Efficient algorithms for least squares wave equation migration and source signature estimation

by

Nasser Kazemi Nojadeh

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

Doctor of Philosophy
in
Geophysics

Department of Physics

University of Alberta

©Nasser Kazemi Nojadeh, 2017

Abstract

Estimating accurate images of the subsurface is one of the end products of seismic data processing. Numerical solutions to the wave equation allow designing linearized forward operators. The adjoint of the linearized forward operator is used to image the interior of the earth. The adjoint operator (migration operator) is sensitive to data sampling and background velocity model. Likewise, migration via adjoint operators produces low-resolution images of the subsurface. Posing seismic imaging as an inverse problem leads to a procedure where the inversion of the linearized forward modelling operator can retrieve an image that honours the seismic observations. Formulating imaging as an inverse problem, besides, allows one to include model space constraints to improve the quality and resolution of subsurface images. This thesis concentrates on the development of efficient and accurate methods for linearized imaging also called least-squares migration. Developing efficient algorithms for least-squares migration is a vital step in better understanding the earth's subsurface structure. However, computational requirements and proper data conditioning are some of the barriers that prevent least-squares migration from becoming a routinely used processing workflow. As part of the development of practical algorithms for least-squares migration, we took advantage of adaptive signal processing strategies and the computational efficiency of preconditioning techniques. We develop scalable algorithms for least-squares migration with less memory and

computation time requirements. Our goal is to achieve similar imaging results without compromising the accuracy of the least-squares migration algorithm. Given that the least-squares migration method is sensitive to the accuracy of the seismic source function, the thesis also examines the pervasive problem of seismic source estimation and provides an algorithm for seismic source estimation that does not require the traditional minimum phase assumption. Imaging tests with synthetic data and a real marine dataset (Mississippi Canyon, north-central Gulf of Mexico south of Louisiana) exemplify the algorithms proposed in this thesis.

Preface

A version of chapter 3 has been published as: Nasser Kazemi and Mauricio D. Sacchi (2015). Block row recursive least-squares migration. Geophysics, 80(5), A95-A101. I was responsible for data simulation and preparation, coding the algorithms, testing the algorithms, and manuscript preparation. M. D. Sacchi was the supervisory author and was involved with concept formation and assisted with manuscript preparation and revision.

Chapter 4 of this thesis is an original work and the results are submitted for publication and currently the paper is under review.

A version of chapter 5 has been published as: Nasser Kazemi and Mauricio D. Sacchi (2014). Sparse multichannel blind deconvolution. Geophysics, 79(5), V143-V152. I was responsible for data simulation and preparation, coding the algorithms, testing the algorithms, and manuscript preparation. M. D. Sacchi was the supervisory author and was involved with concept formation and assisted with manuscript preparation and revision.

Albert Einstein:

The most practical solution is a good theory.

I dedicate this thesis to my parents, my sister and my brother.

Acknowledgements

First of all, I thank my supervisor Dr. Mauricio Sacchi. I am amazed by his ability in breaking the complicated Geophysical problems into small pieces and explaining them in a very simple and understandable way. I am also thankful for his patience and guidance throughout the completion of this thesis work. I thank my teachers Dr. Yu Jeffrey Gu, Dr. Moritz Heimpel, Dr. Douglas Schmitt, and Dr. Christoph Frei for teaching me the different realms of Geophysics and mathematics. I also thank my committee members Dr. Larry Lines, Dr. Sharon Morsink, Dr. Jeff Boisvert, and Dr. Yu Jeffrey Gu for their critical and constructive inputs on my thesis. I really appreciate the time that they took to read the thesis. I also like to acknowledge Sam Kaplan for his previous contribution on least-squares wave equation migration. His thesis was one of the reasons that I got interested in imaging.

Next, I thank all of my fellow friends and group mates at SAIG (Mostafa Naghizadeh, Amsalu Anagaw, Juan I. Sabbione, Nadia Kreimer, Ismael Vera-Rodriguez, JianJun Gao, Jinkun Cheng, Aaron Stanton, Robert Ferner, Fernanda Carozzi, Wenlei Gao, Linan Xu, Ke Chen, Amr Ibrahim, Wubshet Alemie , Sohel Bhuiyan, Ramin Dokht, Mohammad Khalkhali, Elahe Poureslami, Yoones Vaezi, Reza Malehmir, and so many others). I feel lucky that I had a chance to get to know them and learn a lot from their stories and their train of thoughts. I appreciate their encouragements and their good thoughts that helped me realize that with a right attitude everything is doable.

Finally I like to express my deepest gratitudes to my parents. Without their support, patient and constant encouragements this journey would not be possible.

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1 Overview

Geophysical sensing via the seismic exploration method uses seismic wavefields acquired on the surface of the earth to estimate images of the subsurface. These images help geoscientists to conduct structural and stratigraphic interpretation of the subsurface. Seismic images are also required to explore, develop and monitor reservoirs of conventional and unconventional resources. Migration (or Seismic Migration) is the term often used to designate techniques and methodologies for imaging the earth's interior (Claerbout, 1971a). Migration methods are based on the numerical solution of the wave equation. The process of generating synthetic data from an earth model is often called forward modelling. In our case, linearized forward modelling is given by demigration operator. In a nutshell, the input to a demigration operator is a reflectivity model of the subsurface, and the output is the seismic data (prestack data [1]). Conversely, the input to a migration operator is the seismic data, and the output is an image of the subsurface reflectivity.

Migration and demigration operators are not an orthogonal pair. In fact, one is the adjoint [2] of the other. In other words, the application of the demigration operator to an estimated model of the reflectivity does not predict the recorded data. To honour the recorded wavefield and to emphasize good features on the expected image of the reflectivity, one can pose imaging as an inverse problem (Chavent and

[1] Volume of data given as a function of time, source, and receiver coordinates.
[2] Adjoint of a matrix is the complex conjugate of that matrix.

Plessix, 1999; Nemeth et al., 1999; Duijndam et al., 2000; Kuhl and Sacchi, 2003; Plessix and Mulder, 2004; Symes, 2008; Kaplan et al., 2010a,b; Kazemi and Sacchi, 2015). The solution to the aforementioned problem is often called least-squares migration. Least-squares migration permits one to obtain images of the subsurface that honour the data. Also, it allows one to include regularization constraints into the formulation of the imaging problem. Regularization constraints incorporate a priori information about the subsurface or about features that one would like to emphasize in the reflectivity image. To date, running least-squares migration as a routinary process in industrial environments has encountered several problems. For instance, least-squares migration is known for demanding fast computational resources with massive memory (Marfurt and Shin, 1989; Mulder and Plessix, 2004a,c). Numerical drawbacks might also prevent the adoption of least-squares migration algorithms. The latter includes the slow convergence of iterative solvers in the presence of incorrect velocity and source signature, appropriate estimation of tradeoff parameters, and operator mismatch (Nemeth et al., 1999; Fletcher et al., 2016).

1.2 Seismic data

In exploration seismology, for each source, receivers are used to record wavefields on the surface of the earth. The time series recorded by each receiver is often called a seismic trace or a seismogram. The group of traces generated by a given shot is referred to as a common shot gather (Figure 1.1). In a 2D seismic acquisition, one often works with a couple of hundreds of common shot gathers. These gathers contain information about the recorded wavefields reflected by subsurface interfaces. The final goal of seismic exploration is to image earth's interior and to provide information about subsurface geological structures in an area of study. Migration algorithms are capable of estimating these images. In a nutshell, migration algorithms can be applied before stacking (prestack migration) or after stacking (poststack migration). Poststack data are processed seismic data that imitate a data volume acquired with source-receiver pairs at the same position; sometimes referred to as the zero offset experiment. Poststack migration occupied a special place in seismic data processing and imaging until the early eighties. Today, accurate migration algorithms for imaging complex media prefer prestack data. In this thesis, we focus on prestack migration techniques. In prestack seismic migration we first back propagate the

receiver side wavefield and forward propagate the source side wavefield. Then, by cross-correlating, the propagated wavefields at different depths (Claerbout, 1971a) one forms an image of the subsurface. This process, which originates from pure physical intuition, can also emerge from a precise mathematical formulation that adopts linearized wave propagation theory (Stoffa et al., 1990).

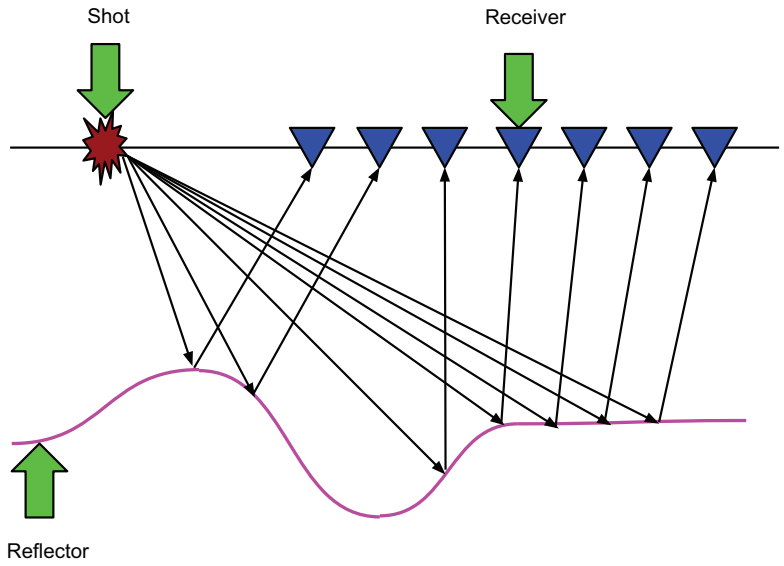

Figure 1.1: Schematic representation of the geometry of source and receivers for a common shot gather (CSG).

1.3 Imaging

Seismic migration aims to produce accurate structural and stratigraphical images of the subsurface. Different migration methods have been proposed and extensively studied by applied geophysicists. Migration methods are divided into two broad categories. The first type is composed of methods based on ray-tracing such as Kirchhoff migration (Schneider, 1978). Kirchhoff migration methods are computa-

tionally efficient and are adaptable to non-regular acquisition geometries. Another category of methods adopts the wave equation to formulate migration as a wavefield extrapolation process. Although wave equation methods are computationally expensive, they provide accurate wavefield extrapolation and high-quality images for complex velocity models. In this approach, one can solve either the one-way or the two-way wave equation (Gazdag, 1978; Stolt, 1978; Gazdag and Sguazzero, 1984; Stoffa et al., 1990; Baysal et al., 1983; McMechan, 1983; Whitmore, 1983). In this thesis, we have adopted one-way wave equation theory to solve the acoustic wave equation and to design propagators for forward modelling (demigration) and migration. We also assume that data measurements contain only the pressure component. Besides, we presume input data that has been properly preprocessed to eliminate free surface and interbed multiple reflections (Guitton, 2005; Griffiths et al., 2011; Weglein et al., 2011).

Imaging can be cast as a linear problem. One can imagine that migrating a seismic dataset is equivalent to multiplying the data (stored in vector form) by a linear operator or matrix. However, it is important to mention that in most cases it is not possible to represent a migration algorithm via a matrix that operates on a vector. Therefore, the action of migration operators on seismic data will be calculated *on the fly* via matrix-free operators. Another factor that we should take into consideration is that migration techniques often introduce artifacts on images. These artifacts arise from the approximations which were made at the time of designing the migration operators and from the incompleteness of the data.

A substantial number of authors have studied the problem of artifacts in seismic migration (Gray et al., 2001; Herron, 2000; Gray, 2013). Artifacts can be minimized by adopting data regularization techniques before imaging (Fomel and Guitton, 2006). They can also be attenuated by least-squares migration algorithms (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,b). In least-squares migration, one tries to find a reflectivity model that fits the seismic observations. In general, we adopt constraints to minimize artifacts that are introduced by data incompleteness, inaccuracies in the velocity model and poor illumination (Kuhl and Sacchi, 2003). Images estimated via least-square migration honour the recorded wavefield and show enhanced vertical and lateral resolution when compared to im-

ages obtained via classical migration methods (Nemeth et al., 1999; Kuhl and Sacchi, 2003; Kazemi and Sacchi, 2015). Last, this thesis investigates least-squares migration with one-way wave equation operators. One-way operators are computationally affordable, and they often provide a reasonable approximation to the propagating wavefield (Mulder and Plessix, 2004b).

In this dissertation, we have designed forward and adjoint operators that are necessary for the least-squares migration. In other words, the migration problem is posed as a linear inverse problem solved via an semi-iterative method: the method of Conjugate Gradient (CG) (Hestenes and Stiefel, 1952). The CG method requires the iterative application of forward and adjoint operators. We have considered forward and adjoint operators that are suitable for realistic prestack migration by allowing for propagation in vertically and laterally varying media. Also, we have constrained the least-squares migration problem by considering smoothness in the L_2 norm of the reflectivity images. It is important to mention that one could have also explored non-quadratic constraints as in Wang and Sacchi (2006). However, non-quadratic constraints lead to non-linear problems that are difficult to solve and, in general, are too unstable for practical industrial applications.

Although least-squares migration is a powerful technique for minimizing migration artifacts, its computational cost requires careful analysis. For today's computer resources, it is challenging to implement least-squares migration methods in their full potential for standardized industrial imaging. The calculation of the inverse of the Hessian [3] dominates the cost of least-squares migration. One could reduce the cost of inverting the Hessian by approximating its inverse. Several interesting articles have tried to address the problem mentioned above (Hu et al., 2001; Etgen, 2002; Guitton, 2004; Yu et al., 2006; Lecomte, 2008; Toxopeus et al., 2008; Naoshi and Schuster, 2009; Kazemi and Sacchi, 2014a). Preconditioning the least-squares migration algorithm, for instance, also results in a substantial reduction of computational cost. Preconditioning is adopted to increase the convergence rate of iterative algorithms such as the Conjugate Gradient method (Rickett, 2003; Symes, 2008; Nammour and Symes, 2009; Herrmann et al., 2009; Demanet et al., 2012; Guitton, 2004; Kazemi and Sacchi, 2014a; Huang et al., 2016).

[3] If the forward or de-migration operator is given by $\mathbf{A}$ and the adjoint or migration operator is given by $\mathbf{A}^T$, the Hessian is defined via $\mathbf{A}^T\mathbf{A}$. Strictly speaking, this definition requires linearity.

1.4 Main contributions

In this thesis we aimed to alleviate the computational cost of least-squares migration algorithm for large data sets. To date, adoption of least-squares migration algorithms by seismic data processing companies has been slow. So far, the developed algorithms could not satisfy the needs of industry. We realized that developing efficient algorithms for least-squares migration is a vital step for improving subsurface imaging. High-resolution images of the subsurface with correct amplitude information can be obtained by the use of least-squares migration. However, the computational cost of least-squares migration makes it unattractive for seismic data processing in industry processing workflow. Even by improving the computational power of computers, it is crucial to develop efficient imaging algorithms as there will always be demand for imaging increasingly dense data acquired over large extensions. We took advantage of adaptive signal processing strategies and preconditioning techniques to develop scalable algorithms for least-squares migration that require less memory and computation time than conventional least-squares migration algorithms.

We have concentrated on two algorithms for least-squares migration as well as one algorithm to solve the ubiquitous problem of seismic source estimation. The first migration algorithm is inspired by work in the field of adaptive signal processing techniques. The method uses adaptive strategies to overcome the memory limitation that could occur when running least-squares migration with a large number of shots. It also links the solutions of blocks of data in a way that the algorithm solves for the original least-squares problem that one would have computed if all the data were loaded into memory at once. The latter is used to address the input/output (I/O) problem that arises when least-squares migration is applied to large data sets. Data sets are often larger than the memory capacity of the computer. In this situation, one may not be able to load and use the whole data at once. Recursive least-squares solutions can offer a solution to the problem above. Groups of shot gathers are read and used to recursively solve the least-squares migration problem via an online strategy borrowed from the field of adaptive signal processing (Ng and Plemmons, 1996). Clearly, the reduction in memory requirement and I/O might not be considered impressive if one limits this research to 2D problems. However, recursive least-squares algorithms might have an impact on 3D imaging. These ideas

could also be adapted to develop (almost) real-time strategies for migrating data while being acquired. Interesting research ideas associated with recursive solutions have immediate application to simultaneous real-time acquisition and imaging of marine data. In other words, acquisition parameters can change as the result of information provided by images that are updated via an (almost) real-time recursive least-squares migration algorithm.

The second proposed algorithm entails incorporating preconditioning to least-squares migration to speed up the convergence of the conjugate gradient method. Preconditioners are matrices (operators) that can cluster the eigenvalues of the matrix that one would like to invert. We designed three types of preconditioners that are capable of clustering the eigenvalues of the Hessian operator. The latter improves the convergence rate of least-squares migration. Consequently, preconditioning reduces the computational cost of least-squares migration.

Data conditioning for imaging requires the application of a series of processing techniques: signal-to-noise ratio enhancement, attenuation of multiple reflections, and velocity model building. An estimator of the seismic source function is needed, as well. Seismic source estimation (also known as wavelet estimation) is a classical problem in seismic data processing. The wavelet is required to deconvolve the seismic traces, and thereby, to enhance the bandwidth of the seismic data. The wavelet is also needed to compute the downward propagating source wavefield in shot profile wave equation migration (Berkhout and Verschuur, 2005; Ribodetti et al., 2011; Fletcher et al., 2016). Having the correct wavelet might not be a critical issue in classical migration methods via adjoint operators. However, in least-squares migration one of the steps involves estimating a model that must fit the data. Fitting the data requires the source wavelet function be estimated a priori and removed from the data. Alternatively, one can incorporate the source function to the forward (demigration) operator. It is evident that the proposed least-squares migration algorithms require source processing; this is the main reason we have also studied the source estimation problem and proposed a new algorithm for blind deconvolution of seismic data (Kazemi and Sacchi, 2014b). Specifically, we have developed a multichannel blind deconvolution algorithm that makes no assumptions about the amplitude and phase spectra of the source signature. In other words, we are avoiding the standard minimum phase wavelet assumption (Robinson and Treitel, 1964). The model assumes that the wavelet is stationary in time and does not consider

attenuation in the formulation.

1.4.1 Organization of thesis

Chapter 1 reviews the field of seismic migration and describes the layout of this thesis.

Chapter 2 investigates the application of the Born approximation and the definition of the forward (demigration) and adjoint (migration) operators. Particular attention is given to one-way operators.

Chapter 3 proposes a block row recursive least-squares algorithm to solve the linear system of equations of pre-stack least-squares depth migration. Recursive estimates of large systems of equations in the context of least-squares fitting is a common problem in different fields of study. For example, recursive adaptive filtering is extensively used in signal processing and control applications (Regalia, 1994; Stearns, 1985). The necessity of recursive solutions to least-squares problems stems from the need for real-time and fast signal processing strategies.

Chapter 4 investigates the design and application of three types of preconditioners. Preconditioning is necessary for improving the convergence of the Conjugate Gradient method. The Conjugate Gradient method is an algorithm for the solution of symmetric and positive definite systems of equations. The Conjugate Gradient method can be regarded as an exact method because it produces the exact solution in a finite number of iterations. In practical terms, however, it behaves as an iterative solver where the solution of the linear system of equations improves in each iteration. The speed at which the iterative solution coverages to the true solution is controlled by the condition number of the system of equations (Shewchuk et al., 1994; Hestenes and Stiefel, 1952). Preconditioning is adopted to replace the original system of equations by one with a smaller condition number. Therefore, the speed of convergence of the algorithm improves. With modern computer resources, least-squares migration is considered a research problem that only in recent years has started to be adopted by industry (Salomons et al., 2014). Least-squares migration could gain acceptance in the industry if one can solve the problem using a small number of iterations. The latter is often not doable, and therefore, preconditioners are introduced as an alternative to approximate the Hessian to a form that can be inverted in a small number of iterations. The thesis introduces three types of

preconditioners. The first two preconditioners belong to the diagonal scaling category, and the third one is a filter-based approach which approximates the Hessian operator by local convolutional filters.

Chapter 5 describes a new source estimation algorithm. The seismic source must be provided as an input parameter to the proposed prestack least-squares depth migration algorithms. In this chapter, we introduce and apply a multichannel blind deconvolution technique to solve for a multichannel impulse response of the subsurface. As a by-product of the algorithm, we can estimate the source signature which is adopted for least-squares migration. By using the estimated source signature, the forward modelled data via the forward (demigration) operator will show similar source signature characteristics to that of the acquired data. The proposed algorithm is called Sparse Multichannel Blind Deconvolution (SMBD)(Kazemi and Sacchi, 2014b). The method is a modification of the multichannel blind deconvolution technique often called Euclid deconvolution where the multichannel impulse response of the earth is estimated by solving a homogeneous system of equations (Xu et al., 1995; Rietsch, 1997a,b). Classical Euclid deconvolution is unstable in the presence of noise and requires the correct estimation of the length of the seismic wavelet. The proposed method, on the other hand, can tolerate moderate levels of noise and does not require a priori knowledge of the length of the wavelet. SMBD solves the homogeneous system of equations arising in Euclid deconvolution by imposing sparsity on the unknown multichannel impulse response.

Finally, Chapter 6 concludes the dissertation and provides advice about potential future research directions.

CHAPTER 2

The Born approximation and forward/adjoint pairs

2.1 Introduction

This chapter investigates the derivation of linearized forward and adjoint operators for shot profile one-way wave equation migration. We will use the Born approximation to derive the forward modelling operator and the adjoint operator. These operators are necessary for developing iterative algorithms for least-squares migration. We will use these operators throughout the thesis.

2.2 The Born approximation: forward/demigration operator and its adjoint

We start with the description of wave motion in an elastic media. The generalized Hooke's law connects stress τ and strain ϵ

$$\tau_{ij} = C_{ijkl}\epsilon_{kl}\,, \tag{2.1}$$

where indices i, j, k, l can take values from one to three. In general, the stiffness matrix $\mathbf{C}$ has 81 coefficients. Considering symmetries and homogeneity, we can decrease the number of the elements of the stiffness matrix to two (Sheriff and Geldart, 1995). For isotropic media, the number of coefficients of the stiffness matrix

reduces to two, which are the well-known *Lame parameters*: λ and μ. Hooke's law for isotropic media reduces to the following expression

$$\tau_{ij} = \lambda\delta_{ij}\epsilon_{kk} + 2\mu\epsilon_{ij}\,. \tag{2.2}$$

To obtain the equation of motion that describes the propagation of waves in elastic media, one needs to replace the stress-strain expression in Newton's second law of motion. The latter leads to the well-known equation of motion for displacements $\mathbf{u}$ for waves propagating in an elastic media

$$\rho\frac{\partial^2\mathbf{u}}{\partial^2 t} = \nabla\lambda(\nabla.\mathbf{u}) + \nabla\mu.[\nabla\mathbf{u} + (\nabla\mathbf{u})^T] + (\lambda+2\mu)\nabla\nabla.\mathbf{u} - \mu\nabla\times\nabla\times\mathbf{u} + \mathbf{f}, \tag{2.3}$$

where $\mathbf{f}$ is the external force, ρ is the density and ∇ and $\nabla\times$ indicate divergence and curl operators, respectively. The first two terms on the right hand side of equation 2.3 contain gradients in *Lame parameters* which stem from the inhomogeneity of subsurface materials. Ignoring these gradients and assuming constant density, leads to a simplified form of equation 2.3

$$\rho\frac{\partial^2\mathbf{u}}{\partial^2 t} = (\lambda+2\mu)\,\nabla\nabla.\mathbf{u} - \mu\,\nabla\times\nabla\times\mathbf{u} + \mathbf{f}, \tag{2.4}$$

or

$$\frac{\partial^2\mathbf{u}}{\partial^2 t} = v_p^2\,\nabla\nabla.\mathbf{u} - v_s^2\,\nabla\times\nabla\times\mathbf{u} + \mathbf{f}, \tag{2.5}$$

where v_p indicates compressional velocity and v_s denotes shear velocity. Moreover, one can separate the S-waves and P-waves by applying divergence and curl operators to equation 2.5 (Dellinger and Etgen, 1990). This leads to equations of motion in terms of potentials. Volumetric waves obey the wave equation for the scalar potential θ

$$\frac{\partial^2\theta}{\partial^2 t} - v_p^2\nabla^2\theta = f_P, \tag{2.6}$$

similarly, shear waves, obey the equation of motion for the vector potential $\mathbf{\Psi}$

$$\frac{\partial^2\mathbf{\Psi}}{\partial^2 t} - v_s^2\nabla^2\mathbf{\Psi} = \mathbf{f}_S, \tag{2.7}$$

where $\theta = \nabla.\mathbf{u}$ and $\mathbf{\Psi} = \nabla\times\mathbf{u}$. In this thesis, we will only focus on equation 2.6. In other words, we will represent the earth by an acoustic and isotropic media with constant density. We consider each layer to be homogenous, however, we allow

variation of velocity in the lateral and vertical directions. This assumption is frequently valid for imaging purposes. However, it is clear that a full elastic treatment is required if our goal is to estimate subsurface material properties (Mora, 1987). Moreover, our assumptions are not valid for geological structures composed of, for instance, anisotropic shales, as shown by Vestrum and Muenzer (1997) and others. We should also point out that the Born approximation assumes that multiples have been suppressed in the seismograms.

Considering a two dimensional constant density acoustic and isotropic medium with explosive source at location $\mathbf{x}_s$, equation 2.6 in the frequency domain reads

$$(\omega^2 s^2 + \nabla^2)\theta = f\ \delta(\mathbf{x} - \mathbf{x}_s), \tag{2.8}$$

where s is slowness (reciprocal of velocity) and ω is temporal frequency. To start the analysis one can assume a background smooth velocity (slowness) field that is known. We can represent the squared slowness and the scalar field in terms of perturbations and backgrounds as follows

$$\begin{aligned} s^2 &= s_0^2 + m, \\ \theta &= \theta_0 + \Delta\theta, \end{aligned} \tag{2.9}$$

where s_0 and θ_0 are background slowness and wavefield, respectively. The parameter m is the perturbation in slowness-squared. Similarly, $\Delta\theta$ is the perturbation in the wavefield due to m. The scattering potential m can also be considered proportional to the seismic reflectivity (Clayton and Stolt, 1981).

By inserting equations 2.9 into equation 2.8 one can now write

$$(\omega^2(s_0^2 + m) + \nabla^2)[\theta_0 + \Delta\theta] = f\ \delta(x - \mathbf{x}_s), \tag{2.10}$$

and by using the fact that

$$(\omega^2 s_0^2 + \nabla^2)\theta_0 = f\ \delta(\mathbf{x} - \mathbf{x}_s), \tag{2.11}$$

equation 2.10 simplifies to

$$(\omega^2 s_0^2 + \nabla^2)\Delta\theta = -\omega^2 m\ \theta, \tag{2.12}$$

where $\Delta\theta = \Delta\theta(\omega, \mathbf{x})$ and $\theta = \theta(\omega, \mathbf{x})$.

Now, by using the Green's function G_0 satisfying the wave equation corresponding to the reference medium

$$(\omega^2 s_0^2 + \nabla^2)\, G_0 = \delta(\mathbf{x} - \mathbf{x}_s), \tag{2.13}$$

the perturbed wavefield can be calculated via

$$\Delta\theta(\omega, \mathbf{x}) = -\int G_0(\mathbf{x}, \omega; \mathbf{x}')\, (\omega^2 m(\mathbf{x}^{'})[\theta_0(\omega, \mathbf{x}^{'}) + \Delta\theta(\omega, \mathbf{x}^{'}])d\mathbf{x}^{'}\,. \tag{2.14}$$

The last equation is non-linear and it can be linearized by reinserting $\Delta\theta$ in the integral and by keeping first order terms

$$\Delta\theta(\omega, \mathbf{x}) \approx -\int G_0(\mathbf{x}, \omega; \mathbf{x}')\, \omega^2 m(\mathbf{x}^{'})\, \theta_0(\omega, \mathbf{x}^{'})\, d\mathbf{x}^{'}\,. \tag{2.15}$$

In general, if the explosive source is at position $\mathbf{x}_s$ and the receivers are at spatial coordinates $\mathbf{x}_r$, equation 2.15 can be written as

$$d(\omega, \mathbf{x}_r, \mathbf{x}_s) = \Delta\theta(\omega, \mathbf{x}_r, \mathbf{x}_s) \approx -\int G_0(\mathbf{x}_r, \mathbf{x}_s, \omega; \mathbf{x}')\, \omega^2 m(\mathbf{x}^{'})\, \theta_0(\omega, \mathbf{x}^{'})\, d\mathbf{x}^{'}, \tag{2.16}$$

which is the forward wavefield modelling operator. In other words, we have

$$d(\omega, \mathbf{x}_r, \mathbf{x}_s) = -\int G_0(\mathbf{x}_r, \mathbf{x}_s, \omega; \mathbf{x}')\, \omega^2 m(\mathbf{x}^{'})\, \theta_0(\omega, \mathbf{x}^{'})\, d\mathbf{x}^{'}. \tag{2.17}$$

The last equation can be written in compressed matrix-times-vector multiplication form as follows

$$\mathbf{d} = \mathbf{A}\mathbf{m}, \tag{2.18}$$

where $\mathbf{A}$ is our shot profile forward modelling/de-migration operator, $\mathbf{d}$ denotes the seismic measurements represented by a vector and the vector $\mathbf{m}$ stands for the acoustic potential. Clearly, writing linear operators in matrix-vector form simplifies our notation but it is important to stress that the matrix $\mathbf{A}$ is never formed. Applying $\mathbf{A}$ to the vector $\mathbf{m}$ is equivalent to applying forward modelling. In other words, the data $d(\omega, \mathbf{x}_r, \mathbf{x}_s)$ is computed by operating on a subsurface model $m(\mathbf{x})$.

If $\mathbf{A}$ is the forward operator, one can also define the adjoint operator $\mathbf{A}^T$ which is the transpose of the matrix $\mathbf{A}$. In function form, the adjoint is given by the following

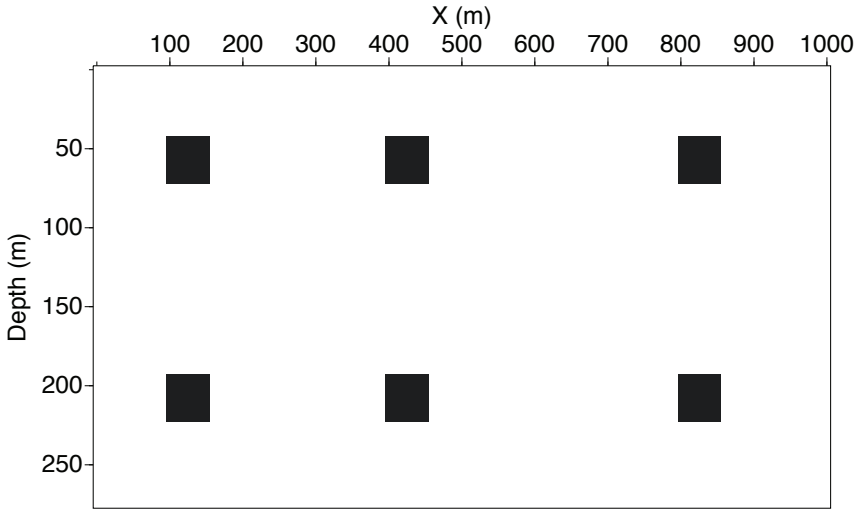

Figure 2.1: True point diffractor model representing the reflectivity of the subsurface.

expression

$$m_{adj}(\mathbf{x}) = = -\int_{\mathbf{x}_s'}\int_{\mathbf{x}_r'}\int_{\omega}(\omega^2\theta_0^*(\mathbf{x},\omega;\mathbf{x}_s')\; G_0^*(\mathbf{x}_s',\mathbf{x}_r',\omega;\mathbf{x})\; d(\mathbf{x}_r',\mathbf{x}_s',\omega))\; d\omega\; d\mathbf{x}_r'\; d\mathbf{x}_s', \tag{2.19}$$

which is equivalent, in matrix-vector form, to

$$\mathbf{m}_{adj} = \mathbf{A}^T\mathbf{d}\,. \tag{2.20}$$

One can combine equations 2.18 and 2.20 to build a relationship between the adjoint migrated image and the true scattered potential (reflectivity)

$$\mathbf{m}_{adj} = \mathbf{A}^T\,\mathbf{A}\mathbf{m}\,. \tag{2.21}$$

The last equation shows that the image obtained via the adjoint operator is not equal to the true image $\mathbf{m}$ unless $\mathbf{A}^T\mathbf{A} = \mathbf{I}$. The latter is not true because $\mathbf{A}^T$ is not the inverse of $\mathbf{A}$. However, in classical migration one adopts adjoint operators

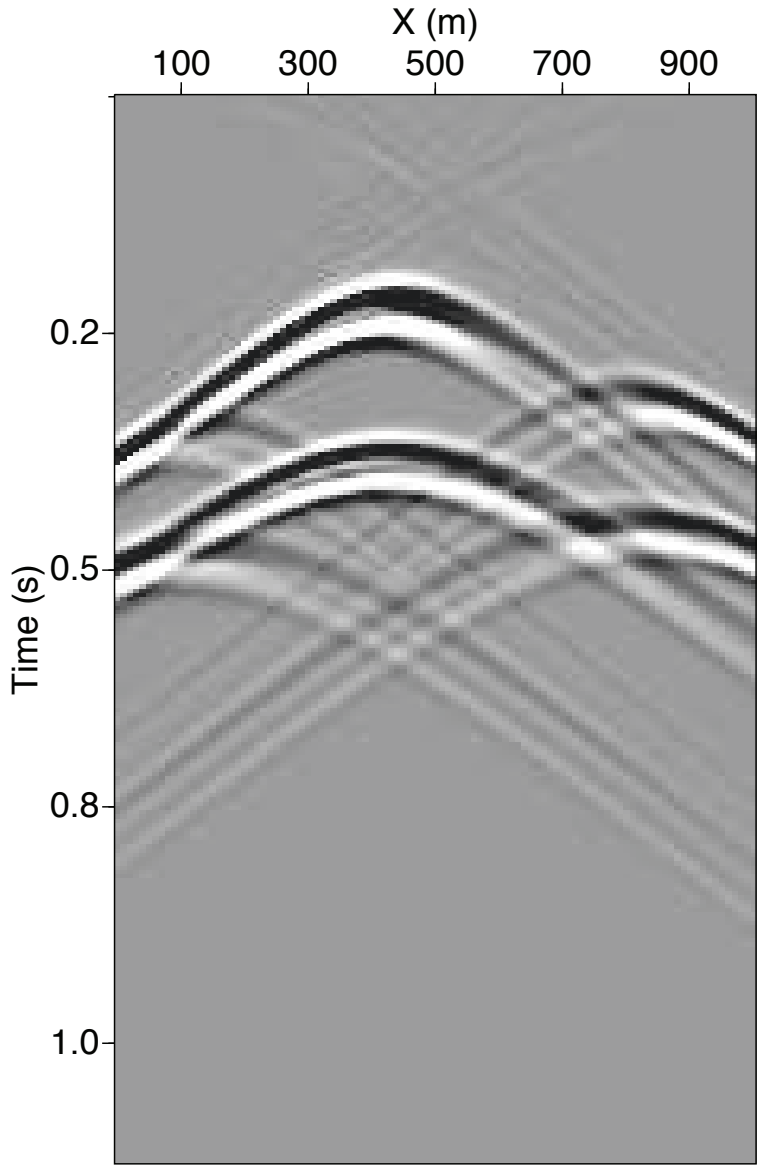

Figure 2.2: Simulation of the recored data using equation 2.17.

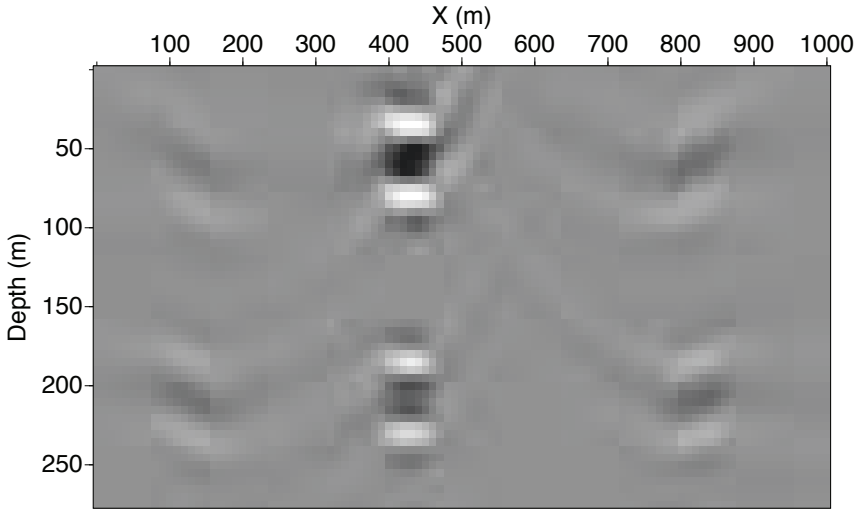

Figure 2.3: Adjoint migrated image using equation 2.19.

for imaging by considering that $\mathbf{A}^T\mathbf{A}$ is close to a diagonally dominant operator and, therefore, within a scale factor, one can assume that $\mathbf{m}_{adj} \approx \mathbf{m}$. One could also say that $\mathbf{m}_{adj}$ is a blurred version of $\mathbf{m}$ where the blurring operator is given by the Hessian operator $\mathbf{H} = \mathbf{A}^T\mathbf{A}$. Therefore, the solution of our problem involves inverting the Hessian operator

$$\mathbf{m} = \mathbf{H}^{-1}\mathbf{m}_{adj} \,. \tag{2.22}$$

The latter will be the core of this thesis. However, we stress that the direct inversion of $\mathbf{H}$ is not possible. Therefore, we will provide a semi-iterative solution that includes constraints.

2.3 Examples

To have a better understanding of the application of forward and adjoint operators, we generated a simple model composed of diffractors (Figure 2.1). Now consider a point source in the middle of the model at the surface. We will try to simulate the data by using our forward modelling operator (i.e., by using equation 2.17).

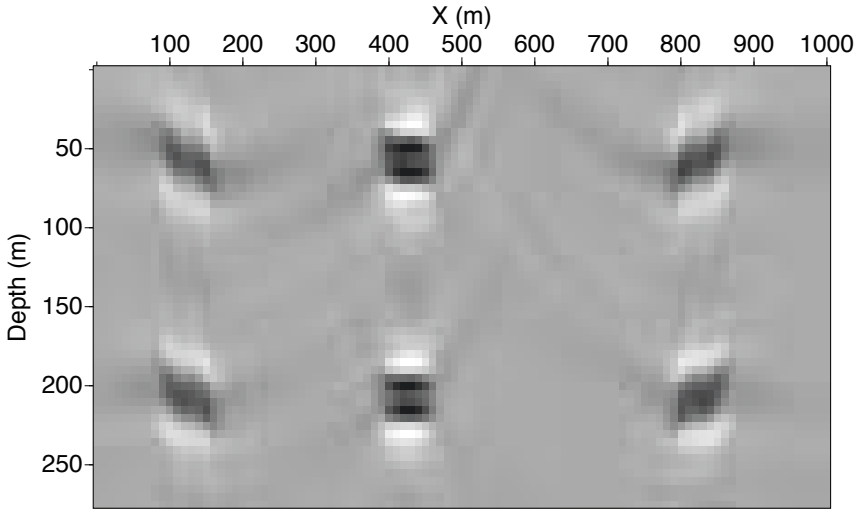

Figure 2.4: Inverted image using equation 2.22.

Receivers are located on the surface with an interval of 10 m and they are all recording the reflected wavefield generated by the point source in the middle of the model. Figure 2.2 shows the modelled data for the numerical experiment. The source function is a Ricker wavelet with a dominant frequency of 30 Hz and time sampling interval of 4 ms.

Equation 2.19 was adopted to migrate the recorded data. Figure 2.3 shows the migrated image. As it is evident from the image, the migrated image resembles the true reflectivity model, but there are amplitude and resolution problems. To balance the amplitudes and increase the resolution we need to remove the action of Hessian operator from the migrated image. By using equation 2.22 we can achieve the goal above. The algorithms for the solution of equation 2.22 will be explained in next chapter. Figure 2.4 shows the high-resolution image that was obtained after removing the effect of the Hessian from the migrated image. One can argue that the inverted model in Figure 2.4 honours the recorded data, shown in Figure 2.2, better than the adjoint migrated image (Figure 2.3). To prove this behaviour, we applied equation 2.17 to the migrated and inverted images. Figure 2.5 shows the predicted data computed by modelling the migrated image and Figure 2.6 depicts the predicted data obtained by modelling the inverted image. As it is evident from

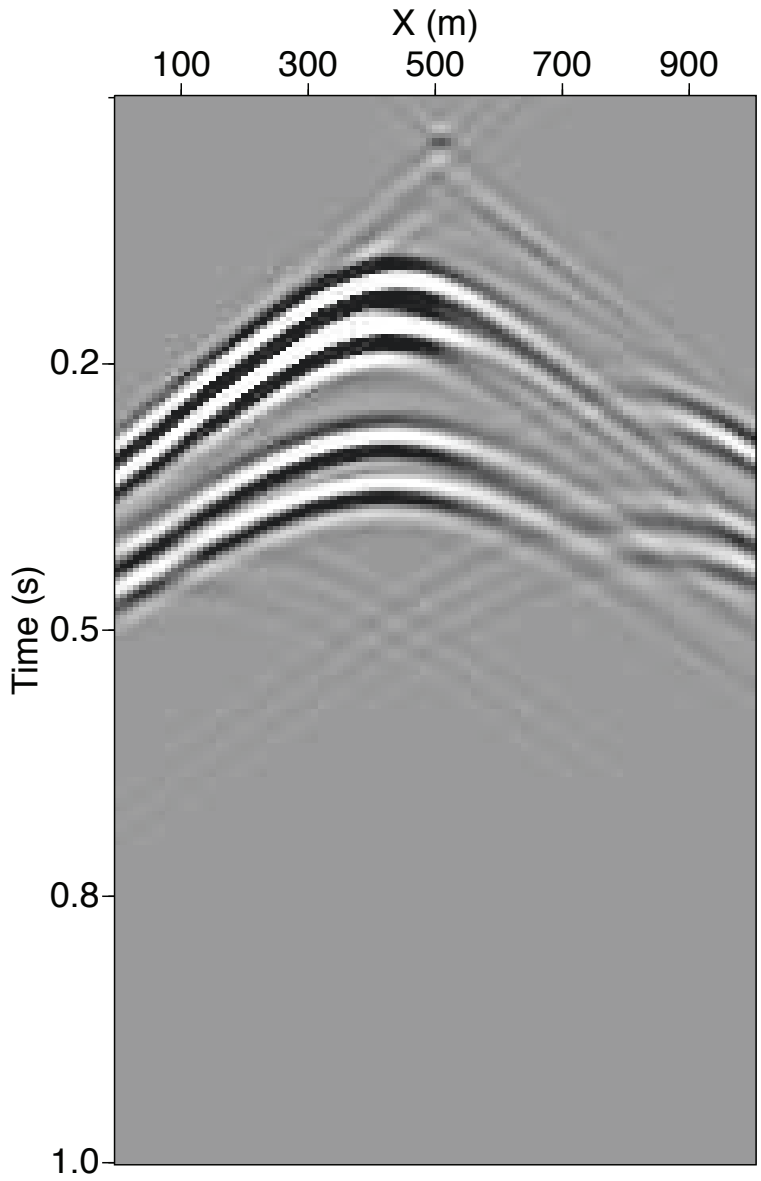

Figure 2.5: Predicted data using equation 2.17 and adjoint image as a subsurface reflectivity.

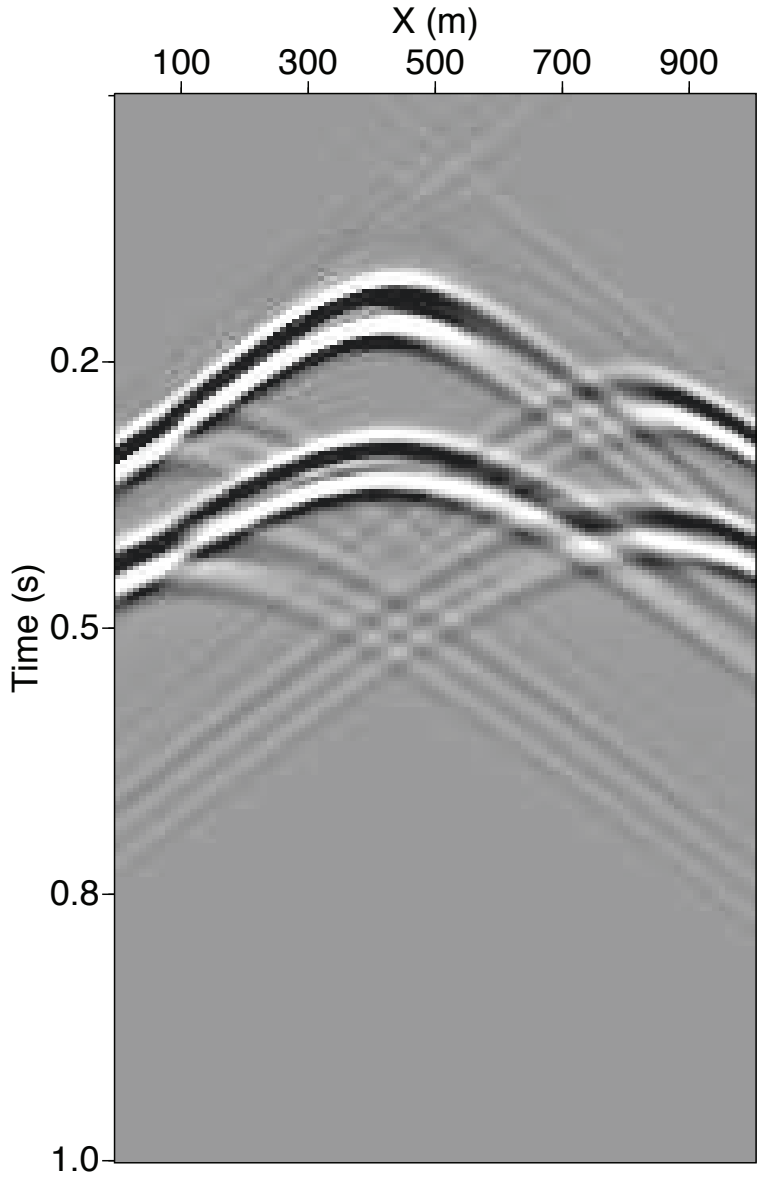

Figure 2.6: Predicted data using equation 2.17 and inverted image as a subsurface reflectivity.

the images, the inverted model can better predict the original recorded data.

2.4 Summary

The Born approximation was adopted to derive forward and adjoint operators for shot profile least-squares migration algorithms. The adjoint and forward pairs are necessary for communicating from data to model and model to data spaces. These operators, in the upcoming chapters, are the engines for semi-iterative solutions to the least-squares migration problem.

CHAPTER 3

Block row recursive least-squares migration[1]

3.1 Introduction

Seismic migration aims to produce true structural and stratigraphical images of the subsurface. Different migration algorithms have been studied in the geophysical literature. Migration methods can be divided into two main categories. The first category contains migration techniques that adopt ray-tracing (e.g., Kirchhoff migration (Schneider, 1978)). Migration based on the ray-tracing method is computationally efficient, and it is easily adaptable to a non-regular acquisition geometry. Another category contains methods that adopt one-way and two-wave wave equation propagators (wave equation migration methods) . Although wave equation methods are computationally expensive, they provide accurate wavefield extrapolation and high-quality images for complex areas. In this approach we solve for one-way or two-way wave equations (Gazdag, 1978; Stolt, 1978; Gazdag and Sguazzero, 1984; Stoffa et al., 1990; Baysal et al., 1983; McMechan, 1983; Whitmore, 1983). The action of these migration operators will be calculated on the fly because it is not possible to have them in explicit matrix form. Moreover, it is worth mentioning that each one of the migration techniques has its own artifacts. These artifacts arise from approximations which were made at the time of designing the migration operators and also from the lack of optimal data density.

[1]A version of this chapter has been published in the journal Geophysics. Nasser Kazemi and Mauricio D. Sacchi (2015). Block row recursive least-squares migration. Geophysics, 80(5), A95-A101.

These artifacts can be removed by adopting data regularization techniques before imaging (Fomel and Guitton, 2006). Artifacts can also be attenuated by implementing least-squares migration (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,b). In least-squares migration, we try to fit the data by inverting the demigration operator and in general, we adopt constraints to minimize the artifacts produced by data incompleteness. However, the computational cost associated to solving a least-squares problem is high and with the present computer resources it is really difficult to implement the least-squares migration method in its full potential for industrial applications. In other words, the direct inverse computation of the Hessian is expensive, and we need to approximate its inverse (Hu et al., 2001; Etgen, 2002; Guitton, 2004; Yu et al., 2006; Lecomte, 2008; Toxopeus et al., 2008; Naoshi and Schuster, 2009; Kazemi and Sacchi, 2014a). Another way of reducing the computational cost of least-squares migration is by adopting encoding methods (Dai et al., 2011; Wei et al., 2010). The other way to look at the problem is to apply least-squares migration in a recursive fashion. In this approach, we can implement the technique on a limited memory machine and in a fast way.

Recursive least-squares algorithms are extensively used for adaptive filtering in the case of dynamic and stationary environments. The main idea is to implement infinite memory algorithms by solving the problem via introducing one data point at a time to the system of equations (e.g., adding one row at the time to the data matrix, or in our case, to the demigration operator). By infinite memory algorithm, we mean a recursive method that operates at a given time on a small segment of data but without forgetting the influence of previous segments of data in the current solution. However, little effort has been made by the geophysical community to adapt this technique for different practical applications. This is mainly because recursive least-squares algorithms require actual matrices. In large scale problems (e.g., migration) everything will be done *on the fly* and also to add or remove one row from the system of equations (e.g., demigration operator) has no real physical meaning. Moreover, the one data point update scheme in the recursive least-squares approach causes stability issues. To tackle these shortcomings, one can update the system of equations in blocks by introducing more than one data point at each step, making the algorithm faster and more stable. This kind of algorithms belongs to the family of block row or block column recursive least-squares. However, in this family,

we need to explore structures in the data matrix to make the algorithms faster and also update the solution in a way that in the end, we solve for the original least-squares problem with sufficient accuracy. However, in most practical geophysical inverse problems we do not have explicit matrices. Of course, there are cases such as in Autoregressive (AR) filtering applications where one can explore unique structures in the data matrix to derive highly efficient recursive algorithms. For example, Naghizadeh and Sacchi (2009) used a rank one update of the recursive least-squares fitting with an exponentially weighted forgetting factor for $f - x$ adaptive filtering in the context of seismic interpolation. It is evident that this technique cannot be applied to least-squares migration.

In this chapter, we propose a block row recursive method to solve the least-squares migration problem. The method, in essence, is close to the work of Ng and Plemmons (1996). We consider recursive least-squares solutions of the wave equation migration with sliding windows involving several rank K down dating and updating computations. The least-squares estimator can be found by solving a small least-squares migration problem in each step, recursively. From the practical point of view, to name a few, the blocks can be a group of shot gathers, group of frequency slices or group of offset classes.

The outline of the chapter is as follows. First, least-squares migration will be explained. Then, we will introduce a block row recursive algorithm. Moreover, we will discuss some of the practical considerations for the proposed method and compare the computational cost of the method with full least-squares migration and examine the efficiency of the proposed method on a simple toy example and the Marmousi model. Finally, the main conclusions will be summarized.

3.2 Least-squares migration

A data generating model under the action of the demigration operator $\mathbf{A}$ can be written as

$$\mathbf{d} = \mathbf{A}\mathbf{m} + \mathbf{n}, \tag{3.1}$$

where $\mathbf{d}$ is $N \times 1$ vectorized version of the recorded data at the surface, $\mathbf{m}$ is the migrated model with $M \times 1$ and $\mathbf{n}$ is the additive noise and, often, also a term to absorb waves not modelled by the demigration operator. Using the adjoint operator

of demigration, one can estimate the migrated model

$$\hat{\mathbf{m}} = \mathbf{A}^T \mathbf{d}, \tag{3.2}$$

where $\hat{\mathbf{m}}$ is the adjoint estimated image of the earth and $\mathbf{A}^T$ is the migration operator. While $\hat{\mathbf{m}}$ can capture the main structures of the true geological model $\mathbf{m}$, the produced model does not honour the data. In other words, application of the demigration operator on the migrated image yields a poor data prediction. Also, the migrated image contains blurring and sampling artifacts. These artifacts come from the fact that migration and demigration operators are not orthogonal and the energy of the signal in the complimentary image space of the operator will be zeroed out. To tackle the problem, we will solve

$$\tilde{\mathbf{m}} = \underset{\mathbf{m}}{\operatorname{argmin}} \; ||\mathbf{A}\ \mathbf{m} - \mathbf{d}||_2^2 + \lambda\ ||\mathbf{m}||_2^2, \tag{3.3}$$

where $\mathbf{m}$ is desired model, $\mathbf{A}$ is demigration operator, $\mathbf{d}$ is the recorded data and λ is a regularization parameter (Nemeth et al., 1999; Kuhl and Sacchi, 2003; Kaplan et al., 2010b). The cost function of equation 3.3 is convex and has a closed form solution

$$\tilde{\mathbf{m}} = (\mathbf{A}^T\mathbf{A} + \lambda\ \mathbf{I})^{-1}\ \mathbf{A}^T\mathbf{d}. \tag{3.4}$$

To solve the problem efficiently, researchers take advantage of semi-iterative algorithms such as well-known Conjugate Gradient algorithm (see appendix A.1). We usually stop the iterations whenever the misfit of the normal equation reaches below a predefined threshold (for the convergence behaviour of the Conjugate Gradient algorithm see Appendix B.1). However, the computational cost of the least-squares migration solution can be high and to make the algorithm faster, one can approximate the Hessian (i.e., $\mathbf{A}^T\mathbf{A}$) inverse (Hu et al., 2001; Kazemi and Sacchi, 2014a). In next section, we will propose a sliding window scheme for solving equation 3.3 in a recursive fashion. The main motivations behind the method are: reducing the computational cost and at the same time producing migrated images that honour the recorded data using memory limited resources.

3.3 Block row recursive least-squares migration

In this section we will follow the recursive least-squares solution via rank K updating and rank K downdating procedure introduced by Ng and Plemmons (1996). However, there are some differences between the proposed method in Ng and Plemmons (1996) with our technique. We are not considering near Toeplitz structure for the data matrix and also in the updating procedure, we propose to use the previous solution of the block row setup as an initial solution for the next sliding window.

To explain the block row recursive least-squares method, let us consider again the problem of equation 3.3. In recursive least-squares computations, it is required to calculate $\mathbf{m}$ while observations are successively added to, or deleted from the system of equations. Suppose we have estimated the model with the first set of measurement points $\mathbf{d}_0$ in least-squares sense

$$\mathbf{m}_0 = (\mathbf{A}_0^T \mathbf{A}_0)^{-1}\, \mathbf{A}_0^T \mathbf{d}_0, \tag{3.5}$$

Now, the question we should ask is that by introducing new data points to the system of equations, can the best estimate for the combined system $\mathbf{A}_0\ \mathbf{m} = \mathbf{d}_0$ and $\mathbf{A}_1\ \mathbf{m} = \mathbf{d}_1$ be estimated using only $\mathbf{m}_0$ and $\mathbf{d}_1$? We define a new matrix

$$\mathbf{P}_1^{-1} = \mathbf{A}_0^T \mathbf{A}_0 + \mathbf{A}_1^T \mathbf{A}_1, \tag{3.6}$$

then we have

$$\mathbf{m}_1 = \mathbf{P}_1\ (\mathbf{A}_0^T \mathbf{d}_0 + \mathbf{A}_1^T \mathbf{d}_1), \tag{3.7}$$

note that $\mathbf{m}_1$ is the best model for the combined system of equations. At this point we need to eliminate $\mathbf{d}_0$ term from equation 3.7. Let us rewrite equation 3.6 as

$$\mathbf{P}_1^{-1} = \mathbf{P}_0^{-1} + \mathbf{A}_1^T \mathbf{A}_1, \tag{3.8}$$

and after few algebraic manipulations, one can show that equation 3.7 changes to

$$\mathbf{m}_1 = \mathbf{m}_0 + \mathbf{P}_1 \mathbf{A}_1^T (\mathbf{d}_1 - \mathbf{A}_1 \mathbf{m}_0), \tag{3.9}$$

where $\mathbf{P}_1 \mathbf{A}_1^T$ is the gain factor. In the case of one data point update we can use the Matrix Inversion Lemma (MIL) and calculate the gain factor without direct inver-

sion. In the case of a block wise update (more than one data point update) to apply fast calculations of the gain matrix, one could explore MIL with QR decomposition. However, none of these approaches are applicable to least-squares migration. This is because in seismic migration, linear operators are not given by actual matrices.

However, it is easy to show that if the system of equations satisfies some assumptions, we can relax the gaining factor term from equation 3.9 and solve the problem recursively on overlapping windows via the method of Conjugate Gradients (CG). Let us explain the method step by step. The least-squares estimator at step i can be found by solving for the $M \times 1$ vector $\mathbf{m}(b_i)$ in

$$\tilde{\mathbf{m}}(b_i) = \underset{\mathbf{m}(b_i)}{\operatorname{argmin}} \, ||\mathbf{A}(b_i)\ \mathbf{m}(b_i) - \mathbf{d}(b_i)||_2^2 + \lambda\ ||\mathbf{m}(b_i) - \tilde{\mathbf{m}}(b_{i-1})||_2^2, \tag{3.10}$$

where $\tilde{\mathbf{m}}(b_i)$ is the least-squares solution of the model till block i and $\mathbf{d}(b_i)$ is the recorded data corresponding to block i with size $Q \times 1$ and $\mathbf{A}(b_i)$ is the $Q \times M$ data matrix of block i and Q is the length of sliding window. Figure 3.1 shows the schematic representation of the setup. To update the solution recursively, we will add K data points to the system of equations and remove K data points from the beginning of the previous data vector. We call this step the rank K updating and downdating (see Figure 3.1 for more information). For this new configuration, one can use

$$\tilde{\mathbf{m}}(b_{i+1}) = \tilde{\mathbf{m}}(b_i) + (\mathbf{A}^T(b_{i+1})\mathbf{A}(b_{i+1}) + \lambda\mathbf{I})^{-1}\ \mathbf{A}^T(b_{i+1})[\mathbf{d}(b_{i+1}) - \mathbf{A}(b_{i+1})\tilde{\mathbf{m}}(b_i)]\,, \tag{3.11}$$

to estimate the best model in the least-squares sense that fits the whole system of equations till step $i+1$. The second term in equation 3.11 can be interpreted as the least-squares solution of the unpredicted part of the new data set by the solution of the previous step. It is worth mentioning that we use CG to solve equation 3.11 and we never calculate the inverse of the Hessian (i.e., $(\mathbf{A}^T(b_{i+1})\mathbf{A}(b_{i+1}) + \lambda\mathbf{I})^{-1}$). Using the previous solution as an initial model for the next block has some advantages. This warm start for the new sliding window will result in fast convergence of the CG method. This is mainly because the nearby data points are highly correlated in the seismic acquisition. It is also good to point out that in each setup we must use proper preconditioners and regularization term to cluster the eigenvalues of the partial Hessians around one. Ng and Plemmons (1996) proved the convergence of this recursive least-squares technique in the probabilistic terms. They

showed that the method would converge superlinearly with probability one provided the underlying process satisfies some assumptions. First of all, the input discrete-time stochastic process should be stationary. Secondly, the auto-covariances of the kernels in each step should be absolutely summable. This, in turn, will assure the invertibility of the processes in each windowed setup. Thirdly, the variances between the auto-covariances of the kernels between different setups should be bounded. Finally, the stationary process has zero mean. All of these assumptions are valid for many time series analysis problems but it is not clear if they are fully applicable to migration. However, in some cases one can show that these assumptions more or less are valid for migration (e.g., the acquisition system has full aperture and sufficiently fine sampling in time and spatial directions). Moreover, the physical properties of the medium such as slowness should be spatially invariant (Stolk, 2000). For this idealized situation, one can follow research by Gelius et al. (2002), Sjoeberg et al. (2003) and Lecomte (2008) where it is shown that the action of the Hessian can be approximated by convolutional operators. Note that this relationship is only valid for spatially invariant media (i.e., constant slowness in spatial directions).

Nevertheless, the ideas in Ng and Plemmons (1996) can be explored for migration applications, and we will show that this method works quite well even for a complex medium. In next section, we will show the efficiency of the proposed method using a simple toy example and the Marmousi model.

3.4 Practical considerations

It is worth mentioning that in this chapter we assume that the velocity field is known and there is no need for residual velocity analysis. From the application point of view, the proposed method can be implemented via different configurations. The key element to keep in mind is the fact that we need overlapping between consecutive blocks to assure smooth changes in the Hessians. This method can be applied on a subset of shot gathers in the context of shot profile migration or in the frequency domain using a subset of frequency slices with overlap. The subwavefieldset of shots can be chosen with regular patterns (e.g., in 2D from left to right or right to left). We cannot use a simple stochastic gradient method with random selection of shot gathers because we need overlapping blocks. In the frequency domain, one could start from low frequencies and move to high frequencies. Application of the proposed

technique using near offsets (near angles) and moving to far offsets (high angles) is also straightforward. Block row recursive least-squares method can also be extended to 3D imaging. One has the options of using a subset of shots or offset classes as a block, provided that a proper migration operator is chosen.

3.5 Computational cost analysis

To compare the computational cost of full least-squares migration and the proposed block row recursive method let us assume that the number of grid points in one coordinate direction is n, hence in 2D coordinates, there are $O(n^2)$ points. We will borrow some of the complexity estimates from Marfurt and Shin (1989) and Mulder and Plessix (2004a) and Mulder and Plessix (2004c). In general, one-way wave equation has a complexity of $O(n_s n_\omega n^2)$ for two-dimensional acquisition system where n_s is the total number of shots and n_ω is the number of frequency realizations in data space. Considering the scheme described in Collino and Joly (1995), the complexity of one-way wave equation in three dimensions increases by the factor of n (i.e., the complexity is $O(n_s n_\omega n^3)$). Moreover, in the CG method in each iteration, we need to call migration and demigration operators. Let us say CG will converge in N iterations. So, the computational cost of the CG method for the 2D case is about $O(2N n_s n_\omega n^2)$. On the other hand, the cost of block row recursive least-squares migration in the frequency domain is $O(2\bar{N} n_s n_b n_\omega^b n^2)$ where $\bar{N}$ is the average number of iterations per block, n_b is the number of blocks and n_ω^b is the number of frequency realizations in each block. In the case of a subset of shot gathers as a block, the cost will be $O(2\bar{N} n_b n_s^b n_\omega n^2)$ where n_b is the number of shot gather blocks and n_s^b is the number of shot gathers per block. Moreover, to make the algorithm more efficient, one can image a subset of the model for each block and reduce the cost even more. Hence, the new cost will be $O(2\bar{N} n_b n_s^b n_\omega n n_m)$ where n_m is the number of grid points of the model in the horizontal direction that covers the illumination aperture corresponding to the geometry of the subset of shots.

In the case of 3D imaging, the I/O cost can also be drastically decreased. In the block row recursive approach we need to load the data for each subset of shot gathers only once, rather than repeatedly as is the case of CG applied to the full least-squares migration problem when the data size exceeds core storage, which is routinely the case in 3D imaging.

3.6 Examples

In the following examples, we used shot profile wave equation adjoint and forward operators with the split step correction in the context of adjoint, least-squares and block row recursive least-squares algorithms.

To test the performance of the proposed method, we generated a 2D reflectivity model (Figure 3.2a) and we used the demigration operator to produce the data set with a 10 meters shot interval and a 5 meters receiver interval. We used a Ricker wavelet with a dominant frequency of 30 Hz and the receivers were active for all of the shots. Then, we applied the adjoint operator to migrate the data set (Figure 3.2b). Figures 3.2c and d show the least-squares and the block row recursive approach migrated models, respectively. In the case of the block row recursive approach, we used 10 consecutive shot gathers in each group, and we deleted 5 shot gathers from the beginning of the previous windowed setup and added 5 new shot gathers to the end of the new windowed setup. The stopping criterion, for both methods, was set to be equal to the drop of the data residual of the first iteration on the order of 10^6 and the regularization parameter was set to $\lambda = 100$. Please note that in the case of the block row recursive method the stopping criterion was defined for each block separately. The full least-squares approach converged after 13 iterations, and in the case of the block row recursive method, the average iteration number per block was approximate $\bar{N} = 3$. Hence, the speed up for this experiment was about a factor of two. It is worth mentioning that for this experiment all the receivers were active and we cannot use a second cost criterion (e.g., $O(2\bar{N}n_b n_s^b n_\omega n n_m)$) for the block row recursive method. The block row recursive approach did a good job in recovering the true reflectivity model, and the result is comparable to that of least-squares migration with the whole data at once. To show how they honour the recorded wavefield, we used the recovered reflectivity models to predict the data set. Figure 3.3a shows the true near offset section of the data set and Figure 3.3b is the predicted near offset section using adjoint migrated model. Finally, Figures 3.3c and d show the predicted near offset sections using least-squares migration and the block row recursive least-squares migration, respectively. It is clear that the predicted near offset sections via the least-squares approach and the block row recursive methods are in good accordance with the true near offset section.

Next, we applied the method to the Marmousi model. It is worth mentioning that

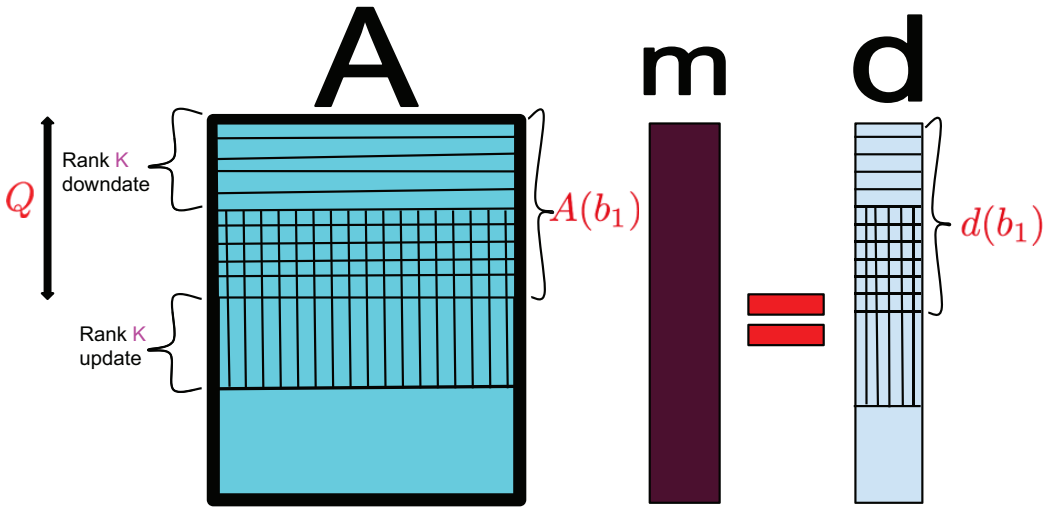

Figure 3.1: Schematic representation of the block row recursive least-squares migration algorithm.

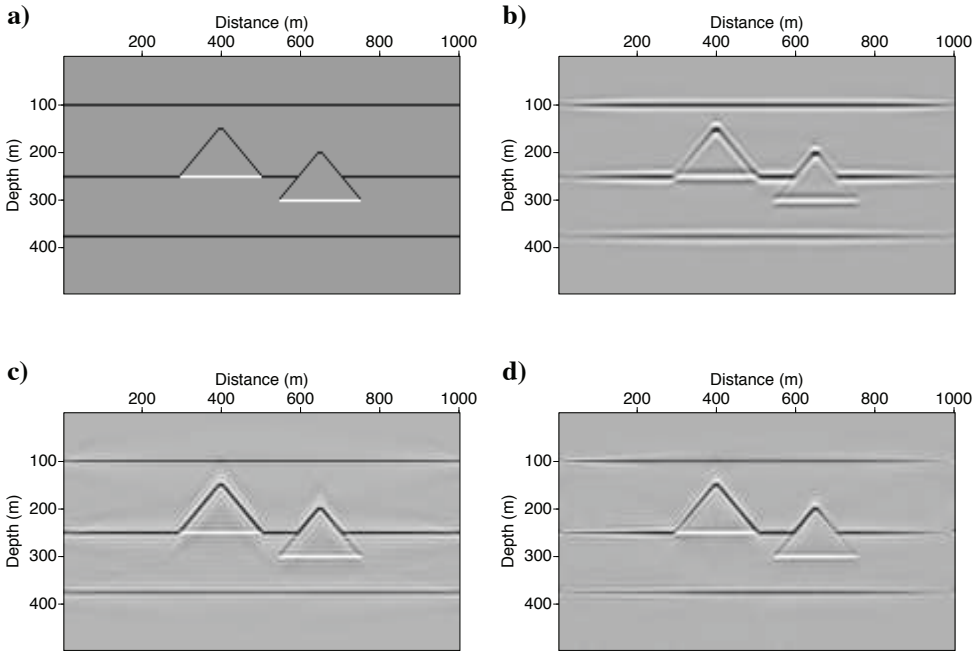

Figure 3.2: True and migrated reflectivity models. a) True reflectivity. b) Image obtained via the adjoint operator (classical migration). c) Image obtained via least-squares migration. d) Image obtained with the proposed block row recursive least-squares migration algorithm.

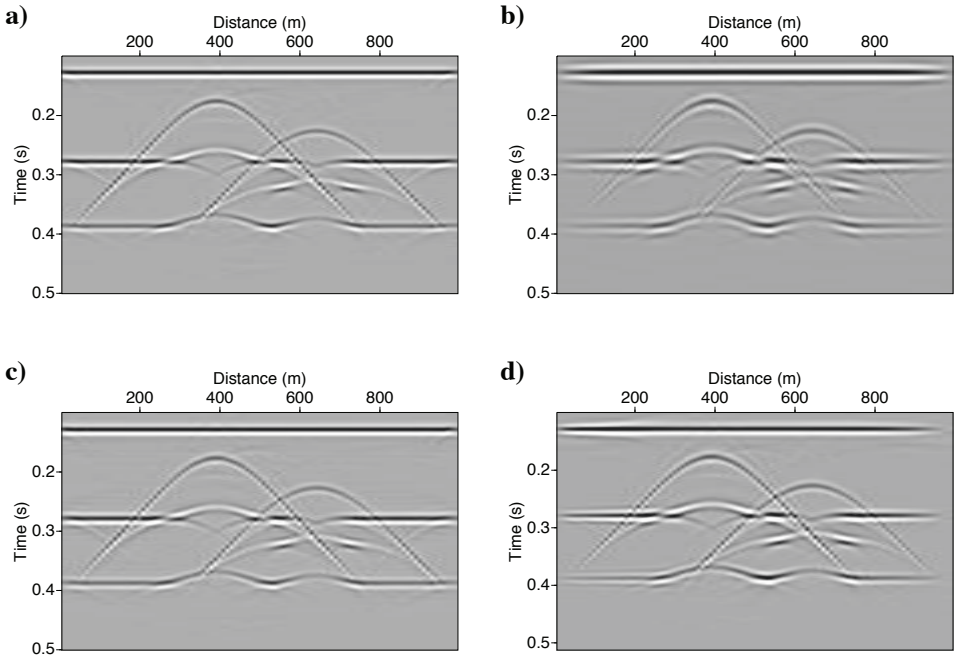

Figure 3.3: True and predicted near offset sections. a) True near offset section. b) Adjoint predicted section. c) Least-squares predicted section. d) Block row recursive least-squares predicted section.

the data set was generated by the finite difference method with a Ricker wavelet with a dominant frequency of 20 Hz. The data set consist of 240 shot gathers with 25 m shot interval that modelled with an off end survey with receivers to the left of the source being pulled towards the right. Each shot gather consists of 96 traces with the smallest offset being 200 m and receiver intervals are 25 m. Figure 3.4 shows the near offset section of the dataset. The non-smooth velocity model of the Marmousi is shown in Figure 3.5a. Figure 3.5b represents the adjoint migrated image of the Marmousi data set using shot profile wave equation migration with split step correction. Figure 3.6a shows the least-squares migrated image of the Marmousi data set after 15 iterations using split step Fourier migration and demigration operators as adjoint and forward operators. The stopping criterion for full least-squares solution was set to be equal to the drop of the data residual of the first iteration of the order of 10^3 and the regularization parameter was set to $\lambda = 1$. Finally, Figure 3.6b shows the migrated image produced by the block row recursive approach. In the case of the block row recursive approach we tested different configurations, and finally, we used 5 consecutive shot gathers in each group, and we deleted 3 shot gathers from the beginnings of the previous windowed setup and added 3 new shot gathers to the end of new windowed setup. The stopping criterion for block row recursive approach was set to be equal to the drop of the data residual of the first iteration on the order of 10^3 and the regularization parameter was set to $\lambda = 1$. Please note that the stopping criterion was defined for each block separately. The average iteration number per block was approximately $\bar{N} = 4$ and we chose n_m to be one-third of the model size in the horizontal direction. Hence, the speed up for this experiment was approximately by the factor of six. It is worth mentioning that we obtained the same results using different configurations. As we mentioned earlier, it is also possible to use a subset of frequency realizations as a block. To do so, we first apply the algorithm on the low-frequency band and then we move to mid frequencies without overlapping between the previous and present frequency block and so on. After doing several tests, we came to the conclusion that in the case of blocks as frequency realizations we do not need overlapping between the blocks. However, we use the solution of the previous blocks (i.e., the migrated model in hand) as a warm start so that we could increase the convergence rate of the algorithm. Figure 3.7 shows the recursive algorithm's performance using frequency blocks. We chose three bandwidths as a low, mid and high-frequency range blocks. Figure 3.7a depicts the migration result after using the first block (i.e., low frequency range) and Figure 3.7b represents the result after

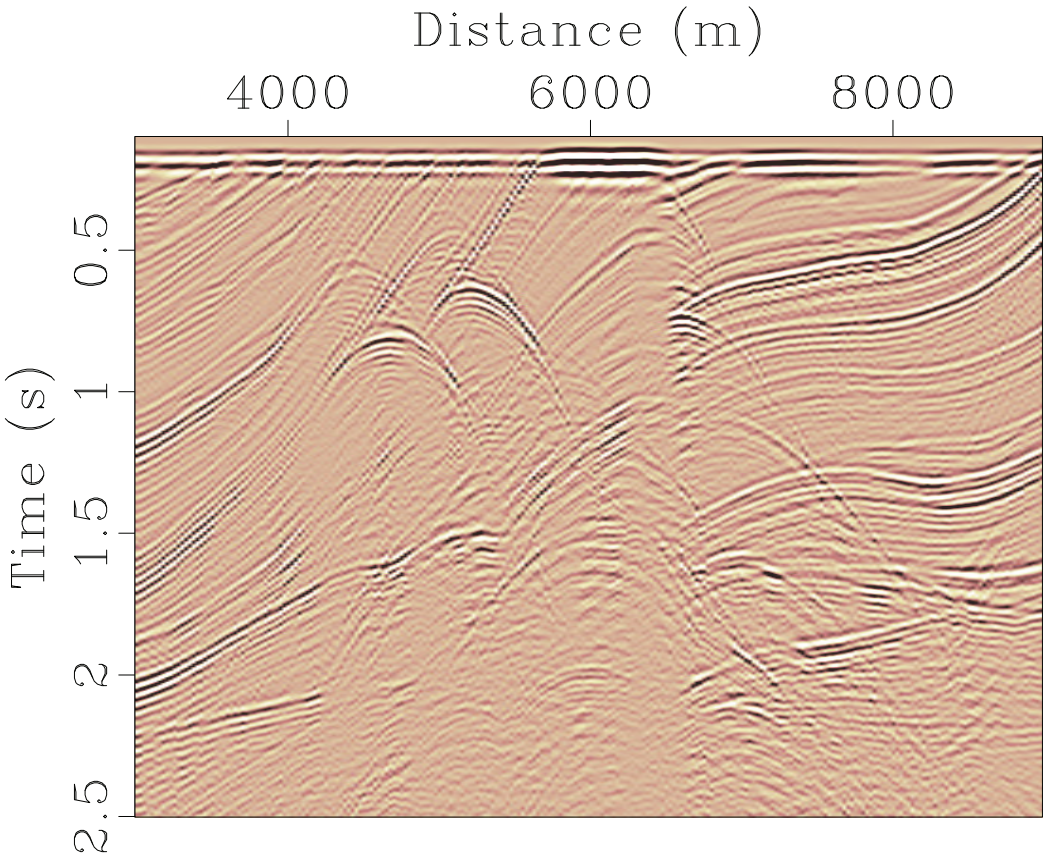

Figure 3.4: True near offset section of the Marmousi model.

recursively updating the model using the previous solution and mid range frequency block. Finally Figure 3.7c shows the result after using all three blocks. As it is clear from the figure the recursive result using frequency blocks resembles the result of a recursive algorithm using blocks of shots. In both cases, the block row recursive approach did a good job in recovering the true reflectivity model, and the results are comparable to that of the least-squares migration algorithm. Analyzing the results, the block row recursive approach did a reasonable job in preserving the amplitude of the reflectors and removed some of the defocusing problems of the image computed with the adjoint operator.

To show how they honour the recorded wavefield, we used the recovered reflectivity models to predict the data set. Figure 3.8a shows the adjoint predicted near offset section of the data set and Figure 3.3b is the predicted near offset section using the

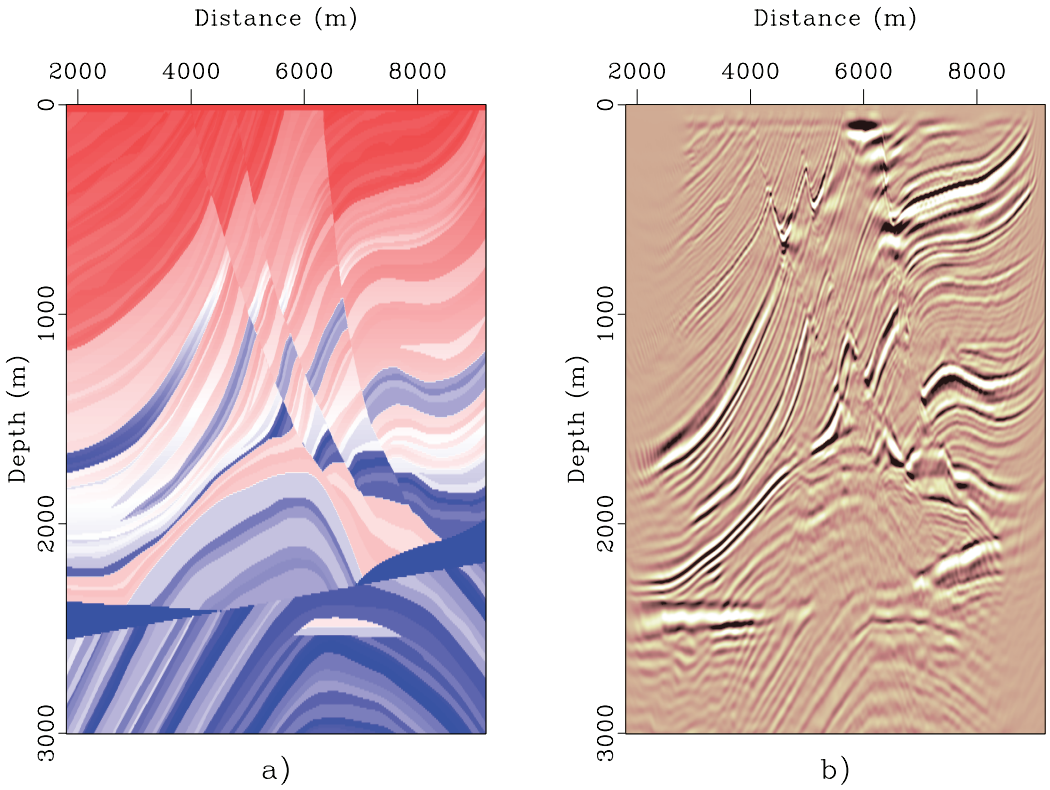

Figure 3.5: Velocity field of the Marmousi model and corresponding shot profile wave equation migrated image. a) True velocity field. b) Image obtained via migration (adjoint operator).

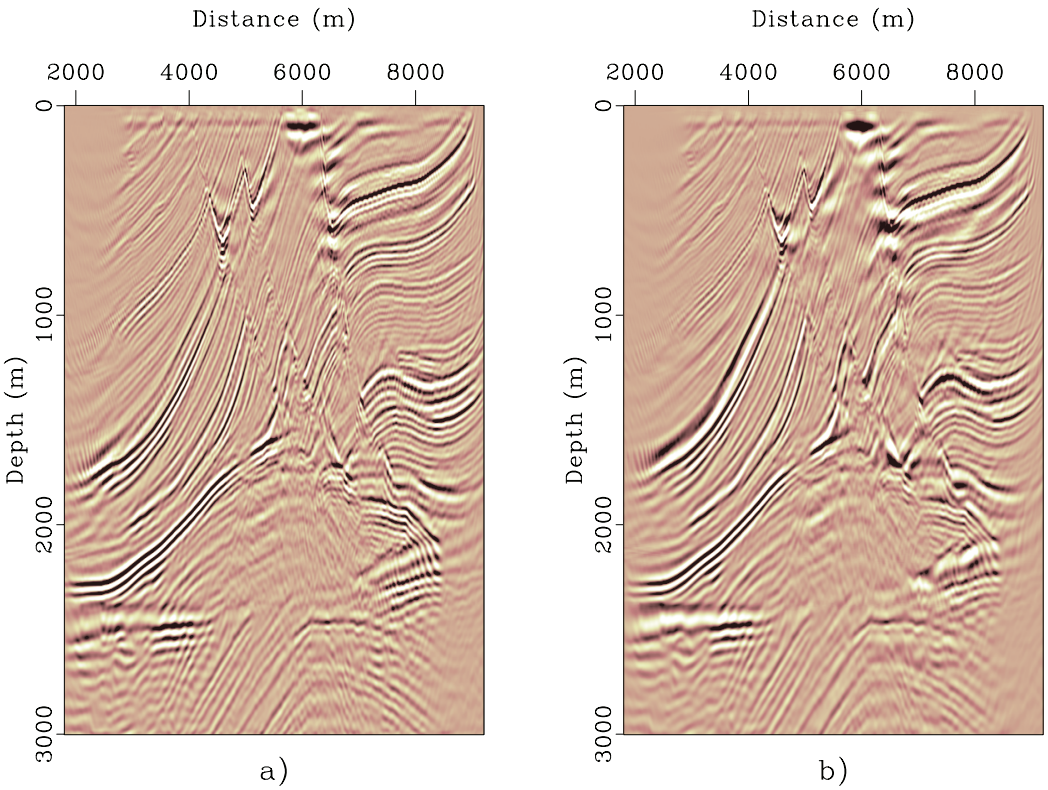

Figure 3.6: Comparison of the results of full least-squares migration and the proposed method on the Marmousi model. a) Full least-squares migrated model. b) Block row recursive least-squares migration with blocks of shot gathers.

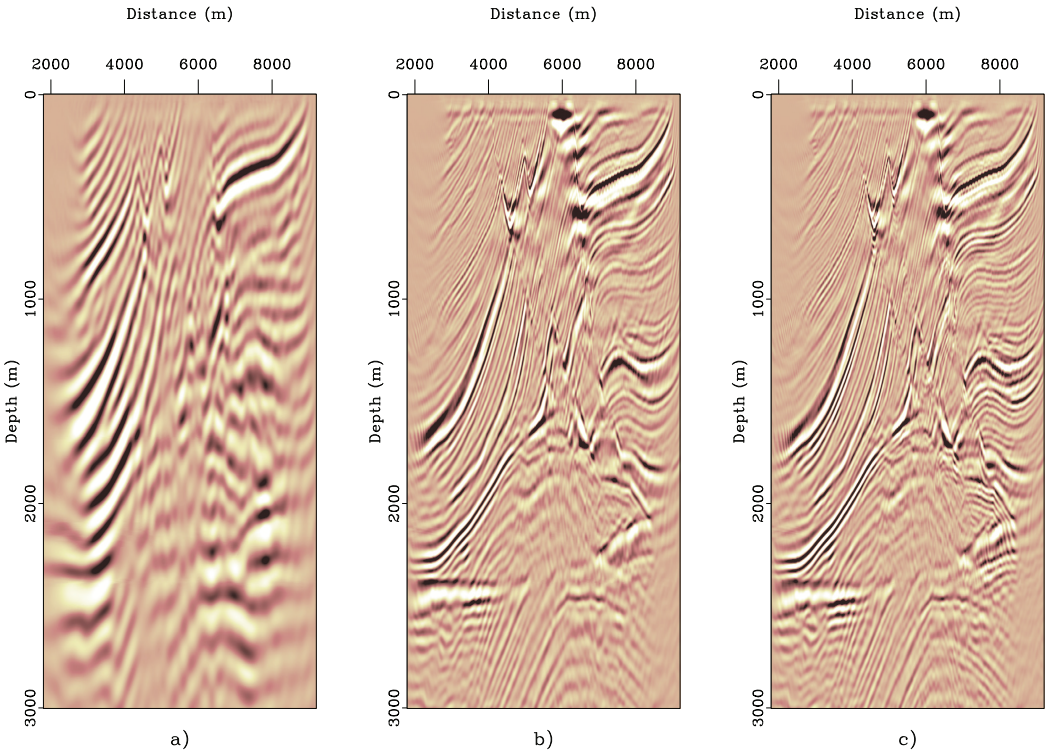

Figure 3.7: Block row recursive algorithm using the frequency blocks. a) Migration result after using low frequency block. b) Migration result after using low and mid frequency blocks. c) Migration result after using low, mid, and high frequency blocks.

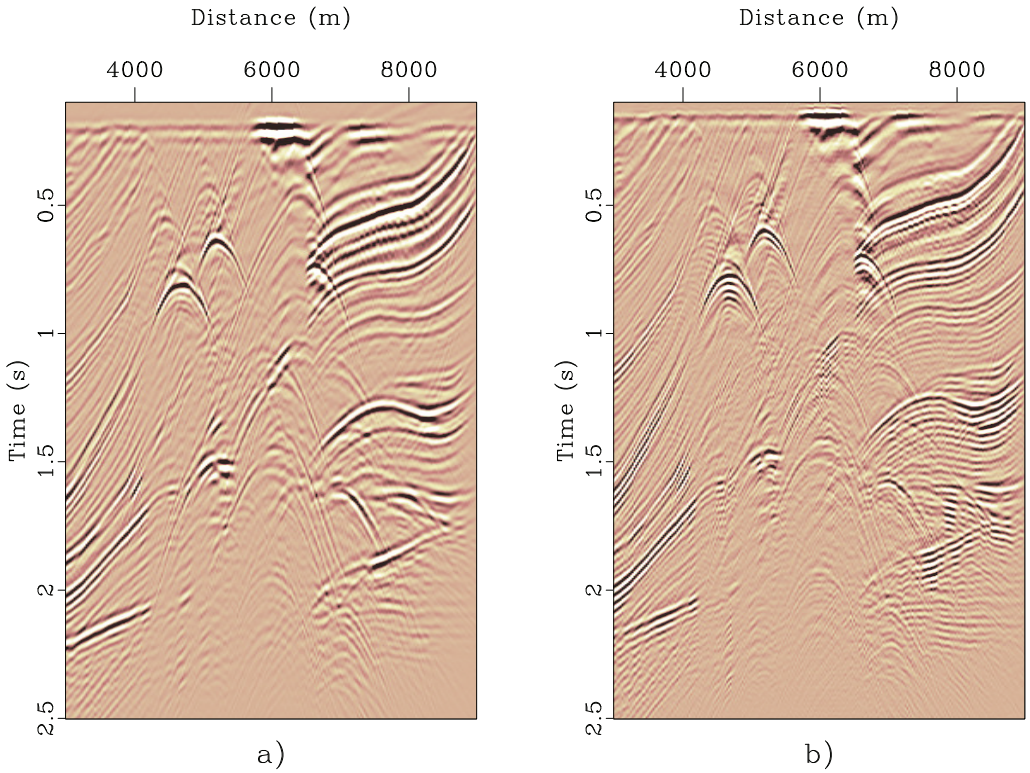

Figure 3.8: Adjoint migration and least-squares migration predicted near offset sections. a) Adjoint predicted section. b) Least-squares predicted section.

conventional least-squares migrated model. Finally, Figures 3.9a and b show the predicted near offset sections using block row recursive approach models with shot and frequency blocks, respectively. It is clear that the predicted near offset sections via the least-squares approach and the block row recursive methods are in good accordance with the true near offset section. Moreover, in both cases (i.e., subset of shots and frequency bands as blocks) the amplitudes are well preserved and also the detailed features are better predicted than the adjoint predicted dataset and the results are in good accordance with the true near offset section shown in Figure 3.4.

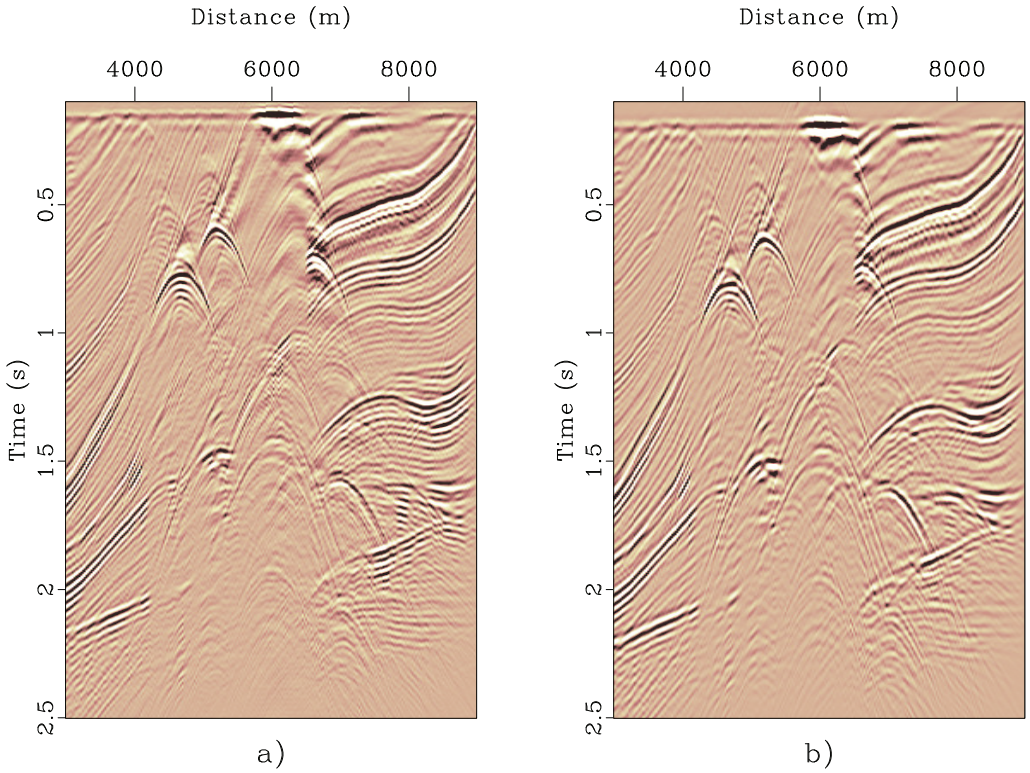

Figure 3.9: Block row recursive predicted near offset sections. a) Predicted section using the reflectivity model inverted via blocks of shots. b) Predicted section using the reflectivity model inverted via blocks of frequency realizations.

3.7 Summary

Recursive estimates of large systems of equations in the context of least-squares fitting is a common practice in different fields of study. For example, recursive adaptive filtering is extensively used in signal processing and control applications. The necessity of solving least-squares problems recursively stems from the need for fast real-time signal processing strategies. The computational cost of least-squares solvers can also limit the applicability of this technique in geophysical problems. We consider a recursive least-squares solution for least-squares wave equation migration with sliding windows involving several fixed rank downdating and updating computations. If we use enough data in each windowed setup the spectrum of the preconditioned system is clustered around one and the method will converge superlinearly with probability one. Numerical experiments are reported in order to illustrate the effectiveness of the technique for least-squares migration. The method uses block wise update of the demigration operator via rank K update and downdate in each setup while the new data points are successively added to the data vector. In each windowed setup, CG algorithm is used to solve the system of equations in a least-squares sense. To have fast convergence, the solution of one block is used as an initial solution for the next block. This warm start will result in fast convergence of the CG algorithm. This is supported by the fact that nearby blocks are highly correlated. The results of applying this technique on a simple toy and on the Marmousi model convinced us that the block row recursive method could be a practical tool for improving the spatial resolution of migrated images.

CHAPTER 4

Model domain pre-conditioners for extended least-squares wave equation migration

4.1 Introduction

Migrating the regularized data (Fomel and Guitton, 2006), or implementing least-squares algorithms (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,b) can alleviate artifacts introduced by acquisition footprint, and the results will honour the recorded wavefield. However, the computational cost of the least-squares migration approach can overshadow its merits. Application of least-squares algorithms on blended data (Dai et al., 2011; Wei et al., 2010; Cheng et al., 2016) or in a recursive fashion (Kazemi and Sacchi, 2015) are some of the alternative ways of reducing its computational cost. Authors also tried to approximate the inverse of the Hessian (Hu et al., 2001; Etgen, 2002; Guitton, 2004; Yu et al., 2006; Lecomte, 2008; Toxopeus et al., 2008; Naoshi and Schuster, 2009; Kazemi and Sacchi, 2014a), instead of adopting direct inversion methods. Interesting research, aiming to approximate the inverse of Hessian operator, can be classified under the umbrella of equation (4.12). The following researchers have tried to approximate the action of the Hessian via coefficients which are a function of phase and space. For instance, Claerbout and Nichols (1994) were able to extract information about the inverse of Hessian from a single application the Hessian on synthetic images. The latter was also followed by Rickett (2003) and resulted in similar solutions. Later on,

Guitton (2004) introduced a method based on non-stationary convolutions. In his method, the coefficients are only a function of phase. To have a proper homogeneity behaviour in phase domain, Symes (2008) invoked space dependent coefficients with weights that are proportional to the inverse of spatial wavenumbers. Nammour and Symes (2009) applied angle dependent scaling scheme to approximate the action of Hessian and its inverse on prestack migrated images. The coefficients, in this case, are a function of space and angle. Finally, Herrmann et al. (2009) explored the action of Hessian as a diagonal operator in Curvelet domain and Demanet et al. (2012) approximated the action of Hessian with coefficients computed by the action of Hessian on randomized test functions built in Curvelet domain. Demanet et al. (2012) argue that the randomness of the test functions can capture the action of the Hessian on a much larger linear subspace than is usually the case with a deterministic method (i.e., applying the Hessian on an adjoint migrated image). Recently, Huang et al. (2016) approximated the Hessian via a pseudo-differential scaling operator which is a function of space and dip.

Using preconditioners can also reduce the computational cost of least-squares migration. If we define an operator that clusters the eigenvalues of the Hessian and reduces the condition number of it, that operator is considered an optimal preconditioner. An approximated inverse of the Hessian operator has the potential of clustering the eigenvalues of the operator. By this logic, we will try to use the approximated version of the inverse of Hessian operator as a preconditioner to accelerate the convergence rate of least-squares migration. We will use three different preconditioners. The first preconditioner is built via a row averaging operator applied to the Hessian. The second preconditioner is designed by exploring the action of the Hessian operator on random models. Finally, we use the filter-based method introduced by Kazemi and Sacchi (2014a) to derive a new class of preconditioners. The filter-based preconditioner, in essence, is close to research proposed by Guitton (2004) and Nammour and Symes (2009). It considers the action of the partial Hessian operators in the context of shot profile least-squares migration as stationary convolutional operators. This follows from the fact that each partial Hessian experiences the spatially invariant portion of the medium under its action. Intuitively speaking, application of these convolutional filters on prestack migrated image will correct amplitudes and increase the resolution of the partial images.

The outline of the chapter is as follows. First, we formulate the extended least-

squares shot profile migration. Next, we introduce the three proposed preconditioners and use them to develop an effective Preconditioned Conjugate Gradient algorithm to increase the convergence rate of least-squares migration. Finally, we evaluate the performances of the techniques (i.e., least-squares migration with and without preconditioners) on a synthetic dataset (Sigsbee2a salt model) and a real dataset from the Gulf of Mexico (Mississippi Canyon, north-central Gulf of Mexico).

4.2 Extended least-squares migration in shot profile domain

It is easy to show that extended shot profile wave equation migration with split step correction operator is linear and can be written in matrix-vector notations. This, in turn, will help us to define the adjoint operator of migration (i.e., demigration operator). These two operators are necessary for communicating from data to model and model to data space in the context of extended shot profile least-squares migration.

To do so, let us consider an extended $2D$ physical model with length $M = n_s n_z n_x$ and a corresponding prestack dataset in frequency domain with length $N = n_s n_\omega n_g$, where n_z is the number of samples of model in depth direction, n_x is the number of samples in horizontal direction, n_s is the total number of shots, n_g is the number of channels per shot and n_ω is the number of frequency realizations in data space. We define the vectorized version of the physical model and the dataset as $\mathbf{m} = [\mathbf{m}_1\ \mathbf{m}_2 \dots \mathbf{m}_{n_s}]^T$ and $\mathbf{d} = [\mathbf{d}_1\ \mathbf{d}_2 \dots \mathbf{d}_{n_s}]^T$, respectively. Then, shot profile wave equation migration with split step correction in matrix-vector notation can be written as follow

$$\mathbf{m} = \mathbf{A}^T\, \mathbf{d}, \tag{4.1}$$

where

$$\mathbf{A}^T = \begin{pmatrix} \mathbf{A}_1 & & & & \\ & \mathbf{A}_2 & & & \\ & & \ddots & & \\ & & & \mathbf{A}_{n_s-1} & \\ & & & & \mathbf{A}_{n_s} \end{pmatrix}^T, \tag{4.2}$$

and

$$\mathbf{A}_j^T = \begin{pmatrix} \mathbf{S}_0^{0,j}\mathbf{P}_0^{0,j} & \mathbf{S}_0^{1,j}\mathbf{P}_0^{1,j} & \cdots & \mathbf{S}_0^{k,j}\mathbf{P}_0^{k,j} & \cdots & \mathbf{S}_0^{n_\omega-1,j}\mathbf{P}_0^{n_\omega-1,j} \\ \mathbf{S}_1^{0,j}\mathbf{P}_1^{0,j}\mathbf{P}_0^{0,j} & \mathbf{S}_1^{1,j}\mathbf{P}_1^{1,j}\mathbf{P}_0^{1,j} & \cdots & \mathbf{S}_1^{k,j}\mathbf{P}_1^{k,j}\mathbf{P}_0^{k,j} & \cdots & \mathbf{S}_1^{n_\omega-1,j}\mathbf{P}_1^{n_\omega-1,j}\mathbf{P}_0^{n_\omega-1,j} \\ \mathbf{S}_2^{0,j}\mathbf{P}_2^{0,j}\mathbf{P}_1^{0,j}\mathbf{P}_0^{0,j} & \mathbf{S}_2^{1,j}\mathbf{P}_2^{1,j}\mathbf{P}_1^{1,j}\mathbf{P}_0^{1,j} & \cdots & \mathbf{S}_2^{k,j}\mathbf{P}_2^{k,j}\mathbf{P}_1^{k,j}\mathbf{P}_0^{k,j} & \cdots & \mathbf{S}_2^{n_\omega-1,j}\mathbf{P}_2^{n_\omega-1,j}\mathbf{P}_1^{n_\omega-1,j}\mathbf{P}_0^{n_\omega-1,j} \\ \vdots & & \ddots & & \vdots & \\ \mathbf{S}_{n_z-2}^{0,j}\prod_{i=0}^{n_z-2}\mathbf{P}_i^{0,j} & \mathbf{S}_{n_z-2}^{1,j}\prod_{i=0}^{n_z-2}\mathbf{P}_i^{1,j} & \cdots & \mathbf{S}_{n_z-2}^{k,j}\prod_{i=0}^{n_z-2}\mathbf{P}_i^{k,j} & \cdots & \mathbf{S}_{n_z-2}^{n_\omega-1,j}\prod_{i=0}^{n_z-2}\mathbf{P}_i^{n_\omega-1,j} \\ \mathbf{S}_{n_z-1}^{0,j}\prod_{i=0}^{n_z-1}\mathbf{P}_i^{0,j} & \mathbf{S}_{n_z-1}^{1,j}\prod_{i=0}^{n_z-1}\mathbf{P}_i^{1,j} & \cdots & \mathbf{S}_{n_z-1}^{k,j}\prod_{i=0}^{n_z-1}\mathbf{P}_i^{k,j} & \cdots & \mathbf{S}_{n_z-1}^{n_\omega-1,j}\prod_{i=0}^{n_z-1}\mathbf{P}_i^{n_\omega-1,j} \end{pmatrix}, \tag{4.3}$$

where $\mathbf{A}^T$ is an M by N matrix which can be interpreted as the migration operator and $\mathbf{P}_i^{k,j}$'s are n_x by n_x square block matrices. To be more precise, the $\mathbf{P}_i^{k,j}$'s matrices are receiver side single depth split step operators where i denotes the depth sample, k is representative of frequency slice and j is the shot number. The field $\mathbf{S}_i^{k,j}$ is the k^{th} monochromatic downward continued source side wavefield with the shot j to the i^{th} depth sample. It is worth mentioning that the crosscorrelation imaging condition is applied in frequency domain by summing over the frequencies (Claerbout, 1971b; Schleicher et al., 2008).

Moreover, the data generating model under the action of shot profile demigration operator $\mathbf{A}$ can be written as follow

$$\mathbf{d} = \mathbf{A}\mathbf{m} + \mathbf{n}, \tag{4.4}$$

where the term $\mathbf{n}$ is the noise and, often, also a term to absorb waves not modelled by the demigration operator. Using the migration operator, which is the adjoint of de-migration operator, we have

$$\mathbf{m}^{adj} = \mathbf{A}^T\mathbf{d}, \tag{4.5}$$

where $\mathbf{m}^{adj}$ is the migrated image. While $\mathbf{m}^{adj}$ can capture main structures of the true geological model $\mathbf{m}$, the produced model does not honour the data. In other words, the application of demigration operator on the migrated image yields a poor data prediction. To tackle the problem, We will minimize the following cost function

$$J = ||\mathbf{A}\mathbf{m} - \mathbf{d}||_2^2 + \mu\,||\mathbf{m}||_2^2, \tag{4.6}$$

where μ is a regularization parameter that balances the importance of the misfit

functional and regularization term. The cost function J is convex and has a closed form solution

$$\mathbf{m}^{LS} = (\mathbf{A}^T\mathbf{A} + \mu\ \mathbf{I})^{-1}\ \mathbf{A}^T\mathbf{d}, \tag{4.7}$$

where $\mathbf{m}^{LS}$ is least-squares inverted model. However, the direct inversion of $(\mathbf{A}^T\mathbf{A} + \mu\ \mathbf{I})$ is not feasible because of of the fact that $\mathbf{A}$ and $\mathbf{A}^T$ are linear operators, not matrices. The cost J can be minimized via the Conjugate Gradient method. In the Conjugate Gradient algorithm, we only need the action of $\mathbf{A}$ and $\mathbf{A}^T$ on vectors. The Conjugate Gradient method is an exact method. However, in practical terms, it is considered an iterative technique, and we will stop iterations whenever the changes in the normal system of equations reaches below of a predefined threshold.

While minimizing the cost J via the Conjugate Gradient method produces reasonable models that honour the recorded data, the computational cost of the Conjugate Gradient algorithm is high. This fact sometimes forces geophysicists to only use the adjoint operator. To evaluate the computational cost of this method interested readers are referred to (Marfurt and Shin, 1989; Collino and Joly, 1995; Mulder and Plessix, 2004a,c; Kazemi and Sacchi, 2015). One can reduce the computational cost of the method of Conjugate Gradients by applying preconditioners that can cluster the eigenvalues of the Hessian operator. In other words, one prefers to solve

$$J = ||\mathbf{A}\mathbf{P}^{-1}\mathbf{y} - \mathbf{d}||_2^2 + \mu\ ||\mathbf{P}^{-1}\mathbf{y}||_2^2, \tag{4.8}$$

where $\mathbf{m} = \mathbf{P}^{-1}\mathbf{y}$, and $\mathbf{P}$ is an $M \times M$ preconditioner. Equation 4.8 can be solved via the Preconditioned Conjugate Gradient algorithm (see Appendix C.1). We will show that by designing and applying a proper preconditioner $\mathbf{P}$, we can increase the convergence rate of the Conjugate Gradient algorithm.

In next section, We will introduce three types of preconditioners for approximating the inverse of the Hessian operator and implement those preconditioners in the context of extended shot profile least-squares migration. The primary motivations behind the proposed methods are the reduction of the computational cost of least-squares shot profile migration and ability to estimate images that honour the recorded data.

4.3 Preconditioners

4.3.1 Diagonal scaling via row averaging the Hessian

In this section we will show a simple procedure that results in finding the row averaged version of the Hessian. In other words, we will try to replace the diagonal of the Hessian with its row averaged values and use it as the diagonal scaling preconditioner necessary for our Preconditioned Conjugate Gradient algorithm. The row averaging of an $M \times M$ matrix can be computed by multiplying the matrix with an $M \times 1$ vector of ones

$$\mathbf{H}^{ra} = diag\ (\mathbf{H}\ \mathbf{1}^T), \tag{4.9}$$

where $diag\ (*)$ operator reshapes the $M \times 1$ vector to $M \times M$ diagonal matrix and $\mathbf{H}^{ra}$ is the row averaged version of the Hessian. To be more specific the entries of $\mathbf{H}^{ra}$ are

$$\mathbf{H}^{ra} = \mathbf{P}^T\mathbf{P} = \begin{pmatrix} \mathbf{a}_1^{ra} & & & & \\ & \mathbf{a}_2^{ra} & & & \\ & & \ddots & & \\ & & & \mathbf{a}_{n_s-1}^{ra} & \\ & & & & \mathbf{a}_{n_s}^{ra} \end{pmatrix} \tag{4.10}$$

where $\mathbf{a}_i^{ra} = diag\ ([\mathbf{A}_i^T\mathbf{A}_i]\ \mathbf{1}^T)$. It is worth mentioning that we never build matrices and everything should be applied *on the fly*. To do so, this preconditioner can be easily calculated via application of the forward and adjoint operators on a model that consists of all ones ($\mathbf{1}^T$).

4.3.2 Diagonal scaling via random trial model

Here, we will attempt to build a preconditioner by computing the action of the Hessian on a $M \times 1$ random test model $\mathbf{r}^T$ with identically distributed samples drawn from a Gaussian distribution

$$\mathbf{H}_{rand}^{ra} = \mathbf{P}^T\mathbf{P} = diag\ (\mathbf{H}\ \mathbf{r}^T). \tag{4.11}$$

To calculate the $\mathbf{H}_{rand}^{ra}$ we need to apply forward and adjoint operators on a random model which is equivalent to the cost of one iteration of the Conjugate Gradient

algorithm.

4.3.3 Filter-based preconditioner

To overcome the computational cost of calculating the inverse of Hessian (i.e., $\mathbf{H} = \mathbf{A}^T\mathbf{A}$), several studies have focused on approximating the action of the Hessian via simple basis functions (Beylkin, 1985; Rakesh, 1988; ten Kroode et al., 1998; Stolk, 2000). By taking advantage of properties of the Hessian, we can represent it as a compressible pseudo-differential operator (i.e., Stolk (2000))

$$\mathbf{H}\ \mathbf{m}(\mathbf{x}) = \sum_{\mathbf{k}} c(\mathbf{x},\mathbf{k})\ \exp\left(j\mathbf{x}.\mathbf{k}\right)\ \hat{\mathbf{m}}(\mathbf{k}), \tag{4.12}$$

where $\mathbf{k}$ is spatial wavenumber vector, $j = \sqrt{-1}$ and $\hat{\mathbf{m}}$ is the Fourier pair of $\mathbf{m}$ and $c(\mathbf{x},\mathbf{k})$ are some coefficients in phase-space domain that control the illumination.

The coefficients are smooth if the acquisition system has a full aperture, densely sampled in time and space and a delta function is implemented as a source wavelet. Moreover, the physical properties of the medium such as slowness should be spatially invariant (e.g., Stolk (2000)). For this idealized situation, Gelius et al. (2002), Sjoeberg et al. (2003) and Lecomte (2008) have shown that the action of the Hessian can be approximated via convolutional operators. In other words, we can ignore the spatial dependency of the coefficients

$$\mathbf{H}\ \mathbf{m}(\mathbf{x}) = \sum_{\mathbf{k}} c(\mathbf{k})\ \exp\left(j\mathbf{x}.\mathbf{k}\right)\ \hat{\mathbf{m}}(\mathbf{k}). \tag{4.13}$$

Equation (4.13) can be interpreted as the convolution of the spatially invariant coefficients with the migrated image. Note that this is only valid for spatially invariant medium (i.e., constant slowness in spatial directions).

In extended shot profile wave equation migration, we need to sort the data in common shot gathers and migrate each shot gather independently. Hence, in least-squares extended shot profile wave equation migration the Hessian can be written

as

$$\mathbf{H} = \mathbf{A}^T\mathbf{A} = \begin{pmatrix} \mathbf{A}_1^T\mathbf{A}_1 & & & & \\ & \mathbf{A}_2^T\mathbf{A}_2 & & & \\ & & \ddots & & \\ & & & \mathbf{A}_{n_s-1}^T\mathbf{A}_{n_s-1} & \\ & & & & \mathbf{A}_{n_s}^T\mathbf{A}_{n_s} \end{pmatrix}. \tag{4.14}$$

To explain how one can approximate the Hessian in equation 4.14, let us consider the i^{th} recorded shot gather (e.g., $\mathbf{d}_i$). The least-squares solution of this dataset is

$$\mathbf{m}_i^{LS} = (\mathbf{A}_i^T\ \mathbf{A}_i)^{-1}\ \mathbf{m}_i^{adj}, \tag{4.15}$$

where $\mathbf{m}_i^{adj} = \mathbf{A}_i^T\ \mathbf{d}_i$ and $\mathbf{H}_i = \mathbf{A}_i^T\ \mathbf{A}_i$ is the i^{th} partial Hessian designed for the i^{th} shot gather. We will calculate the partial Hessians for all of the shot gathers.

Before doing so, it is worth mentioning that the partial Hessian operators corresponding to different shot geometries are not the same in a spatially variant medium. However, each partial Hessian operator contains information from the spatially limited parts of the medium. This, in turn, allows us to use the spatial invariant behaviour of the medium under the action of each partial Hessian operator. Hence, we can implement equation (4.13) and approximate each partial Hessian via 2D convolutional operators.

Let us consider equation (4.15) as a starting point. Equation (4.15) can be rewritten as follows

$$\mathbf{m}_i^{LS} = \mathbf{H}_i\ \mathbf{m}_i^{adj} \simeq \mathbf{F}_i\ \mathbf{m}_i^{adj} \tag{4.16}$$

where $\mathbf{F}_i$ is a 2D convolutional matrix that approximates the Hessian corresponding to the geometry of the i^{th} shot gather. The $\mathbf{F}_i$ filters can be calculated by solving

$$\hat{\mathbf{f}}_i = \underset{\mathbf{f}_i}{\operatorname{argmin}}\ ||\mathbf{M}_i^{adj}\ \mathbf{f}_i - \mathbf{m}_i^{LS}||_2^2, \tag{4.17}$$

where $\mathbf{M}_i^{adj}$ is the 2D convolution matrix corresponding to $\mathbf{m}_i^{adj}$ with a proper size and $\mathbf{f}_i$'s are vectors that can be reshaped to make the appropriate convolutional filters (i.e., $\mathbf{F}_i$'s). After solving equation (4.17) and building the convolutional filters

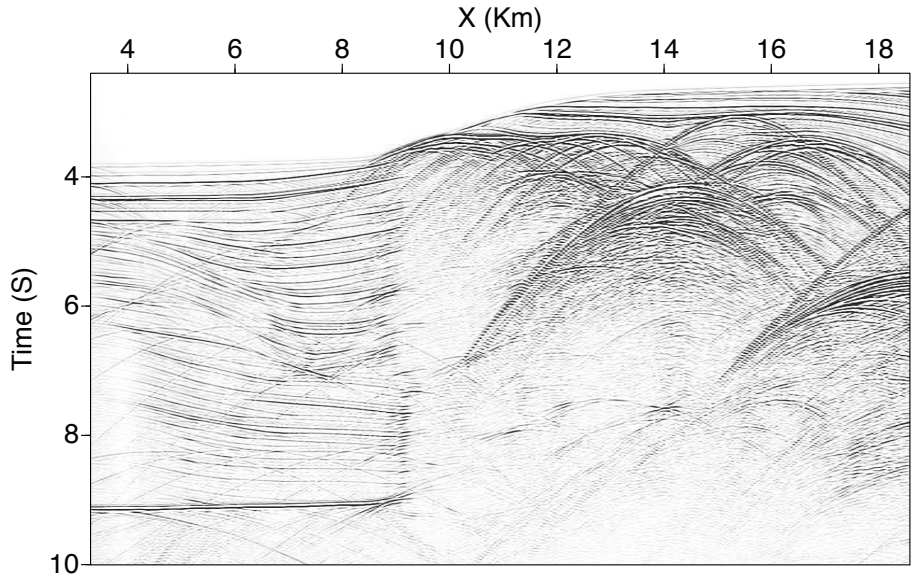

Figure 4.1: True near offset section of Sigsbee2a model.

for all the shots, equation (4.14) can be expressed as

$$\mathbf{H} \simeq \mathbf{P}^T\mathbf{P} = \begin{pmatrix} \mathbf{F}_1 & & & & \\ & \mathbf{F}_2 & & & \\ & & \ddots & & \\ & & & \mathbf{F}_{n_s-1} & \\ & & & & \mathbf{F}_{n_s} \end{pmatrix}, \tag{4.18}$$

where $\mathbf{P}$ is the filter-based preconditioner. Given that the diagonal blocks of $\mathbf{P}$ are built from convolutional matrices, it can be easily inverted in the wavenumber domain via simple Fourier domain deconvolution.

In next section, we will show the efficiency of the proposed preconditioners in reducing the computational cost of extended least-squares migration in shot profile domain using the Sigsbee2a model and real data example from the Gulf of Mexico.

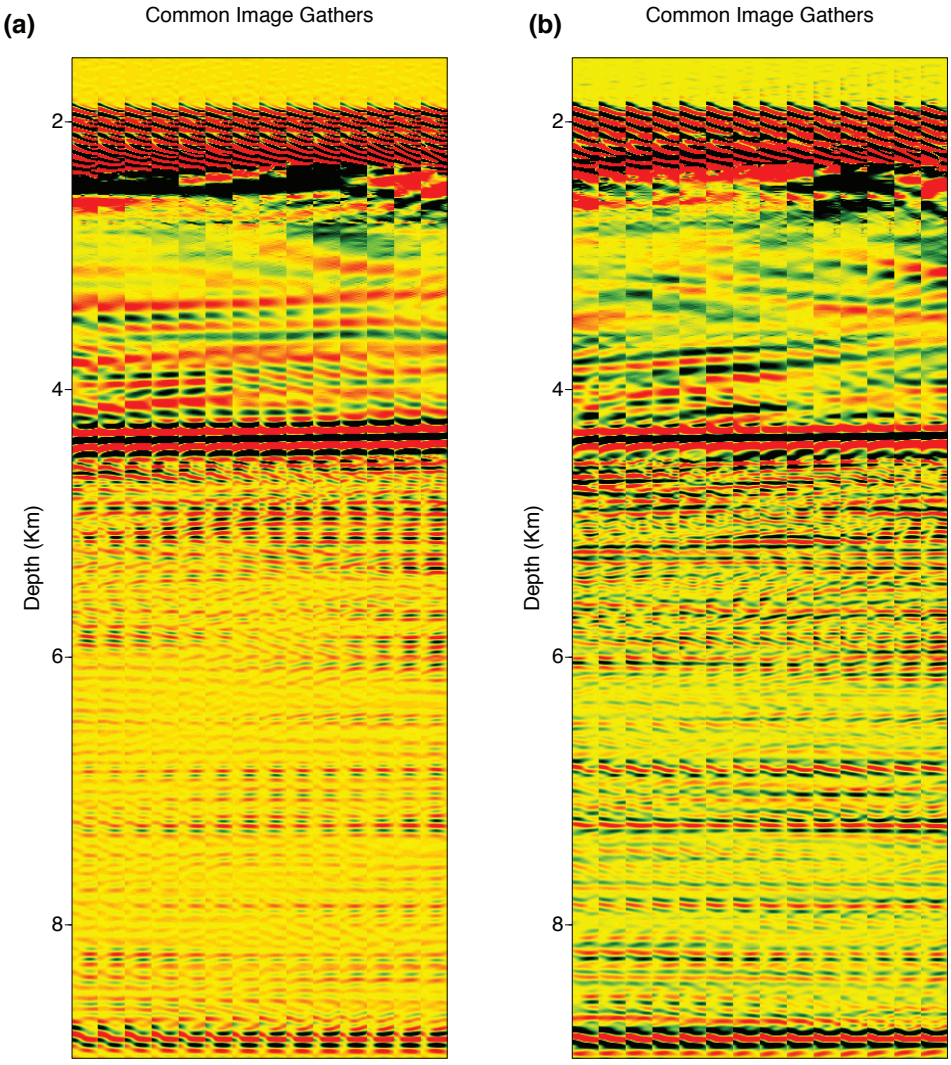

Figure 4.2: Subset of the common image gathers. a) Adjoint. b) Least-squares without preconditioner.

4.4 Synthetic example

To evaluate the performance of preconditioners in the context of extended shot profile least-squares migration we run the algorithms on the part of the Sigsbee2a model and on real data example from the Gulf of Mexico. For Sigsbee2a model, we have 335 sources with a spacing of 45.7 m. The signature of the source is a Ricker wavelet with a dominant frequency of 20 Hz. Sources are moving from right to left of the model and acquisition system is off end. For each shot gather there are 348 receivers with spacing of 22.86 m. Figure 4.1 shows the near offset section of the data. In all of the least-squares results, we choose $\mu = 100$ and run the algorithm for 15 iterations. The output of our migration algorithm is a cube of partial migrated images which each of them belongs to a shot gather (i.e., cube's coordinates are x, z and shot-number). A slice of the cube passing through a fixed horizontal location (i.e., x_0) will be the common image gather at that location. Figure 4.2 depicts the subset of the common image gathers of the adjoint and least-squares migration without preconditioner. The quality of the common image gathers in the case of least-squares migration are better than the quality of the images obtained via the adjoint operator. Moreover, the events are more continuous, and the amplitudes are more balanced comparing to those obtained via the adjoint operator. Finally stacking all of the common image gathers results in final migrated images (Figure 4.3). Figure 4.3 shows that the final migrated image computed via least-squares migration without preconditioner has higher resolution and a more balanced amplitude response in comparison to images computed via the adjoint operator.

Although the adjoint migrated image can capture the main features of the subsurface structure, it does not honour the data. In other words, the application of demigration algorithm on the extended images does not predict the data. Figure 4.4a shows the adjoint predicted near offset section of the Sigsbee2a model and Figure 4.4b is the least-squares predicted near offset section. There are amplitude and resolution problems in the adjoint predicted section comparing to the true near offset section shown in Figure 4.1. On the other hand, the solution computed via least-squares migration can predict data that honours the observation.

As we showed earlier, to honour the recorded data and improve the resolution of common image gathers and the final migrated image, the application of the least-squares migration algorithms is necessary. However, the computational cost of least-

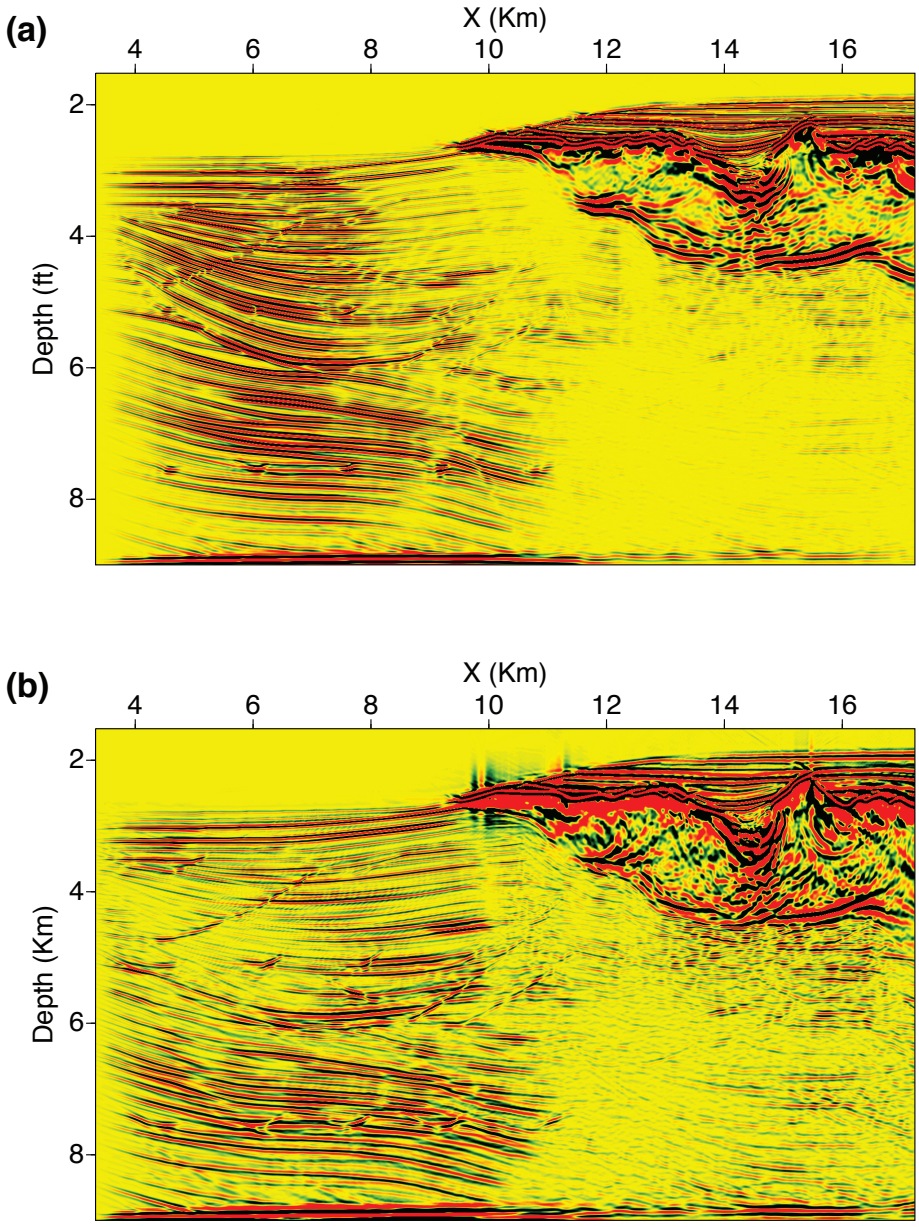

Figure 4.3: Migration and least-squares migration of the Sigsbee2a model after stacking all of the common image gathers. a) Migration or adjoint operator. b) Least-squares migration without preconditioner. The images are representatives of scattering potential/reflectivity. The black colour is for negative reflectivity, red is for positive reflectivity and yellow is for no reflectivity.

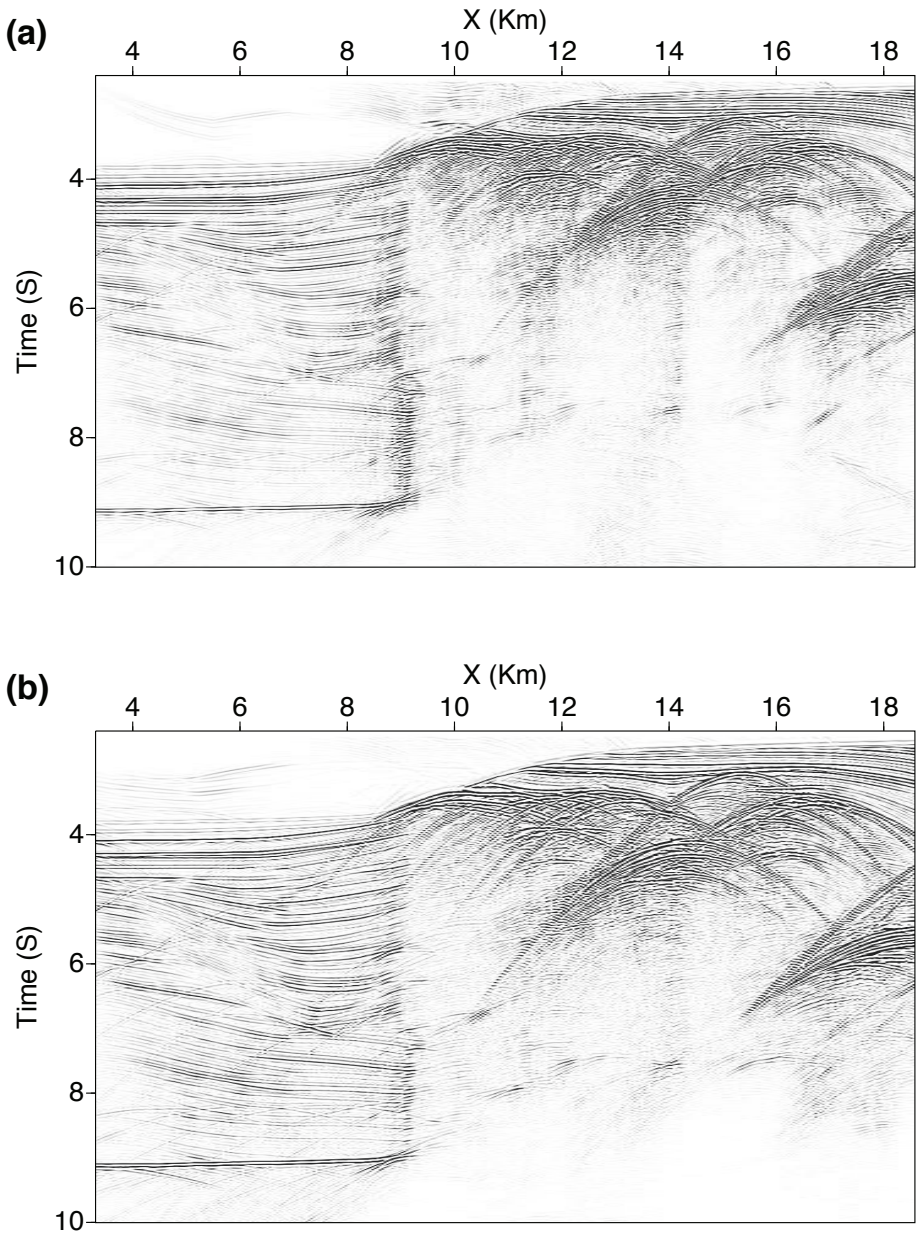

Figure 4.4: Adjoint and least-squares predicted near offset sections of Sigsbee2a model. a) Adjoint. b) Least-squares without preconditioner.

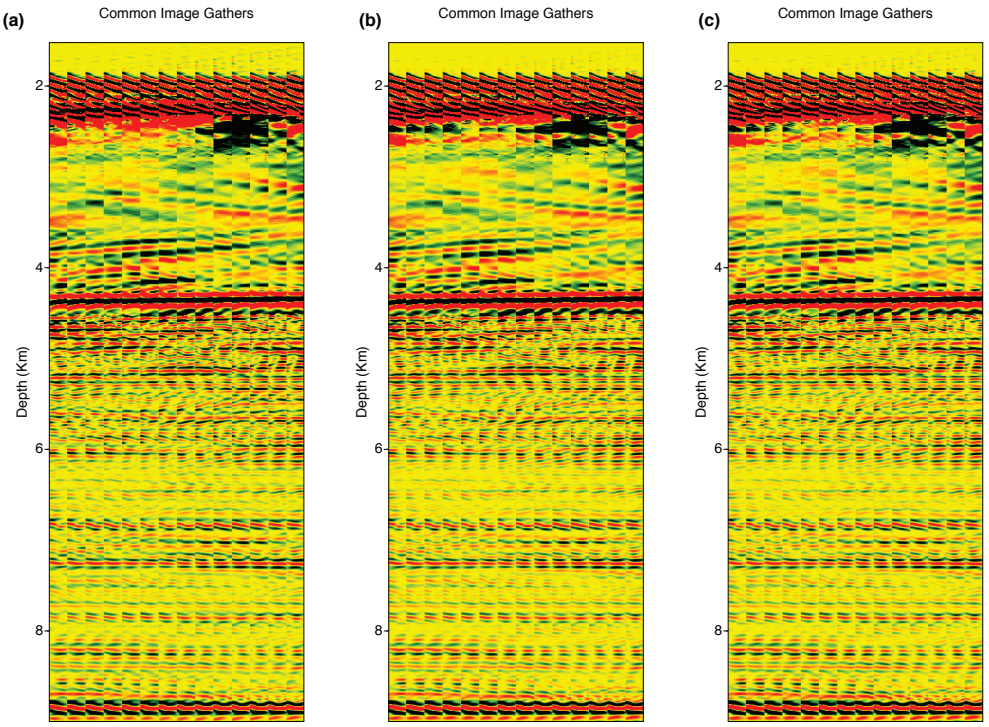

Figure 4.5: Subset of the common image gathers of preconditioned extended least-squares migration on the Sigsbee2a model. a) Least-squares migration with row averaging preconditioner. c) Least-squares with random trail preconditioner. c) Least-squares migration with filter-based preconditioner.

squares migration can overshadow its improvements. To alleviate this problem, we now introduce the afore-described preconditioners. To do so, we implemented the Preconditioned Conjugate Gradient method with the three different preconditioners that were introduced in previous sections. Figure 4.5 compares the common image gathers of preconditioned extended least-squares migration results. The common image gathers are at the same location as the one we showed in Figure 4.2a. Figure 4.5 shows that algorithms improved the amplitude and resolution of events comparing to that of the adjoint. Besides, the results are similar to those obtained via least-squares migration without preconditioning.

However, from the common image gathers we can see that the filter-based method did a slightly better job than the other two preconditioners. Figure 4.6 shows the final migrated images of different preconditioned least-squares algorithms after

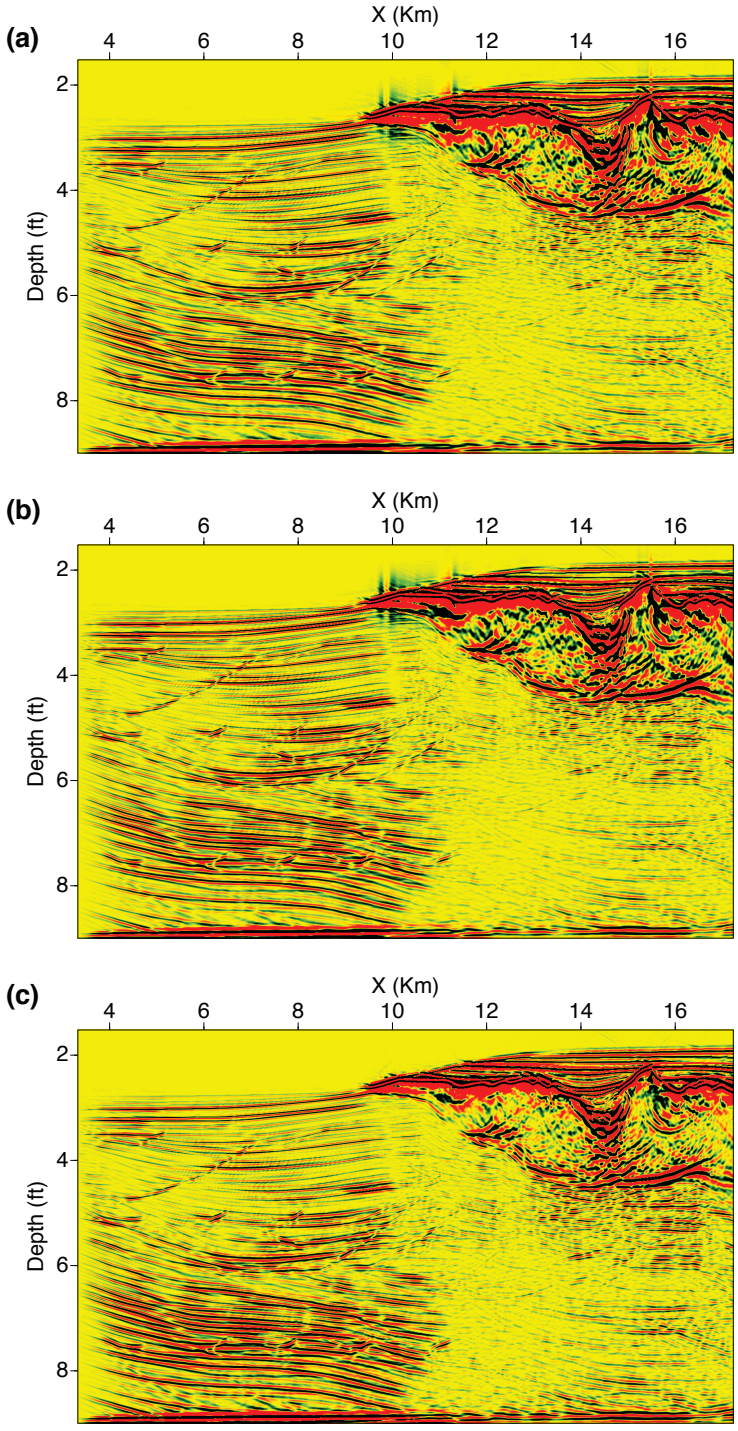

Figure 4.6: Final migrated images of Sigsbee2a model after stacking all of the common image gathers of preconditioned extended least-squares migration. a) Least-squares migration with row averaging preconditioner. b) Least-squares migration with random preconditioner. c) Least-squares migration with filter-based preconditioner.

stacking all of the common images gathers. It is evident from the images that the filter-based algorithm was able to increase the resolution of the migrated image. This is because the filter-based preconditioner approximates the Hessian better than the other two preconditioners. To show the goodness of fit of the algorithms, we forward modelled the extended images and compared results. Figure 4.7 represents the predicted near offset sections of the Sigsbee2a model using preconditioned extended least-squares migration. Comparing to the prediction computed with the adjoint operator, preconditioned least-squares migration has honoured the data with high accuracy.

4.4.1 Figures of merit

We provide three types of measurements to compare the performances of different approaches. The first figure of merit is the normalized cost. Cost functions are normalized to the cost of the first iteration. Figure 4.8 compares the convergence behaviour of the extended least-squares algorithm with and without preconditioning.

The dashed line in Figure 4.8 is the race line. From the race line, we can compare the number of iterations required by different algorithms to give the same performance that one might get after the final iteration of least-squares migration without preconditioning. For example, in the case of least-squares migration with row-averaging preconditioning, the algorithm's convergence beats least-squares migration without preconditioning after 5 iterations. It means that by applying the row-averaged preconditioner we can increase the convergence rate by factor of three. The second figure of merit is a residual panel between the true and predicted near offset sections using different techniques. Figure 4.9 shows the residual panels between the predicted and true near offset sections. As it is clear from the figure, the adjoint method does not honour the data. On the other hand, the least-squares migration solution is able to fit the data. Moreover, Figurer 4.10 depicts the performances of different preconditioning techniques in fitting the data. It can be seen from the figure that all of the preconditoners did a good job in reducing the misfit between the predicted and true data sections. The third figure of merit is the time that takes for each method to complete their tasks. One of the measurements of efficiency of the algorithm is that how fast an algorithm can finish the required computations

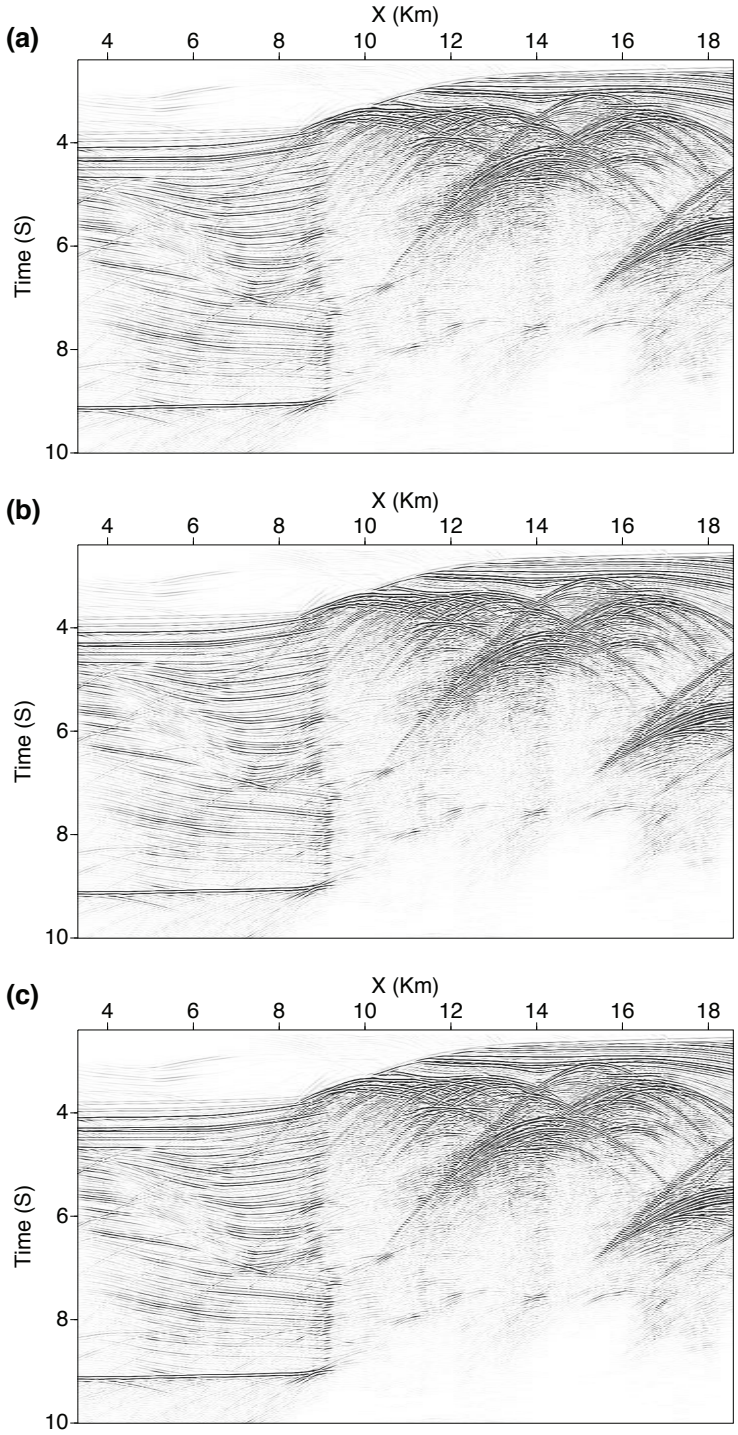

Figure 4.7: Preconditioned extended least-squares predicted near offset sections of Sigsbee2a model. a) Least-squares with row averaging preconditioner. b) Least-squares with random trail preconditioner. c) Least-squares with filter-based preconditioner.

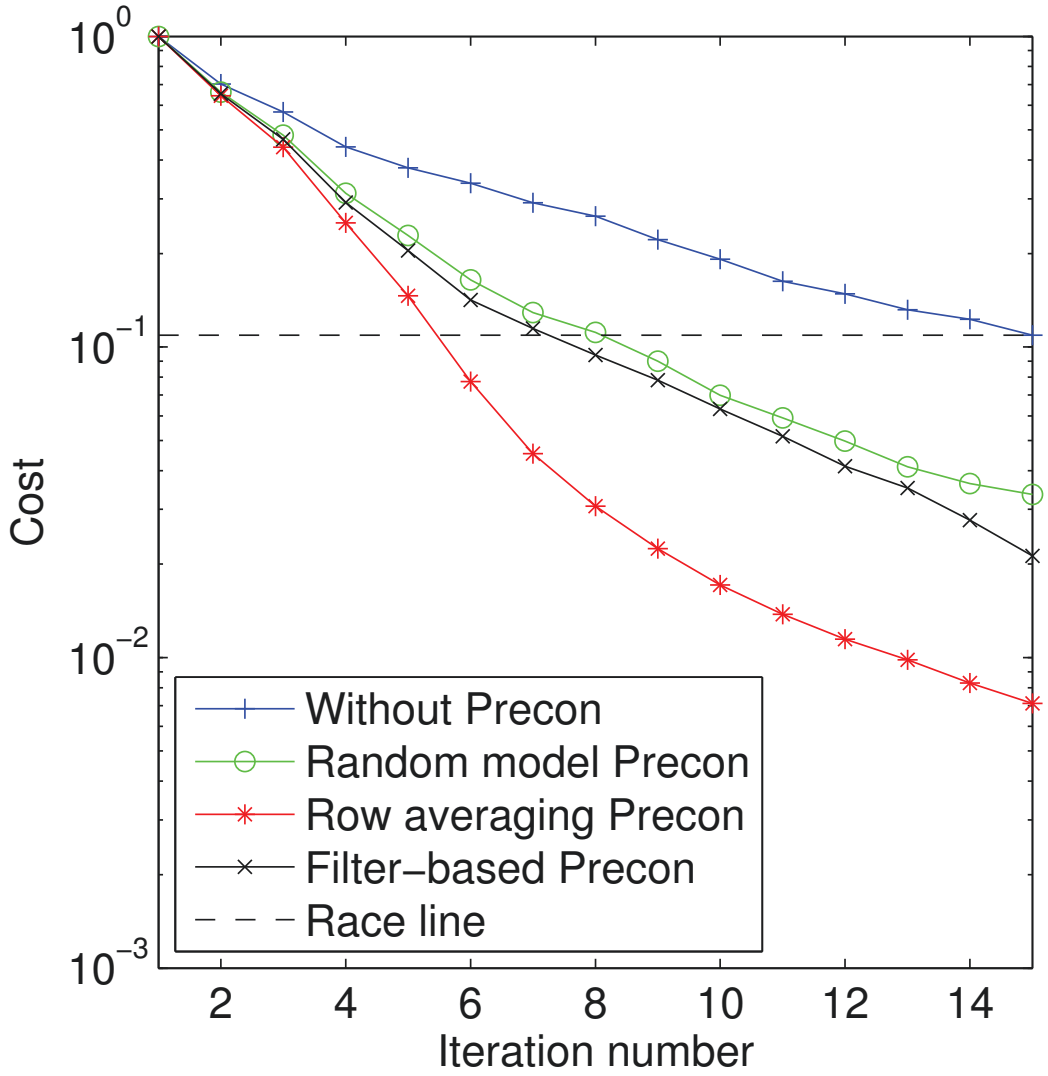

Figure 4.8: Comparison of the convergence behaviour of extended least-squares algorithms with and without preconditioning for the Sigsbee2a model.

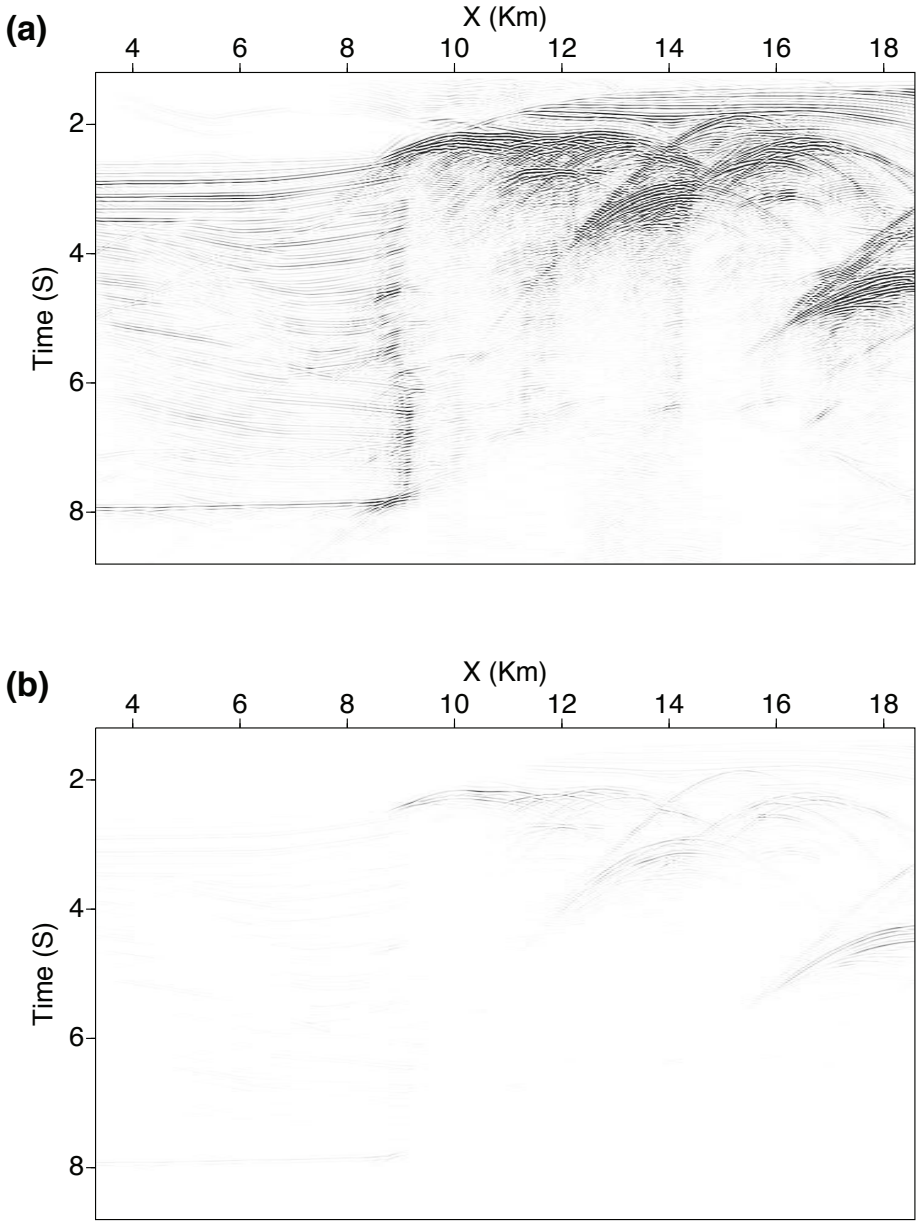

Figure 4.9: Adjoint and least-squares predicted residual panel between the predicted and true near offset sections of the Sigsbee2a model. a) Adjoint. b) Least-squares without preconditioning.

Algorithm	T_{precon} [day]	T_{race} [day]	T_{total} [day]
LS without precon	0	32.7	32.7
LS with row avaraging	2.18	10.9	13.08
LS with random trial	2.18	15.26	17.44
LS with filter-based	4.36	17.44	21.8

Table 4.1: Comparing the computation time for different algorithms with reference to the race line in Figure 4.8 for Sigsbee2a model. All of the times are computed using the same input parameters and only using one core. T_{precon} is the computation time for building the preconditioner, T_{race} is the computation time to reach the race line, and T_{total} is the total computation time.

using the same resources. We tested the algorithms on a machine with an Intel(R) Xeon(R) CPU E5-2650 0 @ 2.00GHz processor and all of the algorithms are codes in C and complied with same compiler. To make a fair comparison all of the input parameters to the programs were kept the same. Table 4.1 shows the time that it takes for different algorithms to reduce the cost to the level defined by our race line in Figure 4.8.

4.5 Real data example

We also applied the algorithms on a 2D marine dataset from the Gulf of Mexico, Mississippi Canyon dataset. Before implementing prestack depth migration, we pre-processed the data. We applied free surface multiple suppression so that we are consistent with the one-way propagation model of our migration operator. We also build the depth velocity model via semblance analysis and time-to-depth conversion. Finally, we estimated the source signature by the $SMBD$ method introduced by Kazemi and Sacchi (2014b). For the Gulf of Mexico dataset, we have 800 sources with a spacing of 26.67 m. Sources are moving from right to left of the model and acquisition system is off end. For each shot gather there are 183 receivers with a spacing of 26.6 m. Figure 4.11 shows the near offset section of the data. In all of the least-squares results, we choose $\mu = 100$ and run the algorithm for 10 iterations.

We will compare the performances of different algorithms and analyze results in terms of resolution and amplitude balance of common image gathers corresponding

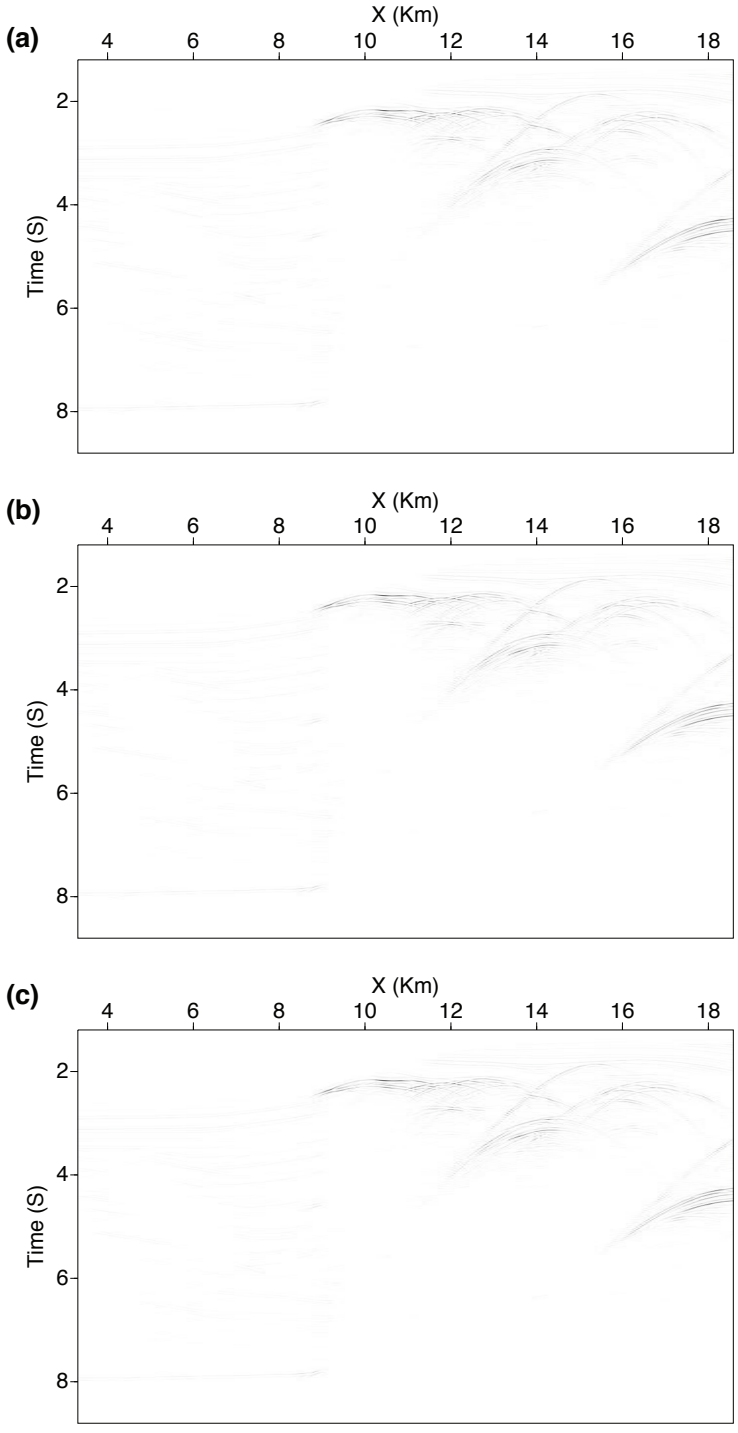

Figure 4.10: Preconditioned extended least-squares residual panel between the predicted and true near offset sections of the Sigsbee2a model. a) Least-squares with row averaging preconditioning. b) Least-squares with random preconditioning. c) Least-squares with filter-based preconditioning.

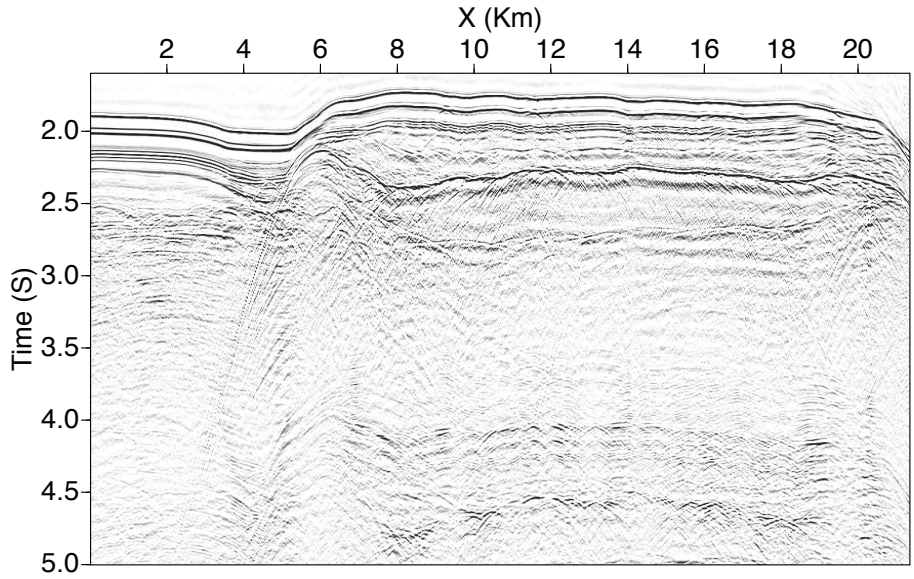

Figure 4.11: True near offset section of Gulf of Mexico data.

to three portions of the dataset. Figures 4.12, 4.13 and 4.14 depict the subset of the common image gathers of the Gulf of Mexico images calculated via adjoint and least-squares migration without preconditioners methods for left, middle and right portions of the model, respectively. The continuity, amplitude and resolution of common image gathers in the case of least-squares migration are better than those obtained via adjoint migration. Stacking all of the common images gathers results in final migrated image of the subsurface. Again, we compare the results of the adjoint technique with that of least-squares migration without preconditioners in Figure 4.15. Comparing Figures 4.15a and b reveals that our least-squares migration algorithm did a good job, comparing to that of adjoint, in improving the resolution and balancing the amplitude of the final migrated image.

To reduce the computational cost of the least-squares migration algorithm, we applied our previously explained preconditioners. Figures 4.17 ,4.18 and 4.19 show the subset of the common image gathers of the Gulf of Mexico migrated images calculated via preconditioned least-squares methods for left, middle and right portions of the model, respectively.

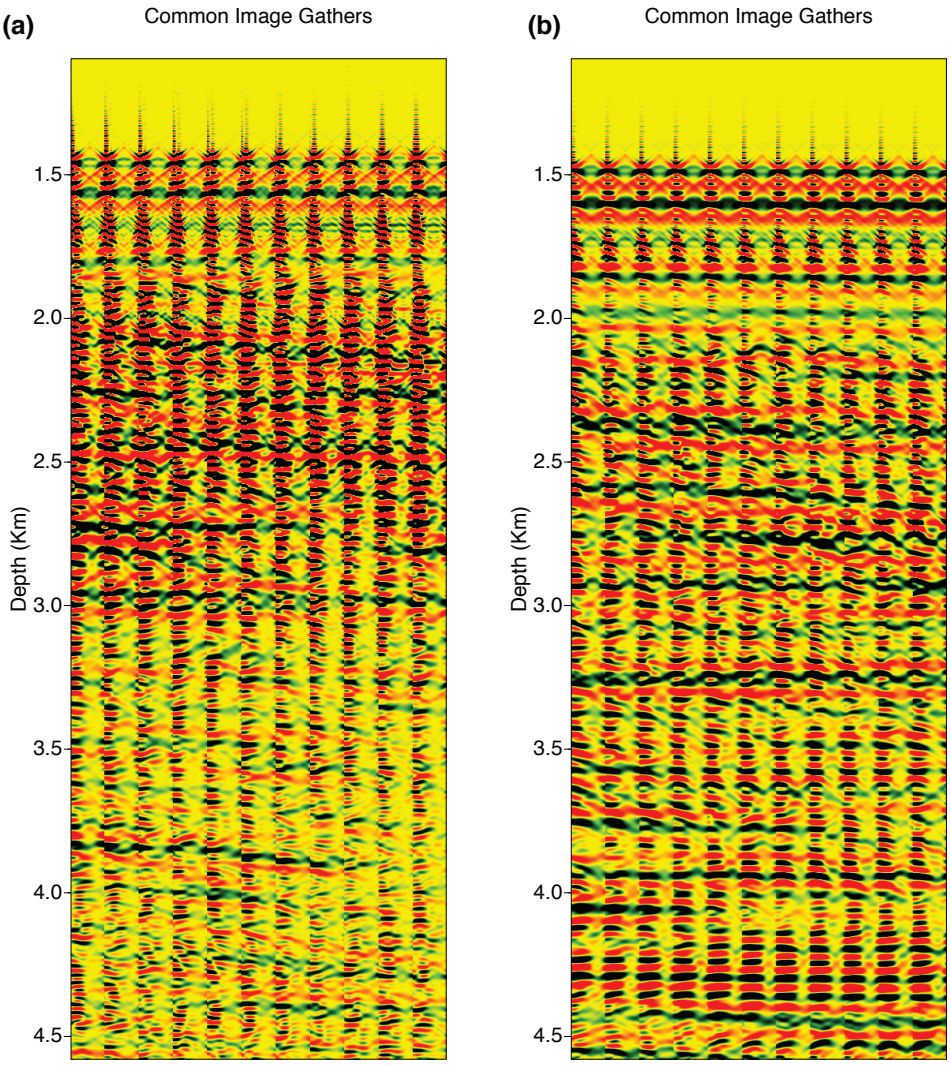

Figure 4.12: Subset of the common image gathers of extended migrated algorithm of Gulf of Mexico dataset corresponding to the left part of the model. a) Adjoint. b) Least-squares without preconditioner.

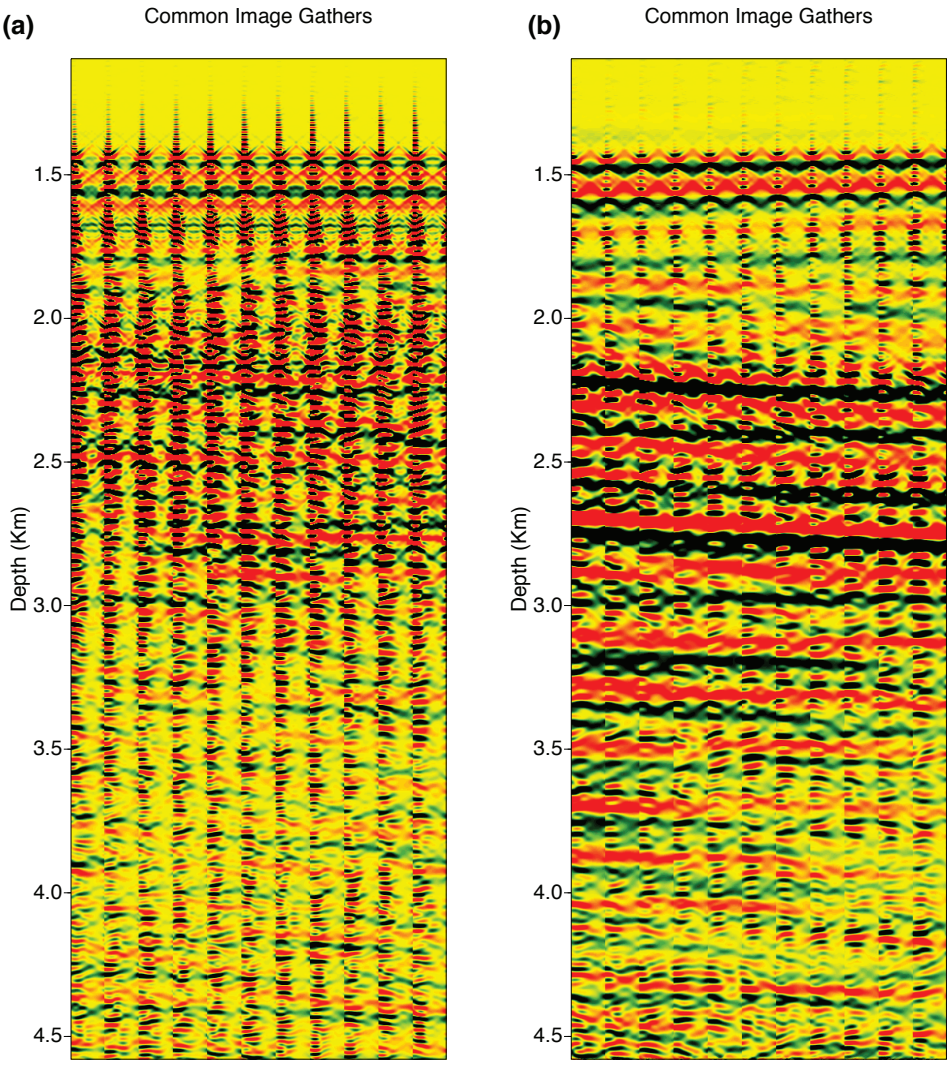

Figure 4.13: Subset of the common image gathers of extended migrated algorithm of Gulf of Mexico dataset corresponding to the middle part of the model. a) Adjoint. b) Least-squares without preconditioner.

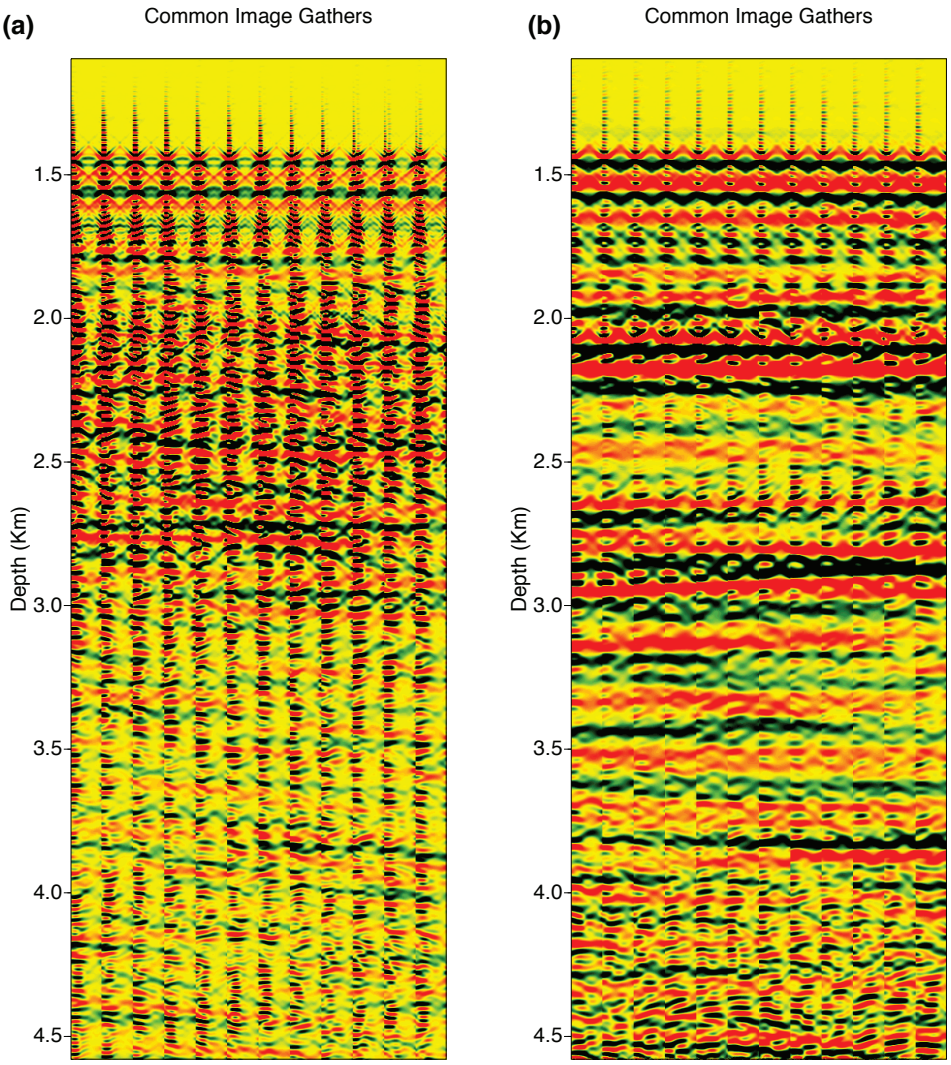

Figure 4.14: Subset of the common image gathers of extended migrated algorithm of Gulf of Mexico dataset corresponding to the right part of the model. a) Adjoint. b) Least-squares without preconditioner.

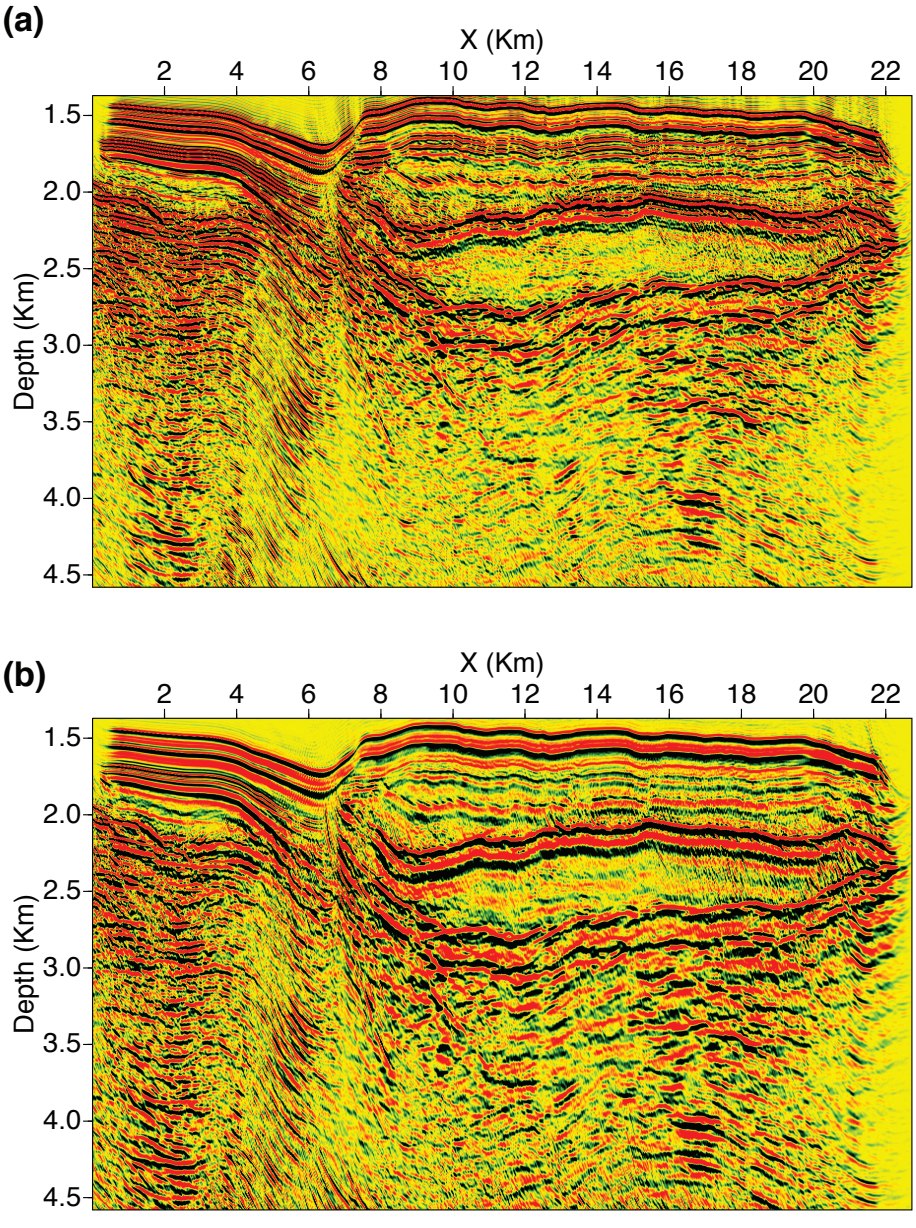

Figure 4.15: Adjoint and least squares migrated images of Gulf of Mexico dataset after stacking all of the common image gathers. a) Adjoint. b) Least-squares without preconditioner.

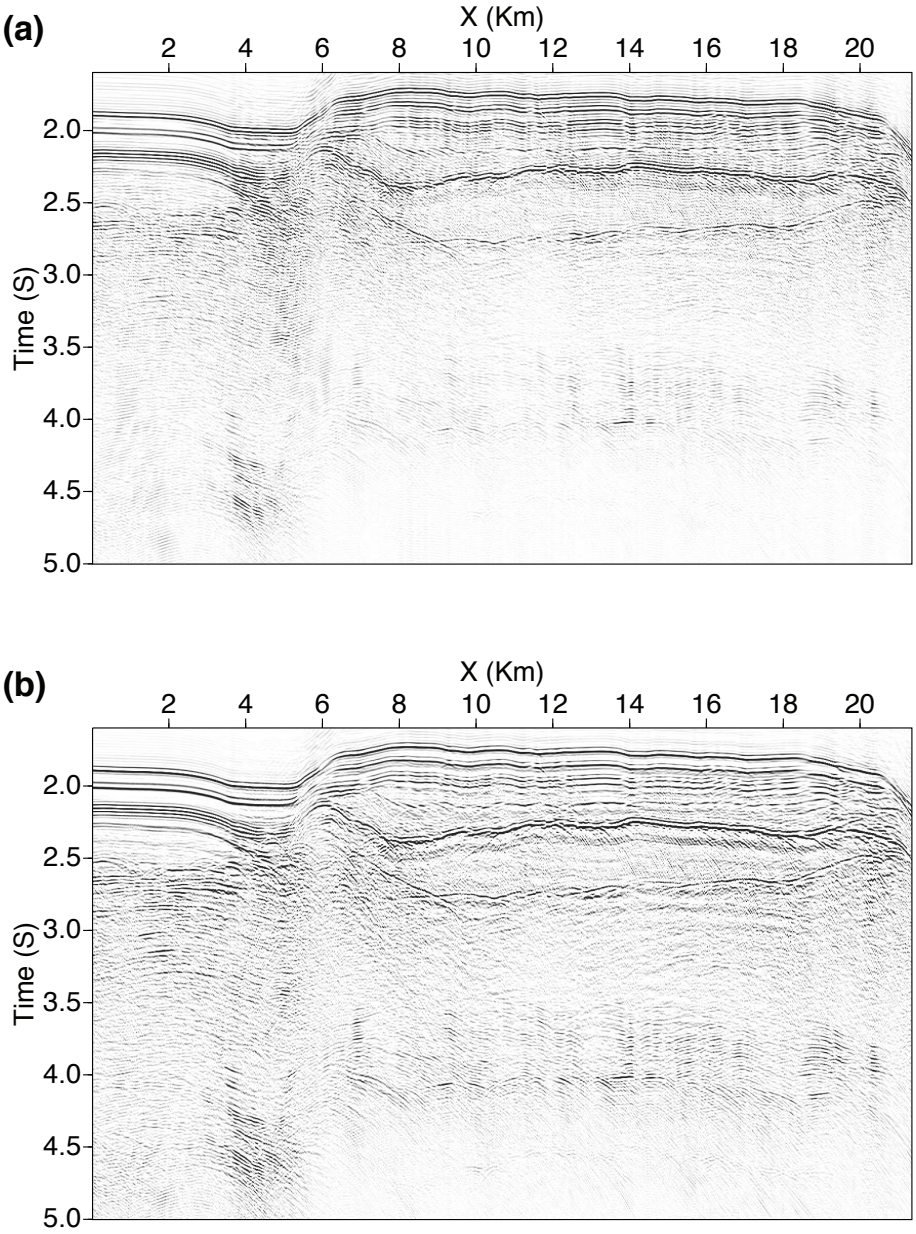

Figure 4.16: Adjoint migration and least-squares migration predicted near offset sections of Gulf of Mexico data. a) Adjoint. b) Least-squares without preconditioner.

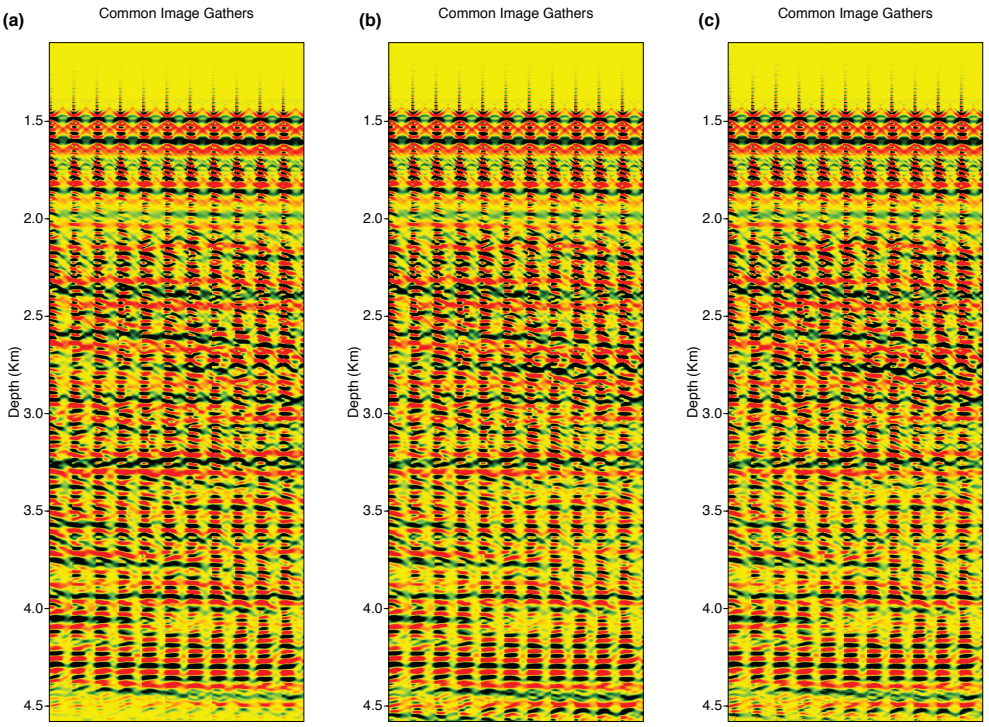

Figure 4.17: Subset of the common image gathers of preconditioned extended least-squares migration applied to a Gulf of Mexico dataset. The gathers correspond to the left part of the model. a) Least-squares migration with row averaging preconditioner. b) Least-squares migration with random preconditioner. c) Least-squares migration with filter-based preconditioner.

Although the results of least-squares migration with and without preconditioners are similar overall the resolution of filter-based images are slightly better than those obtained via least-squares migration without preconditioning. This difference is also noticeable in the final migrated images (Figure 4.20). Figure 4.21 represents the predicted near offset sections of Gulf of Mexico dataset using preconditioned extended least-squares migration. As it is clear from the figure, preconditioned least-squares algorithms honoured the data better than the adjoint operator and the predicted near offset section show higher correlation with true section shown in Figure 4.11.

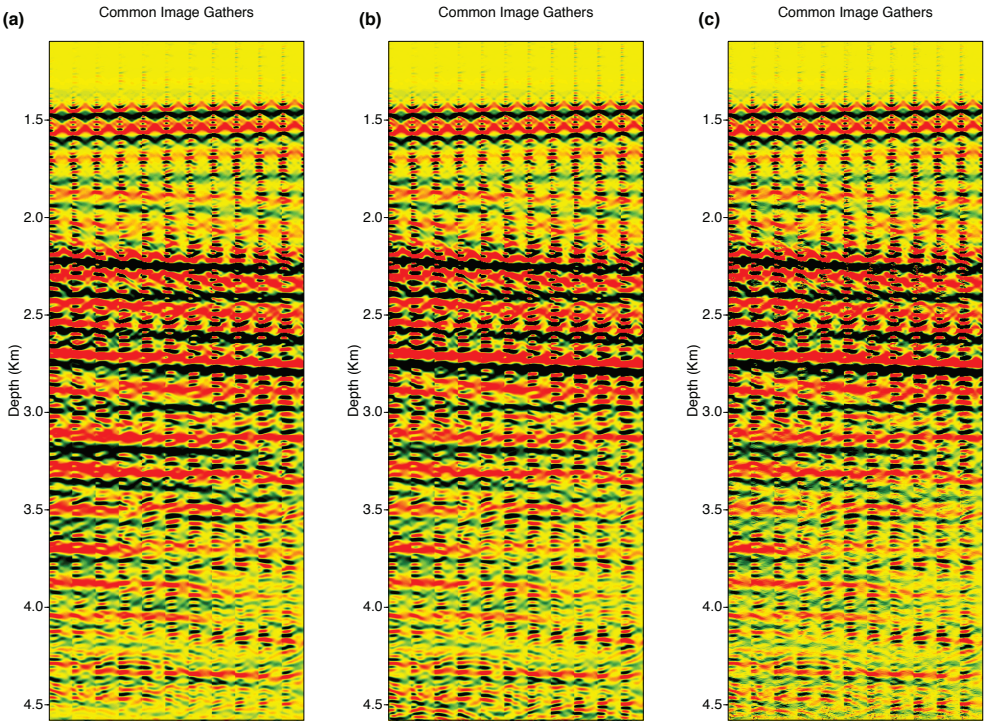

Figure 4.18: Subset of the common image gathers of preconditioned extended least-squares migrated algorithms on Gulf of Mexico data corresponding to the middle part of the model. a) Least-squares with row averaging preconditioner. b) Least-squares with random trail preconditioner. c) Least-squares with filter-based preconditioner.

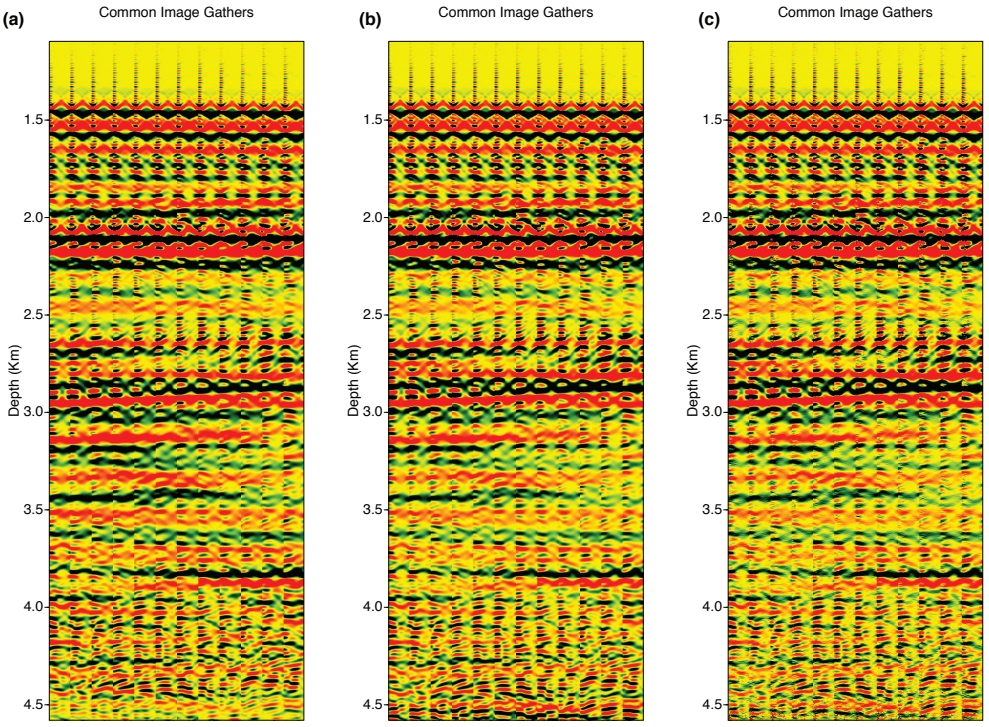

Figure 4.19: Subset of the common image gathers of preconditioned extended least-squares migration. Gulf of Mexico data corresponding to the right part of the model. a) Least-squares migration with row averaging preconditioner. b) Least-squares migration with random model preconditioner. c) Least-squares migration with filter-based preconditioner.

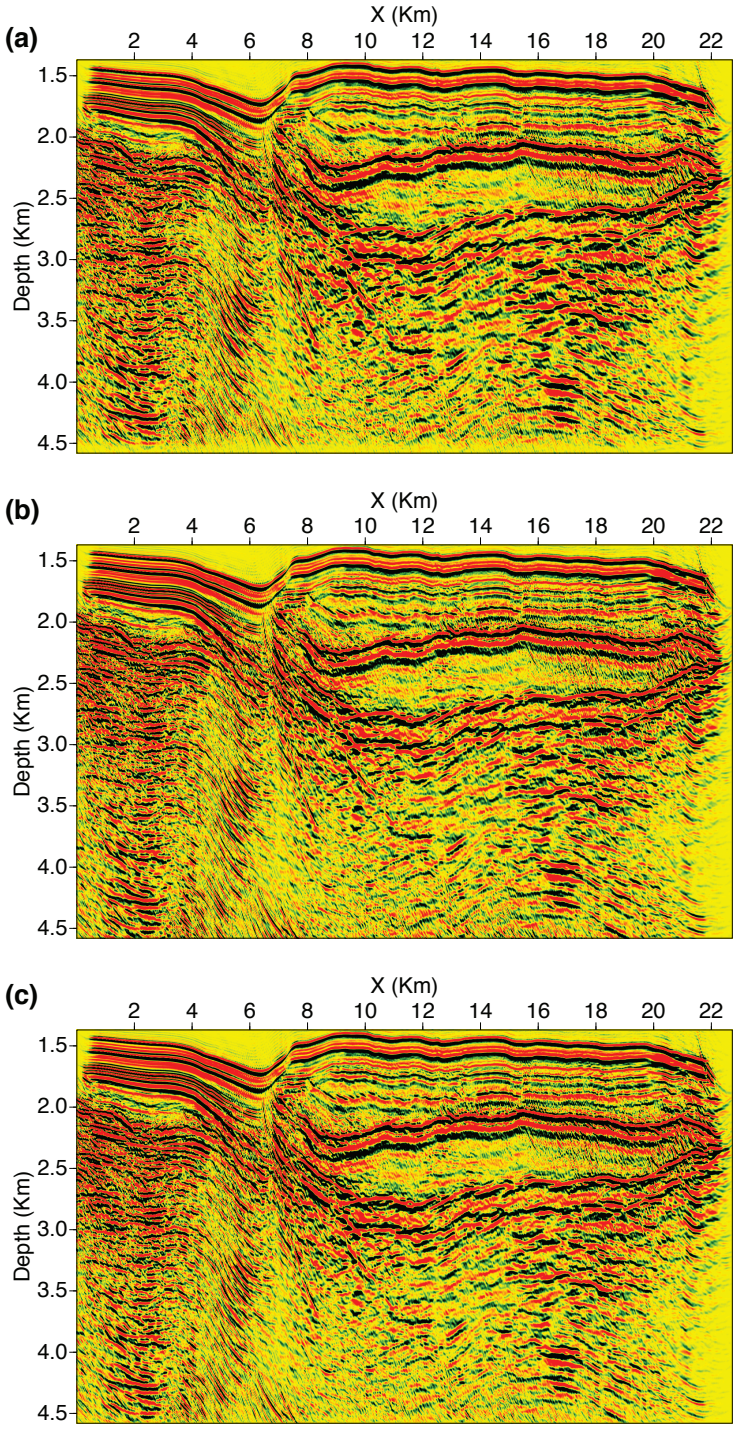

Figure 4.20: Final migrated images of Gulf of Mexico data after stacking all of the common image gathers of preconditioned extended least-squares migrated algorithm. a) Least- squares migration with row averaging preconditioning. b) Least-squares migration with random model preconditioning. c) Least-squares migration with filter-based preconditioning.

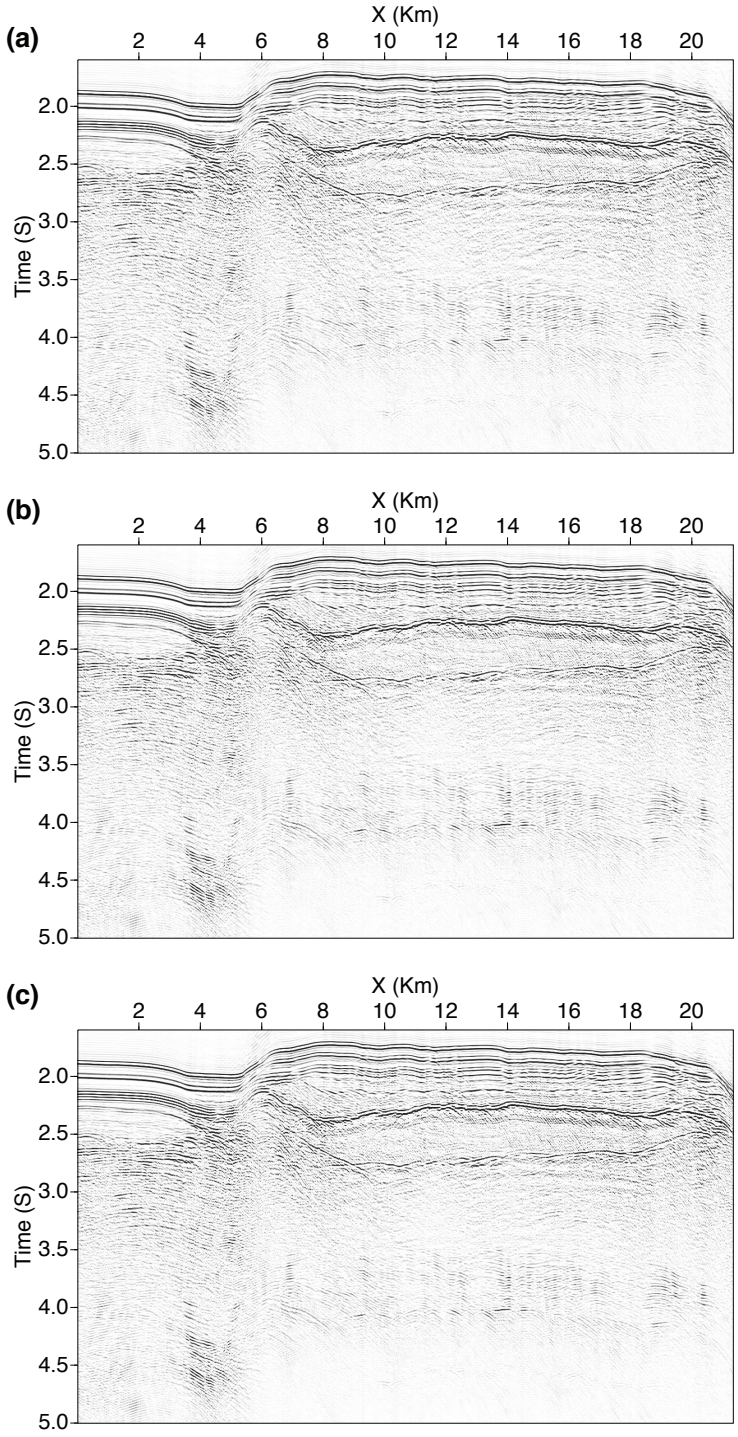

Figure 4.21: Preconditioned extended least-squares predicted near offset sections of Gulf of Mexico data. a) Least-squares with row averaging preconditioner. b) Least-squares with random trail preconditioner. c) Least-squares with filter-based preconditioner.

Algorithm	T_{precon} [day]	T_{race} [day]	T_{total} [day]
LS without precon	0	23.92	23.92
LS with row avaraging	2.39	9.57	11.96
LS with random trial	2.39	11.96	14.35
LS with filter-based	4.78	16.74	21.52

Table 4.2: Comparing the computation time for different algorithms with reference to the race line in Figure 4.22 for Gulf of Mexico data set. All of the times are computed using the same input parameters and only using one core. T_{precon} is the computation time for building the preconditioning operator, T_{race} is the computation time to reach the race line, and T_{total} is the total computation time.

4.5.1 Figures of merit

Again, we provide the aforementioned figures of merit. The first figure of merit is normalized cost. Figure 4.22 compares the convergence behaviour of the extended least-squares algorithm with and without preconditioning. The race line marks the number of iterations that are needed to surpass the performance of the least-squares migration without preconditioning. The second figure of merit is a residual panel between the true and predicted near offset sections using different techniques.

Figure 4.23 shows the residual panels between the predicted and true near offset sections. As it is clear from the figure, the adjoint method does not honour the data. On the other hand, least-squares approach is able to fit the data. Moreover, Figure 4.24 depicts the performances of different preconditioned approaches in fitting the data. It can be seen from the figure that all of the preconditioners did a good job in reducing the misfit between the predicted and true data sections. The third figure of merit is the time that takes for each method to complete their tasks. One of the measurements of efficiency of the algorithm is that how fast an algorithm can finish the required computations using the same resources. We tested the algorithms on a machine with an Intel(R) Xeon(R) CPU E5-2650 0 @ 2.00GHz processor. All of the algorithms are codes in C and compiled with same compiler. To have fair comparisons all of the input parameters to the programs were kept the same. Table 4.2 shows the time that it takes for different algorithms to reduce the cost to the level defined by our race line in Figure 4.22.

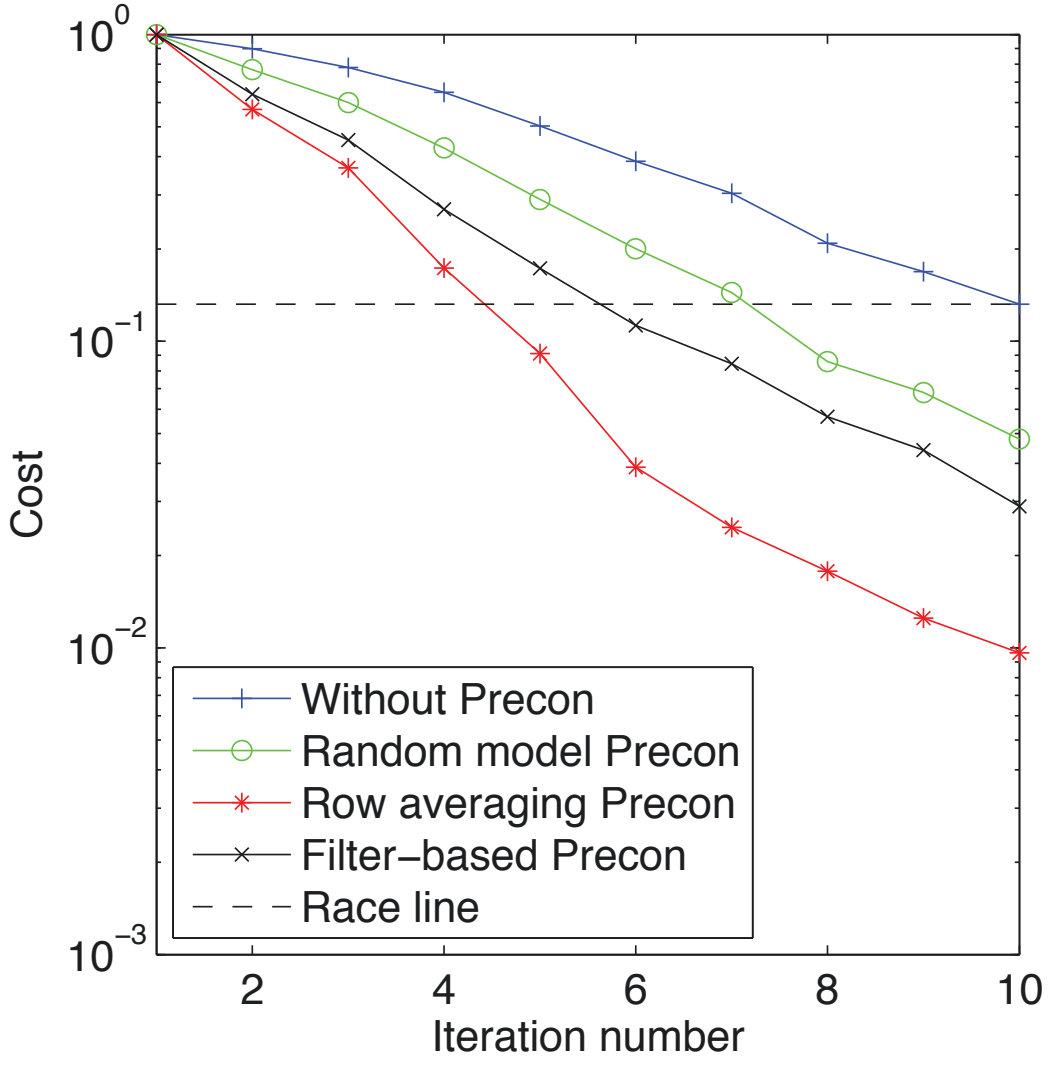

Figure 4.22: Comparison of the convergence behaviour of extended least-squares algorithms migration with and without preconditioning for the Gulf of Mexico dataset.

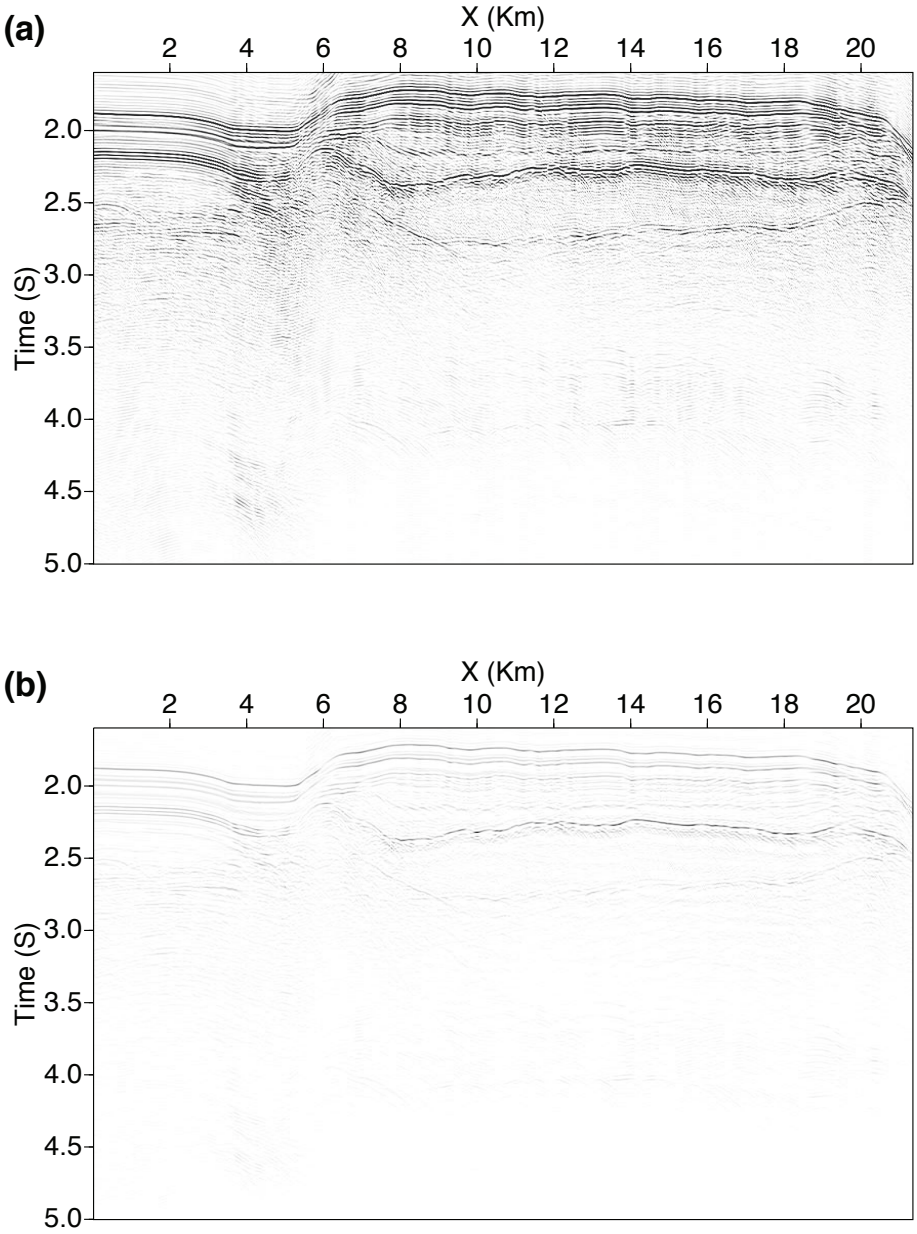

Figure 4.23: Adjoint migration and least-squares migration residual panel between the predicted and true near offset sections of Gulf of Mexico data. a) Adjoint. b) Least-squares without preconditioning.

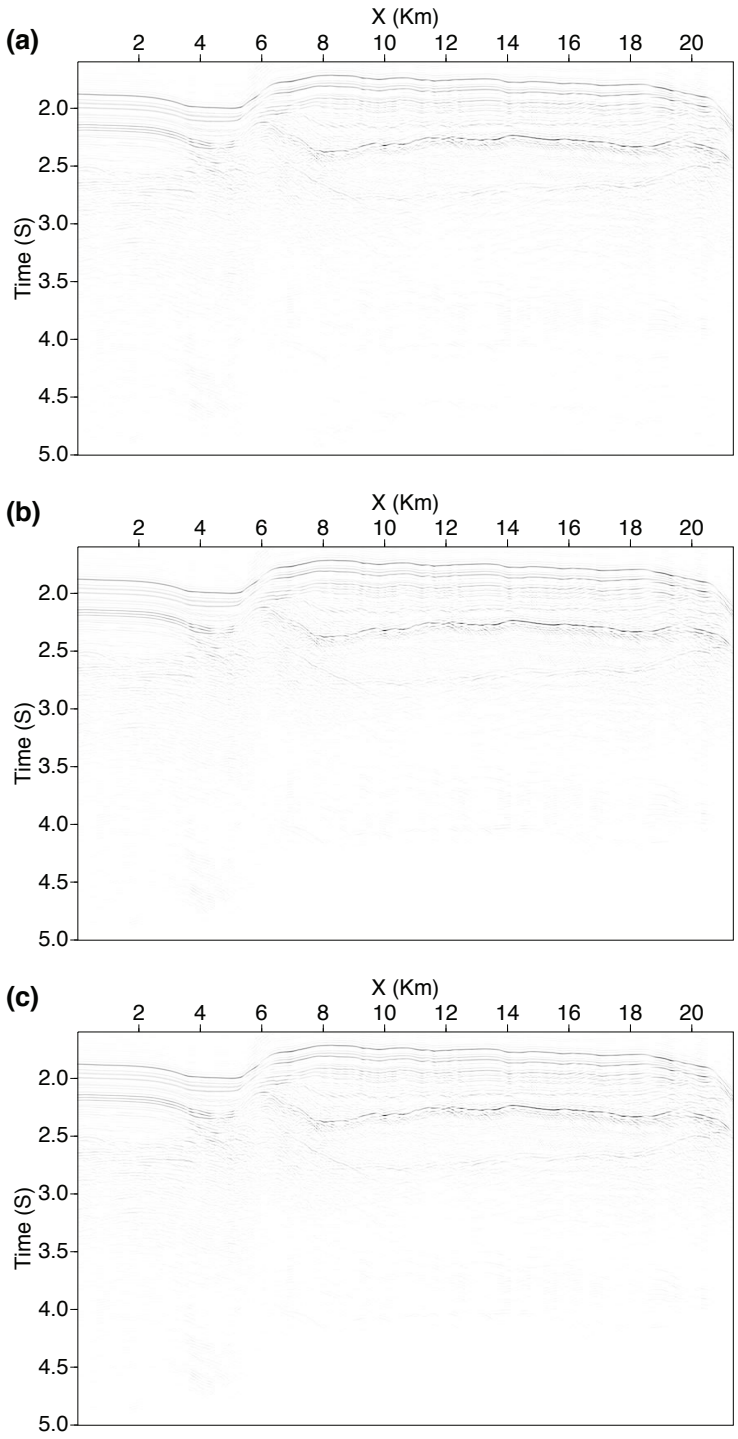

Figure 4.24: Preconditioned extended least-squares residual panel between the predicted and true near offset sections of Gulf of Mexico data. a) Least-squares with row averaging preconditioner. b) Least-squares with random preconditioning. c) Least-squares with filter-based preconditioning.

4.6 Summary

We studied the application of preconditioners in the context of extended least-squares migration in shot profile domain. Three different preconditioners were introduced. Row-averaging and random model preconditioners belong to the category of diagonal scaling preconditioners. The third preconditioner, named filter-based, approximates the Hessian via 2D convolutional filters. In other words, we consider the action of partial Hessian operators in the context of shot profile least-squares migration as stationary convolutional operators. This follows from the fact that each partial Hessian experiences the spatially invariant portion of the medium under its action. It is shown that the application of preconditioners can dramatically reduce the computational cost of least-squares migration algorithms. We tested the algorithms on the Sigsbee2a synthetic model and a Gulf of Mexico real marine data example. In both cases the preconditioned least-squares algorithm was effective in increasing the convergence rate of the conjugate gradient method and in the case of the filter-based preconditioner, we could see cleaner results with higher resolution than the results obtained via least-squares migration without preconditioning.

CHAPTER 5

Source signature estimation[1]

5.1 Introduction

Deconvolution is an important and recurrent topic in seismic data processing. Many signals and images can be represented via the convolution of an unknown signal of interest and a blurring kernel. In general, the process that permits to remove the effects of the blurring kernel on the observed signal or image is called deconvolution. When both the signal of interest and the blurring kernel are unknown, the aforementioned process is denominated blind deconvolution (Shalvi and Weinstein, 1990). In seismic data processing, the signal of interest is the impulse response of the earth and the blurring kernel is the seismic wavelet. In general, the seismic wavelet is considered unknown and therefore, exploration seismologists are often faced with a typical blind deconvolution problem (Ulrych et al., 1995). We usually use convolutional models and statistical deconvolution methods. The difficulty with statistical deconvolution is that we are essentially trying to solve one equation and two unknowns (Ziolkowski, 1991). To solve the problem, we need to add more information and assumptions. Accordingly, the validity of wavelet estimation depends on the validity of our assumptions.

In seismic data processing, deconvolution is part of early efforts to enhance resolution of seismic data and an important component of the transition from analog

[1] A version of this chapter has been published in the journal of Geophysics. Nasser Kazemi and Mauricio D. Sacchi (2014). Sparse multichannel blind deconvolution. Geophysics, 79(5), V143-V152.

analysis of seismic data to digital seismic data processing (Robinson, 1967). Early work using linear prediction theory solves the seismic blind deconvolution problem by making two fundamental assumptions: the earth's reflectivity series (impulse response of the earth) is a white sequence and the seismic wavelet is a minimum phase sequence (Robinson and Treitel, 1964). These two assumptions permit to estimate a causal and stable inverse filter that is applied to the data to estimate the impulse response of the earth. Many deconvolution methods have been proposed to overcome the minimum phase assumption. Two early attempts are homomorphic deconvolution based on the work by Oppenheim and Schafer (1968) and Oppenheim et al. (1976) and implemented for the first time in exploration seismology by Ulrych (1971). A comprehensive theoretical and practical study of the application of homomorphic systems to deconvolution and suppression of air gun reverberations is provided in Stoffa et al. (1974) and Buhl et al. (1974). Similarly, practical methods to implement homomorphic blind deconvolution on real data have been proposed by Otis and Smith (1977) and, more recently, by Herrera and van der Baan (2012). The Minimum Entropy Deconvolution (MED) algorithm is another method that avoids the minimum phase assumption (Wiggins, 1978). MED assumes that the reflectivity is a sparse sequence. The MED algorithm estimates a non-minimum phase filter by maximizing a measure of sparsity of the seismic trace (Donoho, 1981). The measure of sparsity is the varimax norm that is also equivalent to an estimate of kurtosis (White, 1988; Longbottom et al., 1988). The maximization of the varimax norm, or the equivalent kurtosis, is analogous to minimization of a measure of entropy. The latter is coincident with Wiggins's interpretation of minimum disorder or minimum entropy as a synonym of sparsity (Sacchi et al., 1994). The convolution of the estimated MED inverse filter with the seismogram yields the reflectivity and the inverse of the MED filter is an estimator of the seismic source wavelet. Practical algorithms inspired in the MED method have been proposed for seismic data dephasing. Maximum kurtosis phase correction, for instance, can be utilized to find a phase rotation term that maximizes the kurtosis of the seismic trace (Levy and Oldenburg, 1987). It can be shown that, under ideal conditions, the phase rotation that maximizes the kurtosis of the seismic trace also de-phases the seismic wavelet (Longbottom et al., 1988; Cambois and Hargreaves, 1994).

The homomorphic deconvolution, MED and maximum kurtosis phase estimation methods suffer from a variety of shortcomings. For instance, homomorphic de-

convolution is inclined to instability due to phase unwrapping and by its inherent inability to incorporate an additive noise term into its formulation. MED deconvolution often tends to annihilate small reflection coefficients (Ooe and Ulrych, 1979; Walden, 1985) and can become unstable in the presence of noise and highly sensitive to operator length (Nickerson et al., 1986). Maximum kurtosis phase estimation is sensitive to the bandwidth of the wavelet (White, 1988; Xu et al., 2012).

The history of seismic deconvolution is populated by interesting statistical methods for blind deconvolution. These methods, however, often only work under ideal signal conditions. For instance, an important excitement was generated by methods based on fourth-order cumulant matching (Hargreaves, 1994; Velis and Ulrych, 1996) and homomorphic deconvolution via forth-order cumulants (Sacchi et al., 1996). Fourth order cumulants are attractive because they can be computed from the seismic trace and they do preserve the phase of the wavelet when the reflectivity is a sparse sequence with non-vanishing kurtosis (Sacchi and Ulrych, 2000). However, the conditions for robust wavelet estimation required by cumulant-based methods are not often satisfied by real seismic data (Stogioglou et al., 1996).

Last, we also mention parametric methods based on maximum likelihood estimation. These methods attack the blind deconvolution via a two-stage procedure. First the wavelet is assumed know and the reflectivity is estimated by maximizing likelihood. Then, the reflectivity is fixed and the likelihood function is maximized with respect to the wavelet (Mendel, 1983; Kaaresen and Taxt, 1998; Canadas, 2002).

In this chapter we studied a multichannel blind deconvolution algorithm often called Euclid's deconvolution. The method was first discussed in the geophysical literature by Rietsch (1997a) and tested with real data examples in Rietsch (1997b). The method has been previously investigated by Xu et al. (1995) for blind channel estimation in communication systems. The method has also been utilized to improve speech recognition by Liu and Malvar (2001). The idea can be summarized as finding common factors of the Z-transform of the source wavelet embedded in a group of seismograms with different reflectivity sequences. The problem leads to the estimation of the multichannel seismic reflectivity via the solution of homogeneous system of equations (Rietsch, 1997a; Mazzucchelli and Spagnolini, 2001). In the ideal case, the eigenvector associated to the minimum non-zero eigenvalue of the homogenous system of equations is an estimator of the multichannel reflectivity. However, small level of noise impinges on the correct identification of the eigenvector associated to

the impulse response of the earth. This problem is examined in detail by Rietsch (1997a,b).

A variant of Euclid deconvolution was recently utilized by Royer et al. (2012) to separate source signatures from propagation effects in teleseismic data. Similar approaches to Euclid deconvolution were also proposed for image restoration (Harikumar and Bresler, 1999; Sroubek and Milanfar, 2012; Harikumar and Bresler, 1999).

Our main contribution is an improvement to Euclid deconvolution to make it applicable to real data processing. The proposed method can tolerate moderate amounts of noise and does not require a priori knowledge of the length of the seismic wavelet. In the proposed method, the homogeneous system of equation is satisfied by a sparse solution (sparse impulse responses). In other words, we are assuming a reflectivity sequence that is sparse. The problem leads to a non-quadratic minimizing technique where the solution must be constrained to lie on the unit sphere. We discuss a steepest descent method that permits to obtain accurate estimates of the seismic reflectivity and wavelet in the presence of a moderate amount of noise.

5.2 Theory

5.2.1 Multichannel Blind Deconvolution

The earth's impulse response can be modelled as a linear system (Robinson and Treitel, 1980). The input-output relationship for this system, assuming a stationary source wavelet and no noise, can be written as follows

$$d_j[n] = \sum_k w[n-k] r_j[k]\,, \quad j = 1 \dots J \tag{5.1}$$

where the multi-channel seismic data is given by $\mathbf{d}_j = (d_j[1], d_j[2], \dots, d_j[N])^T$. Similarly, the impulse response for channel j is given by $\mathbf{r}_j = (r_j[1], r_j[2], \dots, r_j[M])^T$, and the seismic source wavelet via the vector $\mathbf{w} = (w[1], w[2], \dots, w[L])^T$. We stress that $N = M+L-1$. We also remind the readers that convolution can be represented via the Z-transform as follows

$$D_j(z) = W(z)\, R_j(z)\,, \quad j = 1 \dots J\,. \tag{5.2}$$

By virtue of equation (5.2), it is easy to show that

$$D_p(z)R_q(z) - D_q(z)R_p(z) = 0\,, \quad \forall\ p, q\,. \tag{5.3}$$

The latter can be rewritten in matrix-vector form as follows

$$\mathbf{D}_p\ \mathbf{r}_q - \mathbf{D}_q\ \mathbf{r}_p = \mathbf{0}, \tag{5.4}$$

where $\mathbf{D}_p$ and $\mathbf{D}_q$ in equation (5.4) represent the convolution matrices of channels p and q, respectively. The combination of all possible equations leads to the following homogeneous system of equations

$$\mathbf{A}\ \mathbf{x} = \mathbf{0}, \tag{5.5}$$

where

$$\mathbf{A} = \begin{pmatrix} \mathbf{D}_2 & -\mathbf{D}_1 & & & & & \\ \mathbf{D}_3 & & -\mathbf{D}_1 & & & & \\ \mathbf{D}_4 & & & -\mathbf{D}_1 & & & \\ \vdots & & & & \ddots & & \\ & \mathbf{D}_3 & -\mathbf{D}_2 & & & & \\ & \mathbf{D}_4 & & -\mathbf{D}_2 & & & \\ & \vdots & & & \ddots & & \\ & & & & \mathbf{D}_J & & -\mathbf{D}_{J-2} \\ & & & & & \mathbf{D}_J & -\mathbf{D}_{J-1} \end{pmatrix} \tag{5.6}$$

and

$$\mathbf{x} = \left[\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \ldots, \mathbf{r}_J\right]^T. \tag{5.7}$$

The classical formulation of Euclid deconvolution estimates the reflectivity by estimating the eigenvector associated to the minimum non-zero eigenvalue of $\mathbf{A}^T\mathbf{A}$ (Rietsch, 1997a). A small amount of noise in the data makes the solution impractical for real data applications (Rietsch, 1997b).

5.2.2 Sparse Multichannel Blind Deconvolution

In the previous analysis we did not consider noise. The addition of a noise term in our signal model leads to the following expression

$$D_j(z) = W(z)\, R_j(z) + N_j(z) \qquad j = 1, \ldots, J \tag{5.8}$$

and

$$\begin{aligned} D_p(z)\; R_q(z) - \quad & D_q(z)\; R_p(z) = \\ & N_p(z)\; R_q(z) - N_q(z)\; R_p(z)\,, \quad \forall\, p, q \end{aligned}$$

or in matrix form,

$$\mathbf{D}_p\mathbf{r}_q - \mathbf{D}_q\mathbf{r}_p = \mathbf{N}_p\mathbf{r}_q - \mathbf{N}_q\mathbf{r}_p\,, \tag{5.9}$$

The last expression is expressed as follows

$$\mathbf{A}\mathbf{x} = \mathbf{e}\,. \tag{5.10}$$

We will assume that $\mathbf{e}$ is white and Gaussian which is clearly a hypothesis that permits us to develop an algorithm, but one understands that $\mathbf{e}$ is not necessarily white and Gaussian (Sroubek and Milanfar, 2012). Therefore, we propose to find a solution $\mathbf{x}$ that minimizes the l_2 norm of the error term $\mathbf{e}$ with the requirement that $\mathbf{x}$ is sparse. To avoid the trivial solution, we must provide an extra constraint (i.e., $\mathbf{x}^T\mathbf{x} = 1$). To summarize the problem, we propose to find the solution by minimizing the following cost function

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}}\, J(\mathbf{x}), \quad \text{subject to} \quad \mathbf{x}^T\mathbf{x} = 1 \tag{5.11}$$

where

$$J(\mathbf{x}) = \frac{1}{2}||\mathbf{A}\,\mathbf{x}||_2^2 + \lambda\mathcal{R}_\epsilon(\mathbf{x}) \tag{5.12}$$

and

$$\mathcal{R}_\epsilon(\mathbf{x}) = \sum_i (\sqrt{x_i^2 + \epsilon^2} - \epsilon). \tag{5.13}$$

The trade-off parameter λ balances the importance of sparseness of the reflectivity and data fitting. $\mathcal{R}_\epsilon(\mathbf{x})$ is the regularization term that promotes a sparse solution. For estimation of sparse solutions, one could have adopted an l_1 norm. However, we prefer to use the hybrid l_1/l_2 norm, $\mathcal{R}_\epsilon(\mathbf{x})$, because it is differentiable and therefore it enables us to use simple optimization methods based on steepest descent techniques (Bube and Langan, 1997; Lee et al., 2006; Schmidt et al., 2007; Li et al., 2012). Figure 5.1 shows the functional $\mathcal{R}_\epsilon$ for $\epsilon = 0.01$ and 0.05 and the classical l_1 norm.

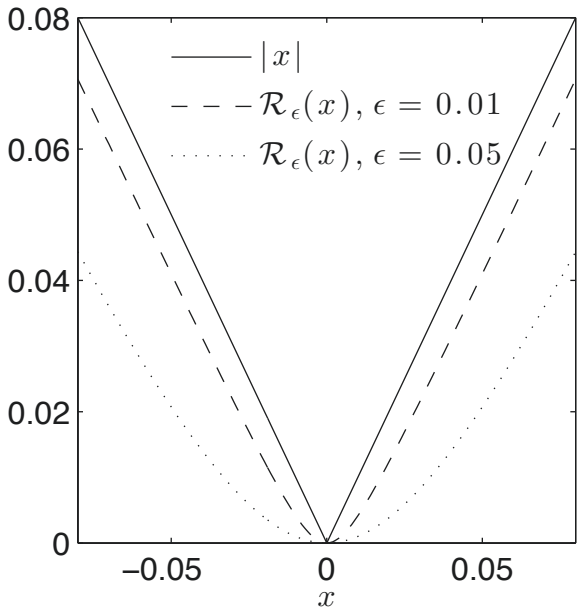

Figure 5.1: Comparison of the $l_1 = |x|$ norm and its approximation $\mathcal{R}_\epsilon(\mathbf{x})$ for $\epsilon = 0.01$ and 0.05.

The constrained optimization problem involves minimizing the following cost function

$$\mathcal{L}(\mathbf{x}) = J(\mathbf{x}) - \frac{\eta}{2}(\mathbf{x}^T\mathbf{x} - 1), \tag{5.14}$$

with Lagrange multiplier given by the scalar η. The condition for minimum must satisfy

$$\nabla\mathcal{L}(\mathbf{x}) = \mathbf{g} - \eta\mathbf{x} = 0, \tag{5.15}$$

where $\mathbf{g} = \nabla J(\mathbf{x})$. Multiplying equation (5.15) by $\mathbf{x}^T$ and using the condition $\mathbf{x}^T\mathbf{x} = 1$ yields

$$\eta = \mathbf{x}^T\mathbf{g}\,. \tag{5.16}$$

Hence, the projection of the gradient on the sphere is given by

$$\nabla\mathcal{L}(\mathbf{x}) = \mathbf{g} - (\mathbf{x}^T\mathbf{g})\,\mathbf{x}\,. \tag{5.17}$$

The steepest descent algorithm can be expressed via the classical update rule $\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k\,\mathbf{h}_k$ with normalized gradient $\mathbf{h}_k = \nabla\mathcal{L}(\mathbf{x}_k)/|\nabla\mathcal{L}(\mathbf{x}_k)|$ where α_k is the adaptive step size and k indicates iteration. A step in the direction of steepest descent might move $\mathbf{x}_{k+1}$ off the unit sphere. This is illustrated by Figure 5.2a. Therefore, we prefer to use an educated step that was derived from Rodrigues' rotation formula (Murray et al., 1994)

$$\mathbf{x}_{k+1} = cos(\theta_k)\,\mathbf{x}_k + sin(\theta_k)\,\mathbf{h}_k\,. \tag{5.18}$$

It is easy to show via a few algebraic manipulations that with this expression the updated solution $\mathbf{x}_{k+1}$ is on the unit sphere (See Appendix D.1). Notice that by choosing a small angle $\theta_k < 0$ we obtain the desired small step in the direction of steepest descent because $\sin(\theta_k) \approx \theta_k < 0$. In equation (5.18), the effect of the term $\cos(\theta_k)$ is to shrink the current position $\cos(\theta_k)\mathbf{x}_k$ in a way that the updated solution lies on the unit sphere. This is illustrated by Figure 5.2b. Our problem has been reduced to a one-dimensional minimization on the sphere. The minimization is carried out by a simple update of θ_k that guarantees that $J(\mathbf{x}_k) > J(\mathbf{x}_{k+1})$ (See Appendix D.2).

It is important to point out that the steepest descent algorithm must be initialized by a solution that is close to the final reflectivity. This is because the problem is non-linear and multi-modal. Our test indicates that starting with the data as an initial solution always leads to sparse estimate of the reflectivity. The last assessment is supported because the data is structurally close to the true sparse reflectivity.

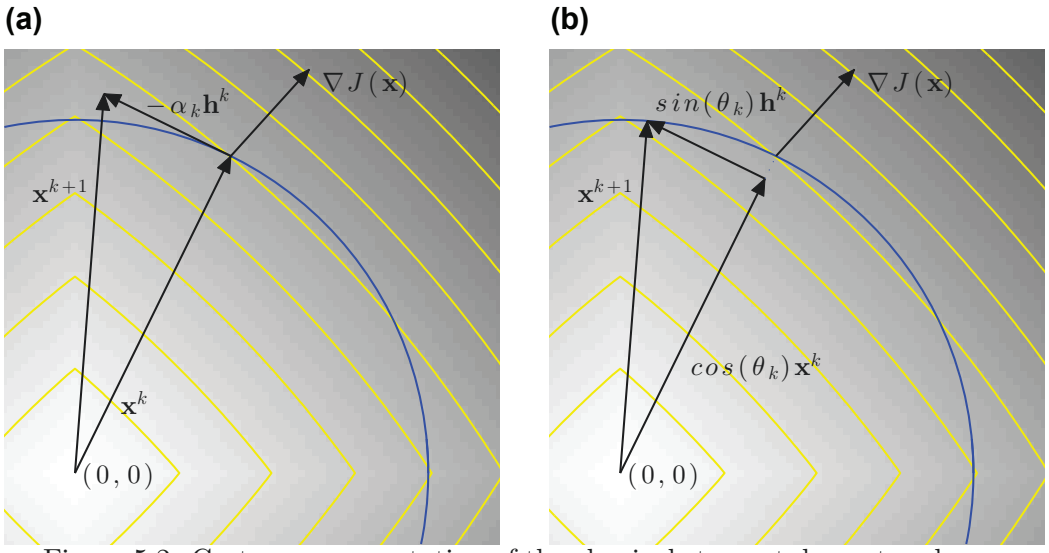

Figure 5.2: Cartoon representation of the classical steepest descent and proposed approach. a) Steepest descent algorithm via classical update rule $\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k\,\mathbf{h}_k$, α_k is the step length. b) Proposed steepest descent algorithm via the update rule $\mathbf{x}_{k+1} = cos(\theta_k)\,\mathbf{x}_k + sin(\theta_k)\,\mathbf{h}_k$, the step length is given by $sin(\theta_k)$. In this case the updated position is guaranteed to lie on the unit sphere.

5.3 Examples

To examine the performance of the proposed method, we introduce two figures of merit for both the estimated wavelet and the estimated reflectivity series. For instance, if a true generic signal is denoted by $\mathbf{y}_0$ and the estimated signal by $\mathbf{y}$, we define the quality of the reconstruction, Q, as follows

$$Q = 10 \log \frac{\|\mathbf{y}_0\|_2^2}{\|\mathbf{y}_0 - \mathbf{y}\|_2^2}\,. \tag{5.19}$$

Our second figure of merit is the Normalized Correlation Coefficient,

$$NCC = \frac{\mathbf{y}^T \mathbf{y}_0}{\|\mathbf{y}_0\|_2\,\|\mathbf{y}\|_2}\,. \tag{5.20}$$

We stress that blind deconvolution algorithms can only determined unscaled versions of the seismic wavelet and reflectivity series. Therefore all our estimators must be re-

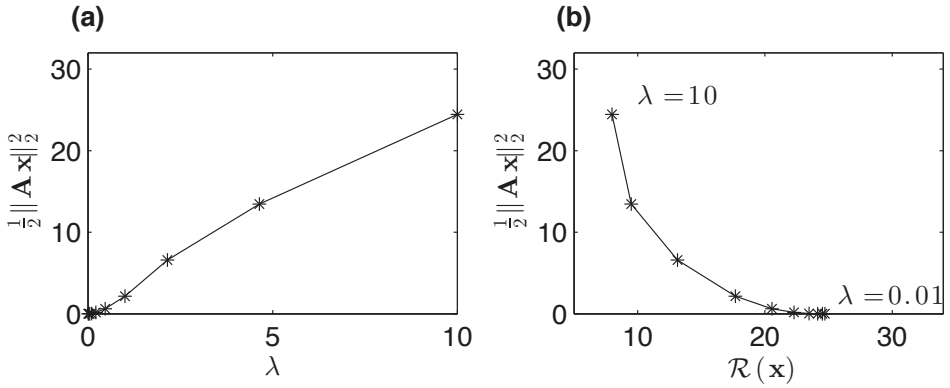

Figure 5.3: Exploring the importance of trace-off parameter λ in SMBD method using synthetic data with $SNR = 100$. a) Trade-off parameter versus the l_2 norm of homogeneous system of equations. b) Trade-off curve using different values of regularization parameter.

scaled prior to computating Q. For this purpose, the estimated signal $\mathbf{y}$ is multiplied by a scalar α such that $\|\mathbf{y}_0 - \alpha\mathbf{y}\|_2^2$ is minimum. It is clear that $\mathbf{y}$ in equation (5.19) is replaced by $\alpha\mathbf{y}$. It is also clear that the aforementioned scaling is not needed for the computation of NCC. We will denote NCC_w and NCC_x the normalized correlation coefficients for the wavelet and the reflectivity, respectively. Similarly, we will use Q_w and Q_x to indicate the quality of the reconstruction of the wavelet and reflectivity, respectively.

5.3.1 Simulations

To test the method, we first run a synthetic example with high signal-noise-ratio ($SNR = 100$). We run the algorithm for different values of the parameter λ to explore the trade-off curve. The latter is portrayed in Figure 5.3. With different experiments we concluded that $\lambda = 4$ yields the best result for different levels of SNR. The true sparse reflectivity is displayed in Figure 5.4a in conjunction with the data (Figure 5.4b). The sampling interval for this exercise was chosen $\Delta t =$ 2 ms. We also used a Ricker wavelet of central frequency 40 Hz and a 50 degrees phase rotation. The estimated reflectivity for this example is shown in Figure 5.4c. The estimated reflectivity and the seismic traces were used to estimate the seismic wavelet via multichannel frequency domain deconvolution. The estimated wavelet and the true input wavelet are shown side by side in Figure 5.4d. In this example, the

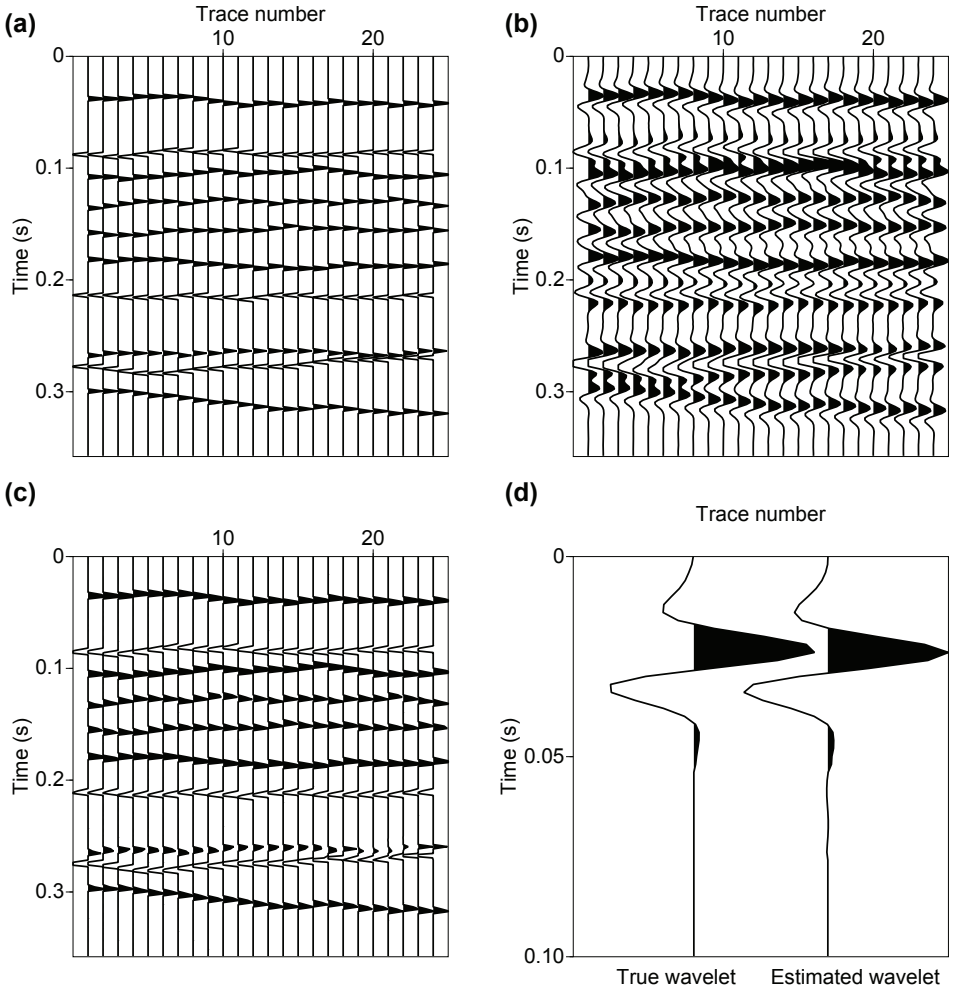

Figure 5.4: Performance of the SMBD method using synthetic data with $SNR = 100$. a) True synthetic reflectivity sequences. b) Seismic traces with $SNR = 100$. b) Estimated sparse reflectivity sequences. d) True and estimated wavelets.

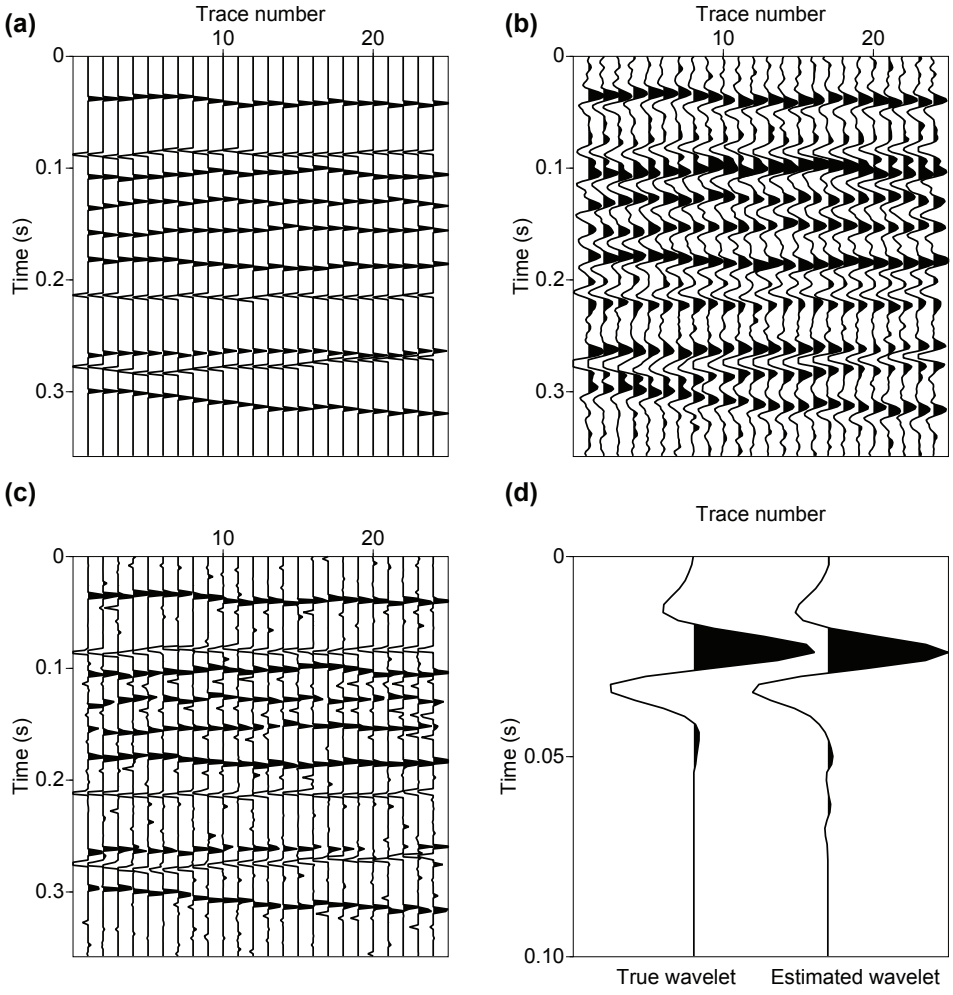

Figure 5.5: Performance of the SMBD method using synthetic data with $SNR = 4$. a) True synthetic reflectivity sequences. b) Seismic traces with $SNR = 4$. b) Estimated sparse reflectivity sequences. d) True and estimated wavelets.

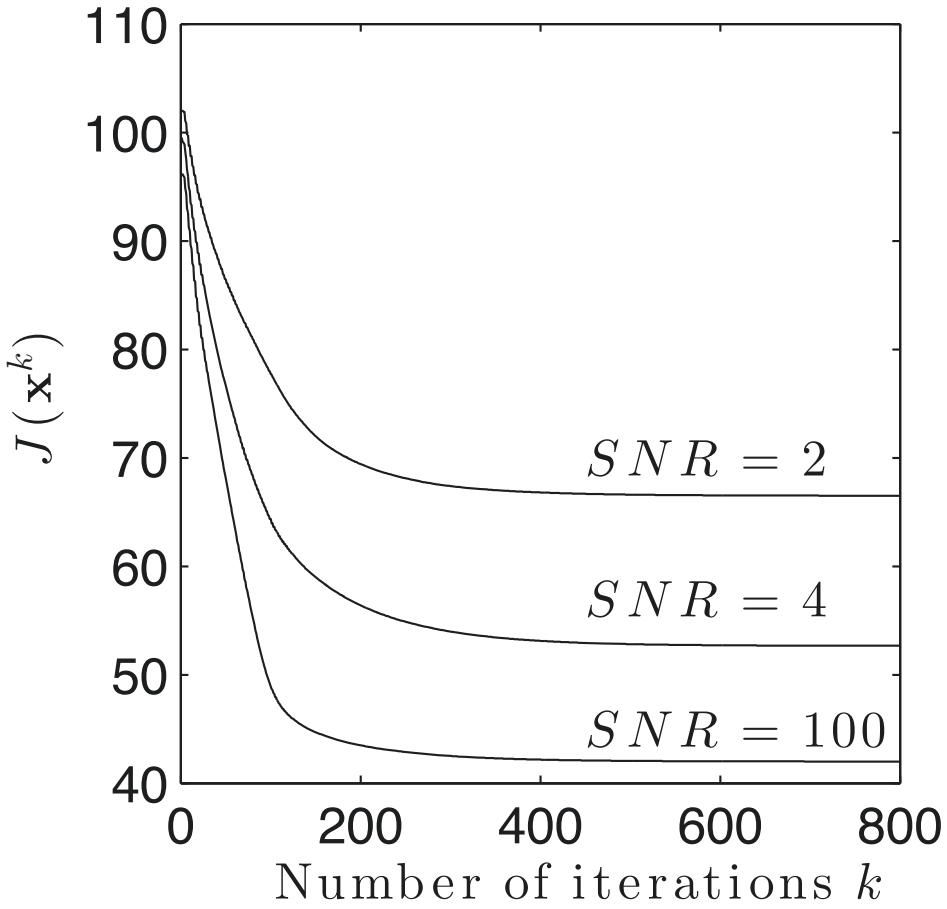

Figure 5.6: Convergence behaviour of SMBD method using synthetic data with different levels of noise.

quality of reconstruction for the wavelet is $Q_w = 14$ dB. Similarly, the quality of the reconstruction for the reflectivity is $Q_x = 5$ dB. Normalized correlation coefficients for the wavelet and reflectivity are given be $NCC_w = 0.96$ and $NCC_x = 0.82$, respectively.

We rerun the synthetic example with additive noise $SNR = 4$. In this case we obtain a reconstruction quality of $Q_w = 13$ dB and a normalized correlation coefficient $NCC_w = 0.89$. Similarly, we obtain $Q_x = 3.8$ dB and $NCC_x = 0.75$ for the reflectivity series. The results for this simulation are shown in Figure 5.5. Finally, we also provide convergence curves for the algorithm for $SNR = 2, 4, 100$ in Figure 5.6.

To analyze the stability of the $SMBD$ method under different levels of noise, we run a Monte Carlo simulation with 20 different realizations of noise and seismic reflectivity for a given SNR and parameter λ. Each realization of the reflectivity is similar to Figure 4a. We have been careful is producing realizations with 10 reflectors each and similar temporal and spatial variability. This was done by taking a random reflectivity composed of 10 impulses of random amplitude for the first trace and randomly perturbing the times to generate the reflectivity of the remaining traces. The amplitude of the reflection coefficients for a given reflector were also allowed to vary in space by a very small amount. The 20 realizations where used to estimate 20 wavelets and 20 multichannel reflectivity estimators that were used to estimate averages and standard errors for NCC_w, NCC_x, Q_w and Q_x.

At this point it is important to mention that the computation of the standard error of the normalized correlation coefficients NCC_w and NCC_x requires special attention. Normalized correlation coefficients are bounded by unity and therefore, they are not normally distributed (Weatherburn, 1949). To compute the mean and standard error of the normalized correlation coefficient, we first apply the Fisher's transform to create a new variable $F = \frac{1}{2}\ln[(1 + (NCC))/(1 - (NCC))]$ that is distributed almost normally. We can now compute the mean and standard error of the variable F which are then inverse transformed to obtained the desired mean and standard error of the normalized correlation coefficient (VanDecar and Crosson, 1990; Herrera and van der Baan, 2012).

The average for 20 realizations of our two figures of merit and their standard deviations are shown in Figures 5.7, 5.8 and 5.9 for $\lambda = 1, 4$ and 10, respectively. In these figures, the diamonds ($\diamond$) are used to indicate NCC_w and Q_w values for

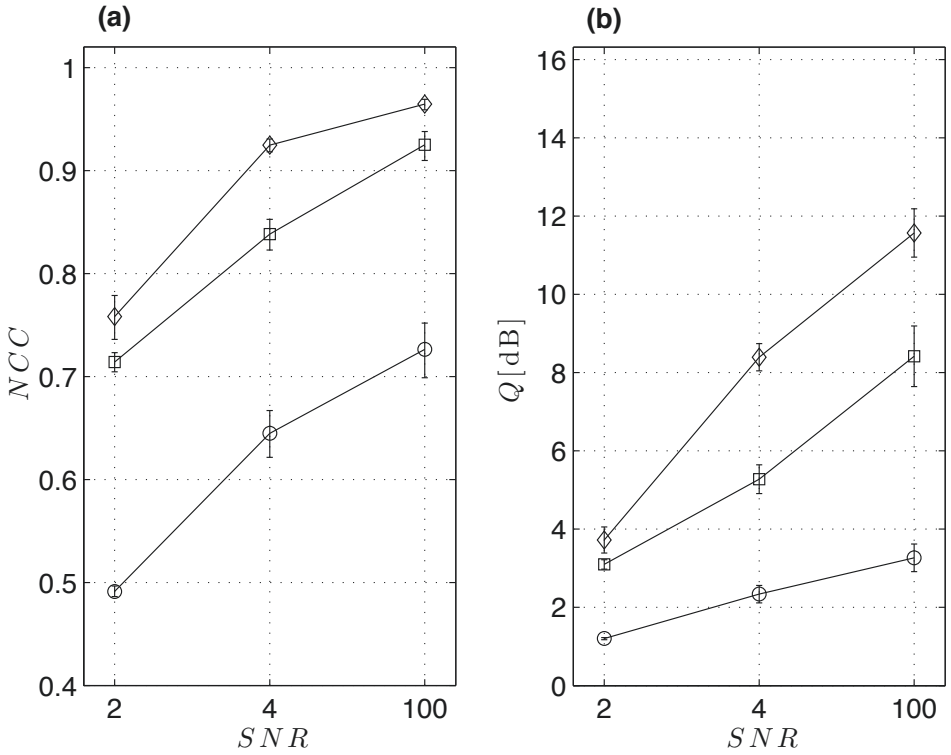

Figure 5.7: a) Mean and standard error of normalized correlation coefficients versus SNR. (b) Mean and standard error of the quality of the reconstruction versus SNR. These results were obtained by running SMBD on 20 realizations of reflectivity models that are similar to the reflectivity shown in Figure 5.4a. Diamonds ($\diamond$) are used to indicate NCC_w and Q_w values for the estimated seismic wavelet. Similarly, circles ($\circ$) are used to indicate the NCC_x and Q_x for the estimated reflectivity. Squares ($\square$) are used to indicate the NCC_x and Q_x computed after applying an Ormsby trapezoidal filter to the true reflectivity and the estimated reflectivity. The trade-off parameter of the SMBD method is $\lambda = 1$. Note that the SNR values are not linearly spaced.

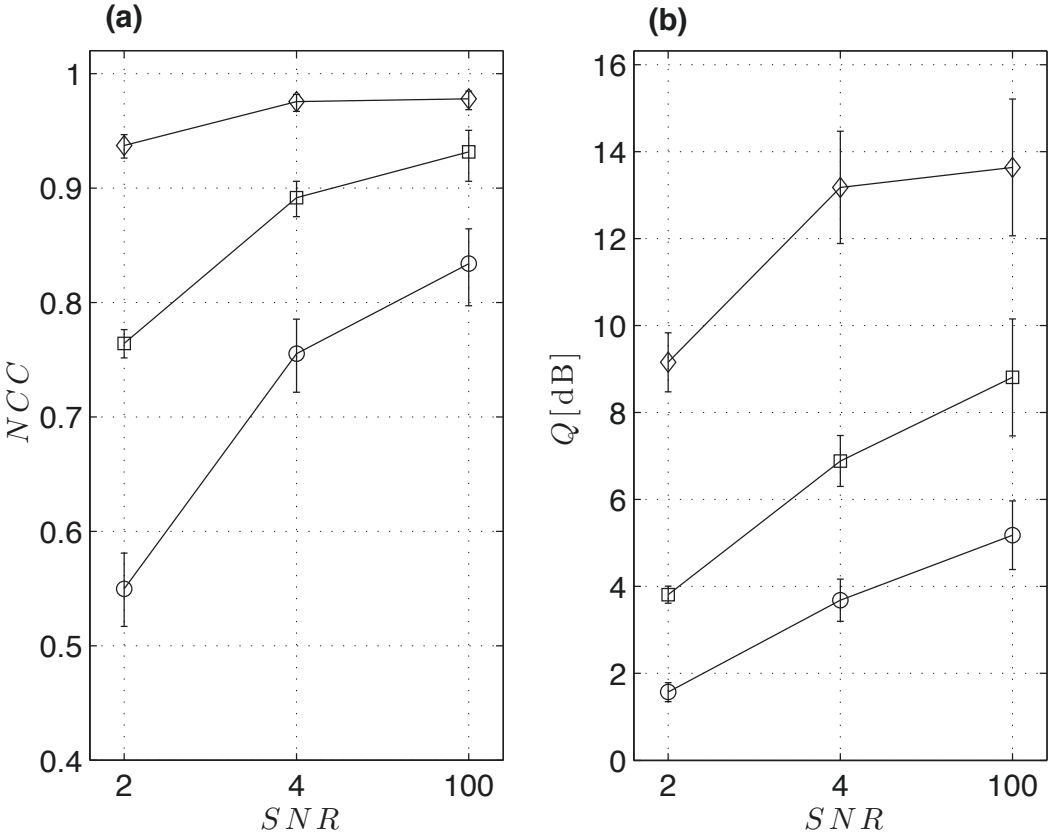

Figure 5.8: Similar to Figure 5.7 but with $\lambda = 4$.

the estimated seismic wavelet. Similarly, the circles ($\circ$) were used to indicate the NCC_x and Q_x values for the estimated seismic reflectivity. We observe that the proposed deconvolution scheme performs better at estimating the wavelet than the reflectivity. The reflectivity is a full band sequence with low and high frequencies annihilated by the bandpass character of the seismic wavelet. Therefore, recovery of frequencies outside the natural band imposed by the seismic wavelet fully relies on the sparse reflectivity assumption (Sacchi et al., 1994). Our synthetic data were sampled at 2 ms and therefore, hoping to recover a full band reflectivity with spectral contributions from 0 Hz to 250 Hz (Nyquist frequency) is definitely an unworkable cause. The latter is reflected by the low values of Q_x in comparison to those of Q_w. To gain critical understanding of the limits of our algorithm, we also compute Q_x values for bandpassed versions of the true and estimated reflectivity series. In this case, we utilize a Ormsby trapezoidal filter (Sheriff, 2002) defined by four corner frequencies $0, 1, 100, 125$ Hz to restrict the bandwidth of the true reflectivity and estimated reflectivity before computing Q_x and NCC_x. The results are also included in Figure 5.7, 5.8 and 5.9 and are indicated with squares ($\square$). One can observe an improvement in NCC_x and Q_x when high frequencies are excluded from the recovered reflectivity.

Finally, Figures 5.10 and 5.11 portray the resulting reflectivity inversion in time and frequency domain for one realization and for the first seismic trace of a group of 24 traces. The example also corresponds to a parameter $\lambda = 4$ and $SNR = 4$. Figure 5.10a presents the seismic trace, Figure 5.10b the true reflectivity series, Figure 5.10c the full band-reflectivity estimated by SMBD, and Figure 5.10d is the estimated reflectivity after applying the Ormsby trapezoidal filter. Figure 5.11 shows, in the same order, the power spectral density of the signals portrayed in Figure 5.10. It is evident from Figure 5.11 that the unfiltered spectra match well the true spectra up to about 100 Hz.

5.3.2 Real data example

Unfortunately, we do not have an automatic way of estimating the trade-off parameter λ needed by SMBD. The simulations in the previous section were used to explore the behaviour of the algorithm in terms of the parameter λ. Based on the analysis presented in Figures 5.7, 5.8 and 5.9, we have selected $\lambda = 4$ because it

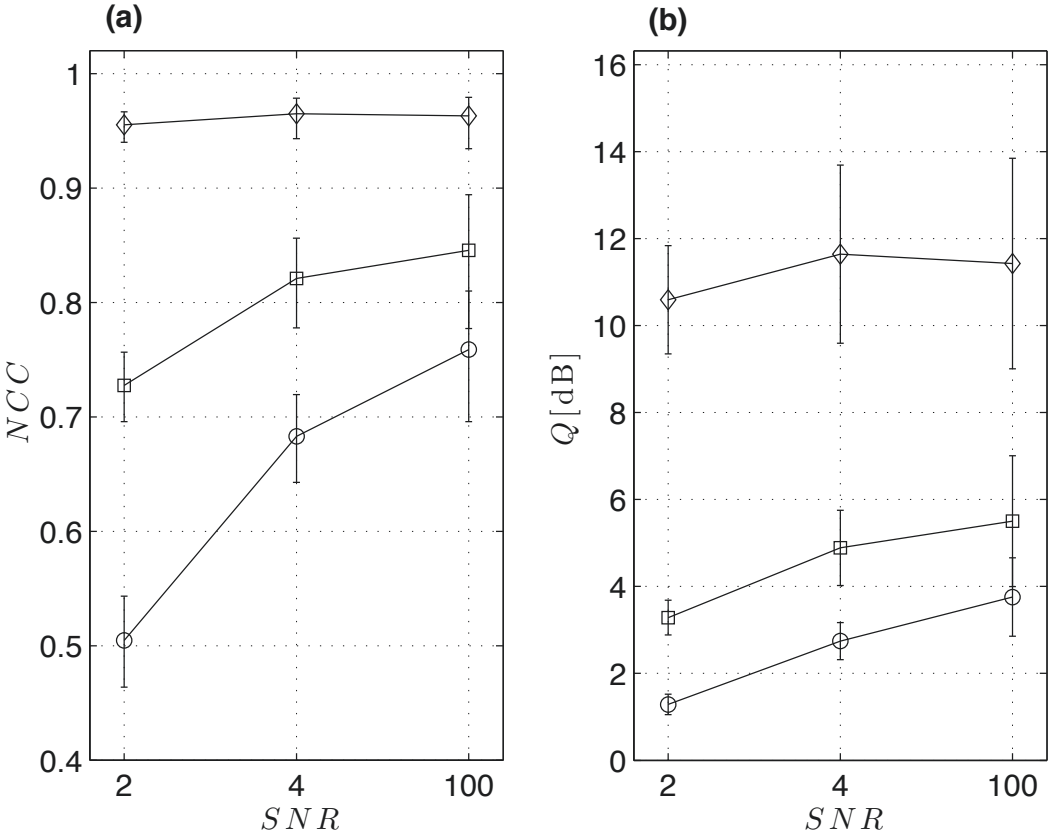

Figure 5.9: Similar to Figure 5.7 but with $\lambda = 10$.

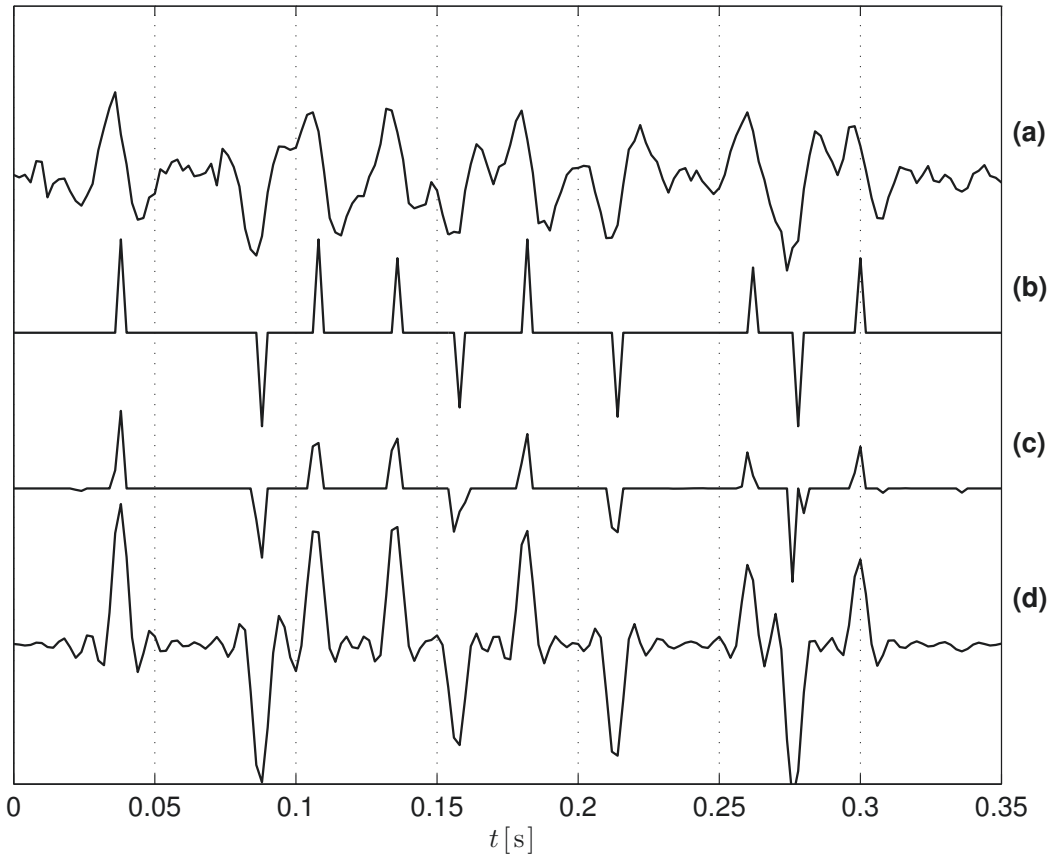

Figure 5.10: a) First trace from one realization of multichannel data similar to Figure 5.5b. b) True reflectivity series. c) Estimated reflectivity via SMBD. d) The estimated reflectivity after applying an Ormsby trapezoidal filter with corner frequencies $0, 1, 100, 125\,\mathrm{Hz}$. The simulation corresponds to values $SNR = 4$ and $\lambda = 4$.

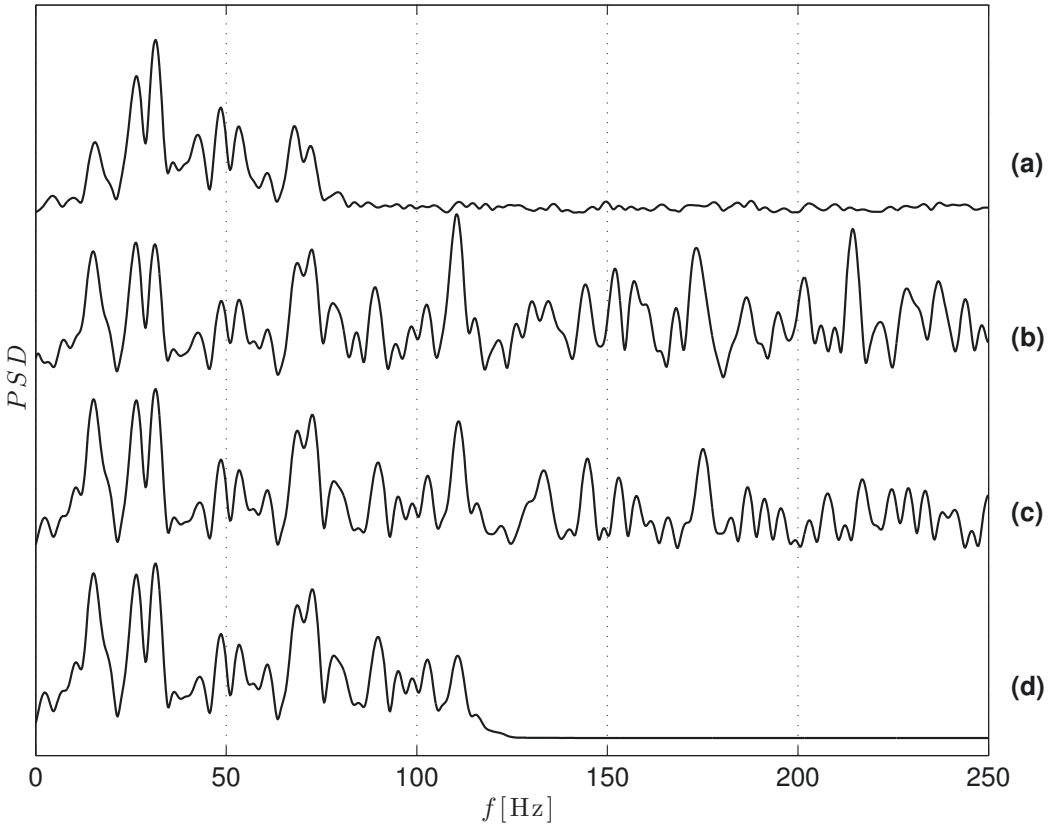

Figure 5.11: Power spectral density of the data portrayed in Figure 5.10. a) Seismic trace. b) True reflectivity series. c) Estimated reflectivity via SMBD. d) The estimated reflectivity after applying an Ormsby trapezoidal filter with corner frequencies $0, 1, 100, 125$ Hz.

provided the best reconstruction of the wavelet and reflectivity for a moderate level of noise ($SNR = 4$).

Our real data test uses the Gulf of Mexico, Mississippi Canyon dataset. These data have been extensively used for testing multiple suppression algorithms (see, for instance, Verschuur and Prein (1999)). SMDB was run on the near offset section of the Mississippi Canyon dataset. The input data and the estimated sparse impulse response are shown in Figures 5.12a and b, respectively. We also show details of the seismic sections before and after deconvolution in Figures 5.13a and b. Notice that Figure 5.13b is the resulting sparse impulse response estimated by SMDB.

This dataset is contaminated by multiples. Therefore, our blind deconvolution algo-

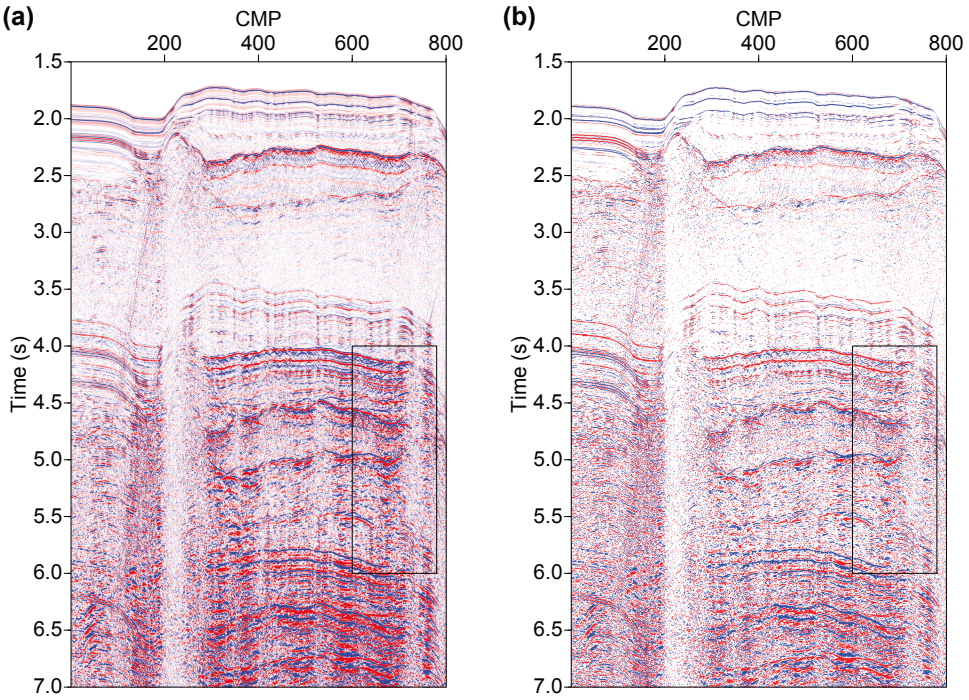

Figure 5.12: a) Near offset section of data set from the Gulf of Mexico. b) Estimated sparse reflectivity.

rithm was used to estimate the full impulse response including multiples rather than the primary only impulse response (reflectivity). We have run SMBD in windows of 1 second in time and 200 traces with 10% overlap in time and space. The windows were patched back together to produce Figure 5.12b.

We used the estimated impulse response of the whole near offset section to estimate the wavelet via a multichannel frequency domain deconvolution. Wavelets computed from individual windows where similar and this is why we have decided to compute one wavelet for the whole near offset section. The seismic wavelet is portrayed in Figure 5.14. We also displayed the estimated wavelet obtained by aligning and averaging the first break. The resemblance of the two wavelets is remarkable, with normalized correlation coefficient $NCC_w = 0.92$.

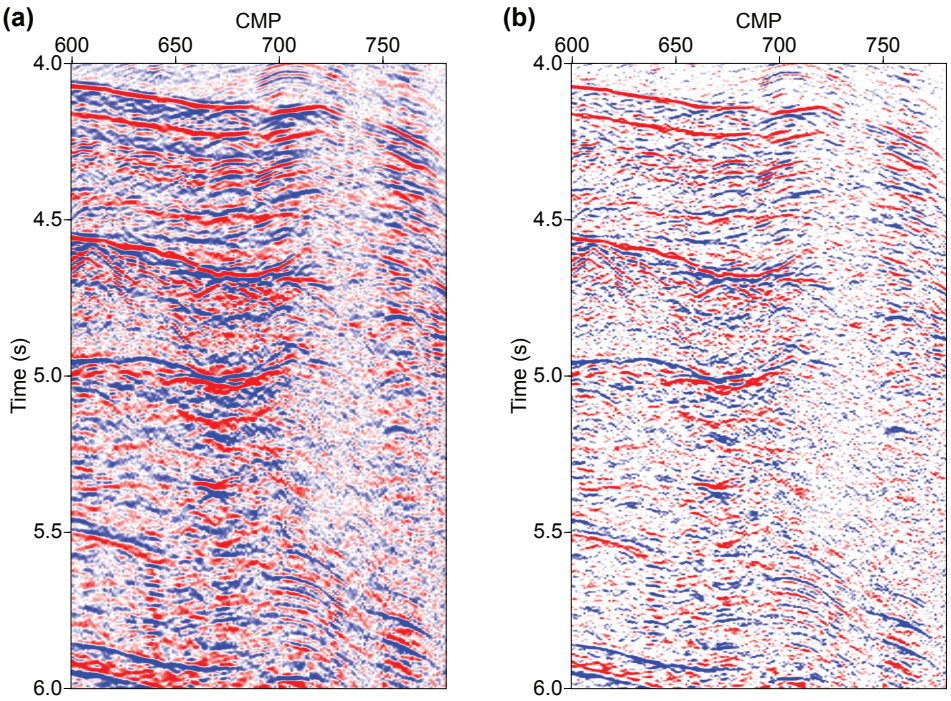

Figure 5.13: A zoom into the black rectangles marked in Figure 5.12. a) Before deconvolution. b) After deconvolution.

5.4 Summary

We have presented an algorithm inspired by Euclid deconvolution that permits us to estimate the seismic reflectivity without a priori knowledge of the seismic wavelet. The seismic wavelet is computed as a byproduct of the process via a multichannel frequency domain deconvolution between traces and estimated reflectivity sequences. The core of the algorithm is the estimation of the reflectivity by the solution of the multichannel homogeneous system of equations with sparsity constraints.

An optimization problem that uses the method of steepest descent was developed. To avoid trivial solutions the reflectivity vector was constrained to have unit norm. The latter leads to a constrained optimization problem where one attempts to estimate a sparse signal that fits a multichannel homogeneous system of equations and, in addition, the signal lies on the unit sphere. This optimization problem was solved by the method of steepest descent with an update rule that keeps current estimates of the sparse reflectivity on the unit sphere. The method is stable under a variety of noise levels and for different values of the trade-off parameter λ. We stress that like in every deconvolution scenario, the method works well when it honours certain assumptions. In this case, not only the wavelet needs to be stationary for all traces but the reflectivity must be sparse.

We used both synthetic and real data examples to evaluate the method. Synthetic examples permitted us to assess the viability of the method in terms of noise. The method gives reasonable estimates of wavelet and reflectivity series with $SNR = 4$ and higher. We have obtained workable results for $SNR = 2$. However, the results of the method clearly deteriorate when we try to push it to work on data severely contaminated with noise. We have observed that the quality of the estimated wavelet is superior to the quality of the estimated reflectivity for the same SNR.

We also applied the method to near offset section of Gulf of Mexico dataset. Our estimated wavelet has a remarkable semblance with the wavelet estimated by aligning and stacking first breaks.

We emphasize that the $SMBD$ method does not consider coherent noise in the convolutional model. In this regard, we believe that its application to onshore data will require extensive preconditioning to remove coherent noise. This is likely true for all blind deconvolution methods.

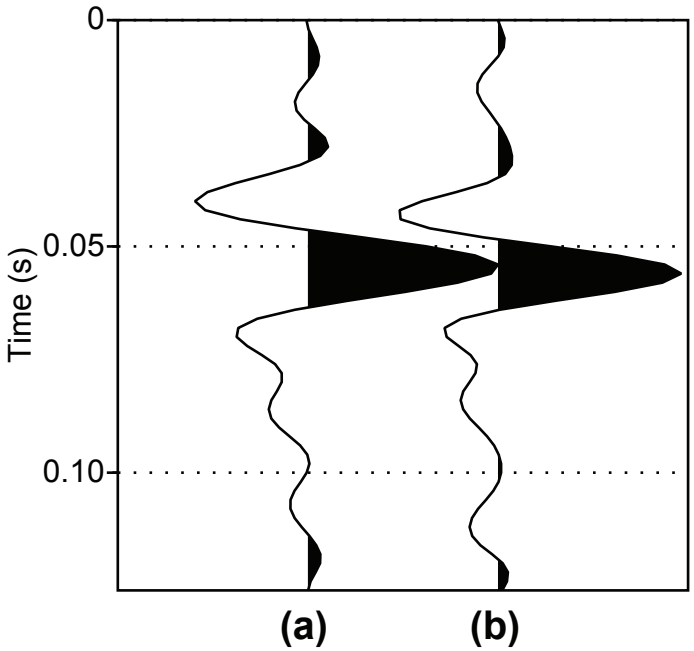

Figure 5.14: Estimated wavelet for Golf of Mexico data set. a) Estimated wavelet using the SMBD method. b) Estimated wavelet obtained by averaging the first beak after alignment.

CHAPTER 6

Conclusions

6.1 Summary

The primary focus of this dissertation was about designing and applying efficient least-squares migration algorithms. I implemented prestack one-way wave equation depth migration. In conventional migration (i.e., adjoint migration) the data fidelity is compromised. In other words, the migrated model does not honour the recorded wavefield. To honour the wavefield and improve the resolution of migrated image, application of the least-squares migration algorithms is necessary. However, the merit of posing migration as an inverse problem encounters several difficulties. First of all, the computational cost of iterative algorithms is high. The second major issue is that the designed operators for communications from model to data and data to the model domain are simplified, and the operators are only able to consider part of the wavefield. I used one-way wave equation to define the forward/migration and adjoint/de-migration operators and from there I used the forward/migration and adjoint/de-migration pairs to solve the least-squares migration problem. The one-way wave equation only considers acoustic wavefields (i.e., P-P reflections) and it ignores shear waves, free surface, internal multiples and turning waves. To acquire an acceptable migrated image, we need to preprocess the data before application of least-squares migration algorithms.

In chapter 2, I used the Born approximation to define and design the forwards/de-migration operator.

In chapter 3, I took advantage of adaptive signal processing techniques and solve the least-squares migration problem in a recursive fashion. This methodology gives us fast and memory efficient algorithms to handle least-squares migration. The necessity of solving least squares problems recursively stems from the need for fast real-time signal processing strategies. The computational cost of least squares solvers can also limit the applicability of this technique in many geophysical problems. I have considered recursive least squares solutions for least-squares wave equation migration with sliding windows involving several fixed rank downdating and updating computations. If we use enough data in each windowed setup, the spectrum of the preconditioned system is clustered around one and the method will converge super-linearly with probability one.

Chapter 4 considers the application of preconditioners in the context of least-squares migration. Preconditioners are operators that can cluster the eigenvalues of the Hessian which ultimately results in improving the convergence of iterative algorithms. Improving the convergence rate is equivalent to reducing the computational cost of the algorithm. In other words, by using preconditioners we will need fewer iterations to reach the predefined accuracy in the migrated image. Preconditioners are used as an alternative tool to change the linear system of equations is a way that it is easier for the Conjugate Gradient algorithm to handle the problem. Diagonal scaling of the Hessian and an approximated inverse of the Hessian are good candidates for preconditioners. I introduced and applied three different preconditioners. The first two preconditioners belong to the diagonal scaling category, and the third one is a filter-based approach which approximates the partial Hessian operators by local convolutional filters. I showed that these preconditioners are effective operators to improve the convergence rate of the least-squares migration algorithms.

In chapter 5, I focused on the source signature estimation problem. As we know, one of the input requirements of pre-stack least-squares migration algorithms is the source signature. To estimate the source signature from the data, I developed a Sparse Multichannel Blind Deconvolution (SMBD) method. The method is a modification of the multichannel blind deconvolution technique often called Euclid deconvolution where the multichannel impulse response of the earth is estimated by solving a homogeneous system of equations. Classical Euclid deconvolution is unstable in the presence of noise and requires the correct estimation of the length of the seismic wavelet. The proposed method, on the other hand, can tolerate

moderate levels of noise and does not require a priori knowledge of the length of the wavelet. SMBD solves the homogeneous system of equations arising in Euclid deconvolution by imposing sparsity on the unknown multichannel impulse response. Trivial solutions to the aforementioned homogeneous system of equations are avoided by seeking sparse solutions on the unit sphere. Synthetic examples were used to judge the viability of the method in terms of noise. In this case of real data (i.e., Gulf of Mexico dataset), the estimated wavelet was compared to a wavelet estimated by averaging first breaks. The estimated wavelet showed a noticeable resemblance to the average first break with a normalized correlation coefficient of 0.92.

6.2 Future work

In this dissertation, I implemented a pre-stack one-way wave equation migration and posed the migration as an inverse problem. However, the designed operators only considered acoustic wavefields and ignored converted wave, free surface multiples and internal multiples. In the future, one can apply accurate forward/de-migration and adjoint/migration pairs like those derived for reverse-time migration (two-way wave equation propagators). The second problem worthy of investigation is the implementation of different regularization techniques to emphasize good features in the migrated image. The latter should aim at improving the interpretability of subsurface images. Moreover, in this thesis, I used an extended imaging approach that yield common shot index gathers, however, implementing prestack algorithm with subsurface offset or Poynting vectors (De Bruin et al., 1990; Prucha et al., 1999; Biondi and Symes, 2004) will result in angle gathers that could lead to more accurate and interpretable amplitude responses. Last, I would suggest exploring and extending preconditioning based on the filter-based approach to least-squares elastic migration (Feng and Schuster, 2017; Ren et al., 2017). Significant results could be obtained by designing filter-based preconditioning strategies that can alleviate multi-parameter crosstalk contamination in elastic least-squares migration.

Bibliography

Baysal, E., D. D. Kosloff, and J. W. C. Sherwood, 1983, Split-step fourier migratio: Geophysics, **48**, 1514– 1524.

Berkhout, A. J., and D. J. Verschuur, 2005, *in* Multiple technology: Part 2, Migration of multiple reflections: 1497–1500.

Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized radon transform: Journal of Mathematical Physics, **26**, 99–108.

Biondi, B., and W. W. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging: Geophysics, **69**, 1283–1298.

Bube, K., and R. Langan, 1997, Hybrid minimization with applications to tomography: Geophysics, **62**, no. 4, 1183–1195.

Buhl, P., P. Stoffa, and G. Bryan, 1974, The application of homomorphic deconvolution to shallow water marine seismology Part I: Real data: Geophysics, **39**, no. 4, 417–426.

Cambois, G., and N. Hargreaves, 1994, Zero-phase conversion of marine data using one-parameter phase filters and kurtosis maximization: SEG Technical Program Expanded Abstracts, **433**, 1591–1594.

Canadas, G., 2002, A mathematical framework for blind deconvolution inverse problems: SEG Technical Program Expanded Abstracts, 2202–2205.

Chavent, G., and R. E. Plessix, 1999, An optimal true-amplitude least squares prestack depth-migration operator: Geophysics, **64**, 508– 515.

Cheng, J., N. Kazemi, and M. Sacchi, 2016, Least-squares migration via a gradient projection method-application to seismic data deblending: Presented at the 78th EAGE Conference and Exhibition 2016.

Claerbout, J., 1971a, Toward a unified theory of reflector mapping: Geophysics, **36**, 467–481.

Claerbout, J., and D. Nichols, 1994, Spectral preconditioning: Stanford Exploration Project, **82**.

Claerbout, J. F., 1971b, Toward a unified theory of reflector mapping: Geophysics, **36**, 467–481.

Clayton, R. W., and R. H. Stolt, 1981, A born-wkbj inversion method for acoustic reflection data: Geophysics, **46**, 1559–1567.

Collino, F., and P. Joly, 1995, Splitting of operators, alternate directions, and paraxial approximations for the three-dimensional wave equation: SIAM Journal on Scientific Computing, **16**, 1019– 1048.

Dai, W., X. Wang, and G. T. Schuster, 2011, Least-squares migration of multisource data with a deblurring filter: Geophysics, **76**, R135–R146.

De Bruin, C., C. Wapenaar, and A. Berkhout, 1990, Angle-dependent reflectivity by means of prestack migration: Geophysics, **55**, 1223–1234.

Dellinger, J., and J. Etgen, 1990, Wave-field separation in two-dimensional anisotropic media: Geophysics, **55**, 914–919.

Demanet, L., P. Ltourneau, N. Boumal, H. Calandra, J. Chiu, and S. Snelson, 2012, Matrix probing: A randomized preconditioner for the wave-equation hessian: Applied and Computational Harmonic Analysis, **32**, 155 – 168.

Donoho, D., 1981, On minimum entropy deconvolution: Academic Press, **109**, no. 6, 2831–9.

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, ExpandedAbstracts, 1012– 1015.

Etgen, J., 2002, Recent advances and the road ahead: An invited talk presented at the 72nd Annual International Meeting.

Feng, Z., and G. T. Schuster, 2017, Elastic least-squares reverse time migration: Geophysics, **82**, S143–S157.

Fletcher, R. P., D. Nichols, R. Bloor, and R. T. Coates, 2016, Least-squares migration-data domain versus image domain using point spread functions: The Leading Edge, **35**, 157–162.

Fomel, S., and A. Guitton, 2006, Regularizing seismic inverse problems by model reparameterization using plane-wave construction: Geophysics, **71**, A43– A47.

Gazdag, J., 1978, Wave equation migration with the phase-shift method: Geo-

physics, **43**, 1342– 1351.

Gazdag, J., and P. Sguazzero, 1984, Migration of seismic data by phase shift plus interpolation: Geophysics, **49**, 124– 131.

Gelius, L.-J., I. Lecomte, and H. Tabti, 2002, Analysis of the resolution function in seismic prestack depth imaging: Geophys. Prosp., **50**, 505– 515.

Gray, S. H., 2013, Spatial sampling, migration aliasing, and migrated amplitudes: Geophysics, **78**, S157–S164.

Gray, S. H., J. Etgen, J. Dellinger, and D. Whitmore, 2001, Seismic migration problems and solutions: Geophysics, **66**, 1622–1640.

Griffiths, M., J. Hembd, and H. Prigent, 2011, Applications of interbed multiple attenuation: The Leading Edge, **30**, 906–912.

Guitton, A., 2004, Amplitude and kinematic corrections of migrated images for nonunitary imaging operators: Geophysics, **69**, 1017– 1024.

———, 2005, Multiple attenuation in complex geology with a pattern-based approach: Geophysics, **70**, V97–V107.

Hargreaves, N., 1994, 432, *in* Wavelet estimation via fourth-order cumulants: SEG, 1588–1590.

Harikumar, G., and Y. Bresler, 1999, Perfect blind restoration of images blurred by multiple filters: Theory and efficient algorithms: IEEE Transaction on Image Processing, **8**, 202–219.

Herrera, R. H., and M. van der Baan, 2012, Short-time homomorphic wavelet estimation: Journal of Geophysics and Engineering, **9**, 674–680.

Herrmann, F., C. Brown, Y. Erlangga, and P. Moghaddam, 2009, Curvelet-based migration preconditioning and scaling: GEOPHYSICS, **74**, A41–A46.

Herron, D. A., 2000, Pitfalls in seismic interpretation: Depth migration artifacts: The Leading Edge, **19**, 1016–1017.

Hestenes, M. R., and E. Stiefel, 1952, Methods of Conjugate Gradients for Solving Linear Systems: Journal of Research of the National Bureau of Standards, **49**, 409–436.

Hu, J., G. T. Schuster, and P. A. Valasek, 2001, Poststack migration deconvolution: Geophysics, **66**, 939– 952.

Huang, Y., R. Nammour, and W. Symes, 2016, Flexibly preconditioned extended least-squares migration in shot-record domain: Geophysics, **81**, S299–S315.

Kaaresen, K., and T. Taxt, 1998, Multichannel blind deconvolution of seismic signals: Geophysics, **63**, no. 6, 2093–2107.

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.

Kazemi, N., and M. Sacchi, 2014a, Filter-based least squares wave equation shot profile migration: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, We P02 03.

———, 2014b, Sparse multichannel blind deconvolution: Geophysics, **79**, V143–V152.

———, 2015, Block row recursive least-squares migration: Geophysics, **80**, A95–A101.

Kuhl, H., and M. D. Sacchi, 2003, Least-squares wave-equation migration for AVP/AVA inversion: Geophysics, **68**, 262– 273.

Lecomte, I., 2008, Resolution and illumination analyses in PSDM : A raybased approach: The Leading Edge, **27**, 650– 663.

Lee, S. I., H. Lee, P. Abbeel, and A. Y. Ng, 2006, Efficient l1 regularized logistic regression: AAAI, 401–408.

Levy, S., and D. Oldenburg, 1987, Automatic phase correction of common midpoint stacked data: GEOPHYSICS, **52**, no. 1, 51–59.

Li, Y., Y. Zhang, and J. Claerbout, 2012, Hyperbolic estimation of sparse models from erratic data: Geophysics, **77**, no. 1, V1–V9.

Liu, J., and H. S. Malvar, 2001, Blind deconvolution of reverberated speech signals via regularization: **5**, 3037 –3040.

Longbottom, J., A. T. Walden, and R. E. White, 1988, Principles and application of maximum kurtosis phase estimation: Geophysical Prospecting, **36**, 115–138.

Marfurt, K., and C. Shin, 1989, *in* The future of iterative modeling in geophysical exploration, in E. Eisner, Ed., Supercomputers in seismic exploration: Pergamon Press, Inc, 203– 228.

Mazzucchelli, P., and U. Spagnolini, 2001, Least squares multichannel deconvolution: EAGE 63rd Conference and Technical Exhibition.

McMechan, G. A., 1983, Migration by extrapolation of time-dependent boundary values: Geophysics, **31**, 413– 420.

Mendel, J. M., 1983, Optimal seismic deconvolution: An estimation based approach: Academic Press.

Mora, P., 1987, Nonlinear two-dimensional elastic inversion of multioffset seismic data: Geophysics, **52**, 1211–1228.

Mulder, W. A., and R. E. Plessix, 2004a, A comparison between one-way and two-way wave-equation migration: Geophysics, **69**, 1491– 1504.

———, 2004b, A comparison between one-way and two-way wave-equation migration: Geophysics, **69**, 1491–1504.

———, 2004c, How to choose a subset of frequencies in frequency domain finite-difference migration: Geophysical Journal International, **158**, 801– 812.

Murray, R. M., Z. Li, and S. S. Sastry, 1994, A mathematical introduction to robotic manipulation: CRC Press.

Naghizadeh, M., and M. Sacchi, 2009, f-x adaptive seismic-trace interpolation: Geophysics, **74**, V9–V16.

Nammour, R., and W. Symes, 2009, Approximate constant density acoustic inverse scattering using dip-dependent scaling: SEG Technical Program Expanded Abstracts 2009, 2347–2351.

Naoshi, A., and G. T. Schuster, 2009, Fast least-squares migration with a deblurring filter: Geophysics, **74**, WCA83– WCA93.

Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, 208– 221.

Ng, M., and R. Plemmons, 1996, Fast recursive least squares adaptive filtering by fast fourier transform-based conjugate gradient iterations: SIAM Journal on Scientific Computing, **17**, 920–941.

Nickerson, W. A., T. Matsuoka, and T. J. Ulrych, 1986, Optimum-lag minimum-entropy deconvolution: SEG Annual Meeting.

Ooe, M., and T. J. Ulrych, 1979, Minimum entropy deconvolution with an exponential transformation: Geophysical Prospecting, **27**, 458–473.

Oppenheim, A., G. Kopec, and J. Tribolet, 1976, Signal analysis by homomorphic prediction: IEEE Transactions on Acoustics, Speech and Signal Processing, **24**, 327 – 332.

Oppenheim, A., and R. Schafer, 1968, Homomorphic analysis of speech: IEEE Transactions on Audio and Electroacoustics, **16**, 221 – 226.

Otis, R., and R. Smith, 1977, Homomorphic deconvolution by log spectral averaging: Geophysics, **42**, no. 6, 1146–1157.

Plessix, R. E., and W. A. Mulder, 2004, Frequency-domain finite-difference amplitude-preserving migration: Geophysical Journal International, **157**, 975–

987.

Prucha, M. L., B. L. Biondi, and W. W. Symes, 1999, Angle-domain common image gathers by wave-equation migration, *in* SEG Technical Program Expanded Abstracts 1999: Society of Exploration Geophysicists, 824–827.

Rakesh, 1988, A linearized inverse problem for the wave equation: Communications in Partial Differential Equations, **13**, 573– 601.

Regalia, P., 1994, Adaptive iir filtering in signal processing and control: CRC Press.

Ren, Z., Y. Liu, and M. K. Sen, 2017, Least-squares reverse time migration in elastic media: Geophysical Journal International, **208**, 1103–1125.

Ribodetti, A., S. Operto, W. Agudelo, J.-Y. Collot, and J. Virieux, 2011, Joint ray + born least-squares migration and simulated annealing optimization for target-oriented quantitative seismic imaging: Geophysics, **76**, R23–R42.

Rickett, J., 2003, Illumination-based normalization for wave-equation depth migration: Geophysics, **68**, 1371–1379.

Rietsch, E., 1997a, Euclid and the art of wavelet estimation, part i: Basic algorithm for noise-free data: Geophysics, **62**, no. 6, 1931–1938.

———, 1997b, Euclid and the art of wavelet estimation, part ii: Robust algorithm and field-data examples: Geophysics, **62**, no. 6, 1939–1946.

Robinson, E., 1967, Predictive decomposition of time series with application to seismic exploration: Geophysics, **32**, no. 3, 418–484.

Robinson, E., and S. Treitel, 1964, Principles of digital filtering: Geophysics, **29**, no. 3, 395–404.

Robinson, E. A., and S. Treitel, 1980, Geophysical signal analysis: Prentice-Hall, Inc.

Royer, A., M. Bostock, and E. Haber, 2012, Blind deconvolution of seismograms regularized via minimum support: Inverse Problems, **28**, 125010 – 125027.

Sacchi, M., and T. Ulrych, 2000, Nonminimum-phase wavelet estimation using higher order statistics: The Leading Edge, **19**, no. 1, 80–83.

Sacchi, M., D. Velis, and A. Comínguez, 1994, Minimum entropy deconvolution with frequency-domain constraints: Geophysics, **59**, no. 6, 938–945.

Sacchi, M., D. R. Velis, and T. J. Ulrych, 1996, *in* Wavelets via Polycepstra: SEG, 1583–1586.

Salomons, B., M. Kiehn, J. Sheiman, B. Strawn, and F. Ten Kroode, 2014, High fidelity imaging with least squares migration: Presented at the 76th EAGE Conference and Exhibition 2014.

Scales, J. A., 1987, Tomographic inversion via the conjugate gradient method: Geophysics, **52**, 179–185.

Schleicher, J., J. C. Costa, and A. Novais, 2008, A comparison of imaging conditions for wave-equation shot-profile migration: GEOPHYSICS, **73**, S219–S227.

Schmidt, M., G. Fung, and R. Rosales, 2007, Fast optimization methods for l1 regularization: A comparative study and two new approaches, *in* Machine Learning: ECML 2007: Springer Berlin Heidelberg, 286–297.

Schneider, W. A., 1978, Integral formulation for migration in two and three dimensions: Geophysics, **43**, 49– 76.

Shalvi, O., and E. Weinstein, 1990, New criteria for blind deconvolution of nonminimum phase systems (channels): IEEE Transactions on Information Theory, **36**, 312 –321.

Sheriff, R. E., 2002, Encyclopedic dictionary of applied geophysics, fourth ed.: Society of Exploration Geophysicists.

Sheriff, R. E., and L. P. Geldart, 1995, Exploration seismology: Cambridge university press.

Shewchuk, J. R., et al., 1994, An introduction to the conjugate gradient method without the agonizing pain.

Sjoeberg, T. A., L. Gelius, and I. Lecomte, 2003, 2d deconvolution of seismic image blur: SEG Technical Program Expanded Abstracts, 1055–1058.

Sroubek, F., and P. Milanfar, 2012, Robust multichannel blind deconvolution via fast alternating minimization: IEEE Transacton on Image Processing, **21**, 1687 – 1700.

Stearns, S. D., 1985, of aldapfive signal processing.

Stoffa, P., P. Buhl, and G. Bryan, 1974, The application of homomorphic deconvolution to shallow water marine seismology Part I: Models: Geophysics, **39**, no. 4, 401–416.

Stoffa, P. L., J. T. Fokkema, M. de Luna Freir, and W. P. Kessinger, 1990, Split-step fourier migration: Geophysics, **55**, 410– 421.

Stogioglou, A., S. McLaughlin, and A. Ziolkowski, 1996, Asymptotic performance analysis for fourth-order cumulant based deconvolution: SEG Technical Program Expanded Abstracts, **422**, 1587–1590.

Stolk, C. C., 2000, Microlocal analysis of a seismic linearized inverse problem: Wave Motion, **32**, 267 – 290.

Stolt, R. H., 1978, Migration by fourier transform: Geophysics, **43**, 23– 48.

Symes, W. W., 2008, Approximate linearized inversion by optimal scaling of prestack depth migration: Geophysics, **73**, R23– R35.

ten Kroode, A. P. E., D.-J. Smit, and A. Verdel, 1998, A microlocal analysis of migration: Wave Motion, **28**, 149 – 172.

Toxopeus, G., J. Thorbecke, K. Wapenaar, S. Petersen, E. Slob, and J. Fokkema, 2008, Simulating migrated and inverted seismic data by filtering a geologic model: Geophysics, **73**, T1– T10.

Ulrych, T., D. Velis, and M. D. Sacchi, 1995, Wavelet estimation revisited: The Leading Edge, **14**, no. 11, 1139–1143.

Ulrych, T. J., 1971, Application of homomorphic deconvolution to seismology: Geophysics, **36**, no. 6, 650–660.

VanDecar, J. C., and R. D. Crosson, 1990, Determination of tele seismic relative phase arrival tomes using multi-channel cross-correlation and least squares: Bulletin of the Seismological Society of America, **80**, 150–169.

Velis, D., and T. Ulrych, 1996, Simulated annealing wavelet estimation via fourth-order cumulant matching: Geophysics, **61**, no. 6, 1939–1948.

Verschuur, D. J., and R. J. Prein, 1999, Multiple removal results from delft university: The Leading Edge, **18**, no. 1, 86–91.

Vestrum, R., and K. Muenzer, 1997, Anisotropic depth imaging below dipping shales: Presented at the CSEG 1997 National Meeting.

Walden, A. T., 1985, Non-gaussian reflectivity, entropy, and deconvolution: Geophysics, **50**, no. 12, 2862–2888.

Wang, J., and M. D. Sacchi, 2006, High-resolution wave-equation amplitude-variation-with-ray-parameter (avp) imaging with sparseness constraints: Geophysics, **72**, S11–S18.

Weatherburn, C. E., 1949, A first course in mathematical statistics: Cambridge University Press.

Weglein, A. B., S.-Y. Hsu, P. Terenghi, X. Li, and R. H. Stolt, 2011, Multiple attenuation: Recent advances and the road ahead (2011): The Leading Edge, **30**, 864–875.

Wei, D., C. Boonyasiriwat, and G. T. Schuster, 2010, 611, *in* 3D Multisource Least-squares Reverse Time Migration: 3120–3124.

White, R. E., 1988, Maximum kurtosis phase correction: Geophysical Journal, **95**, no. 2.

Whitmore, N. D., 1983, Iterative depth migration by backward time propagation:

53rd Ann. Internat. Mtg., Soc. Expl. Geophys, 382– 385.

Wiggins, R. A., 1978, Minimum entropy deconvolution: Geoexploration, **16**, 21–35.

Xu, G., H. Liu, L. Tong, and T. Kailath, 1995, A least-squares approach to blind channel identification: EEE Transactions on Signal Processing, **43**, 2982 –2993.

Xu, Y., P. Thore, and S. Duchenne, 2012, The reliability of the kurtosis-based wavelet estimation: SEG Technical Program Expanded Abstracts, **493**, 1–5.

Yu, J., J. Hu, G. T. Schuster, and R. Estill, 2006, Prestack migration deconvolution: Geophysics, **71**, S53– S62.

Ziolkowski, A., 1991, Why don't we measure seismic signatures?: Geophysics, **56**, 190–201.

APPENDIX A

Generalized Conjugate Gradient algorithm

A.1 Generalized Conjugate Gradient algorithm

In this section we show how one can use Conjugate Gradient algorithm to solve

$$J = ||\mathbf{Am} - \mathbf{d}||_2^2 + \mu\, ||\mathbf{m}||_2^2. \tag{A.1}$$

Following Hestenes and Stiefel (1952) and Scales (1987) we know how to solve $||\underset{\sim}{\mathbf{A}}\mathbf{x} - \underset{\sim}{\mathbf{d}}||_2^2$. To be able to take advantage of the same algorithm, we concatenate the operators in equation A.1 so that the new cost function resembles the conventional least-squares minimization

$$J = ||\begin{bmatrix}\mathbf{A}\\ \sqrt{\mu}\end{bmatrix}\mathbf{x} - \begin{bmatrix}\mathbf{d}\\ \mathbf{0}\end{bmatrix}||_2^2 = ||\underset{\sim}{\mathbf{A}}\mathbf{x} - \underset{\sim}{\mathbf{d}}||_2^2, \tag{A.2}$$

where $\underset{\sim}{\mathbf{A}} = \begin{bmatrix}\mathbf{A}\\ \sqrt{\mu}\end{bmatrix}$ and $\underset{\sim}{\mathbf{d}} = \begin{bmatrix}\mathbf{d}\\ \mathbf{0}\end{bmatrix}$.

After doing so, the Generalized Conjugate Gradient algorithm for cost function of equation A.2 reads

where $< *, * >$ means inner product. The worst convergence of Conjugate Gradient algorithm defined as

$$||\mathbf{e}_k||_{\underset{\sim}{\mathbf{A}}} \leq 2(\frac{\sqrt{C}-1}{\sqrt{C}+1})^k ||\mathbf{e}_0||_{\underset{\sim}{\mathbf{A}}}, \tag{A.3}$$

Algorithm 1 Generalized Conjugate Gradient algorithm

choose $\mathbf{y}_0$,
Put $\mathbf{s}_0 = \underset{\sim}{\mathbf{d}} - \underset{\sim}{\mathbf{A}}\mathbf{x}_0 = \begin{bmatrix} \mathbf{d} - \mathbf{A}\mathbf{x}_0 \\ -\sqrt{\mu}\ \mathbf{x}_0 \end{bmatrix}$,
Set $\mathbf{r}_0 = \mathbf{p}_0 = \underset{\sim}{\mathbf{A}}^T(\underset{\sim}{\mathbf{d}} - \underset{\sim}{\mathbf{A}}\mathbf{x}_0) = \mathbf{A}^{-T}(\mathbf{b} - \mathbf{A}\mathbf{x}_0) - \mu\mathbf{x}_0$,
Set $\mathbf{q}_0 = \underset{\sim}{\mathbf{A}}\mathbf{x}_0$,
Initialize iteration $k = 0$,
while $(||\underset{\sim}{\mathbf{A}}\mathbf{x} - \underset{\sim}{\mathbf{d}}||_2^2 > tol)$ **do**,
  $\alpha_{k+1} = < \mathbf{r}_k, \mathbf{r}_k > / < \mathbf{q}_k, \mathbf{q}_k >$,
  $\mathbf{y}_{k+1} = \mathbf{x}_k + \alpha_{k+1}\ \mathbf{p}_k$,
  $\mathbf{s}_{k+1} = \mathbf{s}_k - \alpha_{k+1}\ \mathbf{q}_k$,
  $\mathbf{r}_{k+1} = \underset{\sim}{\mathbf{A}}^T\mathbf{s}_{k+1}$,
  $\beta_{k+1} = < \mathbf{r}_{k+1}, \mathbf{r}_{k+1} > / < \mathbf{r}_k, \mathbf{r}_k >$,
  $\mathbf{p}_{k+1} = \mathbf{r}_{k+1} + \beta_{k+1}\mathbf{p}_k$,
  $\mathbf{q}_{k+1} = \underset{\sim}{\mathbf{A}}\mathbf{p}_{k+1}$,
  $k \leftarrow k + 1$,
end while
Set $\mathbf{x} = \mathbf{x}_k$,

where $\mathbf{e}_k$ is residual at k^{th} iteration, $||*||_{\underset{\sim}{\mathbf{A}}} = \underset{\sim}{\mathbf{A}}^T * \underset{\sim}{\mathbf{A}}$ and C is the condition number of $\underset{\sim}{\mathbf{A}}^T\underset{\sim}{\mathbf{A}}$ matrix.

APPENDIX B

Convergence of Conjugate Gradient algorithm

B.1 Convergence of Conjugate Gradient algorithm

Consider a linear system of equations with a data matrix $\mathbf{A}$ of $n \times n$. Here we will solve the least-squares minimization problem represented by equation A.1. Theoretically, the number of iterations which Conjugate Gradient algorithm takes till convergence is equivalent to the number of non-zero and distinct eigenvalues of that matrix. However, in real world applications with big sparse data matrices, nobody runs the algorithm till convergence. In practical problems, one can show the convergence behaviour of Conjugate Gradient algorithm by probing the properties of *Krylov subspace*. *Krylov subspace* for CG algorithm reads

$$\begin{aligned} \mathcal{D}_i &= \{\mathbf{d}_0, \mathbf{A}\mathbf{d}_0, \mathbf{A}^2\mathbf{d}_0, \ldots, \mathbf{A}_{i-1}\mathbf{d}_0\}, \\ \mathcal{D}_i &= \{\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, \mathbf{A}^2\mathbf{r}_0, \ldots, \mathbf{A}_{i-1}\mathbf{r}_0\}. \end{aligned} \tag{B.1}$$

One can show that the error term has the form

$$\mathbf{e}_i = (\mathbf{I} + \Sigma_{p=1}^{i} \Theta_p \mathbf{A}^p)\mathbf{e}_0, \tag{B.2}$$

where Θ_p coefficients are related to β_i and γ_i terms. One can represent the Θ_p coefficients as polynomials. In this case, if the eigenvalue spectrum of the data matrix is in set $\lambda \in \Lambda$ and we use polynomials $\mathbf{P}_i(\lambda)$, then the convergence of CG

reads

$$||\mathbf{e}_i||_{\mathbf{A}} \leq \underset{\mathbf{P}_i}{\text{argmin}} \quad \underset{\lambda \in \Lambda}{\text{argmax}} \quad \mathbf{P}_i(\lambda)^2 ||\mathbf{e}_0||_{\mathbf{A}}, \tag{B.3}$$

which shows that Conjugate Gradient finds the polynomials in such a way that inequality B.3 is minimized. For example, consider the well known *Chebyshev polynomials* that can minimize equation B.3 over the eigenvalue spectrum of data matrix (i.e., $[\lambda_{min} \quad \lambda_{max}]$). The *Chebyshev polynomials* of degree i can be expressed as

$$Q_i(\omega) = \frac{1}{2}[(\omega + \sqrt{\omega^2 + 1})^i + (\omega - \sqrt{\omega^2 - 1})^i]. \tag{B.4}$$

It can be shown that if we choose polynomials of equation B.3 as

$$\mathbf{P}_i(\lambda) = \frac{Q_i(\frac{\lambda_{min}+\lambda_{max}-2\lambda}{\lambda_{max}-\lambda_{min}})}{Q_i(\frac{\lambda_{min}+\lambda_{max}}{\lambda_{max}-\lambda_{min}})}, \tag{B.5}$$

then the convergence of CG reads

$$\begin{aligned} ||\mathbf{e}_i||_{\mathbf{A}} &\leq Q_i(\frac{\lambda_{min} + \lambda_{max}}{\lambda_{max} - \lambda_{min}})^{-1} ||\mathbf{e}_0||_{\mathbf{A}}, \\ &= Q_i(\frac{C+1}{C-1})^{-1} ||\mathbf{e}_0||_{\mathbf{A}}, \\ &= 2[(\frac{\sqrt{C}+1}{\sqrt{C}-1})^i + (\frac{\sqrt{C}-1}{\sqrt{C}+1})^i]^{-1} ||\mathbf{e}_0||_{\mathbf{A}}, \end{aligned} \tag{B.6}$$

where C is condition number of data matrix. Moreover, by considering the fact that the term $(\frac{\sqrt{C}-1}{\sqrt{C}+1})^i$ goes to zero as iteration number grows, we finally get

$$||\mathbf{e}_i||_{\mathbf{A}} \leq 2(\frac{\sqrt{C}-1}{\sqrt{C}+1})^i ||\mathbf{e}_0||_{\mathbf{A}}. \tag{B.7}$$

It is easy to show that if the eigenvalues of matrix $\mathbf{A}$ is clustered then CG algorithm will easily find the polynomials. Figure B.1 shows Conjugate Gradient convergence as a function of condition number.

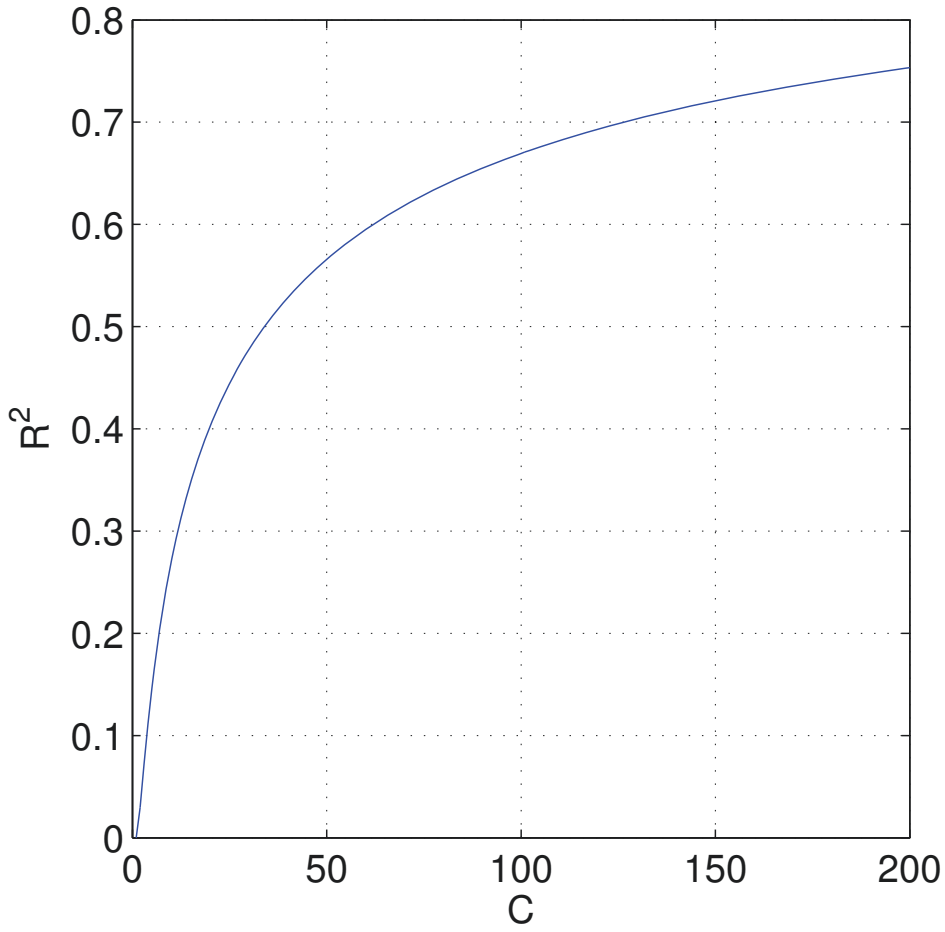

Figure B.1: Worst convergence behaviour of Conjugate Gradient algorithm, where $R^2 = (\frac{\sqrt{C}-1}{\sqrt{C}+1})^2$.

APPENDIX C

Generalized Preconditioned Conjugate Gradient algorithm

C.1 Generalized Preconditioned Conjugate Gradient algorithm

In this section we show how one can use Conjugate Gradient algorithm to solve

$$J = ||\mathbf{A}\mathbf{P}^{-1}\mathbf{y} - \mathbf{d}||_2^2 + \mu\, ||\mathbf{P}^{-1}\mathbf{y}||_2^2, \tag{C.1}$$

where $\mathbf{m} = \mathbf{P}^{-1}\mathbf{y}$, and $\mathbf{P}$ is an $M \times M$ preconditioner. Following Hestenes and Stiefel (1952) and Scales (1987) we know how to solve the conventional least squares minimization problem (i.e., $||\underset{\sim}{\mathbf{A}}\mathbf{x} - \underset{\sim}{\mathbf{d}}||_2^2$). To take advantage of the same algorithm, we concatenate the operators in equation C.1 so that the new cost function resembles the conventional least-squares minimization

$$J = ||\begin{bmatrix} \mathbf{A}\mathbf{P}^{-1} \\ \sqrt{\mu}\,\mathbf{P}^{-1} \end{bmatrix}\mathbf{y} - \begin{bmatrix} \mathbf{d} \\ \mathbf{0} \end{bmatrix}||_2^2 = ||\underset{\sim}{\mathbf{A}}\mathbf{y} - \underset{\sim}{\mathbf{d}}||_2^2, \tag{C.2}$$

where $\underset{\sim}{\mathbf{A}} = \begin{bmatrix} \mathbf{A}\mathbf{P}^{-1} \\ \sqrt{\mu}\,\mathbf{P}^{-1} \end{bmatrix}$ and $\underset{\sim}{\mathbf{d}} = \begin{bmatrix} \mathbf{d} \\ \mathbf{0} \end{bmatrix}$. Note that $\mathbf{x} = \mathbf{P}^{-1}\mathbf{y}$.

After doing so, the generalized Preconditioned Conjugate Gradient algorithm for cost function of equation C.2 reads

Algorithm 2 Generalized Preconditioned Conjugate Gradient algorithm

choose $\mathbf{y}_0$,
Put $\mathbf{s}_0 = \underset{\sim}{\mathbf{d}} - \underset{\sim}{\mathbf{A}}\mathbf{y}_0 = \begin{bmatrix} \mathbf{d} - \mathbf{A}\mathbf{P}^{-1}\mathbf{y}_0 \\ -\sqrt{\mu}\,\mathbf{P}^{-1}\mathbf{y}_0 \end{bmatrix}$,
Set $\mathbf{r}_0 = \mathbf{p}_0 = \underset{\sim}{\mathbf{A}}^T(\underset{\sim}{\mathbf{d}} - \underset{\sim}{\mathbf{A}}\mathbf{y}_0) = \mathbf{P}^{-T}\mathbf{A}^{-T}(\mathbf{b} - \mathbf{A}\mathbf{P}^{-1}\mathbf{y}_0) - \mu\mathbf{P}^{-T}\mathbf{P}^{-1}\mathbf{y}_0$,
Set $\mathbf{q}_0 = \underset{\sim}{\mathbf{A}}\mathbf{p}_0$,
Initialize iteration $k = 0$,
while $(||\underset{\sim}{\mathbf{A}}\mathbf{y} - \underset{\sim}{\mathbf{d}}||_2^2 < tol)$ **do**,
  $\alpha_{k+1} = < \mathbf{r}_k, \mathbf{r}_k > / < \mathbf{q}_k, \mathbf{q}_k >$,
  $\mathbf{y}_{k+1} = \mathbf{y}_k + \alpha_{k+1}\,\mathbf{p}_k$,
  $\mathbf{s}_{k+1} = \mathbf{s}_k - \alpha_{k+1}\,\mathbf{q}_k$,
  $\mathbf{r}_{k+1} = \underset{\sim}{\mathbf{A}}^T\mathbf{s}_{k+1}$,
  $\beta_{k+1} = < \mathbf{r}_{k+1}, \mathbf{r}_{k+1} > / < \mathbf{r}_k, \mathbf{r}_k >$,
  $\mathbf{p}_{k+1} = \mathbf{r}_{k+1} + \beta_{k+1}\mathbf{p}_k$,
  $\mathbf{q}_{k+1} = \underset{\sim}{\mathbf{A}}\mathbf{p}_{k+1}$,
  $k \leftarrow k + 1$,
end while
Set $\mathbf{y} = \mathbf{y}_k$,
Calculate $\mathbf{x} = \mathbf{P}^{-1}\mathbf{y}$.

where $< *, * >$ means inner product. The worst convergence of Conjugate Gradient algorithm defined as

$$||\mathbf{e}_k||_{\underset{\sim}{\mathbf{A}}} \leq 2(\frac{\sqrt{C} - 1}{\sqrt{C} + 1})^k ||\mathbf{e}_0||_{\underset{\sim}{\mathbf{A}}}, \tag{C.3}$$

where $\mathbf{e}_k$ is residual at k^{th} iteration, $||*||_{\underset{\sim}{\mathbf{A}}} = \underset{\sim}{\mathbf{A}}^T * \underset{\sim}{\mathbf{A}}$ and C is the condition number of $\underset{\sim}{\mathbf{A}}^T\underset{\sim}{\mathbf{A}}$ matrix.

Comparing the concatenated versions of equations A.1 and C.1, it is easy to show that if the condition number of $\mathbf{P}^{-T}\mathbf{A}^T\mathbf{A}\mathbf{P}^{-1} + \mu\mathbf{P}^{-T}\mathbf{P}^{-1}$ matrix is smaller than $\mathbf{A}^T\mathbf{A} + \mu\mathbf{I}$, then the convergence rate of the Preconditioned Conjugate Gradient algorithm will be faster than Conjugate Gradient algorithm.

APPENDIX D

Convergence of SMBD algorithm

D.1 Curvilinear line search on the unit sphere

This section proves that curvilinear line search on the unit sphere is equal to the equation (5.18). In our problem, we need a rotation matrix that preserves the sparse solution on the unit sphere. This is equivalent to have a rotation matrix $\mathcal{R}_k$

$$\mathbf{x}_{k+1} = \mathcal{R}_k \, \mathbf{x}_k, \tag{D.1}$$

such that $||\mathbf{x}_{k+1}||_2 = ||\mathbf{x}_k||_2 = 1$. Using angle axis representation of rotation matrix via aka Rodrigues' formula, we have

$$\mathcal{R}_k \, \mathbf{x}_k = \mathbf{x}_k + sin(\theta_k) \, (\mathbf{r}_k \times \mathbf{x}_k) + (1 - cos(\theta_k)) \, (\mathbf{r}_k \mathbf{r}_k^T - \mathbf{I}) \, \mathbf{x}_k, \tag{D.2}$$

where $\mathbf{r}_k$ is rotation axis, θ_k is rotation angle, $\mathbf{I}$ is identity matrix and k is iteration. We need to choose $\mathbf{r}_k$ in such a way that it lets the search go in the direction of the projection of the gradient into the tangent plane on the sphere. Hence, the only option is $\mathbf{r}_k = \mathbf{x}_k \times \mathbf{h}_k$ or in other words

$$\mathbf{h}_k = \mathbf{r}_k \times \mathbf{x}_k, \tag{D.3}$$

where $\mathbf{h}_k$ is the normalized projected gradient on the sphere. Note that the length of $\mathbf{r}_k$, $\mathbf{x}_k$ and $\mathbf{h}_k$ are equal to one. Now this is easy to show that rotation axis is

orthogonal to the current solution. Hence, we have

$$\mathbf{r}_k^T \mathbf{x}_k = \mathbf{0}. \tag{D.4}$$

Inserting equations (D.3) and (D.4) into the equation (D.2) yields

$$\mathcal{R}_k \, \mathbf{x}_k = sin(\theta_k) \, \mathbf{h}_k + cos(\theta_k) \, \mathbf{x}_k, \tag{D.5}$$

as it is required.

D.2 Convergence behaviour of SMBD algorithm

This section proves the convergence of the proposed steepest descent method. For $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{m}_k$ as long as $-\mathbf{g}_k^T \mathbf{m}_k > 0$ it is gradient descent. $\mathbf{g}_k$ is the gradient of the cost function at iteration k. In gradient descent the step size should be small enough to guarantee the convergence. We will use the same concept to prove the convergence of the proposed technique.

Again, consider the angle axis representation of rotation matrix via aka Rodrigues' formula

$$\mathcal{R}_k \, \mathbf{x}_k = \mathbf{x}_k + sin(\theta_k) \, (\mathbf{r}_k \times \mathbf{x}_k) + (1 - cos(\theta_k)) \, (\mathbf{r}_k \mathbf{r}_k^T - \mathbf{I}) \, \mathbf{x}_k, \tag{D.6}$$

by analogy we have

$$\mathbf{m}_k = sin(\theta_k) \, (\mathbf{r}_k \times \mathbf{x}_k) + (1 - cos(\theta_k)) \, (\mathbf{r}_k \mathbf{r}_k^T - \mathbf{I}) \, \mathbf{x}_k. \tag{D.7}$$

It is very interesting that unlike conventional steepest descent, the evolving direction depends on the step size θ_k. Now, we need to check if $-\mathbf{g}_k^T \mathbf{m}_k > 0$. We should emphasize that the steepest descent is valid only for sufficiently small step sizes. Assuming small angles, we have $sin(\theta) \approx \theta$ and $(1 - cos(\theta)) \approx 0$. Hence, equation (D.7) simplifies to

$$\mathbf{m}_k \approx \theta_k \, (\mathbf{r}_k \times \mathbf{x}_k) = \theta_k \, \mathbf{h}_k. \tag{D.8}$$

Obviously $-\mathbf{g}_k^T \, \theta_k \, \mathbf{h}_k > 0$. Hence it is gradient descent. It is worth mentioning that to satisfy the constraint we used $\mathbf{m}_k = sin(\theta_k) \, (\mathbf{r}_k \times \mathbf{x}_k) + (1 - cos(\theta_k)) \, (\mathbf{r}_k \mathbf{r}_k^T - \mathbf{I}) \, \mathbf{x}_k$

but to prove the convergence of the proposed steepest descent we have $\mathbf{m}_k \approx \theta_k \, \mathbf{h}_k$.