

Gradient projection methods with applications to simultaneous source
seismic data processing

by

Jinkun Cheng

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Geophysics

Department of Physics
University of Alberta

©Jinkun Cheng, 2017

Abstract

Simultaneous source acquisition, or blended acquisition, has become an important strategy to reduce the cost of seismic surveys by allowing overlapping between different sources. The major technical challenge associated with this acquisition design is the strong interferences caused by the closely fired shots.

This thesis focuses on the separation or the deblending of simultaneous source data via constrained inversion methods. The cost function is the misfit between the predicted and the observed blended data. The constraint is that the desired signal is coherent in the common receiver, common offset, and common midpoint domains when the fire time delay corresponding to each shot is corrected. The simultaneous source interferences would appear incoherent when the randomized source fire scheme is applied. In this thesis, I assume that the desired coherent signal can be represented via a low-rank matrix. The randomly distributed interferences would increase the rank. The coherence constraint for deblending can be implemented effectively by a low-rank constraint in the corresponding data domain. I adopt the gradient projection method that iteratively solves this low-rank constrained inverse problem for deblending. The projection filters are the f-x-y eigenimage filter (Chapter 3) and the Singular Spectrum Analysis filter (Chapter 4) that suppress the source cross-talk artifacts while preserving the unblended signal. Fast implementations of the two reduced-rank filters are achieved via randomized rank-reduction methods. The gradient projection framework is then extended to the direct imaging of simultaneous data via shot-profile least-squares migration.

Preface

A version of the work in chapter 3 of this thesis has been published in a journal paper: Cheng, J., and M. D. Sacchi, 2016, Fast dual-domain reduced-rank algorithm for 3D deblending via randomized QR decomposition, *Geophysics*, 81, V57-V66.

A version of the work in chapter 4 of this thesis has been published in a journal paper: Cheng, J., and M. D. Sacchi, 2015, Separation and reconstruction of simultaneous source seismic data via iterative rank reduction, *Geophysics*, 80, V89-V101.

In these publications, I was responsible for designing and programming the algorithms, preparing data examples and writing the manuscripts. Dr. Sacchi was the supervisory author and was involved in concept formulation and manuscript editing.

To my grandma, Toyng Li (1933-2017)

Acknowledgements

First and foremost, I am deeply indebted to my supervisor, Dr. Mauricio Sacchi, who has provided me invaluable advice and has given me more support than I can ever ask for. His wealth of knowledge and clarity of thoughts have made working with him an exceptional experience to me. I am also grateful to Prof. Doug Schmidt, Prof. Moritz Heimpel, Prof. Al Meldrum, and Prof. Vakhtang Putkaradze for their trust in me, and to Prof. Johan Robertsson for being on my thesis committee. I would like to thank everyone in the Signal Analysis and Imaging Group (SAIG) for those enjoyable conversations and the fruitful discussions. Furthermore, I appreciate the sponsors of SAIG for their financial support and constructive feedbacks from the annual meetings. Also, I wish to thank Prof. Jianjun Gao, Dr. Li Han, Dr. Jiajun Han and Dr. Yuanyin Zhang for their guidance and supports during the early stage of my Ph.D. study. I would like to thank my friends Yunfeng Chen, Wei Xie, Jingying Zeng, Li Li and Xiaotian Yu for their company during my stay at the University of Alberta. I would like to express my special thanks to Min Zhang, who has always been there putting up with me. Finally, I wish to dedicate this piece of work to my family, especially my parents and my grandma, who provided me with trust, understandings, and encouragements throughout my academic pursuits.

Contents

1	Introduction	1
1.1	Simultaneous source/Blended acquisition	2
1.1.1	Blended acquisition design	2
1.1.2	Numerical simulation of blended acquisition	6
1.2	Review of previous work	8
1.2.1	Direct imaging of blended data	8
1.2.2	Simultaneous source separation	8
1.3	Contributions of this thesis	10
1.4	Thesis overview	11
2	The gradient projection method	13
2.1	The method of gradient descent	15
2.1.1	Convergence analysis	15
2.1.2	The step-size rule	17
2.2	The method of gradient projection	18
3	Deblending via iterative rank reduction	20
3.1	Introduction	20
3.2	Theory	21
3.2.1	The low-rank constraint for deblending	21
3.2.2	Rank constrained minimization	23

3.2.3	Randomized QR decomposition	25
3.3	Examples	28
3.3.1	Example with synthetic 3D VSP data set	28
3.3.2	Example simulated from a real data set	37
3.4	Conclusions	38
4	Deblending via iterative SSA	41
4.1	Introduction	41
4.2	Theory	42
4.2.1	Singular spectrum analysis	42
4.2.2	The SSA projection operator	45
4.2.3	Joint separation and reconstruction of seismic sources	47
4.3	Examples	48
4.3.1	Comparison of projection operators	48
4.3.2	Deblending: SAIG sythetic data set	49
4.3.3	Joint deblending and reconstruction: Gulf of Mexico data set	56
4.4	An analysis of simultaneous source fire time delays	59
4.4.1	Evaluation of different statistical models	60
4.4.2	Nonlinear optimization of fire time delays	63
4.5	Conclusions	65
5	Blended shot-profile least-squares migration	68
5.1	Introduction	68
5.2	Theory	68
5.2.1	Preliminaries	68
5.2.2	Shot-profile migration and de-migration of blended data	71
5.3	Least-squares migration via gradient projection	73
5.4	Examples	75

5.4.1	Two layer model	75
5.4.2	SAIG velocity model	77
5.4.3	Acoustic Marmousi2 model	81
5.5	Conclusions	86
6	Conclusions	91
6.1	Summary	91
6.2	Contributions	93
6.3	Future developments	94
	Bibliography	95
	Appendices	
A	Theory of random projections	103
B	Fast and memory efficient SSA	107
B.1	introduction	107
B.2	Method	107
B.3	Performance of fast and memory efficient SSA	112
B.4	Conclusions	112
C	Deblending via RPCA	113
C.1	Introduction	113
C.2	Theory	113
C.2.1	Principal Component Analysis	113
C.2.2	Robust Principal Component Analysis	114
C.2.3	RPCA seismic data noise attenuation	116
C.3	Examples	117
C.3.1	Erratic noise elimination	117
C.3.2	Simultaneous source noise suppression	119
C.4	Conclusions	121

List of Tables

1.1	Summary of simultaneous source seismic data acquisition methods	4
4.1	Quality of deblending Q_S and performance I for the projected gradient method with different projection operators: rank reduction (proposed method), $f - x$ deconvolution (Peng and Liu, 2013) and $f - k$ thresholding (Abma et al., 2010). These results correspond to the comparison tests portrayed in Figure 4.2.	49
B.1	Comparison of computational accuracy and efficiency for different SSA methods: conventional 3D MSSA (MSSA), randomized MSSA (RMSSA), Lanczos Bidagonalization (Lanczos), and the proposed fast and memory efficient SSA (FSSA). In this experiment, the subset size for FSSA is three times of the desired rank.	112

List of Figures

1.1	Comparison of the acquisition geometry between (a) conventional seismic acquisition, and (b) simultaneous source acquisition.	3
1.2	Illustration of phase-encoding of dual source simultaneous source acquisition using vibroseis (Meunier, 2011).	3
1.3	Illustration of early marine simultaneous source acquisition with two shots well separated in space (Beasley et al., 1998).	5
1.4	Illustration of gathers with each trace corresponding to a different source location. The figures on the left shows the source and receiver pairs of (a) common receiver, (b) common offset, and (c) common midpoint gathers. The figures on the right shows the illustration of a (d) common receiver, (e) a common offset, and (f) a common midpoint gather in 2D seismic acquisition.	5
1.5	Illustration of numerical blending using two shots. The time delay for shot 2 is denoted by τ	7
1.6	Illustration of pseudo-deblending of two shots. The time delay τ is shifted back for shot 2. However, the interferences cannot be removed by pseudo-deblending.	7
1.7	Blending noise of pseudo-deblended data set in (a) common receiver, (b) common offset, and (c) common midpoint domain. The desired signal is coherent while the blending noise appears incoherent.	9
2.1	Illustration of the gradient descent method. The method starts at \mathbf{x}^0 and converges to the minimum solution \mathbf{x}^o in the contour map.	16
3.1	Distribution of singular values of spatial data in normal (a) and logarithmic (b) scale at 20Hz from the true unblended common receiver gather in $f - x - y$ domain (red). We also portray the distribution of singular values for data contaminated with source interferences in a pseudo-deblended common receiver gather (blue).	22

3.2	The processing time of rank reduction versus the size of matrix both in logarithmic scale. The blue curve shows the processing time utilizing the truncated SVD while the red is the processing time of the RQRD method. We choose P equal $3k$ and the RQRD algorithm is about 10 times faster than the conventional tSVD.	27
3.3	Distribution of sources and receivers of the 3D VSP data set. The sources are deployed on a regular grid in the surface and the receivers are deployed between 1350m and 1850m in the subsurface. Each red point represents a group of 10 sources and each blue triangle represents 5 receivers.	28
3.4	Distribution of the firing time of seismic sources in the conventional acquisition (Blue) and the one-vessel simultaneous source acquisition (Red). In this example, STR equals 2 and 50 % of acquisition time is saved by simultaneous source acquisition	30
3.5	The quality of deblending versus rank using both tSVD and RQRD. The blue curve shows the results utilizing the truncated SVD as the low rank projection operator while the red is the results corresponding to the RQRD method. The survey time ratio is 10 and the firing time remains the same for all trials. Compared to the truncated SVD which relies on the selection of rank K , a broad range of subset sizes in RQRD, $p \in [1.5K, 5K]$, ensures the success of deblending.	31
3.6	Probability map for different rank and survey time ratio. For each specific rank and survey time ratio, 50 realizations of random firing time delays were generated with uniform distribution. If the quality factor after deblending (Q_s) is larger than 20, we consider the method successfully removed the simultaneous source crosstalk. The white area in the figure indicates an area where the proposed algorithm succeeded for all 50 trials.	32
3.7	Results of simultaneous source separation in common receiver domain when STR equals to 2: (a) The real unblended time slice at 1.2 s. (b) Pseudo-deblended time slice. (c) Deblended time slice after after 20 iterations of the proposed algorithm. (d) Differences between (a) and (c). In this example, the signal-to-noise ratio after separation is 36.5dB.	33
3.8	Results of simultaneous source separation in common receiver domain when STR equals to 2: (a) The real unblended common receiver gather (centre receiver). (b) Pseudo-deblended common receiver gather. (c) Deblended common receiver gather after 20 iterations. (d) Differences between (a) and (c). In this example, the signal-to-noise ratio after separation is 36.5dB.	33
3.9	Results of simultaneous source separation in common shot domain when STR equals to 2: (a) The real unblended common shot gather (centre shot). (b) Pseudo-deblended shot record. (c) Deblended shot record after 20 iterations. (d) Differences between (a) and (c). In this example, the signal-to-noise ratio after separation is 36.5dB.	34

3.10	Results of simultaneous source separation in common receiver domain when STR equals to 10: (a) The real unblended time slice at 1.2 s. (b) Pseudo-deblended time slice. (c) Deblended time slice via the proposed algorithm. (d) Differences between (a) and (c). In this example, the signal-to-noise ratio after separation is 28.4dB.	34
3.11	Results of simultaneous source separation in common receiver domain when STR equals to 10: (a) The real unblended common receiver gather (centre receiver). (b) Pseudo-deblended common receiver gather. (c) Deblended common receiver gather via the proposed algorithm. (d) Differences between (a) and (c). In this example, the signal-to-noise ratio after separation is 28.4dB.	35
3.12	Results of simultaneous source separation in common shot domain when STR equals to 10: (a) The real unblended common shot gather (centre shot). (b) Pseudo-deblended shot record. (c) Deblended shot record via the proposed algorithm. (d) Differences between (a) and (c).In this example, the signal-to-noise ratio after separation is 28.4dB.	35
3.13	Results of simultaneous source separation in common receiver domain when STR equals to 20: (a) The real unblended time slice at 1.2 s. (b) Pseudo-deblended time slice. (c) Deblended time slice via the proposed algorithm. (d) Differences between (a) and (c). In this example, the signal-to-noise ratio after separation is 20.2dB.	36
3.14	Results of simultaneous source separation in common receiver domain when STR equals to 20: (a) The real unblended common receiver gather (centre receiver). (b) Pseudo-deblended common receiver gather. (c) Deblended common receiver gather via the proposed algorithm. (d) Differences between (a) and (c). In this example, the signal-to-noise ratio after separation is 20.2dB.	36
3.15	Results of simultaneous source separation in common shot domain when STR equals to 20: (a) The real unblended common shot gather (centre shot). (b) Pseudo-deblended shot record. (c) Deblended shot record via the proposed algorithm. (d) Differences between (a) and (c).In this example, the signal-to-noise ratio after separation is 20.2dB.	37
3.16	Results of simultaneous source separation of a 2D marine line: (a) The real unblended CMP gather. (b) CMP gather sorted from Pseudo-deblended data. (c) Deblended results via the proposed algorithm. (d) Differences between (a) and (c).In this example, the signal-to-noise ratio after separation is 14.2dB.	39
3.17	Results of simultaneous source separation of a 2D marine line: (a) The real unblended common offset gather. (b) (b) Common offset gather sorted from Pseudo-deblended data. (c) Deblended result via the proposed algorithm. (d) Differences between (a) and (c).	40
4.1	The windowing operator and its adjoint operator. The windowing operator \mathcal{W} extracts small patches of data from an entire gather. The adjoint operator \mathcal{W}^* combines all the processed small patches back into a gather. A Gaussian taper is used to combine areas with overlap.	46

4.2	Comparisons of deblending results for numerically blended synthetic data with linear events. (a) The original unblended synthetic data. (b) Pseudo-deblended gather. (c) Deblending result with the proposed iterative rank reduction method. (d) Deblending result using iterative $f - x$ deconvolution. (e) Deblending results using iterative $f - k$ domain thresholding. (f) Difference section between a and c. (g) Difference section between a and d. (h) Difference section between a and e.	50
4.3	Velocity model utilized to simulate blended data for our examples via finite difference modelling. We synthesize a data set with 4 ms time interval and a 20 Hz central-frequency Ricker wavelet. We also overlaid the source (*) and receiver (▲) geometry in this plot.	51
4.4	Spatial and temporal distribution of firing times for conventional seismic acquisition (Blue) and 2D simultaneous source acquisition. The multi-vessel scenario is portrayed in red. In each round, 5 sources fire with small random time delays. The spatial distance between sources are fixed in each round. . .	51
4.5	Convergence of the iterative SSA filtering deblending algorithm. Blue line indicates the l_2 norm of the difference between blended observations and the synthesized blended observation versus iteration. We also portray in red the difference between the unblended data and the true data versus iteration. . .	52
4.6	Results of the proposed iterative rank reduction deblending method in common shot domain. (a) The original shot gather from synthetic data. (b) Pseudo-deblended shot gather after numerical blending. (c) The deblended shot gather after 30 iterations of the proposed algorithm. The quality of data has been improved to 15.1 dB with respect to 0.5 dB. (d) Difference section between original and deblended data.	53
4.7	Results of the proposed iterative rank reduction deblending method in common receiver domain. (a) The original common receiver gather from synthetic data. (b) Pseudo-deblended common receiver gather after numerical blending. (c) The deblended common receiver gather after 30 iterations of the proposed method. (d) Difference section between original and deblended data.	54
4.8	Results of the proposed iterative rank reduction deblending method for the near-offset section. (a) The original near-offset section from synthetic data. (b) Pseudo-deblended near-offset section after numerical blending. (c) The deblended near-offset section after 30 iterations.	55
4.9	Spatial and temporal distribution of firing time for a conventional seismic acquisition (Blue) and 2D simultaneous source acquisition with 25 missing shots for one vessel scenario (red). (a) Source firing times. (b) Firing time intervals between adjacent sources.	56
4.10	Results of deblending and reconstruction for a common shot gather. (a) The original shot gather from a Gulf of Mexico dataset. (b) Pseudo-deblended shot gather after numerical blending and sampling. The desired source is missing. (c) The deblended and reconstructed shot gather after 30 iterations. The quality of data has been improved to 10.8 dB from -1.5 dB. (d) Difference section between original and deblended and reconstructed data.	57

4.11	Results of deblending and reconstruction for a common receiver gather. (a) The original common receiver gather from a Gulf of Mexico dataset. (b) Pseudo-deblended common receiver gather after numerical blending. (c) Deblended and reconstructed common receiver gather after 30 iterations. (d) Difference section between original and deblended data.	58
4.12	Results of deblending and reconstruction for whole data volume. (a) The original data from the Gulf of Mexico. (b) Pseudo-deblended data. (c) Deblended and reconstructed data after 30 iterations. (d) Difference section between original and deblended data.	59
4.13	Evaluation of deblending based on fire time delay distributions: (a), (b) and (c) shows the conventional fire time (blue) over the simultaneous source fire time distribution (red) for uniform, exponential, and normal distribution, respectively. (d), (e) and (f) display the quality of deblending versus the inverse of STR when fire time delays are generated from Uniform, exponential, and normal distribution, respectively.	62
4.14	Thermal state energy versus iteration numbers in Simulated Annealing inversion method.	64
4.15	Optimal time delays inverted via fast simulated annealing. In this example, the STR equals to 2.	65
4.16	Results of deblending for a common receiver gather using optimal source fire time delays. (a) The original common receiver gather from the Gulf of Mexico dataset. (b) Pseudo-deblended common receiver gather (c) Deblended common receiver gather with iterative SSA filtering (d) Difference section between original and deblended data.	66
5.1	Distribution of firing times for simultaneous source acquisition (red) compared to conventional acquisition (blue). In this example about 75% of acquisition time is saved by source blending. The computational efficiency exhibits a similar pattern for the proposed least squares migration.	75
5.2	Convergence of the gradient projection algorithm for the two-layer example. The algorithm converges within 30 iterations.	76
5.3	Migration results of the two-layer model: a shot-index common image gather ($x = 600$). (a) A shot-index common receiver gather from conventional unblended shot-profile migration; (b) A shot-index common receiver gather from blended shot-profile migration; (c) A shot-index common receiver gather from blended least-squares shot-profile migration.	77
5.4	Migration results of the two-layer model: partial image for shot 30. (a) A partial image from conventional unblended shot-profile migration; (b) A partial image from the blended shot-profile migration; (c) A partial image from the blended least-squares shot-profile migration.	78
5.5	Migration results of the two-layer model: stacked image (a) Stacked image from conventional unblended shot-profile migration; (b) Stacked image from the blended shot-profile migration; (c) Stacked image from the blended least-squares shot-profile migration.	79

5.6	Results of deblending via the proposed algorithm for the two-layer model. (a) Ideal unblended shot gather (center shot) (b) Pseudo-deblended shot gather (c) Deblended shot gather after 30 iterations of gradient projection in the least-squares migration. The interferences are effectively suppressed.	80
5.7	The SAIG velocity model.	81
5.8	Convergence of the gradient projection algorithm for the SAIG example. The algorithm converges within 60 iterations.	82
5.9	Migration results for SAIG velocity model: a shot-index common image gather($x = 2.5km$) (a) A shot-index common receiver gather from conventional unblended shot-profile migration; (b) A shot-index common receiver gather from blended shot-profile migration; (c) A shot-index common receiver gather from blended least-squares shot-profile migration.	82
5.10	Migration results for SAIG velocity model: a partial image for a shot ($S_x = 2.5km$). (a) A partial image from conventional unblended shot-profile migration; (b) A partial image from the blended shot-profile migration; (c) A partial image from the blended least-squares shot-profile migration.	83
5.11	Migration results for SAIG velocity model: stacked image (a) Stacked image from conventional unblended shot-profile migration; (b) Stacked image from the blended shot-profile migration; (c) Stacked image from the blended least-squares shot-profile migration.	84
5.12	Deblending results via the proposed algorithm for the SAIG velocity model: (a) Ideal unblended shot gather (centre shot); (b) Pseudo-deblended shot gather; (c) Deblended shot gather after 60 iterations of least-squares migration; (d) Difference between (a) and (c).	85
5.13	The Acoustic Marmousi2 model.	86
5.14	Convergence of the gradient projection algorithm for the Marmousi2 example. The algorithm converges within 80 iterations.	87
5.15	Migration results for Marmousi2 model: a shot-index common image gather($x = 2km$) (a) A shot-index common receiver gather from conventional unblended shot-profile migration; (b) A shot-index common receiver gather from blended shot-profile migration; (c) A shot-index common receiver gather from blended least-squares shot-profile migration.	87
5.16	Migration results for Marmousi2 model: a partial image for a shot ($S_x = 2.5km$). (a) A partial image from conventional unblended shot-profile migration; (b) A partial image from the blended shot-profile migration; (c) A partial image from the blended least-squares shot-profile migration.	88
5.17	Migration results for Marmousi2 model: stacked image (a) Stacked image from conventional unblended shot-profile migration; (b) Stacked image from the blended shot-profile migration; (c) Stacked image from the blended least-squares shot-profile migration.	89

5.18	Deblending results for the Marmousi2 model. (a) Ideal unblended shot gather (centre shot); (b) Pseudo-deblended shot gather; (c) Deblended shot gather after 100 iterations of least-squares migration. (d) Difference between (a) and (c).	90
C.1	Results of Robust de-noising for the synthetic data set. (a) The ideal CMP gather of the 3D data cube. (b) The CMP gather contaminated with Gaussian and erratic noise. (c) De-noised gather with $f - x - y$ eigenimage filtering. (d) CMP gather after RPCA De-noising. (e) Differences between (a) and (c). (f) Difference between (a) and (d).	118
C.2	RPCA de-noising for the Alaska data. (a) The original common offset gather before RPCA denoising. (b) Common offset gather after applying RPCA. (c) Difference between (a) and (b). (d) The original common midpoint gather before RPCA denoising. (e) Common midpoint gather after applying RPCA. (f) Difference between (d) and (e).	119
C.3	Deblending via RPCA. (a) The unblended common offset gather. (b) Pseudo-deblended common offset gather. (c) Common offset gather after RPCA de-noising. (d) Differences between (a) and (c).	120
C.4	Deblending via RPCA. (a) The unblended common midpoint gather. (b) Pseudo-deblended common midpoint gather. (c) Common midpoint gather after RPCA de-noising. (d) Differences between (a) and (c).	121

List of abbreviations and symbols

MSSA	Multichannel Singular Spectrum Analysis
PCA	Principal Component Analysis
POCS	Projection onto convex sets
RPCA	Robust Principal Component Analysis
RQRD	Randomized QR Decomposition
SSA	Singular Spectrum Analysis
STR	Survey time ratio
SVD	Singular Value Decomposition
tSVD	Truncated Singular Value Decomposition
VSP	Vertical Seismic Profile
b	The blended data
D	The data matrix
B	The blending operator
\mathcal{G}	Wavefield propagation operator
τ	Source fire time delay
K	The desired rank/number of dips
m	Migrated image
P	Size of random subset for RQRD
Q_S	Quality of separation
S	Sorting operator

T	Cooling temperature for simulated annealing
t	Source initiation time
W_R	Receiver side wavefield
W_S	Source side wavefield
α	The largest eigenvalue of an operator
\mathbf{A}	A matrix
\mathbf{C}	A circulant matrix
\mathbf{d}	The data vector
\mathbf{g}	The gradient
\mathbf{H}	Hankel matrix constructed from frequency-space data
\mathbf{N}	The blending noise
\mathbf{T}	A Toeplitz matrix
\mathbf{x}, \mathbf{y}	Vectors
\mathbf{x}^o	The optimal solution
\mathcal{A}	An operator
\mathcal{B}^*	The pseudo-deblending operator
\mathcal{C}	The coherent constraint
\mathcal{F}	The Fourier transform
\mathcal{F}^{-1}	The inverse Fourier transform
\mathcal{L}	The de-migration operator
\mathcal{L}^*	The migration operator
\mathcal{P}_C	The projection operator
\mathcal{T}	The sampling operator
$\mathcal{W}, \mathcal{W}^*$	Data patching and unpatching operator
λ	The step length
\mathcal{D}	The data tensor
ν	The iteration number
$\mathbf{\Omega}$	Matrix composed by independent random vectors
σ, v	Singular values
\tilde{D}	Data in frequency space domain
\mathbf{L}	A low-rank matrix
\mathbf{S}	A sparse matrix

CHAPTER 1

Introduction

A seismic survey is a primary tool to explore the structure and the properties of the earth subsurface. The simplest seismic survey consists of initiating a seismic source and deploying an array of receivers to measure ground motion. In this experiment, seismic waves emitted from the source propagates down into Earth's subsurface. The energy is scattered, transmitted, and reflected to Earth's surface before being recorded by receivers. One can repeat the process by firing a number of shots at different locations. In conventional methods of seismic data acquisition, sufficient time gaps are deployed between different sources. The goal is to avoid signal overlapping between shot records. The acquired seismic data then go through a sequence of processes, and the ultimate goal of a seismic survey is to acquire an accurate image that reflects the structures of the subsurface. To achieve this, undesired signals, such as noise and multiple arrivals need to be eliminated, and a step named migration is required to map the recorded energy to the corresponding locations in the subsurface.

In a successful seismic survey, seismic sources and detectors are placed with a careful design in accordance with the acquisition environment. For land acquisition, dynamite and vibroseis are the most commonly used seismic sources. Because of the complexity of the environment, the acquisition crew often face various logistics obstacles and accessibility problems. Sources and receivers are often distributed on an irregular grid. For marine seismic data acquisition, dynamite sources are not allowed due to the strong shock wave that destructs the environment for marine lives. Airguns become the most popular seismic source due to their reliability and repeatability (Meunier, 2011). The detectors for marine acquisition can be towed streamers behind the vessels or the ocean bottom nodes deployed on the seabed. For both land and marine acquisition, one can choose among vertical component geophones, multi-component geophones, and hydrophones based on different exploration purposes.

Besides, great attention is required if the target for the seismic survey involves challenging

areas, such as deep water, salt domes, and thick carbonates in the near surface. The density of the source and receiver grids needs to be significantly increased to capture enough reflections for the details of the target. The offset and azimuthal coverage should also be enlarged to fully illuminate the area. For instance, the wide-azimuth marine acquisition has been proposed to improve the quality of seismic survey for subsalt exploration. In this situation, conventional seismic acquisition methods can be extremely expensive because of the source non-overlapping constraint and the huge amount of sources that are required.

1.1 Simultaneous source/Blended acquisition

Simultaneous source acquisition techniques have been gaining popularity as a low-cost strategy to improve seismic data acquisition (Beasley et al., 1998; Berkhout, 2008). In the configuration of simultaneous source acquisition, instead of firing one shot at a time and imposing large time intervals between different shots, several seismic sources fire at close time intervals. The responses are then recorded by the same set of receivers. Figure 1.1 shows the comparison between conventional and simultaneous source acquisition. By allowing overlaps between the closely fired shots, we can acquire multiple shot records during the period that one could have used to acquire one conventional shot record. Meanwhile, adding simultaneous sources could increase the source density and thus improve the quality of seismic data without extra costs. The major technical challenge for simultaneous source seismic data processing lies in the strong interferences introduced by the closely fired shots.

1.1.1 Blended acquisition design

Early work on simultaneous source techniques focused on land acquisition with two groups of vibroseis operating at the same time at different spatial positions (Silverman, 1979; Garotta, 1983). Figure 1.2 shows the phase-encoding technique for separating the simultaneous sources: one group of the vibroseis generates a signal with positive polarity, and the other group generates the negatively polarized signal. The sources are then separated by summing and differencing the two records (Meunier, 2011). The high-fidelity vibratory seismic (HFVS) acquisition generalizes the method with different phase encoding schemes to take the source crosstalk under control (Sallas, 1989). A comprehensive review of land simultaneous source acquisition can be found in Bagaini et al. (2012).

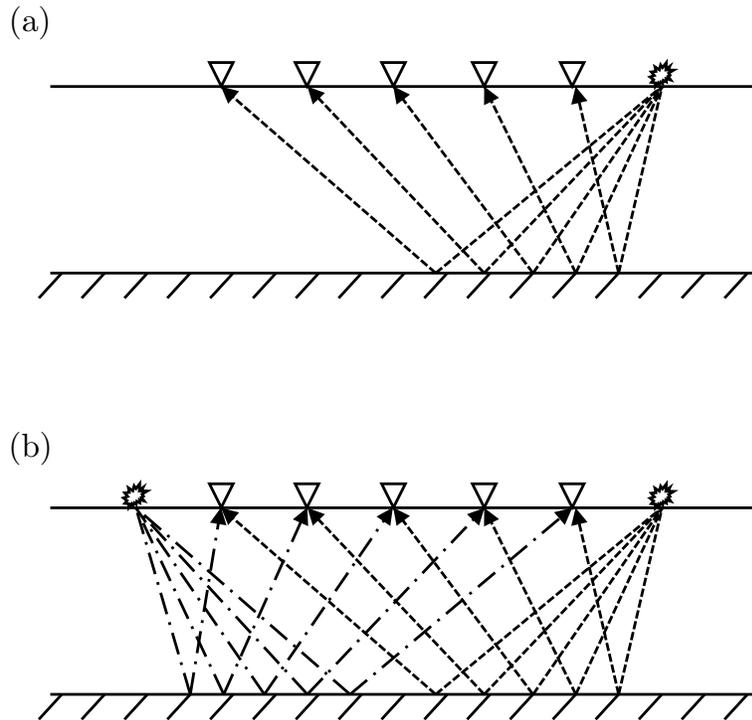


Figure 1.1: Comparison of the acquisition geometry between (a) conventional seismic acquisition, and (b) simultaneous source acquisition.

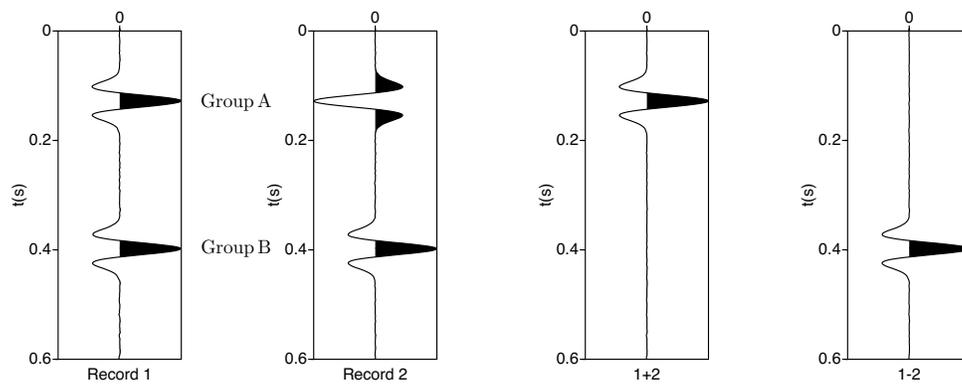


Figure 1.2: Illustration of phase-encoding of dual source simultaneous source acquisition using vibroseis (Meunier, 2011).

This thesis focuses on marine simultaneous source acquisition. The phase-encoding methods are not applicable to impulsive sources like airguns. Separation of marine simultaneous sources relies on the source geometric distance and the specific source initiation schemes. Figure 1.3 shows an early simultaneous shooting experiments with an additional source vessel firing behind towed streamers (Beasley et al., 1998). The sources are separated using the moveout information as they are well separated in space. Other methods utilize the random source initiation time for simultaneous source acquisition. Stefani et al. (2007) and Hampson et al. (2008) proposed the source dithering scheme with small randomized source initiation time. In the dithered simultaneous source acquisition, one vessel fire with a fixed time intervals and another vessel fire at a random time around the first source. In the common receiver, common midpoint, and common offset data domain, each trace corresponds to a different source location (Figure 1.4). When we align the signal from one source in these domains, the signal from the other source would appear incoherent. Separation of simultaneous sources is achievable utilizing coherence constraints. Moldoveanu et al. (2012) proposed a multi-vessel simultaneous source acquisition with each vessel sails along a circle. In this case, the sources are geometrically separated and are dithered by small random time delays. Howe et al. (2008) and Abma et al. (2012) proposed the independent simultaneous sweeping (ISS) technique that allows multiple (more than two) vessels to work independently without synchronizing the source activities. Compared to the dithering method, which is based on small random time shifts, ISS produces sufficiently large time delays between shots. Besides randomized source initiation schemes, Robertsson et al. (2016) proposed to use periodical source modulation functions to encode sources that allows wavefield apparition in the spectral domain. The features of some typical designs for simultaneous source acquisition are summarized in Table 1.1 .

Method	Environment	Authors	Year	Feature
Multi-vibroseis	Land	Silverman	1979	Orthogonal phase-encoding
HFVS	Land	Sallas	1989	Improved phase-encoding
Two vessels	Marine	Beasley	1998	Extra source behind streamers
Dithered sources	Marine	Stefani et al.	2007	Small random fire time delays
ISS	Land/Marine	Howe, Abma	2008	Effective fire time delays
Multi-vessel coil	Marine	Moldoveanu et al.	2012	Circular source trajectory
Apparition	Marine	Robertsson et al.	2016	Periodical source modulation

Table 1.1: Summary of simultaneous source seismic data acquisition methods

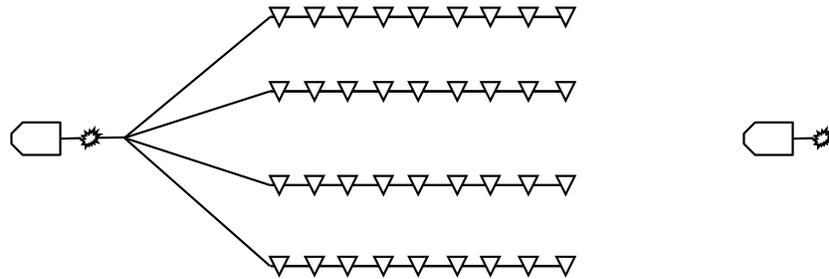


Figure 1.3: Illustration of early marine simultaneous source acquisition with two shots well separated in space (Beasley et al., 1998).

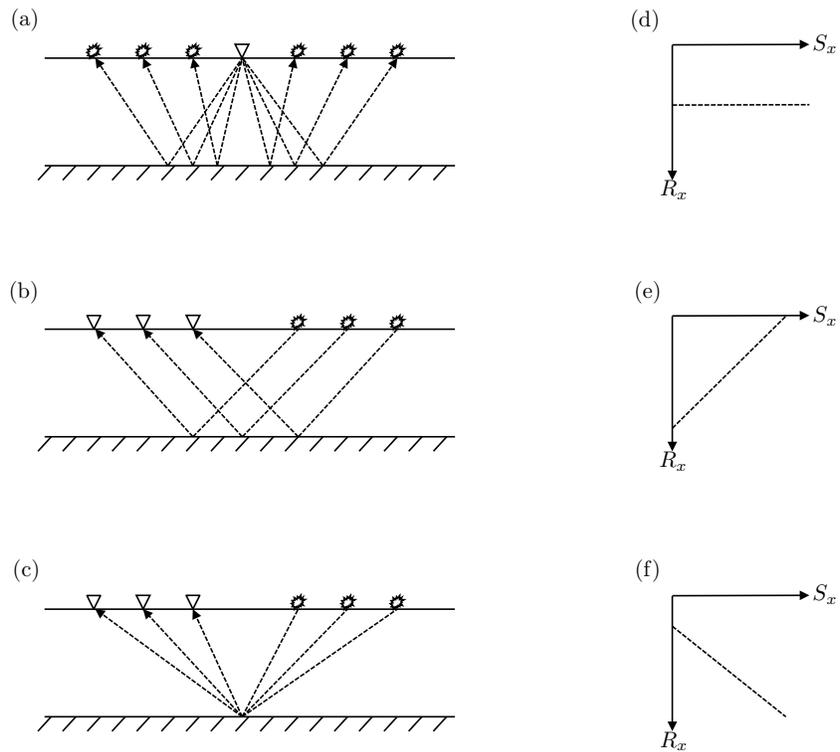


Figure 1.4: Illustration of gathers with each trace corresponding to a different source location. The figures on the left shows the source and receiver pairs of (a) common receiver, (b) common offset, and (c) common midpoint gathers. The figures on the right shows the illustration of a (d) common receiver, (e) a common offset, and (f) a common midpoint gather in 2D seismic acquisition.

1.1.2 Numerical simulation of blended acquisition

We assume seismic data acquired via a conventional seismic acquisition survey are denoted as $D(t, r, s)$, where t, r, s are used to indicate the time, receiver, and source indices, respectively. In simultaneous source acquisition, the trace recorded by the j -th receiver (r_j) can be simulated via

$$b(t, r_j) = \sum_{i \in \mathbb{S}} D_{r_j}(t - \tau_i, r_j, s_i), \quad (1.1)$$

where \mathbb{S} indicates a group of shots with fire time and location pairs (τ_i, s_i) . One can rewrite Equation 1.1 using its operator form as follows

$$\mathbf{b} = \mathcal{B} \mathbf{D}, \quad (1.2)$$

where \mathcal{B} symbolizes the blending operator (Berkhout, 2008), \mathbf{b} is the blended data and \mathbf{D} is the ideal unblended common receiver gather for the receiver j . To avoid cluttering our notation, we will drop the sub-index j and understand that the blending process operates on all receivers. Figure 1.5 illustrates the numerical blending process in the common shot domain.

Equation 1.2 provides a linear system of equations as the blending operator shifts each shot record according to the fire time and then superposes the shot records into a blended shot gather. In the context of the inverse problem, one can adopt the blending operator as the forward model operator. Estimating the unblended data from the blended observation can be achieved by minimizing the following function

$$J = \|\mathbf{b} - \mathcal{B} \mathbf{D}\|_2^2. \quad (1.3)$$

In our analysis, the problem is ill-posed as the trace recorded by a single receiver has the responses from multiple shots. There does not exist a single and direct solution. However, we are aware that below certain frequency, the problem is well-posed and over-determined (Andersson et al., 2017).

The backward operator, or the adjoint operator, of the blending operator is the pseudo-deblending operator defined by Berkhout (2008). We denote the pseudo-deblending operator as follows

$$\hat{\mathbf{D}} = \mathcal{B}^* \mathbf{b}. \quad (1.4)$$

Pseudo-deblending entails shifting the fire time delays back to each single shot and then truncating the blended shot record to the recording interval of the conventional shot. The pseudo-deblending process is demonstrated in Figure 1.6. Since pseudo-deblending extracts each shot in the blended record, the pseudo-deblended data set has the same format as the

unblended data set. Pseudo-deblended data is equivalent to the minimum norm solution of Equation 1.3 (Berkhout, 2008; Wapenaar et al., 2012) given by

$$\hat{\mathbf{D}} = \mathcal{B}^*(\mathcal{B}^*\mathcal{B})^{-1}\mathbf{b}. \quad (1.5)$$

However, as illustrated in Figure 1.6, Pseudo-deblending cannot remove source interferences (Mahdad et al., 2011; van Borselen et al., 2012).

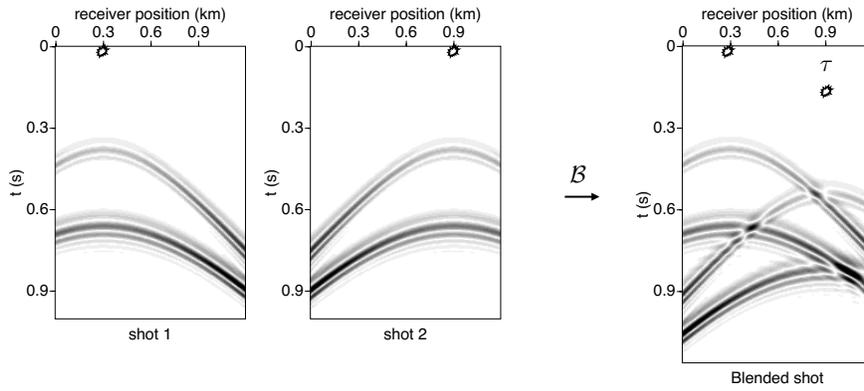


Figure 1.5: Illustration of numerical blending using two shots. The time delay for shot 2 is denoted by τ .

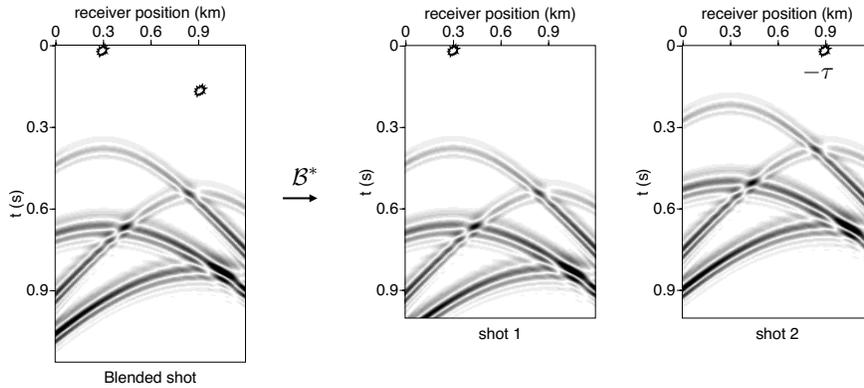


Figure 1.6: Illustration of pseudo-deblending of two shots. The time delay τ is shifted back for shot 2. However, the interferences cannot be removed by pseudo-deblending.

1.2 Review of previous work

In recent years, simultaneous source acquisition has become an important method to lower the costs of seismic acquisition. There has been a good deal of research for processing the blended data from both industry and academia. Current blended seismic data processing methods can be classified into two major categories. The groups are listed in the subsections below.

1.2.1 Direct imaging of blended data

The first category of processing methods entails applying direct imaging and waveform inversion to the blended data without special processing for the source crosstalk. On the one hand, processing steps such as stacking and prestack migration are considered as passive methods that can attenuate the source crosstalk artifacts in the migrated image (Krey, 1987; Romero et al., 1999). On the other hand, as one can propagate the wavefields for multiple shot records at the same time, using blended sources has seen advantages in reducing the computational cost of seismic imaging, especially in least-squares migration (Tang, 2007; Dai et al., 2011) and full waveform inversion (Krebs et al., 2009; Herrmann and Li, 2012; Anagaw and Sacchi, 2014). Source phase encoding schemes are designed to ensure that the stacking over all the partial images can sufficiently suppress source cross-talk artifacts (Schuster et al., 2011; Godwin and Sava, 2013). Regularization on the migrated image has also been utilized to further suppress sources interferences in least-squares migration (Dai and Schuster, 2012; Xue et al., 2014). However, direct imaging of blended data might not be the optimal solution for prestack and amplitude sensitive analyses, such as amplitude versus offsets (AVO) inversion and time-lapse seismic analyses. This is because when stacking over all the partial images, the amplitude information is lost (Ayeni et al., 2011).

1.2.2 Simultaneous source separation

The second category of methods for processing simultaneous source data entails introducing an additional processing step, which is referred to as simultaneous source separation, or deblending, into the conventional processing flow. The goal is to separate the responses from each shot and to eliminate the simultaneous source interferences. After deblending, the data should be comparable to the ideal unblended data that one could have acquired from the conventional seismic acquisition.

The separation of simultaneous sources can be treated as a blind signal separation problem. Ikelle (2007) analogies seismic source blending to the ‘cocktail party problem’ and adopted

independent component analysis (ICA) for the separation. Many researchers also treat deblending as a noise removal problem and proposed deblending methods based on the randomization of fire time delays. The pseudo-deblending process shifts the fire time delay back for each shot. In the common receiver, common offset, and common midpoint domains of the pseudo-deblended data, the desired signal would appear coherent due to the source geometric pattern. The interferences from the blended shots would appear random as they are perturbed by the randomized fire time delays. Figure 1.7 (a) exhibits the common receiver gather of the pseudo-deblended data set. Figures 1.7 (b) and (c) show the common offset gather, as well as the common midpoint gather, respectively. One can separate the desired signal from the interferences based on the coherency in all three domains.

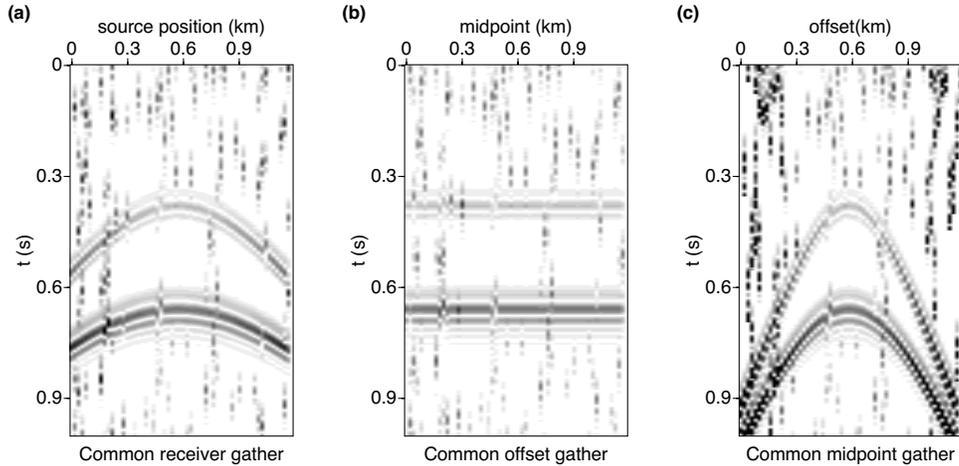


Figure 1.7: Blending noise of pseudo-deblended data set in (a) common receiver, (b) common offset, and (c) common midpoint domain. The desired signal is coherent while the blending noise appears incoherent.

Many researchers treat deblending as an noise removal problem by studying the following problem

$$\mathbf{D}^{obs} = \mathbf{D} + \mathbf{N}, \quad (1.6)$$

where $\mathbf{D}^{obs} = \mathcal{B}^* \mathbf{b}$ is the pseudo-deblended data and \mathbf{N} is called the blending noise. As is discussed above, in common receiver, common offset, and common midpoint domains, the desired unblended data is coherent while the blending noise would appear random. Incoherent noise removal methods can be used to suppress the blending noise in these domains. For instance, Huo et al. (2009) proposed to use a multi-dimensional vector median filters in common midpoint domain to suppress the interferences. Kim et al. (2009) suggested to model the blending noise in common offset domain and then subtract the modelled noise

from the data. Methods based on Radon transforms have also been suggested as a stacking tool in common receiver gathers for source crosstalk attenuation (Moore et al., 2008). Radon transforms are usually combined with sparse inversion to improve the quality of deblending (Akerberg et al., 2008; Ibrahim and Sacchi, 2014). The advantage of the de-noising based deblending methods is that the methods are usually fast and easy to apply.

Compared to the de-noising based deblending methods, deblending methods based on inversion techniques are more favorable in eliminating simultaneous source interferences. The reason is that the simultaneous source crosstalk is often as strong as, or even stronger than the desired signal. The one-time filtering scheme might not be optimal to suppress the large amplitude blending noise. Inversion methods, on the other hand, iteratively solve for an estimate of the data that honours the blending acquisition system. The solution is given by

$$\mathbf{b} = \mathcal{B}[\mathcal{C}(\mathbf{D})], \quad (1.7)$$

where \mathcal{C} is the constraint that only keeps the coherent signal in the solution (Abma et al., 2010). A variety of methods can be considered as the coherent constraint \mathcal{C} to separate simultaneous sources. For example, Mahdad et al. (2011) proposed to embed the $f-k$ filters into the adaptive subtraction inversion approach to suppress the interferences. Similarly, van Borselen et al. (2012) designed an inversion approach with the constraint that the records produced by the nearby sources should appear similar. A sparse model constraint can also be used to estimate the unblended data, if the data are transformed into an auxiliary domain, such as the Fourier domain (Abma et al., 2010), Curvelet domain (Mansour et al., 2012) and Seislet domain (Chen et al., 2014).

1.3 Contributions of this thesis

The main contributions of this thesis are summarized as follows:

- We propose an iterative rank-reduction method to the problem of deblending/separation of simultaneous sources in the field of seismic data processing. We pose deblending as a low-rank constrained inverse problem and adopt the projected gradient method to iteratively separate simultaneous sources.
- We incorporate the frequency-space domain eigenimage filtering method and Singular Spectrum Analysis method as the projection operators in the gradient projection framework for suppressing the simultaneous source interferences in common receiver and midpoint-offset data domains.

- We achieve important improvements in the computational efficiency of matrix rank reduction and Singular Spectrum Analysis via random projection and fast Hankel matrix vector products. As part of my research, I have also developed fast algorithms for rank reduction of Hankel forms.
- We provide a new method for shot-profile least-squares migration of simultaneous source seismic data. The simultaneous source crosstalk artifacts are suppressed in the shot-index common image domain.

1.4 Thesis overview

Chapter 2 derives the method of gradient descent for solving least-squares problems. We then extend the derivation to the gradient projection method which is an extension of the gradient descent method for solving constrained optimization problems. Convergence analysis, as well as the choice of step-size, are also discussed in this chapter.

Chapter 3 introduces a fast dual domain algorithm that is based on matrix rank reduction for separating simultaneous source seismic data. The proposed algorithm operates on 3D common receiver gathers. We propose an inversion scheme that minimizes the misfit between predicted and observed blended data in time domain subject to a low-rank constraint that is applied to data in the frequency-space domain. The inverse problem is solved by the gradient projection method with rank-reduction performing as a projection operator. We implement the low-rank constraint via the randomized QR decomposition. The latter allows nearly one order of processing time improvement on the truncated SVD and is less stringent on the selection of the rank of the data. We adopt synthetic and real data examples to test the performance of the proposed source separation algorithm.

In **Chapter 4**, I introduce the Singular Spectrum Analysis (SSA) filtering method as the projection filter in the gradient projection deblending framework. SSA is capable of suppressing the interferences generated by simultaneous source acquisition in small windows of common receiver gathers. Through tests with a synthetic example, we show that the interference can be effectively suppressed by the proposed method. We also show that the proposed algorithm can be modified to cope with deblending and data recovery simultaneously. A real survey acquired in the Gulf of Mexico is utilized to mimic a simultaneous source acquisition with missing shot locations. The algorithm was able to recover missing shot gathers from the blended acquisition with an improvement of the signal quality of about 12 dB. The impact of the source fire time on the source separation results are also studied in this chapter.

In **chapter 5**, we adopt the gradient projection method for the least-squares migration of the simultaneous source seismic data. We present a new scheme that computes the contributions of each single unblended shot in the migrated image of the blended data by correlating the blended receiver-side wavefield with each crosstalk free and time-shifted single source-side wavefield. The method leads to partial images that one could acquire via shot-profile migration. We notice that, in the shot-index image domain, the simultaneous source interferences are erratically distributed whereas the desired signal is coherent. This observation is used to incorporate a coherence constraint in a least-squares migration formulation of the deblending problem. We incorporate the SSA filter in the shot-index domain as the projection operator to solve for a volume of artifacts-reduced shot-index gathers that honours the observed blended data. We adopt synthetic examples to test the proposed method.

Chapter 6 contains the conclusions of this thesis. We identify the contributions and limitations of the objects of this thesis and provide the recommendations for the future work.

CHAPTER 2

Gradient projection method

The gradient projection method, or the projected gradient method, was first proposed by Goldstein (1964) and Levitin and Polyak (1966) independently for solving the following constrained optimization problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in \mathbb{C}, \end{aligned} \tag{2.1}$$

where $f(x)$ is a differentiable function that one would like to minimize. The set \mathbb{C} denotes a closed convex set in Hilbert space (Bertsekas, 1976). The method adopts the gradient descent iterations and assumes that “one can constructively project points onto the convex sets” (Goldstein, 1964). The goal is to find the intersection or the minimum distance between the set defined by the cost function $f(x)$ and the set defined by the constraint \mathbb{C} .

In this chapter, the gradient descent solution for the unconstrained least-squares problem is derived. The method is then extended to the gradient projection method for the constrained least-squares problem. Let us first consider the linear system of equations as follows

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon}, \tag{2.2}$$

where \mathbf{A} denotes a square matrix of size n that maps the model \mathbf{x} to the datum of the problem \mathbf{y} . The vector $\boldsymbol{\epsilon}$ denotes the noise caused by imperfect measurements. If \mathbf{A}^{-1} exists, Equation 2.2 has an explicit solution

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{A}^{-1}\mathbf{y} \\ &= \mathbf{A}^{-1}(\mathbf{A}\mathbf{x} + \boldsymbol{\epsilon}) \\ &= \mathbf{x} + \mathbf{A}^{-1}\boldsymbol{\epsilon}. \end{aligned} \tag{2.3}$$

The effect of the noise ϵ can be revealed using the Singular Value Decomposition (SVD) (Eckart and Young, 1936)

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{i=1}^n \sigma_i \mathbf{u}_i \mathbf{v}_i^T \quad (2.4)$$

where \mathbf{U} and \mathbf{V} are two orthogonal unitary matrices and $\mathbf{\Sigma}$ is a diagonal matrix with each element σ_i called a singular value. It can be shown that

$$\mathbf{A}^{-1} = \sum_{i=1}^n \frac{1}{\sigma_i} \mathbf{u}_i^T \mathbf{v}_i. \quad (2.5)$$

Substituting Equation 2.5 to Equation 2.3, we have

$$\hat{\mathbf{x}} = \mathbf{x} + \sum_{i=1}^n \frac{\mathbf{u}_i^T \epsilon \mathbf{v}_i}{\sigma_i}. \quad (2.6)$$

If the observation matrix \mathbf{A} is well conditioned and the noise is relatively small, the inverted solution is close to the true solution $\hat{\mathbf{x}} \approx \mathbf{x}$. However, if \mathbf{A} is ill-posed, the singular values can be much smaller than the noise. The inverted solution is not stable and the noise in the observation system will be magnified.

If \mathbf{A} is a singular matrix, the explicit solution does not exist. One can transform Equation 2.2 to the Euler form (Piana and Bertero, 1997)

$$\mathbf{A}^* \mathbf{A} \mathbf{x} = \mathbf{A}^* \mathbf{y}, \quad (2.7)$$

where \mathbf{A}^* denotes the adjoint of \mathbf{A} . Note that the noise term ϵ is omitted in Equation 2.7. The solution can be then computed using a fixed-point scheme as follows (Piana and Bertero, 1997)

$$\mathbf{x} = \mathbf{x} - \mathbf{A}^* (\mathbf{A} \mathbf{x} - \mathbf{y}). \quad (2.8)$$

Inserting a simple relaxation parameter λ to Equation 2.8 leads to the Landweber iteration (Landweber, 1951) for solving the linear problem

$$\mathbf{x}^{\nu+1} = \mathbf{x}^{\nu} - \lambda \mathbf{A}^* (\mathbf{A} \mathbf{x}^{\nu} - \mathbf{y}), \quad (2.9)$$

where \mathbf{x}^{ν} denotes the solution at the ν -th iteration. The Landweber method is an iterative method that converges to the fixed point of the linear system of equations from any initial guess. We will discuss later that the Landweber method is a particular case of the gradient descent method for solving least-squares problems.

2.1 The method of gradient descent

We now turn our attention to the minimization of an unconstrained least-squares problem given by

$$\min J = \frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|_2^2. \quad (2.10)$$

In the method of gradient descent, one starts to search for a solution that minimizes J from an arbitrary initial point \mathbf{x}_0 . As it was shown in Figure 2.1, a series of steps is taken until the estimation $\mathbf{x}^{\nu+1}$ is close enough to the desired solution. The gradient descent search usually takes the following form

$$\mathbf{x}^{\nu+1} = \mathbf{x}^\nu + \lambda^\nu \mathbf{g}^\nu, \quad (2.11)$$

where \mathbf{g}^ν denotes the search direction that one would like to take at the step ν . The parameter $\lambda^\nu > 0$ is called the step length that determines how much we want to progress along the search direction. Generally, we want to choose a direction that makes J decrease most dramatically. That is the opposite direction to the gradient of J ,

$$\mathbf{g}^\nu = -\nabla J = -\mathbf{A}^*(\mathbf{Ax}^\nu - \mathbf{y}). \quad (2.12)$$

Combining Equation 2.11 with Equation 2.12, we can find that the gradient descent solution is given by

$$\mathbf{x}^{\nu+1} = \mathbf{x}^\nu - \lambda \mathbf{A}^*(\mathbf{Ax}^\nu - \mathbf{y}). \quad (2.13)$$

We will fix the step size of the gradient search with $\lambda^\nu = \lambda$ for the analysis of convergence. Equation 2.13 is identical to the Landweber method.

2.1.1 Convergence analysis

We adopted the proof of convergence that Boyd and Mutapic (2007) discussed for the subgradient method. Instead of ensuring the cost function is decreasing at each step, we study the Euclidean distance from $\mathbf{x}^{\nu+1}$ to the optimal solution, \mathbf{x}^o , of Equation 2.10 as follows

$$\begin{aligned} \|\mathbf{x}^{\nu+1} - \mathbf{x}^o\|_2^2 &= \|\mathbf{x}^\nu - \lambda \mathbf{g}^\nu - \mathbf{x}^o\|_2^2 \\ &= \|(\mathbf{x}^\nu - \mathbf{x}^o) - \lambda \mathbf{g}^\nu\|_2^2 \\ &= \|\mathbf{x}^\nu - \mathbf{x}^o\|_2^2 + \lambda^2 \|\mathbf{g}^\nu\|_2^2 - 2\lambda \mathbf{g}^\nu (\mathbf{x}^\nu - \mathbf{x}^o). \end{aligned} \quad (2.14)$$

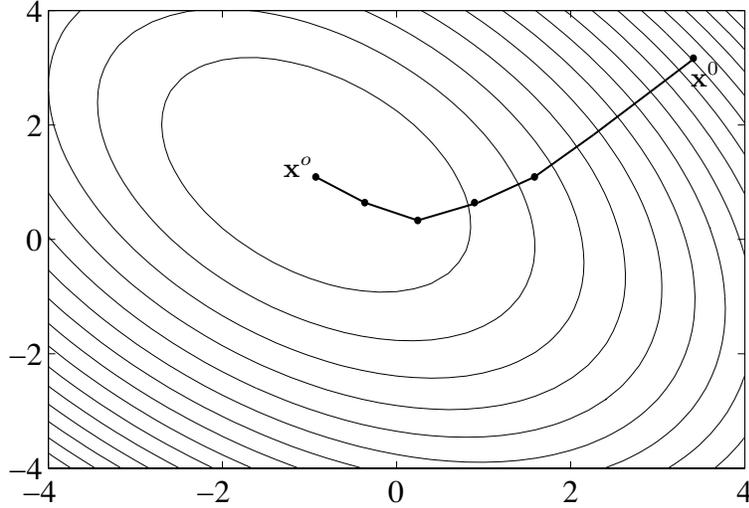


Figure 2.1: Illustration of the gradient descent method. The method starts at \mathbf{x}^0 and converges to the minimum solution \mathbf{x}^o in the contour map.

Since J is the least-squares function, we have the following inequality based on the definition of the gradient

$$J(\mathbf{x}^o) \geq J(\mathbf{x}^\nu) + \mathbf{g}^\nu (\mathbf{x}^o - \mathbf{x}^\nu). \quad (2.15)$$

Substituting Equation 2.15 into Equation 2.14, we have

$$\|\mathbf{x}^{\nu+1} - \mathbf{x}^o\|_2^2 \leq \|\mathbf{x}^\nu - \mathbf{x}^o\|_2^2 + \lambda^2 \|\mathbf{g}^\nu\|_2^2 - 2\lambda (J(\mathbf{x}^o) - J(\mathbf{x}^\nu)). \quad (2.16)$$

Equation 2.16 can be expanded recursively to the initial solution \mathbf{x}^1 as follows

$$\|\mathbf{x}^{\nu+1} - \mathbf{x}^o\|_2^2 \leq \|\mathbf{x}^1 - \mathbf{x}^o\|_2^2 + \sum_{j=1}^{\nu} \lambda^2 \|\mathbf{g}^j\|_2^2 - 2\lambda \sum_{j=1}^{\nu} (J(\mathbf{x}^o) - J(\mathbf{x}^j)). \quad (2.17)$$

The convergence of the method is guaranteed if the error is bounded by

$$J(\mathbf{x}^o) - J(\mathbf{x}^\nu) \leq \frac{\|\mathbf{x}^1 - \mathbf{x}^o\|_2^2 + \sum_{j=1}^{\nu} \lambda^2 \|\mathbf{g}^j\|_2^2}{2\lambda\nu}, \quad (2.18)$$

where $J(\mathbf{x}^\nu)$ is the minimized cost function at iteration ν . The righthand side converges to $\frac{\lambda G^2}{2}$, where G is a constant that is larger than the maximum value of the gradient.

2.1.2 The step-size rule

We now derive the step-size rule that guarantees the convergence of the gradient descent method by computing the residual of the cost function at the $(\nu + 1)$ -th iteration as follows

$$r^{\nu+1} = \|\mathbf{Ax}^{\nu+1} - \mathbf{y}\|_2^2. \quad (2.19)$$

Substituting Equation 2.13 to Equation 2.19, we have

$$\begin{aligned} \|\mathbf{Ax}^{\nu+1} - \mathbf{y}\|_2^2 &= \|\mathbf{A}[\mathbf{x}^\nu - \lambda \mathbf{A}^*(\mathbf{Ax}^\nu - \mathbf{y})] - \mathbf{y}\|_2^2 \\ &= \|(\mathbf{Ax}^\nu - \mathbf{y}) - \lambda \mathbf{AA}^*(\mathbf{Ax}^\nu - \mathbf{y})\|_2^2. \end{aligned} \quad (2.20)$$

Equation 2.20 can be further computed via

$$\begin{aligned} \|\mathbf{Ax}^{\nu+1} - \mathbf{y}\|_2^2 &= \|\mathbf{Ax}^\nu - \mathbf{y}\|_2^2 + \lambda^2 \|\mathbf{AA}^*(\mathbf{Ax}^\nu - \mathbf{y})\|_2^2 - 2\lambda \langle \mathbf{Ax}^\nu - \mathbf{y}, \mathbf{AA}^*(\mathbf{Ax}^\nu - \mathbf{y}) \rangle \\ &= \|\mathbf{Ax}^\nu - \mathbf{y}\|_2^2 + \lambda^2 \|\mathbf{AA}^*(\mathbf{Ax}^\nu - \mathbf{y})\|_2^2 - 2\lambda \|\mathbf{A}^*(\mathbf{Ax}^\nu - \mathbf{y})\|_2^2. \end{aligned} \quad (2.21)$$

We can then rearrange

$$\|\mathbf{Ax}^{\nu+1} - \mathbf{y}\|_2^2 = \|\mathbf{Ax}^\nu - \mathbf{y}\|_2^2 + \lambda \|\mathbf{A}^*(\mathbf{Ax}^\nu - \mathbf{y})\|_2^2 (\lambda \|\mathbf{A}\|_2^2 - 2). \quad (2.22)$$

To ensure that the error is decreasing, we must let

$$\|\mathbf{Ax}^{\nu+1} - \mathbf{y}\|_2^2 \leq \|\mathbf{Ax}^\nu - \mathbf{y}\|_2^2 \quad (2.23)$$

Combining Equation 2.22 with Equation 2.23, we have

$$\lambda \|\mathbf{A}\|_2^2 - 2 \leq 0. \quad (2.24)$$

Therefore, the convergence of the Landweber iteration is guaranteed when the step-size λ satisfies

$$0 < \lambda \leq \frac{2}{\|\mathbf{A}\|_2^2}. \quad (2.25)$$

The least-squares functional is not increasing when the step-size λ is small enough. In other words, when the step-size satisfies Equation 2.25, the convergence of the gradient method is guaranteed. The gradient descent method belongs to the category of first order methods for minimizing a differentiable function. The performance of the algorithm largely depends on the scaling and conditioning of the problem (Boyd and Mutapcic, 2007). Compared to the second order methods, such as the interior-point method and the Newton method, gradient descent can be much more expensive. However, as we will discuss later, the method usually requires less memory and can be immediately adapted to a variety of large-scale problems,

where constraints are applied to the solution.

2.2 The method of gradient projection

In the context of inverse problems, the use of the prior information usually plays an important role that ensures the solution is stable and physically sound. In many cases, the prior information can be exploited by requiring that the solution belongs to some set (Piana and Bertero, 1997). To incorporate such constraints, one can reformulate Equation 2.10 into the constrained least-squares problem as follows

$$\begin{aligned} \min \quad & J = \|\mathbf{Ax} - \mathbf{y}\|_2^2 \\ \text{s.t.} \quad & \mathbf{x} \in \mathbb{C}. \end{aligned} \tag{2.26}$$

The gradient projection method is an extension of the gradient descent method for solving Equation 2.14. The formulation of the gradient projection method is given by

$$\begin{aligned} \mathbf{z}^\nu &= \mathbf{x}^\nu - \lambda \mathbf{A}^* (\mathbf{Ax}^\nu - \mathbf{y}) \\ \mathbf{x}^{\nu+1} &= \mathcal{P}_C[\mathbf{z}^\nu]. \end{aligned} \tag{2.27}$$

In iteration ν , we search the solution along the gradient descent direction to acquire an estimation \mathbf{z}^ν . The estimation is then projected by its Euclidean projection \mathcal{P}_C to ensure that the solution belongs to set \mathbb{C} . If the projection is a linear map within the same set, the method reduces to the conditional gradient method by Frank and Wolfe (1956).

The convergence analysis and the step-size rule for the gradient descent method can be extended to the gradient projection method (Eicke, 1992; Boyd and Mutapcic, 2007). One can compute the geometrical distance from \mathbf{z}^ν to the optimal solution \mathbf{x}^o as follows

$$\|\mathbf{z}^\nu - \mathbf{x}^o\|_2^2 \leq \|\mathbf{x}^\nu - \mathbf{x}^o\|_2^2 + \lambda^2 \|\mathbf{G}^\nu\|_2^2 - 2\lambda (J(\mathbf{x}^o) - J(\mathbf{x}^\nu)). \tag{2.28}$$

The proof of convergence proceeds exactly as the gradient descent method if the projection \mathcal{P}_C satisfies the following condition

$$\|\mathbf{x}^{\nu+1} - \mathbf{x}^o\|_2^2 = \|\mathcal{P}_C[\mathbf{z}^\nu] - \mathbf{x}^o\|_2^2 \leq \|\mathbf{z}^\nu - \mathbf{x}^o\|_2^2. \tag{2.29}$$

Equation 2.29 ensures that the projection moves the gradient estimation closer to the optimal solution and it is naturally satisfied when \mathbb{C} is a convex set (Eicke, 1992). However, a variety of projection operators can be selected based on the constraint that one would like to impose to the least-squares problem. In addition, the gradient projection method can be extended

to operators as follows

$$\begin{aligned}\mathbf{z}^\nu &= \mathbf{x}^\nu - \lambda \mathcal{A}^*(\mathcal{A}\mathbf{x}^\nu - \mathbf{y}) \\ \mathbf{x}^{\nu+1} &= \mathcal{P}_C[\mathbf{z}^\nu],\end{aligned}\tag{2.30}$$

where \mathcal{A} is a linear and continuous operator that maps \mathbf{x} into the observation \mathbf{y} . The operator \mathcal{A}^* denotes the adjoint operator that takes \mathbf{y} back to the domain of \mathbf{x} . In this situation, the step size λ needs to satisfy the following condition

$$0 < \lambda \leq \frac{2}{\alpha}.\tag{2.31}$$

where α is the maximum eigenvalue of $\mathcal{A}^*\mathcal{A}$ that can be computed via the power iterations (Golub and van Loan, 1996).

CHAPTER 3

Deblending via iterative rank reduction¹

3.1 Introduction

Research on seismic signal processing via rank reduction techniques has been applied to the enhancement of signal-to-noise ratio of seismic data. For instance, Freire and Ulrych (1988) illustrated how matrix rank reduction methods could be utilized to eliminate incoherent noise from seismic records in time domain. A related family of methods, the Karhunen-Loeve transform, has also been introduced to improve the signal-to-noise ratio of prestack gathers (Al-Yahya, 1991). Trickett (2003) proposed the $f - x - y$ eigenimage filtering method that applies matrix rank reduction in the frequency-space domain for de-noising dipping events. Kreimer and Sacchi (2012) extended the $f - x - y$ eigenimage filtering method to multi-dimensional seismic data by a tensor rank reduction method called High-order SVD. The assumption of the rank-reduction based de-noising methods is that in the offset-midpoint domain, the ideal noiseless data can be represented via a superposition of plane waves. Only the first K eigenimages are needed to express the desired data for a section containing K dipping events (Trickett, 2003). Since incoherent noise spreads equally along all the singular values, a truncated SVD that only keeps the K leading singular values is effective for suppressing the incoherent noise.

In this chapter, we propose an iterative rank reduction method for the separation of simultaneous source data. Notably, we pose deblending as a rank constrained minimization problem. The objective function is a least-squares functional that ensures the deblended estimated data reproduce the acquired blended data. A low-rank constraint is applied to the deblended data in the frequency-space domain for preserving the coherent signal while

¹A version of this chapter is published in Cheng J. and M. D. Sacchi, 2016, Fast dual-domain reduced-rank algorithm for 3D deblending via randomized QR decomposition, GEOPHYSICS 81(1): V89-V101

suppressing the incoherent simultaneous source crosstalk. The solution of the problem is obtained via the gradient projection method with matrix rank reduction acting as a projection operator. Also, computing matrix rank reduction via the Singular Value Decomposition could result in an expensive algorithm, especially for large-scale problems (Golub and van Loan, 1996). We present a fast algorithm named randomized QR decomposition for computing the low-rank approximation (Halko et al., 2011; Chiron et al., 2014). Our synthetic examples simulated with a 3D VSP data set, as well as from a real seismic data set, show that the proposed algorithm can suppress the crosstalk generated by simultaneous source acquisition. Our research clearly demonstrates the advantage of adopting the randomized QR decomposition as a strategy to improve the computational cost of iterative rank-reduction algorithms.

3.2 Theory

3.2.1 The low-rank constraint for deblending

We study the low-rank property of common receiver gathers of multi-dimensional seismic data in the frequency-space domain. We remind the readers that the desired unblended data in time domain $d(t, x_l, y_l)$ can be written in terms of its temporal domain Fourier transform as follows

$$d(t, x_l, y_l) = \int \tilde{D}(\omega, x_l, y_l) e^{i\omega t} d\omega, \quad l \in \mathbb{S}, \quad (3.1)$$

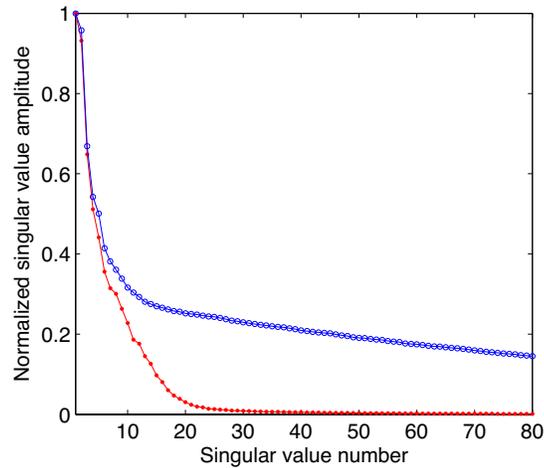
where $i = \sqrt{-1}$. As we consider a regular distribution of sources in the $x - y$ plane, at a given monochromatic frequency ω , one can express $\tilde{D}(\omega, x_l, y_l)$ in terms of a spectral matrix \mathbf{D}_ω of size $N_{S_x} \times N_{S_y}$, where the total number of sources is given by $N_S = N_{S_x} \times N_{S_y}$. The Singular Value Decomposition (SVD) of \mathbf{D}_ω is given by

$$\mathbf{D}_\omega = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H, \quad (3.2)$$

where \mathbf{U} and \mathbf{V} are matrices with orthonormal columns and $\mathbf{\Sigma}$ is a diagonal matrix with elements called the singular values of the matrix (Golub and van Loan, 1996). Figure 3.1 shows the distribution of singular values of the ideal spectral matrix at 20 Hz (blue). We also portray in red the singular value distribution of the spectral matrix from a pseudo-deblended common receiver gather at the same frequency. The desired signal which is coherent in the common receiver domain contributes to the k largest singular values whereas interferences due to random time delays boost up all the singular values. In other words, the coherence pass constraint for deblending can be implemented by a low-rank constraint applied to the pseudo-deblended data. However, in the common receiver domain, the number of singular

values k that one would like to keep does not equal to the number of dipping events. Moreover, since the interferences caused by simultaneous source acquisition are as strong as, or even stronger than the desired reflections, a one-time rank reduction strategy fails to remove the crosstalk completely.

(a)



(b)

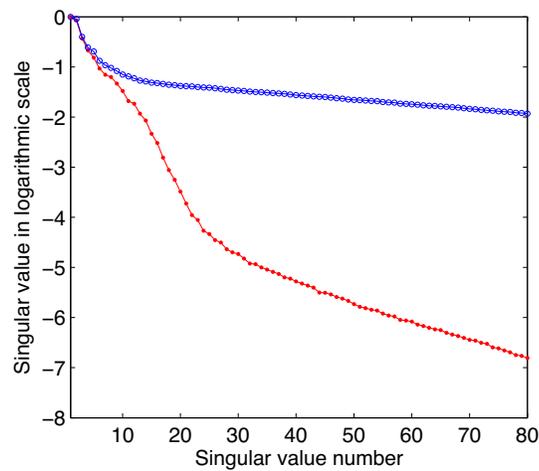


Figure 3.1: Distribution of singular values of spatial data in normal (a) and logarithmic (b) scale at 20Hz from the true unblended common receiver gather in $f - x - y$ domain (red). We also portray the distribution of singular values for data contaminated with source interferences in a pseudo-deblended common receiver gather (blue).

3.2.2 Rank constrained minimization

To iteratively suppress the simultaneous source interferences, we propose an inversion scheme that minimizes the rank of the aforescribed spectral matrices while honouring the blended acquisition. Combining Equation 1.2 and Equation 3.1 in the operator form, we have

$$\mathbf{b} = \mathcal{B}\mathcal{F}^{-1}\tilde{\mathcal{D}}, \quad (3.3)$$

where \mathcal{F}^{-1} denotes the inverse temporal Fourier transform that maps data in frequency domain to time domain. The forward Fourier transform that transforms data from time domain to frequency domain is denoted by \mathcal{F} . We use the tensor notation $\tilde{\mathcal{D}}$ to denote the $f-x-y$ domain data cube. Considering the low rank constraint for deblending, the desired unblended data is estimated by

$$\begin{aligned} \forall \omega : \min \text{rank}(\mathbf{D}_\omega), \\ \text{s.t. } \mathbf{b} = \mathcal{B}\mathcal{F}^{-1}\tilde{\mathcal{D}}, \end{aligned} \quad (3.4)$$

or by the equivalent form

$$\begin{aligned} \min J = \|\mathbf{b} - \mathcal{B}\mathcal{F}^{-1}\tilde{\mathcal{D}}\|_2^2, \quad \text{s.t.} \\ \forall \omega : \mathbf{D}_\omega \in C(K) = \{\mathbf{D}_\omega : \text{rank}(\mathbf{D}_\omega) \leq K\}. \end{aligned} \quad (3.5)$$

where K is the desired rank of \mathbf{D}_ω . The low-rank constraint is imposed to each frequency slice of the $f-x-y$ data cube. However, the objective function is computed by minimizing the misfit function between observations and predicted blended data in $t-x-y$ domain.

We show that the rank minimization problem in Equation 3.5 can be tackled via the gradient projection method. We consider to successively update a current estimate $\tilde{\mathcal{D}}^\nu$ to minimize the objective function. This is done by modifying $\tilde{\mathcal{D}}^\nu$ in the opposite direction of the gradient

$$\tilde{\mathcal{D}}^{\nu+1} = \tilde{\mathcal{D}}^\nu - \lambda \mathcal{F}\mathcal{B}^*(\mathcal{B}\mathcal{F}^{-1}\tilde{\mathcal{D}}^\nu - \mathbf{b}). \quad (3.6)$$

The optimal solution to Equation 3.5 is also the optimal solution to the following cost function

$$\begin{aligned} J_2 = \|\tilde{\mathcal{D}} - \tilde{\mathcal{D}}^{\nu+1}\|_2^2, \quad \text{s.t.} \\ \forall \omega : \mathbf{D}_\omega \in C(K) = \{\mathbf{D}_\omega : \text{rank}(\mathbf{D}_\omega) \leq K\}, \end{aligned} \quad (3.7)$$

only if $\tilde{\mathcal{D}}^{\nu+1}$ converges to the optimal solution of the least-squares functional in Equation 3.5 (Ye and Ji, 2009; Cai et al., 2010; Ma et al., 2011). At a given temporal frequency ω ,

Equation 3.7 reduces to

$$\begin{aligned} J(\omega) &= \|\mathbf{D}_\omega - \mathbf{D}_\omega^{\nu+1}\|_2^2 \\ \text{s.t. } \mathbf{D}_\omega &\in C(K) = \{\mathbf{D}_\omega : \text{rank}(\mathbf{D}_\omega) \leq K\}. \end{aligned} \quad (3.8)$$

Equation 3.8 entails finding a low rank approximation of $\mathbf{D}_\omega^{\nu+1}$ at a given frequency ω . The classic solution is the well studied truncated SVD (tSVD). If we use \mathcal{P}_r to denominate a projection operator that projects each frequency slice of data to a low-rank matrix, then the solution for separating the simultaneous sources can be expressed via

$$\tilde{\mathbf{D}}^{\nu+1} = \mathcal{P}_r[\tilde{\mathbf{D}}^\nu - \lambda \mathcal{F} \mathcal{B}^* (\mathcal{B} \mathcal{F}^{-1} \tilde{\mathbf{D}}^\nu - \mathbf{b})]. \quad (3.9)$$

This gradient projection algorithm entails searching for a solution in the gradient descent direction. The solution is then projected to a set of low-rank matrices. In practice, it is more convenient to adapt the above algorithm to the following form

$$\begin{aligned} \mathcal{Z} &= \mathcal{D}^\nu - \lambda \mathcal{B}^* (\mathcal{B} \mathcal{D}^\nu - \mathbf{b}), \\ \mathcal{D}^{\nu+1} &= \mathcal{F}^{-1} \mathcal{P}_r \mathcal{F} [\mathcal{Z}] = \mathcal{P}_{fr} [\mathcal{Z}]. \end{aligned} \quad (3.10)$$

We remind the reader that \mathcal{D}^ν and, therefore \mathcal{Z} are deblended data at iteration ν in the $t-x-y$ domain. However, the rank-reduction constraint \mathcal{P}_{fr} must be applied in the $f-x-y$ domain. The algorithm is initialized with the pseudo-deblended data $\mathcal{B}^* \mathbf{b}$. The reason is that the pseudo-deblended data contain exactly the information of the desired unblended signal. The convergence of the algorithm is guaranteed when $\lambda < 2/\alpha$, where α is the maximum eigenvalue of the operator $\mathcal{B}^* \mathcal{B}$ (Ma et al., 2011). The latter can be computed via the power method. Algorithm 1 shows the framework of the gradient projection algorithm for the low-rank constrained inversion method.

Algorithm 2 illustrates the projection operator \mathcal{P}_{fr} . We transform the current estimation \mathcal{Z} in $t-x-y$ domain to data $\tilde{\mathcal{Z}}$ in $\omega-x-y$ domain. Then, for each temporal frequency ω , we perform rank-reduction on the spatial data \mathbf{Z}_ω . Before introducing the fast rank reduction method based on random projection, the conventional truncated SVD (tSVD) is utilized. The latter entails keeping the largest k singular values of \mathbf{Z}_ω while setting the other singular values to zero. The data $\hat{\mathbf{Z}}_\omega$ are then reconstructed with the new set of singular values. To continue with the algorithm, the rank-reduced data, $\hat{\tilde{\mathcal{Z}}}$, in $\omega-x-y$ domain is transformed back to the time domain to obtain a new estimate of the deblended data $\mathcal{D}^{\nu+1}$. Therefore the gradient descent algorithm operates in $t-x-y$ domain whereas the rank constraints are applied in the $\omega-x-y$ domain.

Algorithm 1 Dual Domain Rank Reduction Deblending Algorithm

Inputs:

Blending operator \mathcal{B} and its adjoint \mathcal{B}^*
 Observed blended trace \mathbf{b}
 Stopping criterion ϵ
 Step size λ

Initialize:

$\mathcal{D}^0 = \mathcal{B}^* \mathbf{b}$; $\nu = 0$;

repeat

$\mathcal{Z} = \mathcal{D}^\nu - \lambda \mathcal{B}^* (\mathcal{B} \mathcal{D}^\nu - \mathbf{b})$

$\nu = \nu + 1$

$\mathcal{D}^\nu = \mathcal{P}_{fr}[\mathcal{Z}]$ (See Algorithm 2 and 3)

until $\|\mathbf{b} - \mathcal{B} \mathcal{D}^\nu\|_2^2 < \epsilon$

$\mathcal{D} = \mathcal{D}^\nu$

Algorithm 2 Projection operator for tSVD $\mathcal{P}_{fr}^K[\mathcal{Z}]$:

Initialize:

$\tilde{\mathcal{Z}} \leftarrow \mathcal{Z}$ (transform to frequency domain)

for $\omega = \omega_{min} : \omega_{max}$ **do**

$[\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}] = \text{svd}[\mathbf{Z}_\omega]$

if $k \leq K$ **then**

$\hat{\Sigma}_{k,k} = \Sigma_{k,k}$

else

$\hat{\Sigma}_{k,k} = 0$

end if

$\hat{\mathbf{Z}}_\omega = \mathbf{U} \hat{\Sigma} \mathbf{V}^H$

end for

$\mathcal{D}^\nu \leftarrow \hat{\mathcal{Z}}$ (transform back to time)

3.2.3 Randomized QR decomposition

Finding the low-rank approximation of a given matrix is ubiquitous in the areas of applied mathematics, numerical analysis and a variety of scientific computing areas (Liberty et al., 2007). The conventional SVD method requires an order of $O(mn^2)$ operations, where m and n denote the size of a given matrix. Alternative rapid rank reduction methods, such as Lanczos bidiagonalization and randomized SVD, were recently applied for seismic data

reconstruction and denoising (Oropeza and Sacchi, 2011; Gao et al., 2013). In this article, we present a method named randomized QR decomposition (RQRD) to improve the efficiency of matrix rank reduction. The idea of using dimensionality reduction for matrix approximation was proposed by Papadimitriou et al. (2000), where he suggested compressing the range of an input matrix with a random subspace. Rokhlin et al. (2009) then developed an efficient dimensionality reduction algorithm using a Gaussian transformation matrix. They also combined dimensionality reduction with the power iteration method that leads to a very efficient method for large scale problems. Halko et al. (2011) provides a comprehensive review of matrix low-rank approximations based on random projection.

Let us project the spectral matrix, \mathbf{Z}_ω , by a set of P random normalized vectors given by the columns of the matrix $\mathbf{\Omega}$:

$$\mathbf{M}_{N_{S_x} \times P} = \mathbf{Z}_\omega_{N_{S_x} \times N_{S_y}} \mathbf{\Omega}_{N_{S_y} \times P}. \quad (3.11)$$

Owing to the randomness, the vectors in the matrix \mathbf{M} are linearly independent. Since the unblended data in \mathbf{D}_ω are low rank, only a number of P random vectors will be required to span the full range of the desired signal (Appendix A). As $P \ll N_{S_y}$, the random projection reduces the size of matrix for rank reduction. Then we compute the orthonormalized basis \mathbf{Q} with the economy size QR decomposition of the matrix \mathbf{M}

$$\mathbf{Q}_{N_{S_x} \times P} \mathbf{R}_{P \times P} = \mathbf{M}_{N_{S_x} \times P}. \quad (3.12)$$

The low rank approximation is computed via the following expression

$$\hat{\mathbf{Z}}_\omega = \mathbf{Q}\mathbf{Q}^H \mathbf{Z}_\omega. \quad (3.13)$$

The RQRD procedure is repeated for each frequency slice in the $f - x - y$ common receiver cube, which makes the RQRD projection operator shown in algorithm 3. Unlike the truncated SVD method which directly solves for the closest low-rank approximation of a given matrix, the random projection methods does not constrain the rank strongly like the SVD. In other words, in RQRD, a subset size equal to the exact rank of the given matrix usually cannot ensure that the solution is a rank K approximation. We consider the size of the random subset, P , as a relaxation of the exact rank K of the matrix (Halko et al., 2011). P is usually determined by factors including the size of the matrix, the singular spectrum and the type of random matrix. In practice, we usually select $P = [1.5K, 3K]$ (Chiron et al., 2014). This usually allows us to achieve better results when the singular values do not decay dramatically and when the precise rank of the matrix is not known. Figure 3.2 shows the comparison of processing time for matrix rank reduction using tSVD (blue) and

RQRD (red). In this example, we choose P equal $3K$ and the RQRD algorithm is about 10 times faster than the conventional tSVD.

Algorithm 3 Projection operator for RQRD $\mathcal{P}_{fr}^P[\mathcal{Z}]$:

Initialize:

$\tilde{\mathcal{Z}} \leftarrow \mathcal{Z}$ (transform to frequency domain)

for $\omega = \omega_{min} : \omega_{max}$ **do**

$\mathbf{M} = \mathbf{Z}_\omega \mathbf{\Omega}$ (random projection)

$[\mathbf{Q}, \mathbf{R}] = \text{qr}[\mathbf{M}]$

$\hat{\mathbf{Z}}_\omega = \mathbf{Q}\mathbf{Q}^H \mathbf{Z}_\omega$

end for

$\mathcal{D}^\nu \leftarrow \hat{\mathcal{Z}}$ (transform back to time)

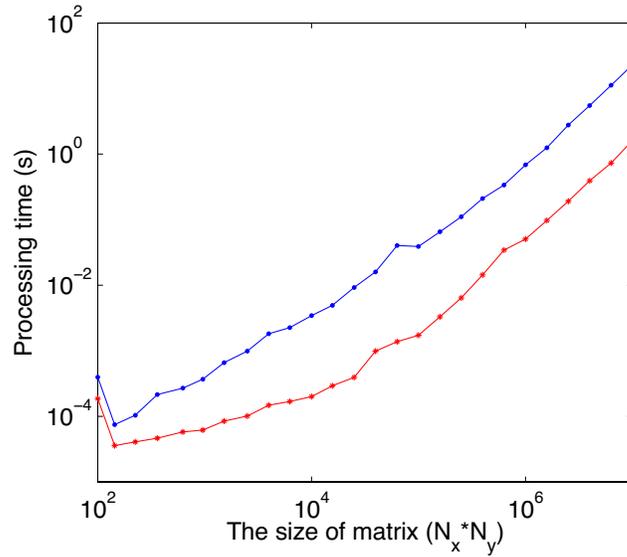


Figure 3.2: The processing time of rank reduction versus the size of matrix both in logarithmic scale. The blue curve shows the processing time utilizing the truncated SVD while the red is the processing time of the RQRD method. We choose P equal $3k$ and the RQRD algorithm is about 10 times faster than the conventional tSVD.

3.3 Examples

3.3.1 Example with synthetic 3D VSP data set

We use a synthetic 3D vertical seismic profile (VSP) data set to mimic the process of simultaneous source acquisition. The data set contains 205 source lines with 205 source positions on each line. The interval of each source position is 16.67m, and the line spacing is also 16.67m (Figure 3.3). In this example, 31 downhole detectors are deployed at depth 1350m to 1850m with 16.67m intervals in the centre of this $x - y$ grid. The model is structurally simple without near surface complexity (O'Brien, 2010). The highest frequency that contains useful information is about 40Hz as a 15.4Hz Ricker is utilized to generate data.

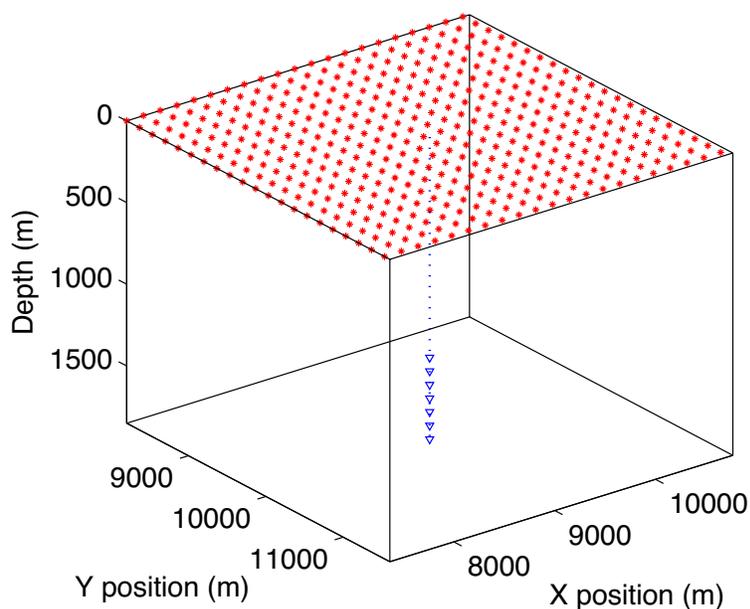


Figure 3.3: Distribution of sources and receivers of the 3D VSP data set. The sources are deployed on a regular grid in the surface and the receivers are deployed between 1350m and 1850m in the subsurface. Each red point represents a group of 10 sources and each blue triangle represents 5 receivers.

The sources are blended in an acquisition corresponding to the single-vessel simultaneous source acquisition where we assume that seismic shots are fired by one vessel. Since the air gun sources are impulsive, the vessel keeps traveling without waiting for new records until it covers the whole survey area. As a result, only the adjacent sources are blended with a given time delay τ . This acquisition design resembles the self-simultaneous shooting except that we extend the concept to 3D seismic acquisition (Abma et al., 2012). Although our data configuration it is not realistic, it allows studying the relationship between firing time delays and the performance of deblending algorithms. A more realistic scenario requires more than one vessel firing at random intervals at different $x - y$ locations as described by Moldoveanu et al. (2012). We measure the efficiency of this acquisition in terms of the survey time ratio (STR), which is defined by the ratio of conventional acquisition time and blended acquisition time (Berkhout, 2008). We understand that our measurement of acquisition time does not consider many realistic aspects in a real seismic survey. However, it allows us to measure how much the data has been compressed by the blended acquisition. For instance, in Figure 3.4, the STR equals to 2 and the acquisition time with blended sources is 50% of the conventional acquisition. The quality of deblending is measured in dB units via

$$Q_S = 10 \log \frac{\|\mathbf{d}_{true}\|_2^2}{\|\mathbf{d}_{true} - \mathbf{d}_S\|_2^2}, \quad (3.14)$$

where \mathbf{d}_{true} is the true synthetic data from a conventional common receiver gather and \mathbf{d}_S stands for the separated common receiver gather via iterative rank reduction. A large Q_s corresponds to fewer simultaneous source interferences in the deblended results.

Performance of the algorithms

We tested the effectiveness of the deblending algorithm under different rank (tSVD), as well as different subset size P in the RQRD algorithm. In this example, we fixed the firing time delay to $STR = 10$. Figure 3.5 shows the quality of deblending versus rank. It can be observed that the tSVD method (blue) presents the highest quality after deblending only when the selected rank is very close to the exact rank K of the data. In other words, prior information would be required in order to choose the optimal rank. The RQRD method (red), on the other hand, exhibits a broad region of high deblending quality. We can achieve reasonable results when the subset size is in the range $P \in [1.5K, 5K]$. The test provides evidence that the subset size P in the RQRD algorithm is a relaxation of the desired rank K .

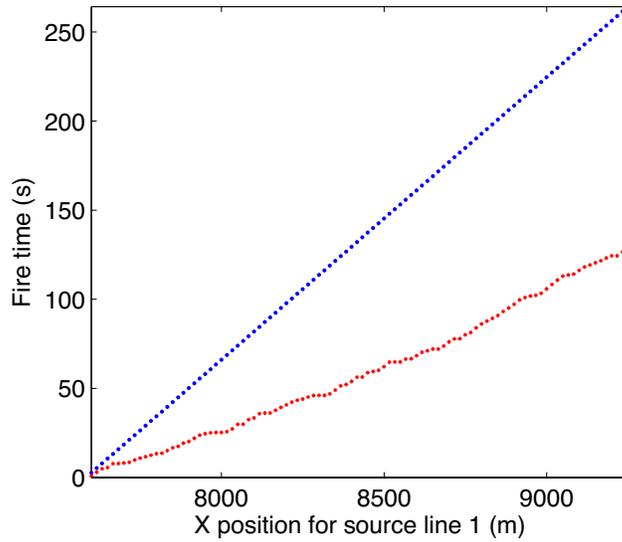


Figure 3.4: Distribution of the firing time of seismic sources in the conventional acquisition (Blue) and the one-vessel simultaneous source acquisition (Red). In this example, STR equals 2 and 50 % of acquisition time is saved by simultaneous source acquisition

Effect of rank and STR

It is important to stress that unlike the other rank reduction based deblending methods that operate in offset-midpoint domain (Maraschini et al., 2012; Wason et al., 2014), where the rank is equivalent to the number of plane waves, in common receiver domain there is no theoretical justification for selecting an optimal rank. We choose to use the RQRD method to test the effects of different rank and the survey time ratio on the proposed deblending algorithm.

In this experiment, we used a STR up to 25 and then tested a wide selection of subset sizes from 1 to 120. For each specific rank and STR , we generated 50 realizations with the same distribution for setting the firing schedule. For each trial, we adopted a relatively small step size to ensure the convergence of the algorithm. When the quality factor after separation, Q_s , reaches a number higher than 20 dB, we consider the proposed algorithm successfully separated the responses from the blended sources. The definition of success can also be interpreted as the percentage error as follows

$$\frac{\|\mathbf{d}_{true} - \mathbf{d}_S\|_2^2}{\|\mathbf{d}_{true}\|_2^2} \leq 0.01. \quad (3.15)$$

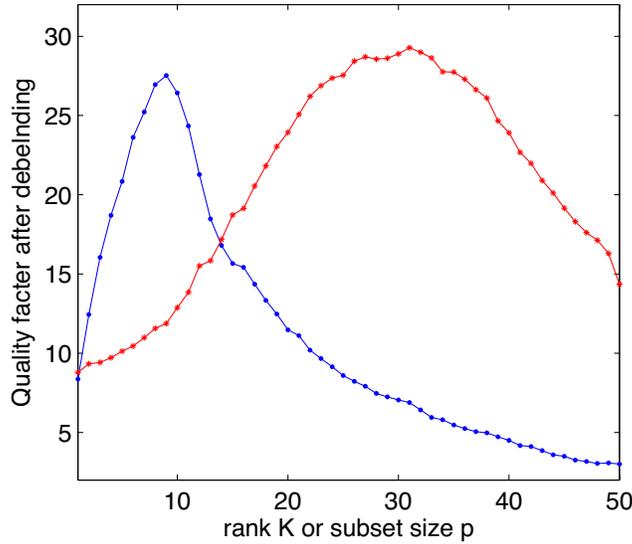


Figure 3.5: The quality of deblending versus rank using both tSVD and RQRD. The blue curve shows the results utilizing the truncated SVD as the low rank projection operator while the red is the results corresponding to the RQRD method. The survey time ratio is 10 and the firing time remains the same for all trials. Compared to the truncated SVD which relies on the selection of rank K , a broad range of subset sizes in RQRD, $p \in [1.5K, 5K]$, ensures the success of deblending.

In other words, a 1% relative means squared error is our threshold for a successful run. Figure 3.6, which is very similar to the Tanner-Donoho plot (Donoho and Tanner, 2009), shows the percentage of successful runs regarding different selection of rank and STR. At a given STR and rank, the white colour means all 50 runs of the algorithm successfully removed the interferences. In contrast, a point in the dark area indicates a combination of rank and STR for which all the trials of the deblending algorithm failed to improve the data quality above 20 dB. It is clear from the figure that for a relatively small STR, a broad range of rank could be adopted to ensure successful separations. However, as the STR grows, we need to select the optimal rank for simultaneous source separation. Also, there always exists a limitation of the blended acquisition regarding the source fire scheme. This is because when the shots are fired in a very condensed fashion, the source interferences will no longer appear incoherent in the common receiver, common offset, and common midpoint domains. For this particular model and acquisition design, the algorithm fails to separate sources when STR is greater than 21.

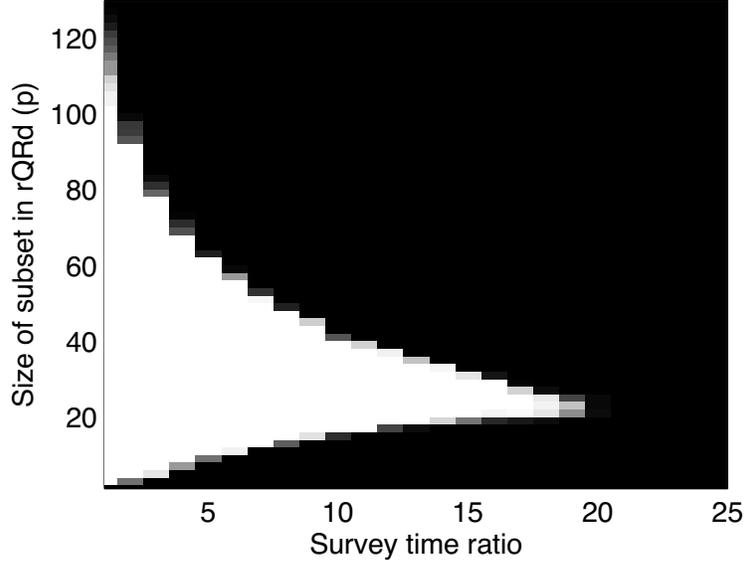


Figure 3.6: Probability map for different rank and survey time ratio. For each specific rank and survey time ratio, 50 realizations of random firing time delays were generated with uniform distribution. If the quality factor after deblending (Q_s) is larger than 20, we consider the method successfully removed the simultaneous source crosstalk. The white area in the figure indicates an area where the proposed algorithm succeeded for all 50 trials.

We show the deblending results for the RQRD method as well. In this case, we used different STRs but fixed the values of the parameter $P = 24$. The proposed algorithm has eliminated the noise and re-established a distribution of singular values similar to the distribution of singular values of the ideal data. In the first example (Figure 3.7-3.9), the STR equals to 2. Figure 3.7 shows a time slice of the results after 25 iterations. Figure 3.8 shows the deblending result for the centre receiver and Figure 3.9 shows the result for the centre shot. Interferences from simultaneously fired shots are effectively suppressed. We improve the quality factor of the pseudo-deblended dataset to 36.5 dB. As a result, the unblended solution becomes comparable with the true shot record. Then, the STR continues to increase by a factor of 10. Figure (3.10-3.12) show the deblending results in different gathers. The deblending algorithm is still robust with a quality of deblending improved to 28.4 dB. When the STR reaches a factor of 20 (Figure 3.13-3.15), we start to see leaking energy in the difference panels, and the quality of deblending reduces to 20.2 dB.

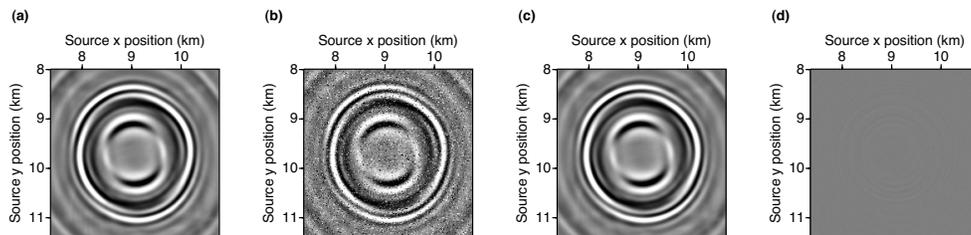


Figure 3.7: Results of simultaneous source separation in common receiver domain when STR equals to 2: (a) The real unblended time slice at 1.2 s. (b) Pseudo-deblended time slice. (c) Deblended time slice after 20 iterations of the proposed algorithm. (d) Differences between (a) and (c). In this example, the signal-to-noise ratio after separation is 36.5dB.

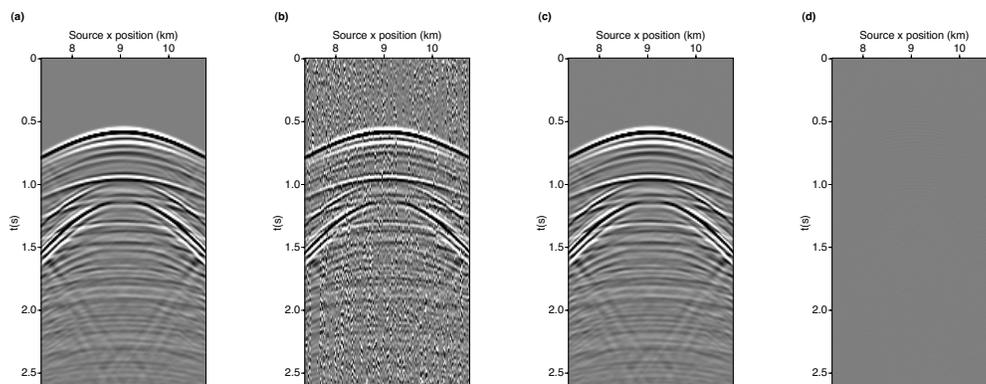


Figure 3.8: Results of simultaneous source separation in common receiver domain when STR equals to 2: (a) The real unblended common receiver gather (centre receiver). (b) Pseudo-deblended common receiver gather. (c) Deblended common receiver gather after 20 iterations. (d) Differences between (a) and (c). In this example, the signal-to-noise ratio after separation is 36.5dB.

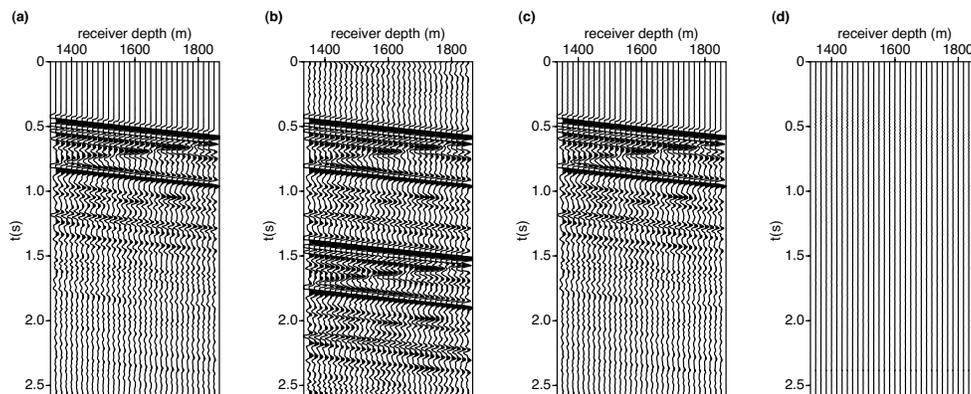


Figure 3.9: Results of simultaneous source separation in common shot domain when STR equals to 2: (a) The real unblended common shot gather (centre shot). (b) Pseudo-deblended shot record. (c) Deblended shot record after 20 iterations. (d) Differences between (a) and (c). In this example, the signal-to-noise ratio after separation is 36.5dB.

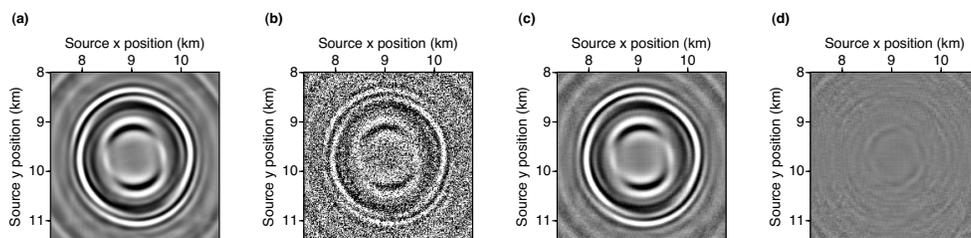


Figure 3.10: Results of simultaneous source separation in common receiver domain when STR equals to 10: (a) The real unblended time slice at 1.2 s. (b) Pseudo-deblended time slice. (c) Deblended time slice via the proposed algorithm. (d) Differences between (a) and (c). In this example, the signal-to-noise ratio after separation is 28.4dB.

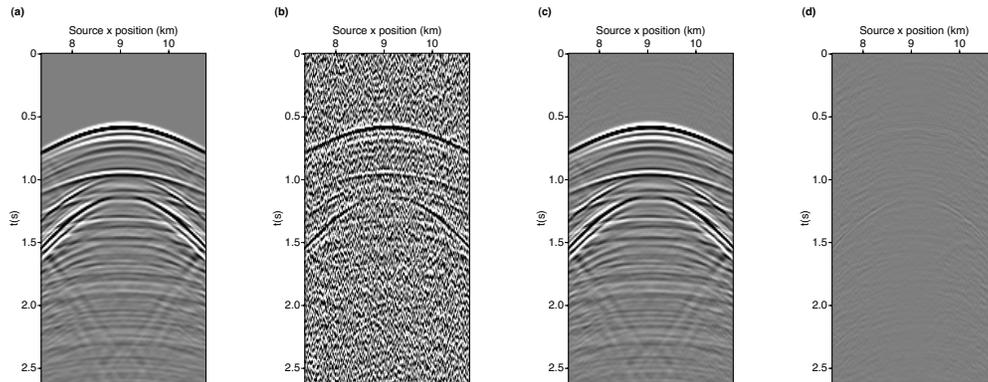


Figure 3.11: Results of simultaneous source separation in common receiver domain when STR equals to 10: (a) The real unblended common receiver gather (centre receiver). (b) Pseudo-deblended common receiver gather. (c) Deblended common receiver gather via the proposed algorithm. (d) Differences between (a) and (c). In this example, the signal-to-noise ratio after separation is 28.4dB.

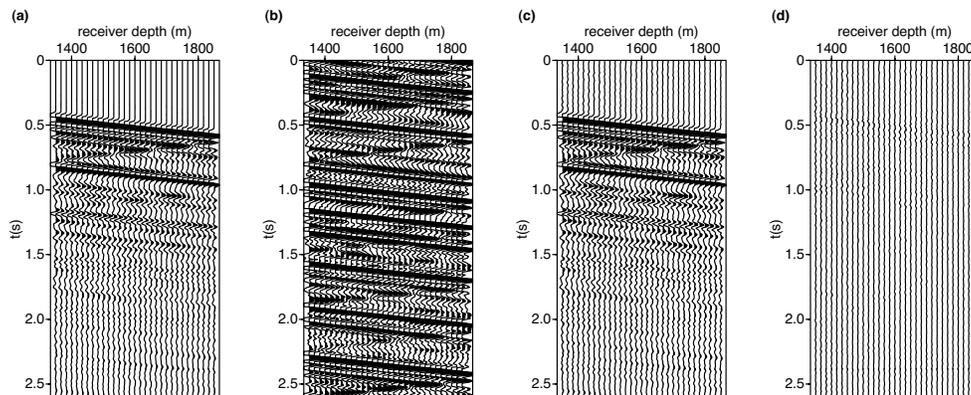


Figure 3.12: Results of simultaneous source separation in common shot domain when STR equals to 10: (a) The real unblended common shot gather (centre shot). (b) Pseudo-deblended shot record. (c) Deblended shot record via the proposed algorithm. (d) Differences between (a) and (c). In this example, the signal-to-noise ratio after separation is 28.4dB.

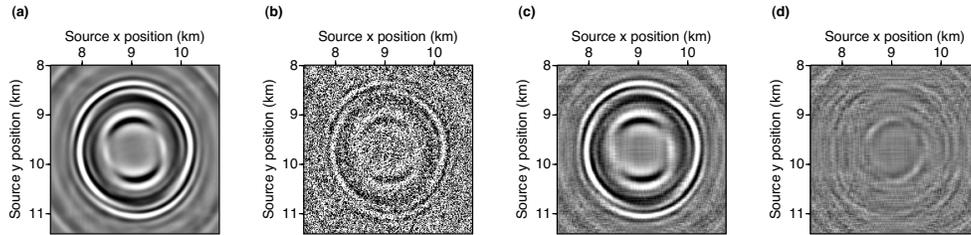


Figure 3.13: Results of simultaneous source separation in common receiver domain when STR equals to 20: (a) The real unblended time slice at 1.2 s. (b) Pseudo-deblended time slice. (c) Deblended time slice via the proposed algorithm. (d) Differences between (a) and (c). In this example, the signal-to-noise ratio after separation is 20.2dB.

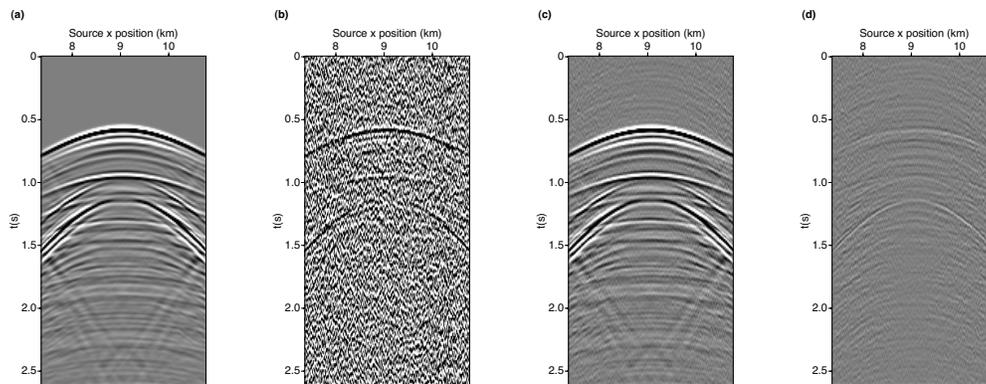


Figure 3.14: Results of simultaneous source separation in common receiver domain when STR equals to 20: (a) The real unblended common receiver gather (centre receiver). (b) Pseudo-deblended common receiver gather. (c) Deblended common receiver gather via the proposed algorithm. (d) Differences between (a) and (c). In this example, the signal-to-noise ratio after separation is 20.2dB.

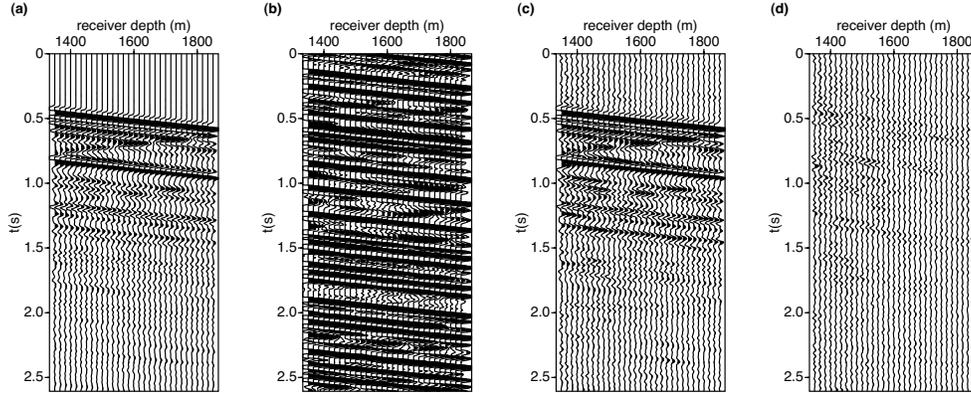


Figure 3.15: Results of simultaneous source separation in common shot domain when STR equals to 20: (a) The real unblended common shot gather (centre shot). (b) Pseudo-deblended shot record. (c) Deblended shot record via the proposed algorithm. (d) Differences between (a) and (c). In this example, the signal-to-noise ratio after separation is 20.2dB.

3.3.2 Example simulated from a real data set

We adopted a 2D marine seismic dataset from the North Viking Graben, North Sea, to test the efficacy of the proposed source separation method. The area contains simple geology (Eggenberger et al., 2016) and both sources and receivers are sampled with an interval of 25m (Keys and Foster, 1998). The data are then numerically blended assuming the receivers are ocean bottom nodes. Amplitude gaining has been applied before source blending. The sources are fired according to the single-vessel simultaneous source acquisition design, and the STR equals 2. We then applied the dual domain rank reduction algorithm to separate the blended data.

For 2D seismic data sets, the frequency slices of the common receiver gathers are vectors. Matrix rank reduction methods can no longer be applied directly on the data. The key is to sort data from source-receiver domain to offset-midpoint domain before applying rank reduction in each iteration. This is because in both common offset and common midpoint domains, the interferences from simultaneously fired shots are perturbed by random source initiation time. The coherence constraint for deblending is valid for a 3D offset-midpoint gather. The low-rank constraint is effective in the offset-midpoint domain (Trickett, 2003). The projected gradient step that operates on 2D seismic data sets for simultaneous source

separation is as follows

$$\begin{aligned}\mathcal{Z} &= \mathcal{D}^\nu - \lambda \mathcal{B}^*(\mathcal{B}\mathcal{D}^\nu - \mathbf{b}), \\ \mathcal{D}^{\nu+1} &= \mathcal{S}^* \mathcal{F}^{-1} \mathcal{P}_r \mathcal{F} \mathcal{S} [\mathcal{Z}] = \mathcal{P}_{sfr} [\mathcal{Z}],\end{aligned}\tag{3.16}$$

where \mathcal{S} denotes the operator that sorts data from shot-receiver domain to offset-midpoint domain. In each iteration, the blending and pseudo-deblending operator is applied to data in source-receiver domain, whereas the f-x-y eigenimage filtering operates on each frequency slice of the offset-midpoint data cube. Figure 3.16 shows the deblending results in common midpoint domain. The interferences are effectively suppressed by the proposed algorithm. Figure 3.17 shows a near offset gather. The quality of the separation (Q_S) has been improved to a factor of 14.2dB.

3.4 Conclusions

This chapter illustrates an inversion scheme that can be utilized for separation of simultaneous source data. The method relies on the randomization of firing time delays and operates in the common receiver domain of a multidimensional seismic dataset. The cost function is defined by the source blending system. As the interferences perturbed by firing time delays increase the rank of each frequency slice of a common receiver cube, a low-rank constraint has been imposed to enforce the coherence of solution. The gradient projection algorithm has been adopted for solving the rank constrained inverse problem. We presented a fast rank reduction algorithm based on random projection and randomized QR decomposition to improve the processing speed. Through a test with a synthetic 3D VSP data set, we showed the algorithm effectively separated the responses from simultaneous sources. We also tested the performance of the algorithm versus rank and survey time ratio. For low survey time ratio values, a broad range of rank would ensure the success of the deblending algorithms. At high survey time ratios, prior knowledge of the model is required for selecting the optimal rank.

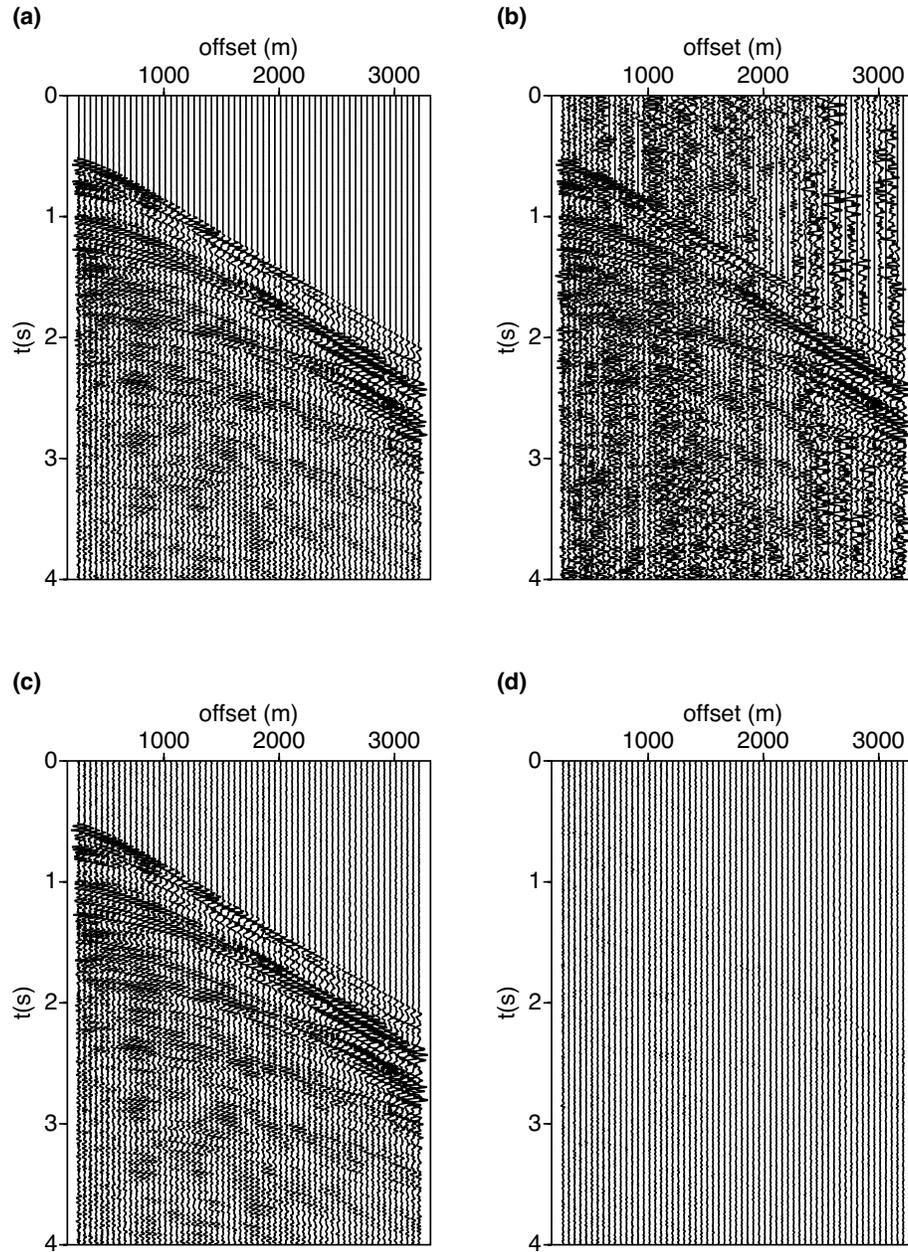


Figure 3.16: Results of simultaneous source separation of a 2D marine line: (a) The real unblended CMP gather. (b) CMP gather sorted from Pseudo-deblended data. (c) Deblended results via the proposed algorithm. (d) Differences between (a) and (c). In this example, the signal-to-noise ratio after separation is 14.2dB.

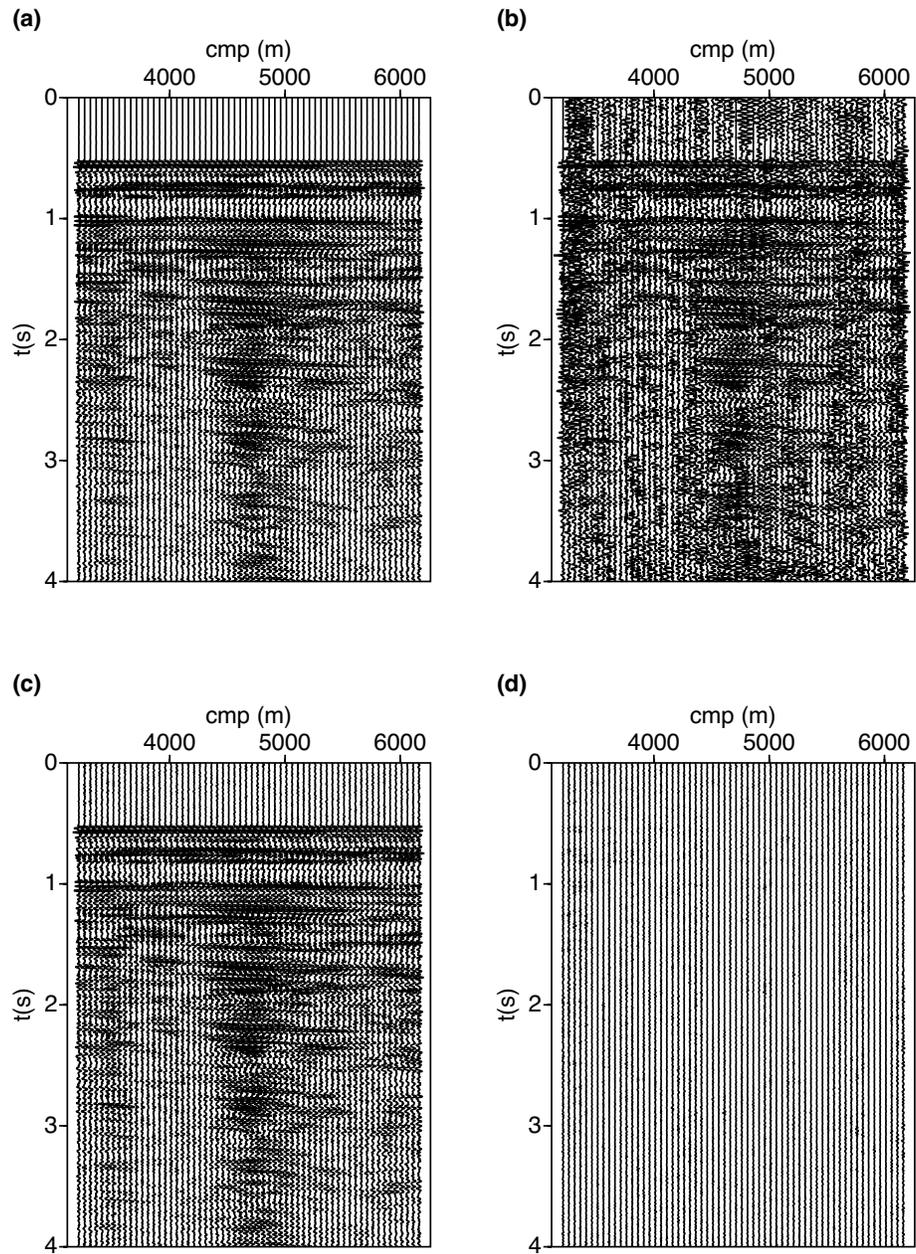


Figure 3.17: Results of simultaneous source separation of a 2D marine line: (a) The real unblended common offset gather. (b) Common offset gather sorted from Pseudo-deblended data. (c) Deblended result via the proposed algorithm. (d) Differences between (a) and (c).

CHAPTER 4

Deblending via iterative SSA¹

4.1 Introduction

Singular Spectrum Analysis (SSA), which is also known as Cadzow filtering, denotes a family of methods where rank-reduction is applied to Hankel matrices formed from noisy and often incomplete observations. These methodologies are found under different names in signal and image processing (Cadzow, 1988), time-series and spectral analysis (Vautard and Ghil, 1989; Hua, 1992) and dynamical systems (Broomhead and King, 1986). For instance, researchers in the field of time series analysis and dynamical systems frequently adopt the name Singular Spectrum Analysis. On the contrary, researchers in the field of communications often used the name Cadzow's iterative denoising (Blu et al., 2008). In the area of seismic signal processing, Cadzow or Singular Spectrum Analysis has been developed for the enhancement of the signal-to-noise ratio and the reconstruction of seismic records (Trickett and Burroughs, 2009; Sacchi, 2009; Trickett et al., 2010). The method entails organizing the spatial data at a given monochromatic frequency into a Hankel matrix. For the ideal noise-free data, the Hankel matrix is of low-rank. The missing traces and noise will increase the rank of the Hankel matrix. Therefore, rank-reduction filtering is an effective way to annihilate noise while preserving the unblended signal. In this work, we will utilize the designation adopted by Sacchi (2009), Oropeza and Sacchi (2011) and Gao et al. (2013), and therefore, we will call our rank-reduction denoising method *SSA filtering*. However, we are aware that equivalent denoising and reconstruction algorithms have been extensively studied in seismic data processing by Trickett and Burroughs (2009) who utilized the name *Cadzow filtering*.

¹A version of this chapter is published in Cheng J. and M. D. Sacchi, 2015, Separation and reconstruction of simultaneous source data via iterative rank reduction, *GEOPHYSICS* 80(4): V57-V66

In this chapter, instead of directly applying rank-reduction on the spatial data, we introduce the SSA reduced-rank filter as the projection filter in the gradient projection method for suppressing interferences that arise in simultaneous source acquisition. Especially, SSA is applied to the small patches extracted from the common receiver gathers of the pseudo-deblended data. This is because SSA is effective in suppressing incoherent noise while preserving linear-events. In small patches, the simultaneous source interferences would appear incoherent and therefore would be suppressed by the SSA filter. The desired data, on the other hand, can be approximated by linear events and would be preserved. Through tests with a synthetic example, we show that the interference can be effectively suppressed by the proposed method. Also, we also show that the proposed algorithm can be modified to cope with deblending and data recovery simultaneously. A real survey acquired in the Gulf of Mexico is utilized to mimic a simultaneous source acquisition with missing shot locations. The algorithm was able to recover the missing shot gathers from the blended acquisition with an improvement of the signal quality. We also study the separability of simultaneous source data based on different distributions of fire time delays. We analyze the relationship between different firing schemes and the quality of the separation via the iterative SSA filtering method. Insights can be gained from these tests towards an optimal acquisition design for simultaneous source acquisition.

4.2 Theory

4.2.1 Singular spectrum analysis

We provide a short review of reduced-rank filtering for noise attenuation. The method can be found in the literature as Cazdow filtering or Singular Spectrum Analysis (SSA). The details associated to the implementation of SSA for seismic noise attenuation and seismic data reconstruction can be found in (Oropeza and Sacchi, 2011). We discuss the 2-D (t - x) implementation of reduce-rank filtering via SSA. However, we make the point that SSA for multidimensional volumes have been extensively discussed in Trickett et al. (2010), Oropeza and Sacchi (2011) and Gao et al. (2013). Seismic data in a small window can be represented in the frequency-space domain via the superposition of plane waves

$$\tilde{d}(\omega)_j = \sum_{k=1}^K A_k(\omega) e^{i\omega\eta_{kj}\Delta x}, \quad (4.1)$$

where $i = \sqrt{-1}$, $j = 1, 2, \dots, N$ is the trace index in the spatial axis and ω represents temporal frequency. In this equation we assume that the data are composed of K linear events with distinct dips with each dip denoted by η_k . We denote $A_k(\omega)$ the complex amplitude of the k -th plane wave and Δx indicates the spatial interval between seismograms. SSA of seismic data in the frequency-space domain entails the following steps:

- We first embed the complex amplitudes at a given monochromatic frequency $\tilde{\mathbf{d}}_\omega = [\tilde{d}_1, \tilde{d}_2, \dots, \tilde{d}_N]^T$ into a Hankel structured trajectory matrix \mathbf{H} as follows

$$\mathbf{H} = \mathcal{H}[\tilde{\mathbf{d}}_\omega] = \begin{bmatrix} \tilde{d}_1 & \tilde{d}_2 & \cdots & \tilde{d}_M \\ \tilde{d}_2 & \tilde{d}_3 & \cdots & \tilde{d}_{M+1} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{d}_{N-M+1} & \tilde{d}_{N-M+2} & \cdots & \tilde{d}_N \end{bmatrix}, \quad (4.2)$$

where \mathcal{H} denotes the operator for constructing the Hankel matrix. For convenience, we choose $M = \lfloor \frac{N}{2} \rfloor + 1$ to make the Hankel matrix approximately square. Forming the Hankel structured matrix is computational efficient but this step requires the storage of a matrix of size $M \times (N - M + 1)$.

To show that the \mathbf{H} is a low-rank matrix, we assume the data are composed of a single linear event. Equation 4.1 can be simplified as follows

$$d_{j+1} = Ae^{i\omega\eta(j+1)\Delta x} = Ae^{i\omega\eta j\Delta x}e^{i\omega\eta\Delta x} = d_j\Phi, \quad (4.3)$$

where $\Phi = e^{i\omega\eta\Delta x}$ is a constant corresponding to the phase-shift caused by Δx . We can apply Equation 4.3 recursively before substituting to Equation 4.2

$$\mathbf{H} = \begin{bmatrix} \tilde{d}_1 & \tilde{d}_1\Phi & \cdots & \tilde{d}_1\Phi^{M-1} \\ \tilde{d}_1\Phi & \tilde{d}_1\Phi^2 & \cdots & \tilde{d}_1\Phi^M \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{d}_1\Phi^{N-M} & \tilde{d}_1\Phi^{N-M+1} & \cdots & \tilde{d}_1\Phi^{N-1} \end{bmatrix}. \quad (4.4)$$

Definitely, \mathbf{H} is a rank-1 matrix as all its rows can be represented by a scalar multiplies the first row. For data composed by the superposition of K plane waves, one can show that $\text{rank}(\mathbf{H}) = K$ (Hua, 1992; Yang and Hua, 1996). Since the additive noise will increase the rank of matrix \mathbf{H} , one can use rank reduction to attenuate the additive noise.

- The second step of SSA entails finding a low-rank approximation of the Hankel trajectory matrix. This is usually done by the Singular Value Decomposition (Golub and

van Loan, 1996)

$$[\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}] = \text{SVD}[\mathbf{H}], \quad (4.5)$$

where \mathbf{U} and \mathbf{V} are orthonormal matrices and $\mathbf{\Sigma}$ is a diagonal matrix. Note that this step can be extremely expensive for large matrices. The computational complexity of the Singular Value Decomposition is of the order of $\mathcal{O}(M^3)$ (Golub and van Loan, 1996). A new set of singular values $\hat{\mathbf{\Sigma}}$ are computed via

$$\begin{aligned} \hat{\Sigma}_{k,k} &= \Sigma_{k,k} & k \leq K \\ \hat{\Sigma}_{k,k} &= 0 & k > K \end{aligned} \quad (4.6)$$

The low-rank approximation of the Hankel matrix is then computed via

$$\hat{\mathbf{H}} = \mathbf{U}\hat{\mathbf{\Sigma}}\mathbf{V}. \quad (4.7)$$

The regrouping of the Hankel matrix also yield a complexity $\mathcal{O}(M^2K)$.

- In the last step, one needs to recover the filtered data from the reduced-rank matrix $\hat{\mathbf{H}}$. Anti-diagonal averaging of the reduced-rank matrix is applied as follows

$$\hat{d}_j = \begin{cases} \frac{1}{j} \sum_{i=1}^j \hat{H}_{i, j-i+1}, & 1 \leq j \leq M, \\ \frac{1}{M} \sum_{i=1}^M \hat{H}_{i, j-i+1}, & M \leq j \leq (N - M + 1), \\ \frac{1}{N-j+1} \sum_{i=j-N+M}^M \hat{H}_{i, j-i+1}, & (N - M + 1) \leq j \leq N, \end{cases} \quad (4.8)$$

where i and j here are the indices of the Hankel matrix. Since the anti-diagonal averaging is the backward process of constructing the Hankel matrix, we can rewrite Equation 4.8 in its operator form as follows

$$\hat{\mathbf{d}}_\omega = \mathcal{H}^*[\mathbf{H}], \quad (4.9)$$

where \mathcal{H}^* denotes the operator for anti-diagonal averaging.

The procedures discussed above can be summarized using the following expression

$$\hat{\mathbf{d}}_\omega = \mathcal{H}^*[\mathcal{R}[\mathcal{H}[\tilde{\mathbf{d}}_\omega]]]$$

where \mathcal{R} is the rank-reduction operator that approximates \mathbf{H} by a rank- K matrix in the frequency-space domain. For our analysis, it is more convenient to represent the SSA filter

in the $t - x$ domain

$$\begin{aligned}\hat{\mathbf{d}} &= \mathcal{F}^{-1} \mathcal{H}^*[\mathcal{R}[\mathcal{H}\mathcal{F}[\mathbf{d}]]] \\ &= \mathcal{P}_{SSA} \mathbf{d}\end{aligned}\tag{4.10}$$

where \mathcal{F} and \mathcal{F}^{-1} represent forward and inverse Fourier operators to transform data from $t - x$ domain to $f - x$ domain and from $f - x$ domain to $t - x$ domain, respectively. The construction of Hankel matrix, rank-reduction of Hankel matrix, and the anti-diagonal averaging of the Hankel matrix are carried out for all frequencies. The operator \mathcal{P}_{SSA} represents the SSA rank-reduction filter that will be used by our source separation algorithm. Compared with other filtering based methods, SSA increases the redundancy of data by forming Hankel matrices. The latter enhances the capability for noise suppression while preserving the desired signal. However, the cost of the SVD and the storage of trajectory matrices make the algorithm unfeasible for large multidimensional seismic data (Gao et al., 2013). In Appendix B, we present a fast and memory efficient implementation of SSA that does not require building Hankel matrices. We will adopt the fast and memory efficient SSA as the projection operator.

4.2.2 The SSA projection operator

Since the linear events assumption for SSA is only valid in a small patch of seismic data, we will decompose the ideal common receiver gather that one would have obtained via standard acquisition into small overlapping windows (Figure 4.1). We assume that each window is composed by a superposition of a finite number of linear dips. The synthesis of the data in small windows is written as follows

$$\mathbf{d} = \sum_{l=1}^L \mathcal{W}_l \mathbf{d}_l,\tag{4.11}$$

where \mathbf{d}_l is the l -th data window and \mathcal{W}_l represents an operator that translates data windows with proper tapering in the areas where adjacent windows overlap. Similarly, we define an adjoint operator of the form

$$\mathbf{d}_l = \mathcal{W}_l^* \mathbf{d}, \quad l = 1, \dots, L,\tag{4.12}$$

where \mathcal{W}_l^* represents an operator that extracts the window l from the data with proper tapering in the areas where adjacent windows overlap. The operators \mathcal{W}_l and \mathcal{W}_l^* are written in the explicit form, and special attention has been taken to guarantee that they pass the dot product test. The latter ensures that expressions 4.11 and 4.12 are a forward-

adjoint pair (Claerbout, 1992). The aforementioned windowing strategy can be extended to multidimensional seismic data.

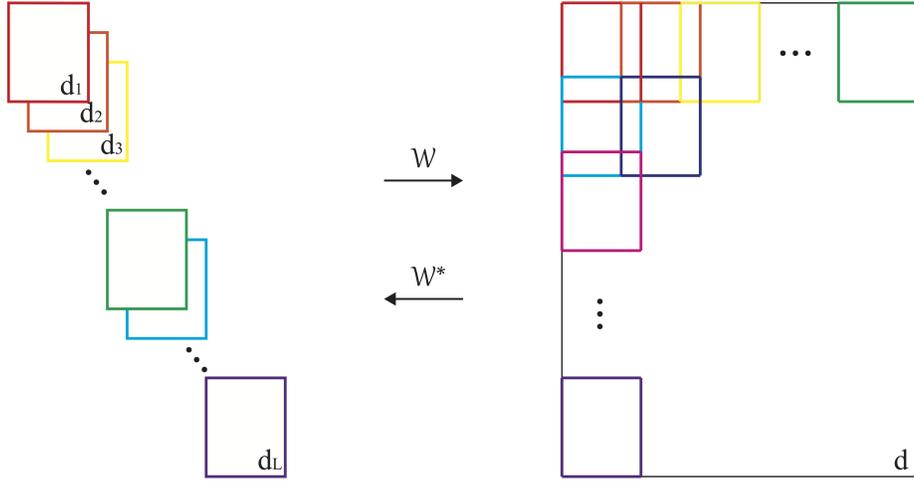


Figure 4.1: The windowing operator and its adjoint operator. The windowing operator \mathcal{W} extracts small patches of data from an entire gather. The adjoint operator \mathcal{W}^* combines all the processed small patches back into a gather. A Gaussian taper is used to combine areas with overlap.

Combining Equation 1.2 with Equation 4.11 and after considering additive noise in the blended data,

$$\mathbf{b} = \mathcal{B} \sum_l \mathcal{W}_l \mathbf{d}_l + \mathbf{n}. \quad (4.13)$$

The separation of sources can be expressed as finding the solution of

$$\min J = \|\mathbf{b} - \mathcal{B} \sum_l \mathcal{W}_l \mathbf{d}_l\|_2^2, \quad l = 1, \dots, L. \quad (4.14)$$

Clearly, once $\mathbf{d}_l, l = 1, \dots, L$ are found, Equation 4.11 is used to reconstruct the common receiver gather. However, the cost function (Equation 4.14) does not have a unique solution. Therefore, an additional constraint is needed to solve the problem. Our constraint is given by

$$\mathbf{d}_l = \mathcal{P}_{SSA} \mathbf{d}_l, \quad l = 1, \dots, L. \quad (4.15)$$

The constrained least-squares problem by Equation 4.14 and Equation 4.15 can be solved

via the gradient projection method described in Chapter 2. In this case, we define the gradient of Equation 4.14 by

$$\mathbf{g}_l = \mathcal{W}_l^* \mathcal{B}^* (\mathcal{B} \sum_m \mathcal{W}_m \mathbf{d}_m - \mathbf{b}), \quad l, m = 1, \dots, L. \quad (4.16)$$

The projected gradient update rule can be written in the following way

$$\mathbf{d}_l^{[\nu+1]} = \mathcal{P}_{SSA}[\mathbf{d}_l^{[\nu]} - \lambda \mathcal{W}_l^* \mathcal{B}^* (\mathcal{B} \sum_m \mathcal{W}_m \mathbf{d}_m^{[\nu]} - \mathbf{b})], \quad l, m = 1, \dots, L. \quad (4.17)$$

where ν denotes iteration number. In each iteration, we compute the difference between the blended estimation and the observed data. The difference is transformed to the pseudo-deblended domain and is distributed to each small patch of the previous estimation by the windowing process. Each small patch is filtered by the projection filter \mathcal{P}_{SSA} to separate the signal and the interferences. The filtered patches are then grouped and blended for computing the difference in the next iteration. Similar to deblending via the iterative rank reduction method described in Chapter 3, we choose the pseudo-deblended data as the initial solution. The convergence of the algorithm can be guaranteed by utilizing Equation 2.25 and diminishing step lengths (Bertsekas, 1999; Nedic and Bertsekas, 2001). In the examples below, the step size is decreased according to $\lambda^{[\nu]} = \lambda^{[0]}/\sqrt{\nu}$. We choose $\lambda^{[0]} = 2/\alpha$, where α is the largest eigenvalue of $\mathcal{B}^* \mathcal{B}$. Our selection of step size guarantees the convergence. However, there might be other step size schedules that might speed up the convergence of the algorithm (Mahdad et al., 2011).

4.2.3 Joint separation and reconstruction of seismic sources

We turn our attention to the case where the seismic data are not regularly sampled in the source coordinate. In other words, we assume that some sources were not fired. The aforementioned source separation algorithm will encounter problems since our $f-x$ domain model is effective only when the sources satisfy a regular spatial distribution. However, the data in the unblended domain can be approximated as the entrywise product of the complete dataset and a sampling operator \mathcal{T} . The operator \mathcal{T} multiplies traces that are alive by 1, whereas dead traces are multiplied by 0 (Liu and Sacchi, 2004). Without losing generality, the new problem is solved via

$$\mathbf{d}_l = \underset{d_l}{\operatorname{argmin}} \|\mathbf{b} - \mathcal{B} \mathcal{T} \sum_l \mathcal{W}_l \mathbf{d}_l\|_2^2, \quad l = 1, \dots, L. \quad (4.18)$$

Therefore, the projected gradient method turns into

$$\mathbf{d}_l^{[\nu+1]} = \mathcal{P}[\mathbf{d}_l^{[\nu]} - \lambda \mathcal{W}_l^* \mathcal{T}^* \mathcal{B}^* (\mathcal{B} \mathcal{T} \sum_m \mathcal{W}_m \mathbf{d}_m^{[\nu]} - \mathbf{b})], \quad l, m = 1, \dots, L. \quad (4.19)$$

where it is easy to show that $\mathcal{T} = \mathcal{T}^*$ and $\mathcal{T}^* \mathcal{T} = \mathcal{T}$ (Liu and Sacchi, 2004; Naghizadeh and Sacchi, 2010).

4.3 Examples

4.3.1 Comparison of projection operators

To test the performance of the proposed deblending method, we first synthesized an example with three linear crossing events to mimic a small patch of a noise-free common receiver gather (Figure 4.2 (a)). We then introduced severe blending noise (Figure 4.2 (b)). In addition to source separation via the proposed iterative rank reduction approach, we also tested the iterative $f - x$ deconvolution method and the iterative $f - k$ thresholding algorithm. In other words, we have replaced the rank-reduction operator \mathcal{P}_{SSA} in Equation 4.10 by an $f - x$ deconvolution operator (Peng and Liu, 2013) and by the $f - k$ hard-thresholding operator (Abma et al., 2010; Chen et al., 2014).

We adopted a rank $K = 3$ for each iteration for the iterative rank reduction deblending method. The $f - x$ deconvolution was implemented with a 15 point filter. As to the $f - k$ domain thresholding method, we adopted the exponential schedule proposed by Gao et al. (2010) to slowly decrease the amplitude threshold. We also tuned the step size to make sure that the three method follow similar convergence curves. The quality of pseudo-deblending Q_{PD} and deblending Q_S is calculated in dB units via the following two expressions

$$\begin{aligned} Q_{PD} &= 10 \log \frac{\|\mathbf{d}_{true}\|_2^2}{\|\mathbf{d}_{true} - \mathbf{d}_{PD}\|_2^2} \\ Q_S &= 10 \log \frac{\|\mathbf{d}_{true}\|_2^2}{\|\mathbf{d}_{true} - \mathbf{d}_S\|_2^2}, \end{aligned} \quad (4.20)$$

where \mathbf{d}_{PD} denotes the pseudo-deblended common receiver gather, \mathbf{d}_{true} is the true synthetic data (a conventional common receiver gather) and \mathbf{d}_S stands for the separated common receiver gather via the iterative projected gradient algorithm. A large Q value corresponds to fewer interferences. The performance of the method is given by $I = Q_S - Q_{PD}$.

Figure 4.2 (c) shows the deblended data obtained via the proposed iterative rank-reduction method. Figure 4.2 (f) shows the difference section for the iterative rank-reduction method. Figure 4.2 (d) and Figure 4.2 (g) portray the deblended data and the difference section for

the projected gradient algorithm with projection operator given by $f - x$ deconvolution, respectively. Similarly, Figure 4.2 (e) and Figure 4.2 (h) show the deblended data and the difference section for the gradient projection algorithm with projection operator given by $f - k$ thresholding. For completeness, we have added Table 4.1 to indicate Q_S and I for our examples ($Q_{PD} = -2.3$ dB). Our results indicate that for this particular example, three linear events can be optimally modelled via a Hankel matrix of rank 3 and therefore rank reduction provides the highest attenuation of source interference noise.

Projection	Q_S [dB]	$I = Q_S - Q_{PD}$ [dB]
Rank reduction	20.2	22.5
$f - x$ deconvolution	10.1	12.4
$f - k$ thresholding	12.0	14.3

Table 4.1: Quality of deblending Q_S and performance I for the projected gradient method with different projection operators: rank reduction (proposed method), $f - x$ deconvolution (Peng and Liu, 2013) and $f - k$ thresholding (Abma et al., 2010). These results correspond to the comparison tests portrayed in Figure 4.2.

4.3.2 Deblending: SAIG sythetic data set

We also test the proposed deblending algorithm with a 2D synthetic example. In this example, we use an acoustic finite difference modelling code to simulate a prestack marine data set for the SAIG velocity model. The velocity model, as well as the source-receiver geometry, are shown in Figure 4.3. The sources and receivers are distributed every 20 meters. A total number of 350 sources were simulated. Each source fires into a fixed array of 375 receivers. The receivers are deployed at 500 meter depth to simulate ocean bottom nodes. A Ricker wavelet with central frequency of 20Hz has been utilized for generating the data.

We use groups of 5 shots with fixed spacing (1400 m) that were blended with time intervals generated from the uniform distribution from 0 to 2 seconds. Then all the five sources moved to the next position and are blended again. A total number of 70 blended shots were generated. Figure 4.4 shows the spatial and temporal distribution of sources for this particular acquisition. As a result, the total acquisition time was compressed to 20% of the conventional acquisition time. Then we apply the proposed algorithm to recover the

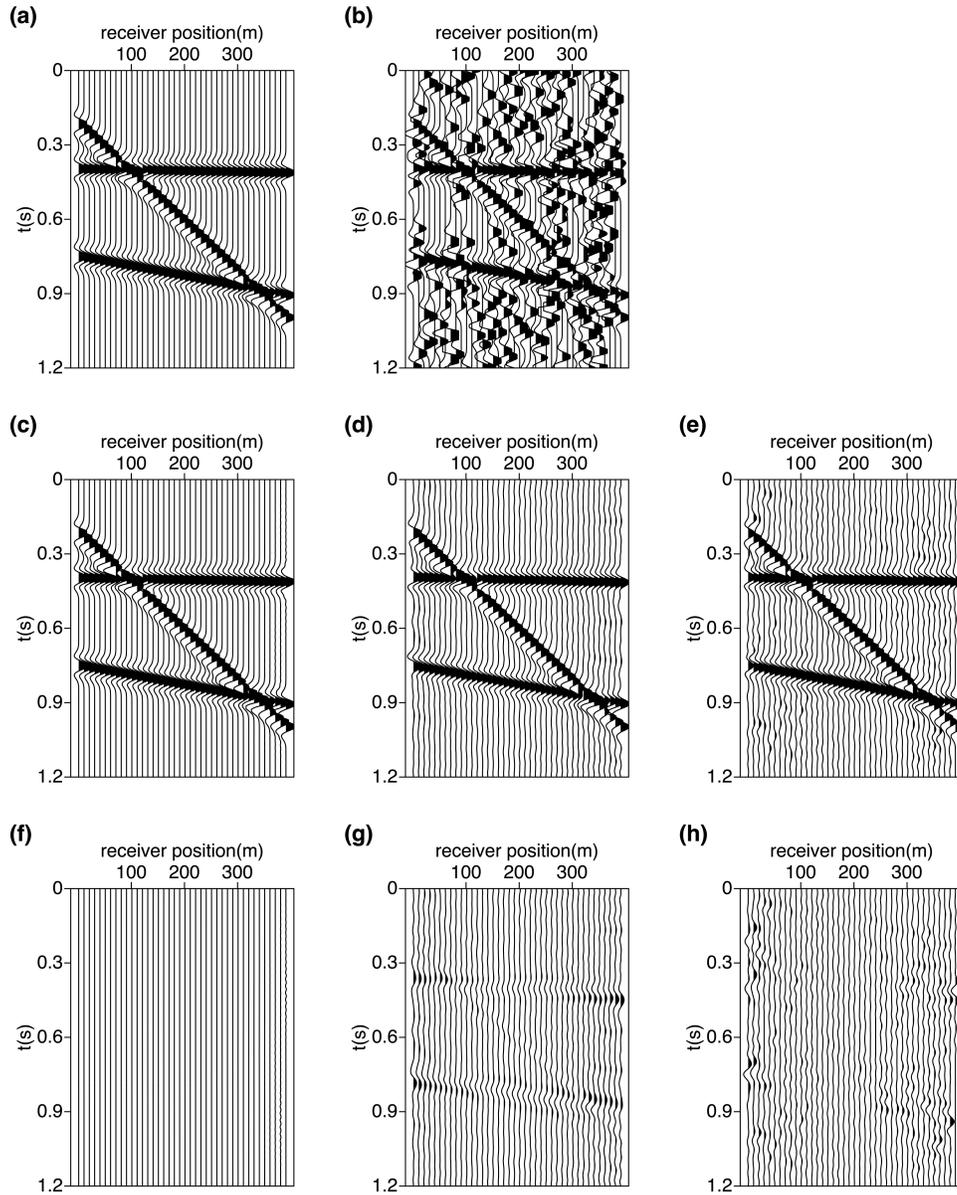


Figure 4.2: Comparisons of deblending results for numerically blended synthetic data with linear events. (a) The original unblended synthetic data. (b) Pseudo-deblended gather. (c) Deblending result with the proposed iterative rank reduction method. (d) Deblending result using iterative $f - x$ deconvolution. (e) Deblending results using iterative $f - k$ domain thresholding. (f) Difference section between a and c. (g) Difference section between a and d. (h) Difference section between a and e.

common receiver gathers and to form separated source gathers for the whole volume. The operator \mathcal{W} extracts windows of size 100 samples in time and 40 traces. The overlap is of 25 samples in time and 10 traces in space. A cosine taper was applied in both time and space.

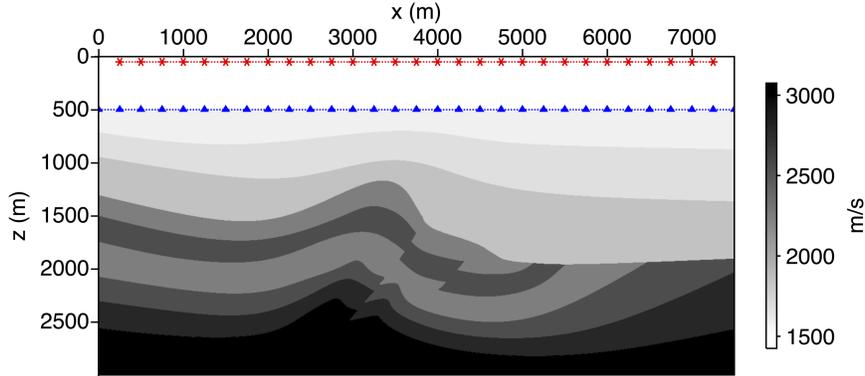


Figure 4.3: Velocity model utilized to simulate blended data for our examples via finite difference modelling. We synthesize a data set with 4 ms time interval and a 20 Hz central-frequency Ricker wavelet. We also overlaid the source (*) and receiver (▲) geometry in this plot.

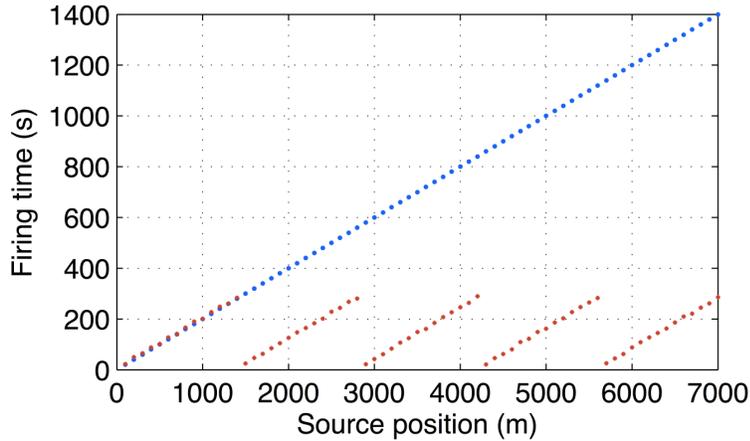


Figure 4.4: Spatial and temporal distribution of firing times for conventional seismic acquisition (Blue) and 2D simultaneous source acquisition. The multi-vessel scenario is portrayed in red. In each round, 5 sources fire with small random time delays. The spatial distance between sources are fixed in each round.

Figure 4.5 shows the misfit between the solution and the observed blended data for one receiver. The curve also shows the difference between the solution at a given iteration versus the true answer. The algorithm is comparatively effective as both curves reach convergence after about 15 iterations. The misfit did not reduce to zero due to the strategy we have adopted for selecting the rank. It is important to mention that, for each window, we let the rank of the SSA filter increase with iterations. At early iterations, we can apply a small rank to eliminate strong crosstalk and then, gradually increase the rank to allow retrieving details that depart from the linear event model. This is analogous to setting the threshold schedule in projection-onto-convex sets (POCS) interpolation and deblending methods (Abma et al., 2010).

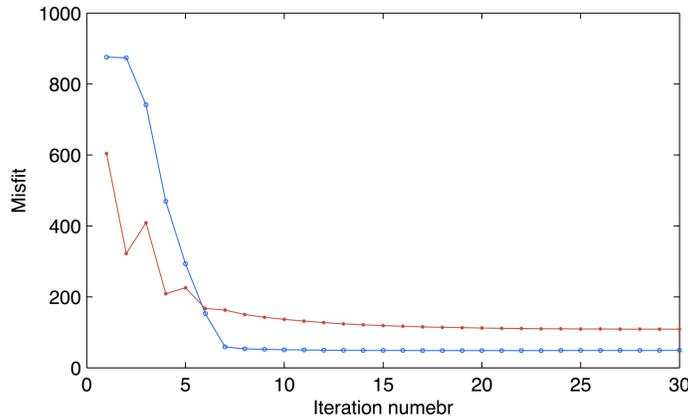


Figure 4.5: Convergence of the iterative SSA filtering deblending algorithm. Blue line indicates the l_2 norm of the difference between blended observations and the synthesized blended observation versus iteration. We also portray in red the difference between the unblended data and the true data versus iteration.

We carried out a large number of simulations to determine rank selection schedules. A strategy that works for us entailed starting with a small rank $k = 3$ and to then increase the rank by one every five iterations. In short, the final rank for our example after 30 iterations is $k = 9$. Figure 4.6 and Figure 4.7 show the results of the separation after 30 iterations for shot and receiver number 175. In this example, the proposed algorithm improved the quality from 0.5 dB to a factor of 15.1 dB ($I = 14.6$ dB). The unblended solution becomes comparable to the true conventional data set. Figure 4.8 shows the near offset section for this data set. The interferences from simultaneously fired shots are effectively suppressed. It is important to note that high amplitude direct waves can severely affect the performance of the algorithm. In fact, the proposed rank reduction algorithm is suited for scenarios where the direct waves are not extremely strong. This is consistent with the results presented by Maraschini et al. (2012).

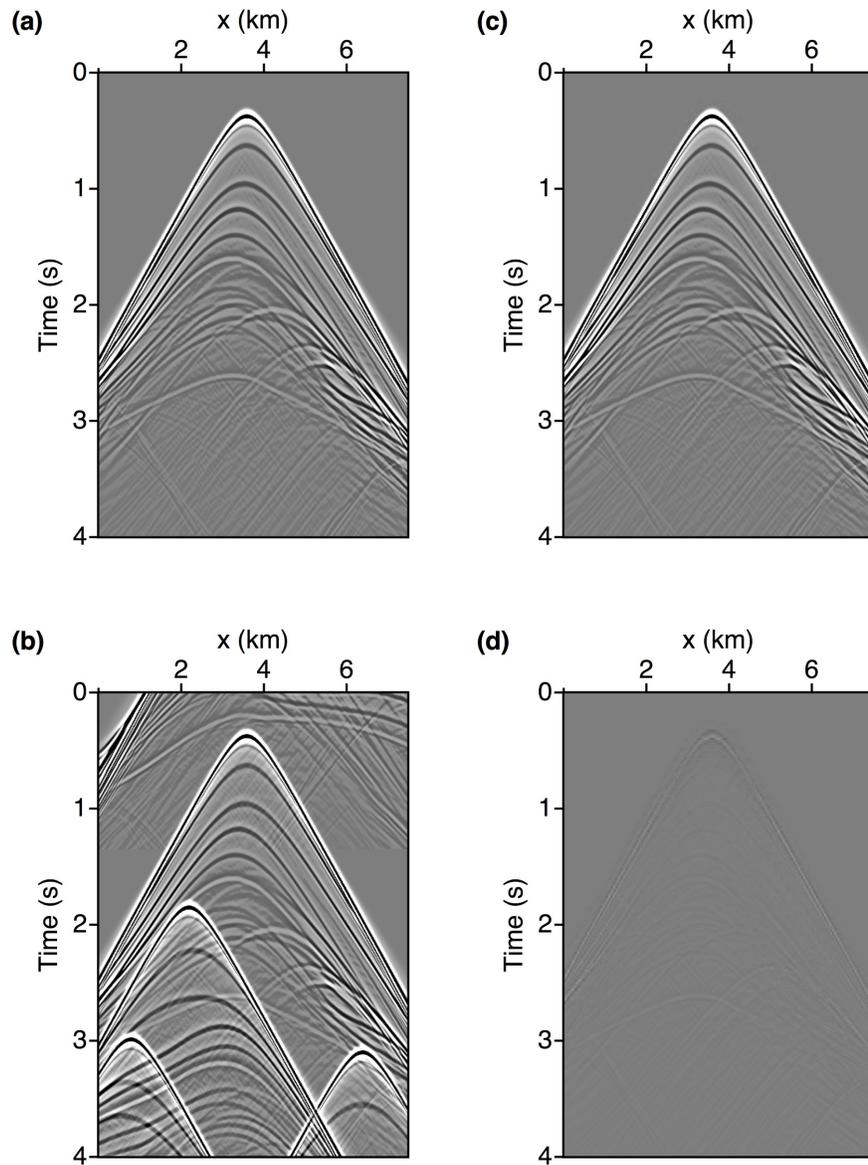


Figure 4.6: Results of the proposed iterative rank reduction deblending method in common shot domain. (a) The original shot gather from synthetic data. (b) Pseudo-deblended shot gather after numerical blending. (c) The deblended shot gather after 30 iterations of the proposed algorithm. The quality of data has been improved to 15.1 dB with respect to 0.5 dB. (d) Difference section between original and deblended data.

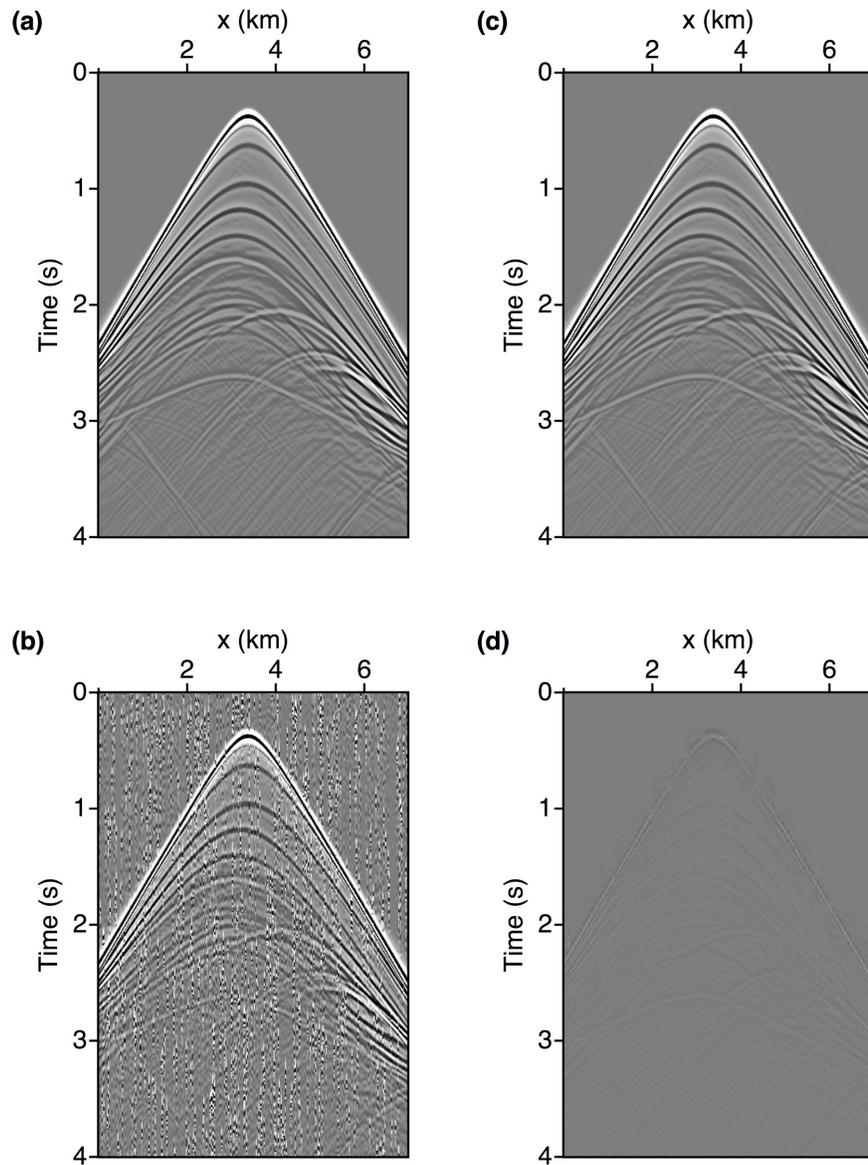


Figure 4.7: Results of the proposed iterative rank reduction deblending method in common receiver domain. (a) The original common receiver gather from synthetic data. (b) Pseudo-deblended common receiver gather after numerical blending. (c) The deblended common receiver gather after 30 iterations of the proposed method. (d) Difference section between original and deblended data.

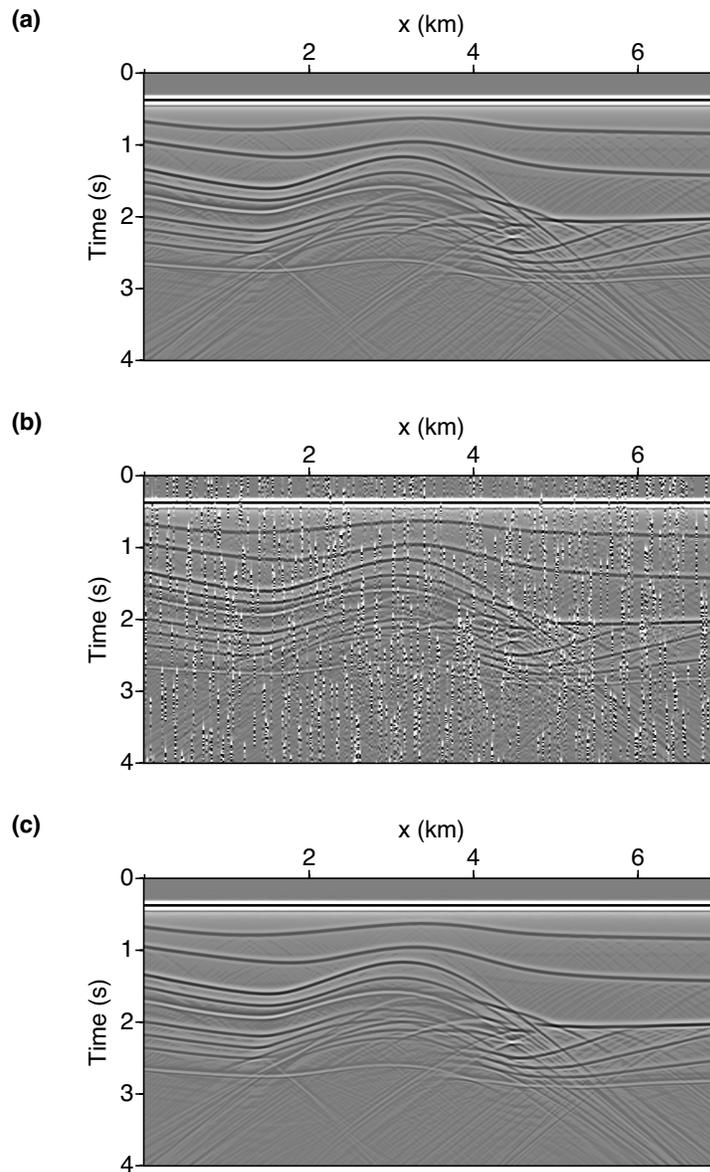


Figure 4.8: Results of the proposed iterative rank reduction deblending method for the near-offset section. (a) The original near-offset section from synthetic data. (b) Pseudo-deblended near-offset section after numerical blending. (c) The deblended near-offset section after 30 iterations.

4.3.3 Joint deblending and reconstruction: Gulf of Mexico data set

We also use a conventional data set collected from the Gulf of Mexico to simulate a blended acquisition. To synthesize the blending acquisition, we assume a self-simultaneous source scenario with one vessel firing. The receivers are ocean bottom nodes. Since the air gun sources are impulsive, the vessel keeps traveling without waiting for the recordings. Under this scenario, random fire time delays usually lead to an irregular distribution of source positions. This problem is tackled by moving the exact source position to the nearest grid point (Li et al., 2013). In this example, the spatial and temporal sampling intervals are set to be 26.6 m and 4 ms, respectively. As is shown in Figure 4.9, the firing time intervals between adjacent sources follow a uniform distribution. The overall acquisition time is reduced by 50%. We consider the case where 25 out of a total number of 80 sources are missing in an irregular pattern. Figure 4.10 shows a common shot gather recovered after 30 iterations. Note that we intentionally picked a common shot gather with the desired source missing. The pseudo-deblended shot gathers only contains interferences from nearby sources. Figure 4.11 shows the deblended common receiver gather, the quality of the reconstruction has been improved to $Q_S = 10.8$ dB with respect to $Q_{PD} = -1.5$ dB ($I = 12.3$ dB). We also display the whole volume for this small dataset in Figure 4.12. Unfortunately, the low fold of this example precludes us displaying a realistic near offset section.

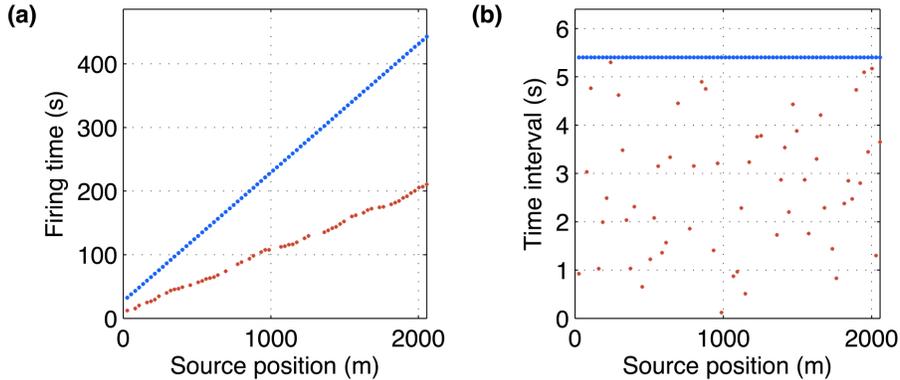


Figure 4.9: Spatial and temporal distribution of firing time for a conventional seismic acquisition (Blue) and 2D simultaneous source acquisition with 25 missing shots for one vessel scenario (red). (a) Source firing times. (b) Firing time intervals between adjacent sources.

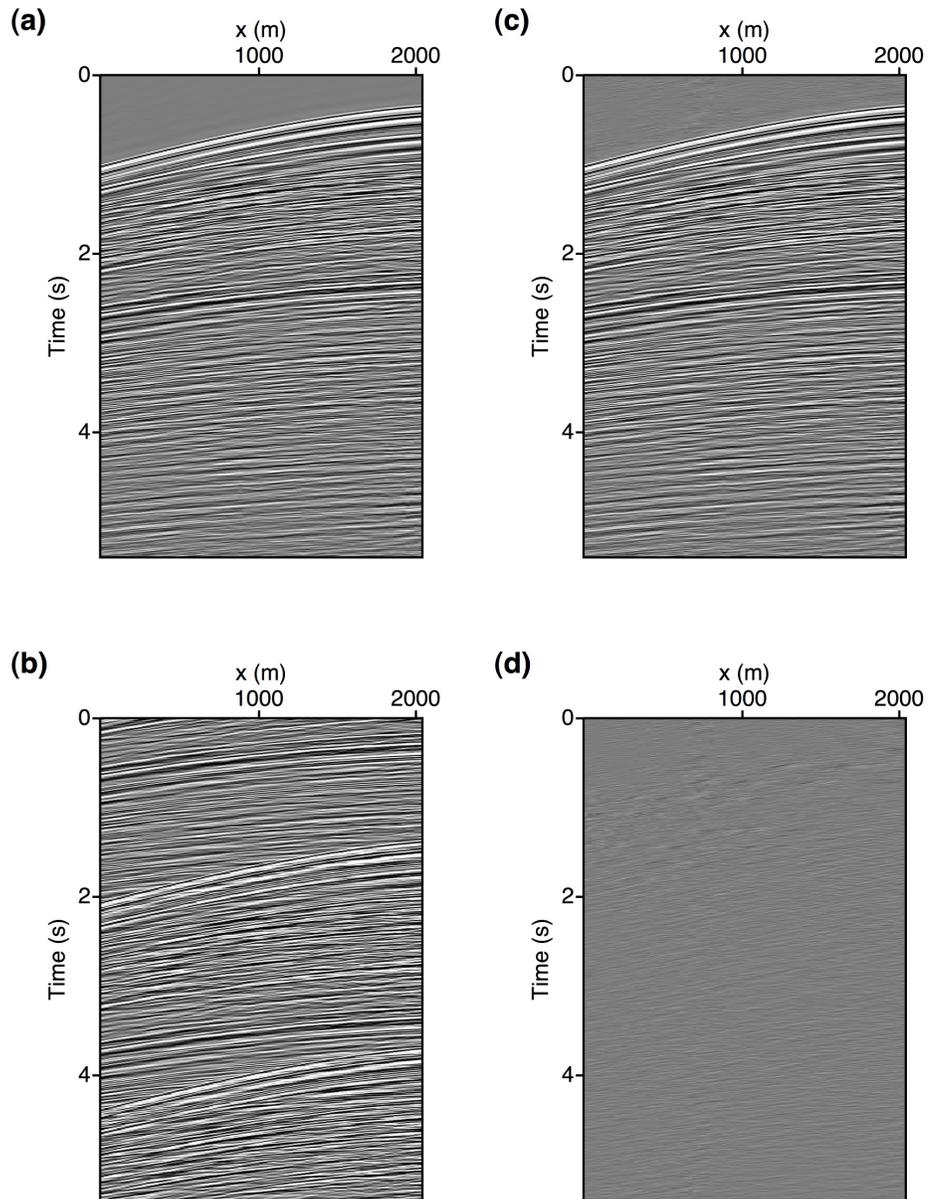


Figure 4.10: Results of deblending and reconstruction for a common shot gather. (a) The original shot gather from a Gulf of Mexico dataset. (b) Pseudo-deblended shot gather after numerical blending and sampling. The desired source is missing. (c) The deblended and reconstructed shot gather after 30 iterations. The quality of data has been improved to 10.8 dB from -1.5 dB. (d) Difference section between original and deblended and reconstructed data.

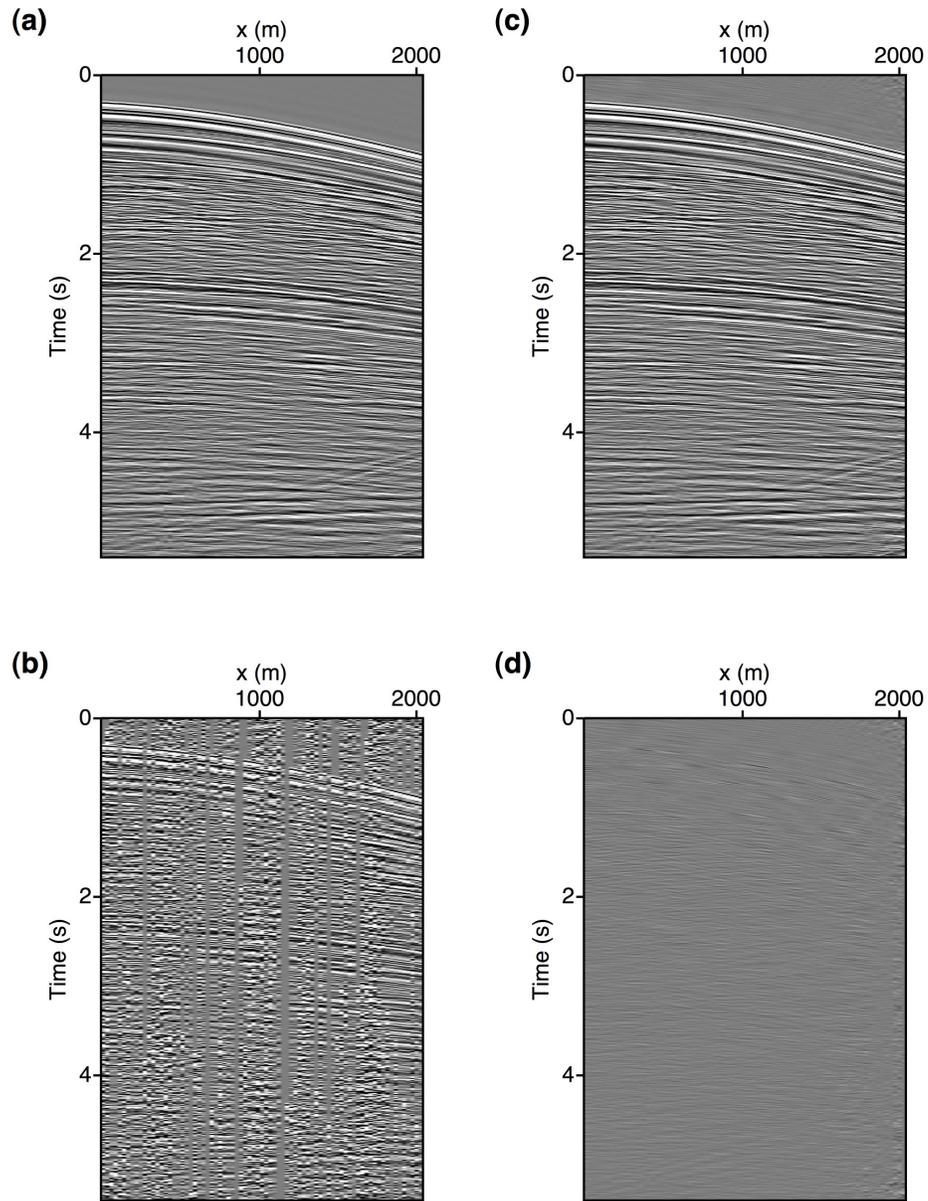


Figure 4.11: Results of deblending and reconstruction for a common receiver gather. (a) The original common receiver gather from a Gulf of Mexico dataset. (b) Pseudo-deblended common receiver gather after numerical blending. (c) Deblended and reconstructed common receiver gather after 30 iterations. (d) Difference section between original and deblended data.

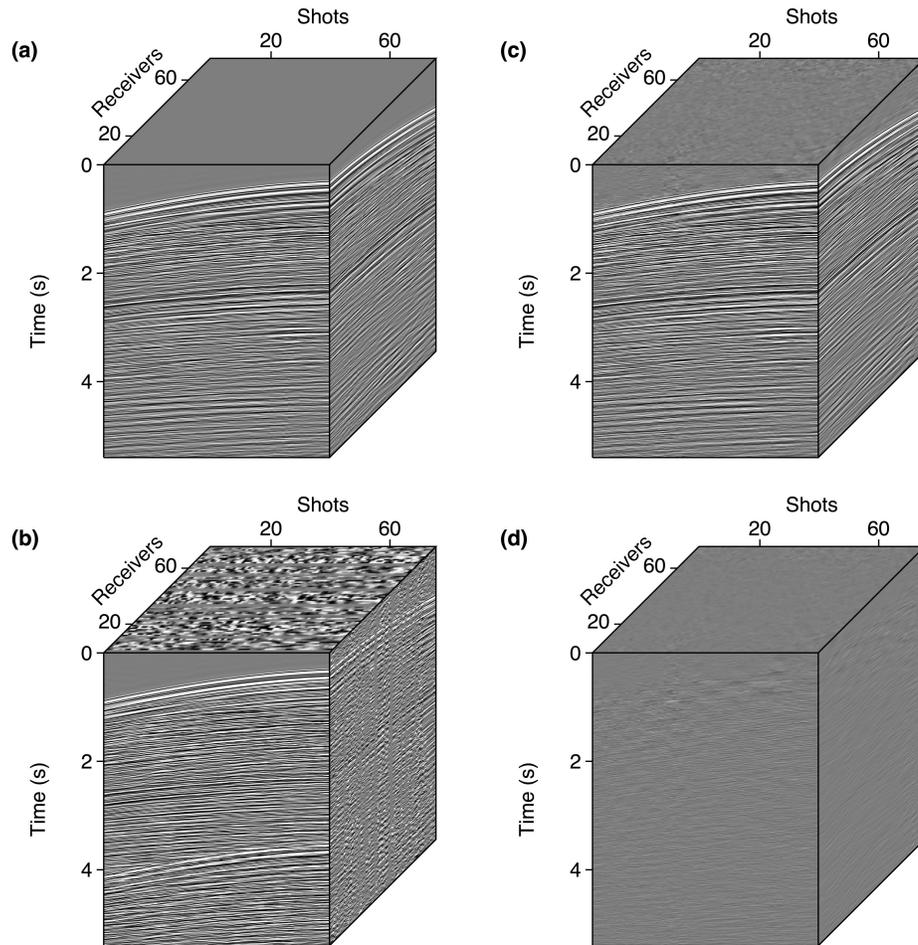


Figure 4.12: Results of deblending and reconstruction for whole data volume. (a) The original data from the Gulf of Mexico. (b) Pseudo-deblended data. (c) Deblended and reconstructed data after 30 iterations. (d) Difference section between original and deblended data.

4.4 An analysis of simultaneous source fire time delays

We study the separability of simultaneous source data based on different statistical models of the random source initiation time. The study is based on the data set acquired from the Gulf of Mexico and we assume the observed data is acquired by the one-vessel simultaneous

source acquisition. In this acquisition, the initiation time of the n -th source is defined by

$$t_n = t_{n-1} + \tau_n = \sum_{i=1}^n \tau_i, \quad (4.21)$$

where τ_i is the time delay for the i -th shot.

Our analysis of the fire time delays are based on different survey time ratios (STR). STR is a measurement of the efficiency of simultaneous source acquisition defined by

$$\text{STR} = \frac{t_{n_0}}{t_n}, \quad (4.22)$$

where t_{n_0} denotes the initiation time of the n -th shot in a conventional seismic acquisition using one vessel. The source initiation time interval, which is denoted as τ_0 , is often fixed in the conventional acquisition. Equation 4.22 can be further calculated by

$$\text{STR} = \frac{\tau_0}{\bar{\tau}}, \quad (4.23)$$

where $\bar{\tau}$ denotes the expectation of fire time intervals for simultaneous source acquisition. As a result, the efficiency of the one-vessel simultaneous source acquisition is determined by the expectation of the fire time delays. Our analyses are carried out based on two different aspects, listed in the subsections below.

4.4.1 Evaluation of different statistical models

In the first aspect, we generate the simultaneous source fire time delays by three types of statistical distributions: uniform distribution, exponential distribution and normal distribution. The probability density function of the uniform distribution can be expressed as follows

$$P_U(\tau_i) = \begin{cases} 0 & \tau_i < 0 \\ \frac{\text{STR}}{2\tau_0} & 0 < \tau_i < \frac{2\tau_0}{\text{STR}} \\ 0 & \frac{2\tau_0}{\text{STR}} < \tau_i. \end{cases} \quad (4.24)$$

The uniform distribution has a constant probability among a fixed interval $[0, \frac{2\tau_0}{\text{STR}}]$. Figure 4.13 (a) shows an example of the the fire time delays generated by uniform distribution when STR equals 2. The fire time delays shows a nearly random pattern within the interval $[0, \tau_0]$.

The probability density function of the exponential distribution is given by

$$P_E(\tau_i) = \begin{cases} 0 & \tau_i < 0 \\ \frac{\text{STR}}{\tau_0} e^{-\frac{\text{STR}}{\tau_0} \tau_i} & 0 < \tau_i. \end{cases} \quad (4.25)$$

The expected value of the exponential distributed fire time delays is given by $\frac{\tau_0}{\text{STR}}$. Figure 4.13 (b) shows an example of the the fire time delays generated by exponential distribution when STR equals 2. Compared to the fire time delays generated by the uniform distribution, there are more chances to generate a small time delay from the exponential distribution. The unfrequent large time delays would lead to time gaps between different groups of closely fired sources.

The probability density function of the normal distribution is given by

$$P_N(\tau_i) = \frac{1}{\sqrt{2\pi v^2}} e^{-\frac{(\tau_i - \bar{\tau})^2}{2v^2}} \quad (4.26)$$

where we set the mean $\bar{\tau} = \frac{\tau_0}{\text{STR}}$ and the variance $v^2 = \bar{\tau}(1 - \frac{\bar{\tau}}{\tau_0})$. The time delays generated are focused on the mean value and vary around the mean within a small range (Figure 4.13 (c)). Choosing the normal distribution can be risky as the strong interferences might be concentrated near weak reflections in common receiver gathers. This firing scheme resembles the dithering approach proposed by Hampson et al. (2008).

For each one of the distributions, we let the STR varies from 1 to 20. For each value of the STR, we generalized 20 realizations of fire time delay based on the same distribution. For each realization, we apply numerical blending to an unblended common receiver gather from the Gulf of Mexico dataset and then utilize the iterative SSA filtering to separate the blended sources. Figure 4.13 (d), (e) and (f) shows the quality of the deblending versus $1/\text{STR}$ corresponding to the uniform, exponential and normal distribution respectively.

For all three distributions, the quality of deblending increases as we spend more time in the field to avoid severe interferences. In contrast, as we increase the STR, the total acquisition time becomes shorter. The data set is compressed heavily by the blending operator, and it is more difficult for the algorithm to achieve satisfactory separation result. Compared to the exponential and normal distributions, using the uniform distribution to generate fire time delays would lead to more stable deblending results. This is because the quality factor corresponding to the uniform distribution does not vary as much as the exponential and normal distributions. Acceptable deblending results are ensured for this specific data set and acquisition design when the STR is smaller than 2.

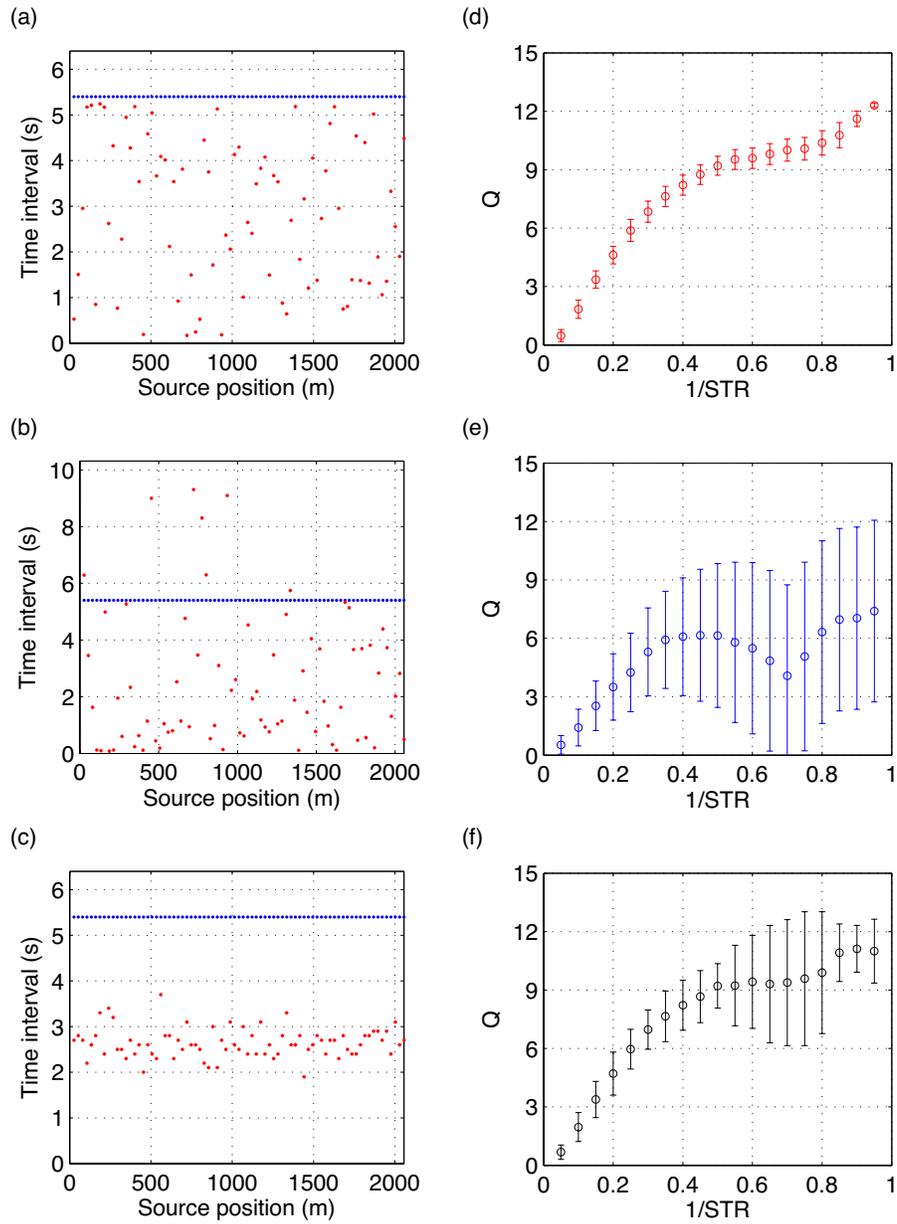


Figure 4.13: Evaluation of deblending based on fire time delay distributions: (a), (b) and (c) shows the conventional fire time (blue) over the simultaneous source fire time distribution (red) for uniform, exponential, and normal distribution, respectively. (d), (e) and (f) display the quality of deblending versus the inverse of STR when fire time delays are generated from Uniform, exponential, and normal distribution, respectively.

4.4.2 Nonlinear optimization of fire time delays

In the second aspects of our analyses, we posed the fire time optimization as a global optimization problem and adopted simulated annealing to solve for a set of fire time delays that enables the optimal separation result. The optimization is based on the assumption that the unblended data are priorly known. The assumption is not realistic. However, in many situations, practical synthetics are approachable, and the method allows to obtain an impression of the optimal set of fire time delays.

In physics, annealing involves controlled heating and cooling of the material to maximize the crystallization. A solid is first heated until all its particles are arranged randomly in the liquid phase. The heating process is followed by a cooling process in which the temperature is lowered slowly to ensure that all the particles are settled in the state of minimum energy (Chaikin and Lubensky, 2000; Bóna et al., 2009). The annealing process analogies the problem of non-linear optimization and therefore, simulated annealing can be utilized to find the optimal fire time delays that minimize the following cost function

$$E = \|\mathbf{d}_S - \mathbf{d}_{true}\|_2^2, \quad (4.27)$$

where E analogies the particle energy that one would like to minimize during the simulated annealing.

In each iteration of the simulated annealing method, perturbations of the model are drawn from a temperature scaled Cauchy distribution (Szu and Hartley, 1987; Ryden and Park, 2006) as follows

$$\mathbf{t}^{\nu+1} = \mathbf{t}^{\nu} + \Delta t \left(\frac{T_k}{T_0} \right) (\boldsymbol{\eta}_1 \tan(\frac{\boldsymbol{\eta}_2 \pi}{2})), \quad (4.28)$$

where \mathbf{t} is a vector that denotes firing times for all sources. T_k denotes the cooling schedule that slowly decreases the initial temperature T_0 in the annealing process

$$T_k = \boldsymbol{\beta}^k T_0. \quad (4.29)$$

$\boldsymbol{\eta}_1$, $\boldsymbol{\eta}_2$, and $\boldsymbol{\beta}$ are vectors with random numbers generated between 0 and 1 in each iteration. The perturbation of the fire time delay is a random number bounded by Δt for each source. The large time gaps between sources are avoided by imposing the aforementioned bounds. Once the fire time delays are perturbed, we can numerically blend the data set and apply iterative SSA filtering to deblend the simultaneous source data. The thermal energy E can be then updated using the deblended data. The acceptance of model update is based on a criterion proposed by Metropolis et al. (1953), which imitates the evolution of a solid at a given temperature. If the resulting thermal energy E is lower than the previous energy, the

perturbation are accepted. Otherwise, the model may also be accepted only if

$$r < e^{-\Delta E/T_k}, \quad (4.30)$$

where r is a random number between 0 and 1 and ΔE is the difference between the previous and current thermal energy. The method converges at low temperatures where no further improvement in the cost function is achieved (Velis and Ulrych, 1996). As we show in Figure 4.14, the algorithm accepts various perturbations in early stages to ensure convergence to a global minimum. As the algorithm progresses, the temperature is lower until the solution reaches the global minimum. We have tuned all the parameters of our simulated annealing code to achieve convergence in about 1000 iterations. Figure 4.15 shows the resulting fire time delays after optimization. Figure 4.16 shows a common receiver gather after IRR deblending with the optimized firing scheme. In this example, we fixed the STR as 2 with around 50 % of acquisition time is saved by simultaneous source acquisition. After separation, the solution becomes comparable to the unblended data with Q_S equal to 12.2.

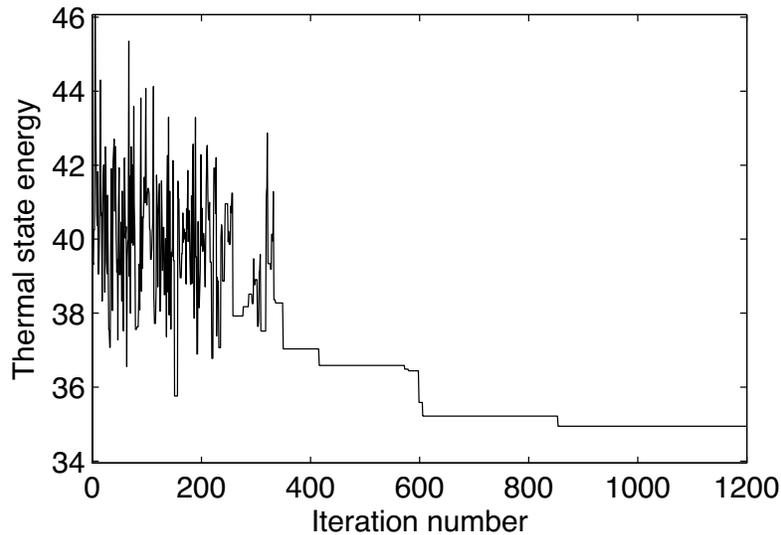


Figure 4.14: Thermal state energy versus iteration numbers in Simulated Annealing inversion method.

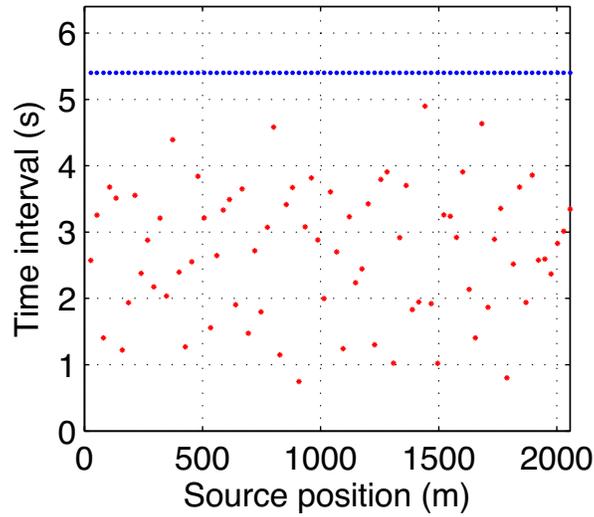


Figure 4.15: Optimal time delays inverted via fast simulated annealing. In this example, the STR equals to 2.

4.5 Conclusions

This chapter illustrates an iterative rank reduction algorithm based on SSA filtering for separating simultaneous source data. The proposed algorithm can be classified among the family of deblending methods via inversion. By implementing rank reduction with a projection operator, the SSA filter, solutions are constrained to be low-rank in Hankel matrices extracted from small spatial-temporal windows in common receiver gathers. The latter is important because the SSA method is a valid denoising and reconstruction technique for a superposition of plane waves. In a small window, the data can be approximated via a limited number of dips plus incoherent interferences caused by the blending process. Convergence of this algorithm can be achieved if the pseudo-deblended data are adopted as the initial solution. Through tests with synthetic examples made by blending a traditional marine acquisition, we show that the interferences of the wavefields can be effectively suppressed. This algorithm could also see applications in multidimensional cases by adopting high order SVD or tensor-based dimensionality reduction methods.

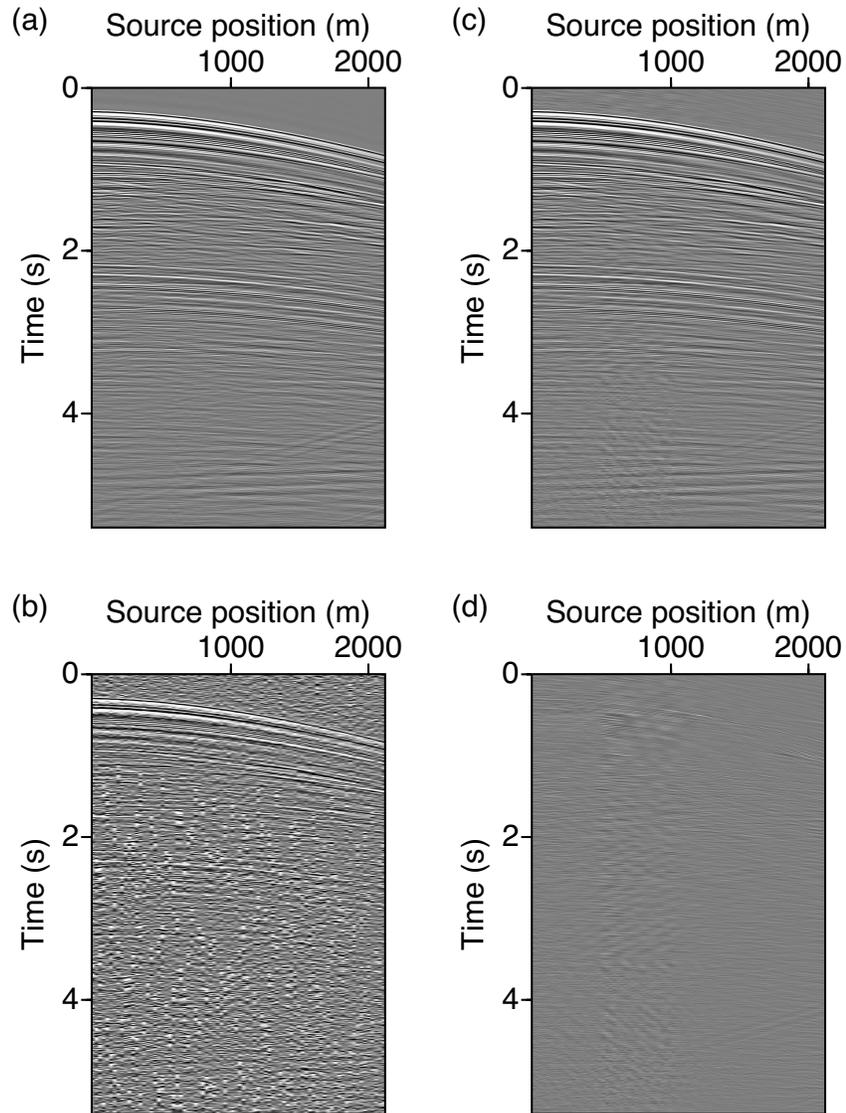


Figure 4.16: Results of deblending for a common receiver gather using optimal source fire time delays. (a) The original common receiver gather from the Gulf of Mexico dataset. (b) Pseudo-deblended common receiver gather (c) Deblended common receiver gather with iterative SSA filtering (d) Difference section between original and deblended data.

We also studied the impact of the firing scheme on the separation of simultaneous source data. Through tests with different survey time ratios and distributions of fire time delays, we can examine the dependence of deblending with the distribution of time delays. The quality of source separation would be improved as the STR value decreases for all uniform, exponential and normal distributed time intervals. Compared to the exponential or normal distribution of firing time delays, the uniform distribution shows a more stable performance. We also applied fast simulated annealing algorithm to optimize the firing scheme assuming we have prior information about the unblended data. However, we believe that a more sophisticated cost function than the one we have adopted could be used to characterize the incoherency of the time delays and lead to optimal deblending results. This work can be regarded as a starting point towards the optimization of the spatial and temporal distribution of sources for blended acquisition.

CHAPTER 5

Shot-profile least-squares migration of blended data

5.1 Introduction

In this chapter, we present a new scheme for the migration of the simultaneous source seismic data that uses the least-squares migration formalism (Chavent and Plessix, 1999; Nemeth et al., 1999; Duijndam et al., 2000; Kuhl and Sacchi, 2003; Plessix and Mulder, 2004; Symes, 2008; Kaplan et al., 2010a). To compute the contributions of each single unblended shot in the migrated image of the blended data, one could correlate the blended receiver-side wavefield with each crosstalk free and time-shifted unblended source-side wavefield. The method leads to partial images that one could acquire via shot-profile migration (Shen and Symes, 2008; Kaplan et al., 2010a). Also, we adopt the previously discussed gradient projection method for the problem of least-squares migration in which a projection is used to eliminate residual cross-talk in shot-index migrated gathers. The noise rejection projection of our least-squares migrations is the SSA filter in the shot-index common image domain. Through synthetic tests, we find the method effectively suppresses simultaneous source crosstalk and improves the quality of the migrated shot-index image gathers. The outputs of our method are a crosstalk-free migrated image and a deblended seismic volume that can be used in a subsequent processing flow.

5.2 Theory

5.2.1 Preliminaries

We first present a brief review of direct migration of blended seismic data and define operators and notations that are adopted in the proposed algorithm for least-squares migration

of blended data.

Migration of conventional seismic data

Considering seismic data collected from the conventional acquisition, one partial image (m_i) in shot-profile wave equation migration can be written as

$$m_j(x, z) = \sum_{\omega} W_{S_j}(x, z, \omega) W_{R_j}^*(x, z, \omega), \quad (5.1)$$

where W_{S_j} denotes the source-side wavefield and $W_{R_j}^*$ is the conjugate transpose of the receiver-side wavefield for the j -th shot. The source-side wavefield W_{S_i} is computed by

$$W_{S_j}(x, z, \omega) = \mathcal{G}^+ s_j(x, z, \omega), \quad (5.2)$$

where $s_j(x, z, \omega) = f(\omega)\delta(x - x_{s_j}, z - z_{s_j})$ is the source signature function composed of the source location x_{s_j}, z_{s_j} and the source wavelet $f(\omega)$. The image is acquired by summation over the frequency ω . The operator \mathcal{G}^+ stands for the wavefield extrapolator that forward propagates the source signature at the earth's surface to any depth. Similarly, the receiver-side wavefield is computed by applying the backward propagation operator \mathcal{G}^- directly on the j -th shot record $d_j(x, \omega)$

$$W_{R_j}(x, z, \omega) = \mathcal{G}^- d_j(x, \omega). \quad (5.3)$$

The final image is acquired by summing all the partial images as follows (Claerbout, 1992)

$$m(x, z) = \sum_{j=1}^{N_S} m_j(x, z), \quad (5.4)$$

where N_S denotes the total number of common source gathers.

The blended migration

In simultaneous source acquisition, seismic data are acquired by firing multiple shots with small random time delays. The response to the blended sources is called a super shot. At this stage, for simplicity of the notations, we assume the N_s shots are fired at the same time with small random time delays and only one super shot gather is acquired. However, the method can be adapted to the cases that the total number of shots can be divided into groups and the shots in each group are blended into a super shot. The blended supershot

gather of N_S shots can be written as follows

$$s^b(x, z, \omega) = \sum_{j=1}^{N_S} s_j(x, z, \omega) e^{-i\omega\tau_j}, \quad j = 1, \dots, N_S, \quad (5.5)$$

where $i = \sqrt{-1}$ and τ_j indicates the time delay of source j . Combining Equation 5.5 with Equation 5.2, one can acquire the expression for the blended source-side wavefield as follows

$$\begin{aligned} W_S^b(x, z, \omega) &= \mathcal{G}^+ s^b(x, z, \omega) \\ &= \mathcal{G}^+ \sum_{j=1}^{N_S} s_j(x, z, \omega) e^{-i\omega\tau_j} \\ &= \sum_{j=1}^{N_S} \mathcal{G}^+ s_j(x, z, \omega) e^{-i\omega\tau_j} \\ &= \sum_{j=1}^{N_S} W_{S_j}(x, z, \omega) e^{-i\omega\tau_j}. \end{aligned} \quad (5.6)$$

The blended source-side wavefield is equivalent to the summation of each phase-shifted unblended sources-side wavefield.

The computation of the receiver-side wavefield of simultaneous source data is similar to the blended source-side wavefield. Since the responses from the blended sources are recorded by the same set of receivers, the recorded blended data $b(x, \omega)$ can be modelled via

$$b(x, \omega) = \sum_{k=1}^{N_S} d_k(x, \omega) e^{-i\omega\tau_k}, \quad k = 1, \dots, N_S, \quad (5.7)$$

where $d_k(x, \omega)$ denotes the unblended shot record that one would have acquired from the conventional acquisition. The blended receiver-side wavefield is computed by substituting Equation 5.7 into Equation 5.3 as follows

$$\begin{aligned} W_R^b(x, z, \omega) &= \mathcal{G}^- b(x, \omega) \\ &= \mathcal{G}^- \sum_{k=1}^{N_S} d_k(x, \omega) e^{-i\omega\tau_k}, \\ &= \sum_{k=1}^{N_S} \mathcal{G}^- d_k(x, \omega) e^{-i\omega\tau_k}, \\ &= \sum_{k=1}^{N_S} W_{R_k}(x, z, \omega) e^{-i\omega\tau_k}. \end{aligned} \quad (5.8)$$

The blended source-side wavefield can also be acquired via the summation of each phase-

shifted unblended receiver-side wavefield.

The final image of the supershot gather, m^b , can be acquired applying the cross-correlation imaging condition as follows

$$\begin{aligned} m^b &= \sum_{\omega} W_S^b(x, z, \omega) W_R^{b*}(x, z, \omega) \\ &= \sum_{\omega} \sum_{j=1}^{N_S} \sum_{k=1}^{N_S} W_{S_j}(x, z, \omega) W_{R_k}^*(x, z, \omega) e^{-i\omega(\tau_j - \tau_k)}. \end{aligned} \quad (5.10)$$

In the cases where the indices $j = k$, the phase-shift term from source-side wavefield and receiver-side wavefield cancel each other. Equation 5.15 reduces to the summation of migrated images of conventional seismic data. However, when indices $j \neq k$ artifacts are generated by crosstalk. To summarize, the image acquired by the blended supershot can be expressed as follows

$$m_i^b(x, z) = m(x, z) + e(x, z), \quad (5.11)$$

where $e(x, z)$ stands for the source crosstalk artifacts given by

$$e(x, z) = \sum_{\omega} \sum_{j=1}^{N_S} \sum_{k=1}^{N_S} W_{S_j}(x, z, \omega) W_{R_k}^*(x, z, \omega) e^{-i\omega(\tau_j - \tau_k)}, \quad k \neq j. \quad (5.12)$$

These source crosstalk artifacts $e(x, z)$ can be attenuated by the blended migration when random phase encoding is applied (Krey, 1987; Romero et al., 1999). Detailed phase-encoding analyses can be found in Schuster et al. (2011) and by Godwin and Sava (2013). Besides, the source blending process significantly decreases the computational cost of migration. In the blended migration, the wavefield extrapolation is only applied to the super shot gather instead of the N_s single shots as in conventional seismic migration. Therefore, blending has become as a popular strategy to speed up processes such as least-squares migration (Tang, 2007; Dai and Schuster, 2012; Herrmann and Li, 2012) and full waveform inversion (Krebs et al., 2009; Anagaw and Sacchi, 2014).

5.2.2 Shot-profile migration and de-migration of blended data

In this chapter of the thesis, we study the shot-profile migration of the simultaneous source seismic data. The motivation is to avoid stacking the partial images that correspond to the unblended shot records. The stacking collapses information from different source-receiver pairs, and therefore prestack information is lost. Besides, in the least-squares migration, stacking of partial images leads to an underdetermined inverse problem, which will limit the

capability of data fitting (Kaplan et al., 2010b). In the shot-profile migration of the blended data, the source-side wavefield corresponding to the j -th shot is defined as follows

$$W_{S_j}^b(x, z, \omega) = W_{S_j}(x, z, \omega) e^{-i\omega\tau_j} \quad j = 1, \dots, N_S. \quad (5.13)$$

In other word, we shift the unblended source-side wavefield of shot j according to its initiation time. Compared to the blended source-side wavefield (Equation 5.6), the summation over j to the total number of shots is avoided. We then use $W_{S_j}^b$ to compute the contribution of the j -th shot in the blended image as follows

$$m_j^b(x, z) = \sum_{\omega} W_{S_j}^b(x, z, \omega) W_R^{b*}(x, z, \omega) \quad (5.14)$$

$$= \sum_{\omega} W_{S_j}(x, z, \omega) e^{-i\omega\tau_j} \sum_{k=1}^{N_S} W_{R_k}^*(x, z, \omega) e^{i\omega\tau_k}, \quad (5.15)$$

Similarly, when the indices $k = j$, the phase-shift term from the j -th single source-side wavefield and the receiver-side wavefield cancel each other. Equation 5.15 reduces to the contribution of the j -th unblended shot in the conventionally migrated image. However, when $k \neq j$, severe artifacts have been generated. The partial image of the j -th shot in the blended migration can be expressed as follows

$$m_j^b(x, z) = m_j(x, z) + e_j(x, z), \quad (5.16)$$

where e_j denotes the source crosstalk artifacts in the partial image given by

$$e_j(x, z) = \sum_{\omega} \sum_{k=1}^{N_S} W_{S_j}(x, z, \omega) W_R^{b*}(x, z, \omega) e^{-i\omega(\tau_j - \tau_k)}, \quad j \neq k. \quad (5.17)$$

$e_j(x, z)$ has less artifacts compared to $e(x, z)$ in the blended migration (equation 5.12) as the single source-side wavefield is cleaner than the blended source-side wavefield. As we will discuss later, the source crosstalk artifacts can be effectively eliminated via the least squares migration. One can repeat the process for each of the unblended shot and an volume of image can be computed as follows,

$$\begin{bmatrix} m_1^b(x, z) \\ m_2^b(x, z) \\ \vdots \\ m_N^b(x, z) \end{bmatrix} = \begin{bmatrix} \sum_{\omega} W_{S_1}^b(x, z, \omega) \\ \sum_{\omega} W_{S_2}^b(x, z, \omega) \\ \vdots \\ \sum_{\omega} W_{S_N}^b(x, z, \omega) \end{bmatrix} [W_R^{b*}(x, z, \omega)]. \quad (5.18)$$

In the blended shot-profile migration, the wavefield propagation of the source-side wavefield are computed N_S times and the wavefield propagation for the receiver-side wavefield are computed only once for all the shots.

Let us assume the sources are fired with randomized time delays. In the shot-index common image domain, since the fire time for the j -th shot has been corrected when computing the source-side wavefield ($j = k$ in Equation 5.15), the desired reflections would remain coherent along the shot-index axis. The source crosstalk artifacts in the partial images will appear random as the fire time delays τ_k are randomized. Therefore, a coherence constraint can be applied to suppress the incoherent source crosstalk artifacts while preserving the desired coherent signal. The coherence constraint in the shot-index domain will not be affected by the accuracy of the migration velocity. However, with the correct velocity, the desired signal in shot-index common image domain will appear less complex (flat). The suppression of the source interferences would be easier compared to using inaccurate migration velocities.

With the migration operator in mind we now study the forward operator that models the blended data from the image volume, the latter is called de-migration in the context of least-squares migration. One can solve for the blended data using Equation 5.8 as follows

$$\hat{b}(x, \omega) = \mathcal{G}^+ \hat{W}_R^b(x, z, \omega), \quad (5.19)$$

where the blended receiver-side wavefield $W_R^b(x, z, \omega)$ can be acquired by reversing the processes denoted in Equation 5.19 as follows

$$\hat{W}_R^b(x, z, \omega) = \begin{bmatrix} W_{S_1}^b(x, z, \omega) & W_{S_2}^b(x, z, \omega) & \cdots & W_{S_N}^b(x, z, \omega) \end{bmatrix} \begin{bmatrix} \mathcal{R}_\omega[m_1^b(x, z)] \\ \mathcal{R}_\omega[m_2^b(x, z)] \\ \vdots \\ \mathcal{R}_\omega[m_N^b(x, z)] \end{bmatrix}. \quad (5.20)$$

The operator \mathcal{R}_ω entails duplicating the partial image into all the frequency slices. The duplicated volume is multiplied by the unblended and phase-shifted source-side wavefield. The blending between the simultaneous sources are denoted via matrix multiplication in this notation. Similar to the blended shot-profile migration, in de-migration, the source-side wavefield are computed N_s times corresponding to the total number of blended shots. The backward propagation from receiver-side wavefield to data is computed only once.

5.3 Least-squares migration via gradient projection

Since no stacking is applied to the shot-profile migration of the blended data, the simultaneous source artifacts will not be suppressed. We propose to use the least-squares migration

to suppress the source crosstalk and at the same time to improve the quality of the image. From now on, we will adopt the operator formulation for the blended shot-profile migration and de-migration as follows

$$\mathbf{m} = \mathcal{L}^* \mathbf{b}, \quad (5.21)$$

$$\hat{\mathbf{b}} = \mathcal{L} \mathbf{m}, \quad (5.22)$$

where $\mathbf{m} = [m_1^b(x, z), m_2^b(x, z), \dots, m_N^b(x, z)]^T$ denotes an array of partial images in which each element of the array corresponds to contribution of each shot. \mathcal{L}^* denotes the adjoint operator which is the migration operator that maps the blended data into the array of partial images. \mathcal{L} is the forward operator that simulates the blended data from \mathbf{m} . To suppress the simultaneous source artifacts in the image volume and to improve the quality of the images, the least-squares shot-profile migration of blended data is introduced. The least-squares migration intends to find an optimal \mathbf{m} that honours the blended observations. It is equivalent to finding the minimum of the cost function

$$J(\mathbf{m}) = \|\mathbf{b} - \mathcal{L} \mathbf{m}\|_2^2. \quad (5.23)$$

To suppress the simultaneous source artifacts, we adopt the previously discussed coherent constraint \mathcal{C} that the desired image is coherent along the shot-index common image domain as follows

$$J(\mathbf{m}) = \|\mathbf{b} - \mathcal{L} \mathbf{m}\|_2^2 \quad s.t. \quad \mathbf{m} = \mathcal{C}[\mathbf{m}] \quad (5.24)$$

The solution to Equation 5.24 is acquired via the gradient projection method as follows

$$\mathbf{m}^{\nu+1} = \mathcal{P}_C[\mathbf{m}^\nu - \lambda \mathcal{L}^*(\mathcal{L} \mathbf{m}^\nu - \mathbf{b})], \quad (5.25)$$

where λ denote the step size. As is discussed in Chapter 2, when the step size λ is sufficiently small, the convergence of the gradient projection method is guaranteed. In our method, we choose $\lambda = 2/\alpha$, where α denotes the largest eigenvalue of $\mathcal{L}^* \mathcal{L}$. \mathcal{P}_C is the projection operator which in our algorithm is the SSA reduced-rank filter discussed in Chapter 4. Gradient descent iterations are adopted in order to search for an optimal model that best fits the blended observation. In each iteration, the current solution is projected to a set where \mathbf{m} is coherent along the shot-index axis.

5.4 Examples

5.4.1 Two layer model

We first adopt a simple two-layer model to test the performance of the proposed method. A synthetic dataset was generated by the born modelling method (i.e. an inverse crime is committed) (Shen and Symes, 2008). We adopted a Ricker wavelet with central frequency equals 20Hz when generating the data. Both sources and receivers are deployed on a regular grid. The receiver interval is 10m, and the source interval is 20m. The data are then numerically blended. Figure 5.1 shows the acquisition design for the blending experiment. In each round, 5 shots are fired with small random time delays. The responses are recorded by the same set of receivers. Then all the shots are moved to the next adjacent location and fire again for the next round. A more realistic design of simultaneous source acquisition has been discussed in (Abma et al., 2012).

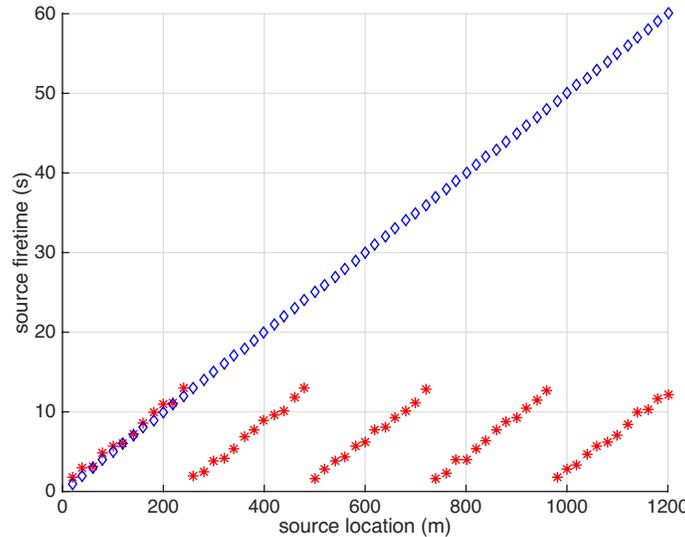


Figure 5.1: Distribution of firing times for simultaneous source acquisition (red) compared to conventional acquisition (blue). In this example about 75% of acquisition time is saved by source blending. The computational efficiency exhibits a similar pattern for the proposed least squares migration.

We then apply the proposed shot-profile least-squares migration to suppress source crosstalk. The forward and adjoint operator pairs are Gazdag de-migration and migration operators

(Gazdag, 1978). In each iteration, we fixed the rank ($r = 2$) for singular spectrum analysis and applied the filter along each shot-index common image gather as the projection operator. We adopted a fixed step size $\lambda = 2.5$ and the algorithm converges after about 30 iterations (Figure 5.2). We show the migration results from direct migration of blended data using source extended imaging condition and the results from the proposed least-squares migration. Figure 5.3 shows the shot-index common image gathers at $x = 600$ and Figure 5.4 shows the partial images corresponding to the centre shots. The source crosstalk has also been effectively annihilated in these domains. However, in Figure 5.5, the differences between the blended migrated stack section and the stack from least-squares migration are not obvious. Stacking is effective in suppressing random interferences, and the two-layer model is comparatively simple. However, next examples illustrate the benefits of the least-squares migration. Figure 5.6(b) shows a pseudo-deblended shot gather and Figure 5.6(c) shows the modelled data after 30 iterations. The source interferences are effectively suppressed, and the modelled data becomes comparable to the data that one would acquire via conventional acquisition. We assume the method could deblend simultaneous source data in the presence of inaccurate velocity and multiples. In the shot-index common image domain, improperly migrated primaries and multiples appear as curved events. The coherence constraint for deblending can separate the coherent multiples from source crosstalk. The migration and de-migration operators can be viewed as a particular transform that separate signal and the blending noise in shot-index domains.

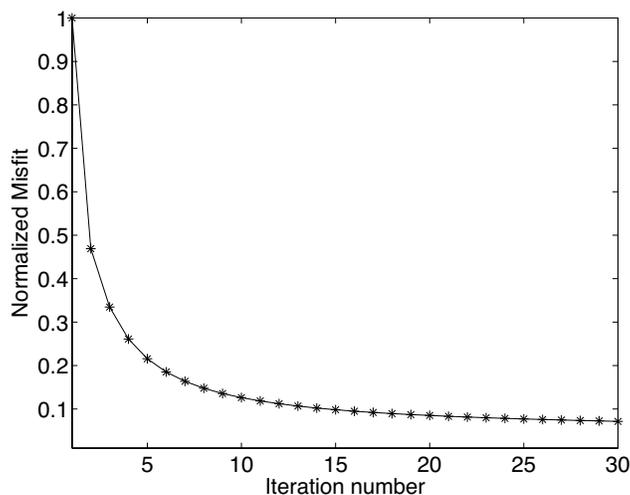


Figure 5.2: Convergence of the gradient projection algorithm for the two-layer example. The algorithm converges within 30 iterations.

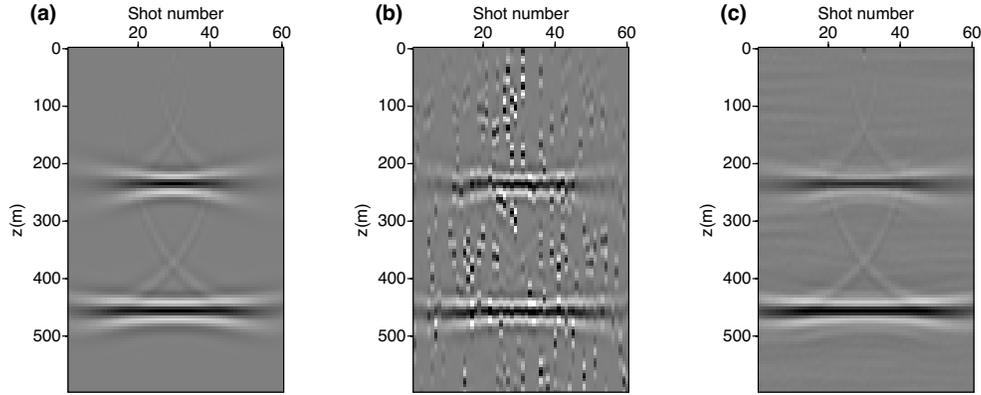


Figure 5.3: Migration results of the two-layer model: a shot-index common image gather ($x = 600$). (a) A shot-index common receiver gather from conventional unblended shot-profile migration; (b) A shot-index common receiver gather from blended shot-profile migration; (c) A shot-index common receiver gather from blended least-squares shot-profile migration.

5.4.2 SAIG velocity model

For the second example, Figure 5.7 shows the velocity model that is utilized for generating seismic data. The sources are sampled at the earth surface with an interval of 20 meters from 1km to 5km. The receivers are deployed on the first layer of the model also with 20m intervals to simulate the ocean bottom nodes. Similar to the previous example, we assume in each group, 5 shots are fired with small random time delays, and a total number of 20 super shots are generated via simultaneous source acquisition. We adopted the split-step migration (Thomson and Chapman, 1983) and de-migration operators with the true given migration velocity (Figure 5.7). The data set is generated via Born modelling and we adopted a Ricker wavelet with central frequency equals 20Hz when generating the data.

In this example, we divide the shot-index common image gathers into small patches of size 60×40 . As the signal in each patch would resemble the common image gathers of the previously shown two-layer experiment, SSA filter is effective in suppressing the simultaneous source artifacts. A similar SSA strategy is discussed by Cheng and Sacchi (2015) and Xiang et al. (2016). We also fixed the step-size for the least-squares migration and the method converges in 60 iterations (Figure 5.8). Figure 5.9 show the shot-index common image gathers at the centre location of the model ($x = 2.5km$). In addition to the suppression of the source crosstalk artifacts, the least-squares migration also extends the aperture of the events. Figure 5.10 compares the migrated image of one single shot ($S_x = 2.5km$). The

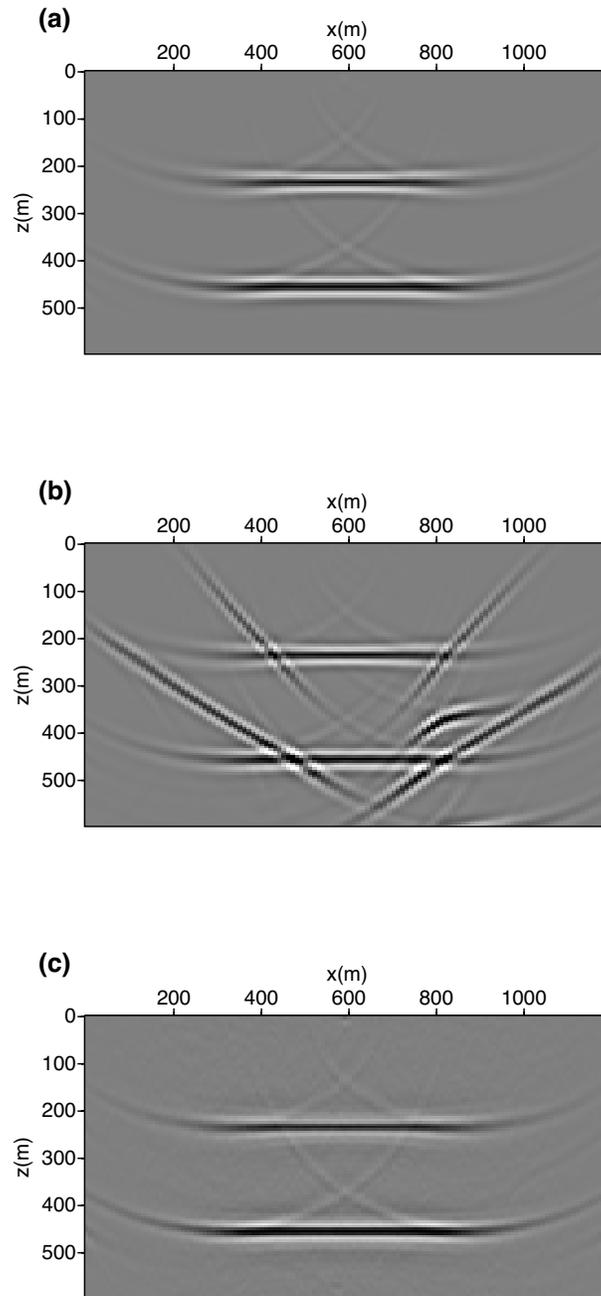


Figure 5.4: Migration results of the two-layer model: partial image for shot 30. (a) A partial image from conventional unblended shot-profile migration; (b) A partial image from the blended shot-profile migration; (c) A partial image from the blended least-squares shot-profile migration.

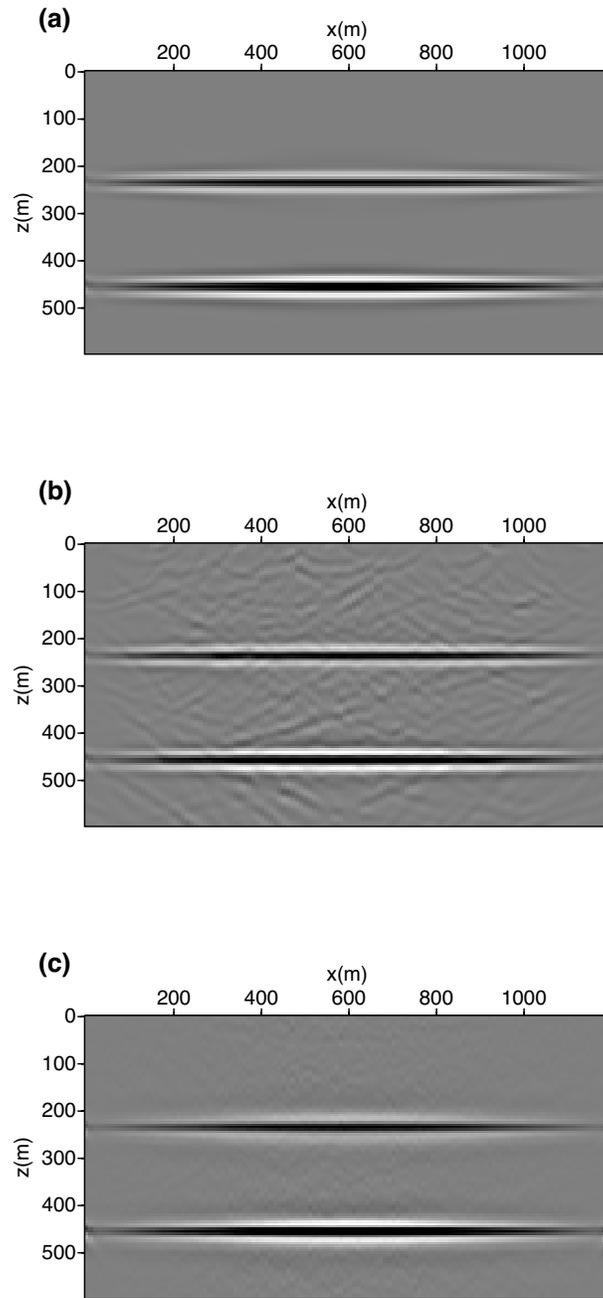


Figure 5.5: Migration results of the two-layer model: stacked image (a) Stacked image from conventional unblended shot-profile migration; (b) Stacked image from the blended shot-profile migration; (c) Stacked image from the blended least-squares shot-profile migration.

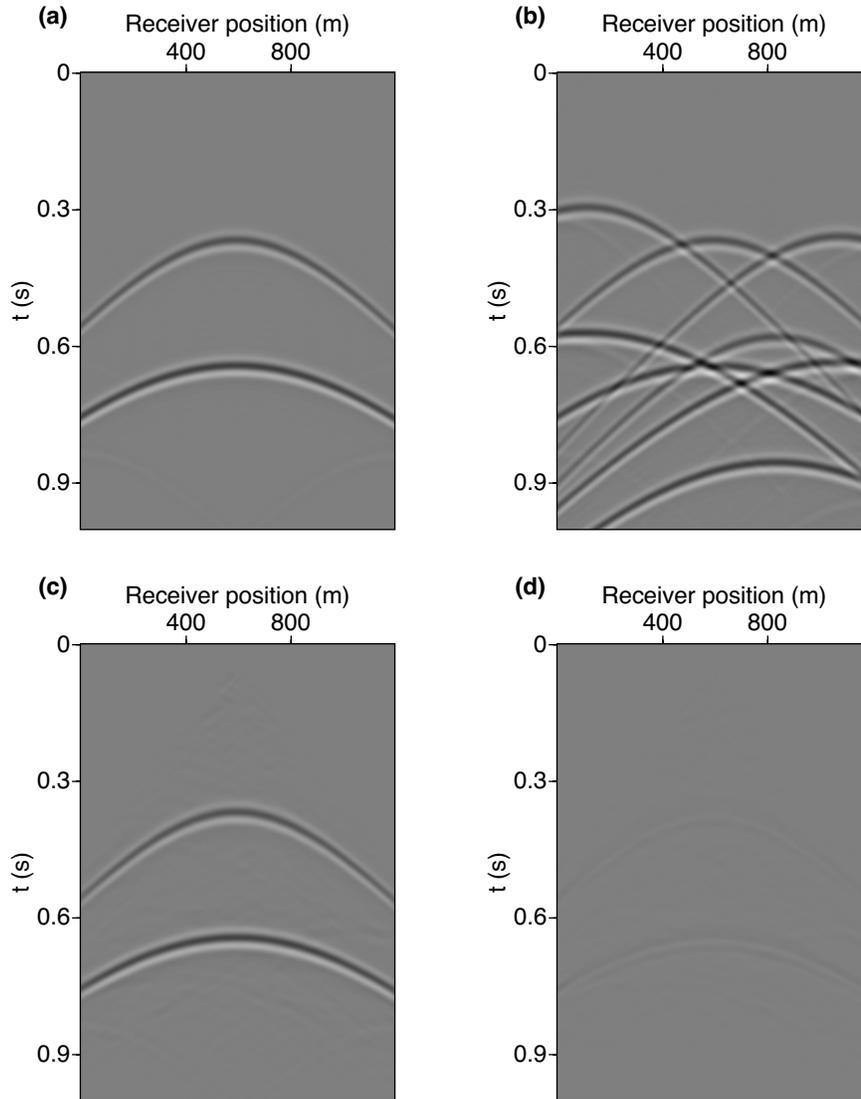


Figure 5.6: Results of deblending via the proposed algorithm for the two-layer model. (a) Ideal unblended shot gather (center shot) (b) Pseudo-deblended shot gather (c) Deblended shot gather after 30 iterations of gradient projection in the least-squares migration. The interferences are effectively suppressed.

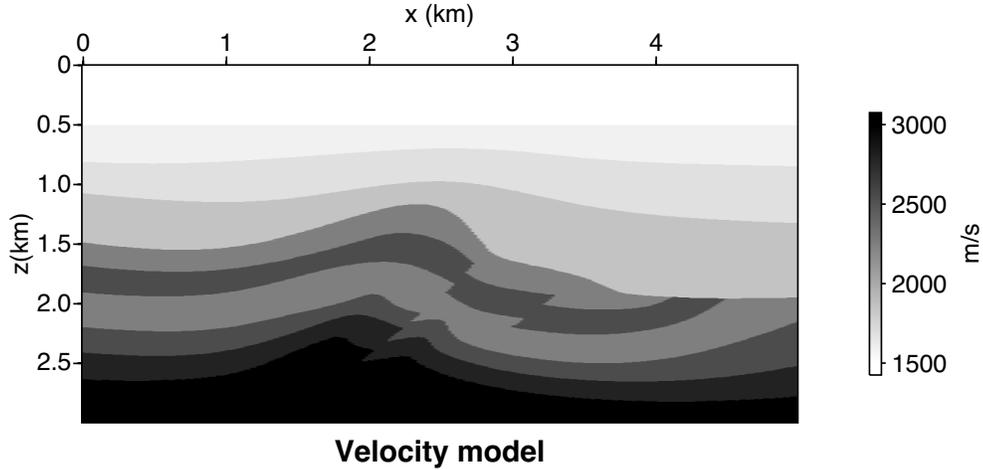


Figure 5.7: The SAIG velocity model.

source crosstalk artifacts have not been suppressed by stacking. Compared to the blended migrated image, the partial image from the least-squares migration significantly eliminates source crosstalk artifacts. Figure 5.11 shows stacked images from conventional migration, blended migration, and the least-squares migration, respectively. The crosstalk artifacts are suppressed by stacking the partial images. The least-squares migration provides an image with equalized amplitudes and a better suppression of blending artifacts. However, we are aware that, to improve the quality of the final stacked image, blended least-squares reverse time migration methods should be considered (Dai and Schuster, 2012; Xue et al., 2014). Figure 5.12 compares the deblending result at the central shot gather with the ideal unblended data (a) and pseudo-deblended shot gather (b). We find that the simultaneous source crosstalk artifacts are effectively suppressed.

5.4.3 Acoustic Marmousi2 model

We also tested the proposed algorithm with the Marmousi2 model (Martin et al., 2002). We modified the original P-velocity in Marmousi2 model, and reduce it size to 801×400 grid points (Figure 5.13). We assume both sources and receivers deployed on a regular grid throughout the earth surface. The source spatial interval is $50m$, and the receiver spatial interval is $5m$. We adopted a Ricker wavelet with central frequency equals $20Hz$ and used Born Modeling when generating the data. Again we blend 5 shot records, and a total number of 18 supershots are generated. For this example, we also applied the SSA rank reduction

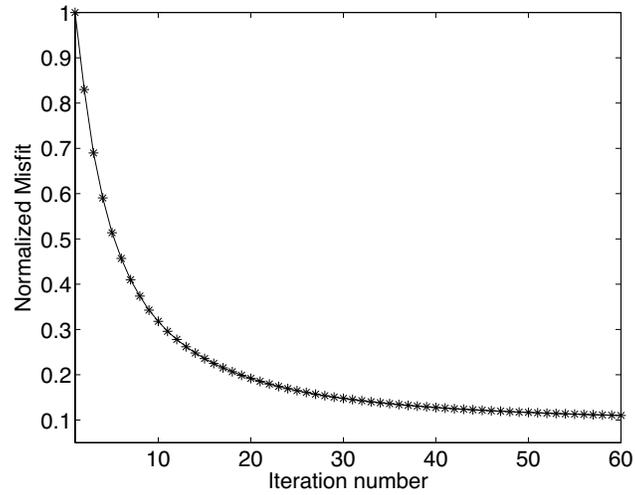


Figure 5.8: Convergence of the gradient projection algorithm for the SAIG example. The algorithm converges within 60 iterations.

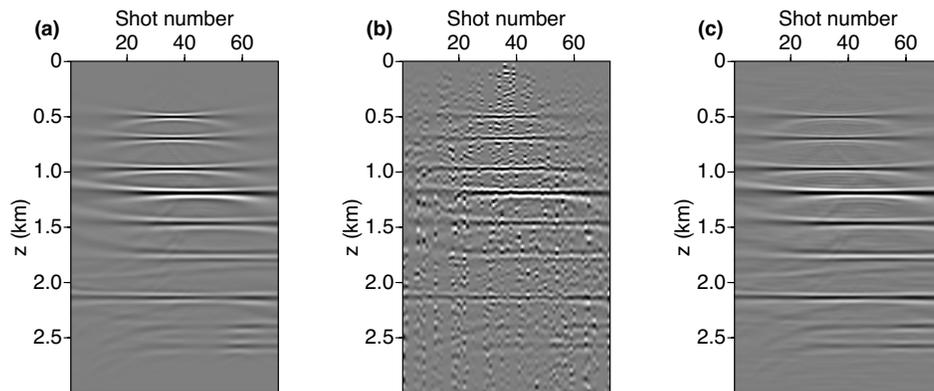


Figure 5.9: Migration results for SAIG velocity model: a shot-index common image gather ($x = 2.5\text{km}$) (a) A shot-index common receiver gather from conventional unblended shot-profile migration; (b) A shot-index common receiver gather from blended shot-profile migration; (c) A shot-index common receiver gather from blended least-squares shot-profile migration.

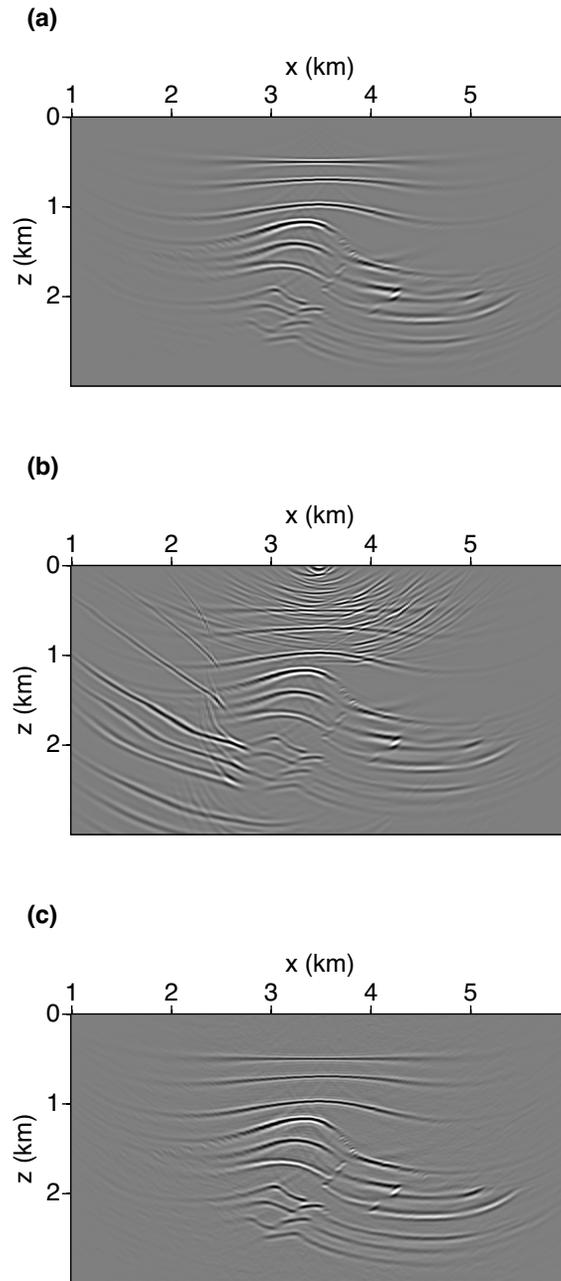


Figure 5.10: Migration results for SAIG velocity model: a partial image for a shot ($S_x = 2.5\text{km}$). (a) A partial image from conventional unblended shot-profile migration; (b) A partial image from the blended shot-profile migration; (c) A partial image from the blended least-squares shot-profile migration.

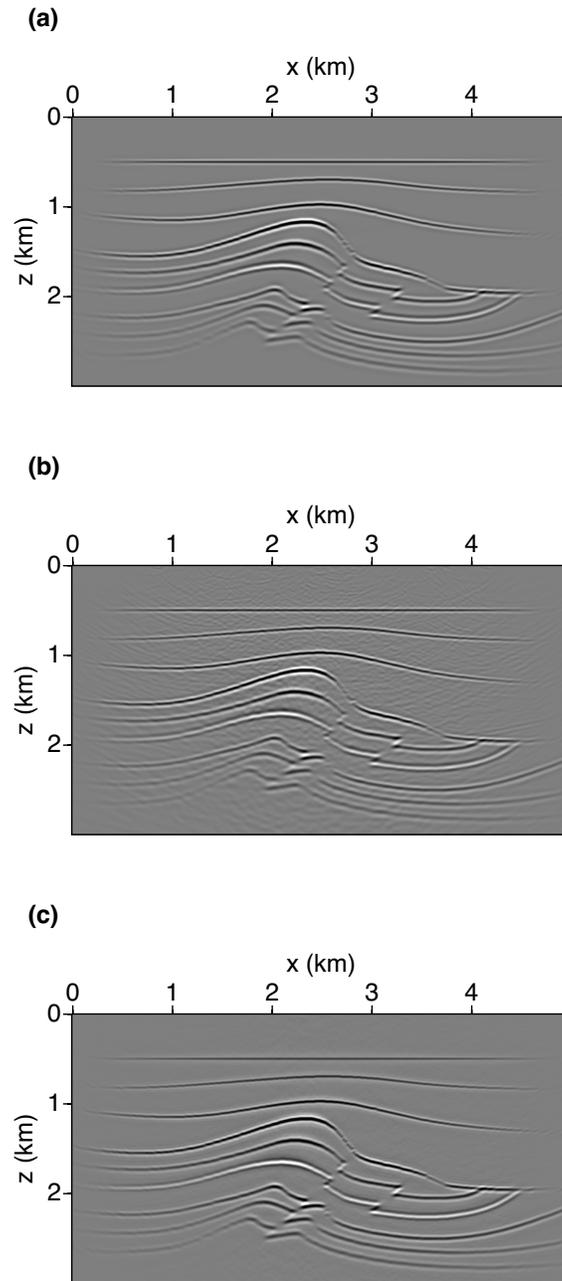


Figure 5.11: Migration results for SAIG velocity model: stacked image (a) Stacked image from conventional unblended shot-profile migration; (b) Stacked image from the blended shot-profile migration; (c) Stacked image from the blended least-squares shot-profile migration.

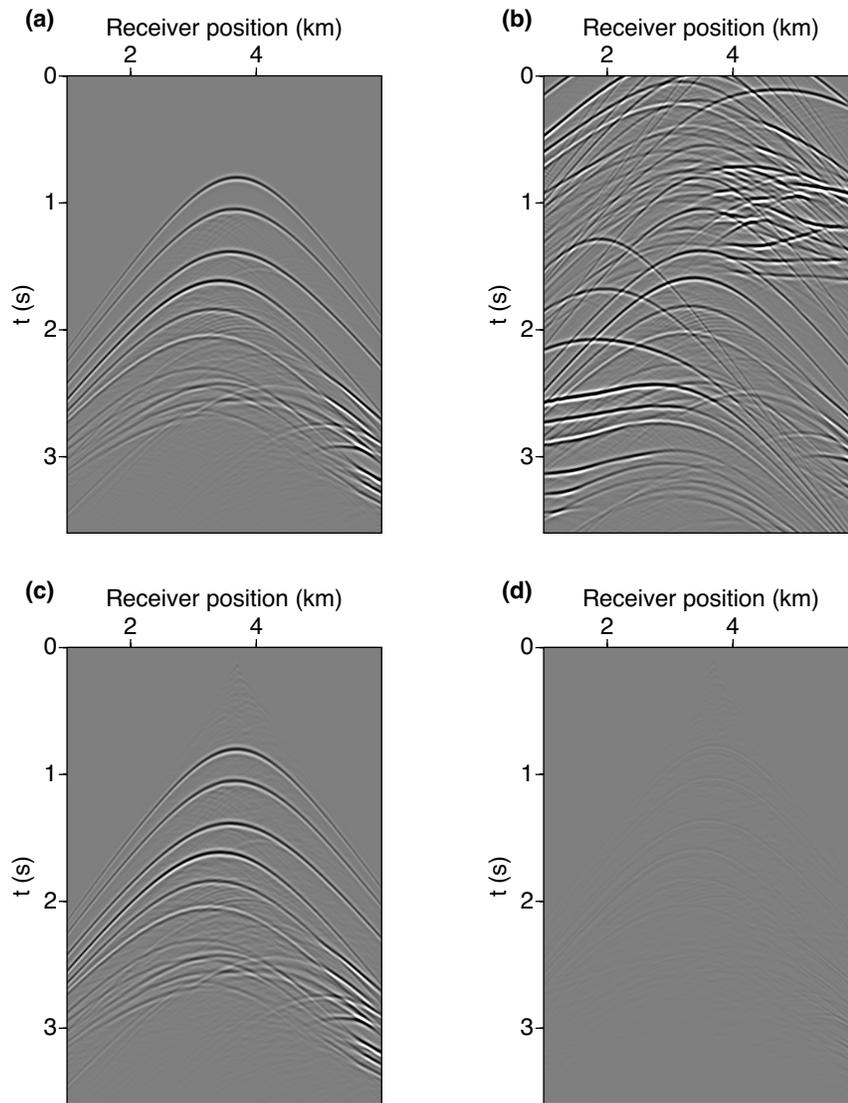


Figure 5.12: Deblending results via the proposed algorithm for the SAIG velocity model: (a) Ideal unblended shot gather (centre shot); (b) Pseudo-deblended shot gather; (c) Deblended shot gather after 60 iterations of least-squares migration; (d) Difference between (a) and (c).

filters in small patches in shot-index common image gathers. Figure 5.15-5.17 compares the migration results for conventional migration, blended migration and the blended least-squares migration, respectively. The blended least-squares migration enhances the resolution and provides equalized amplitudes. Figure 5.18 shows the results of deblending for shot number 40 after 80 iterations of least-squares migration. The proposed algorithm effectively suppressed the simultaneous source interferences.

5.5 Conclusions

We have presented a projected gradient method to solve the least-squares migration problem with constraints to suppress simultaneous source crosstalk artifacts. The proposed algorithm adopts the projected gradient descent method and iteratively searches for a solution that fits the observed data. The forward and adjoint operators are the shot-profile de-migration and migration of the blended data. The latter leads to shot-index common image gathers. Through tests with synthetic examples, we find that the method effectively suppresses source crosstalk artifacts in the image domain. Compared to the conventional migration methods, the least-squares migration also enhances the resolution and leads to images with balanced amplitudes.

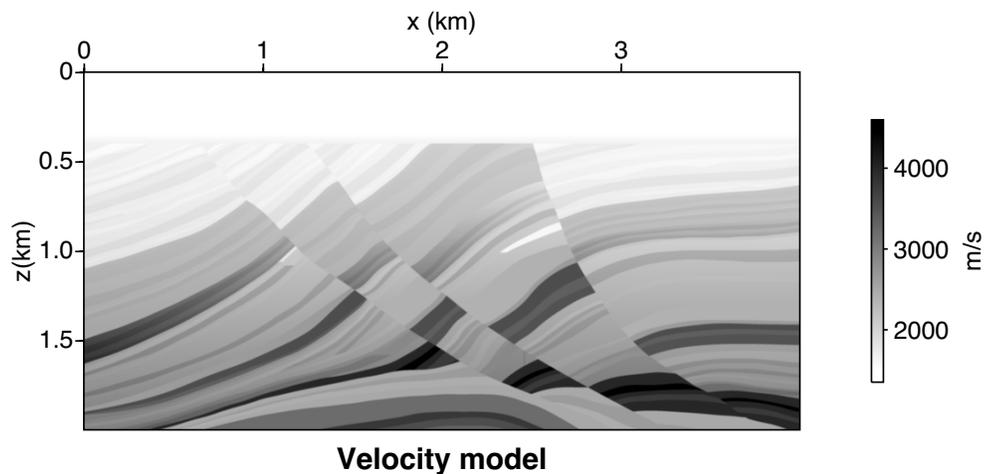


Figure 5.13: The Acoustic Marmousi2 model.

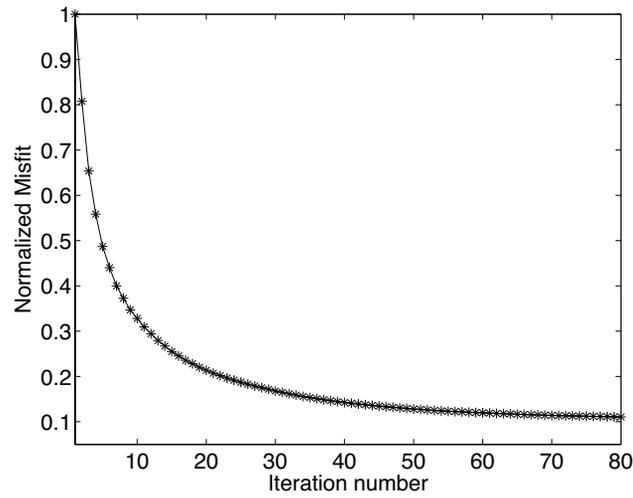


Figure 5.14: Convergence of the gradient projection algorithm for the Marmousi2 example. The algorithm converges within 80 iterations.

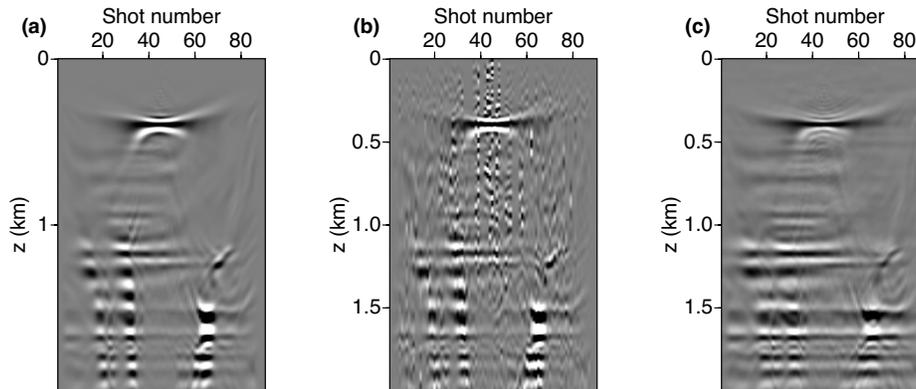


Figure 5.15: Migration results for Marmousi2 model: a shot-index common image gather ($x = 2\text{ km}$) (a) A shot-index common receiver gather from conventional unblended shot-profile migration; (b) A shot-index common receiver gather from blended shot-profile migration; (c) A shot-index common receiver gather from blended least-squares shot-profile migration.

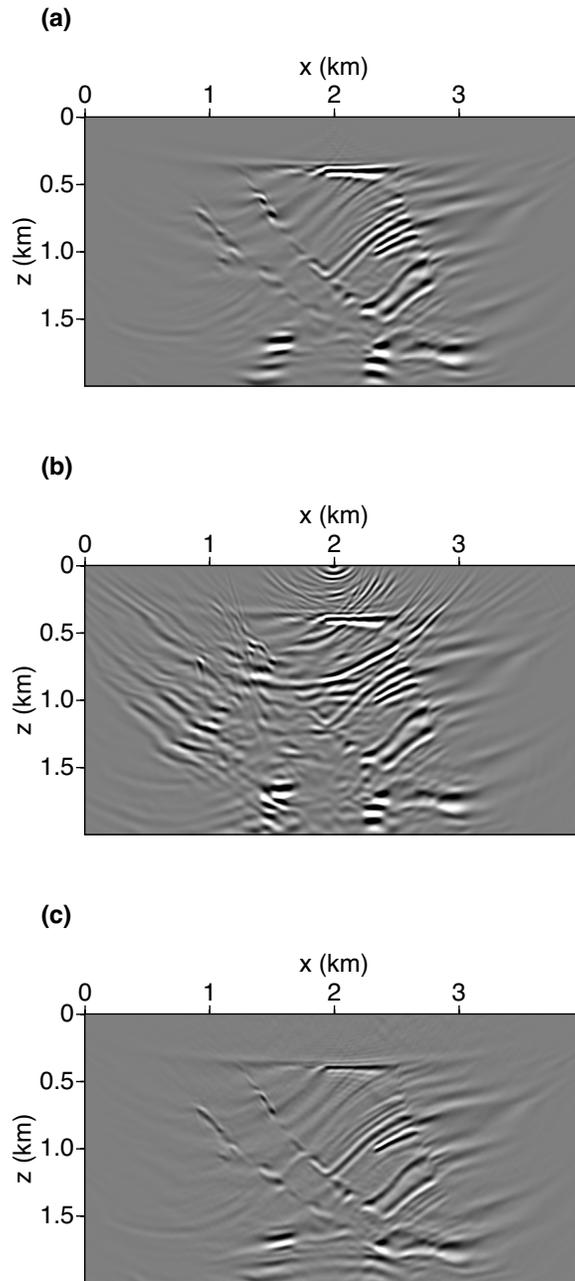


Figure 5.16: Migration results for Marmousi2 model: a partial image for a shot ($S_x = 2.5\text{km}$). (a) A partial image from conventional unblended shot-profile migration; (b) A partial image from the blended shot-profile migration; (c) A partial image from the blended least-squares shot-profile migration.

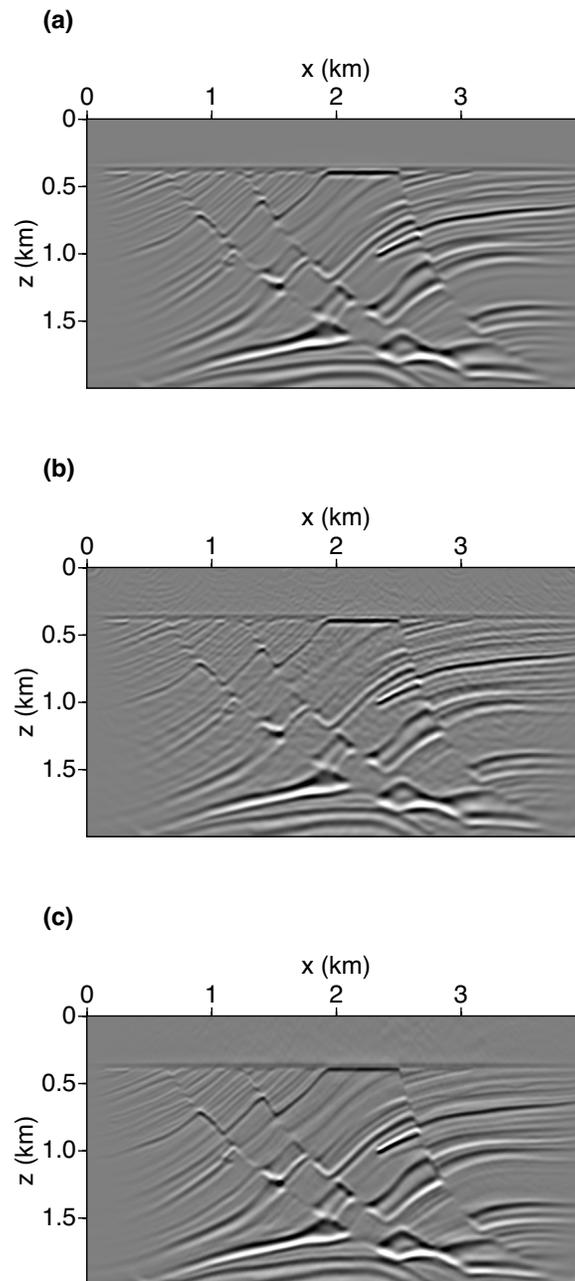


Figure 5.17: Migration results for Marmousi2 model: stacked image (a) Stacked image from conventional unblended shot-profile migration; (b) Stacked image from the blended shot-profile migration; (c) Stacked image from the blended least-squares shot-profile migration.

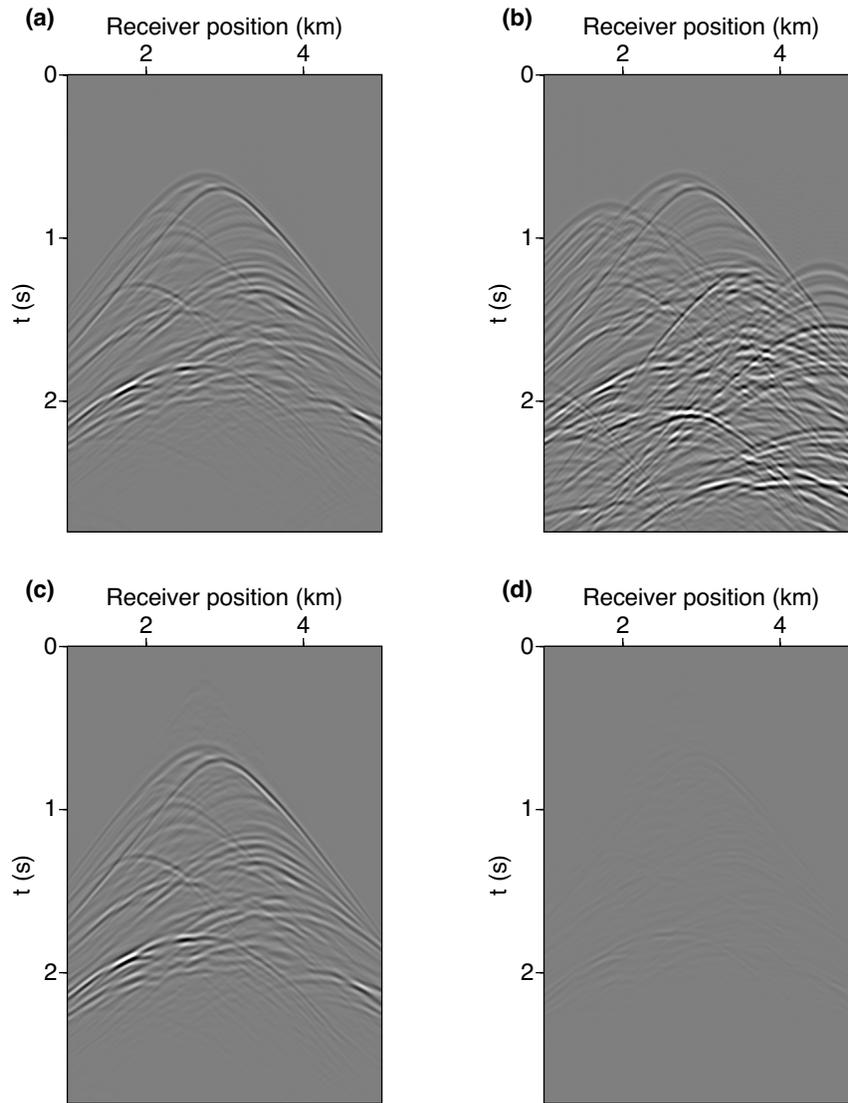


Figure 5.18: Deblending results for the Marmousi2 model. (a) Ideal unblended shot gather (centre shot); (b) Pseudo-deblended shot gather; (c) Deblended shot gather after 100 iterations of least-squares migration. (d) Difference between (a) and (c).

CHAPTER 6

Conclusions

6.1 Summary

Simultaneous source acquisition can significantly improve the efficiency and the quality of seismic data acquisition. In the configuration of simultaneous seismic acquisition, two or more seismic sources are excited in an overlapping fashion. However, severe source interferences are generated by the closely fired shots. Research into simultaneous source seismic data processing falls into two different categories. The first category of methods introduces an extra step called deblending to separate the responses from the blended records. The conventional seismic data processing methods can then be applied to the deblended data. In the second category of methods, the blended data are directly migrated into an image of the earth's subsurface. This thesis aims to introduce the projected gradient method for the separation and the least-squares migration of simultaneous source data. The gradient projection method is an optimization technique for solving constrained inverse problems. The method adopts the gradient descent iteration to minimize an objective function, and in each step, the constraint is implemented via a projection operator. Chapter 2 studies the performance of the projected gradient method and provides the condition that guarantees the convergence of the algorithm.

In Chapter 3, separation of simultaneous source seismic data is formulated as a coherence constrained inverse problem. We assume that the fire time delays between the simultaneously fired shots are random. In common receiver, common offset and common midpoint domain of pseudo-deblended data, the desired signal is coherent, and the source interferences are perturbed by the randomized fire time. The objective function is given by the difference between the blended observation and the predicted blended data. The coherence constraint is that each frequency of the desired data the $f-x-y$ domain can be represented

via a low-rank matrix. Therefore, I incorporated $f - x - y$ eigenimage filtering into the gradient projection framework for the separation of simultaneous source data. To compute the low-rank approximation efficiently, I used the randomized QR decomposition to replace the truncated Singular Value Decomposition. We tested this method using numerically blended synthetic and real data examples. Our analysis showed that the method is effective in suppressing the simultaneous source interferences while preserving the desired signal. We also studied the impact of the parameter rank in the algorithm and the survey time ratio in simultaneous source acquisition design on the separation results.

In Chapter 4, I introduced another reduced-rank projection operator based on Singular Spectrum Analysis (SSA). Instead of performing matrix rank reduction directly to the spatial data in the frequency-space domain, in SSA, rank reduction is applied on the Hankel matrix constructed from spatial data. Since the SSA method is only valid when linear events are present in the seismic data, we apply the method in small patches from the pseudo-deblended common receiver gathers. Compared to the $f - x - y$ eigenimage filter, the Singular Spectrum Analysis filter is more efficient in annihilating the source interferences while preserving the desired unblended data. We tested the method on a numerically blended Gulf of Mexico data set. In this example, we assume around 30% of shots are missing, and the algorithm can reconstruct the missing data while separating the simultaneous sources. In this chapter, we generated the random source fire time delays based on different statistical models. Compared to exponential and normal distribution, we obtained better deblending results when the source fire time delays are generated by a uniform distribution.

In chapter 5, we studied the simultaneous source separation using the least-squares migration formulation. In the blended least-squares migration, the forward and adjoint operators are the shot-profile de-migration and migration operators of the blended data. To compute the contribution in the image from each unblended shot, we used the blended receiver-side wavefield to correlate with the unblended, fire time corrected source-side wavefield. In the shot-index common image domain, the desired signal is coherent and the source crosstalk artifacts appear incoherent. Therefore we introduced a coherence constraint to the least-squares migration and solved the problem via the gradient projection method. The projection operator is the Singular Spectrum Analysis filter that is applied in the shot-index common image domain to eliminate the crosstalk artifacts.

Appendix B analyzes the computation complexity of the Singular Spectrum Analysis method and includes a fast and memory efficient implementation of the SSA method. In the computational efficient SSA method, the low-rank approximation is computed via the randomized QR decomposition. Hankel matrix-vector product is computed via Fast Fourier transform by constructing circulant matrices. The anti-diagonal averaging of reduced-rank Hankel matrix is computed efficiently via convolution. As a result, the construction of the Hankel

matrices is avoided in the SSA filter. The latter is of great importance when applying the method to large-scale and multi-dimensional seismic data.

In Appendix C, separation of simultaneous source seismic data is expressed as an incoherent noise removal problem. A robust principal component analysis method is adopted to suppress the incoherent source crosstalk artifacts in pseudo-deblended data. The method can also be applied to suppress erratic noise in seismic records.

6.2 Contributions

The contributions of this dissertation are as follows. Firstly, the gradient projection method is incorporated into the constrained inversion based simultaneous source separation problem. The objective function is the misfit between the blended observation and the predicted blended data. The condition for source separation is that the desired signal is coherent in common receiver, common offset and common midpoint data domains. The projected gradient method is well suited for the deblending problem. In each iteration, we use the gradient descent iteration to minimize the cost function. A variety of filters can act as a projection operator for constraining the solution. Secondly, this thesis exploits the low-rank property of the desired unblended seismic data, which is different from the transformation based deblending methods. Several matrix rank reduction based methods, including $f - x - y$ eigenimage filtering method, Singular Spectrum Analysis filtering method, and Robust Principal Component Analysis method, are incorporated to suppress the crosstalk artifacts generated by simultaneous source acquisition. Thirdly, in Chapter 5 of this thesis, we proposed a new scheme for migration of blended seismic data. Using the fire time corrected unblended wavefield to correlate with the blended receiver-side wavefield, we can compute the contribution of each unblended shot in the image. We also find that in the shot-index common image domain, the desired signal is coherent while the source crosstalk artifacts appear random. The latter allows us to formularize deblending as a least-squares migration problem and to apply the projected gradient method for solving the problem. Finally, randomized QR decomposition and fast and memory efficient Singular Spectrum Analysis are proposed. The randomized QR decomposition significantly decreases the computation time of matrix rank reduction. The fast and memory efficient Singular Spectrum Analysis provides substantial improvements in the computational efficiency by avoiding the Hankel structured matrix.

6.3 Future developments

Future investigations of simultaneous source data acquisition are moving towards more sophisticated source encoding methods. For instance, Abma and Ross (2013) introduced popcorn shooting technique that varies the activation time of individual airgun within an array. The goal is to reduce the peak amplitude of the airgun sources and to provide more variability in the blended acquisition. Berkhout (2013) discussed to use a narrow band source system with different central frequencies to enhance the incoherency between the blended shots. The new acquisition techniques would request new algorithms to take the interferences under control. Also, most simultaneous source separation methods are based on the randomization of the source fire time delays. Some other aspects, such as the geometrical distribution of sources and receivers, can be considered to improve the separation results. For example, van Borselen et al. (2012) exploited the similarity between the nearby shots for deblending. Robertsson et al. (2016) proposed a periodical source modulation method that enables the apparition of the wavefield from different sources in the $f - k$ domain. Moreover, the gradient projection method can be adapted to different applications of seismic data processing. The projection operator can be combined with non-linear gradient methods and a variety of filters and operators can be selected to constrain the estimation. Finally, the fast and memory efficient Singular Spectrum Analysis method can be extended to Multi-dimensional Singular Spectrum Analysis for 5D seismic data interpolation. The Hankel matrix vector product can be extended to multi-level block Hankel matrix via $N - D$ Fast Fourier transform. The anti-diagonal averaging of a multi-level block Hankel matrix can also be computed via $N - D$ convolution. The method significantly reduces the memory requirement and the computational time for multi-dimensional SSA based seismic data reconstruction.

Bibliography

- Abma, R., T. Manning, M. Tanis, J. Yu, and M. Foster, 2010, High quality separation of simultaneous sources by sparse inversion: 72nd Annual International Conference and Exhibition, EAGE, Expanded Abstracts.
- Abma, R., and A. Ross, 2013, Popcorn shooting: Sparse inversion and the distribution of airgun array energy over time: 83rd Annual International Meeting, SEG, Expanded Abstracts, 31–35.
- Abma, R., Q. Zhang, A. Arogunmati, and G. Beaudoin, 2012, An overview of BP’s marine independent simultaneous source field trials: 82nd Annual International Meeting, SEG, Expanded Abstracts Annual International Meeting, SEG, Expanded Abstracts, 1–5.
- Akerberg, P., G. Hampson, J. Rickett, H. Martin, and J. Cole, 2008, Simultaneous source separation by sparse Radon transform: 78th Annual International Meeting, SEG, Expanded Abstracts, 2801–2085.
- Al-Yahya, K., 1991, Application of the partial Karhunen-Loève transform to suppress random noise in seismic sections: *Geophysical Prospecting*, **39**, 77–93.
- Anagaw, A. Y., and M. D. Sacchi, 2014, Comparison of multifrequency selection strategies for simultaneous-source full-waveform inversion: *Geophysics*, **79**, R165–R181.
- Andersson, F., J. O. Robertsson, D.-J. van Manen, J. Wittsten, K. Eggenberger, and L. Amundsen, 2017, Flawless diamond separation in simultaneous source acquisition by seismic apparition: *Geophysical Journal International*, **209**, 1735–1739.
- Ayeni, G., A. Almonmin, and D. Nichols, 2011, On the separation of simultaneous-source data by inversion: 81st Annual International Meeting, SEG, Expanded Abstracts, 20–25.
- Bagaini, C., M. Daly, and I. Moore, 2012, The acquisition and processing of dithered slip-sweep vibroseis data: *Geophysical Prospecting*, **60**, 618–639.
- Beasley, C. J., R. E. Chambers, and Z. Jiang, 1998, A new look at simultaneous sources: 68th Annual International Meeting, SEG, Expanded Abstracts, 133–135.
- Beck, A., and M. Teboulle, 2009, A fast iterative shrinkage-thresholding algorithm for linear inverse problems: *SIAM J. Imaging Sciences*, **2**, 183–202.
- Berkhout, A. J., 2008, Changing the mindset in seismic data acquisition: *The Leading Edge*,

- 27**, 924–938.
- , 2013, Decentralized blended acquisition: 83rd Annual International Meeting, SEG, Expanded Abstracts, 7–11.
- Bertsekas, D. P., 1976, On the Goldstein-Levitin-Polyark gradient projection method: Automatic Control, IEEE Transactions on, **21**, 174–184.
- , 1999, Nonlinear programming: Cambridge university press.
- Blu, T., P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, 2008, Sparse sampling of signal innovations: Signal Processing Magazine, IEEE, **25**, 31–40.
- Bóna, A., M. A. Slawinski, and P. Smith, 2009, Ray tracing by simulated annealing: Bending method: Geophysics, **74**, T25–T32.
- Boyd, S., and A. Mutapcic, 2007, Subgradient methods. Course notes.
- Broomhead, D., and G. King, 1986, Extracting qualitative dynamics from experimental data: Physica D, **20**, 217–236.
- Cadzow, J., 1988, Signal enhancement-a composite property mapping algorithm: Acoustics, Speech and Signal Processing, IEEE Transactions on, **36**, 49–62.
- Cai, J., E. J. Candes, and Z. Shen, 2010, A singular value thresholding algorithm for matrix completion: SIAM Journal on Optimization, **20**, 1956–1982.
- Candès, E. J., X. Li, Y. Ma, and J. Wright, 2011, Robust principal component analysis?: Journal of the ACM (JACM), **58**, 11.
- Chaikin, P. M., and T. C. Lubensky, 2000, Principles of condensed matter physics: Cambridge university press.
- Chavent, G., and R. E. Plessix, 1999, An optimal true-amplitude least squares prestack depth-migration operator: Geophysics, **64**, 508–515.
- Chen, K., and M. D. Sacchi, 2014, Robust reduced-rank filtering for erratic seismic noise attenuation: Geophysics, **80**, V1–V11.
- Chen, Y., S. Fomel, and J. Hu, 2014, Iterative deblending of simultaneous-source seismic data using seislet-domain shaping regularization: Geophysics, **79**, V179–V189.
- Cheng, J., and M. D. Sacchi, 2015, Separation and reconstruction of simultaneous source data via iterative rank reduction: Geophysics, **80**, V57–V66.
- , 2016, Fast dual-domain reduced-rank algorithm for 3D deblending via randomized QR decomposition: Geophysics, **81**, V89–V101.
- Chiron, L., M. A. van Agthoven, B. Kieffer, C. Rolando, and M.-A. Delsuc, 2014, Efficient denoising algorithms for large experimental datasets and their applications in fourier transform ion cyclotron resonance mass spectrometry: Proceedings of the National Academy of Sciences, **111**, 1385–1390.
- Claerbout, J. F., 1992, Earth soundings analysis: Processing versus inversion: Blackwell Scientific Publications.
- Dai, W., and J. Schuster, 2012, Multisource least-squares reverse time migration: Geophys-

- ical Prospecting, **60**, 681–695.
- Dai, W., X. Wang, and G. T. Schuster, 2011, Least-squares migration of multisource data with a deblurring filter: *Geophysics*, **76**, R135–R146.
- Dasgupta, S., and A. Gupta, 2003, An elementary proof of a theorem of Johnson and Lindenstrauss: *Random Structures & Algorithms*, **22**, 60–65.
- Donoho, D., and J. Tanner, 2009, Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **367**, 4273–4293.
- Duijndam, A. J. W., A. W. F. Volker, and P. M. Zwartjes, 2000, Reconstruction as efficient alternative for least squares migration: 70th Annual International Meeting, SEG, Expanded Abstracts, 1012–1015.
- Eckart, C., and G. Young, 1936, The approximation of one matrix by another of lower rank: *Psychometrika*, **1**, no. 3, 211–218.
- Eggenberger, K., J. O. Robertsson, F. Andersson, D. van Manen, and L. Amundsen, 2016, Signal apparition applied to towed marine simultaneous sources – a case study on synthesized real data from the Viking Graben: *First Break*, **34**, 81–88.
- Eicke, B., 1992, Iteration methods for convexly constrained ill-posed problems in Hilbert space: *Numerical Functional Analysis and Optimization*, **13**, 413–429.
- Fazel, M., 2002, Matrix rank minimization with applications: Master’s thesis, Ph.D. thesis, Stanford University.
- Frank, M., and P. Wolfe, 1956, An algorithm for quadratic programming: *Naval research logistics quarterly*, **3**, 95–110.
- Freire, S. L. M., and T. J. Ulrych, 1988, Application of singular value decomposition to vertical seismic profiling: *Geophysics*, **53**, 778–785.
- Gao, J., X. Chen, J. Li, and J. Ma, 2010, Irregular seismic data reconstruction based on exponential threshold model of POCS method: *Applied Geophysics*, **7**, 229–238.
- Gao, J., M. D. Sacchi, and X. Chen, 2013, A fast reduced-rank interpolation method for prestack seismic volumes that depend on four spatial dimensions: *Geophysics*, **78**, V21–V30.
- Garotta, R., 1983, Simultaneous recording of several vibroseis seismic lines: 53rd Annual International Meeting, SEG, Expanded Abstracts, 308–310.
- Gazdag, J., 1978, Wave equation migration with the phase-shift method: *Geophysics*, **43**, 1342–1351.
- Godwin, J., and P. Sava, 2013, A comparison of shot-encoding schemes for wave-equation migration: *Geophysical Prospecting*, **61**, 391–408.
- Goldstein, A. A., 1964, Convex programming in Hilbert space: *Bulletin AMS*, **70**, 709–710.
- Golub, G., and C. van Loan, 1996, *Matrix computations*, third edition ed.: London: The

- Johns Hopkins University Press.
- Halko, N., P.-G. Martinsson, and J. A. Tropp, 2011, Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions: *SIAM review*, **53**, 217–288.
- Hampson, G., J. Stefani, and F. Herkenhoff, 2008, Acquisition using simultaneous sources: *The Leading Edge*, **27**, 918–923.
- Herrmann, F. J., and X. Li, 2012, Efficient least-squares migration with sparsity promotion and compressive sensing: *Geophysical Prospecting*, **60**, 696–712.
- Howe, D., M. Foster, T. Allen, B. Taylor, and I. Jack, 2008, Independent simultaneous sweeping – a method to increase the productivity of land seismic crews: 78th Annual International Meeting, SEG, Expanded Abstracts, 2826–2830.
- Hua, Y., 1992, Estimating two-dimensional frequencies by matrix enhancement and matrix pencil: *Signal Processing, IEEE Transactions on*, **40**, 2267–2280.
- Huo, S., Y. Luo, and P. Kelamis, 2009, Simultaneous sources separation via multi-directional vector-median filter: 79th Annual International Meeting, SEG, Expanded Abstracts, 150–173.
- Ibrahim, A., and M. D. Sacchi, 2014, Simultaneous source separation using a robust Radon transform: *Geophysics*, **79**, V1–V11.
- Ikelle, L., 2007, Coding and decoding: Seismic data modeling, acquisition and processing: 77th Annual International Meeting, SEG, Expanded Abstracts, **26**, 66–70.
- Indyk, P., and R. Motwani, 1998, Approximate nearest neighbors: towards removing the curse of dimensionality: *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, ACM, 604–613.
- Johnson, W. B., and J. Lindenstrauss, 1984, Extensions of lipschitz mappings into a hilbert space: *Contemporary mathematics*, **26**, 1.
- Jolliffe, I., 2010, *Principle Component Analysis*, 2nd ed.: Springer Series in Statistics, Springer.
- Kaplan, S. T., M. Naghizadeh, and M. D. Sacchi, 2010a, Data reconstruction with shot-profile least squares migration: *Geophysics*, **75**, WB121–WB136.
- Kaplan, S. T., P. S. Routh, and M. D. Sacchi, 2010b, Derivation of forward and adjoint operators for least-squares shot-profile split-step migration: *Geophysics*, **75**, S225–S235.
- Keys, R. G., and D. J. Foster, 1998, Comparison of seismic inversion methods on a single real data set: 68th Annual International Meeting, SEG, Expanded Abstracts.
- Kim, Y., I. Gruzinov, M. Guo, and S. Sen, 2009, Source separation of simultaneous source obc data: 79th Annual International Meeting, SEG, Expanded Abstracts, 51–55.
- Korobeynikov, A., 2010, Computation- and space-efficient implementation of ssa: *Statistics and Its Interface*, **3**, 357–368.
- Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D.

- Lacasse, 2009, Fast full-wavefield seismic inversion using encoded sources: *Geophysics*, **74**, WCC177–WCC188.
- Kreimer, N., and M. D. Sacchi, 2012, A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation: *Geophysics*, **77**, V133–V122.
- Krey, T. C., 1987, Attenuation of random noise by 2-d and 3-d cdp stacking and kirchhoff migration: *Geophysical Prospecting*, **35**.
- Kuhl, H., and M. D. Sacchi, 2003, Least-squares wave-equation migration for AVP/AVA inversion: *Geophysics*, **68**, 262–273.
- Landweber, L., 1951, An iterative formula for Fredholm integral equations of the first kind: *Amer. J. Math.*, **73**, 615–624.
- Levitin, E. S., and B. T. Polyak, 1966, Constrained minimization methods: *USSR comp Math and Math Phys*, **6**, 1–50.
- Li, C., C. C. Mosher, L. C. Morley, Y. Ji, and J. D. Brewer, 2013, Joint source deblending and reconstruction for seismic data: 83rd Annual International Meeting, SEG, Expanded Abstracts, 82–87.
- Liberty, E., 2007, The random projection method. Lecture notes.
- Liberty, E., F. Woolfe, P.-G. Martinsson, V. Rokhlin, and M. Tygert, 2007, Randomized algorithms for the low-rank approximation of matrices: *Proceedings of the National Academy of Sciences*, **104**, 20167–20172.
- Liu, B., and M. Sacchi, 2004, Minimum weighted norm interpolation of seismic records: *Geophysics*, **69**, 1560–1568.
- Ma, S., D. Goldfarb, and L. Chen, 2011, Fixed point and Bregman method for matrix rank minimization and matrix completion: *Mathematical Programming Series A*, **128**, 321–353.
- Mahdad, A., P. Doulgeris, and G. Blacquiere, 2011, Separation of blended data by iterative estimation and subtraction of blending interference noise: *Geophysics*, **76**, 9–17.
- Mansour, H., H. Wason, T. Lin, and F. J. Herrmann, 2012, Randomized marine acquisition with compressive sensing: *Geophysical Prospecting*, **60**, 648–662.
- Maraschini, M., R. Dyer, K. Stevens, and D. Bird, 2012, Source separation by iterative rank reduction - theory and applications: 72nd Annual International Conference and Exhibition, EAGE, Expanded Abstracts.
- Martin, G. S., K. J. Marfurt, and S. Larsen, 2002, Marmousi2: An updated model for the investigation of avo in structurally complex areas: 72nd Annual International Meeting, SEG, Expanded Abstracts.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, 1953, Equation of state calculations by fast computing machines: *The journal of chemical physics*, **21**, 1087–1092.
- Meunier, J., 2011, Seismic acquisition from yesterday to tomorrow: Society of Exploration

Geophysicists.

- Moldoveanu, N., Y. Ji, and C. J. Beasley, 2012, Multivessel coil shooting acquisition with simultaneous sources: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1–6.
- Moore, I., B. Dragoset, T. Ommundsen, D. Wilson, W. C., and D. Eke, 2008, Simultaneous source separation using dithered sources: 78th Annual International Meeting, SEG, Expanded Abstracts, 2806–2810.
- Naghizadeh, M., and M. D. Sacchi, 2010, On the sampling functions and Fourier reconstruction methods: *Geophysics*, **75**, WB137–WB151.
- Nedic, A., and D. P. Bertsekas, 2001, Incremental subgradient methods for nondifferentiable optimization: *SIAM J. OPTIM*, **12**, 109–138.
- Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: *Geophysics*, **64**, 208–221.
- O’Brien, M. J., 2010, Mounting an offense against poor-quality shear data: 80th Annual International Meeting, SEG, Expanded Abstracts, 177–181.
- O’Leary, D., and J. Simmons, 1981, A bidiagonalization-regularization procedure for large scale discretizations of ill-posed problems: *Siam J. Sci. Stat. Comput*, **2**, 474–489.
- Oropeza, V., and M. Sacchi, 2011, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis: *Geophysics*, **76**, V25–V32.
- Papadimitriou, C. H., P. Raghavan, H. Tamaki, and S. Vempala, 2000, Latent semantic indexing: A probabilistic analysis: *J. Comput. System Sci.*, **61**, 217–235.
- Peng, C., and B. Liu, 2013, Deblending of simultaneous sources using an iterative approach: An experiment with variable-depth streamer data: 83rd Annual International Meeting, SEG, Expanded Abstracts, 4278–4282.
- Piana, M., and M. Bertero, 1997, Projected Landweber method and preconditioning: *Inverse Problems*, **13**, 441–463.
- Plessix, R. E., and W. A. Mulder, 2004, Frequency-domain finite-difference amplitude-preserving migration: *Geophysical Journal International*, **157**, 975–987.
- Recht, B., M. Fazel, and P. A. Parrilo, 2010, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization: *SIAM Review*, **52**, 471–501.
- Robertsson, J. O., L. Amundsen, and Å. S. Pedersen, 2016, Signal apparition for simultaneous source wavefield separation: *Geophysical Journal International*, **206**, 1301–1305.
- Rokhlin, V., A. Szlam, and M. Tygert, 2009, A randomized algorithm for principal component analysis: *SIAM J. Matrix Anal. Appl.*, **31**, 1100–1124.
- Romero, L. A., D. C. Ghiglia, C. C. Ober, and S. A. Morton, 1999, Phase encoding of shot records in prestack migration: *Geophysics*, **65**, 426–436.
- Ryden, N., and C. B. Park, 2006, Fast simulated annealing inversion of surface waves on pavement using phase-velocity spectra: *Geophysics*, **71**, R49–R58.

- Sacchi, M. D., 2009, F-x singular spectrum analysis: CSPG CSEG CWLS Convention, 392–395.
- Sallas, J. J., 1989, Enclosure for suppression of air-coupled noise produced by seismic vibrator: U. S. Patent 4922473.
- Schuster, G. T., X. Wang, Y. Huang, W. Dai, and C. Boonyasirawat, 2011, Theory of multisource crosstalk reduction by phase-encoded statics: *Geophysical Journal International*, **184**, 1289–1303.
- Shen, P., and W. W. Symes, 2008, Automatic velocity analysis via shot profile migration: *Geophysics*, **73**, VE49–VE59.
- Silverman, D., 1979, Method of three dimensional seismic prospecting: U.S. Patent, 4,159,463.
- Stefani, J., G. Hampson, and E. Herkenhoff, 2007, Acquisition using simultaneous sources: 69th Annual International Conference and Exhibition, EAGE, Expanded Abstracts.
- Symes, W. W., 2008, Approximate linearized inversion by optimal scaling of prestack depth migration: *Geophysics*, **73**, R23–R35.
- Szu, H., and R. Hartley, 1987, Fast simulated annealing: *Physics letters A*, **122**, 157–162.
- Tang, Y., 2007, Least-squares migration of incomplete data sets with regularization in the subsurface-offset domain: *SEG-125*, 105–121.
- Tao, M., and X. Yuan, 2011, Recovering low-rank and sparse components of matrices from incomplete and noisy observations: *SIAM Journal on Optimization*, **21**, 57–81.
- Thomson, D., and N. Chapman, 1983, A wide-angle split-step algorithm for the parabolic equation: *The Journal of the Acoustical Society of America*, **74**, 1848–1854.
- Trickett, S., L. Burroughs, and A. Milton, 2012, Robust rank-reduction filtering for erratic noise: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1–5.
- Trickett, S., L. Burroughs, A. Milton, L. Walton, and R. Dack, 2010, Rank-reduction-based trace interpolation: 80th Annual International Meeting, SEG, Expanded Abstracts Annual International Meeting, SEG, Expanded Abstracts, 3829–3833.
- Trickett, S. R., 2003, F-x-y eigenimage noise suppression: *Geophysics*, **68**, 751–759.
- Trickett, S. R., and L. Burroughs, 2009, Prestack rank-reducing noise suppression: 79th Annual International Meeting, SEG, Expanded Abstracts Annual International Meeting, SEG, Expanded Abstracts.
- van Borselen, R., R. Baardman, T. Martin, B. Goswami, and E. Fromyr, 2012, An inversion approach to separating sources in marine simultaneous shooting acquisition - application to a Gulf of Mexico data set: *Geophysical Prospecting*, **60**, 640–647.
- Vautard, R., and M. Ghil, 1989, Singular spectrum analysis in non linear dynamics, with applications to paleoclimatic time series: *Physica D*, **35**, 395–424.
- Velis, D. R., and T. J. Ulrych, 1996, Simulated annealing two-point ray tracing: *Geophysical Research Letters*, **23**, 201–204.

- Wapenaar, K., J. van der Neut, and J. Thorbecke, 2012, Deblending by direct inversion: *Geophysics*, **77**, A9–A12.
- Wason, H., R. Kumar, F. J. Herrmann, and A. Y. Aravkin, 2014, Source separation via svd-free rank minimization in the hierarchical semi-separable representation: 84th Annual International Meeting, SEG, Expanded Abstracts, 120–126.
- Xiang, K., X. Chen, H. Chen, and Y. Chen, 2016, Least-squares reverse time migration for blended data with a local low-rank constraint: 78th Annual International Conference and Exhibition, EAGE, Expanded Abstracts.
- Xue, Z., Y. Chen, S. Fomel, and J. Sun, 2014, Imaging incomplete data and simultaneous-source data using least-squares reverse-time migration with shaping regularization: 84th Annual International Meeting, SEG, Expanded Abstracts, 3991–3996.
- Yang, H. H., and Y. Hua, 1996, On rank of block Hankel matrix for 2-D frequency detection and estimation: *Signal Processing, IEEE Transactions on*, **44**, 1046–1048.
- Ye, J., and S. Ji, 2009, An accelerated gradient method for trace norm minimization: *Proceedings of the 26th Annual International Conference on Machine Learning*.
- Zhou, Z., X. Li, J. Wright, E. Candes, and Y. Ma, 2010, Stable principal component pursuit: *Information Theory Proceedings (ISIT), IEEE International symposium*, 1518–1522.

APPENDIX A

Theory of random projections

Random projection, or random embedding, is one of the major precursors to the randomized QR decomposition method discussed in Chapter 3. The method involves taking any collection of n points in a high dimension space and mapping it into a lower dimensional space. The goal is to reduce the dimension while providing approximate preservation of the properties of the collection. The existence of such projection is guaranteed by the Johnson-Lindenstrass lemma (Johnson and Lindenstrauss, 1984). The latter ensures that the random projection maintains the pairwise distance between points and can be proved via the probabilistic methods (Johnson and Lindenstrauss, 1984; Dasgupta and Gupta, 2003). In this appendix, rather than directly prove the Johnson-Lindenstrauss lemma, we study the properties of a linear mapping using a Gaussian matrix. The main results of our analysis are that the random projection preserves the norm as well as the rank with probabilities.

Let us start with a vector $\mathbf{x} \in \mathbb{R}^d$, where \mathbb{R}^d denotes a real set that has d dimensions. The following random projection is applied

$$\mathbf{y} = \frac{1}{\sqrt{K}} \mathbf{\Omega} \mathbf{x}, \tag{A.1}$$

where $\mathbf{\Omega}$ is a random matrix composed by K random vectors of length d . Each entry of the random vector are sampled independently from a Gaussian distribution $N(0, 1)$. We now show that the random projection will preserve the length of the vector (Indyk and Motwani,

1998). The expectation of \mathbf{y} can be computed via

$$\begin{aligned} E(\|\mathbf{y}\|^2) &= E\left(\sum_{i=1}^K \left(\sum_{j=1}^d \frac{1}{\sqrt{K}} \Omega_{ij} x_j\right)^2\right) \\ &= \sum_{i=1}^K \frac{1}{K} \sum_{j=1}^d E(\Omega_{ij}^2) E(x_j)^2. \end{aligned} \quad (\text{A.2})$$

Since each column of $\mathbf{\Omega}$ is drawn from $N(0, 1)$, it is not difficult to show that

$$\sum_{j=1}^d E(\Omega_{ij}^2) = 1. \quad (\text{A.3})$$

Substituting Equation A.3 to Equation A.2, one can acquire

$$\begin{aligned} E(\|\mathbf{y}\|^2) &= \sum_{i=1}^K \frac{1}{K} E\|\mathbf{x}\|^2 \\ &= E\|\mathbf{x}\|^2. \end{aligned} \quad (\text{A.4})$$

Therefore the random projection preserves the norm with probability. Using the probabilistic method, Indyk and Motwani (1998) shows that when $K > \frac{9 \log n}{\epsilon^2 - \epsilon^3}$, the error are bounded by

$$(1 - \epsilon)\|\mathbf{x}\|^2 \leq \|\mathbf{y}\|^2 \leq (1 + \epsilon)\|\mathbf{x}\|^2. \quad (\text{A.5})$$

With this in mind, we will now show that the random projection will preserve the rank of a matrix following the analysis provided in Liberty et al. (2007) and Halko et al. (2011). Now let us consider a rank- K matrix \mathbf{A} of size $m \times n$. The SVD of \mathbf{A} is given by

$$\mathbf{A}_K = \sum_{i=1}^K \sigma_i \mathbf{u}_i \mathbf{v}_i^T. \quad (\text{A.6})$$

A random projection can be applied as follows

$$\mathbf{B} = \frac{1}{\sqrt{K}} \mathbf{\Omega} \mathbf{A}, \quad (\text{A.7})$$

where $\mathbf{\Omega}$ is the random set of size $K \times m$ and is composed by P random vectors drawn from Gaussian distribution. The SVD of \mathbf{B} can be computed via

$$\mathbf{B}_K = \sum_{i=1}^K v_i \mathbf{a}_i \mathbf{b}_i^T, \quad (\text{A.8})$$

where \mathbf{a}_i and \mathbf{b}_i are unitary vectors that are orthogonal. In the randomized QR decomposition, one can acquire a low-rank approximation of \mathbf{A} via the following expression

$$\hat{\mathbf{A}}_K = \mathbf{A} \sum_{i=1}^K \mathbf{b}_i \mathbf{b}_i^T. \quad (\text{A.9})$$

The accuracy of $\hat{\mathbf{A}}_K$ can be estimated via $\|\mathbf{A} - \hat{\mathbf{A}}_K\|_F^2$. Using the properties of orthonormal basis \mathbf{b}_i , the following computation can be written

$$\begin{aligned} \|\mathbf{A} - \hat{\mathbf{A}}_K\|_F^2 &= \sum_{i=1}^n \|(\mathbf{A} - \hat{\mathbf{A}}_K)\mathbf{b}_i\|^2 \\ &= \sum_{i=1}^n \|(\mathbf{A}\mathbf{b}_i - \hat{\mathbf{A}}_K\mathbf{b}_i)\|^2 \\ &= \sum_{i=1}^n \|(\mathbf{A}\mathbf{b}_i - \mathbf{A} \left(\sum_{j=1}^K \mathbf{b}_j \mathbf{b}_j^T \right) \mathbf{b}_i)\|^2 \\ &= \sum_{i=K+1}^n \|\mathbf{A}\mathbf{b}_i\|^2 \\ &= \|\mathbf{A}\|_F^2 - \sum_{i=1}^K \|\mathbf{A}\mathbf{b}_i\|^2. \end{aligned} \quad (\text{A.10})$$

Equation A.10 can be further manipulated as follows (Liberty, 2007)

$$\begin{aligned} \|\mathbf{A} - \hat{\mathbf{A}}_K\|_F^2 &= \|\mathbf{A}\|_F^2 - \|\mathbf{A}_K\|_F^2 + \|\mathbf{A}_K\|_F^2 - \sum_{i=1}^K \|\mathbf{A}\mathbf{b}_i\|^2 \\ &= \|\mathbf{A} - \mathbf{A}_K\|_F^2 + (\|\mathbf{A}_K\|_F^2 - \sum_{i=1}^K \|\mathbf{A}\mathbf{b}_i\|^2). \end{aligned} \quad (\text{A.11})$$

It is not difficult to show that

$$\|\mathbf{A}_K\|_F^2 = \sum_{i=1}^K \sigma_i^2. \quad (\text{A.12})$$

One can utilize the properties of the orthonormal vectors \mathbf{b}_i and \mathbf{v}_i as follows

$$\begin{aligned} \sum_{i=1}^K v_i^2 &= \sum_{i=1}^K \|\mathbf{B}\mathbf{b}_i\|^2 \\ &= \sum_{i=1}^K \left\| \frac{1}{\sqrt{K}} \mathbf{\Omega} \mathbf{A} \mathbf{b}_i \right\|^2 \\ &\approx \sum_{i=1}^K \|\mathbf{A}\mathbf{b}_i\|^2, \end{aligned} \quad (\text{A.13})$$

at the same time, we can compute

$$\begin{aligned}
\sum_{i=1}^K v_i^2 &= \sum_{i=1}^K \|\mathbf{B}\mathbf{v}_i\|^2 \\
&= \sum_{i=1}^K \left\| \frac{1}{\sqrt{K}} \boldsymbol{\Omega} \sum_{j=1}^n \sigma_j \mathbf{u}_j \mathbf{v}_j^T \mathbf{v}_i \right\|^2 \\
&\approx \sum_{i=1}^K \sigma_i^2
\end{aligned} \tag{A.14}$$

Now combining Equation A.12, Equation A.13 and Equation A.14, we have

$$\|\mathbf{A} - \hat{\mathbf{A}}_K\|_F^2 \approx \|\mathbf{A} - \mathbf{A}_K\|_F^2. \tag{A.15}$$

The error is bounded by Equation A.6. In other words, the random projection will preserve the rank of the matrix with probability.

APPENDIX B

Fast and memory efficient SSA

B.1 introduction

Despite the efficacy of SSA, its computational cost has always been a concern because the rank reduction is usually implemented via Singular Value Decomposition (SVD). Efforts have been made by replacing the SVD by more efficient algorithms, such as randomized Singular Value Decomposition (Oropeza and Sacchi, 2011) and Lanczos Bidiagonalization (Gao et al., 2013). For instance, Gao et al. (2013) utilized a fast Hankel matrix-vector product in a Lanczos Bidiagonalization algorithm to avoid building Hankel trajectory matrices. The method is very efficient in calculating the low-rank approximation of the trajectory matrices. However, large Hankel matrices are formed in the final anti-diagonal averaging stage to recover the filtered data.

We present a fast and memory efficient implementation for SSA that does not require building Hankel matrices. We use randomized QR decomposition for fast matrix rank reduction (Chiron et al., 2014). Then the final anti-diagonal averaging of the Hankel matrix is computed efficiently via a convolution algorithm (Korobeynikov, 2010). We also extend the method to multi-dimensional cases where the low-rank approximation of block Hankel matrices is adopted (Gao et al., 2013). We adopt the method to the iterative SSA filtering deblending algorithm that is described in Chapter 4.

B.2 Method

The fast and memory efficient SSA entails the following steps:

- Firstly, we replace the SVD of a Hankel matrix by the randomized QR decomposition discussed in Chapter 3. Instead of applying SVD to the Hankel matrix, a random projection is first performed

$$\mathbf{M} = \mathbf{H}\mathbf{\Omega}, \quad (\text{B.1})$$

where $\mathbf{\Omega}$ denotes a random set that is composed of P independent vectors. Since P is much smaller than the size of the Hankel matrix, the random projection shrinks the size of Hankel matrix while keeping as much variability as possible. The random projection is followed by an economic-size QR decomposition that is applied to matrix \mathbf{M}

$$[\mathbf{Q}, \mathbf{R}] = \text{qr}(\mathbf{M}). \quad (\text{B.2})$$

We point out that the QR decomposition is a very stable algorithm to provide the orthonormal basis. Since the QR decomposition is applied on a shrink-sized matrix, the randomized QR decomposition shows promising improvements in computing low-rank estimations of a given matrix (Cheng and Sacchi, 2016). Conversely, Lanczos bidiagonalization (Gao et al., 2013) tends to be unstable, and an expensive re-orthogonalization on Lanczos vectors is often required. The low-rank approximation can be computed by projecting \mathbf{H} onto the orthonormal basis \mathbf{Q}

$$\hat{\mathbf{H}} = \mathbf{Q}(\mathbf{Q}^*\mathbf{H}). \quad (\text{B.3})$$

In the RQRD method, the parameter P in the randomized QR decomposition is a relaxation of the desired rank. In other words, randomized QR decomposition is less stringent on the choice of rank (number of dips). A wide range of P would ensure the accuracy of rank reduction (Chiron et al., 2014). The latter is important as we usually do not have prior information about the rank of seismic data.

- Secondly, we show the random projection in randomized QR decomposition can be computed using a fast Hankel matrix-vector product. The idea is to embed the Hankel matrix into a circulant matrix and then use Fast Fourier transform to compute matrix vector multiplications (O’Leary and Simmons, 1981). A circulant matrix \mathbf{C} multiplies a vector \mathbf{x} is computed via Fast Fourier Transform:

$$\mathbf{C}\mathbf{x} = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{c}) \circ \mathcal{F}(\mathbf{x})), \quad (\text{B.4})$$

where \mathbf{c} is the first column of the circulant matrix (Gao et al., 2013) and \circ denotes the Hadamard (element-wise) product. We refer the readers to Gao et al. (2013), where the authors discussed in details about embedding a Toeplitz matrix into a circulant matrix. To illustrate the embedding, we simply assume $\mathbf{d} = [d_1, d_2, d_3]^T$.

The corresponding circulant matrix is given by

$$\mathbf{C} = \begin{bmatrix} d_2 & d_1 & d_3 \\ d_3 & d_2 & d_1 \\ d_1 & d_3 & d_2 \end{bmatrix} = \begin{bmatrix} \mathbf{T} & * \\ * & * \end{bmatrix}, \quad (\text{B.5})$$

where $\mathbf{T} = \begin{bmatrix} d_2 & d_1 \\ d_3 & d_2 \end{bmatrix}$ is a Toeplitz matrix built from the given data \mathbf{d} . In practice, one can re-order the data vector $[d_1, d_2, d_3]^T$ into the first column of circulant matrix $[d_2, d_3, d_1]^T$ and then adopt Equation B.4 to compute Toeplitz matrix vector product. Therefore, as is discussed in Gao et al. (2013), no Hankel or Toeplitz matrix is required. A Hankel matrix can be easily converted to a Toeplitz matrix by reversing the columns. We conveniently adopt the algorithm in Gao et al. (2013) given the following relation

$$\mathbf{y} = \mathbf{H}\mathbf{x} = \mathbf{T}\mathbf{z}, \quad (\text{B.6})$$

where \mathbf{T} denotes a Toeplitz matrix and \mathbf{x} is a vector (Korobeynikov, 2010). \mathbf{z} is acquired by reversing the order of the entries of \mathbf{x} . \mathbf{y} is the resulting vector of the multiplication. Fast Hankel matrix-vector product is summarized in Algorithm 4. Equation B.1 can be treated as P Hankel matrix vector product and thus can be computed efficiently via Algorithm 4 as no Hankel matrices are formed.

- Thirdly, we show that the anti-diagonal averaging of the Hankel matrix can be efficiently computed via convolution. To clearly demonstrate the method, the Singular Value Decomposition is adopted and we assume that the desired rank of the Hankel trajectory matrix equals 1. We can rewrite Equation 4.7 as follows

$$\hat{\mathbf{H}} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1. \quad (\text{B.7})$$

Since we assume $\hat{\mathbf{H}}$ is a rank-1 matrix, σ_1 is a constant and denotes the largest singular value and \mathbf{u}_1 and \mathbf{v}_1 denotes the first row and the first column of matrix \mathbf{U} and \mathbf{V} , respectively. $\sigma_1 \mathbf{u}_1 \mathbf{v}_1$ is also named an eigenimage of the matrix $\hat{\mathbf{H}}$ (Trickett, 2003). Combining Equation 4.8 and Equation B.7 yields the following expression

$$\hat{d}_j = \begin{cases} \frac{\sigma_1}{i} \sum_{i=1}^j u_{1i} v_{1j-i+1}, & 1 \leq j \leq M \\ \frac{\sigma_1}{M} \sum_{i=1}^M u_{1i} v_{1j-i+1}, & M \leq j \leq (N - M + 1) \\ \frac{\sigma_1}{N-j+1} \sum_{i=j-N+M}^M u_{1i} v_{1j-i+1}, & (N - M + 1) \leq j \leq N \end{cases}, \quad (\text{B.8})$$

or equivalently,

$$\hat{d}_j = \sigma_1 w_j \sum_{i=1}^N u_{1_i} v_{1_{j-i+1}}, \quad (\text{B.9})$$

where w_j denotes constants that are determined by the size of Hankel matrix and are computed in advance. We can rewrite Equation B.9 in its vector form

$$\hat{\mathbf{d}} = \mathbf{w} \circ [s_1(\mathbf{u}_1 * \mathbf{v}_1)], \quad (\text{B.10})$$

where $\mathbf{u}_1 * \mathbf{v}_1$ denotes the convolution that can be efficiently computed using the Fast Fourier Transform. We can repeat the process to compute each eigenimage corresponding to each desired singular value of the rank- P approximation. The filtered data equals the summation of the P eigenimages

$$\hat{\mathbf{d}} = \mathbf{w} \circ [s_1(\mathbf{u}_1 * \mathbf{v}_1) + s_2(\mathbf{u}_2 * \mathbf{v}_2) + \cdots + s_P(\mathbf{u}_P * \mathbf{v}_P)]. \quad (\text{B.11})$$

The computational complexity reduces to $\mathcal{O}(N \log(P))$.

The strategy can also be adopted for the randomized QR decomposition. Again if we assume $\hat{\mathbf{H}}$ is a rank-1 matrix, Equation B.3 reduces to

$$\hat{\mathbf{H}} = \mathbf{q}_1 \mathbf{t}_1, \quad (\text{B.12})$$

where $\mathbf{t}_1 = \mathbf{q}_1^* \hat{\mathbf{H}}$ and \mathbf{q}_1 denotes the first row of the matrix \mathbf{Q} from the QR decomposition. Apparently \mathbf{t}_1 can be computed via the fast Hankel matrix-vector product. In other words, the computational of the explicit Hankel matrix can be avoided in the anti-diagonal averaging, and thus, in the full SSA algorithm. Equation B.12 then becomes comparable with Equation B.7 with the filtered data given by

$$\hat{\mathbf{d}} = \mathbf{w} \circ (\mathbf{q}_1 * \mathbf{t}_1). \quad (\text{B.13})$$

We can also extend the solution to find the rank- P approximation

$$\hat{\mathbf{d}} = \mathbf{w} \circ [(\mathbf{q}_1 * \mathbf{t}_1) + (\mathbf{q}_2 * \mathbf{t}_2) + \cdots + (\mathbf{q}_P * \mathbf{t}_P)]. \quad (\text{B.14})$$

The resulting fast and memory efficient SSA is summarized in Algorithm 5. Similar to the algorithm by Gao et al. (2013), the proposed fast and memory efficient SSA algorithm can be extended to Multi-dimensional situations. This permits us to use SSA for prestack 5D seismic data reconstruction. In this case, multilevel block Hankel matrices are required but can be multiplied on the flight via multidimensional Fourier transforms. The multilevel anti-diagonal averaging can also be computed via multi-dimensional convolution.

Algorithm 4 Fast Hankel matrix-vector product

Inputs:Seismic data at a given frequency: \mathbf{D} ; Vector \mathbf{x} ; Size of Hankel matrix: L and M **Outputs:**Vector after multiplication: \mathbf{y} $\mathbf{c} = [D_M, D_{M+1}, \dots, D_N, D_1, D_2, \dots, D_{M-1}]^T$ (first row of circulant matrix) $\mathbf{z} = [x_M, x_{M-1}, \dots, x_1, 0, 0, \dots, 0]^T$ (reverse order and padding zeros to length N) $\hat{\mathbf{y}} = \text{ifft}(\text{fft}(\mathbf{c}) \circ \text{fft}(\mathbf{z}))$ $\mathbf{y} = \hat{\mathbf{y}}(1 : L)$

Algorithm 5 Fast and memory efficient SSA

Inputs:Seismic data: \mathbf{D} ; rank: P **Outputs:**SSA filtered data: $\hat{\mathbf{D}}$ **for** $\omega = \omega_{min} : \omega_{max}$ **do** $\mathbf{d} = \mathbf{D}(\omega, :)$ $\mathbf{\Omega} = \text{rand}(M, P)$ (generate random vectors) $\hat{\mathbf{d}} = 0$ **for** $i = 1 : P$ **do** $\mathbf{M}(:, i) = \text{fast_multiply}(\mathbf{d}, \mathbf{\Omega}(:, i))$ (algorithm 4)**end for** $[\mathbf{Q}, \mathbf{R}] = \text{qr}[\mathbf{M}]$ **for** $i = 1 : P$ **do** $\mathbf{q} = \mathbf{Q}(:, i)$ $\mathbf{z} = \text{fast_multiply}(\mathbf{d}, \mathbf{q})$ $\hat{\mathbf{d}} = \hat{\mathbf{d}} + \text{ifft}(\text{fft}(\mathbf{q}) \circ \text{fft}(\mathbf{z}))$ (convolution)**end for** $\hat{\mathbf{D}}(\omega, :) = \hat{\mathbf{d}}$ **end for**

B.3 Performance of fast and memory efficient SSA

To test the efficiency of the proposed Fast SSA algorithm, we compared the performance of different implementations of SSA for random noise attenuation of 3D seismic data patches of different size. SSA rank reduction filter is applied on 50 frequency slices are performed in each run. Table B.1 shows the signal-to-noise ratio of the data after de-noising and the computational time for the conventional multichannel SSA (MSSA) (Oropeza and Sacchi, 2011), Randomized MSSA (Oropeza and Sacchi, 2011), SSA via Lanczos bi-diagonalization (Gao et al., 2013) and the proposed fast and memory efficient SSA, respectively. The four algorithms exhibit similar capabilities in the attenuation of random noise. However, the proposed fast and memory efficient SSA outperforms the other three methods in computational efficiency.

$(N_x \times N_y)$	MSSA		RMSSA		Lanzos		FSSA	
	Time(s)	S/N(dB)	Time(s)	S/N(dB)	Time(s)	S/N(dB)	Time(s)	S/N(dB)
20×20	2.033	5.021	0.667	5.052	0.411	5.001	0.237	5.302
40×40	12.76	9.667	7.669	10.01	1.124	9.801	0.678	10.22
60×60	58.85	12.01	41.12	11.77	2.485	12.56	1.325	11.96
80×80	152.7	14.05	142.5	14.28	6.933	14.77	2.998	14.68

Table B.1: Comparison of computational accuracy and efficiency for different SSA methods: conventional 3D MSSA (MSSA), randomized MSSA (RMSSA), Lanczos Bidagonalization (Lanzcos), and the proposed fast and memory efficient SSA (FSSA). In this experiment, the subset size for FSSA is three times of the desired rank.

B.4 Conclusions

We propose a fast and memory efficient implementation of singular spectrum analysis. The method is based on the randomized QR decomposition to perform matrix rank reduction to a shrink-sized matrix. With the help of fast Hankel matrix-vector product and fast convolution for anti-diagonal averaging, Hankel structured trajectory matrix is avoided. The proposed method significantly improves the computational efficiency of SSA and can be extended to multi-dimensional scenarios.

APPENDIX C

Deblending via Robust Principal Component Analysis

C.1 Introduction

We provide a robust rank reduction method for the separation of simultaneous source seismic data. Unlike the methods discussed in Chapter 3 and Chapter 4, where we treat deblending as an inverse problem, here, we will treat deblending as an incoherent noise removal problem. Due to the strong amplitude of the simultaneous source interferences, we consider the blending noise as a special type of erratic noise in the pseudo-deblended data sets. The proposed method adopts the framework of the $f - x - y$ eigenimage filtering method which applies matrix rank reduction in the frequency space domain. We replace the Singular Value Decomposition in the $f - x - y$ eigenimage filter by the Robust Principal Component Analysis to suppress the source crosstalk artifacts. In this appendix, we first provide a brief review of the Principal Component Analysis (PCA) before introducing the robust PCA (RPCA). We then illustrate the effectiveness of this method via synthetic and real data examples. However, we are aware that better separation results can be achieved by inversion method discussed in Chapter 3 and Chapter 4.

C.2 Theory

C.2.1 Principal Component Analysis

Principal component analysis is an important tool for multivariate analysis in statistics. The idea is to reduce the dimensionality of a data set while preserving as much variability

of data variables as possible (Jolliffe, 2010). Let us consider to recover a low-rank matrix \mathbf{L} from the observed data

$$\mathbf{D} = \mathbf{L} + \mathbf{N}, \quad (\text{C.1})$$

where \mathbf{N} is a matrix representing the additive noise. If we assume \mathbf{N} is composed by small random perturbations, an optimal estimate of \mathbf{L} can be acquired via the following optimization problem

$$\begin{aligned} \min \quad & \|\mathbf{D} - \mathbf{L}\|_2^2 \\ \text{s.t.} \quad & \text{rank}(\mathbf{L}) = K. \end{aligned} \quad (\text{C.2})$$

The problem can be efficiently solved via Singular Value Decomposition (SVD) (Golub and van Loan, 1996). The observed data \mathbf{D} can be decomposed into a group of eigenimages via the SVD. The low-rank component \mathbf{L} can be described with a few eigenimages that are associated to the largest singular values. The noise \mathbf{N} , however, will have energy spread over all the eigenimages (Trickett, 2003). Therefore, methods based on matrix rank reduction are very effective techniques for attenuating random Gaussian noise.

C.2.2 Robust Principal Component Analysis

One problem associated with the PCA-based techniques is that PCA lacks robustness to the erratic noise. In seismic data processing, erratic noise includes swell noise, power line noise and artifacts caused by glitches in recording instruments. Outliers tend to manifest as high-amplitude isolated signals that do not obey the Gaussian distribution. The conventional least-squares error criterion utilized by PCA will perform poorly (Golub and van Loan, 1996; Trickett et al., 2012; Chen and Sacchi, 2014). Therefore, we replace the least-squares constraint by an ℓ_1 norm constraint where we have assumed that the erratic noise is a sparse signal. Finding a low rank approximation matrix subject to an l_1 misfit constraint is a non-convex optimization problem (Fazel, 2002). A practical algorithm can be developed by replacing the non-convex optimization problem by a convex one where a gradient-based optimization method is used. The resulting algorithm is named robust principal component analysis (RPCA) (Candès et al., 2011).

We assume that the erratic noise can be represented using a sparse matrix \mathbf{S} . Only a few entries of \mathbf{S} are non-zero elements and can be arbitrarily large in amplitude (Zhou et al., 2010). Robust principal component analysis suggests the following optimization problem:

$$\begin{aligned} \min \quad & \|\mathbf{S}\|_0 \\ \text{s.t.} \quad & \text{rank}(\mathbf{L}) = K, \quad \mathbf{D} = \mathbf{L} + \mathbf{S}, \end{aligned} \quad (\text{C.3})$$

where $\|\mathbf{S}\|_0$ denotes the ℓ_0 norm of \mathbf{S} , which means the number of non-zero elements in \mathbf{S} . Equation C.3 is a NP-hard problem. To make the problem tractable, we use the ℓ_1 norm, which is defined by the summation of absolute values of the elements of the matrix, to replace the ℓ_0 norm. In the meantime, we consider to replace the low-rank constraint by the nuclear norm of \mathbf{L} which is defined as the sum of all singular values of the matrix \mathbf{L} . One can show that the ℓ_1 norm is the tightest convex relaxation of the ℓ_0 norm. Similarly, the nuclear norm is the tightest convex relaxation to the low rank constraint (Fazel, 2002). We also introduce a *Frobenius* norm constraint, $\|\mathbf{D} - \mathbf{L} - \mathbf{S}\|_F^2$, to tolerate the inclusion of Gaussian noise. The resulting cost function can be written as follows

$$\min J = \frac{1}{2\mu} \|\mathbf{D} - \mathbf{L} - \mathbf{S}\|_F^2 + \beta \|\mathbf{S}\|_1 + \|\mathbf{L}\|_*, \quad (\text{C.4})$$

where β is a trade-off parameter that balances the sparsity and low rank constraints. The scalar μ is a small constant that controls the inclusion of Gaussian noise.

We consider to minimize the Equation C.4 via an iterative scheme to estimate the low-rank data \mathbf{L} as well as the sparse erratic noise \mathbf{S} . We split the cost function into two sub-problems based on the sub-gradient method. The solution of Equation C.4 is equivalent to the solution of the following system of equations

$$\min J_S = \beta \|\mathbf{S}\|_1 + \|\mathbf{S} - \hat{\mathbf{S}}^\nu\|_F^2 \quad (\text{C.5a})$$

$$\min J_L = \|\mathbf{L}\|_* + \|\mathbf{L} - \hat{\mathbf{L}}^\nu\|_F^2 \quad (\text{C.5b})$$

only if $\hat{\mathbf{L}}^\nu$ and $\hat{\mathbf{S}}^\nu$ converge to the solution of

$$\min J_0 = \frac{1}{2\mu} \|\mathbf{D} - \mathbf{L} - \mathbf{S}\|_F^2. \quad (\text{C.6})$$

Therefore, we can calculate $\hat{\mathbf{L}}^\nu$ and $\hat{\mathbf{S}}^\nu$ by updating a current estimation in the opposite direction of the gradient of J_0

$$\begin{aligned} \hat{\mathbf{S}}^\nu &= \mathbf{S}^\nu - \frac{1}{2\mu} (\mathbf{L}^\nu + \mathbf{S}^\nu - \mathbf{D}) \\ \hat{\mathbf{L}}^\nu &= \mathbf{L}^\nu - \frac{1}{2\mu} (\mathbf{L}^\nu + \mathbf{S}^\nu - \mathbf{D}). \end{aligned} \quad (\text{C.7})$$

Equation C.5a is commonly seen in the field of compressive sensing. It leads to a soft-thresholding step to all the entries of the updated solution $\hat{\mathbf{S}}^\nu$ (Beck and Teboulle, 2009). The solution to Equation C.5b can also be found in recent developments of matrix completion. In this case, instead of applying soft-thresholding directly to the entries, a soft-thresholding step are performed to the singular values of the matrix $\hat{\mathbf{L}}^\nu$ (Cai et al., 2010; Recht et al., 2010).

Algorithm 6 RPCA

Inputs:Observed data \mathbf{D} , trade-off parameter λ and stopping criterion ϵ **Initialize:**

$$\mathbf{L}^0 = \mathbf{0}; \mathbf{S}^0 = \mathbf{0}; \nu = 1$$

repeat

$$\hat{\mathbf{S}}^\nu = \mathbf{S}^\nu - \frac{1}{2\mu}(\mathbf{L}^\nu + \mathbf{S}^\nu - \mathbf{D})$$

$$\hat{\mathbf{L}}^\nu = \mathbf{L}^\nu - \frac{1}{2\mu}(\mathbf{L}^\nu + \mathbf{S}^\nu - \mathbf{D})$$

$$\mathbf{S}^{\nu+1} = \max(|\hat{\mathbf{S}}^\nu(i, j)| - \frac{\beta\mu}{2}, 0)$$

$$[\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}] = \text{svd}[\hat{\mathbf{L}}^\nu]$$

$$\hat{\mathbf{\Sigma}} = \max(|\mathbf{\Sigma}(i, i)| - \frac{\beta\mu}{2}, 0)$$

$$\mathbf{L}^{\nu+1} = \mathbf{U} \hat{\mathbf{\Sigma}} \mathbf{V}^*$$

$$\nu = \nu + 1$$

until $\|\mathbf{D} - \mathbf{L}^\nu - \mathbf{S}^\nu\|_F^2 < \epsilon$ **Outputs:**Low-rank estimation $\mathbf{L}^{\nu+1}$;Estimated erratic noise $\mathbf{S}^{\nu+1}$.

The RPCA method is summarized in Algorithm 6. In each iteration, we modify the current estimate of the low-rank data and erratic noise in the opposite direction to the gradient of the quadratic term. Then, we apply two steps of soft-thresholding to the modified estimators. Zhou et al. (2010) have proved that the selection of $\beta = 1/\max(m, n)$ can guarantee high quality recovery of the matrix \mathbf{L} , where m and n are the size of the data matrix \mathbf{D} . The tuning parameter μ can be chosen according to $\mu = 0.1\sqrt{\max(m, n) + 8v\sqrt{\max(m, n)}}$ (Tao and Yuan, 2011) where v is an estimator of the standard error of the additive noise in the data.

C.2.3 RPCA seismic data noise attenuation

The Robust Principal Component Analysis can be adopted to suppress erratic noise present in the seismic data. The de-noising method, which applies robust matrix rank reduction to each frequency slice of the $f - x - y$ data cube, is shown in Algorithm 7. The method resembles the $f - x - y$ eigenimage filtering method except that we use RPCA instead of SVD to compute the low-rank approximations.

Algorithm 7 RPCA de-noising

Inputs:3D seismic volume \mathcal{D} trade-off parameter β

$$\tilde{\mathcal{D}}(\omega, x, y) \leftarrow \mathcal{D}(t, x, y) \quad (\mathcal{F})$$

for $\omega = \omega_{min} : \omega_{max}$ **do**

$$\mathbf{D}_\omega \leftarrow \tilde{\mathcal{D}}(\omega, :, :)$$

$$[\mathbf{L}, \mathbf{S}] = \text{RPCA}[\mathbf{D}_\omega] \quad (\text{Algorithm 6})$$

$$\hat{\tilde{\mathcal{D}}}(\omega, :, :) \leftarrow \mathbf{L}$$

end for

$$\hat{\mathcal{D}}(t, x, y) \leftarrow \hat{\tilde{\mathcal{D}}}(\omega, x, y) \quad (\mathcal{F}^{-1})$$

C.3 Examples

C.3.1 Erratic noise elimination

A synthetic data set is utilized to test the robustness of the proposed de-noising method. Figure C.1 (a) shows the 3D $t - x - y$ data set, which has 30×30 traces and a total time of 1 second. As is shown in Figure C.1 (b), we added Gaussian noise with signal-to-noise ratio equals to 3. We also added erratic noise to randomly selected isolated traces with an amplitude that is about 3 times of the maximum amplitude of desired signal. The processing frequency band ranges from 1 to 40 Hz. The results of the $f - x - y$ eigenimage filtering and the RPCA de-noising were compared. We evaluate the performance of the algorithm via

$$Q = 10 \log_{10} \frac{\|\mathcal{D}\|_F^2}{\|\hat{\mathcal{D}} - \mathcal{D}\|_F^2},$$

where \mathcal{D} is the noise-free data and $\hat{\mathcal{D}}$ denotes the de-noised data. Figure C.1 (c) shows the result of $f - x - y$ eigenimage filtering. The rank was set to 3. Figure C.1 (d) shows the difference between the filtered data and the true data. The erratic noise was not properly removed and the estimated data shows noticeable artifacts. Figure C.1 (e) is the result of the RPCA de-noising. Both the Gaussian noise and the erratic noise are successfully suppressed. The proposed method effectively suppressed the incoherent noise. Figure C.1 (f) shows the error panel corresponding to RPCA de-noising. We improve the quality of the data from $Q = -6.7$ dB to $Q = 14.9$ dB.

We also adopt the proposed method to de-noise a 2D prestack seismic line acquired from Alaska. Figure C.2 (a) and Figure C.2 (d) represent a prestack data section that contains high-amplitude erratic noise. The data are sorted from the source-receiver domain to the offset-midpoint domain before applying the de-noising algorithm. Figure C.2 (b) and Figure C.2 (e) shows the de-noising result of a common offset gather and a common midpoint gather respectively. We can find that the erratic noise is effectively eliminated by the proposed RPCA method. The method also preserves the original signal.

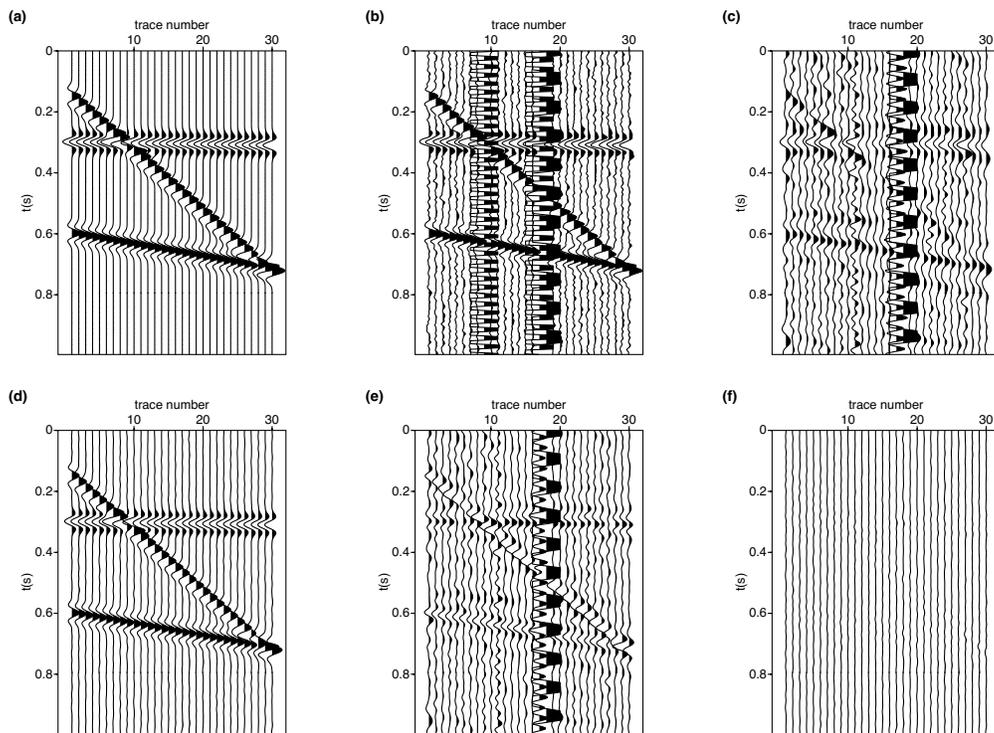


Figure C.1: Results of Robust de-noising for the synthetic data set. (a) The ideal CMP gather of the 3D data cube. (b) The CMP gather contaminated with Gaussian and erratic noise. (c) De-noised gather with $f-x-y$ eigenimage filtering. (d) CMP gather after RPCA De-noising. (e) Differences between (a) and (c). (f) Difference between (a) and (d).

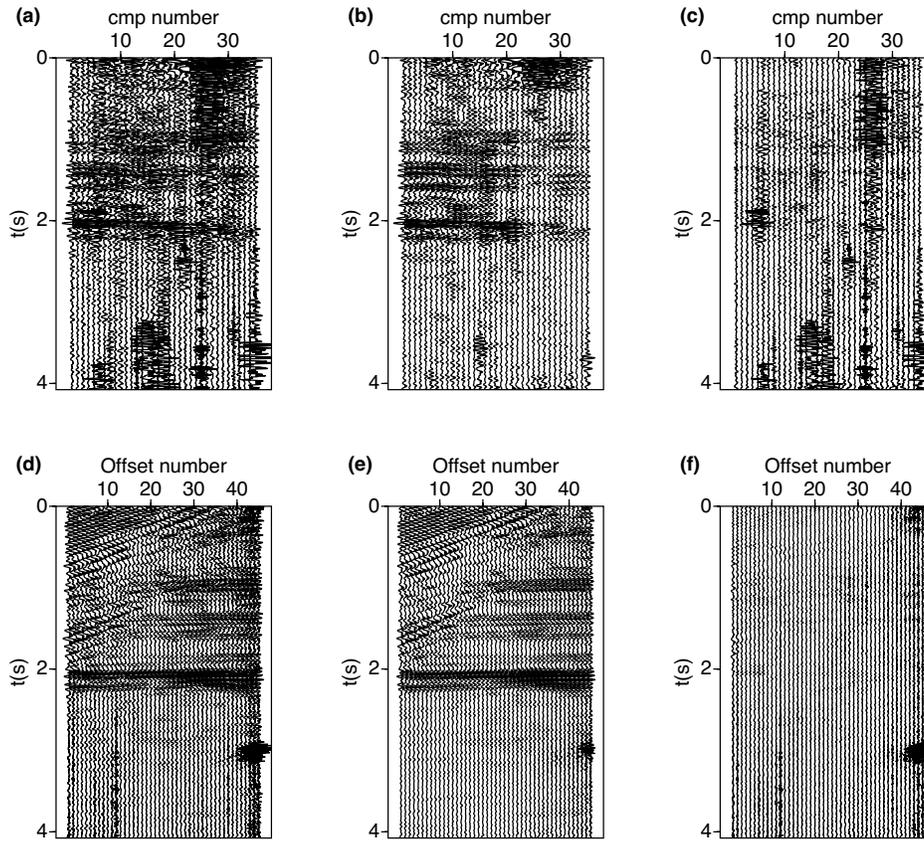


Figure C.2: RPCA de-noising for the Alaska data. (a) The original common offset gather before RPCA denoising. (b) Common offset gather after applying RPCA. (c) Difference between (a) and (b). (d) The original common midpoint gather before RPCA denoising. (e) Common midpoint gather after applying RPCA. (f) Difference between (d) and (e).

C.3.2 Simultaneous source noise suppression

We also applied the proposed algorithm for suppressing the crosstalk brought by simultaneous source acquisition. We adopted a 2D data set acquired from the North Viking Graben. The data set is then numerically blended according to the self-simultaneous source acquisition design. Only one vessel fires with small random time delays generated via a uniform distribution. In this example, 40% of the total acquisition time are saved by the acquisition design. The both sources and receivers are sampled with 20m intervals. The data are sorted into the offset-midpoint domain from source-receiver domain after applying

pseudo-deblending. Figure C.3 (b) shows a pseudo-deblended common offset gather which contains the incoherent blending noise. The amplitude of the crosstalk is as strong as, or even stronger than the desired signal. Figure C.3 (c) shows the de-noising result with the proposed method. The RPCA algorithm effectively removes the interferences and yields a value $Q = 10.8$. Figure C.4 shows the results in common midpoint domain. This example portrays a deblending technique based on robust denoising (Huo et al., 2009; Ibrahim and Sacchi, 2014). However, it is important to mention that deblending methods that are based on inversion could also be adopted.

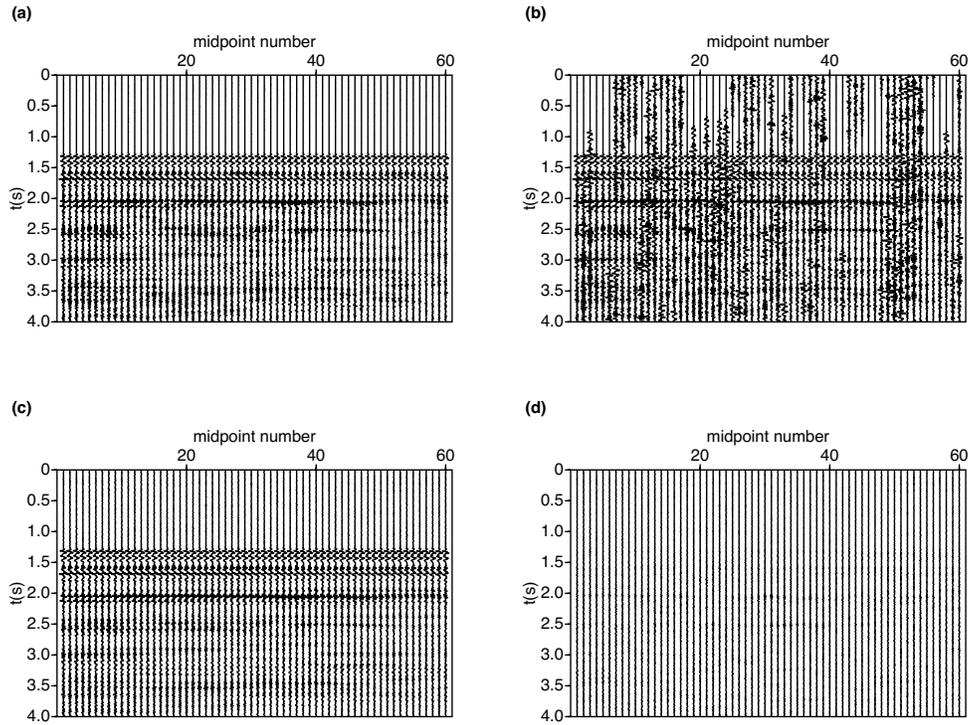


Figure C.3: Deblending via RPCA. (a) The unblended common offset gather. (b) Pseudo-deblended common offset gather. (c) Common offset gather after RPCA de-noising. (d) Differences between (a) and (c).

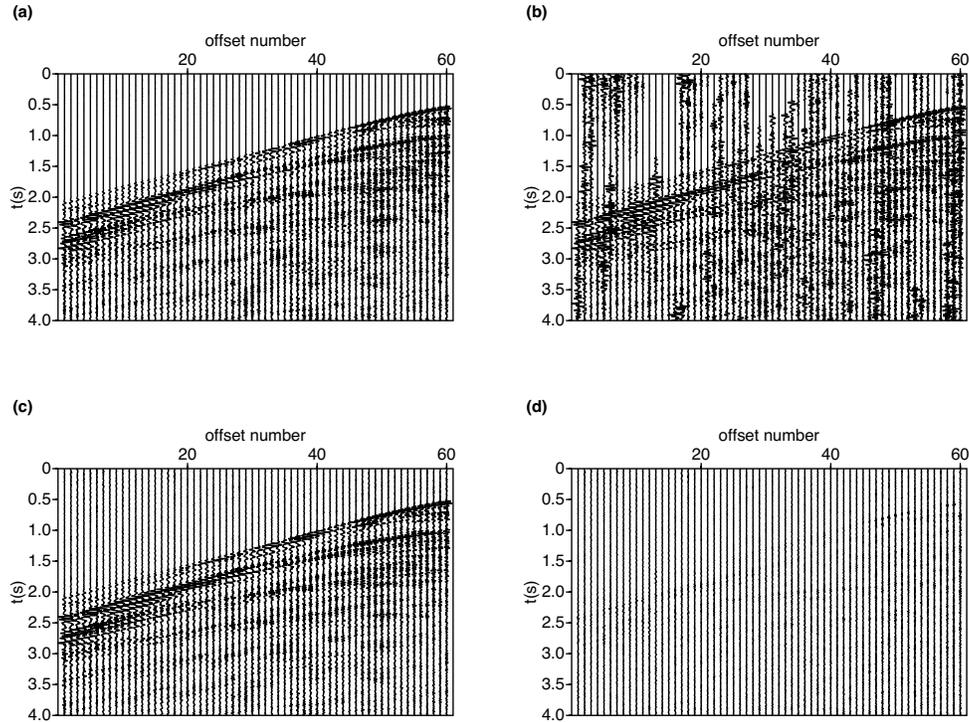


Figure C.4: Deblending via RPCA. (a) The unblended common midpoint gather. (b) Pseudo-deblended common midpoint gather. (c) Common midpoint gather after RPCA de-noising. (d) Differences between (a) and (c).

C.4 Conclusions

In Appendix C, we presented a robust principal component analysis method for suppressing erratic noise that is often present in seismic data. We assume the ideal data can be represented via low-rank matrices in the frequency-space domain and that the erratic noise can be represented via a sparse matrix. A nuclear norm constraint, as well as an l_1 norm constraint, are used to simultaneously recover the data and the erratic noise. We tackled the problem via first order gradient method. Through tests with various examples, the proposed RPCA de-noising method is shown to be able to remove both Gaussian and erratic noise. The method has the potential to be adapted for the de-noising and reconstruction of simultaneous seismic data. Furthermore, higher dimensional version of the algorithm could be developed by interchanging matrices by multilinear arrays (tensors) to represent multi-dimensional spatial data at a given temporal frequency.