

To verify the viability of the alternative sparsity-promoting approach in combination with on-the-fly source estimation, we examine the performance of LB with source estimation on a simplified stylized example. As we can see, Equation 4.8 implies a bilinear dependence of the reflected data on both the filter \mathbf{w} and the curvelet coefficient \mathbf{x} . It is well known that this sort of bilinear dependence can give rise to ambiguities even though the vector \mathbf{x}



(a) The solutions of randomized SPGL1 with different block sizes



(b) The solutions of LB with different block sizes

Figure 4.1: Comparison between the solutions of randomized SPGL1 and LB with different block sizes: figure (a) shows the results of SPGL1 with different block sizes. For small block size, e.g. 20% or 10% of the whole size of \mathbf{A} , randomized SPGL1 obtains results with more noisy elements at wrong locations whereas less amplitudes at the correct locations, indicating the failure in recovering the accurate solutions; Figure (b) shows that LB converges to the correct solution for all the selections of the block size.

is sparse.

We exemplify this seismic bilinear relationship by defining $\mathbf{W}\mathbf{x} = \mathbf{b}$. Now a block of the tall matrix, $\mathbf{A}_i \in \mathcal{R}^{500 \times 10000}$, $i \in [1 \dots 40]$ serves as a proxy for the LB modeling operator \mathbf{J}_i for the i^{th} shot with only one single trace. We implement the trace-by-trace convolution via a Toeplitz matrix defined in terms of the filter $\mathbf{w} \in \mathcal{R}^{500 \times 1}$ acting on each $\mathbf{A}_i\mathbf{x}$. The multiplication of matrix $\mathbf{W} \in \mathcal{R}^{20000 \times 20000}$ to $\mathbf{A}\mathbf{x}$ compactly represents the repeated convolutions of the filter to all traces.

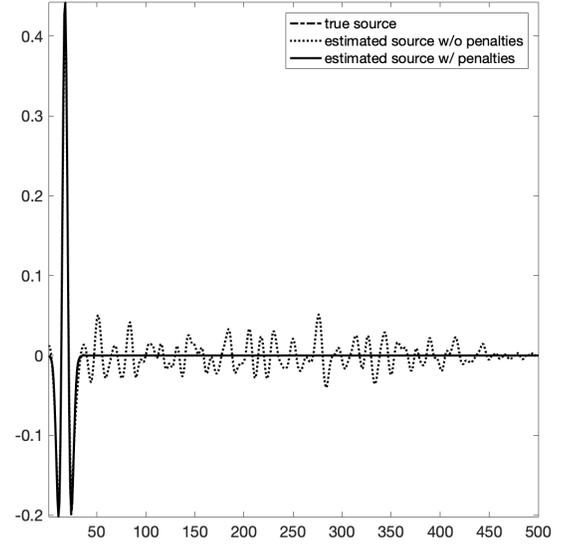
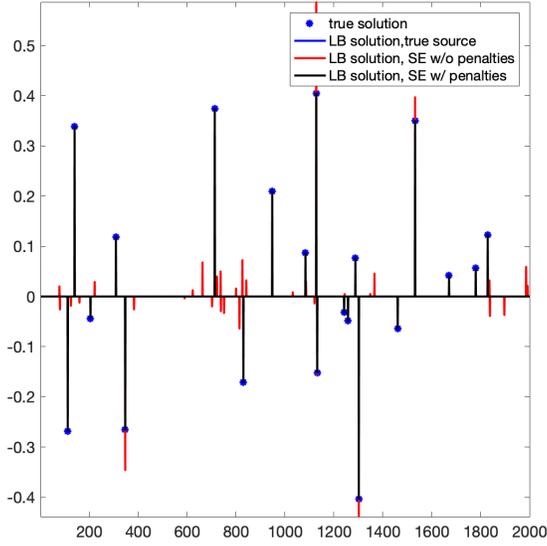
This example, designed to jointly invert \mathbf{x} and \mathbf{w} , aims to exhibit the capability of our Algorithm 8 to carry out seismic imaging and on-the-fly source estimation. To demonstrate the effect of the penalty term in line 10 of Algorithm 8, we compare sparsity-promoting solutions for the fixed true wavelet to solutions with on-the-fly source estimation with and without the additional penalty. During each iteration we randomly choose 10% blocks of the tall matrix \mathbf{A} , which means each \mathbf{A}_k contains 4 unit block \mathbf{A}_i , and we run five passes through the data in total. After some parameter testing, we chose the following values for

the penalty parameters, $\lambda = 1, \nu = 1, \alpha = 8$. We found that different choices for these/these penalty parameter have little effect on our inversion results. Finally, the time parameter t_0 is set according to the approximate duration of \mathbf{w} of the filter, which in this case corresponds to Ricker wavelet since we chose \mathbf{q}_0 to be a delta Dirac. We also initialize the filter \mathbf{w} with a normalized Dirac. Because of the amplitude ambiguity well-known to challenge blind deconvolution problems, we normalize the estimated source.

Pairs of estimated sparse "reflectivities" ($\delta\hat{\mathbf{m}}$) and source functions ($\hat{\mathbf{q}} = \mathbf{w}_{k+1} * \mathbf{q}_0$) after normalization are included in Figure 4.2. We can draw the following conclusions from these results. First, for the noise-free data, the LB iterations are able to recover the sparse reflectivity and source function well modulo a single amplitude factor, which we corrected by normalizing its ℓ_2 -norm. Second, the estimated source function and reflectivity become noisy (cf. the red line in Figure 4.2 a and the dash line in Figure 4.2 b) when we do not include a penalty enforcing the estimated filter to be short in time. Finally, the method is robust with respect to noise as we can see from Figures 4.2 c and 4.2 d where 10% Gaussian noise was added. This result also stresses the importance of including the penalty.

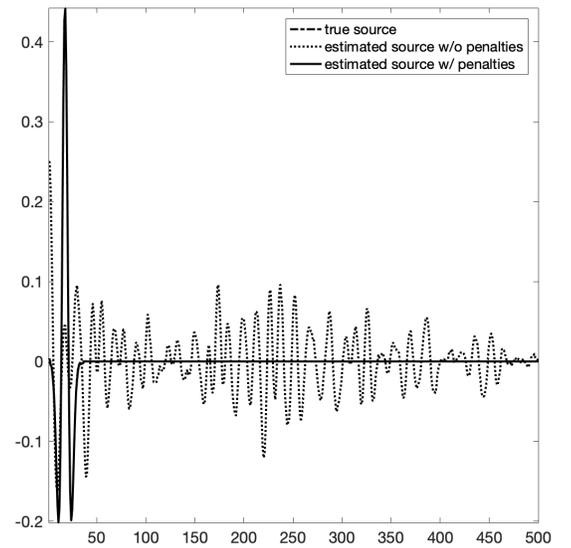
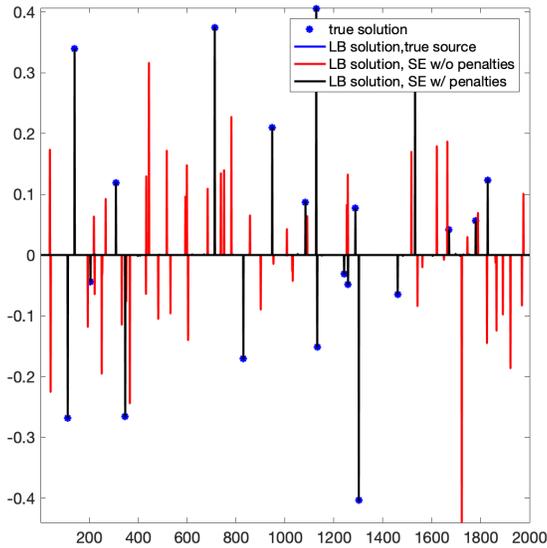
4.6.3 Experiments on the modified Marmousi model

To illustrate the performance and robustness with respect to noise of the proposed SPLS-RTM-SE method for a model with complex layered stratigraphy. We derive this imaging example from the well-known synthetic Marmousi brougois1990marmousi model, which is 3.2 km deep and 8.0 km wide, with a grid size of 5×5 m. To avoid imaging artifacts, we use a background velocity that is sufficiently kinematically accurate. We simulate the response to 320 equally spaced sources positioned at a depth of 25 m. We used a minimum phase source time function with its significant spectrum ranging from 10 to 40 Hz as shown in Figure 4.3. We used this type of source to generate linear data by applying the demigration operator ($\nabla\mathbf{F}_i(\mathbf{m}_0, \mathbf{q}), i = 1 \dots n_s$) to a bandwidth limited medium perturbation $\delta\mathbf{m}$



(a) Solutions of LB with the true source and source estimations with and without penalties for noise-free data

(b) The true source and estimated sources with and without penalties for noise-free data



(c) Solutions of LB with the true source and source estimations with and without penalties for noisy data

(d) The true source and estimated sources with and without penalties for noisy data

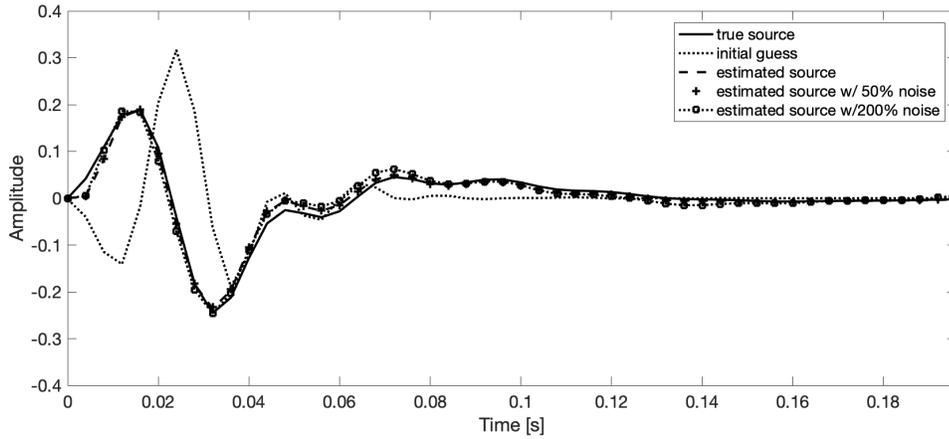
Figure 4.2: Comparison of solutions obtained with the LB iterations (see Algorithm 8) for a fixed true source (denoted by the blue line) and for on-the-fly source estimations with and without penalties. We obtained results with five passes through the data. Our method is well capable of estimating the "reflectivity" (a) and "source function" (b) after normalizing the ℓ_2 norm. The proposed method is also robust with respect to additive noise as we can see in (c) and (d). We added 10% Gaussian noise.

given by the difference between two smoothings of the true medium huang2016flexibly. We record data at 320 equally spaced co-located receivers. To assess the sensitivity to noise, we created two additional data sets by adding zero-centered Gaussian noise with energy ranging from 50% and 200% of the simulated linear data.

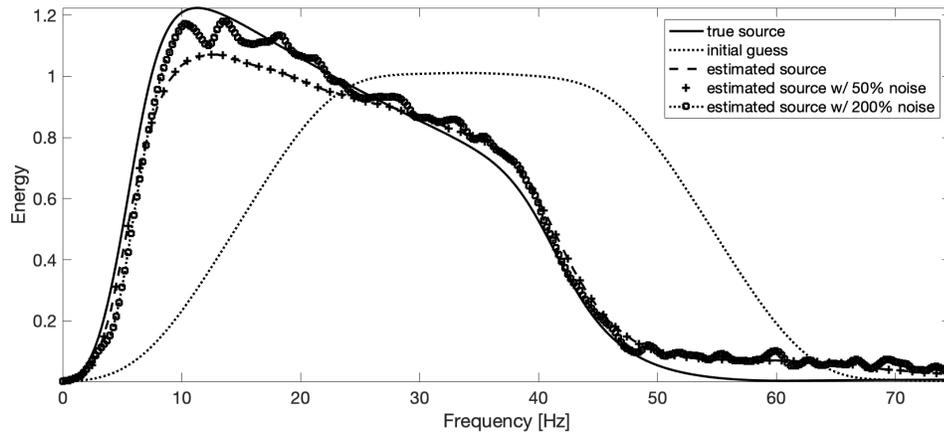
Contrary to source estimation in the frequency domain, we need an initial source function q_0 for the source time function (see Figure 4.3a and 4.3b where the initial source time function and its amplitude spectrum are depicted by the dashed black line). We need this initial source function to make sure that the finite-difference propagators remain stable. To make sure we do not exceed the valid frequency range of our simulations, we chose the frequency band of the initial source time function broad. To circumvent bias, we initialize the time function with a flat amplitude spectrum between 20 – 50 Hz. To allow for a realistic scenario, we applied a phase shift to this initial guess making it mixed phase and non-symmetric .

Before the inversion, we first investigate the importance of the source function to seismic imaging by comparing the RTM images obtained with the true and initial wavelets (Figure 4.4). For more accurate visualization, we apply a depth differentiate on both RTM images fairly. As the phase of the true and initial wavelets differ, RTM with the initial guess locates the reflectors in the wrong positions with apparent wrong phases. Furthermore, the comparison indicates that the initial guess also introduces additional artificial interfaces even though the initial guess sharpens the image resulting from its broader spectrum. This sharp image with artificial interfaces is deceptive and even disastrous. For example, oil and gas companies would prefer the sharper image. As a result, the following geological interpretation and financial decisions based on the imprecise sharp image could have serious repercussions.

To carry out the alternating inversion for the reflectivity and unknown filter w , we run Algorithm 8 for 40 iterations with a batch size of 8— i.e., we use 8 randomly selected sources per iteration without replacement. The total number of wave-equation solves is

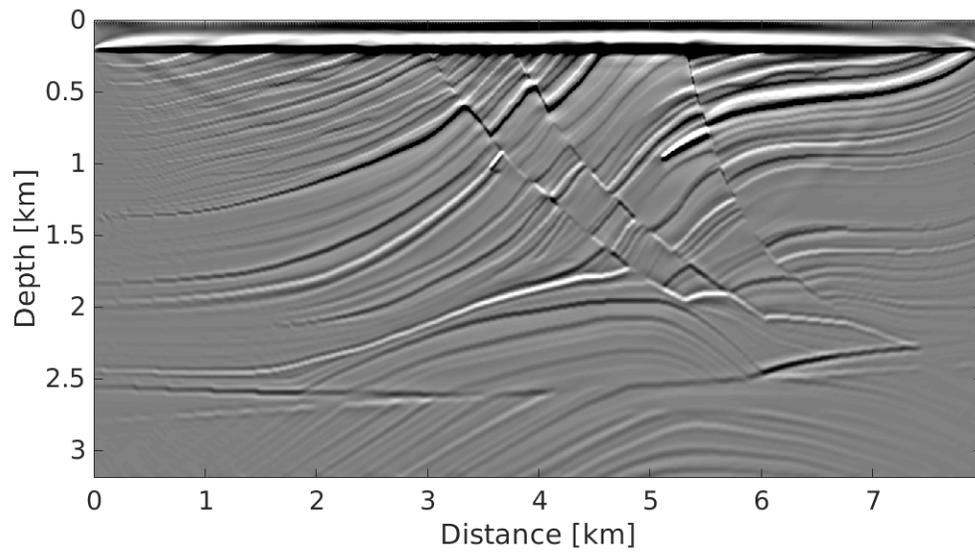


(a) Time signatures

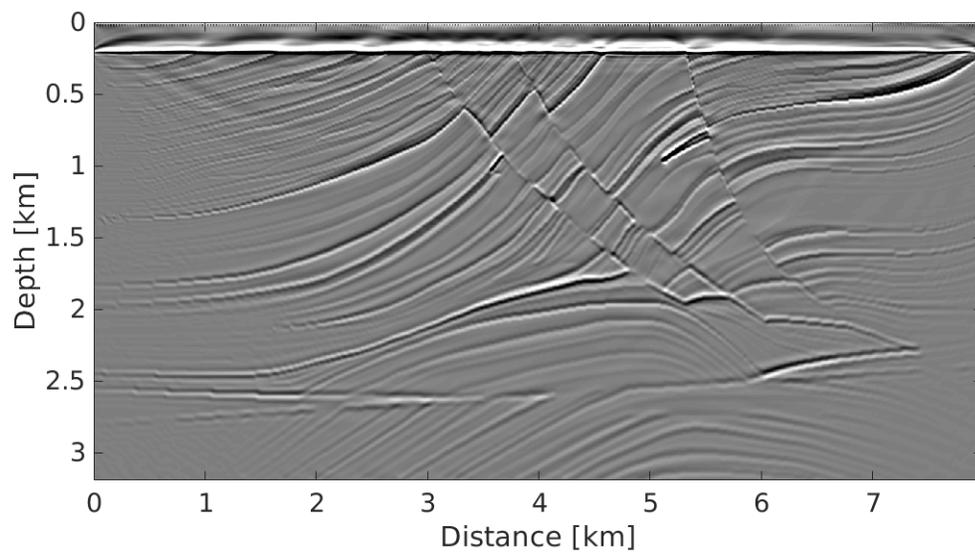


(b) Frequency spectrum

Figure 4.3: Comparison true, initial, and estimated source time functions (\mathbf{q}_0 , $\hat{\mathbf{q}} = \mathbf{w}_{k+1} * \mathbf{q}_0$) and their associated amplitude spectra. (a) the time signatures and (b) the frequency spectra. The estimated source time functions and spectra were obtained from noise-free data and from data to which zero-centered Gaussian noise was added with energy ranging from 50% and 200% of the simulated linear data.



(a) RTM with true source



(b) RTM with the initial source

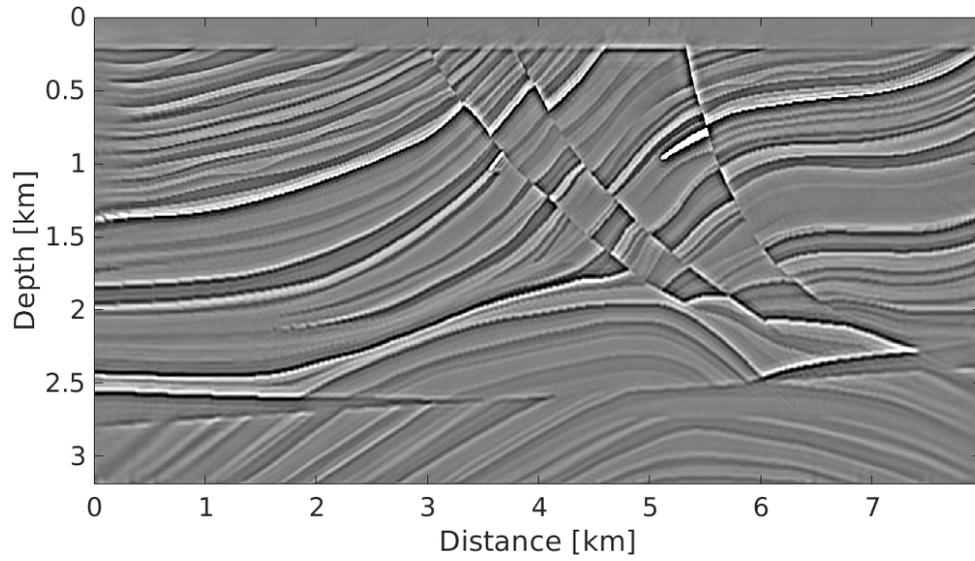
Figure 4.4: (a) and (b) are the RTM images of Marmousi with true source and initial guess of source respectively.

equivalent to touching each shot only once—i.e., we make one pass through the data. To improve the convergence of the inversion, we employ preconditioners in both the data and model domains [see Herrmann, Brown, Erlangga, and Moghaddam 2009 for detail]. To remove the imprint of the sources/receivers on the image, we also included a top mute to our operators. Similarly, we applied a mute to the data to suppress the dominating water bottom reflection and long offsets. Finally, we choose the thresholding parameter λ to be 10% of the maximum value of the first gradient to avoid unnecessary extra iterations resulting from a threshold value that is too large or small.

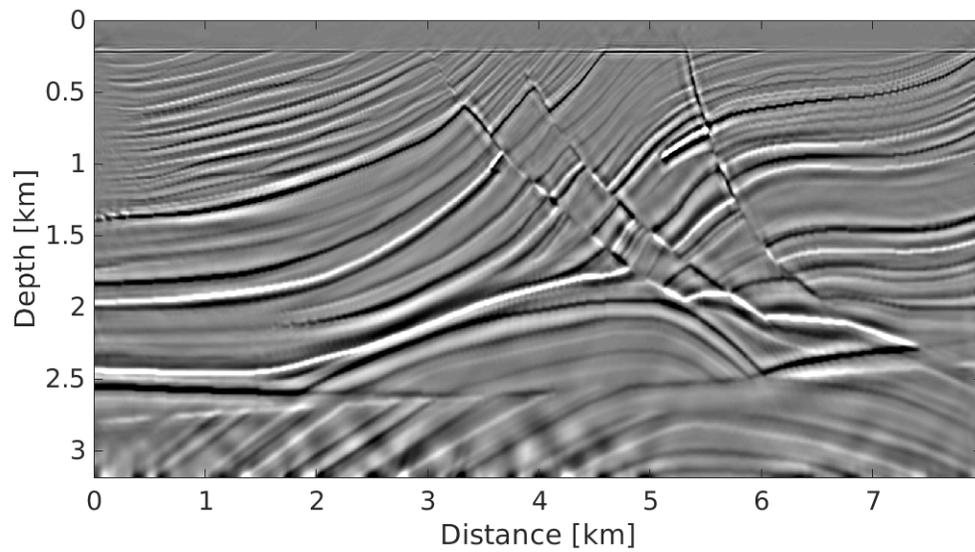
The estimated source functions $\hat{\mathbf{q}} = \hat{\mathbf{w}}_{final} * \mathbf{q}_0$ and their amplitude spectra are after ℓ_2 -norm scaling included in Figure 4.3. Overall we can see that the source functions are well recovered despite the presence of noise. For low noise, the estimated spectrum is the same as the one obtained from the noise-free data while the source function obtained from the high noise data is less smooth but closer to the true source function. Other than that we are dealing with a nonlinear blind deconvolution, we do not have an explanation for this behavior. While the noise dependence of the estimated source functions behaves somewhat aberrant, the recovered reflectivities behave as expected (cf. Figures 4.5a and 4.5b for images obtained with the true source and with the initial guess and images 4.6a – 4.6c obtained with on-the-fly source estimation for noise-free and noisy data.) We can make the following observations from these experiments. First, it is important to image with the correct source even when the data is noise-free. While our sparsity-promoting scheme is able to recover a high-resolution image (see Figure 4.5a) when the source function corresponds to the true source, the image quality deteriorates rapidly if the amplitude and phase spectra of the wavelet are wrong (see Figure 4.5b). Energy is no longer focused and the shape of locations of the imaged reflectors are off. However, the results included in Figure 4.6 demonstrate that good results can be obtained when estimating the source function on the fly. The estimated reflectivity depicted in Figure 4.6a is close to the reflectivity obtained when we image with the true source function (cf. Figures

4.5a and 4.6). Moreover, the estimated images are, as expected, relatively insensitive to noise in the data albeit the imaged reflectivity for the high noise case deteriorated somewhat (cf. Figures 4.6a – 4.6c). Contrary to the imaging result for the wrong initial source function, the reflectors are positioned correctly and have the correct phase, shape, and amplitude even in situations of substantial noise although at the expense some remaining noise, low- and high-frequency artifacts. The latter are related to the use of the curvelet transform and are to be expected. Overall, these results confirm the robustness of our imaging in a situation where there is significant noise.

To arrive at the estimated images in Figure 4.6, we set the penalty parameters $\nu = 1$ and $\alpha = 8$ in Algorithm 8. After the first source estimation in the second iteration, we reset the coefficients \mathbf{z} and \mathbf{x} to zero to avoid spending too many iterations on correcting wrongly located reflectors from the first iteration in which the initial guess of the source wavelet is used. In addition to the visual quality of the estimated images, convergence plots for the relative error for the data residual (the relative ℓ_2 -norm error between the observed data and the demigrated data for estimated reflectivity $\hat{\delta\mathbf{m}}$ convolved with the estimated filter, $\frac{\|\mathbf{w}_k * \tilde{\mathbf{b}}_k - \mathbf{b}_k\|_2}{\|\mathbf{b}_k\|_2}$) and the relative model error (the ℓ_2 -norm error between the true reflectivity and the recovered reflectivity, $\frac{\|\hat{\delta\mathbf{m}}_k - \delta\mathbf{m}\|_2}{\|\delta\mathbf{m}\|_2}$) confirm our observation that Algorithm 8 is capable of providing high quality images in the absence of precise knowledge on the source function and in the presence of substantial noise. Our approach arrives at these least-squares images at the cost of a single data pass. Understandably, the algorithm starts off with a large relative residual and model error due to the wrong initial guess for the source function. As Algorithm 8 progresses, these relative errors continue to decay and are comparable to the convergence plots for the true source function. Because on-the-fly source estimation improves our ability to adapt to the data, the relative data residual for the noise-free case (dashed line) is even better than the relative error in case the source function is known (solid line). While encouraging, these results are obtained for a relatively simple imaging experiment and for data that is obtained with linearized modeling via demigration.

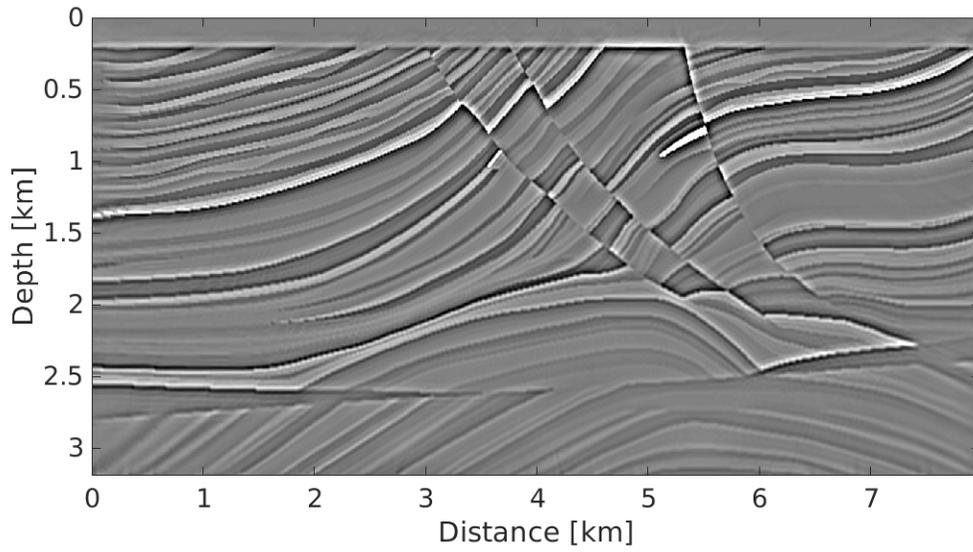


(a) Inverted image with true source

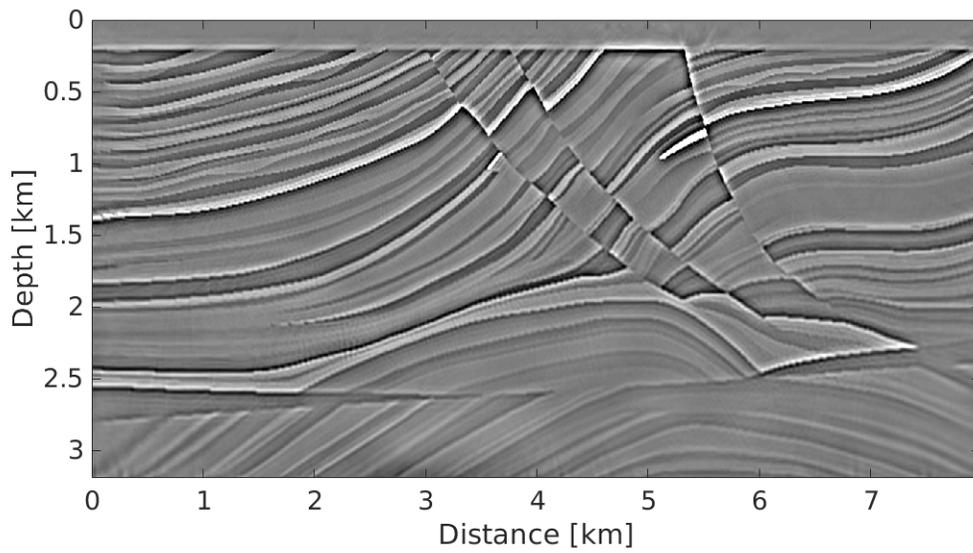


(b) Inverted image with the initial guess

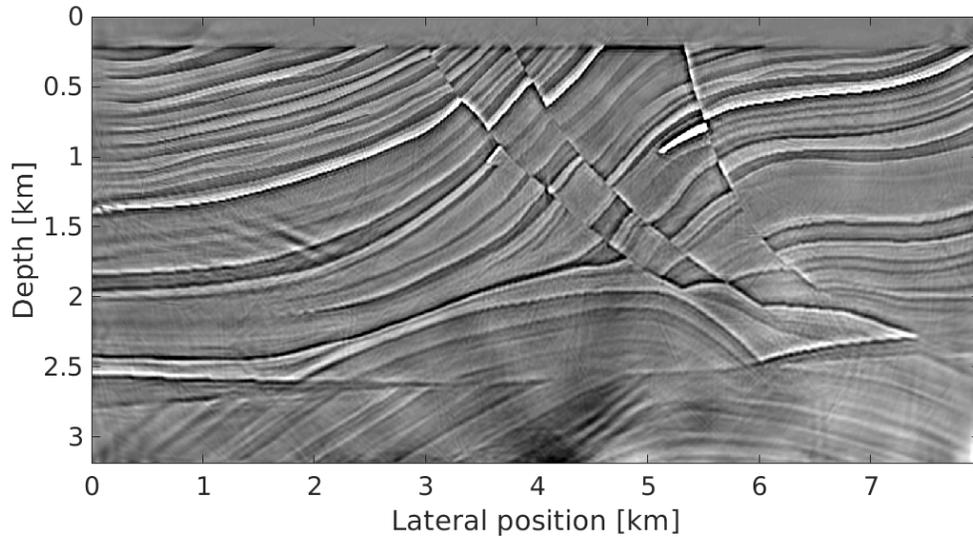
Figure 4.5: Inverted images with (a) the true source and (b) initial source, generated by the Marmousi model.



(a) Inverted image with source estimation for noise-free data



(b) Inverted image with source estimation for data with 50% noise



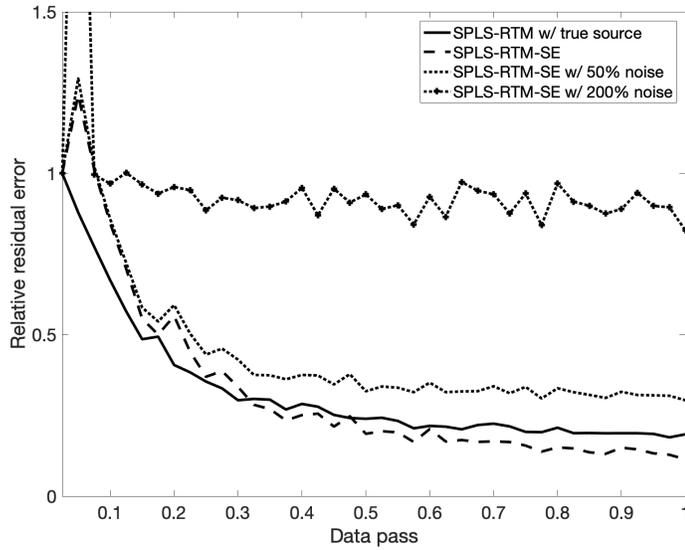
(c) (c) Inverted image with source estimation for data with 200% noise

Figure 4.6: Inverted images with source estimation (Algorithm 8) for (a) noise-free data, (b) data with 50% noise and (c) data with 200% noise, generated from the Marmousi model.

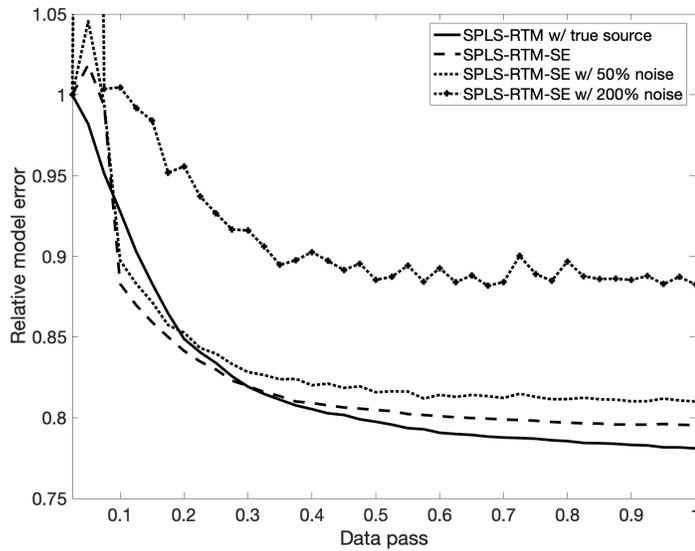
In other words, we commit an inversion crime. In the next section, we will show that the proposed method also performs well in more complicated settings with nonlinear data.

4.6.4 Experiments on Sigsbee model

Sparsity-promoting imaging algorithms such as SPLS-RTM (Algorithm 8) are designed to handle complex imaging scenarios with strong velocity contrasts and strong lateral velocity variations. Examples of such scenarios are salt plays where reflections underneath the salt are of interest. To demonstrate the viability of our imaging approach with on-the-fly source estimation, in this scenario we consider the challenging Sigsbee2A model of size 24.4×9.2 km. This model contains a large salt body and a number of faults and diffractors. To demonstrate the capability of our approach to handle this challenging situation, we simulate nonlinear data for a marine acquisition without a free surface. We model 960 sources in total recorded with an array with 320 receivers sampled at 25 m and with a maximum offset of 8 km and towed at a depth of 15 m. We used a source wavelet with a peak at 15 Hz (see Figure 4.8) and we record for 10 s.



(a) Residual decay



(b) Model error decay

Figure 4.7: Convergence plots for the relative residual error (a)—i.e., the ℓ_2 -norm error between the observed data and the demigrated data for estimated reflectivity $\hat{\delta m}$ and the relative model error (b)—i.e., the ℓ_2 -norm error between the true reflectivity and the recovered reflectivity.

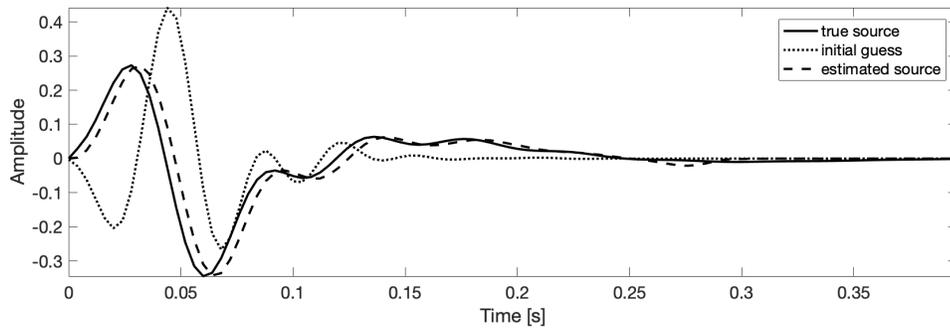
As is customary during imaging under salt, we use a background velocity model that features salt with relatively strong and therefore reflecting boundaries. We approximate linear data by using this background velocity model to generate data, which we subtract from the simulated data in the hard true Sigsbee2A model. Because the presence of salt in the background model, the incident wavefield contains reflections that give rise to unwanted low-frequency tomographic artifacts in the image. This problem is widely reported in the literature (Root, Stolk, and Hoop 2010; Whitmore and Crawley 2012; Witte, Yang, and Herrmann 2017). To remove these imaging artifacts, we replace the conventional imaging condition for RTM by the inverse-scattering imaging condition (Root, Stolk, and Hoop 2010; Whitmore and Crawley 2012; Witte, Yang, and Herrmann 2017). While this condition has proven capable of removing tomographic artifacts during RTM (Whitmore and Crawley 2012; Witte, Yang, and Herrmann 2017) and sparsity-promoting least-squares RTM (Witte, Yang, and Herrmann 2017) it changes the linearized forward operator (the Jacobian ∇F_i), resulting in an inconsistent system. Contrary to RTM with the conventional imaging condition, imaging with the inverse scattering imaging condition corresponds to estimating perturbations in the impedance rather than in the velocity.

Unfortunately, this difference in which quantity is being imaged is problematic for our proposed on-the-fly source estimation, which tries to correct for inconsistencies between "observed" data and predicted data. Contrary to the situations where we use the conventional imaging condition, the data residual now contains contributions from the wrong wavelet and the linearized imaging condition. and this leads to wrong estimates for the unknown source function. We overcome this problem via a hybrid iterative algorithm where we switch imaging conditions during the iterations outlined in Algorithm 8. To estimate the source function, we first iterate with the conventional imaging condition. Since the convergence to the source function is fast, we switch after five iterations to the scattering imaging and keep the estimated source function fixed. Basically, we jump from Algorithm 8 to Algorithm 7.

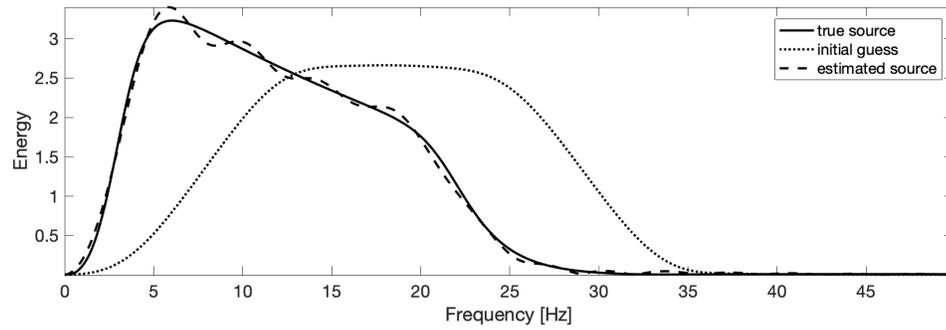
Results of this hybrid approach are summarized in Figures 4.8 – 4.10. As before, we compare our results with on-the-fly source estimation to SPLS-RTM for the true source function. The initial guess and estimated wavelets in Figure 4.8 again confirm the validity of our approach, yielding a reasonably accurate estimate for the source after only five iterations and subsequent normalization of the ℓ_2 -norm. Imaging results obtained after twenty iterations with 10% of the shots, which amounts to two data passes in total, are included in Figures 4.9 and 4.10. Unlike a typical RTM image (Figure 4.9a), images yielded by SPLS-RTM are well resolved and true amplitude. This is because we invert the linearized modeling operator, which compensates for the source, finite aperture, and propagation effects. As before, we included preconditioners and mutes. Comparison of Figure 4.9b, obtained with Algorithm 7 with the true source function, and Figure 4.9c, which we computed with our hybrid method switching from Algorithm 8 to Algorithm 7 after five iterations, shows near identical results confirming the validity of the proposed approach. These observations are confirmed by the trace-by-trace comparisons in Figure 4.10.

To trigger the inversion with source estimation, we follow the same strategy used in the Marmousi experiments, which entailed a mixed-phase wavelet with a plateau between 15 – 25Hz in the spectrum, which is higher than the approximate peak frequency of the true wavelet centered from 5Hz to 20Hz, indicated in Figure 4.8(b).

For both SPLS-RTM with true source and SPLS-RTM-SE, we conduct 20 iterations using the LB method with 10% sources per iteration, which corresponds to two passes through the data. We implement the basic preconditioners, that is, the depth scaling preconditioner and the data topmute, used in the Marmousi experiments. In SPLS-RTM, We also use the linearized inverse scattering imaging condition (Whitmore and Crawley 2012; Witte, Yang, and Herrmann 2017) instead of the traditional cross-correlation imaging condition to deal with low-frequency artifacts in the gradients caused by significant backscattering of the forward modeled wavefield from the strong salt body of the Sigsbee2A model. Both imaging operator \mathbf{J}^\top and modeling operator \mathbf{J} must incorporate the modification of the



(a) Time signatures

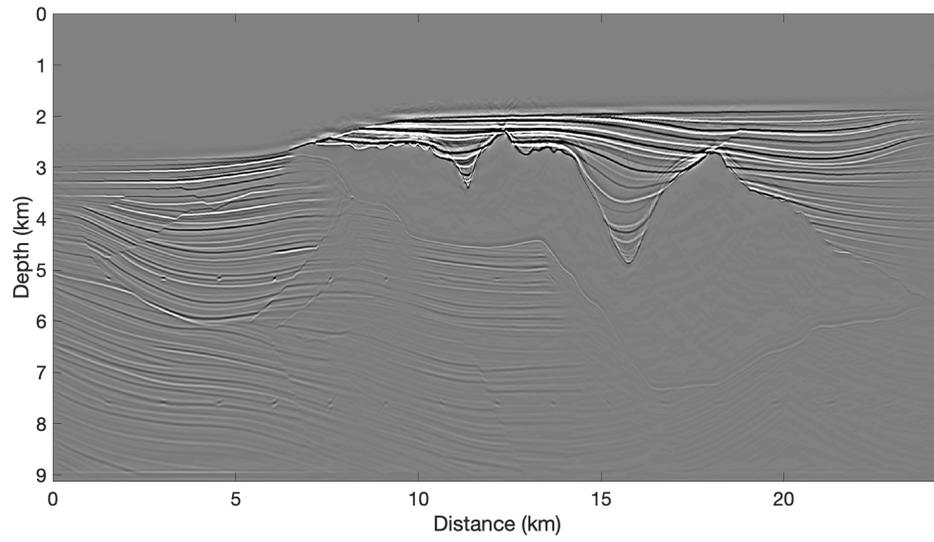


(b) Frequency spectrum

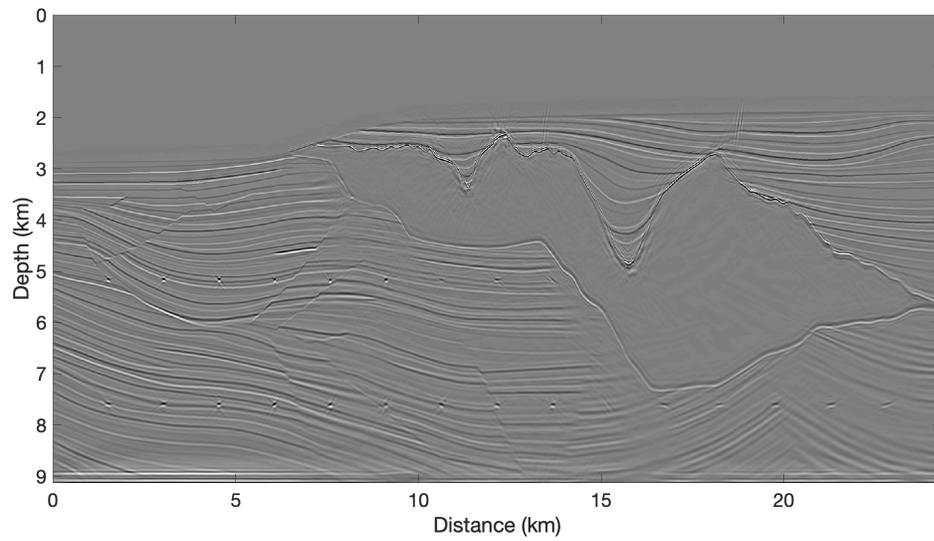
Figure 4.8: Comparison true, initial, and estimated source time functions ($\mathbf{q}_0, \hat{\mathbf{q}} = \mathbf{w}_{k+1}\mathbf{q}_0$) and their associated amplitude spectra. (a) the time signatures and (b) the frequency spectra. The estimated source time functions and spectra were obtained during the first five iterations with the conventional imaging condition.

imaging condition to preserve their adjointness as one pair. In SPLS-RTM-SE, considering that the Jacobian with cross-correlation imaging condition obeys Taylor extension and approximates the nonlinear data better, we use this Jacobian and estimate source in the first 5 iterations. Warmed with these iterations, we switch to the Jacobian with inverse scattering imaging condition.

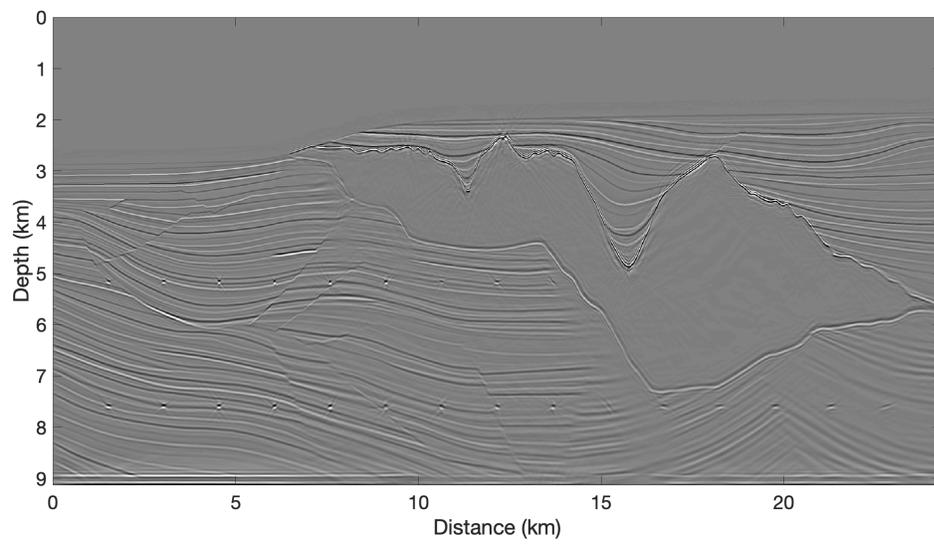
The strategy to choose the parameters λ , ν and α is same as in the Marmousi experiments. Since in our test, the define of the shape of \mathbf{r} is quiet stable, here we still use the vector in Marmousi experiments.



(a) RTM with the true source wavelet

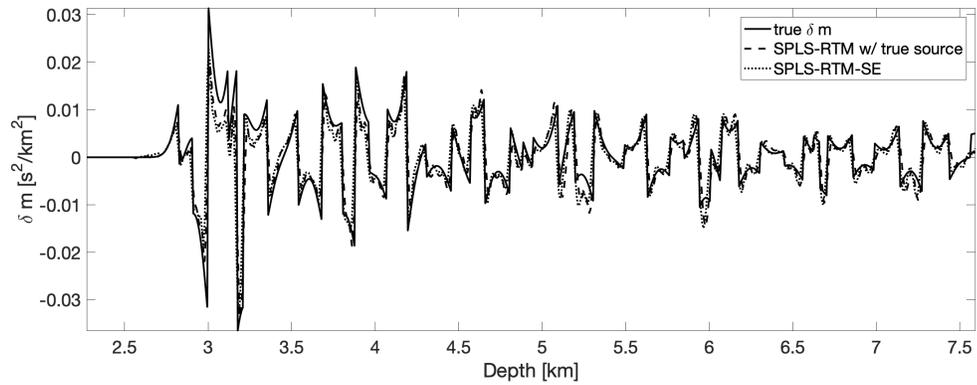


(b) SPLS-RTM with the true source wavelet

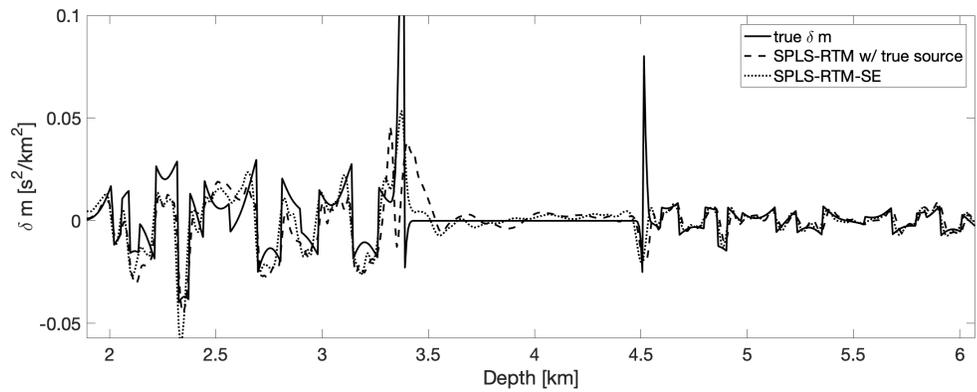


(c) SPLS-RTM with on-the-fly source estimation

Figure 4.9: Comparison between SPLS-RTM with the true source wavelet and SPLS-RTM with on-the-fly source estimation. In both cases, we did total 20 iterations amounting to two data passes.



(a) Traces at $x = 4.5\text{km}$



(b) Traces at $x = 11.3\text{km}$

Figure 4.10: Trace by trace comparisons between the true model perturbation, the images from SPLS-RTM with the true source and with on-the-fly source estimation. The traces in (a) and (b) are extracted from lateral positions $x = 4.5\text{km}$ and $x = 11.3\text{km}$, respectively

4.7 Discussion

Many researchers have shown the capability of LS-RTM in producing high-resolution and true amplitude subsurface images; the usage of this approach, unfortunately, incurs an enormous computational cost. The balance between the computational cost and image quality is a critical problem that must be solved for the successful applications of LS-RTM to industrial-scale problems. This study proposed a time-domain sparsity-promoting LS-RTM that incorporates the art-to-the-state technologies of stochastic optimization, curvelet-based constraints, and the linearized Bregman method. Through random selection of shots during each iteration, this method significantly reduces the computational cost. Meanwhile, employing the curvelet-based constraint, the proposed approach suppresses noise-related and sub-sampling-related artifacts. The LB approach also enables us to redraw a random subset at each iteration and guarantees convergence. As a result, the proposed method generates high-resolution subsurface images with a computational cost of two conventional RTMs.

Another bottleneck for the successful application of LS-RTM is the absence of accurate source signatures. To solve this problem, we embedded an on-the-fly source estimation step into the basic LB workflow. The estimation step involves solving a linear sub-problem in which a number of solutions are possible solutions due to the limited band of the source. To mitigate the non-uniqueness, we introduced both event-based and energy-based penalties to regularize the sub-problem. Several numerical cases illustrate that the usage of the two penalties significantly mitigates the non-uniqueness of the sub-problem and enables us to find the correct source functions.

The proposed time-domain SPLS-RTM-SE may have a strong potentiality for industrial-scales applications for the following reasons. First, frequency-domain approaches are often infeasible for 3D problems because of the complexity of solving the 3D Helmholtz equation with high frequency. With the usage of art-to-the-state technologies including fast stencils

and parallel computing, the computational cost of solving the time-domain wave equation has been reduced to an acceptable level, which results in many 3D RTM practical applications (Baysal, Kosloff, and Sherwood 1983). Since the computational cost of the proposed approach roughly equals to one or two conventional RTM, the proposed approach is computationally feasible for 3D industrial applications. Moreover, since the proposed source estimation technique does not require additional PDE solves, we can embed it into the 3D LS-RTM in a straightforward manner without any additional computational concerns. Thus, in the future, we can expect a practical 3D application of the proposed SPLS-RTM-SE.

However, all the feasible and possible works on the effective inversion we discussed above is under the assumption that there is only primaries in the observed data, which is hard to require in the real marine acquisition, where the free surface together with the ocean bottom will generate strong multiples, i.e. higher order reflectivities, bouncing between these two interfaces. If without proper pre-processes, e.g. demultiple (Weglein, Gasparotto, Carvalho, and Stolt 1997; Biersteker 2001; Verschuur and Berkhout 2005; Hargreaves 2006; Brittan, Martin, Bekara, and Koch 2011; Lin and Herrmann 2016), the imaging condition used to locate the reflectors will cross-correlate the multiples with the primaries to form the mirrored artifacts in the images. In the next chapter we will discuss the extension of the effective sparsity-promotion into the multiples imaging to solve the problem of cross order imaging artifacts.

4.8 Conclusion

We proposed a scalable time-domain approach to sparsity-promoting least-squared reverse time migration with on-the-fly source estimation in principle suitable for industrial 3D imaging problems. The presented approach leverages recently developed techniques from convex optimization and variable projection that greatly reduce costs and the necessity to provide an estimate for the source function. As a result, our approach is capable of

generating high-fidelity true-amplitude images including source estimates at the cost of roughly one to two migrations involving all data.

By means of carefully designed experiments in 2D, we were able to demonstrate that our method is capable of handling noisy data and complex imaging settings such as salt. We were able to image under salt, which is often plagued by low-frequency tomographic artifacts, by switching between applying the conventional imaging condition initially, followed by iterations that apply the inverse-scattering condition. In this way, we estimated the source function first while creating an artifact-free image with later iterations during which the imaging condition was switched while keeping the source function fixed.

Because the presented method relies on time-domain propagators, we anticipate it will be able to scale to large 3D industrial imaging problems. Because 3D imaging with full-azimuthal sparse data typically provided good illumination of the reservoir, we expect the proposed methodology to produce high fidelity results at a cost of roughly one to two reverse time migrations involving all shots.

CHAPTER 5

SPARSITY-PROMOTING LEAST-SQUARES REVERSE TIME MIGRATION WITH MULTIPLES

5.1 Summary

In this chapter, we focus on the least-square reverse time migration for primaries and multiples together. Instead of predict the primaries by pre-processes of demultiples e.g. SRME or EPSI, we use the total up-going data to invert the model perturbation based on the SPLS-RTM framework we propose in chapter 4 by injecting the areal source of the total down-going wavefield. With this methodology, we can remove most imaging artifacts related to the presence of surface-related multiples, introduced by a shallow ocean bottom, which are difficult to predict by traditional demultiples pre-process such as SRME due to the energy leakage from adaptive subtraction. We test our method to one linear data set in time-harmonic domain on part of the Sigsbee2A model, where the multiples are generated by the Born modeling with respect to only the velocity perturbation. We also show the extra illumination by introducing the secondary source (areal source) into the inversion when there is areas missing sources. Then we move to invert the nonlinear data set in the time domain, where the data is generated by iWave with free surface and directive wave deleted, following with up- and down-going wavefield decomposition and extrapolation. By injecting the areal source in the inversion, most of the phantoms from the first order multiples for the deeper part of the model are removed, but there are still some leftover artifacts in the shallower part, and the inverted perturbations at the ocean bottom have some phase errors because we are inverting with only Born with respect to the velocity. We also compare the recovered multiples and primaries with the ideal data that generated with absorbing surface. We conduct one data pass for the inversions in both scenarios, which cost roughly only one

RTM. Our time-domain inversion based on Devito can be easily extend to the further 3D scenarios.

5.2 Introduction

In the chapter 4, we focus on the least-square reverse time migration with only primaries. By fitting the modeled primaries to observed primaries in the least-squares sense, least-squares reverse-time migration (LS-RTM, (Guitton, Kaelin, and Biondi 2006)) can remove the imprint of the source wavelet, limited aperture, and other amplitude effects on migrated images. Since minimizing the ℓ_2 -norm on the data residual attempts to invert a highly overdetermined but inconsistent system, resulting images often suffer from overfitting. One possible way to remove these artifacts is to impose some sort of regularization onto the original LS-RTM formulation. Aside from imaging artifacts associated with possible overfitting, LS-RTM is also computationally prohibitively expensive withstanding its widespread adaptation. Motivated by ideas from Compressive Sensing (Donoho 2006), Herrmann and Li 2012 proposed to solve the expensive overdetermined and inconsistent system of LS-RTM by solving a series of much smaller and therefore much cheaper randomized subproblems. Thanks to this randomized subsampling, we are able to carry out sparsity-promoting LS-RTM at the cost of roughly one-to-three passes through the data. However, the resulting images remained somewhat noisy, a well-known by product of stochastic optimization methods where different subsets of shots are used during each iteration. By replacing the ℓ_1 -norm objective by an elastic net, as proposed by Lorenz, Schopfer, and Wenger 2014 in the field of online Compressive Sensing, these remaining noisy artifacts can be removed as shown by Herrmann, Tu, and Esser 2015a. Following this work, we (Yang, Witte, Fang, and Herrmann 2016) implemented this linearized Bregman method in the time domain with on-the-fly source estimation by variable projection.

So far, this work mostly involved imaging of primaries only. For (shallow) Marine data, this assumption requires separation of primaries from the total down-going wavefield,

which includes surface-related multiples. One way to separate the primaries is to consider the surface-related multiples elimination (SRME) relation (Verschuur, Berkhout, and Wapenaar 1992), which models multiples as a multi-dimensional convolution between the vertical derivative of the surface-free Green's function and the total down-going wavefield. While this relation has resulted in technologies such as SRME and Estimation of Primaries by Sparse Inversion (EPSI), (Lin and Herrmann 2013) that mitigate the adverse effects of surface-related multiples successfully, it is computationally very expensive because it involves multi-dimensional convolutions that correspond to dense matrix-matrix multiplies. Also, SRME requires the sources to be co-located with the receivers, which can be expensive as well. Besides the computational cost, SRME struggles to estimate the source wavelet and therefore the shape of the recovered primaries may get distorted, especially in shallow water acquisitions. EPSI on the other hand, maps the multiples to primaries, which offers the potential usage of these multiples to help with illumination of the subsurface. Lu, Whitmore, Valenciano, and Chemingui 2015 and also Tu and Herrmann 2015a used the fact that the bounce points of surface-related multiples can be considered as secondary sources to improve migrated images. Lu, Whitmore, Valenciano, and Chemingui 2015 carried out these secondary sources into imaging by replacing RTM's cross-correlation-based imaging condition with a deconvolution. Although the later approach has provides results (Lecerf, Hodges, Lu, Valenciano, Chemingui, Johann, and Thedy 2015) spectacular in improving the illumination, the deconvolution image conditions it used can lead to unwanted crosstalk caused by the interference between different orders of multiples.

By integrating the SRME relationship into sparsity-promoting LS-RTM, Tu and Herrmann 2015a was able to properly model surface-related multiples resulting in an inversion procedure where surface-related multiples are mapped to imaged reflectors. His main contribution was that integrating the SRME relation into LS-RTM simply corresponds to adding the down-going wavefield as an areal source. As such Tu and Herrmann 2015a arrived at a result where the multi-dimensional convolutions of EPSI are carried out by the

wave-equation solver while this formulation also no longer requires co-location of sources and receivers during acquisition. While this work (Tu and Herrmann 2015a) has resulted in high-quality multiple-free images for the designed linear data, it relied on an unnatural strong perturbation in the velocity to generate realistic surface-related multiples in the water column. We remove this problem by including density variations at the ocean bottom into our time-domain formulation (Yang, Witte, Fang, and Herrmann 2016). Because our time-stepping formulation is based on Devito (Lange, Kukreja, Louboutin, Luporini, Vieira, Pandolfo, Velesko, Kazakas, and Gorman 2016) — a just in time compiler for stencil-based finite-difference codes — we envision that the proposed approach can readily be extended to 3D seismic.

This chapter is organized as follows. First, we describe multiples prediction with the SRME and formulate the areal source injection into our linearized Born modelling operator mapping velocity and density perturbations to linearized data. Next, we derive an optimization scheme for LS-RTM using a mixed ℓ_1 - ℓ_2 -norm objective function with areal source included. We solve this problem with linearized Bregman. The experiments are set up with part of the synthetic Sigsbee2A model (Paffenholz, Stefani, McLain, and Bishop 2002). We first show some early inverted images that we obtained in time-harmonic domain for linear designed data with density involved in the modeling kernel. Then we implement in time domain based on Devito and compare the inversion results of total down-going wavefield with and without areal source injection for nonlinear data.

5.3 SRME and areal source

All the successful inversion-based imaging methods mentioned in chapter 4 are designed for primary order reflections (primaries) only. However, in the marine surveys, the reflections bounce strongly between the ocean bottom and the free water surface, shown in Figure 5.1, generating strong higher order reflections that declared as multiples (Figure 5.2).

Due to the presence of the surface-related multiples (mentioned as multiples for short

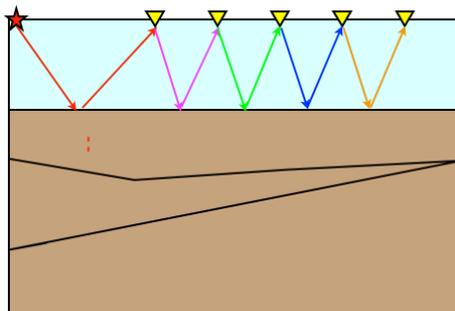


Figure 5.1: The generation of different ordered reflections. The star and the triangles floating in the water volume stand for the source and receivers respectively. From left to right, the lines with arrowhead in different colors indicate the reflected paths of primaries and multiples with higher and higher order.

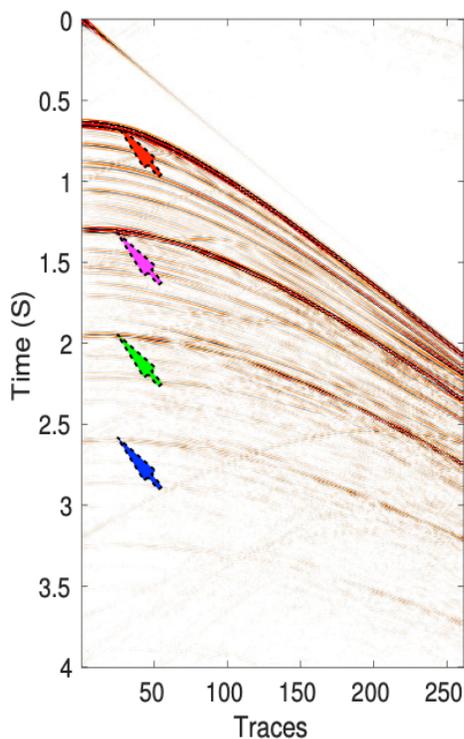


Figure 5.2: One shot gather that contains multiples, obtained by the marine survey shown in Figure 5.1. From top to bottom, different colored arrows mark out the reflection events of different orders.

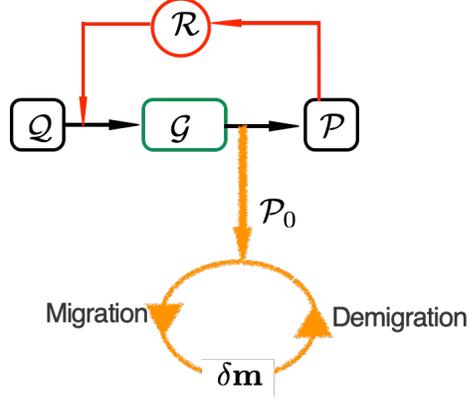


Figure 5.3: Imaging workflow with SRME demultiples. The primaries \mathcal{P}_0 , as the input in the looping migration-demigration sub-workflow, are predicted by SRME

in the following chapters) that generated by the reflections bouncing between the ocean bottom and the free surface in the marine survey, the traditional imaging workflow needs to involve special demultiple pre-processes to predict the primaries. Figure 5.3 displays schematically the imaging workflow combined with SRME (Verschuur, Berkhout, and Wapenaar 1992), the outstanding demultiple processing. There the SRME relation describes the mechanism of multiples generation with a loop (shown in Figure 5.3) by relating the vertical derivative of the surface-free Green's function and the down-going wavefield to the total up-going wavefield. The total up-going wavefield is obtained by receiver-side deghosted and extrapolated to the surface (Wapenaar 1998) on the marine data.

We express the SRME relation (Verschuur, Berkhout, and Wapenaar 1992) as below:

$$\mathcal{P}_i = \mathcal{G}_i(\mathcal{Q}_i + \mathcal{R}_i\mathcal{P}_i), \quad (5.1)$$

where the subscript $i = 1 \dots n_f$, with n_f the total number of discretized frequencies. The monochromatic matrix \mathcal{P}_i stands for the total up-going wavefield at the surface with n_s common-shot gathers in its columns and n_r common-receiver gathers in its rows. Note that this data matrix \mathcal{P}_i does not include direct waves. The matrix \mathcal{G}_i denotes the surface-free dipole Green's function organized in a similar fashion. The matrix \mathcal{Q}_i represents the down-going point source wavefield, \mathcal{R}_i is the reflectivity at surface, and normally is considered

to be -1 . The whole term $\mathcal{R}_i \mathcal{P}_i$ acts as an areal source wavefield, which when multiplied by \mathcal{G}_i produces the total up-going wavefield including surface-related multiples.

In SRME's workflow, the predicted primaries are updated iteratively by $\mathcal{P}_{0j} = \mathcal{P} - \mathcal{R} \mathcal{Q}_j^{-1} \mathcal{P}_{0j} \mathcal{P}$, derived from equation 5.1, where the subscript j indicates the iteration index, and the discretized frequency index i is dropped for simplicity. In the first iteration, \mathcal{P}_{01} is initialized by \mathcal{P} . The unknown source term \mathcal{Q}_j is obtained through solving one least-square problem to minimize the previous primaries' energy $|\mathcal{P} - \mathcal{R} \mathcal{Q}_j^{-1} \mathcal{P}_{0j-1} \mathcal{P}|_2^2$. This adaptive subtraction introduces energy leakage by involving the implicit assumption of lowest energy for primaries, while the fact is the ideal primaries contain the fewest events. Especially in shallow water case where the multiples arrive early and interrupt the group of primary events, this energy leakage could be a serious problem. Figure 5.4 gives an failed example for SRME in shallow water case. Here the total up-going data (Figure 5.4(a)) is obtained by summing the primaries in Figure 5.4(b) and the designed one-order higher multiples together. Given by the exact multiples, the predicted primaries by SRME contains residual events leaked from multiples, therefore the true primaries (marked by arrows) are distorted to minimize the total energy. Further distortions and artifacts in the following imaging or inversion workflows are expected.

5.3.1 EPSI

To break the drawback of minimized energy assumption in SRME, EPSI tries to invert for the sparse Green's function (sketched by the green lines in Figure 5.5) and predict the primaries. Here the sparsity assumption in Green's function physically aims to restrict the primary events as less as possible. However, the computational cost will increase because EPSI is based on the sparsity recovery methods which call for amount of iterations, during each, lots of fast Fourier transform (FFT) and IFFT, and matrix-by-matrix multiplications for the convolution between areal source and Green's function are required.

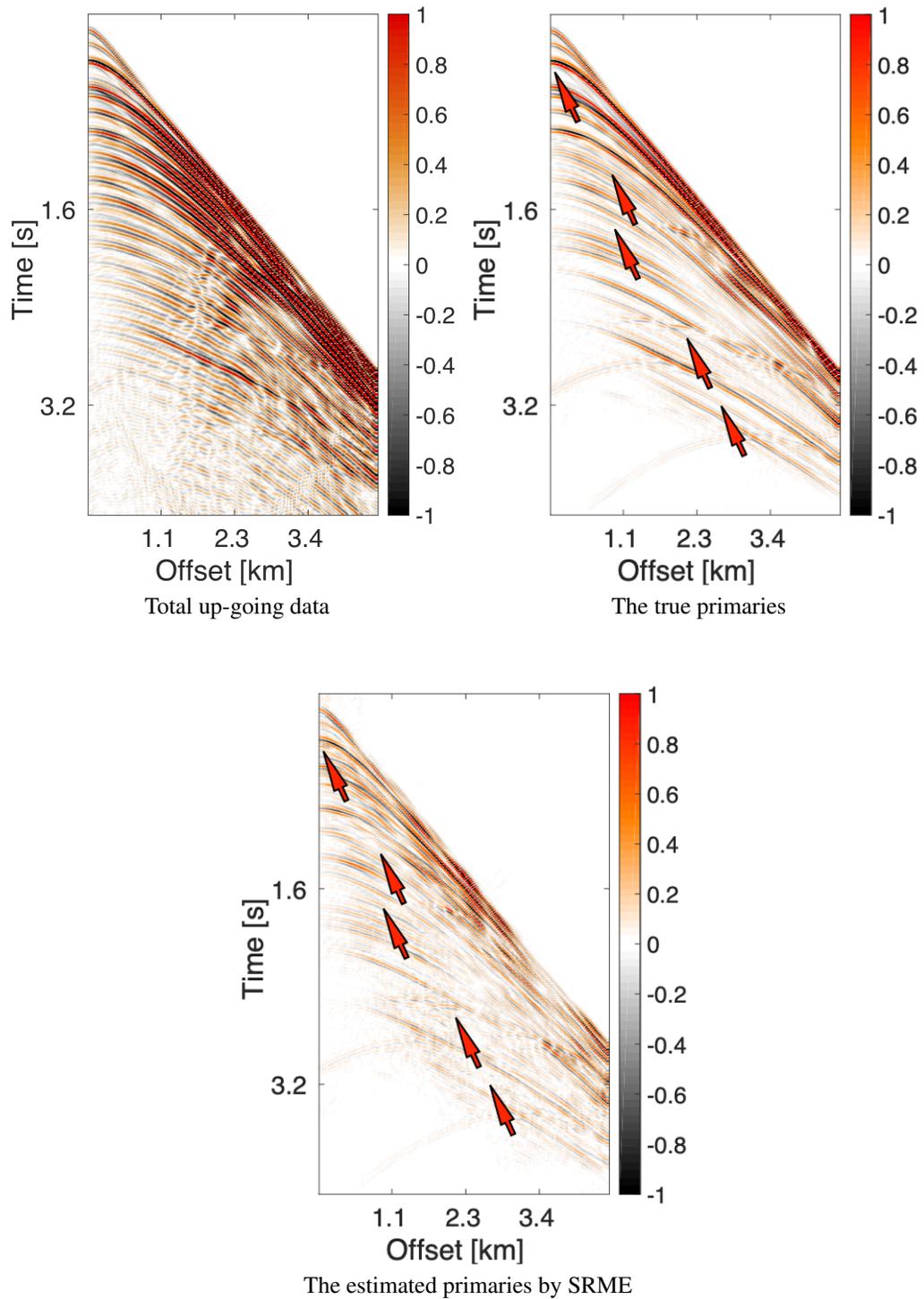


Figure 5.4: Example for SRME's failure in shallow water. (a) The designed total up-going data that generated by summing the designed primaries in (b) with its one-order higher data. (c) The predicted primaries by applying the ideal multiples to the adaptive subtraction.

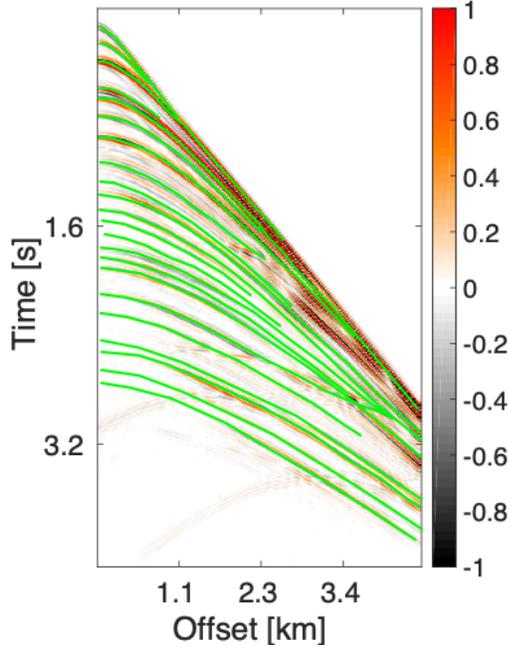


Figure 5.5: Primaries and the Green's function sketched by the green lines

5.3.2 Areal source and its adaption into linearized modeling

The above data-driven methods, i.e. SRME and EPSI that predict primaries with the least energy or sparsest Green's function, both need matrix-by-matrix multiplications to implement the multi-dimensional convolution between the areal source and Green's function. Tu and Herrmann 2015a proposed that the SRME relation can be introduced as the areal source into the linearized Born modelling, which physically carries out the multi-dimensional convolutions avoid these expensive multiplications.

$$\begin{aligned}
 \mathcal{P}_i &\approx \nabla \mathcal{F}_i[\mathbf{m}_0, \delta \mathbf{m}; \mathcal{I}](\mathcal{Q}_i - \mathcal{P}_i) \\
 &= \nabla \mathcal{F}_i[\mathbf{m}_0, \delta \mathbf{m}; \mathcal{Q}_i - \mathcal{P}_i] \\
 &= \nabla \mathcal{F}_i[\mathbf{m}_0; \mathcal{Q}_i - \mathcal{P}_i] \delta \mathbf{m}.
 \end{aligned} \tag{5.2}$$

In this expression, $\nabla \mathcal{F}_i$ represents the vertical derivative of linearized Born modelling, which is linear with the model perturbation $\delta \mathbf{m}$ and the impulsive source array \mathcal{I} . When the background model \mathbf{m}_0 is kinetically correct, $\nabla \mathcal{F}_i$ is a good approximation of the Green's

function \mathcal{G}_i . By using fundamental properties of Green's function the third expression in Equation 5.2 replaces the expensive convolutions in the first expression by including the down-going wavefield as an areal source.

In the time domain, the relationship 5.2 can be formulated directly as

$$\mathbf{P} \approx \nabla \mathbf{F}[\mathbf{m}_0; \mathbf{Q} - \mathbf{P}] \delta \mathbf{m}, \quad (5.3)$$

where \mathbf{P} , \mathbf{Q} , $\nabla \mathbf{F}$ denote the corresponding operators in the time domain.

To generate realistic surface-related multiples in the water column, we introduce density variations ρ at the ocean bottom. Equation 5.3 now becomes

$$\begin{aligned} \mathbf{P} &\approx \nabla \mathbf{F}_{\mathbf{m}, \rho}[\mathbf{m}_0, \rho_0; \mathbf{Q} - \mathbf{P}] \begin{bmatrix} \delta \mathbf{m} \\ \delta \rho \end{bmatrix} \\ &\approx \nabla \mathbf{F}_{\mathbf{m}}[\mathbf{m}_0, \rho_0; \mathbf{Q} - \mathbf{P}] (\delta \mathbf{m} + \delta \mathbf{m}'). \end{aligned} \quad (5.4)$$

In this expression, $\nabla \mathbf{F}_{\mathbf{m}, \rho}[\mathbf{m}_0, \rho_0; \mathbf{Q} - \mathbf{P}]$ corresponds to linearized Born modelling with respect to perturbations in the velocity and density, respectively. The operator $\nabla \mathbf{F}_{\mathbf{m}}[\mathbf{m}_0, \rho_0; \mathbf{Q} - \mathbf{P}]$ corresponds to linearized Born modelling with respect to velocity changes only. Here $\delta \mathbf{m}'$ is the density-induced "velocity" perturbation at the Ocean bottom. By including this additional term, we are able to model realistic surface-related multiples in the water column without relying on unrealistic and numerical problems inducing velocity perturbations. Note, that we assume the density to be constant throughout the remainder of the model so the migrated amplitudes should be interpreted as impedance perturbations in cases where there are strong variations of the density in the subsurface.

5.4 Joint inversion of multiples and primaries based on LB

Tu 2015 demonstrated that sparsity-promoting LS-RTM leads to artifact-free high-resolution images. When using the linearized Bregman method with source subsampling, we obtain

these images by solving

$$\begin{aligned} \min_{\mathbf{x}} \lambda \|\mathbf{x}\|_1 + \frac{1}{2} \|\mathbf{x}\|_2^2 \\ \text{subject to } \sum_j \|\nabla \mathbf{F}_j(\mathbf{m}_0, \rho_0, \mathbf{Q}_j - \mathbf{P}_j) \mathbf{C}^T \mathbf{x} - \mathbf{P}_j\|_2 \leq \sigma, \end{aligned} \quad (5.5)$$

where \mathbf{x} represents the curvelet coefficient vectors for the velocity perturbation $\delta \mathbf{m} + \delta \mathbf{m}'$. $\|\cdot\|_1$ and $\|\cdot\|_2$ stand for ℓ_1 and ℓ_2 -norms, respectively. The sum runs over all shots, σ is the two-norm of the noise. We modify the Algorithm 7 that designed for inversion with only primaries into the Algorithm 9 to invert the curvelet coefficients with the total up-going data by injecting the areal source into line 4. The parameter's setting is following the same strategy discussed in chapter 4.

Algorithm 9 Linearized Bregman for SPLS-RTM with multiples

- 1: Initialize $\mathbf{x}_0 = \mathbf{0}$, $\mathbf{z}_0 = \mathbf{0}$, \mathbf{q} , λ_1 , $\text{batchsize} n'_s \ll n_s$
 - 2: **for** $k = 0, 1, \dots$ **do**
 - 3: Randomly choose shot subsets $\mathcal{I} \subset [1 \dots n_s]$, $|\mathcal{I}| = n'_s$
 - 4: $\mathbf{A}_k = \{\nabla \mathbf{F}_j(\mathbf{m}_0, \rho_0, \mathbf{Q}_j - \mathbf{P}_j) \mathbf{C}^T\}_{j \in \mathcal{I}}$
 - 5: $\mathbf{b}_k = \{\mathbf{P}_j\}_{j \in \mathcal{I}}$
 - 6: $t_k = \|\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k\|_2^2 / \|\mathbf{A}_k^T (\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k)\|_2^2$
 - 7: $\mathbf{z}_{k+1} = \mathbf{z}_k - t_k \mathbf{A}_k^T \mathcal{P}_\sigma(\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k)$
 - 8: $\mathbf{x}_{k+1} = S_{\lambda_1}(\mathbf{z}_{k+1})$
 - 9: **end for**
 - 10: note: $S_{\lambda_1}(\mathbf{z}_{k+1}) = \text{sign}(\mathbf{z}_{k+1}) \max\{0, \|\mathbf{z}_{k+1}\| - \lambda_1\}$
 - 11: $\mathcal{P}_\sigma(\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k) = \max\{0, 1 - \frac{\sigma}{\|\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k\|}\} \cdot (\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k)$
-

5.5 Numerical experiments

To test the performance of this method, we conduct the experiments based on a shallow water model modified from Sigsbee2A model shown in Figure 5.10, which is discretized

with grid size of 5 m, totally 4.15 km deep and 4.48 km wide. The water layer is roughly 100 m deep. The background velocity m_0 is smoothed from the true model and is kinematically correct. The density model is converted from velocity background model by the Gardner relation in the sedimental layers, and the value for water layer is constant 1k/ml. So in the true density model there is only sharp perturbation near the ocean bottom. We use a Ricker wavelet centered at 15 Hz as source wavelet and record the data for 4 s. The 261 shots and receivers are spread over the top of the model with 15 m interval at the depth of 20 m. In the early state, We first test the performance of jointly inversion with areal source injection to a designed linearized data set via our time-harmonic domain modeling kernel. Then we move to the time domain, and test the performance for a nonlinear data set via Devito. For both data sets we totally conduct one data pass in inversion. For the time-domain inversion, we also use the preconditioners mentioned in chapter 4 to accelerate the inversion. The sparsity-promoting inversion framework with areal source injection helps to clean up these cross-ordered artifacts as indicated by the comparison between the SPLS-RTM image for only primaries and the SPLS-RTM image for the total deghosted up-going wavefield with areal source.

5.5.1 Test on linear data set in time-harmonic domain

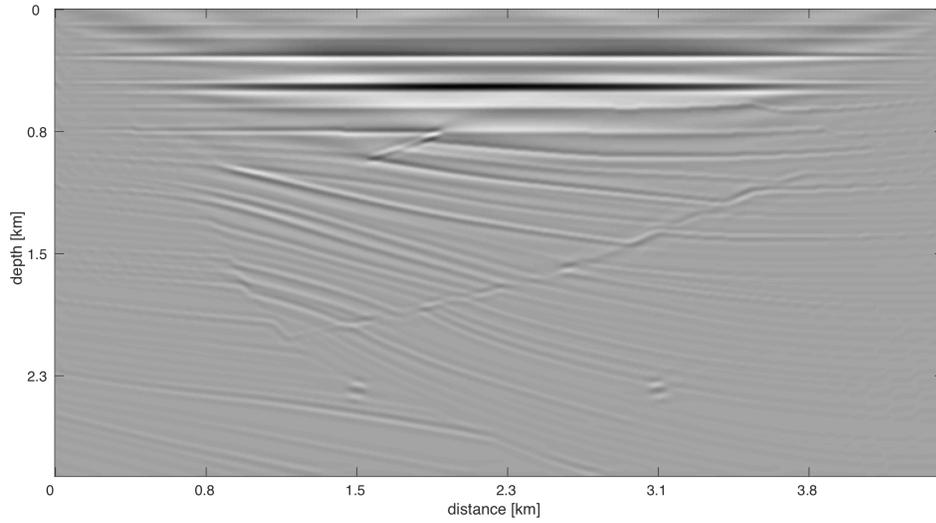
During the test in time-harmonic domain, the multiples in the linear data set is generated by injecting the up-going wavefield of the synthetic primaries into the linearized Born modeling respective to only velocity perturbation with dipole source setting. And the total up-going data is the summation of the synthetic primaries and multiples. Due to the cross-correlation between the events in different orders, the RTM for the total deghosted up-going wavefield with areal source involves more artificial reflectors indicated by yellow arrows in Figure 5.6b compared to the RTM for only primaries in Figure 5.6a. The sparsity-promoting inversion framework with areal source injection helps to clean up these cross-ordered phantoms as indicated by the comparison between the SPLS-RTM for only

primaries (Figure 5.7a) and the SPLS-RTM (Figure 5.7b) for the total deghosted up-going wavefield. Also areal source injected in the inversion helps to increase the illumination as the secondary sources, i.e. the area where with less illumination (Figure 5.9a) due to the lack of source (Figure 5.7) will have better image with areal source injection (Figure 5.9b).

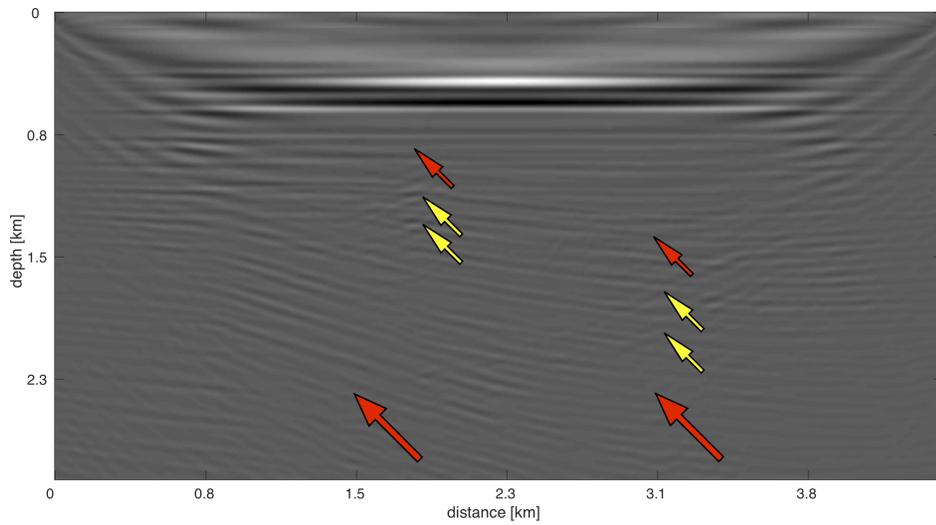
5.5.2 Test on nonlinear data set in the time domain

To test the performance of this proposed framework with filed data in time domain, we generate the nonlinear data with iWave (Symes 2013) by subtracting the data that generated by free-surface with the background models (density model and velocity model as in Figure 5.10) from the data that generated by free-surface with the true models. Then the nonlinear data is decomposed and extrapolated wapenaar1998reciprocity into the up- and down-going wavefields at the surface (one shot gather is shown in Figure 5.11). In SPLS-RTM here, we re-randomize roughly 5% shots of the whole data set during each iteration, and totally conduct one data pass.

The image we get by inverting the total up-going data directly with absorbing surface and dipole sources has interference from the phantoms generated by the cross-correlation between reflections in different orders. Here the phantoms from the cross-correlation between different ordered multiples of the shallowest layer interface are even stronger than those from the primaries of deeper layer interface and the responding first order multiples. The latter ones are recognized as the duplicated phantoms just below the correct reflectors as in Figure 5.12a. And the removed artifacts by injecting the areal sources in the inversion (Figure 5.12c) are mainly the latter phantoms. The former ones are very strong and some of them are still there as in Figure 5.12b around the first interfaces—i.e. the ocean bottom. And the remaining phantoms near the ocean bottom are due to the fact that we invert the density perturbation into a velocity perturbation, which need to be improved in the future. The improvements by areal source injection into the inversion framework are obvious in Figure 5.12c.

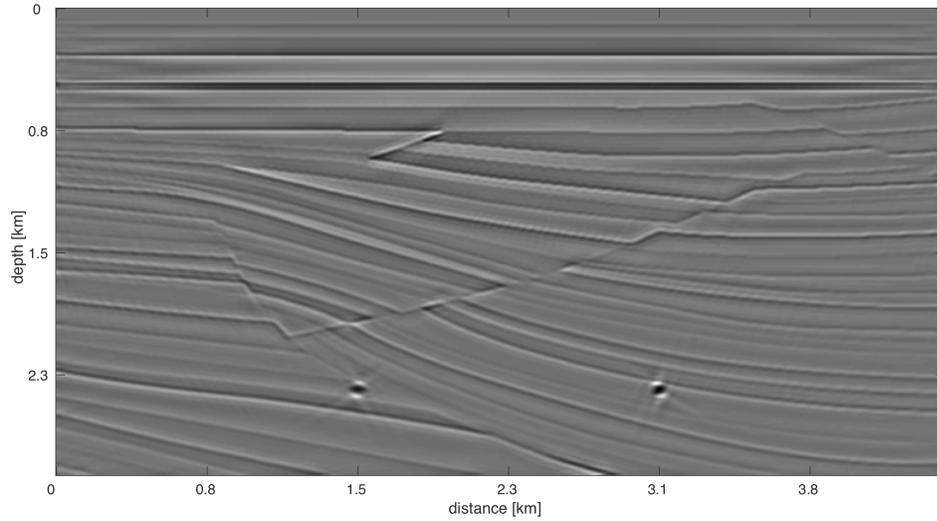


(a) RTM for primaries

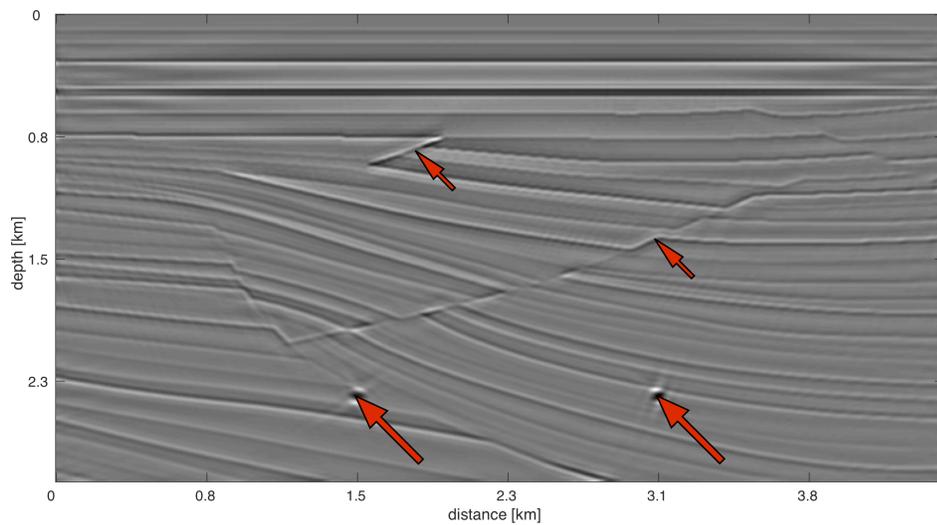


(b) RTM for the total deghosted up-going wavefield data

Figure 5.6: RTM images for linear data sets. (a) The RTM image for only primaries; (b) The RTM image for the total deghosted up-going wavefield with areal source.

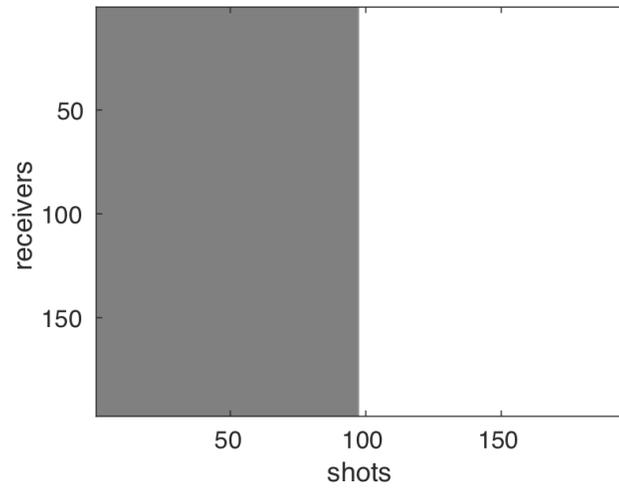


(a) LS-RTM for primaries

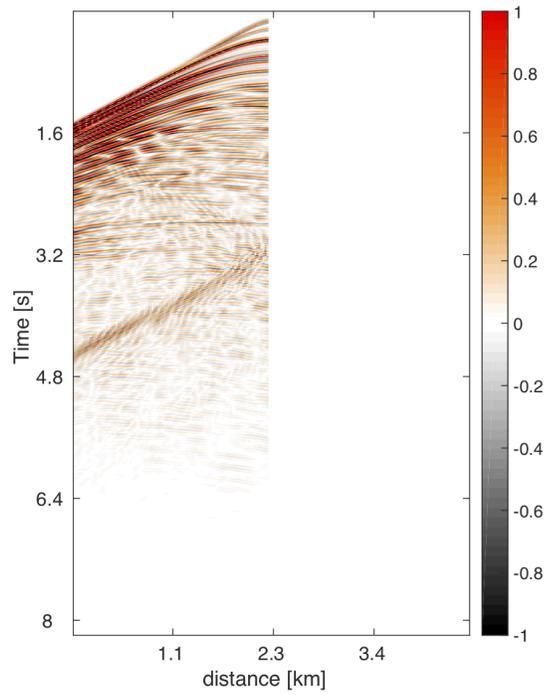


(b) SPLS-RTM for the total deghosted up-going wavefield data with areal source

Figure 5.7: SPLS-RTM images for linear data sets. (a) is the SPLS-RTM image for only primaries; (b) is the SPLS-RTM image for the total deghosted up-going wavefield with areal source.

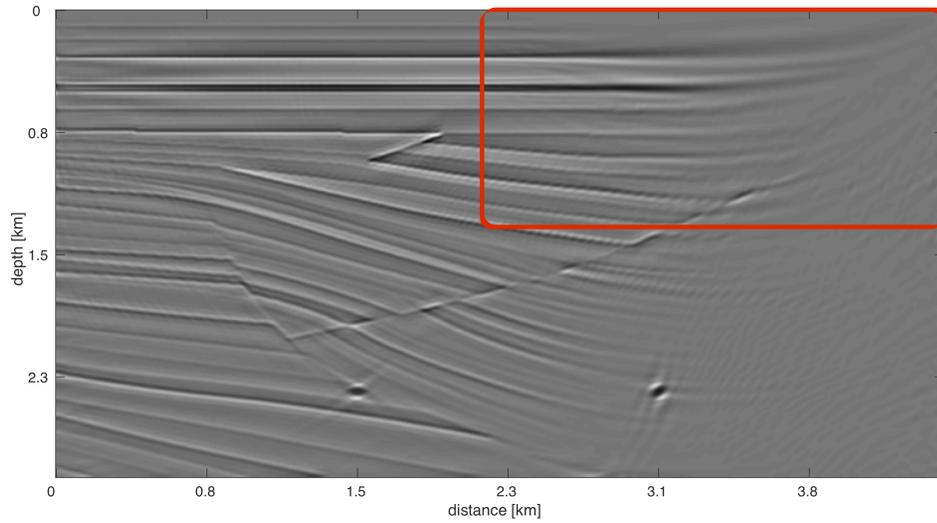


(a) Mask applied on acquisition geometry

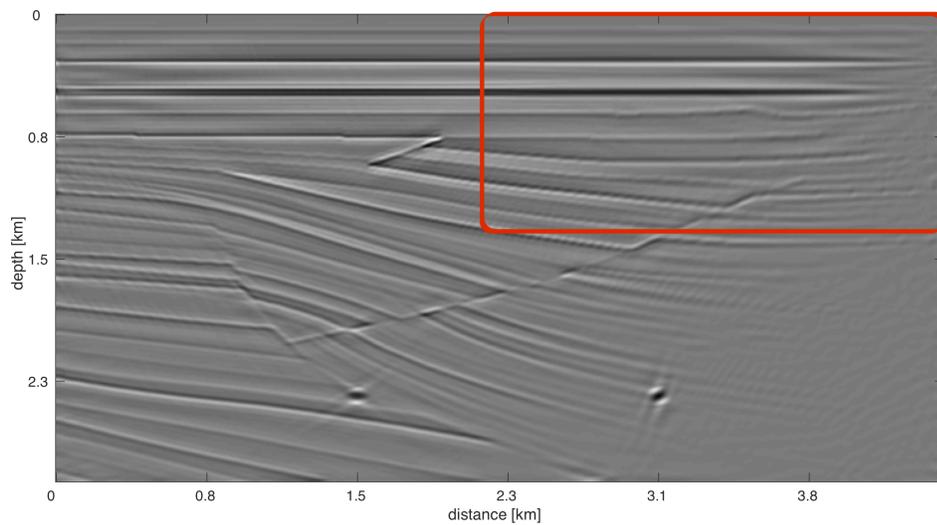


(b) One common-receiver gather with mask

Figure 5.8: The mask applied on the acquisition geometry and the masked data.



(a) LS-RTM for the masked primaries



(b) SMLS-RTM for the masked total deghosted up-going wavefield data with areal source

Figure 5.9: SMLS-RTM images for the masked linear data sets. (a) is the SMLS-RTM image for the masked primaries; (b) is the SMLS-RTM image for the masked total deghosted up-going wavefield with areal source.

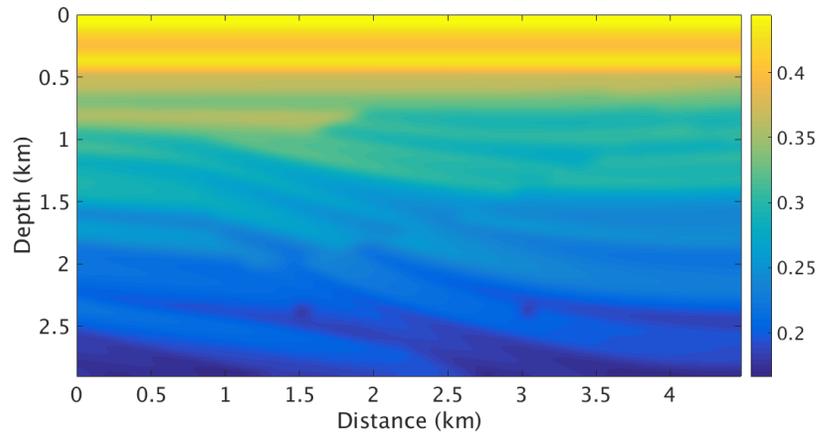


Figure 5.10: Slowness square background model modified from Sigsbee2A.

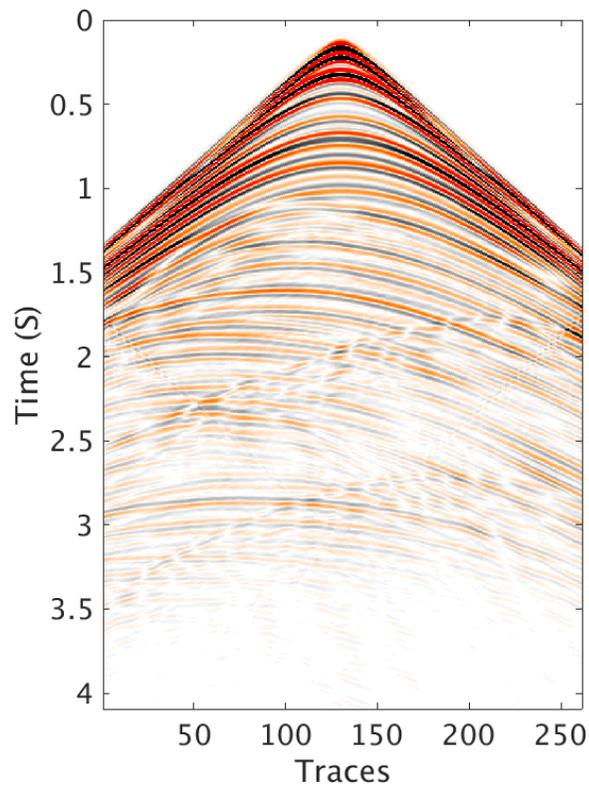
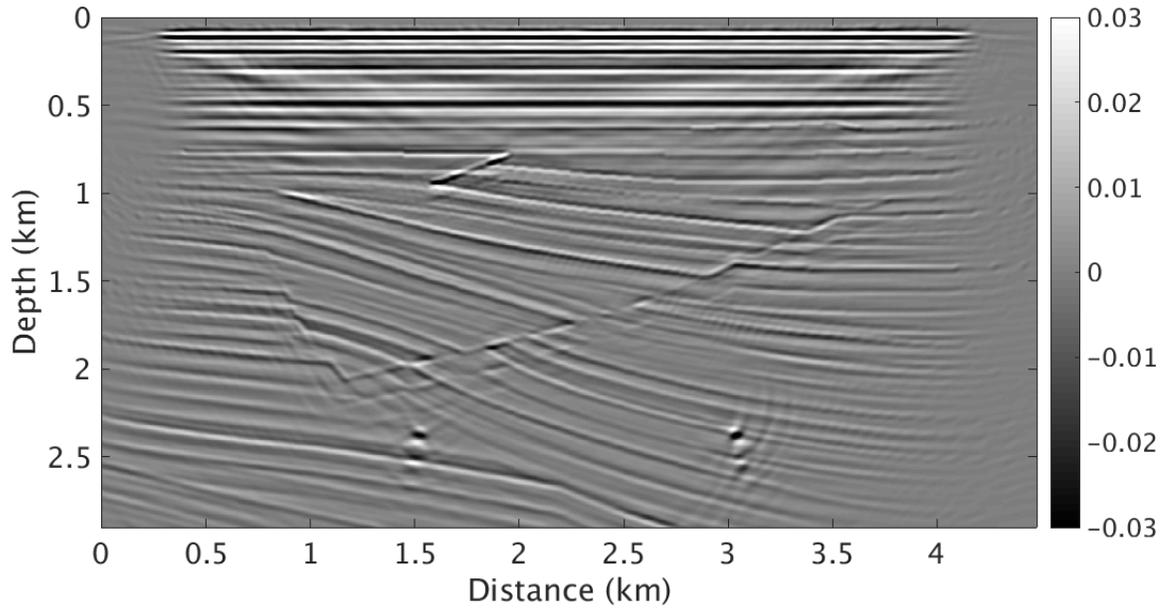
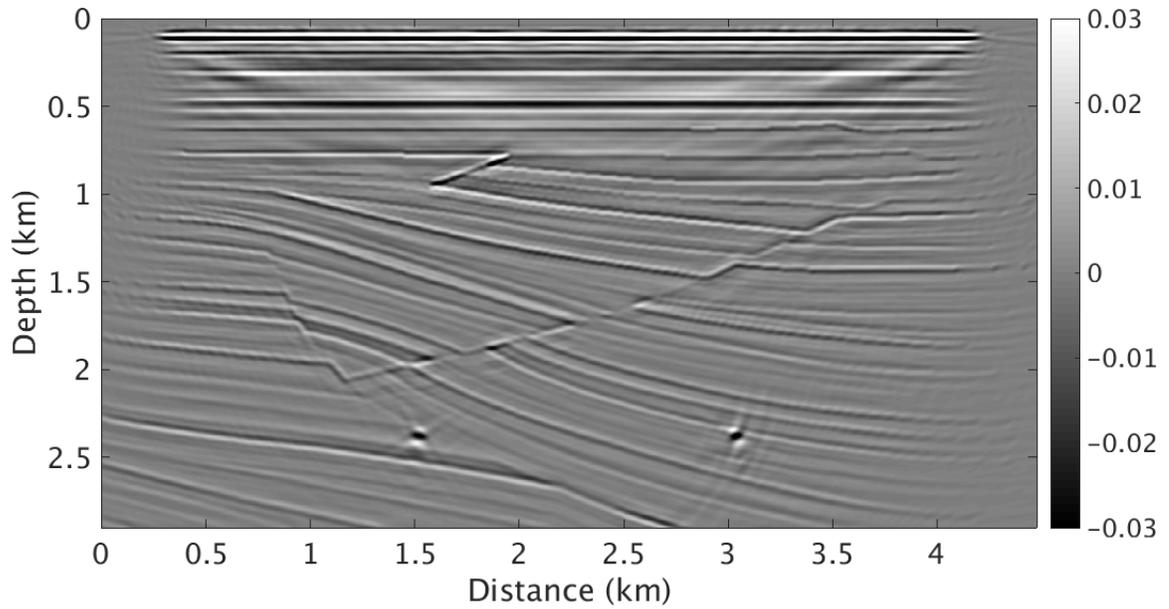


Figure 5.11: One shot record of areal source.

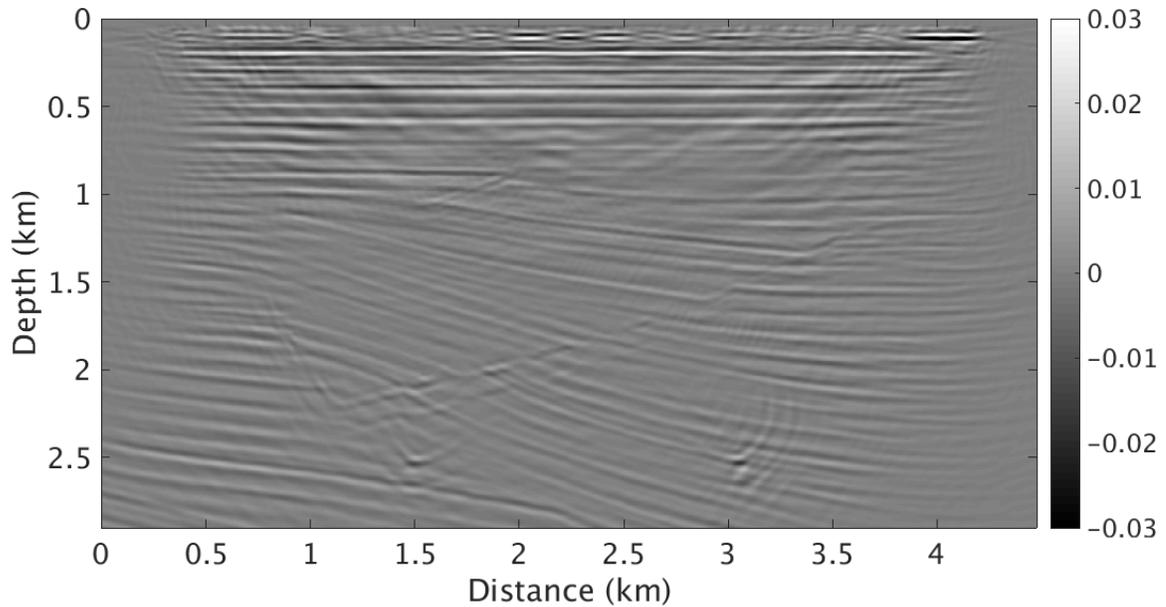
subfigure0



(a) Inverted image with total up-going data without areal source injection



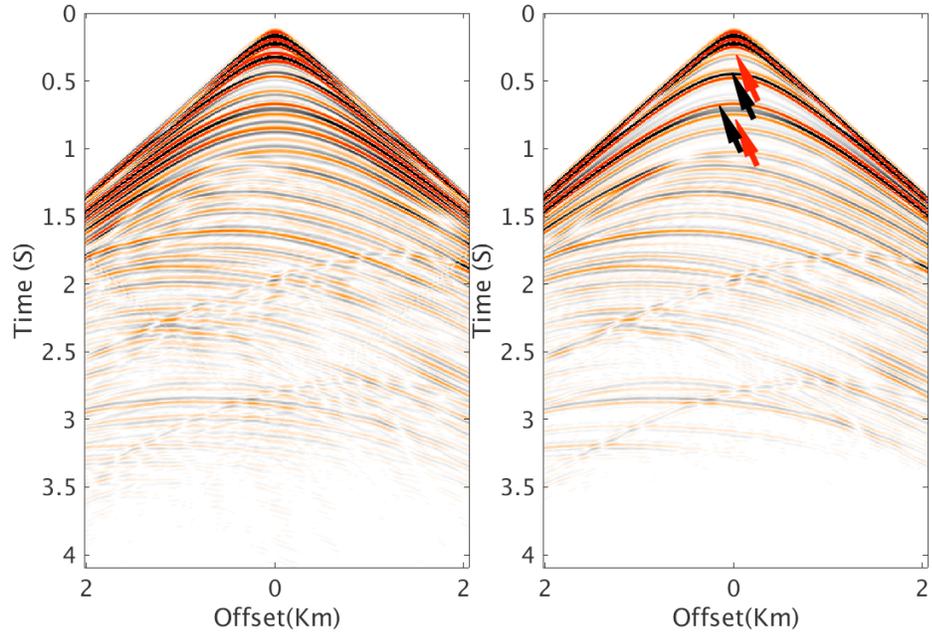
(b) Inverted image with total up-going data with areal source injection



(c) Difference between (a) and (b)

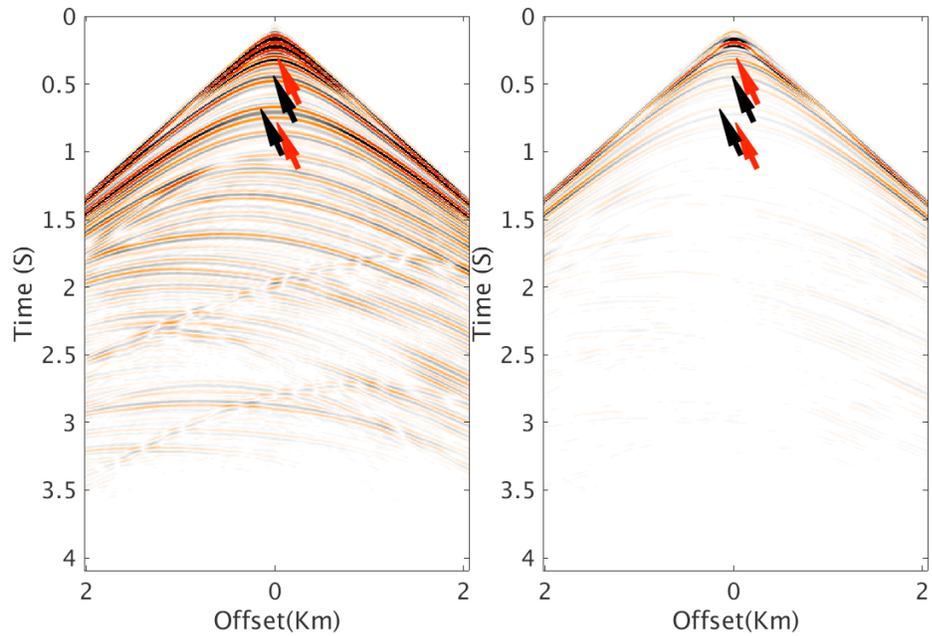
Figure 5.12: (a) and (b) Images by the total up-going data without and with areal source injection, respectively. (c) Differences between (a) and (b).

By comparing the synthetic primaries generated by Born modeling with dipole source and the recovered model perturbations $\delta\mathbf{m}$ with the ideal primaries generated by the true $\delta\mathbf{m}$, it is clear that most of the first order multiples from the deeper part of the model are removed or alleviated, but there are still leftover residuals from multiples recovered primaries. These differences can also be observed between the recovered multiples and the ideal multiples. Most of the leftover energies are from the higher order multiples of the primaries of the ocean bottom.



(a) Deghosted total up-going data

(b) Synthetic primaries



(c) Recovered primaries

(d) Difference between the Synthetic and recovered primaries

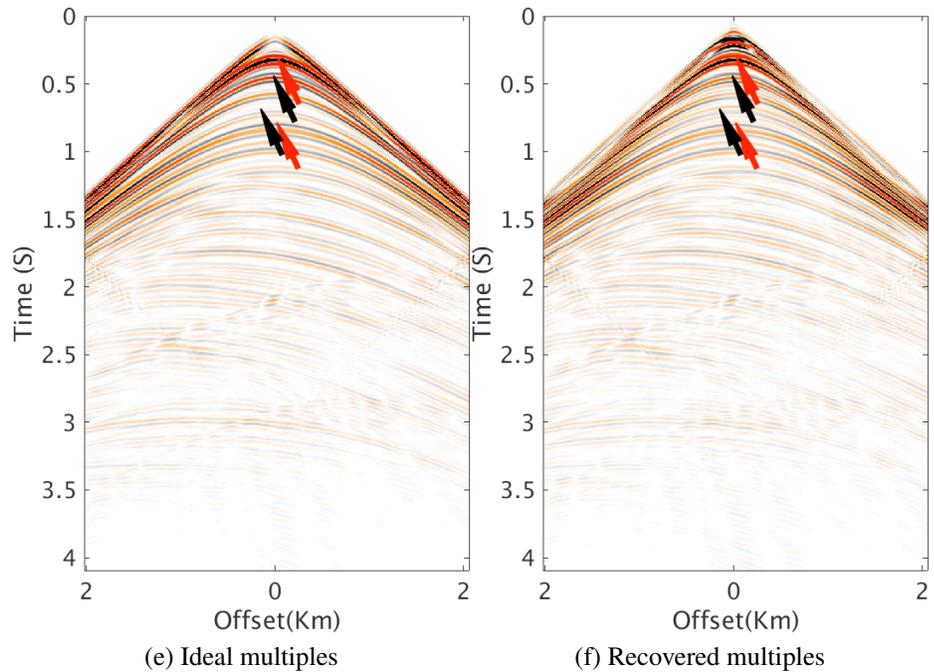


Figure 5.13: Comparison between shot gathers. (a) The deghosted total up-going data, (b) the ideal primaries with absorbing surface and dipole sources mirrored about the surface. (c) The recovered primaries by the inverted δm with absorbing surface and dipole sources mirrored about the surface. (d) The differences between (b) and (c). The red and block arrows indicate that the reflections in the recovered primaries are weaker or stronger than those in the ideal primaries, respectively. (e) The ideal multiples that got by subtracting the synthetic primaries with dipole sources from the deghosted total up-going data. (f) The recovered multiples that generated by areal source.

5.6 Discussion

In this chapter we extend our sparsity-promoting LS-RTM framework in chapter 4 to the jointly inversion of primaries and multiples for shallow water scenario. When the ocean bottom is shallow, the strong surface-related multiples would interrupt the primaries, which will fail the traditional SRME due to the energy leakage in adaptive subtraction, resulting in the distortion in the predicted primaries, so the later imaging. We jointly inverse primaries and multiples by injecting the areal source into the Born modeling. In order to get relief from the strong velocity perturbation at the ocean bottom to generate the strong multiples, we introduce density which is converted from the velocity for the sedimental layer by Gardener relation. In the early test with the designed linear data set based on part of the Sigsbee2A model in time-harmonic domain, the image we invert for the total up-going data with areal source injection has the phantoms cleaned. In the test on the nonlinear data set in the time domain, most of the phantoms from the cross-correlation between different orders of reflections are cleaned. Since the primaries of the ocean bottom is generated by the strong density perturbation but we are inverting via only Born modeling with respect to velocity, there is some phase errors in the estimated perturbation at the ocean bottom, and this will accumulate for the inverted reflectors from higher order multiples. We need to deal with this problem further in the future, either by developing Born modeling with respect to density term or taking advantage of other optimization methods which avoid the overfitting of the total up-going data to the velocity perturbation.

CHAPTER 6

REMOVING DENSITY EFFECTS IN LEAST SQUARE REVERSE TIME MIGRATION WITH A LOW-RANK MATCHED FILTER

6.1 Summary

Least-squares reverse-time migration faces difficulties when it inverts the data containing strong components related to density variation with velocity-only Born modeling operator. The strong density perturbation will be inverted as strong dummy velocity perturbations, which influence the amplitudes and phase of the velocity perturbations in the inverted model. The traditional method is to invert the additional density variations by developing Born operator with respect to both density and velocity or modify the image condition. In this work, we develop a matched-filter based LS-RTM for velocity-only Born modeling operator, which removes the artifacts in the imaging created by the strong density variation. This method doesn't call for extra work of finite difference stencil and is more general. In the experiment part, we use a complex discontinuous layered medium with strong density variations at the ocean bottom, and show the efficacy of the propose formulation.

6.2 Introduction

Least-squares reverse-time migration (LS-RTM, (Guitton, Kaelin, and Biondi 2006; Dai, Fowler, and Schuster 2012; Plessix and Mulder 2002)) tries to fit observed reflection data in a least-squares sense to overcome RTM's shortcomings in producing high resolution and high-fidelity amplitudes. In other words, LS-RTM attempts to invert the linearized Born modeling operator iteratively whereas RTM directly treats the adjoint of that operator as its inverse. So far, LS-RTM has demonstrated an ability to produce high-resolution images in combination with an efficient computational framework (Herrmann, Tu, and Esser 2015a;

Yang, Witte, Fang, and Herrmann 2016), overcoming drawbacks of overfitting artifacts (Herrmann and Li 2012) caused by minimizing the ℓ_2 -norm.

In addition to the developments listed above, people working on LS-RTM made lots of progress on incorporating multi-parameters (elastic parameters (Duan, Sava, and Guitton 2016), visco-acoustic parameters (Dutta and Schuster 2014) and so on) for complex geological structures. The corresponding Born modeling operators are linearized with respect to these elastic parameters, which allow us to mimic the elastic wave-propagation effects during the inversion. While important progress has been made handling elastic effects—e.g. by grouping subsets of elastic parameters that give rise to different radiation patterns (Operto, Gholami, Prioux, Ribodetti, Brossier, Metivier, and Virieux 2013)—working with multiple elastic parameters remains challenging.

Among the different parameters that rule the leading order behaviour of wave propagation, we count velocity and density as the most important pair (Beylkin and Burridge 1987). The products of these two, i.e., the seismic impedance, determines the amplitudes of the seismic waves to leading order. Perturbations in density generate reflection events even for a constant velocity model. This means that if we invert data generated by strong density perturbations with a Born modeling that accounts for velocity changes only we can expect strong artifacts degrading the quality of migrated images (Przebindowska, Kurzman, Köhn, and Bohlen 2012; Plessix, Milcik, Rynja, Stopin, Matson, and Abri 2013). There are two main reasons for these artifacts: *(i)* the wavefields scattered by the velocity and density parameters exhibit similar behaviours for some scattering angles (Operto, Gholami, Prioux, Ribodetti, Brossier, Metivier, and Virieux 2013), and *(ii)* if we only invert for velocity perturbations without incorporating the true density perturbations in Born modeling operator then LS-RTM will try to fit the amplitudes and phase of the observed seismic data in terms of velocities only (Bai and Yingst 2014). This can lead to dummy reflection events in the LS-RTM along with incorrect amplitudes and phase distortions of the true reflectivity yielded by the velocity perturbations.

In this work, we propose to use a matched-filter approach to remove artifacts caused by strong unmodeled density perturbations in the context of imaging with surface-related multiples. Specifically, we are interested in imaging based on linearized inversions that derive from velocity-only acoustic Born modeling that handles strong water-bottom multiples generated by strong density changes at the ocean bottom. We find that the proposed matched-filter, when organized as a matrix, exhibits low-rank structure. This is due to the fact that the matched-filter tries to approximate the difference between radiation patterns of velocity and a (strong ocean bottom) density contrast, which varies smoothly with offset. Inspired by this observation, we propose to simultaneously estimate velocity perturbations and a low-rank matched-filter, which maps nonlinear (observed) data that contains components related to both density and velocity perturbations into "linearized" data close to data generated by Born modeling for perturbations in velocity only. The proposed method does not require the explicit Born operator for density but does need terms that depend on a smoothly varying background density.

The chapter is organized as follows. First, we form the objective function for our extended LS-RTM with a low-rank matrix constraint on the matched filter and conclude by describing a computationally efficient algorithm. Next, we evaluate the performance of the proposed approach on a quasi-layered model with faults where the density varies strongly at the ocean bottom. Finally, we show the benefits of inverting for the matched filter to correctly image both the amplitude and phase of the reflectivity for imaging problems that contain a strong density contrast at the ocean bottom.

6.3 Methodology

We start with a brief overview of LS-RTM, and then propose a matched-filter based formulation to handle strong density-related effects in imaging. As we mentioned before, LS-RTM attempts to minimize the ℓ_2 norm of the data residual between the observed and

synthetic data by solving the following (unconstrained) optimization problem:

$$\min_{\mathbf{x}} f(\mathbf{x}) = \sum_{i=1}^{n_f} \|\mathbf{B}_i - \nabla \mathbf{F}_i(\mathbf{m}_0)(\mathbf{x})\|_F, \quad (6.1)$$

where $\nabla \mathbf{F}_i(\mathbf{m}_0)$ represents the monochromatic Jacobian with respect to velocity for all shots and followed by a matriciation putting monochromatic shots in its columns. The vector \mathbf{x} stands for the unknown velocity perturbations. Finally, the matrix \mathbf{B}_i is the i^{th} frequency slice of the observed (nonlinear) data in the S-R domain (source-receiver domain). In this work, we think of non-linear data as the difference between the response of the "true earth"—i.e., a "hard" model for velocities and densities and the response of a "smoothed background earth" where both velocity and density vary smoothly. The symbol $\|\cdot\|_F$ denotes the Frobenius norm. The above equation entails a "velocity only" linearization, which is accurate in the absence of strong density variations and a good background velocity model with respect to which the linearization is carried out. Under those conditions, equation 6.1 can produce good quality migrated images, which can then be used to perform reservoir characterization.

However, if the observed seismic data contains strong density effects generated by a strong ocean bottom, then the linearization undergridding Equation 6.1 is no longer valid and as a consequence this may lead to a degradation in quality of migrated images (Przebindowska, Kurzmann, Köhn, and Bohlen 2012; Plessix, Milcik, Rynja, Stopin, Matson, and Abri 2013). One way to address this issue is to include density into Equation 6.1 and re-linearize the Born modeling operator with respect to both velocity and density. While this approach is certainly a viable option, it is challenging and perhaps excessive to form a Born modeling operator that includes density, especially because it is well known that simultaneously inverting for both velocity and density is difficult because the two parameters have similar radiation patterns for certain scattering angles (Operto, Gholami, Prioux, Ribodetti, Brossier, Metivier, and Virieux 2013). As a result, LS-RTM runs the risk to map the perturbations in density to the velocity, which results in cross-talk in the imaging (Bai

and Yingst 2014).

To address these issues, we propose to include a matched filter, which allows us to compensate for certain leading order density effects while inverting for the velocity perturbations only. Under the assumption that we can find such a matched filter \mathbf{M} , we modify Equation 6.1 as follows:

$$\min_{\mathbf{x}, \mathbf{M}_i} f(\mathbf{x}, \mathbf{M}_i) = \sum_{i=1}^{n_f} \|\mathbf{B}_i \mathbf{M}_i - \nabla \mathbf{F}_i(\mathbf{m}_0) \mathbf{x}\|_F, \quad (6.2)$$

where \mathbf{M}_i is the matched filter matrix for i^{th} frequency. As in our earlier work involving on-the-fly source estimation (Tu, Aravkin, Leeuwen, and Herrmann 2013; Yang, Witte, Fang, and Herrmann 2016), we use variable projections to solve for \mathbf{M}_i while minimizing the above objective with respect to the velocity perturbations collected in the vector \mathbf{x} . However, contrary to finding a single time signature for the wavelet, the above matched filter involves for each frequency a full wavefield opening the risk of overfitting. To counter this problem, we control the rank k_i for each frequency. We motivate this choice by the fact that the difference in radiation patterns of velocity and the strong density contrasts at the ocean bottom vary smoothly over offset and we use this to stabilize the inversion. We now solve

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{M}_i} f(\mathbf{x}, \mathbf{M}_i) &= \sum_{i=1}^{n_f} \|\mathbf{B}_i \mathbf{M}_i - \nabla \mathbf{F}_i(\mathbf{m}_0) \mathbf{x}\|_F, \\ &\text{s.t. rank}(\mathbf{M}_i) = k_i, \end{aligned} \quad (6.3)$$

For simplicity, we will now focus on solving the above problem for one single frequency and drop the subscript i accordingly and implicitly sum over frequencies, i.e., $\sum_{i=1}^{n_f}$ for the remainder of this section. Since rank-minimization problems are NP hard, we use its convex relaxation instead, i.e., we replace the rank constraint by a nuclear-norm constraint (Recht, Fazel, and Parrilo 2010; Aravkin, Kumar, Mansour, Recht, and Herrmann 2014). Then we reformulate the optimization problem into the following format by introducing

one parameter τ to balance the Frobenius norm and the nuclear norm

$$\min_{\mathbf{x}, \mathbf{M}_i} f(\mathbf{x}, \mathbf{M}_i) = \sum_{i=1}^{n_f} \|\mathbf{B}_i \mathbf{M}_i - \nabla \mathbf{F}_i(\mathbf{m}_0) \mathbf{x}\|_F + \tau \|\mathbf{M}_i\|_* \quad (6.4)$$

As we mentioned before, we solve the above problem 6.4 with variable projections (Golub and Pereyra 2003; Tu, Aravkin, Leeuwen, Lin, and Herrmann 2016; Yang, Witte, Fang, and Herrmann 2016). This involves computing gradient steps with respect to \mathbf{x} that minimize the $\|\cdot\|_F$ norm —i.e.,

$$\mathbf{x}_{k+1} = \mathbf{x}_k + s \nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{M})|_{\mathbf{x}=\mathbf{x}_k, \mathbf{M}=\mathbf{M}_k}, \quad (6.5)$$

where s is the step size. As prescribed by variable projection, we solve for \mathbf{M} by minimizing 6.4 for \mathbf{x}_k fixed. The workflow for solving this optimization problem is summarized in Algorithm 10

Algorithm 10 Workflow for match-filtered LS-RTM

- 1: Initialize $\mathbf{x}_0 = \mathbf{0}, \tau$
 - 2: for $k = 1 \cdots N$
 - 3: for $i = 1 \cdots n_f$
 - 4: $\mathbf{D}_{ki} = \nabla \mathbf{F}_i(\mathbf{m}_0) \mathbf{x}_k$
 - 5: Initialize $\mathbf{Y}_0 = \mathbf{0}, \mathbf{M}_{ki,0}$
 - 6: for $j = 1 \cdots J$
 - 7: $\mathbf{Y}_j = \mathbf{Y}_{j-1} + \delta_j (\mathbf{D}_{ki} - \mathbf{B}_i \mathbf{M}_{ki,j-1})$
 - 8: $\mathbf{M}_{ki,j} = \mathcal{D}_\tau (\mathbf{B}_i^\top \mathbf{Y}_j)$
 - 9: end
 - 10: end
 - 11: $\mathbf{g}_k = \sum_{i=1}^{n_f} \nabla \mathbf{F}_i^\top \text{Vec}(\mathbf{B}_i \mathbf{M}_{ki,J} - \mathbf{D}_{ki})$
 - 12: $\mathbf{x}_k = \mathbf{x}_{k-1} + s_k \mathbf{g}_k$
 - 13: end
 - 14: note: \mathcal{D}_τ is the singular value shrinkage operator, δ_j and s_k are step sizes needed in update
-

6.4 Numerical experiments

To test the performance of the proposed method, we conduct experiments on a quasi layered model with strong density contrast at the ocean bottom (Figure 6.1e). Both velocity and density models are 1km deep and 2km wide and the underlying grid is discretized to 10m. The background velocity model \mathbf{m}_0 and background density model ρ_0 are smooth version of the true velocity and density models respectively and kinematically correct, as shown in Figure 6.1. We use a Ricker wavelet centered at 10Hz as source wavelet and record the data for 4 seconds. The 100 shots and 100 receivers are spread over the model with 10m spacing. We conduct 20 iterations for the LS-RTM with ideal linearized data with respect to only velocity perturbation and 100 iterations to solve the optimization problem we proposed with the nonlinear data, and during each iteration we run 200 iterations to

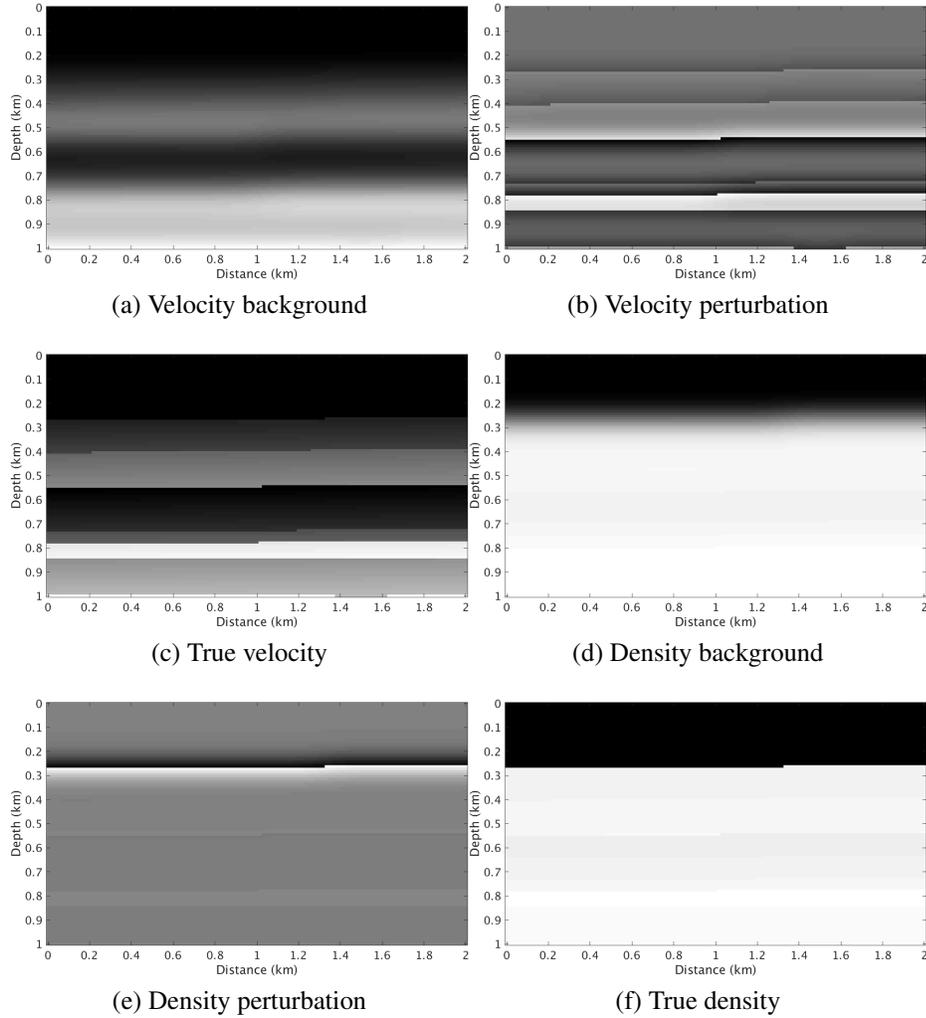


Figure 6.1: Quasi layered model with faults. (a, b, c) background velocity, the corresponding perturbation and the true velocity. (c, d, e) Same for density model.

solve the matched filter.

Given the true velocity and density models and their backgrounds, we generate "observed" nonlinear data (Figure 6.2a,b,c) by subtracting the response of the background models for varying velocity and density from the response yielded by the true velocity and density models. We compare this response with linear data (Figure 6.2d,e,f) obtained by applying the Born modeling operator with respect to only velocity. The monochromatic frequency slices of the data predicted by the recovered low-rank filters M in Figure 6.3a,b,c are shown in Figure 6.2g,h,i. It is clear from the figures that the estimated matched-filter

varies smoothly across sources, thus, exhibits low-rank structure. These results also validate our belief that the difference between radiation patterns of velocity and the strong density contrast exhibits smooth structure over offset.

We can observe from Figure 6.4 that in the last iteration of the out-loop of Algorithm 10, the ranks of the inverted matched-filters increase along iterations, leaving the Frobenius norm of the objective function decreasing (as shown in Figure 6.5), and finally arrive stable rank levels which are still very low compared to the full-rank 100.

Figure 6.6a shows the idealized LS-RTM results using the linearized data. We can see that for the ideal scenario, the layer interfaces are sharp and amplitudes are in the correct range. Also it is clear that the velocity perturbations at shallower interfaces especially at the ocean bottom are much weaker than those at the deeper interfaces. Next, we invert the nonlinear data without any matched-filter approach using the velocity-only Born modeling operator. It is evident from the inverted image (Figure 6.6b that the LS-RTM maps the strong density perturbations to the velocity perturbations, thus, creating the dummy strong reflectors at the ocean bottom. Moreover, the amplitudes and phase of the subsequent deeper reflectors are wrong. Finally, we use the proposed matched-filter approach to perform the LS-RTM (Figure 6.6c. Using the propose method, we are able to remove the effects of the strong density perturbations at both the shallow and deeper sections. Thus, the estimated matched-filter can handle the strong density perturbation related effects while inverting only the velocity perturbation, and can successfully remove the cross-talk created by density perturbations.

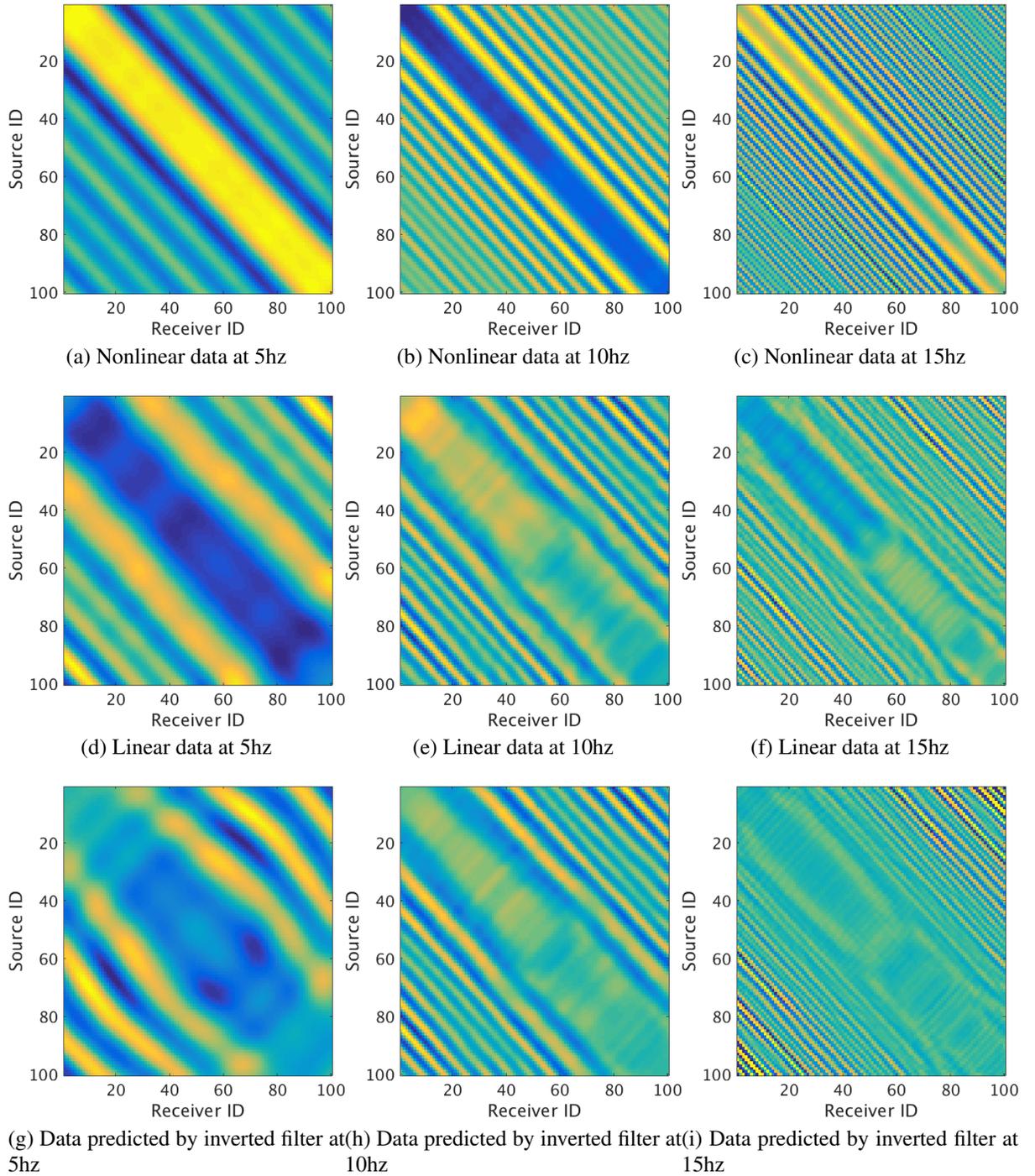


Figure 6.2: Data comparison. (a),(b) and (c) are the nonlinear data slices at 5, 12, 25Hz, respectively. (d),(e) and (f) are the synthetic linear data slices at 5, 12, 25Hz, respectively. (g),(h) and (i) are the predicted linear data slices with the inverted filters at 5, 12, 25Hz respectively.

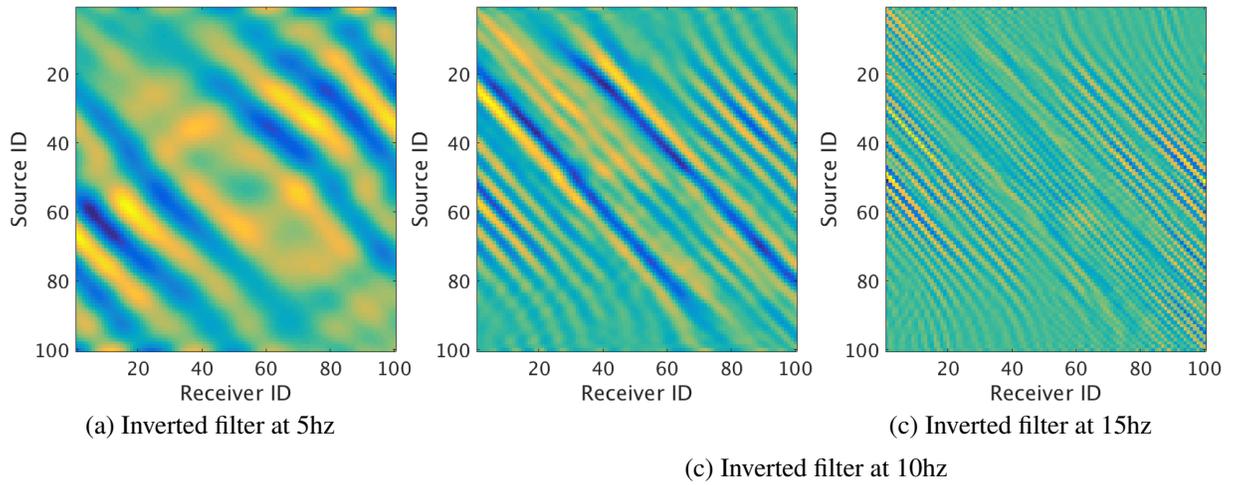


Figure 6.3: Inverted filters at 5, 12, 25Hz.

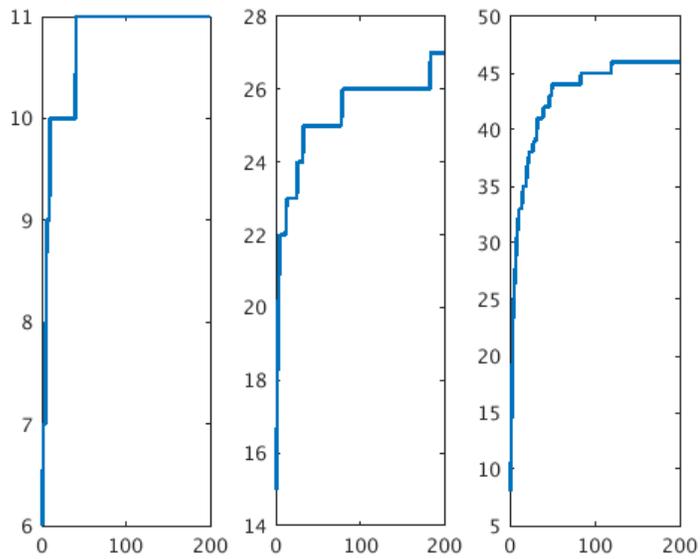


Figure 6.4: In the last iteration of out-loop, ranks of the matched-filters along the inner iterations at 5, 12, 25Hz.

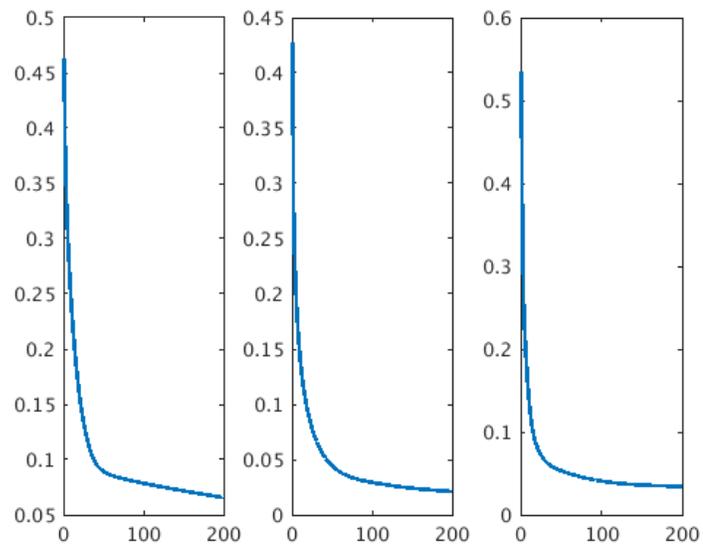
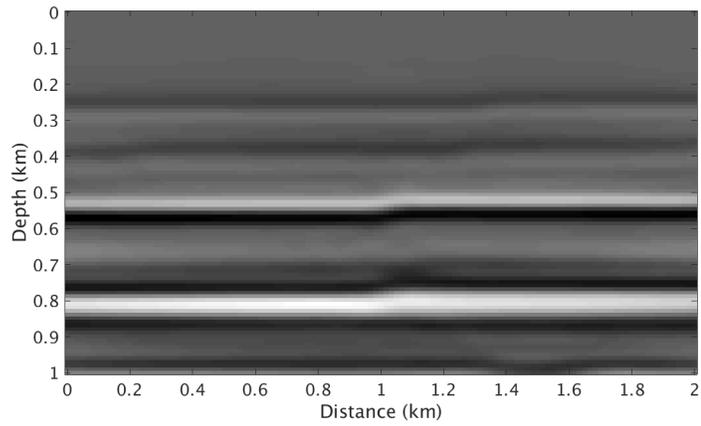
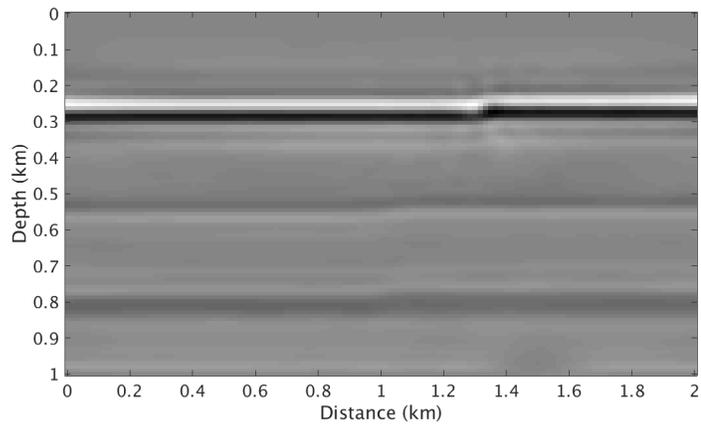


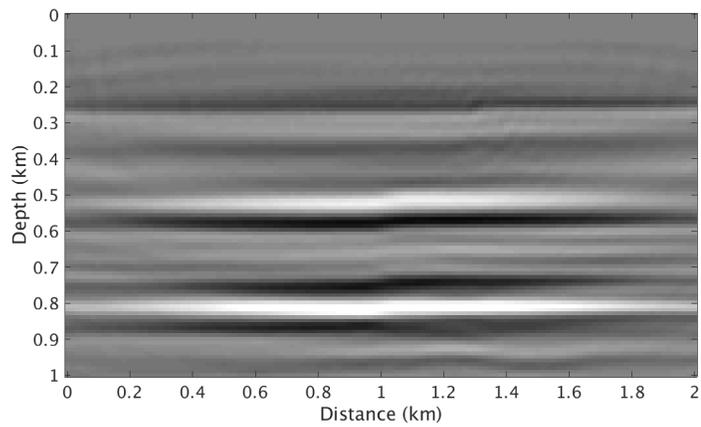
Figure 6.5: In the later iteration of out-loop, the Frobenius residuals of the objective function along inner iterations at 5, 12, 25Hz.



(a) Image inverted from linear data



(b) Image inverted from nonlinear data



(c) Image inverted from nonlinear data using matched-filter approach

Figure 6.6: LS-RTM results using velocity-only Born modeling operator. (a) Idealized linearized data, nonlinear data (b) without, and (c) with matched-filter approach.

6.5 Conclusion

In this chapter, we propose a matched-filter based least-squares reverse time migration formulation to remove the strong density variation related components in the observed data. In contrast to other methods which invert density as one additional output or reform image condition, our method doesn't require the work related to finite difference. Our modified formulation inverts for the matched-filter and velocity perturbation simultaneously that matching the nonlinear observed data to the linear data with respect to velocity-only. In the experiment part, we used a discontinuous layered model with strong density variations at the ocean bottom to test our method. We showed that the proposed matched-filter approach can remove the artifacts from density perturbation and get artifacts-free inverted velocity perturbations. During each iteration of estimating the matched filter, SVD is involved, which would be expensive when there are amount of sources and receivers. We would consider to develop and compare other optimization method to improve the computational efficiency of our method. The further work is to incorporate the surface related multiples into the formulation since strong density variation can leads to strong surface related multiples, which can further enhance the resolution of the inverted images.

CHAPTER 7

CONCLUSION

In summary, this thesis has contributed a computationally efficient method for recovering the low-rank representations of the full subsurface extended image volumes. The method is based on a fast time-stepping propagator and provides a SVD-free approach to mapping the low-rank factors to other factors for velocity variation scenarios. This work has also addressed the topic of source estimation for time-domain sparsity-promoting least-squares reverse-time migration. In a discussion of the influence of converted artifacts from density variations at the ocean bottom, the thesis has extended the work of efficient sparsity-promoting inversion for only primaries to joint inversion of both primaries and multiples. Another contribution is the design of a low-rank filter that matches the density effect from the strong density variations at the ocean bottom in least-squares reverse time migration without using Born modeling with respect to density.

7.1 Low-rank recovery of subsurface extended image volumes based on power-schemed rSVD and mapping via the invariance relationship

In the first part of this thesis (Chapter 2 and 3), in order to get salable access to each element of the full subsurface image volumes (EIV), we proposed a feasible approach to recovering their low-rank representations via a randomized SVD method based on the probing technique in the time domain. To ensure that the randomized probing is feasible and numerically stable in the time domain, we combined the source term in the formulation of EIV with Gaussian random probing vectors to form a bandwidth limited source wavefield. By the utilization of randomness, we limited the wave-equation solves from $2n_s$ to $4n_p$, where $n_p \ll n_s$. In addition, the rSVD method with probing ingeniously reformulates the original costly multi-dimensional convolutions in EIV between the forward and backward

wavefields into relatively cheaper multi-dimensional convolutions between the data record whose size is much smaller than the wavefield and the other gathers in size of $n_s \times n_p \times n_t$, which is also much smaller than all n_s wavefields.

To promote our low-rank recovery method to a realistic problem in which the rank of monochromatic EIV increases along frequency, we combined the power iterations with the basic rSVD method to accelerate the decay of the singular values of EIVs by narrowing down the gap between the neighboring singular values. We compared different power schemes and the basic rSVD with different probing sizes and powers by checking the errors of the recovered singular values and the RTM (i.e., the diagonal elements extracted from the recovered factors of EIV) and concluded that for our seismic problem, the optimal choice was block Krylov iterations with a power of 1. Power iterations improve the recovered accuracy without increasing the probing size; unfortunately, they increase the cost, particularly in wave-equation solves proportional to a power of q . We also proposed several ways of extracting various gathers from the low-rank factors of EIV, including reverse time migration (RTM), common image point gathers (CIPs), common image gathers (CIGs), and geological dip-corrected CIGs without any extra wave-equations solves, and we conducted seismic numerical experiments to compare the recovered image gathers with the traditional gathers that were costly to obtain.

We also proposed factorization-free mapping from one pair of low-rank factors to another pairs via invariance relationship for velocity variation scenarios. Combined with the power scheme-based rSVD, which we used in the initial background model, when the background model was updated, the cost of wave-equations in the mapping process was proportional only to the probing size, namely $4n_p$. In the seismic numerical experiment, we demonstrated the effectiveness of mapping by comparing the image gathers extracted from the updated low-rank factors to those from the factors obtained with kinematically correct model with power iterations.

7.2 Time-domain sparsity promoting least-squares reverse time migration with source estimation

In the second part of this thesis (Chapter 4), we proposed a scalable time-domain approach to sparsity-promoting least-squares reverse time migration with on-the-fly source estimation suitable for industrial 3D imaging problems in principle. The approach leverages recently developed techniques from convex optimization and variable projection that significantly reduce costs and the need to provide an estimate for the source function. As a result, our approach is capable of generating high-fidelity true-amplitude images, including source estimates at a cost of roughly one to two migrations involving all data.

By means of carefully designed experiments in 2D, we were able to demonstrate that our method is capable of handling noisy data and complex imaging settings such as salt. We were able to image under salt, which is often plagued by low-frequency tomographic artifacts, by switching between initially applying the conventional imaging condition and then performing iterations that apply the inverse-scattering condition. In this way, we estimated the source function first while creating an artifact-free image with later iterations, when the imaging condition was switched while keeping the source function fixed.

7.3 Sparsity-promoting least-squares reverse time migration with multiples and removing density effects with a low-rank matched filter

In the third part of this thesis (Chapter 5), we extended our sparsity-promoting LS-RTM framework to the joint inversion of primaries and multiples in a shallow water scenario, where the SRME failed due to the energy leakage in the adaptive subtraction of the primaries prediction. Without any matrix-matrix multiplies, we introduced the multiples by injecting the areal source into the Born modeling, and cleaned the cross-talks between different orders of reflections by our sparsity-promoting inversion framework. In our test on a nonlinear dataset in the time domain, the primaries of the ocean bottom were generated

by the strong density perturbation but inverted via Born modeling only with respect to velocity. Hence, several phase errors in the estimated perturbations were introduced, which accumulated for the inverted reflectors from higher order multiples.

To deal with the converted velocity perturbations from the density variations, we proposed a matched filter-based least-squares reverse time migration formulation instead of inverting the density perturbation as one additional output. Our modified formulation inverts simultaneously for the velocity perturbation and the low-rank filter that matches the nonlinear observed data to the linear data component with respect to velocity only. We conducted preliminary tests based on a discontinuous layered model and demonstrated the capability of removing the artifacts from density perturbation and obtained the artifacts-free migrated image related to only velocity perturbation.

7.4 Current limitations and future work

Several limitations of the work presented in this thesis should be considered in the future and could be addressed by the following research topics. For one, in the low-rank recovery of EIV, we implemented and adapted the randomized probing technology with time-stepping propagator. With regard to the numerical stability, we combined the random probing vectors with the source term into one band-limited source. This reformulation assumed that all the shooting sources have the same wavelet signature and requires modification before it can be applied to real data in the future.

In the work of recovering EIV, we applied fast Fourier transform and its inverse to convert between time-domain tensors to their corresponding monochromatic matrices, which might involve large memory when extended to three dimensions. This limitation could be alleviated if we applied on-the-fly Fourier transform with time stepping (Witte, Louboutin, Luporini, Gorman, and Herrmann 2019), in which each time-stepping propagation involved only one frequency component in the source term, used less memory, but required more forward modeling to cover all the discretized frequencies. The usage of on-the-fly Fourier

transform with time-stepping operators will also allow us apply less probing vectors for lower frequencies, which can save wave-equation solves to some degree.

When the EIV is ill-conditioned, for example, when its singular values are dominated by some specific values derived from the components of the salt body, we could create clearer images by mitigating the ill-conditioning of the EIV. Some recent peer work shows the improvement of spacial differentiation-related preconditioners on removing low-wave number artifacts for only specific image gathers, e.g., CIGs. Since we can access any specific elements from the low-rank factors without any investment in wave-equation solves, we can easily design some preconditioners after the low-rank recovery to get better CIGs. Also we can design some preconditioners applied during the recovery. With the ill-conditioning alleviated, we could expect that with the energy on the tail of the singular values raised and the gap between neighboring singular values enlarged by the preconditioners, the power scheme we proposed would have more obvious improvement than basic rSVD in increasing the recovery accuracy.

In the work of removing the density effect by a matched low-rank filter, we applied least-squares reverse time migration with only primaries. However the SVD investigated in each iteration of estimating the matched low-rank filter could be potentially expensive with large numbers of source and receiver. We will consider about other optimization method to avoid this factorization in the future. We would also introduce the low-rank filter to the joint inversion of primaries and multiples to remove the ocean bottom related artifacts converted from density variations.

REFERENCES

- [1] R. Sheriff and L. Geldart, “Exploration seismology. volume 1: History, theory, and data acquisition,” 1983.
- [2] S. T. Vaage, *Method for acquiring and processing seismic survey data using ocean bottom cables and streamers*, US Patent 6,704,244, 2004.
- [3] E. Baysal, D. D. Kosloff, and J. W. Sherwood, “Reverse time migration,” *Geophysics*, vol. 48, no. 11, pp. 1514–1524, 1983.
- [4] W.-F. Chang and G. A. McMechan, “Reverse-time migration of offset vertical seismic profiling data using the excitation-time imaging condition,” *Geophysics*, vol. 51, no. 1, pp. 67–84, 1986.
- [5] J. E. Rickett and P. C. Sava, “Offset and angle-domain common image-point gathers for shot-profile migration,” *Geophysics*, vol. 67, no. 3, pp. 883–889, 2002.
- [6] T. van Leeuwen and F. J. Herrmann, “Wave-equation extended images: Computation and velocity continuation,” in *EAGE technical program*, EAGE, EAGE, 2012.
- [7] T. van Leeuwen, R. Kumar, and F. J. Herrmann, “Enabling affordable omnidirectional subsurface extended image volumes via probing,” *Geophysical Prospecting*, vol. 65, no. 2, pp. 385–406, 2017.
- [8] R. Kumar, M. Graff-Kray, T. van Leeuwen, and F. J. Herrmann, “Low-rank representation of extended image volumes—applications to imaging and velocity continuation,” in *SEG Technical Program Expanded Abstracts*, (SEG, Anaheim), Oct. 2018, pp. 4418–4422.
- [9] W. W. Symes, “Migration velocity analysis and waveform inversion,” *Geophysical prospecting*, vol. 56, no. 6, pp. 765–790, 2008.
- [10] C. C. Stolk, M. V. de Hoop, and W. W. Symes, “Kinematics of shot-geophone migration,” *GEOPHYSICS*, vol. 74, no. 6, WCA19–WCA34, 2009. eprint: <https://doi.org/10.1190/1.3256285>.
- [11] G. Schuster, *Least-squares cross-well migration: 63rd annual international meeting, seg, expanded abstracts, 110–113*, 1993.
- [12] T. Nemeth, C. Wu, and G. T. Schuster, “Least-squares migration of incomplete reflection data,” *Geophysics*, vol. 64, no. 1, pp. 208–221, 1999.

- [13] A. Guitton, B. Kaelin, B. Biondi, *et al.*, “Least-square attenuation of reverse time migration artifacts,” in *2006 SEG Annual Meeting*, Society of Exploration Geophysicists, 2006.
- [14] T. Keho and W. Beydoun, “Paraxial ray kirchhoff migration,” *Geophysics*, vol. 53, no. 12, pp. 1540–1546, 1988.
- [15] J. F. Claerbout, C. Green, and I. Green, “Imaging the earth’s interior,” 1985.
- [16] A. Berkhout, “Seismic inversion in terms of pre-stack migration and multiple elimination,” *Proceedings of the IEEE*, vol. 74, no. 3, pp. 415–427, 1986.
- [17] D. J. Verschuur, A. Berkhout, and C. Wapenaar, “Adaptive surface-related multiple elimination,” *Geophysics*, vol. 57, no. 9, pp. 1166–1177, 1992.
- [18] D. J. Verschuur, *Seismic multiple removal techniques: Past, present and future*. EAGE publications Netherlands, 2006.
- [19] A Guitton, A Valenciano, D Bevc, and J Claerbout, “Smoothing imaging condition for shot-profile migration: *Geophysics*, 72,” *S149–S154*, p. 54, 2007.
- [20] N. Whitmore, A. Valenciano, W. Sollner, and S. Lu, “Imaging of primaries and multiples using a dual-sensor towed streamer,” in *SEG Technical Program Expanded Abstracts 2010*, Society of Exploration Geophysicists, 2010, pp. 3187–3192.
- [21] S. Lu, N. Whitmore, A. Valenciano, and N. Chemingui, “Imaging of primaries and multiples with 3d seam synthetic,” in *SEG Technical Program Expanded Abstracts 2011*, Society of Exploration Geophysicists, 2011, pp. 3217–3221.
- [22] C. De Bruin, C. Wapenaar, and A. Berkhout, “Angle-dependent reflectivity by means of prestack migration,” *Geophysics*, vol. 55, no. 9, pp. 1223–1234, 1990.
- [23] P. Sava and S. Fomel, “Angle domain common image gathers by wavefield continuation methods,” *Geophysics*, vol. 68, no. 3, pp. 1065–1074, 2003. eprint: <http://library.seg.org/doi/pdf/10.1190/1.1581078>.
- [24] F. ten Kroode, “A wave-equation-based kirchhoff operator,” *Inverse Problems*, vol. 28, no. 11, p. 115013, 2012.
- [25] H. Köhl and M. D. Sacchi, “Least-squares wave-equation migration for avp/ava inversion,” *Geophysics*, vol. 68, no. 1, pp. 262–273, 2003.
- [26] F. Mahmoudian and G. F. Margrave, “A review of angle domain common image gathers,” *CREWES Research Report*, vol. 21, pp. 1–22, 2009.

- [27] R. Dafni and W. W. Symes, “Scattering and dip angle decomposition based on subsurface offset extended wave-equation migration,” *GEOPHYSICS*, vol. 81, no. 3, S119–S138, 2016. eprint: <https://doi.org/10.1190/geo2015-0528.1>.
- [28] R. Dafni and W. W. Symes, “Kinematic artifacts in the subsurface-offset extended image and their elimination by a dip-domain specularly filter,” *GEOPHYSICS*, vol. 81, no. 6, S477–S495, 2016. eprint: <https://doi.org/10.1190/geo2016-0115.1>.
- [29] T. Yang and P. Sava, “Image-domain wavefield tomography with extended common-image-point gathers,” *Geophysical Prospecting*, vol. 63, no. 5, pp. 1086–1096, 2015.
- [30] B. Biondi and W. W. Symes, “Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging,” *Geophysics*, vol. 69, no. 5, pp. 1283–1298, 2004.
- [31] W. W. Symes and J. J. Carazzone, “Velocity inversion by differential semblance optimization,” *Geophysics*, vol. (5), 56, no. 5, pp. 654–663, 1991.
- [32] P. Shen and W. W. Symes, “Automatic velocity analysis via shot profile migration,” *Geophysics*, vol. 73, no. 5, VE49–VE59, 2008.
- [33] N. Kukreja, M. Louboutin, F. Vieira, F. Luporini, M. Lange, and G. Gorman, “Devito: Automated fast finite difference computation,” in *Domain-Specific Languages and High-Level Frameworks for High Performance Computing (WOLFHPC), 2016 Sixth International Workshop on*, IEEE, 2016, pp. 11–19.
- [34] J. F. Claerbout, “Coarse grid calculations of waves in inhomogeneous media with application to delineation of complicated seismic structure,” *Geophysics*, vol. 35, no. 3, pp. 407–418, 1970. eprint: <http://geophysics.geoscienceworld.org/content/35/3/407.full.pdf+html>.
- [35] B. Biondi and W. W. Symes, “Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging,” *Geophysics*, vol. (5), 69, no. 5, p. 1283, 2004.
- [36] P. Sava and I. Vasconcelos, “Extended imaging conditions for wave-equation migration,” *Geophysical Prospecting*, vol. 59, no. 1, pp. 35–55, Jan. 2011.
- [37] R. Kumar, M. Graff-Kray, I. Vasconcelos, and F. J. Herrmann, “Target-oriented imaging using extended image volumes—a low-rank factorization approach,” *Geophysical Prospecting*, vol. 67, no. 5, pp. 1312–1328, 2019, (Geophysical Prospecting).

- [38] N. Halko, P.-G. Martinsson, and J. A. Tropp, “Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions,” *SIAM review*, vol. 53, no. 2, pp. 217–288, 2011.
- [39] C. Musco and C. Musco, “Randomized block krylov methods for stronger and faster approximate singular value decomposition,” in *Advances in Neural Information Processing Systems*, 2015, pp. 1396–1404.
- [40] C. Da Silva, Y. Zhang, R. Kumar, and F. J. Herrmann, “Applications of low-rank compressed seismic data to full waveform inversion and extended image volumes,” *Geophysics*, vol. 84, no. 3, pp. 1–71, 2019.
- [41] C. D. 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, Sep. 2015, (Linear Algebra and its Applications).
- [42] M. Yang, M. Graff, R. Kumar, and F. J. Herrmann, “Low-rank representation of subsurface extended image volumes with power iterations,” in *SEG Technical Program Expanded Abstracts 2019*, Society of Exploration Geophysicists, 2019, pp. 4470–4474.
- [43] R. Kumar, M. Graff-Kray, T. van Leeuwen, and F. J. Herrmann, “Low-rank representation of extended image volumes: Applications to imaging and velocity continuation,” in *SEG Technical Program Expanded Abstracts 2018*, Society of Exploration Geophysicists, 2018, pp. 4418–4422.
- [44] R. Kumar, M. Graff, I. Vasconcelos, and F. J. Herrmann, “Target-oriented imaging using extended image volumes: A low-rank factorization approach,” *Geophysical Prospecting*, vol. 67, no. 5, pp. 1312–1328, 2019.
- [45] F. Luporini, M. Lange, M. Louboutin, N. Kukreja, J. Hüchelheim, C. Yount, P. Witte, P. H. J. Kelly, F. J. Herrmann, and G. J. Gorman, “Architecture and performance of devito, a system for automated stencil computation,” *CoRR*, vol. abs/1807.03032, 2018. arXiv: 1807.03032.
- [46] 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, Mar. 2019, (Geophysics).
- [47] T. van Leeuwen, R. Kumar, and F. J. Herrmann, “Enabling affordable omnidirectional subsurface extended image volumes via probing,” *Geophysical Prospecting*, Aug. 2016, (published online in Geophysical Prospecting).

- [48] P. Sava and S. Fomel, “Time-shift imaging condition in seismic migration,” *GEOPHYSICS*, vol. 71, no. 6, S209–S217, 2006. eprint: <https://doi.org/10.1190/1.2338824>.
- [49] P. Sava and I. Vasconcelos, “Extended imaging conditions for wave-equation migration,” *Geophysical Prospecting*, vol. 59, no. 1, pp. 35–55, 2011.
- [50] P. A. Witte, M. Louboutin, F. Luporini, G. J. Gorman, and F. J. Herrmann, “Compressive least-squares migration with on-the-fly fourier transforms,” *Geophysics*, vol. 84, no. 5, pp. 1–76, 2019.
- [51] W. Hu, A. Abubakar, and T. M. Habashy, “Simultaneous multifrequency inversion of full-waveform seismic data,” *GEOPHYSICS*, vol. 74, no. 2, R1–R14, 2009. eprint: <https://doi.org/10.1190/1.3073002>.
- [52] M. Holmes, A. Gray, and C. Isbell, “Fast svd for large-scale matrices,” in *Workshop on Efficient Machine Learning at NIPS*, vol. 58, 2007, pp. 249–252.
- [53] P. Sava and I. Vasconcelos, “Extended imaging conditions for wave-equation migration,” *Geophysical Prospecting*, vol. 59, no. 1, pp. 35–55, 2011.
- [54] M. Louboutin, M. Lange, F. Luporini, N. Kukreja, P. A. Witte, F. J. Herrmann, P. Velesko, and G. J. Gorman, “Devito: An embedded domain-specific language for finite differences and geophysical exploration,” *arXiv preprint arXiv:1808.01995*, 2018.
- [55] N. Whitmore, “Iterative depth migration by backward time propagation,” in *SEG Technical Program Expanded Abstracts 1983*, Society of Exploration Geophysicists, 1983, pp. 382–385.
- [56] S Dong, J Cai, M Guo, S Suh, Z Zhang, B Wang, and e. Z. Li, “Least-squares reverse time migration: Towards true amplitude imaging and improving the resolution,” in *SEG Technical Program Expanded Abstracts 2012*, Society of Exploration Geophysicists, 2012, pp. 1–5.
- [57] C. Zeng, S. Dong, and B. Wang, “Least-squares reverse time migration: Inversion-based imaging toward true reflectivity,” *The Leading Edge*, vol. 33, no. 9, pp. 962–968, 2014.
- [58] T. van Leeuwen, A. Y. Aravkin, and F. J. Herrmann, “Seismic waveform inversion by stochastic optimization,” *International Journal of Geophysics*, vol. 2011, 2011.
- [59] E. Haber, M. Chung, and F. Herrmann, “An effective method for parameter estimation with PDE constraints with multiple right-hand sides,” *SIAM Journal on Optimization*, vol. 22, no. 3, pp. 739–757, 2012.

- [60] W. B. Powell, “Clearing the jungle of stochastic optimization,” in *Bridging data and decisions*, Informs, 2014, pp. 109–137.
- [61] D. A. Lorenz, F. Schopfer, and S. Wenger, “The linearized bregman method via split feasibility problems: Analysis and generalizations,” *SIAM Journal on Imaging Sciences*, vol. 7, no. 2, pp. 1237–1262, 2014.
- [62] F. J. Herrmann, N. Tu, and E. Esser, “Fast online migration with compressive sensing,” 2015.
- [63] M. Yang, P. Witte, Z. Fang, and F. Herrmann, “Time-domain sparsity-promoting least-squares migration with source estimation,” in *SEG Technical Program Expanded Abstracts 2016*, Society of Exploration Geophysicists, 2016, pp. 4225–4229.
- [64] W. Yin, “Analysis and generalizations of the linearized bregman method,” *SIAM Journal on Imaging Sciences*, vol. 3, no. 4, pp. 856–877, 2010.
- [65] J.-F. Cai, S. Osher, and Z. Shen, “Convergence of the linearized bregman iteration for l_1 -norm minimization,” *Mathematics of Computation*, vol. 78, no. 268, pp. 2127–2136, 2009.
- [66] R. G. Pratt, “Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model,” *Geophysics*, vol. 64, no. 3, pp. 888–901, 1999.
- [67] A. Y. Aravkin, T. van Leeuwen, and N. Tu, “Sparse seismic imaging using variable projection,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2013, pp. 2065–2069.
- [68] Z. Fang, R. Wang, and F. J. Herrmann, “Source estimation for wavefield-reconstruction inversion,” *Geophysics*, vol. 83, no. 4, R345–R359, 2018.
- [69] A. Y. Aravkin, T. van Leeuwen, H. Calandra, and F. J. Herrmann, “Source estimation for frequency-domain fwi with robust penalties,” in *74th EAGE Conference and Exhibition incorporating EUROPEC 2012*, 2012.
- [70] T. van Leeuwen, A. Y. Aravkin, F. J. Herrmann, M. Li, J. Rickett, and A. Abubakar, “Reply to the discussion,” *Geophysics*, vol. 79, no. 3, pp. X11–X15, 2014.
- [71] J. Rickett, “The variable projection method for waveform inversion with an unknown source function,” *Geophysical Prospecting*, vol. 61, no. 4, pp. 874–881, 2013.

- [72] N. Tu, A. Y. Aravkin, T. van Leeuwen, and F. J. Herrmann, “Fast least-squares migration with multiples and source estimation,” in *75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013*, 2013.
- [73] W. A. Mulder and R.-E. Plessix, “A comparison between one-way and two-way wave-equation migration,” *Geophysics*, vol. 69, no. 6, pp. 1491–1504, 2004.
- [74] N. Aoki and G. T. Schuster, “Fast least-squares migration with a deblurring filter,” *Geophysics*, vol. 74, no. 6, WCA83–WCA93, 2009.
- [75] F. J. Herrmann and X. Li, “Efficient least-squares imaging with sparsity promotion and compressive sensing,” *Geophysical prospecting*, vol. 60, no. 4, pp. 696–712, 2012.
- [76] N. Tu and F. J. Herrmann, “Fast imaging with surface-related multiples by sparse inversion,” *Geophysical Journal International*, vol. 201, no. 1, pp. 304–317, 2015.
- [77] E. J. Candès *et al.*, “Compressive sampling,” in *Proceedings of the international congress of mathematicians*, Madrid, Spain, vol. 3, 2006, pp. 1433–1452.
- [78] D. L. Donoho, “Compressed sensing,” *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [79] E. J. Candès and M. B. Wakin, “An introduction to compressive sampling,” *IEEE signal processing magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [80] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.
- [81] N. Tu and F. J. Herrmann, “Fast imaging with surface-related multiples by sparse inversion,” *Geophysical Journal International*, vol. 201, no. 1, pp. 304–317, 2015, (Geophysical Journal International).
- [82] F. J. Herrmann, C. R. Brown, Y. A. Erlangga, and P. P. Moghaddam, “Curvelet-based migration preconditioning and scaling,” *Geophysics*, vol. 74, no. 4, A41–A46, 2009.
- [83] T. J.P.M. O. ’t Root, C. C. Stolk, and M. V. de Hoop, *Linearized inverse scattering based on seismic reverse time migration*, 2010. arXiv: 1012.4278 [math.AP].
- [84] N. Whitmore, S. Crawley, *et al.*, “Applications of rtm inverse scattering imaging conditions,” in *2012 SEG Annual Meeting*, Society of Exploration Geophysicists, 2012.

- [85] P. A. Witte, M. Yang, and F. J. Herrmann, “Sparsity-promoting least-squares migration with the linearized inverse scattering imaging condition,” in *EAGE Annual Conference Proceedings*, (EAGE, Paris), Jun. 2017.
- [86] A. B. Weglein, F. A. Gasparotto, P. M. Carvalho, and R. H. Stolt, “An inverse-scattering series method for attenuating multiples in seismic reflection data,” *Geophysics*, vol. 62, no. 6, pp. 1975–1989, 1997.
- [87] J Biersteker, “Magic: Shell’s surface multiple attenuation technique,” in *SEG Technical Program Expanded Abstracts 2001*, Society of Exploration Geophysicists, 2001, pp. 1301–1304.
- [88] D. Verschuur and A. Berkhout, “Transforming multiples into primaries: Experience with field data,” in *SEG Technical Program Expanded Abstracts 2005*, Society of Exploration Geophysicists, 2005, pp. 2103–2106.
- [89] N. Hargreaves, “Surface multiple attenuation in shallow water and the construction of primaries from multiples,” in *SEG Technical Program Expanded Abstracts 2006*, Society of Exploration Geophysicists, 2006, pp. 2689–2693.
- [90] J Brittan, T Martin, M Bekara, and K Koch, “3d shallow water demultiple—extending the concept,” *First Break*, vol. 29, no. 9, pp. 97–101, 2011.
- [91] T. T. Lin and F. J. Herrmann, “Estimation of primaries by sparse inversion with scattering-based multiple predictions for data with large gaps,” *Geophysics*, vol. 81, no. 3, pp. V183–V197, 2016.
- [92] F. J. Herrmann, N. Tu, and E. Esser, “Fast ”online” migration with Compressive Sensing,” in *EAGE Annual Conference Proceedings*, (EAGE, Madrid), UBC, UBC, Jun. 2015.
- [93] T. T. Lin and F. J. Herrmann, “Robust estimation of primaries by sparse inversion via one-norm minimization,” *Geophysics*, vol. 78, no. 3, R133–R150, 2013.
- [94] S. Lu, D. N. Whitmore, A. A. Valenciano, and N. Chemingui, “Separated-wavefield imaging using primary and multiple energy,” *The Leading Edge*, vol. 34, no. 7, pp. 770–778, 2015.
- [95] D. Lecerf, E. Hodges, S. Lu, A. Valenciano, N. Chemingui, P. Johann, and E. Thedy, “Imaging primaries and high-order multiples for permanent reservoir monitoring: Application to jubarte field,” *The Leading Edge*, vol. 34, no. 7, pp. 824–828, 2015.
- [96] M. Lange, N. Kukreja, M. Louboutin, F. Luporini, F. Vieira, V. Pandolfo, P. Velsko, P. Kazakas, and G. Gorman, “Devito: Towards a generic finite difference dsl

using symbolic python,” in *2016 6th Workshop on Python for High-Performance and Scientific Computing (PyHPC)*, IEEE, 2016, pp. 67–75.

- [97] J. Paffenholz, J Stefani, B McLain, and K Bishop, “Sigsbee_2a synthetic subsalt dataset-image quality as function of migration algorithm and velocity model error,” in *64th EAGE Conference & Exhibition*, 2002.
- [98] K. Wapenaar, “Reciprocity properties of one-way propagators,” *Geophysics*, vol. 63, no. 5, pp. 1795–1798, 1998.
- [99] N. Tu, “Fast imaging with surface-related multiples,” (PhD), phd, The University of British Columbia, Vancouver, Aug. 2015.
- [100] W. W. Symes, *Using iwave*, 2013.
- [101] W. Dai, P. Fowler, and G. T. Schuster, “Multi-source least-squares reverse time migration,” *Geophysical Prospecting*, vol. 60, no. 4, pp. 681–695, 2012.
- [102] R. E. Plessix, W. Mulder, *et al.*, “Amplitude-preserving finite-difference migration based on a least-squares formulation in the frequency domain,” in *2002 SEG Annual Meeting*, Society of Exploration Geophysicists, 2002.
- [103] Y. Duan, P. Sava, A. Guitton, *et al.*, “Elastic least-squares reverse time migration,” in *2016 SEG International Exposition and Annual Meeting*, Society of Exploration Geophysicists, 2016.
- [104] G. Dutta and G. T. Schuster, “Attenuation compensation for least-squares reverse time migration using the viscoacoustic-wave equation,” *Geophysics*, vol. 79, no. 6, S251–S262, 2014.
- [105] S. Operto, Y. Gholami, V Prioux, A. Ribodetti, R Brossier, L Metivier, and J. Virieux, “A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice,” *The Leading Edge*, vol. 32, no. 9, pp. 1040–1054, 2013.
- [106] G. Beylkin, R. Burrige, *et al.*, “Multiparameter inversion for acoustic and elastic media,” in *1987 SEG Annual Meeting*, Society of Exploration Geophysicists, 1987.
- [107] A Przebindowska, A Kurzmann, D Köhn, and T Bohlen, “The role of density in acoustic full waveform inversion of marine reflection seismics,” in *74th EAGE Conference and Exhibition incorporating EUROPEC 2012*, 2012.
- [108] R. Plessix, P Milcik, H Rynja, A Stopin, K Matson, and S Abri, “Multiparameter full-waveform inversion: Marine and land examples,” *The Leading Edge*, vol. 32, no. 9, pp. 1030–1038, 2013.

- [109] J. Bai, D. Yingst, *et al.*, “Simultaneous inversion of velocity and density in time-domain full waveform inversion,” in *2014 SEG Annual Meeting*, Society of Exploration Geophysicists, 2014.
- [110] 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.
- [111] A. Y. Aravkin, R. Kumar, H. Mansour, B. Recht, and F. J. Herrmann, “Fast methods for denoising matrix completion formulations, with applications to robust seismic data interpolation,” *SIAM Journal on Scientific Computing*, vol. 36, no. 5, S237–S266, Oct. 2014, (SISC).
- [112] G. Golub and V. Pereyra, “Separable nonlinear least squares: The variable projection method and its applications,” *Inverse problems*, vol. 19, no. 2, R1, 2003.
- [113] N. Tu, A. Aravkin, T. van Leeuwen, T. Lin, and F. J. Herrmann, “Source estimation with surface-related multiplesfast ambiguity-resolved seismic imaging,” *Geophysical Journal International*, vol. 205, no. 3, pp. 1492–1511, 2016.