MAPPING SUBSURFACE STRUCTURES BY LEAST-SQUARES INVERSION OF SEISMIC DATA

A Dissertation

Presented to

the Faculty of the Department of Earth and Atmospheric Sciences

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Wei Huang

December 2015

MAPPING SUBSURFACE STRUCTURES BY LEAST-SQUARES INVERSION OF SEISMIC DATA

Wei Huang

APPROVED:

Hua-wei Zhou, Chairman
Department of Earth and Atmospheric Sciences

Aibing Li
Department of Earth and Atmospheric Sciences

Yingcai Zheng
Department of Earth and Atmospheric Sciences

Jonathan Liu
Exxon Mobil

Dean, College of Natural Sciences and Mathematics

Acknowledgement

I would like to express my deepest gratitude to my advisor, Dr. Hua-Wei Zhou, for his guidance, support and encouragement throughout this work. I draw constant inspiration from his superb teaching skills and his unbridled passion for research. I also thank the other members of my dissertation committee, Dr. Aibing Li, Dr. Yingcai Zheng, and Dr. Jonathan Liu, for their valuable time and suggestions during the course of the work.

I am very grateful for the financial support from ION Geophysical. I would also thank all my colleagues in ION Geophysical for their help and fruitful discussions, especially Dr. Guoquan Chen, Dr. David Yingst and Paul Farmer. Special thanks to my whole family for supporting and cheering me up. Words can not express how grateful I am to my beloved wife, Yan Huai, who consistently encourages me throughout this experience.

MAPPING SUBSURFACE STRUCTURES BY LEAST-SQUARES INVERSION OF SEISMIC DATA

An Abstract of a Dissertation
Presented to
the Faculty of the Department of Earth and Atmospheric Sciences
University of Houston

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

By
Wei Huang
December 2015

Abstract

Accurate mapping of subsurface structure through seismic techniques is essential in oil and gas exploration. With the development of computational power, there has been an increased focus on data-fitting related seismic-inversion techniques for producing high fidelity seismic velocity models and images, such as full-waveform inversion and least-squares migration. However, more advanced methods, such as data-fitting techniques, are generally formulated in least-squares optimization, and can be less robust and expensive in terms of computational cost. The nonlinearity of inversion problems also pose another issue for successful mapping of subsurface structure. Recently, various techniques to optimize data-fitting seismic-inversion problems have been implemented for the industrial need to better efficiency.

The primary objective of this study is to optimize least-squares techniques for seismic-velocity model building and imaging. This work can be divided into three equally important parts. The first part of this work is developing a new multi-level temporal integration to make full-waveform inversion (FWI) more robust than its classic implementation. The second contribution is to maximize the capability of the least-squares migration through numerical optimization and Hessian preconditioning. The third part is to account for the large amplitude differences between field and modeled data. A new local normalization scheme is proposed for better performance of the least-squares migration. The field examples demonstrate the effectiveness of the proposed methods in generating high quality images and improving the inversion efficiency.

Contents

List of Figures

List of Abbreviations

AGC amplitudegain control

CG conjugate gradient

FWI full-waveform inversion

LSM least-squares migration

LSRTM least-squares reverse-time migration

MTI multi-level temporal integration

PSDM prestack depth migration

RTM reverse-time migration

SCG stochastic conjugate gradient

SG stochastic gradient

SNR signal-to-noise ratio

Chapter 1

Introduction

The comprehensive estimation of subsurface structures is essential for oil and gas exploration, and is normally achieved by inverting the geophysical parameters from seismic signals. Seismic-imaging consists of acquiring seismic data and retrieving the subsurface properties, which can be formulated as inversion problems (Bamberger et al., 1982; Tarantola, 1987; Berkhout, 1982). Fast development of oil and gas resources requires more accurate velocity models and subsurface structure maps. With the development of computational power, there has been an increased focus on least-squares fitting seismic-inversion techniques for high-fidelity seismic-velocity model and image, such as full-waveform inversion (FWI) and least-squares migration (LSM), which is the focus of this study. In this chapter, I briefly outline the general concepts and key developments for least-squares inversion and detailed research goals and contributions.

1.1 Background and overview

Pioneer work of using seismic waves to map subsurface structures for the oil and gas industry was done by (Claerbout, 1971), where mapping subsurface reflectors from different imaging conditions were presented. The mapping of subsurface structure using seismic data is generally done using an estimated velocity model and certain migration techniques.

The early velocity model estimations were based on the assumption that the earth is a layer-caked model. However, it was recognized that more accurate velocity models are needed for the oil and gas industry. With the fast development of acquisition techniques, more accurate velocity techniques are in high demand to maximize the capability of seismic data. Ray-based tomography emerged as model-building tools in the 90s, and now the ray-based tomography has become the standard model-building tool for seismic depth imaging (Woodward et al., 2008). The industry has widely adopted the ray-based, prestack depth migration (PSDM) domain tomographies to perform migration velocity estimation (Stork, 1992).

Ray tomography is based on the high frequency approximation of ray paths, and is generally inadequate in high velocity contrast. Ray-based velocity model building is limited by the high-frequency approximation that fails at large velocity variations (Woodward et al., 2008). Recent industrial and academic efforts have shifted from ray-based tomography to wave-equation based velocity inversion techniques. The biggest advantage wave-equation based techniques have over ray-based ones is wave-equations are more accurate in describing the seismic wave propagation in the presence of large velocity variations (Vigh and Starr, 2008). Among the wave-equation based velocity inversion techniques, full waveform inversion (FWI) is the most investigated technique in recent years (Virieux and Operto, 2009). The FWI has been implemented in different domains, including the time domain (Tarantola, 1984; Vigh and Starr, 2008), the frequency domain (Pratt et al., 1998), and the Laplacian domain (Shin et al., 2014; Ha et al., 2012).

Ray-based tomography uses the phase or travel time information to update velocity model, while the full-waveform inversion, proposed by (Tarantola, 1984) and (Lailly, 1983), inverts the velocity through matching both the phase and amplitude between the field and modeled data. Although there have been several successful industrial-sized FWI projects (Huang et al., 2012; Vigh and Starr, 2008), many challenges still exist for real-world FWI practice. First, this minimization is achieved by a non-linear optimization algorithm that updates the model properties based on back-propagating the differences between the real and modeled data through the model itself. While such a minimization will commonly converge to a particular model realization, it is well known that the problem we are trying to solve in FWI is ill-posed. The physical meaning of ill-posed in this case can be characterized as any given set of data residuals for which multiple wave propagation models which fit the data equally-well (Backus and Glbert, 1968; Jackson, 1972). This result suggests that multiple minima can occur.

When properly implemented, FWI is able to achieve high quality seismic velocity. On the other hand, FWI needs a initial velocity model that is very close to the true velocity; otherwise, FWI can easily get into local minimum (Mulder and

Plessix, 2008). To alleviate the requirement for an accurate initial velocity model, low-frequency and large offset dataset are required by FWI. In general, multi-scale approaches are used (Bunks et al., 1995). In real world FWI, only the refraction part of the seismic data is used to update large scale variation (Huang et al., 2013), while the reflection data is generally neglected to avoid cycle-skipping.

One of the other difficulties is that no single acoustic wave equation can represent real-world wave propagation (Zhou et al., 2012). FWI transfers the difference between modeled and observed data to update velocity, which implies that an accurate modeling operator is the key to a successful velocity inversion. However, it is apparent that the propagation of a real seismic waveform is complicated by elastic effects, such as attenuation and anisotropy (Thomsen, 1986; Alkhalifah and Tsvankin, 1995). The amplitude information can also be distorted by data processing, source signature variations, and background noise. All these factors make it challenging to extract the amplitude information from real-world seismic data. In general, the phase information in seismic waveform is directly related to the velocity structure, which makes it more desirable to remove the impacts from amplitude. To partially remove the source signature effects and 'unknown' physical processes that cause the amplitude difference, a number of authors proposed cross-correlation based objective functions (Luo and Schuster, 1990, 1991; Van Leeuwan and Mulder, 2010). A most recent development used adaptive filtering to suppress the amplitude contribution in velocity inversion (Warner and Guasch, 2014).

Although removing the amplitude impact can make FWI more stable, its conventional format is still limited by its migration resolution kernel (Fichtner and Trampert, 2011). The original FWI derivation is only valid for transmitters and refracted waves with large scale velocity errors (Zhou et al., 2012). Realizing the limitation of FWI, some efforts have been put toward the image-domain wave-equation velocity analysis (Sava, 2004; Sava and Biondi, 2004; Albertin et al., 2006; Yang and Sava, 2011; Shen, 2004). Image-domain wave-equation velocity analysis uses image-focusing as the criteria for velocity correctness (Mackay and Abma, 1992). This can have many different focusing measurements. However, the paucity of real-world successful application of wave-equation velocity analysis is partially due to the high computational cost incurred in the iterative velocity updates.

On the other hand, migration of seismic data is a complement tool for validating velocity model, such that, a valid velocity model should result in a well-focused seismic image. With a predefined velocity model, seismic data can be mapped from their surface location to subsurface structures by different migration algorithms (Claerbout, 1971). Early imaging algorithms relied on ray-based methods, and fails when there are large velocity variations. Wave-equation based imaging methods, such as reverse-time migration (RTM) (Baysal et al., 1983), accommodate high velocity variations in real world analysis (Farmer et al., 2009). For example, only RTM can generate high quality images of the high impedance contrast between the salt-dome flank and surrounding reflectors. In general, a high resolution image is greatly desired for geological interpretation; however, a migrated seismic image is limited by the theoretical resolution (Chen and Schuster, 1999; Berkhout, 1984; Safar, 1985). The migrated image, even generated by an accurate velocity model, can be blurred because of the migration point spread effect (Yu et al., 2006). In real practice, data quality and limited aperture can also strongly affect the resolution of the migrated image. In order to deconvolve the source signature effects and migration spreading, imaging through inversion using least-squares migration (LSM), has been shown by various authors to achieve high-quality images (Nemeth et al., 1999; Dai and Schuster, 2013; Zhang et al., 2013; Huang and Zhou, 2014, 2015). Migration artifacts due to limited aperture and irregular sampling are suppressed through the least-squares inversion process. However, for field data applications, it is not easy to match the recorded amplitudes because of the visco-elastic nature of the earth and inaccuracies in the estimation of seismic sources. Objective functions based on cross-correlation have the advantage of being less-sensitive to the source signature and amplitude variation (Dutta et al., 2013; Zhang et al., 2013).

On the computational side, both FWI and LSM share the same challenge, a high computational cost. In the FWI or LSM, wave-equation based seismic simulations are performed for individual sources and the differences between simulated and observed common-shot-gathers drive the updating of the velocity model or reflectivity (Krebs et al., 2009; Dai and Schuster, 2013; Nemeth et al., 1999). Therefore, the cost of FWI and LSM is proportional to the number of common-shot-gathers, which could be prohibitively high for industrial-sized 3D seismic surveys (Huang et al., 2013).

To improve the efficiency of FWI and LSM, numerous numerical schemes have been suggested to reduce the computational cost. In general, current optimizations for least-squares seismic inversion problems can be categorized into several different groups: super-grouping, source encoding, and stochastic optimization. Super-grouping is more of a standard practice in industrial seismic imaging projects (Huang et al., 2013), where recorded shots are moved spatially to be combined into a giant super-shot. Special treatment is required to compensate the spatial change, called partial move-out. This method has limited accuracy and can generate artifacts by moving field records to designated positions. The second major optimization scheme is the source-encoding technique. Instead of moving shots spatially to join several shots together, the source-encoding scheme is to simulate several shots simultaneously with an assigned random time delay for different shots. This method reduces the computational cost for forward simulation (Krebs et al., 2009). Source-encoding schemes can be efficient in most 2D cases. For real-world large 3D surveys, source-encoding may have several issues, including, an increased number of random shots requires propagation in a larger velocity grid; also, the cross-talk artifact between different shots offsets the huge advantage over conventional non-source-encoding methods.

Stochastic optimization has been a popular algorithm for many applications in machine learning (Schraudolph and Graepel, 2003), where stochastic sampling techniques are used to reduce the data volume required for optimization. Recent development in stochastic optimization has received much attention in seismic inversion. FWI with the stochastic optimization has achieved results comparable to that of conventional methods but with only a fraction of conventional shot-by-shot method (Van Leeuwan et al., 2011). Implementations of stochastic optimization for seismic inversion were also presented by several other authors (Huang and Zhou, 2015, 2014; Wang et al., 2014; Gao et al., 2010).

1.2 Motivation and objective of this study

Different kinds of wave-equation velocity inversion and imaging techniques have been implemented using different methods, and are generally in the form of least-squares inversion. Least-squares inversion based techniques use iterative optimization of a specific objective function. Iteration is typically performed using a local gradient or a steepest descent method to minimize an objective function. An objective function is generally defined as the the least-squares misfit between observed and modeled data. Because of the nature of this objective function, it is well known that multiple minima can occur. In general, least-squares velocity model-building and imaging are not as robust as conventional ray-based techniques and can be computationally expensive for industrial-sized projects.

The major motivation of this research is to reduce the nonuniqueness in solving ill-posed inverse problems and to improve the stability and efficiency of the least-squares inversion method to seismic data. The specific motivation of this study is trying to understand and solve the following problems:

- FWI in its classic form have been proved to be unstable for real world applications, and requires improved efficiently under classic FWI implementation.

- Least-squares migrations (LSM) is highly computational demanding. How can the efficiency of LSM be improved to make it applicable for practical application?

- Real-world least-squares imaging poses more challenges due to the elastic effects in the wave propagation. How can the amplitude variations from the elastic effects be effciently compensated?

Motivated by above mentioned questions, this thesis is organized as follows:

- While reflection FWI is able to provide more a stable result, it is much more computational demanding. FWI, in its classic form, is still more appealing for practical use when a good initial model from a robust ray-based tomography is available. However, FWI is very unstable for most cases, which can lead to erroneous velocity update. In Chapter 2, I presented a multi-level temporal integration objective function to improve the performance of FWI.

- The biggest challenge for least-squares migration is the high computational cost assuming we have already achieved a correct velocity. In Chapter 3, I propose using a stochastic conjugate gradient to improve the efficiency of LSM. In Chapter 4, an approximate Hessian preconditioning is applied to improve the efficiency and performance of least-squares reverse-time migration (LSRTM).

- In order to compensate the big amplitude difference between field and modeled data, I propose a local normalization scheme to suppress the amplitude impact and enhance the phase information. A field example for least-square reverse-time migration with the local normalization is presented to conclude this study.

- Conclusion and potential future work are given in Chapter 6.

Chapter 2

Full waveform inversion with multi-level temporal integration

Generally, stabilizing the performance of FWI can be devided into two equally important categories: First is to change the objective function from a time domain to a different domain. Second, using certain regularization method, e.g. using total variation regularization to make the model more blocky (Wang et al., 2012) or using preconditioning, e.g. smoothing to make the model smooth. In this section, we propose a multi-level temporal integration scheme to stabilize the FWI performance, which belongs to the first category of optimization.

2.1 Introduction

The theory of FWI was brought to the geophysical community by (Tarantola, 1984), where the velocity model was updated through calculating the gradient of the misfit objective function. Due to limited computation power, it was recently possible for the FWI to achieve high fidelity images for industrial-sized projects (Vigh and Starr, 2008; Huang et al., 2013; Mothi and Kumar, 2014). In the full-waveform inversion, wave-equation based seismic simulations are performed for individual sources and the differences between simulated and observed commmon-shot-gathers are used to update the velocity model (Krebs et al., 2009). Therefore, the cost of FWI is proportional to the number of shot gathers, which could be prohibitively high for industrial-sized 3D seismic surveys (Huang et al., 2013).

Numerous numerical schemes have been suggested to reduce the computational cost and improve the effciency of FWI (Huang and Zhou, 2015; Krebs et al., 2009; Van Leeuwan et al., 2011). (Krebs et al., 2009) introduce the shots-encoding scheme, where the modeling and the gradient calculations are optimized through an encoding technique. Stochastic optimization has been a popular algorithm applied to many applications in machine learning (Schraudolph and Graepel, 2003). Stochastic sampling techniques are used to reduce the data volume required for optimization. Recent development in stochastic optimization has received much attention in seismic inversion. FWI with stochastic optimization achieved results comparable to that of conventional methods but with only a few percent of CPU time demanded by conventional shot-by-shot methods (Van Leeuwan et al., 2011).

However, an efficient implementation of FWI does not guarantee the success in velocity update. To avoid the algorithm being trapped into a local minima, FWI requires an initial velocity model that is close to the true velocity model. Low-frequency content and ultra-long-offset are essential for a successful FWI update (Huang et al., 2013; Mothi and Kumar, 2014). However, in most seismic acquisitions, long-offset data are not available, and low-frequency data are contaminated by various noises.

Aside from the norm objective function used in early FWI implementations (Tarantola, 1984), different techniques have been invented to reduce the possibility of being trapped in local minima or cycle skipping in the absence of low frequency and long-offset data. Early efforts have been put toward delineating the phase from the amplitude information, such as the cross-correlation objective function (Luo and Schuster, 1991). (Ha et al., 2012) showed that the Laplacian objective function is less sensitive to the lack of low frequencies. (Luo and Wu, 2013) proposed to use envelope inversion to recover the large-scale component of the model, and the envelope objective function was shown to be more efficient and noise insensitive.

Seismic inversion through a multi-level approach has been presented by a number of authors (Bunks et al., 1995; Zhou, 2003), where inversions were performed from coarse to fine scales. Multi-level approaches are generally more stable in comparison to a single level approach. The techniques of temporal and spatial signal integration have been widely used in acoustic motion detection (Moore and Tan, 2003; Hopkins and Moore, 2007). Temporal integration involves combining information over time to improve detection or discrimination, and can often be thought of as an accumulation process, or energy integration (Moore and Tan, 2003). Apparently, information extracted from one part of a signal can influence the evaluation and interpretation of information extracted from another part at a different time. In this study, I briefly review the theoretical background of conventional FWI and our proposed multi-level temporal integral scheme. I conclude by comparing the results from a conventional implementation to our multi-level temporal integral approach.

2.2 FWI with multi-level temporal integration

The state function for wave propagation with constant density can be described as

$$\frac{1}{m^2(\mathbf{x})}\frac{\partial^2 p(\mathbf{x},t;\mathbf{x}_s)}{\partial t^2} - \nabla^2 p(\mathbf{x},t;\mathbf{x}_s) = f(t;\mathbf{x}_s) \tag{2.1}$$

where $m(x)$ is the velocity model, $p(\mathbf{x},t;\mathbf{x}_s)$ is the state variable and $f(t;\mathbf{x}_s)$ is the source function. Conventional misfit objective function compares the observed and modeled wavefield at the receiver location through standard L^2 norm and is given by

$$J(m(\mathbf{x})) = \sum_{x\in x_s}\int_0^T \|p_{res}(x,\tau;x_s)\| d\tau \tag{2.2}$$

where $\|.\|$ denotes a L^2 norm, T is the total recording time, and the misfit $p_{res}(x,\tau;x_s)$ is the difference between the modeled and measured at the observation location by subtraction

$$p_{res}(x,\tau;x_s) = p(x,\tau;x_s) - d(x,\tau;x_s) \tag{2.3}$$

where $p(x, \tau; x_s)$ is the modeled wavefield and $d(x, \tau; x_s)$ is the observed wavefield. Using the adjoint state method (Symes, 2008), the adjoint source can be readily calculated by the derivative of the objective function $J(m(\mathbf{x}))$ with respect to the state variable, and the gradient of the objective function can be written as

$$\frac{\partial J}{\partial m} = RE \sum_{x \in x_g} \int_0^T \left(\frac{\partial^2 p(x, \tau; x_s)}{\partial t^2} \frac{2}{m^3(\mathbf{x})} \right) a_{res}(x, \tau; x_s) ds \tag{2.4}$$

where $a_{res}(x, \tau; x_s)$ is the residual state variable that generated by the adjoint source and adjoint state function. In this study, a non-linear conjugate gradient method is used to calculate the step length (Hager, 2006).

The temporal integration involves combining information over time, which can improve information detection or discrimination (Moore and Tan, 2003). To maximize the ability of temporal integral, we define multi-level physical variables that are a measurement of the temporal integral through the following deductive definition

$$P_k(\mathbf{x}, t; \mathbf{x}_s) = \int_0^t P_{k-1}(\mathbf{x}, \tau; \mathbf{x}_s) d\tau \tag{2.5}$$

where $P_k(\mathbf{x}, t; \mathbf{x}_s)$ is the k^{th} level temporal integral and $P_{k-1}(\mathbf{x}, \tau; \mathbf{x}_s)$ is the $(k-1)^{th}$ level temporal integral. We also define

$$P_0(\mathbf{x}, t; \mathbf{x}_s) = p(\mathbf{x}, t; \mathbf{x}_s) \tag{2.6}$$

where $p(\mathbf{x}, t; \mathbf{x}_s)$ is the wavefield with the original source injection function. With above definitions, we further define the multi-level temporal integration objective function as

$$J(m(\mathbf{x})) = \sum_{\mathbf{x} \in \mathbf{x}_g} \int_0^T \| P_{k,res}(\mathbf{x}, \tau; \mathbf{x}_s) \| d\tau \tag{2.7}$$

where $P_{k,res}(\mathbf{x}, \tau; \mathbf{x}_s)$ indicates a misfit at k level temporal integration. Under Leibniz integral rule for derivative functions (Harley, 1973), the state function for the new state variable becomes

$$\frac{1}{m^2(\mathbf{x})}\frac{\partial^2 P_k(\mathbf{x}, t; \mathbf{x}_s)}{\partial t^2} - \nabla^2 P_k(\mathbf{x}, t; \mathbf{x}_s) = F_k(t; \mathbf{x}_s) \tag{2.8}$$

where $F_k(t; \mathbf{x}_s)$ is k^{th} level temporal integral of the source wavefield

$$F_k(\mathbf{x}, t; \mathbf{x}_s) = \int_0^t F_{k-1}(\mathbf{x}, \tau; \mathbf{x}_s) d\tau \tag{2.9}$$

where we let $F_0(\mathbf{x}, t; \mathbf{x}_s) = f(\mathbf{x}, t; \mathbf{x}_s)$. Equation 2.1 and Equation 2.8 have identical mathematical formulations, except that the variables have different physical contents. The physical meaning of Equation 2.8 can be explained by the fact that the temporal integrated physical field is the result of temporal integrated source injection. Mathematically, the integral over time t is effectively an operator of $\frac{1}{i\omega}$ in frequency domain plus a DC or extremely low frequency term, which can be removed through a low-pass filter.

With the new physical state variable defined in Equation 2.5 and the state function indicated by Equation 2.8, the new gradient can be readily caculated by:

$$\frac{\partial J}{\partial m} = RE \sum_{\mathbf{x} \in \mathbf{x}_g} \int_0^T \left(\frac{\partial^2 P_k(\mathbf{x}, \tau; \mathbf{x}_s}{\partial t^2} \frac{2}{m^3(\mathbf{x})} \right) A_{res,k}(\mathbf{x}, \tau; \mathbf{x}_s) ds \tag{2.10}$$

where $A_{res,k}(\mathbf{x}, \tau; \mathbf{x}_s)ds$ is the residual state variable which is generated by the adjoint source $P_{res,k}(\mathbf{x}, \tau; \mathbf{x}_s)$ and adjoint state function. Modification of this multi-level temporal integration implementation can be straightforward, such as the cross correlation objective function proposed by Luo and Schuster (1991).

2.3 Synthetic results

To validate the multi-level temporal integral FWI, we performed a synthetic test with the Marmousi model. A 2D synthetic prestack dataset was created by forward modeling which uses a Ricker source wavelet with the dominant frequency at 10 Hz and the Marmousi velocity model shown in Figure 2.1(A) . The velocity ranges from about 1500 m/s to 5500 m/s. The acquisition geometry is similar to that of marine acquisition, which has a maximum offset to 8 km and recording time of 4 seconds.

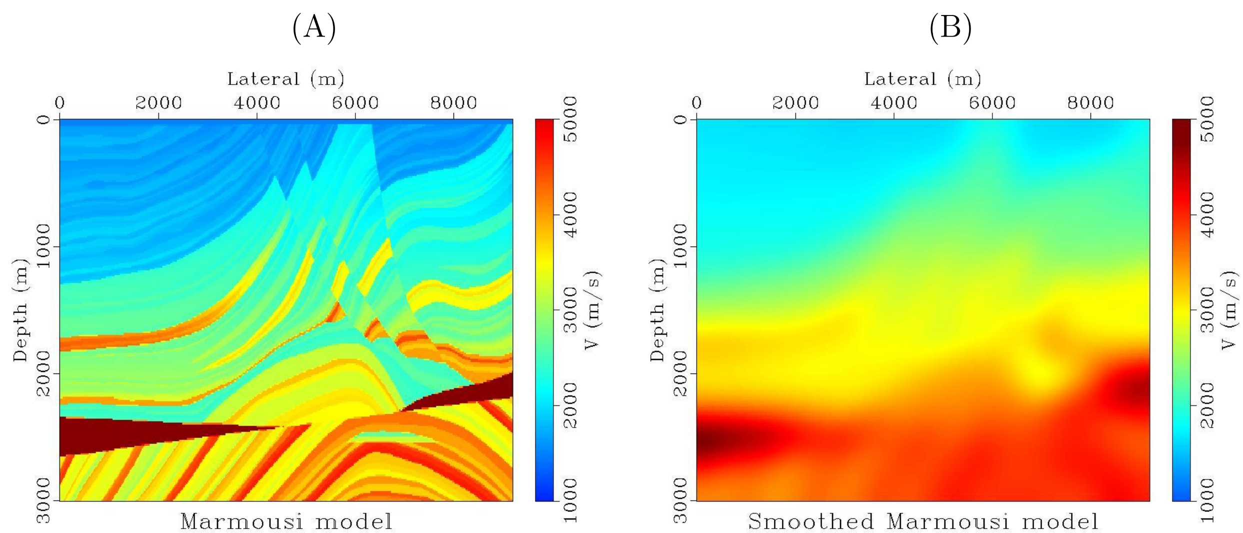


Figure 2.1: Velocity models for the synthetic test. (A) True velocity model for forward modeling (B) Smoothed velocity as the initial model for FWI and MTI FWI update.

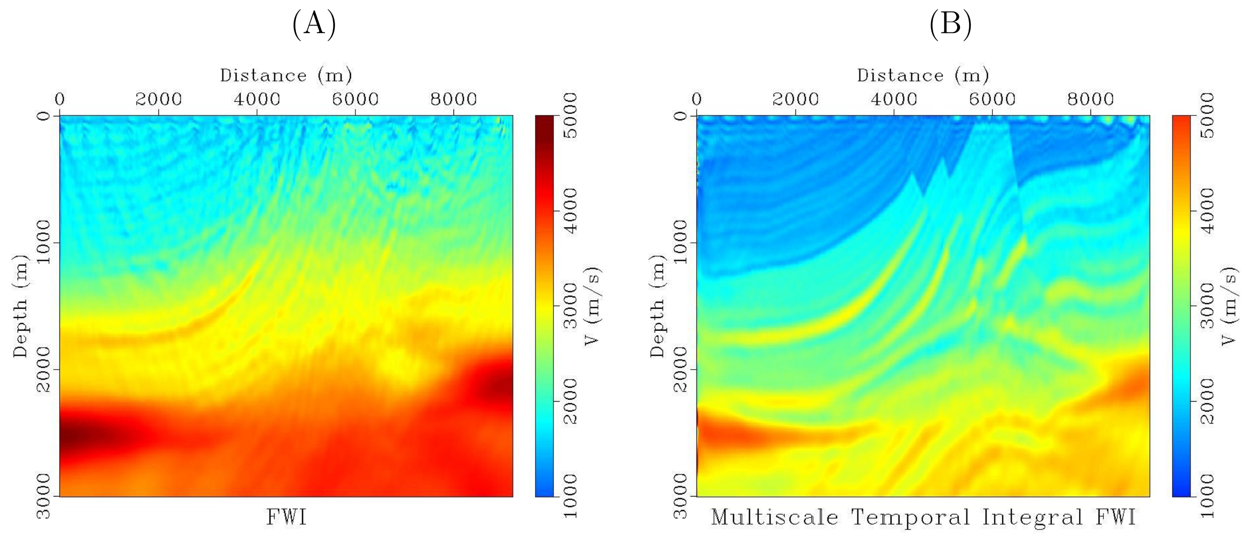


Figure 2.2: Velocity models inverted by two different FWI implementations. (A) Velocity model inverted by the classic FWI (B) Velocity model inverted by MTI FWI.

The initial velocity is an overly smoothed Marmousi model in Figure 2.1(B). We first applied the FWI with its original formulation at 10 Hz for 200 iterations, and the result is shown in Figure 2.2(A). The FWI, in its original form, starts to capture some of the fine features of the sedimentary structures in the true velocity model. However, with a dominant frequency at 10 Hz, this FWI is inadequate to recover the large structure of the velocity or low frequency part of the velocity, especially when the initial velocity is far from the true velocity. These is noticeable when comparing the shallow portion, ranging from 0 to 1000 m depth, of Figure 2.1(A) and Figure 2.2(A).

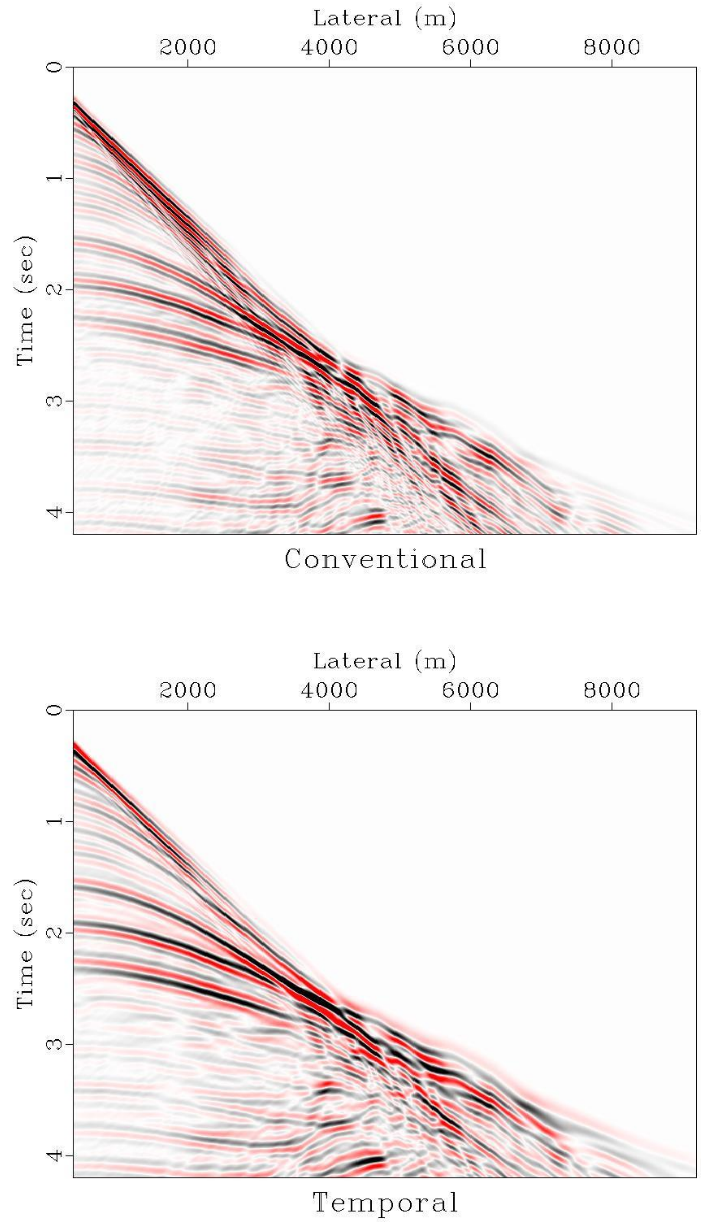


Figure 2.3: Synthetic common-shot-gather with marine acquisition configuration. Top: regular common-shot-gather. Bottom: common-shot-gather generated by level-3 temporal-integral.

We then applied the FWI using the multi-level temporal integration. A total number of 200 iterations are conducted and 50 iterations for consecutive levels from 3 to 0 respectively. A comparison of the original shot and the level 3 temporal integrated shots is presented in Figure 2.3. It is noticeable that the temporal integrated shots demonstrate much lower frequency content, which is clearly demonstrated in the spectra comparison in Figure 2.4. The FWI update starts with the level 3 temporal integral update and then progressively moves to level 0, e.g. we run 50 iterations of FWI with level 3 integration. Followed by run 50 iterations of FWI with level 2.

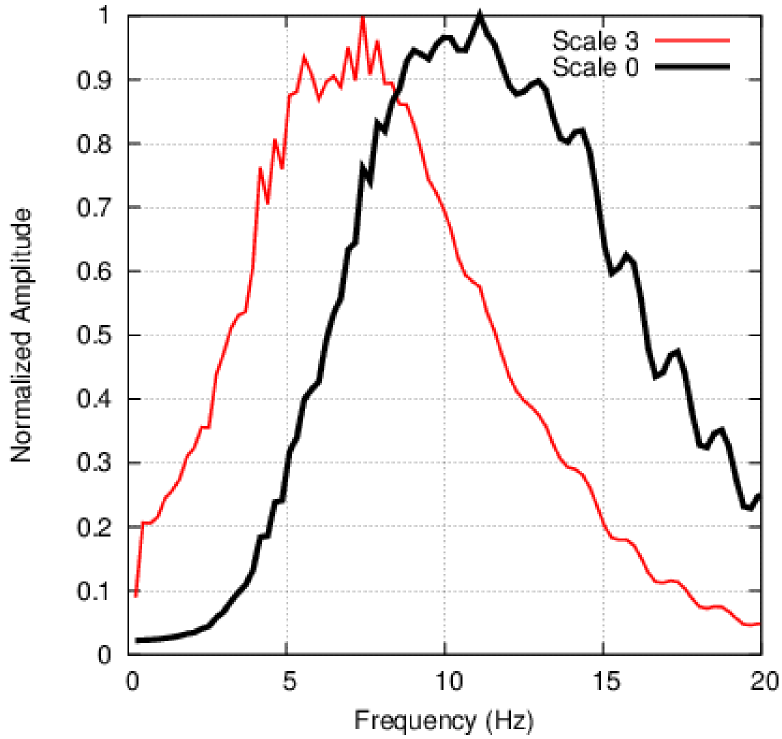


Figure 2.4: Frequency content comparison between the level-3 temporal-integrated common-shot-gather and the regular common-shot-gather.

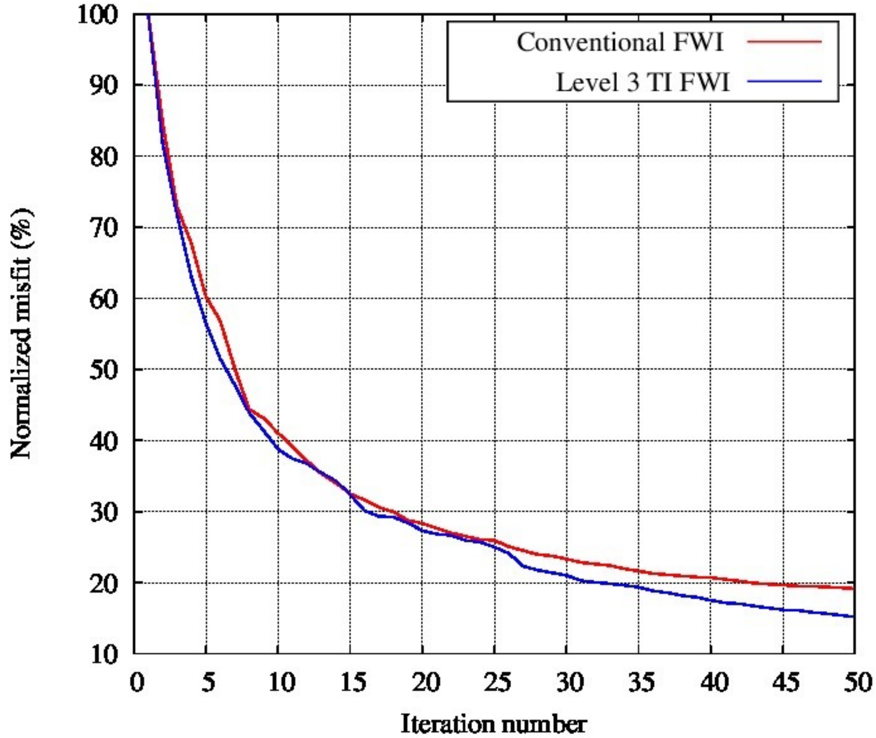


Figure 2.5: Convergnce comparison between the conventional FWI and the level 3 MTI FWI.

The normalized misfits of the conventional FWI and multi-level temporal integration in the first 50 iterations are presented in Figure 2.5. The decreasing rates of the misfit in data space from these two different FWI are about the same for the first 50 iterations, which suggest that the misfit cannot be the unique criteria for a successful FWI update. The final updates from 200 iterations for these FWI methods are presented in Figure 2.2(A) and Figure 2.2(B), respectively. It is clear that conventional implementation fails to recover the bulk properties of the true velocity model. We also compare the velocity profiles at 4 km in lateral distance, which is shown in Figure 2.6. From Figure 2.6, we observe that the FWI in conventional implementation is already trapped in the local minima, which is demonstrated by some bad updates in the shallow region ranging from surface to 1000 m depth. The implementation of the FWI by multi-level temporal integral shows a more stable behavior, and the update recovers most of the feature of the velocity model. However, the deeper update is still very challenging due to large velocity difference between the initial model and the true velocity model. This difference in the updates is seen around the extremely high velocity salt layer at about 2500 m depth. In general, the FWI with multi-level temporal integration performs much better than the conventional FWI. However, a field dataset can be more challenging because of the low-frequency noise in real-world acquisition, e.g. swell noise. Future studies involve the application of this method to the field data.

2.4 Conclusion

We have presented a new numerical scheme to accelerate and stabilize the full-waveform inversion (FWI). This scheme is based on a multi-level temporal integration of the wavefield. The multi-level integration of the original wavefield invokes new state variables and satisfies the wave equation through a temporal integrated source wavefield. The new method is validated using the Marmousi synthetic model. Using simulated data with a dominant frequency at 10 Hz, the classic implementation of the FWI can be easily trapped in local minima; in contrast, the FWI with multi-level temporal integration (MTI) is more robust in recovering the true velocity model.

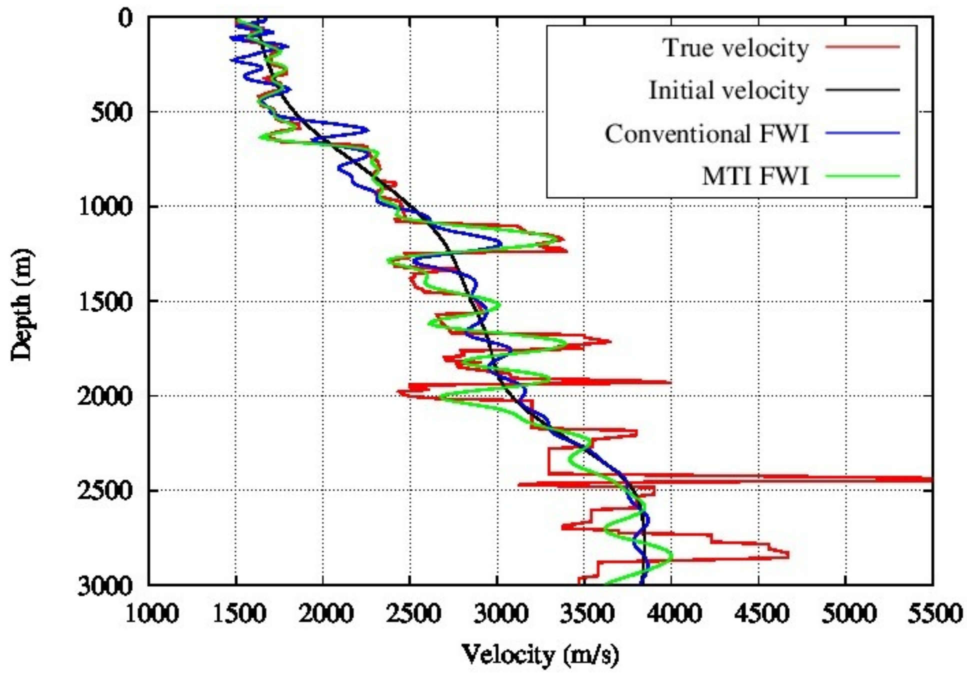


Figure 2.6: Comparison of velocity profiles extracted from the true velocity, the initial velocity, the velocity inverted by conventional FWI, and the velocity inverted by MTI FWI at the center of the velocity model.

Chapter 3

Least-squares seismic-inversion with stochastic conjugate gradient method

With the development of computational power, there has been an increased focus on data-fitting related seismic-inversion techniques for high-fidelity seismic velocity models and images, such as full-waveform inversion and least-squares migration. However, these data-fitting methods can be very expensive in terms of computational cost. Various techniques to optimize these data-fitting seismic-inversion problems have been implemented in recent years to improve efficiency. In this chapter, we propose a stochastic conjugate gradient method for these data-fitting related inverse problems. We first describe the basic theory of the stochastic conjugate gradient method and then give synthetic examples. The numerical experiments illustrate the potential of this method for large-size seismic-inversion applications.

3.1 Introduction

The possibility of using a data-fitting technique for seismic-inversion problems was shown by (Tarantola, 1984) in the 1980s. However, limited by the computational power, only recently the least-squares data-fitting technique was used in industrial imaging projects, e.g., full waveform inversion (FWI), to help velocity model building and imaging for the seismic industry (Vigh and Starr, 2008). In the full-waveform inversion, wave-equation based seismic simulations were performed for individual sources, and the differences between simulated shots and observed shots were used to update the velocity model (Krebs et al., 2009). Therefore, the cost of FWI is proportional to the number of shots, and can be prohibitively high for industrial-sized 3D seismic surveys. Another important seismic-inversion technique that draws a lot of attention in recent years is the least-squares migration with the goal to suppress the migration artifacts and achieve a high-resolution seismic image (Dai and Schuster, 2013; Nemeth et al., 1999). Similar to that of full-waveform inversion, the least- squares migration incurs iterative data-fitting through modeling process, called Born modeling. This modeling can be as huge a computational burden as that of conventional FWI. Recent developments in acquisition technology can provide the exploration industry with high-density and rich-azimuth dataset, which can potentially generate high quality seismic image and velocity model. For example, recent circular-type acquisition can generate datasets with shot density that is several times greater than that of a typical WAZ (wide azimuth) design, resulting in a much higher fold and improved signal-to-noise ratio for subsalt imaging (Huang et al., 2013). Even under current computational power, the size of these datasets makes the data-fitting based inversion computationally formidable. Thus, numerical optimization is highly desired for industrial applications.

In general, current optimizations for least-squares seismic inversion problems can be categorized into several different groups: super-grouping, source encoding and stochastic optimization. Super-grouping is more of a standard practice in industrial seismic imaging projects (Huang et al., 2013), where recorded shots are moved spatially to be combined into a giant super-shot. Special treatment to compensate the spatial change has to be performed, and called a partial move-out. This method has limited accuracy and can bring artifacts by moving field records to a designated position. The second major optimization scheme is the source-encoding technique. Instead of moving shots spatially to join several shots together, the source-encoding scheme is to simulate several shots simultaneously with an assigned random time delay for different shots. This reduces the computational cost for forward simulation (Godwin and Sava, 2010; Krebs et al., 2009). Source encoding schemes can be efficient in most 2D cases. However, source-encoding can have several issues in real-world application, including, an increased number of random shots require to propagate in a larger velocity grid. Also, source-encoding can generate cross-talk artifacts between different shots, which offsets the huge advantage over conventional non-source-encoding methods.

Stochastic optimization is a popular algorithm for many application in machine learning (Schraudolgh and Greapel, 2003), where stochastic sampling techniques are used to reduce computational cost. Recent development in stochastic optimization had received a lot of attention in seismic-inversion problems. (Van Leeuwan et al., 2011) made the application of stochastic optimization application in FWI, and they achieved comparable results to that of the conventional method using only a fraction of computational cost of the conventional method. Rather than combining different shots into one giant super-shot spatially or temporally, the stochastic sampling technique uses different small batches of original data in subsequent iterations to reduce the computational cost. However, due to the stochastic nature of the random sampling over iterations, it is hard to find the conjugate direction for the consecutive iterations, thus a steepest gradient method is actually practiced.

Recent numerical study in function simulation (Jiang and Wilford, 2013) showed the advantage of the stochastic conjugate gradient method (SCG), which could increase the efficiency of the seismic-inversion problems. In this study, we prescribe the stochastic conjugate gradient method for the general least-squares data-fitting seismic-inversion problems. The stochastic conjugate gradient method could potentially increase the convergence rate in comparison to stochastic gradient method. Detailed description of the stochastic conjugate gradient algorithm for least-squares inversion is first given, and numerical results based on a least-squares Kirchhoff migration are presented to prove this methodology. It is conclusive that, by using stochastic conjugate gradient method, the inversion process can reach a faster convergence rate than conventional stochastic sampling methods.

3.2 Theory

3.2.1 Solving of least-squares seismic-inversion problems

The general seismic-inversion problem can be mathematically explained by finding a model vector from following equation

$$\mathbf{d} = \mathbf{L}\mathbf{m} \tag{3.1}$$

where $\mathbf{d}$ is the observed data vector, $\mathbf{L}$ is the forward modeling operator describing a physical process, which is in principle non-linear for most geophysical problems, such as, wave equation modeling or Kirchhoff operator, and $\mathbf{m}$ is the model vector we want to invert. Because of the complexity of the forward modeling operator $\mathbf{L}$, a linear modeling operator is practiced in this study. In general, the solution of Equation 3.1 cannot be achieved directly; least-squares techniques are invoked to solve Equation 3.1. The objective function for inversion of model $\mathbf{m}$ in a least-squares manner is written by

$$J(\mathbf{m}) = \|\mathbf{L}\mathbf{m} - \mathbf{d}\| \tag{3.2}$$

where $J(\mathbf{m})$ is called the objective function or the misfit and $\|.\|$ denotes the L^2 norm. However, because most seismic-inversion problems are ill-posed, different regularization methods have to be used to stabilize the results by least-squares inversion. The most commonly used regularization scheme is the L^2 norm Tikhonov regularization, and the regularized objective function can be written as

$$J(\mathbf{m}) = \|\mathbf{L}\mathbf{m} - \mathbf{d}\| + \epsilon\|\mathbf{D}\mathbf{m}\| \tag{3.3}$$

where $\|\mathbf{Lm} - \mathbf{d}\|$ is the misfit between the modeled and observed data, is the regularization term, and $\mathbf{D}$ denotes the regularization operator, e.g., smoothing operator or unity matrix. There are many different regularization methods other than the Tikhonov regularization method, such as the LASSO method to promote model sparsity (Tibshirani, 1996), which are out the scope of this work. The regularization improves the conditioning of the inverse problem, thus enabling a direct numerical solution, and can be written as shown by (Tarantola, 1984),

$$\mathbf{m} = \left(\mathbf{L}^T\mathbf{L} + \epsilon\mathbf{D}^T\mathbf{D}\right)^{-1}\mathbf{L}^T\mathbf{d} \tag{3.4}$$

where $\mathbf{L}^T$ and $\mathbf{D}^T$ are the adjoint operators to the forward and regularization operator respectively. However, because of the large size of the Hessian matrix $\mathbf{L}^T\mathbf{L}$ for most geophysical problems, it is impractical to invert the Hessian matrix directly. Instead, the model is often solved iteratively through numerical methods. Iterative methods have computational advantages in large-scale geophysical problems when forward and adjoint operators are represented by a sparse matrix and can be computed efficiently (Saad, 2014). One of the simplest numerical methods to solve the least-squares problem is the steepest gradient method. To find the model vector $\mathbf{m}$ that minimizes the objective function prescribed in Equation 3.1, one takes steps proportional to the negative of the gradient or the approximate gradient of $J(\mathbf{m})$. The update of the model vector for the kth iteration can be written as

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha\mathbf{g}_k \tag{3.5}$$

where $\mathbf{g}_k$ denotes the gradient of objective function $J(\mathbf{m})$ and is given by

$$\mathbf{g}_k = \mathbf{L}^T(\mathbf{Lm}_k - \mathbf{d}) + \epsilon\mathbf{D}^T\mathbf{Dm}_k \tag{3.6}$$

Apparently, the magnitude of the update is related to the value α which can be achieved by standard quadratic line search or analytic solutions to reduce the value of the objective function.

The conjugate gradient method is another powerful method to find the solution to Equation 3.3. The fundamental idea of conjugate gradient method is to update the model in a conjugate direction of the current gradient, which can increase the convergence rate in comparison to the steepest descendent method (Schraudolph and Graepel, 2003). After the first iteration in the steepest descendent direction $\nabla J(\mathbf{m})$, the following steps constitute one iteration moving along a subsequent conjugate direction $\mathbf{s}_k$. The update on the model vector can be described as

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha \mathbf{s}_k \tag{3.7}$$

where $\mathbf{g}_0 = \mathbf{s}_0$ for the first iteration, and subsequent conjugate gradient direction can be written as

$$\mathbf{s}_k = \mathbf{g}_k + \beta_k \mathbf{s}_{k-1} \tag{3.8}$$

where different versions of β_k are available, such as the Fletcher-Reeves (FR) formula (Fletcher and Reeves, 1964)

$$\beta_k = \frac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{g}_{k-1}^T \mathbf{g}_{k-1}} \tag{3.9}$$

To solve seismic-inversion by minimizing the objective function, which is denoted as Equation 3.2, multiple iterations of modeling are involved. A series of least-squares seismic-inversion studies using conjugate gradient methods have been performed with success (Dai and Schuster, 2013; Nemeth et al., 1999). The computational cost of least-squares seismic-inversion problems is proportional to the sample space, and one iteration of the gradient based model update usually invokes a set of forward modeling and its adjoint operation. For example, one model update with a standard steepest gradient method, described by Equation 3.5, requires the calculation of gradient, described by Equation 3.6, and two more forward modeling processes are needed to find the optimal step-length through a quadratic line search.

3.2.2 Stochastic conjugate gradient method

Stochastic method is invoked to reduce the computational cost for large geophysical inversion problems. Stochastic optimization for least-squares inversion is used to approximate the objective function in a stochastic sense, where the expectation of the objective function is unchanged in a statistical sense. This can be described as

$$J_s(\mathbf{m}) = \frac{1}{N_s}(\mathbf{Lm} - \mathbf{d})^T \mathbf{W}^T \mathbf{W} (\mathbf{Lm} - \mathbf{d}) \tag{3.10}$$

where $\mathbf{W}$ denotes the random sampling function, normally in the norm distribution format with a zero mean. With the statistical expectation of $\mathbf{W}^T\mathbf{W}$ being N_s, Equation 3.10 can be reduced back to the normal objective function described by Equation 3.2. Apparently, the stochasticity of Equation 3.10 is decreased with the increase of sampling size.

Current stochastic sampling methods for least-squares seismic-inversion propose changing of the sampling subset during each iteration (Van Leeuwan et al., 2011).

In each iteration, a new subset data sample is used to calculate the optimization direction, which is the gradient of the objective function. When fully non-stochastic, where the sampling batch is the whole data, these stochastic sampling methods are reduced to the conventional steepest gradient method. Different random-sampling functions can be used to estimate the objective function in a real-world application, e.g., limiting the sampled data over a prescribed area.

For conventional non-stochastic least-squares inversion problem, the conjugate gradient method has a higher convergence rate than the steepest gradient method. However, solving the stochastic least-squares problem through Equation 3.10 is more difficult than the conventional one (Van Leeuwan et al., 2011), which is described by Equation 3.2. In the case of stochastic sampling, the global minimum is probed by limited stochastic input, giving rise to the noisy estimation of the true Hessian $\mathbf{L}^T\mathbf{L}$ and the gradient $\mathbf{L}^T(\mathbf{L}\mathbf{m}_k - \mathbf{d}) + \epsilon\mathbf{D}^T\mathbf{D}\mathbf{m}$. One of the other difficulties that stochastic sampling may encounter is that conventional conjugate gradient method can fail during iterations because changing of the sampling subset with no overlap breaks the conjugacy of the searching direction over many iterations (Jiang and Wilford, 2013).

To mitigate the breakdown of conjugacy of the searching direction for the stochastic sampling technique, we use a similar approach from that of approximation of function (Jiang and Wilford, 2013) for least-squares seismic-inversion problems. Assuming the sampled subset can still be representative of the large eigenvalues of the original system, we can perform a predefined number of iterations of the conjugate gradient update on each sampled dataset. For each random sampled subset with a sampling function, $\mathbf{W}$, the gradient is calculated by

$$\mathbf{g}_{w,k} = \mathbf{L}^T\mathbf{W}^T\mathbf{W}(\mathbf{L}\mathbf{m}_k - \mathbf{d}) \tag{3.11}$$

where the subscript w denotes a subset of the data space, subscript k is the iteration number within the same batch sample w, and $\mathbf{m}$ is the current updated model. The subsequent conjugate gradient direction can be calculated by

$$\mathbf{s}_{w,k} = \mathbf{g}_{w,k} + \beta_{w,k}\mathbf{s}_{w,k-1} \tag{3.12}$$

where $\mathbf{s}_{w,0} = \mathbf{g}_{w,0}$ for the start of inversion with a current velocity model and current sampled data. The update of the model has a similar format to that of a conventional conjugate gradient, except that the conjugate gradient is calculated stochastically only within the same subset

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha_{w,k}\mathbf{s}_{w,k} \tag{3.13}$$

where $\alpha_{w,k}$ can be calculated through a line search over the sampled subset or analytic solutions for linear forward operator.

The critical difference between the stochastic conjugate gradient and the standard stochastic steepest gradient method is that, stochastic conjugate gradient method performs just a few iterations of conjugate gradient within each stochastically sampled dataset to speed up the convergence. However, in order to keep the stochasticity of the inversion problem defined by Equation 3.10, the inversion must be moved towards a different subset after several iterations to prevent the model from being trapped within a local minimum.

3.3 Synthetic examples

3.3.1 Synthetic setup

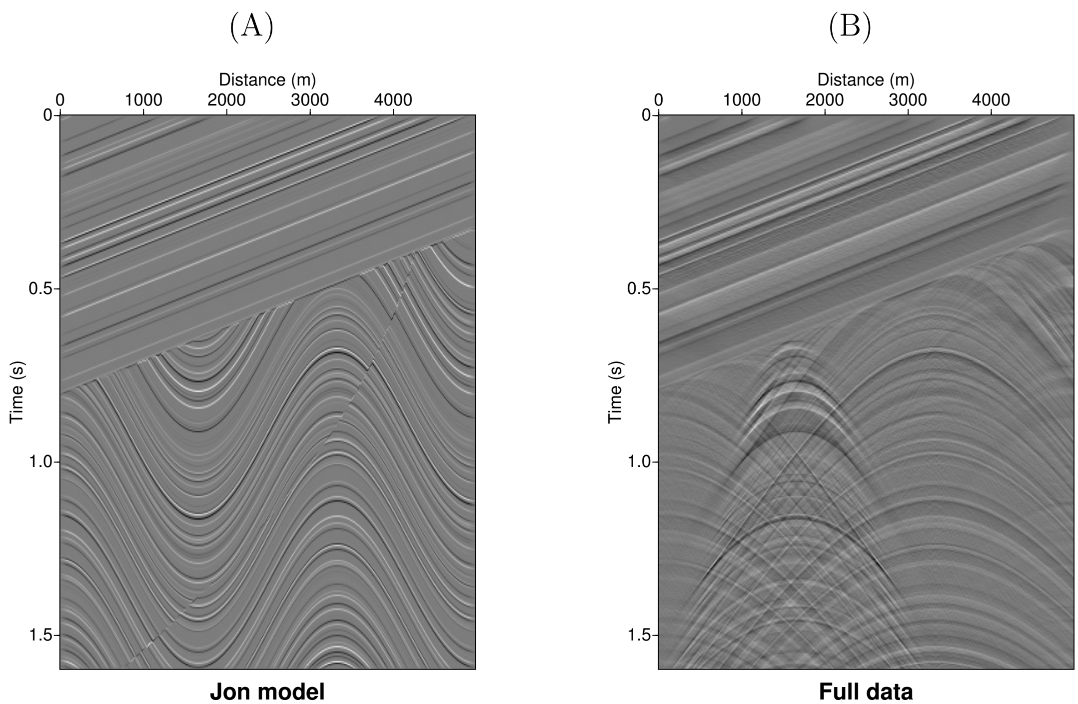


Figure 3.1: Synthetic modeling. (A) Jon's Reflectivity model for generating the synthetic data (B) synthetic time data generated by the reflective model.

To demonstrate the benefits of the stochastic conjugate gradient method, we use a least-squares post-stack Kirchhoff time migration to show this numerical optimization. In order to simplify the idea and separate the benefit of the stochastic conjugate gradient method from other inversion uncertainties, the setup of the test is quite idealized. The earth reflectivity model from (Claerbout, 1985) is selected and is shown in Figure 3.1(A). A constant background velocity of 2000 m/s is used for both the forward and inversion process. The forward modeling and the migration operator are set to be adjoint for this study. The forward operator is the direct Kirchhoff mapping of the reflectivity and adjoint operator is the Kirchhoff time migration. The modeled time data is shown in Figure 3.1(B).

Least-squares Kirchhoff time migration was first carried out by using all the traces with 100 iterations of the regular conjugate gradient as a basis, where the whole dataset was used for each iteration for the model update. Conventional Kirchhoff time migration was also carried out by using the same parameters. Comparison of the results from regular migration and least-squares migration is shown in Figure 3.2, where the advantage of LSM is clearly demonstrated. The least-squares migration results show a much higher resolution, more balanced amplitude and less migration artifact (Dai and Schuster, 2013; Nemeth et al., 1999). The blurring effect from conventional migration is removed through the least-squares process. And LSM can generate a reflectivity map that is close to the true reflectivity, as shown in Figure 3.1(A). The image quality in the deeper boundary of the model is degraded because of the limited migration aperture.

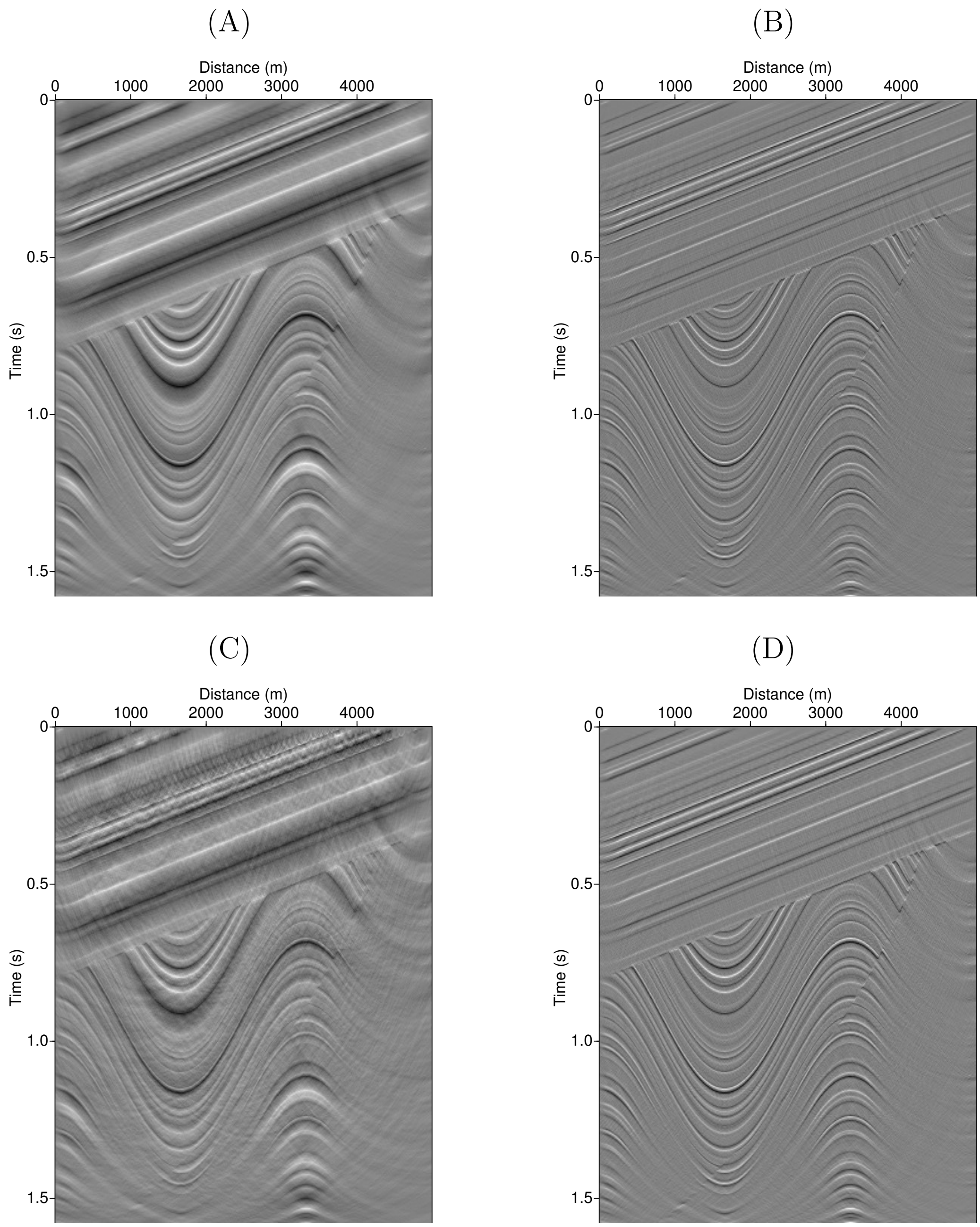


Figure 3.2: Image comparison from LSM and regular migration. (A) Regular migration using the entire traces (B) Least-squares migration using entire traces (C) Regular migration using 50% of the traces (D) Least-squares migration using 50% of the traces.

One of the biggest advantages of the least-squares migration in comparison to conventional migration is the ability to handle incomplete reflection data (Nemeth et al., 1999). To demonstrate the ability of least-squares migration to handle incomplete reflection data, we randomly remove 50% traces from original data, and migrate the data with conventional and the least-squares operator. Figure 3.3 shows the subsets with randomly selected 50% and 20% of the original datasets. The migration results with 50% of original trace using conventional and least-squares migration are demonstrated in Figure 3.2(C) and Figure 3.2(D). From Figure 3.2(C) and 3.2(D) it is conclusive that, conventional migration using incomplete dataset will generate strong migration artifacts, such as the migration swing ; while least-squares migration can efficiently suppress the migration artifacts due to the incompleteness of the input data.

3.3.2 Stochastic conjugate gradient method

To validate the stochastic update scheme, we first apply conventional stochastic approach, which uses a different subset of data over consecutive iterations. Figure 3.4 shows the results from stochastic gradient method, using different sizes of batch sample, which ranges from 5% to 100% of the entire traces. Apparently, the stochastic gradient (SG) method with a 100% trace percentage is identical to that of the conventional gradient method without any computational advantage. From Figure 3.4, it is obvious that, the difference between images from varying the percentage of batch samples is marginal, especially in the shallow section. The trace-by-trace normalized misfit, as a function of the number of iterations, is shown in Figure 3.6.

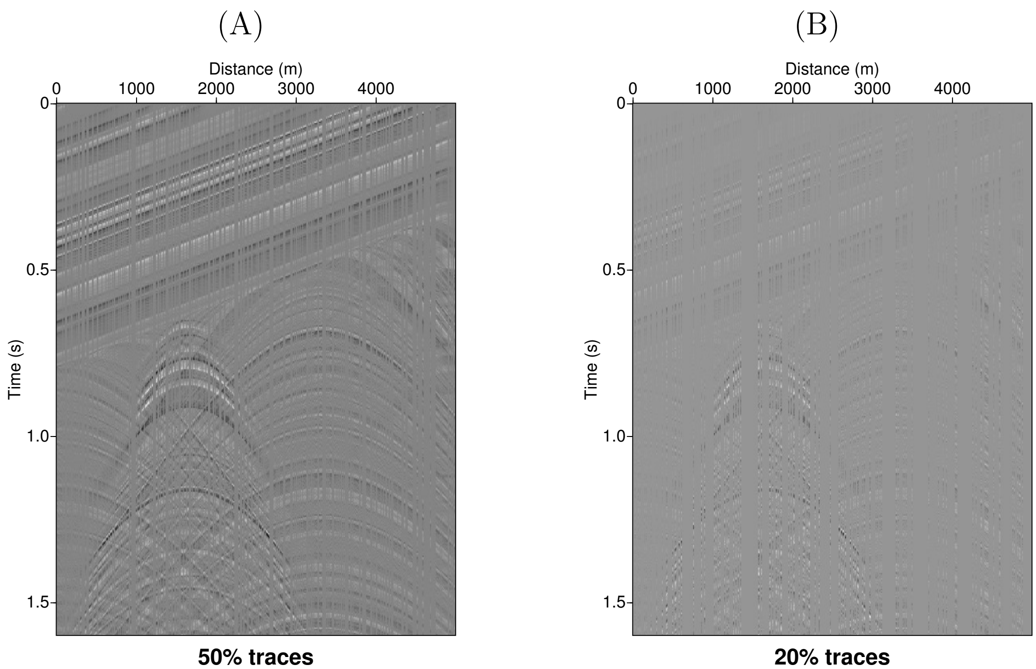


Figure 3.3: Stochastic sampling of data. (A) 50 % of the original data using random sampling (B) 20 % of the original data using random sampling.

It can be observed that the convergence rates of stochastic gradient method with different batch sizes are about the same. The deep reflectors are slightly better imaged when migrated with all the traces, which is observed in Figure 3.4.

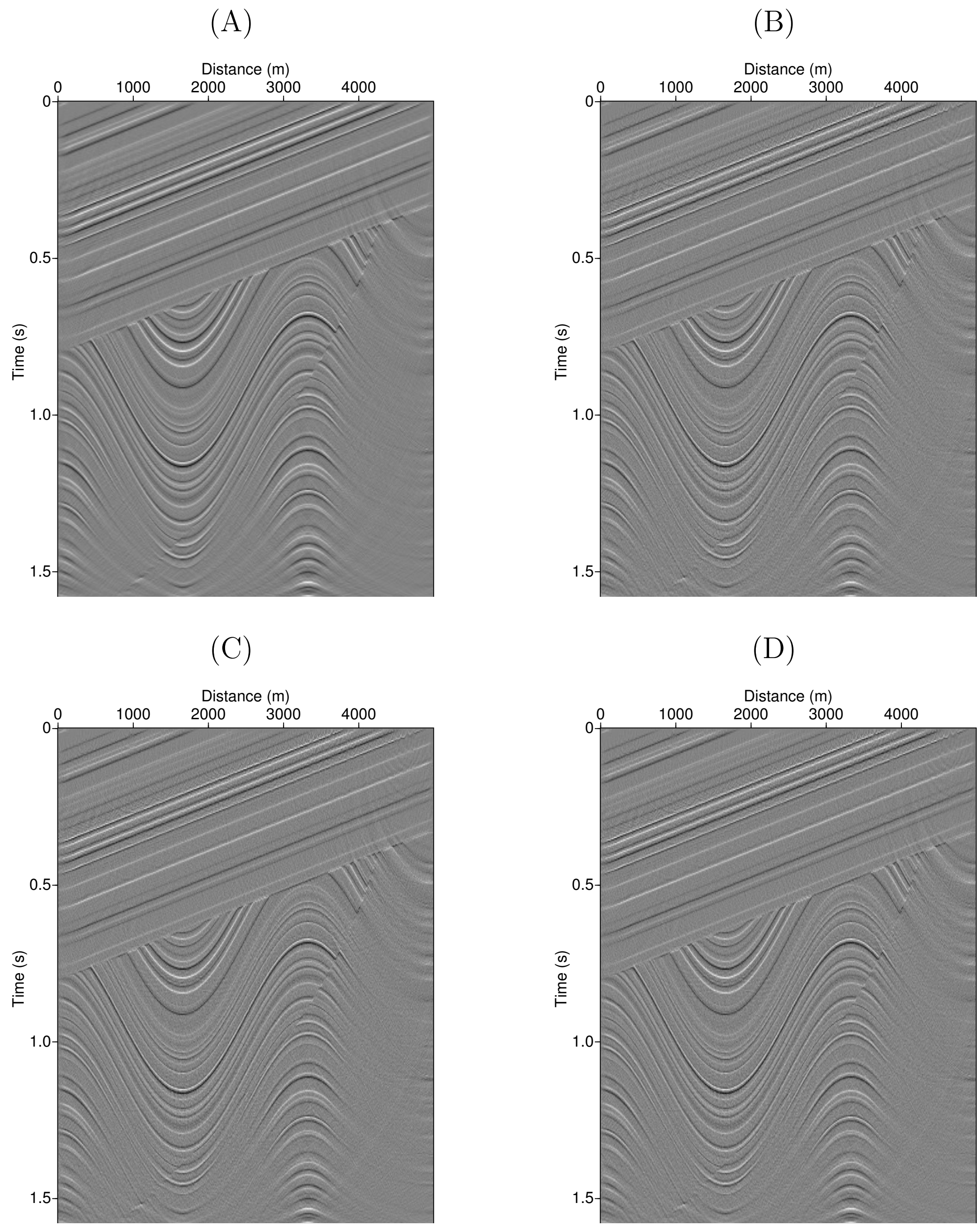


Figure 3.4: Migration image comparison. (A) LSM using the entire traces (B) SCG using 20% of the traces for each iteration (C) SCG using 10% of the traces for each iteration (D) SCG using 5% of the traces for each iteration.

(A) (B)

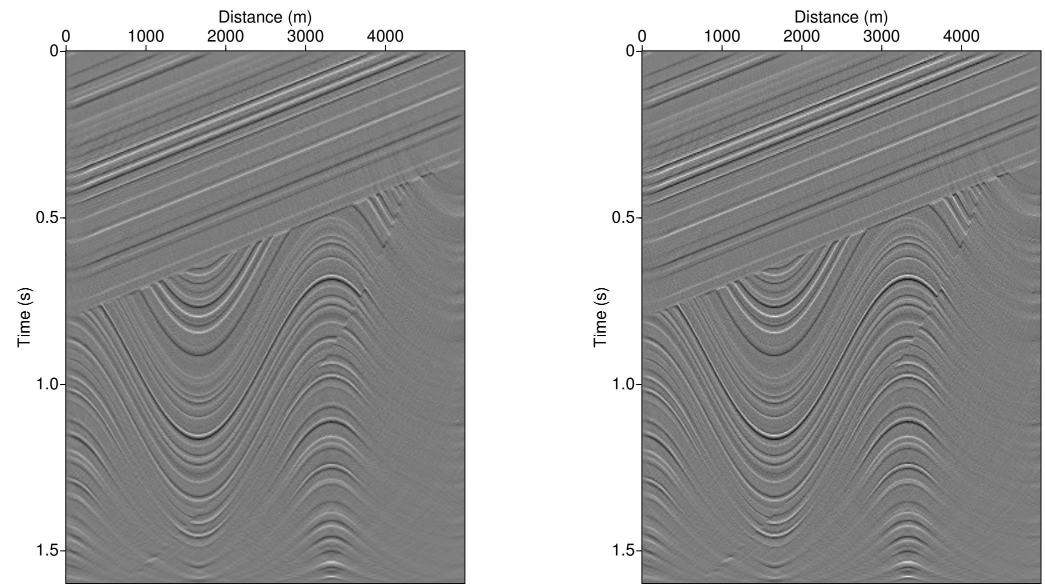


Figure 3.5: Migration image comparison. (A) LSM with stochastic steepest gradient (B) LSM with stochastic conjugate gradient.

To compare the stochastic steepest decent method and stochastic conjugate gradient method, 100 iterations of image updates were performed. The sampled batch size is fixed for both the stochastic conjugate gradient and stochastic gradient method, which is 5% for this test. For stochastic gradient method, different sample traces are selected for consecutive iterations, and standard line search was used in combination with the stochastic gradient to update the reflectivity model. In order to make the results comparable with that of the stochastic gradient, 5 iterations of conjugate updates were performed for each sampled batch, and a total of 20 total sampling batches were used for each update, which makes the computational cost the same as that of stochastic gradient. However, because of the random nature of our sampling scheme, using 20 random batches may not cover the every trace in the data . It can be concluded that using data batches with overlapping traces will not break down the convergence of the stochastic method as long as the sampling function is completely random.

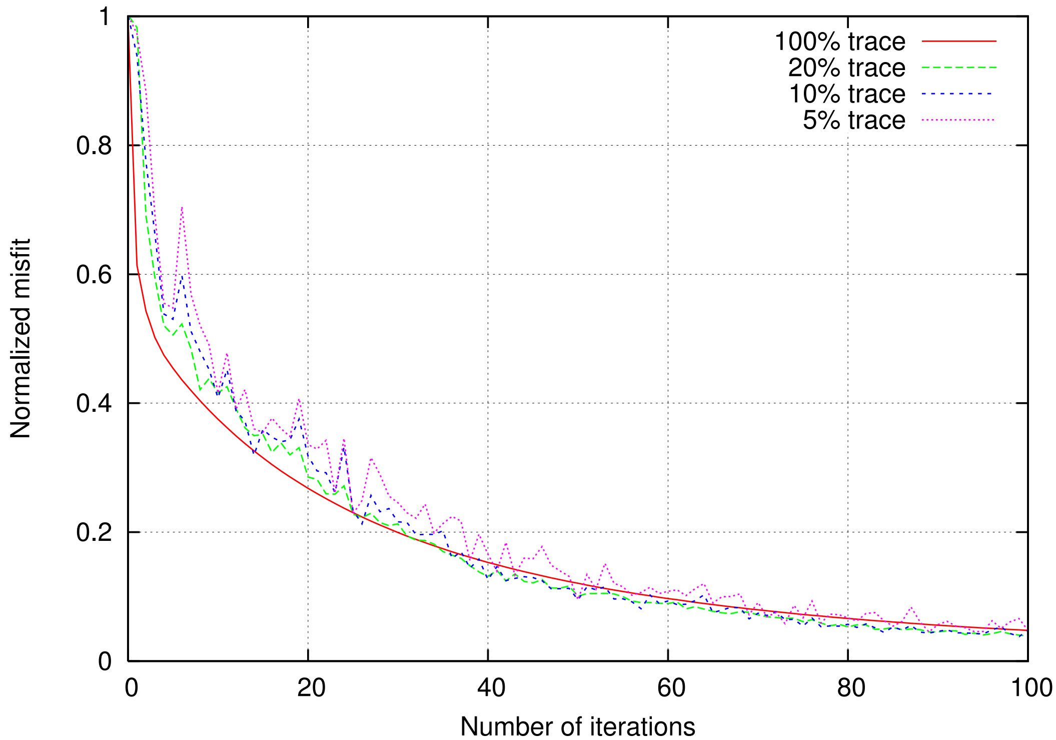


Figure 3.6: Convergence comparison for the least-squares migration with different batch sample sizes.

The results of the stochastic gradient and stochastic conjugate gradient method are shown in Figure 3.5. The difference between the image from SCG and that from SG are marginal, because in this much idealized testing, both methods can achieve good convergence over 100 iterations. It is also clear that stochastic optimization results are quite comparable to that of conventional least-squares method using whole-data space, which is shown by Figure 3.4(A). The differences between these methods are prominent when we present the misfit as a function of iteration number in a log-scale view as shown in Figure 3.7. In general, the misfit drop from stochastic optimization is not as fast as that of regular optimization using the entire data space, because of the stochastic nature of approximating the Hessian and the gradient. Comparing SCG and SG, it is clear that the stochastic conjugate gradient method shows a higher convergence rate than that of stochastic gradient method. The misfit drop for SCG shows a zig-zag pattern, which can be explained by the fact that, when iteration using the same randomly sampled batch, the conjugacy is well maintained thus resulting in a more steep misfit drop or a better model update. The convergence of SG method shows a smoother damping behavior. In particular, stochastic optimizations demand more iteration to achieve the same level of misfit as that of regular inversion using the entire dataset. This partially offsets the saving of simulating smaller dataset required by each iteration. For example, in order to achieve a 90% percent misfit drop, while the conventional least-squares method using the entire data space required around 10 iterations, SCG requires around 40 iterations, and the SG method demanded around 80 iterations.

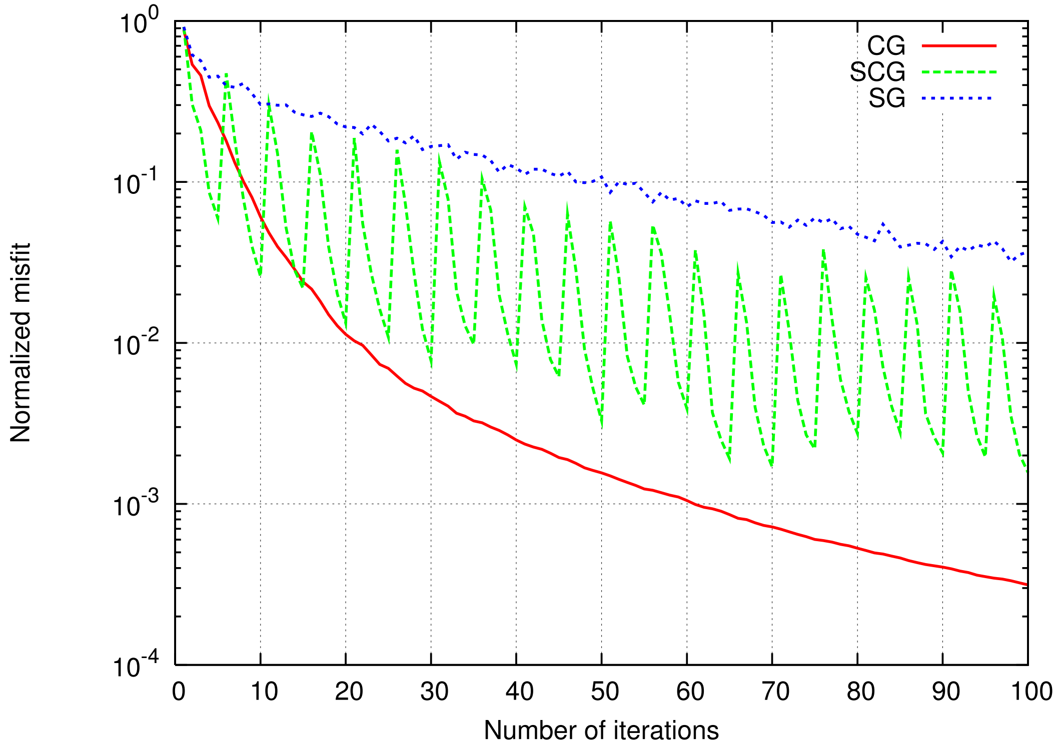


Figure 3.7: Convergence comparison for least-squares migration with conjugate gradient method using 100% trace, stochastic gradient, and stochastic conjugate gradient method.

3.4 Discussion and conclusion

High computational cost has been a big drawback for least-squares seismic-inversion problems. Stochastic sampling techniques have been used to optimize these seismic-inversion problems. In most stochastic sampling applications on seismic imaging and inversion, sample batches are changed over each iteration, which generally invokes steepest decent gradient method. The conjugate gradient method, in its standard form, cannot be carried out when the gradient is calculated stochastically, because of the loss of the conjugacy from non-overlap samples. Stochastic conjugate gradient method for function approximation has shown promise with higher convergence rates than the stochastic steepest gradient method. We further extend this method to general least-squares seismic inversion problems. Rather than changing the sampling subset each iteration, the model update can be performed with a limited number of conjugate gradient updates using the same sampled subset.

We used a least-squares Kirchhoff time migration to prove this method. The results show that the stochastic conjugate gradient method has a higher convergence rate than that of conventional stochastic steepest gradient method. However, for a real-world dataset, the convergence is subject to the uncertainties in the velocity model. The modeling and imaging operators are not exact adjoint, which influence the convergence rate. From the synthetic test, we conclude that it is possible for the stochastic conjugate gradient method to be applied to large-scale seismic-inversion problems, which can out-perform the current stochastic gradient method.

Chapter 4

Least-squares reverse-time migration with Hessian preconditioning

Least-squares reverse-time migration (LSRTM) has been shown to be able to improve image quality over conventional migration method; however, the computational cost of LSRTM prevents it from large scale industrial application. Least-squares inversion process normally invokes a Newton-like method to optimize the objective function, and the full-Hessian is hard to compute for large scale geophysical problems. In this chapter, we developed a numerical scheme using approximate Hessian to reduce the computational cost. Compared with the classic approach, the proposed scheme can generate a more balanced image and make the LSRTM more efficient.

4.1 Introduction

Least-squares inversion techniques has been an important part of estimating the sub-surface velocity, e.g., full-waveform inversion (FWI), and building true amplitude reflectivity model, e.g., least-squares reverse-time migration (LSRTM). Unlike the conventional migration, where the reflectivity image is achieved by applying the adjoint operator to the data (Plessix, 2006), least-squares inversion updates the model through minimizing the objective function. The objective function is generally given as the difference between modeled and observed data.

Solving the least-squares problem generally invokes the gradient method. The gradient formed by cross-correlation suffers from geometrical spreading effects, which results in poor amplitudes for deep reflectors and a slow convergence rate. The approximate Hessian in the Gauss-Newton method ignores the nonlinear term in full Hessian matrix (Pratt et al., 1998), and is too expensive to be calculated directly. (Shin et al., 2001) proposed to use the pseudo-Hessian matrix as a substitution for the approximate Hessian in the full-waveform inversion. The diagonal pseudo-Hessian is limited to balancing the amplitudes due to the source-side spreading and ignores the receiver-side Greens functions. (Tang, 2009) used a phase-encoding technique to calculate the Hessian for a targeted area and performed the inversion in the model space. This method is still expensive. Under the reciprocity theory, the receiver-side Greens functions can be replaced by the source-side Greens functions under the assumption that the sources and receivers are collocated (Plessix and Mulder, 2004). This method forms the double illumination strategy seen in the FWI application.

In this study, we first briefly outline the basics of RTM and LSRTM. We focus on an efficient Hessian approximation, which is based on the reciprocity of sources and receivers. The Hessian preconditioning makes LSRTM more efficient in generating a high quality image. We also employ a smoothing operator as the preconditioner for the Hessian matrix, which is similar to that of smoothing-imaging condition proposed by (Guitton et al., 2006). Synthetic examples of Sigsbee2B velocity model show the efficiency of the proposed method in generating high quality images.

4.2 Born modeling and reverse-time migration

The theory of LSRTM has been established by a number of authors (Dai and Schuster, 2013; Plessix, 2006). In this section, we briefly outline the derivation of the Born modeling operator and reverse-time migration operator. The least-squares inversion using Born modeling is shown as the linear operator relative to the reflectivity model.

4.2.1 Born modeling

The wave equation can be very complex if taking into account the elastic effects. For simplicity, we start with the acoustic wave equation with velocity model $m(\mathbf{x})$ and a constant density, ρ, to derive the Born modeling operator and reverse-time migration operators. The wave-equation can be written as

$$\frac{1}{m(\mathbf{x})^2}\frac{\partial^2 p(\mathbf{x},t;m(\mathbf{x}))}{\partial t^2} - \nabla^2 p(\mathbf{x},t;m) = s(\mathbf{x},t) \tag{4.1}$$

where $s(\mathbf{x}, t)$ is the source signature and $p(\mathbf{x}, t; m)$ is the resulting wavefield. Given the source at $\mathbf{x}'$ and t', the Green's function solution $G(\mathbf{x}, t; m)$ can be written as:

$$\frac{1}{m(\mathbf{x})^2} \frac{\partial^2 G(\mathbf{x}, t; m(\mathbf{x}))}{\partial t^2} - \nabla^2 G(\mathbf{x}, t; m) = \delta(\mathbf{x} - \mathbf{x}')\delta(t - t') \tag{4.2}$$

We can perturb the background model to relate the reflectivity model to seismic data

$$m(\mathbf{x}) = m_0 + \Delta m \tag{4.3}$$

For simplicity, we rewrite the wave equation with Green's function in the frequency domain

$$\left[\frac{\omega^2}{m_0^2} + \nabla^2\right] G(\mathbf{x}_i, \mathbf{x}_s; m_0; \omega) = -\delta(\mathbf{x}_i - \mathbf{x}_s) \tag{4.4}$$

The solution of wave equation with sourcing term $S(x_j; \omega)$ can be written as

$$P_0(\mathbf{x}_i, \mathbf{x}_j; m_0; \omega) = \sum_{j=1}^{M} G(\mathbf{x}_i, \mathbf{x}_j; m_0; \omega) S(\mathbf{x}_j; \omega) \tag{4.5}$$

where P_0 is the seismic response or background wavefield from M different sources with source signature $S(\mathbf{x}_j; \omega)$. When we perturb the velocity model with a small perturbation Δm, the perturbed wavefield $P = P_0 + \Delta P$ is a solution of following Green's function

$$\left[\frac{\omega^2}{(m_0 + \Delta m)^2} + \nabla^2\right] (G(\mathbf{x}_i, \mathbf{x}_s; m_0; \omega) + \Delta G(\mathbf{x}_i, \mathbf{x}_s; m_0 + \Delta m; \omega)) = -\delta(\mathbf{x}_i - \mathbf{x}_s) \tag{4.6}$$

with a small perturbation approximation, by neglecting high order terms, to yield

$$\frac{1}{(m_0 + \Delta m)^2} \approx \frac{1}{m_0^2}\left(1 - \frac{2\Delta m}{m_0}\right) \tag{4.7}$$

which can be plugged into Equation 4.6, and we can have

$$\left[\frac{\omega^2}{m_0^2} + \nabla^2\right] \Delta G(\mathbf{x}_i, \mathbf{x}_s; m_0 + \Delta m; \omega) = 2\omega^2 \frac{r(\mathbf{x})}{m_0^2} G(\mathbf{x}_i, \mathbf{x}_s; m_0; \omega) \tag{4.8}$$

where the reflectivity $r(\mathbf{x})$ is defined as

$$r(\mathbf{x}) = \frac{\Delta m(\mathbf{x})}{m(\mathbf{x})} \tag{4.9}$$

We can see the perturbed Green's function ΔG is the result of the secondary source of $2\omega^2 \frac{r(\mathbf{x})}{m_0^2} G_0(\mathbf{x}_i, \mathbf{x}_s; m_0; \omega)$. With the Green's function solution for the secondary source, we can have the perturbed data as:

$$\Delta P(\mathbf{x}_i, \mathbf{x}_s; \omega) = \sum_{j=1}^{M} S(\omega)\omega^2 \frac{2}{m_0^2} G(\mathbf{x}_i, \mathbf{x}_j; m_0; \omega) G(\mathbf{x}_j, \mathbf{x}_s; m_0; \omega) r(\mathbf{x}_j) \tag{4.10}$$

Equation 4.10 is the Born modeling expressed in the frequency domain, where the Green's function $G(\mathbf{x}_i, \mathbf{x}_j; m_0; \omega)$ and $G(\mathbf{x}_j, \mathbf{x}_s; m_0; \omega)$ denote the impluse response from $\mathbf{x}_i$ to $\mathbf{x}_j$, and $\mathbf{x}_j$ to $\mathbf{x}_s$ respectively. Apparently, Equation 4.10 represents a single scattering process. The interpretation of Equation 4.10 can be straightforward, which is an illustration of Huygens construction. The observed wavefield at $\mathbf{x}_i$ is generated by superposition of wave field scattered from scatters located at $\mathbf{x}_j$, where these scatters have sourcing functions that are proportional to the product of incident source wavefield $S(\omega)G(\mathbf{x}_j, \mathbf{x}_s; m_0; \omega)$ and the reflectivity $r(\mathbf{x}_j)$ at point $\mathbf{x}_j$.

4.2.2 Reverse-time migration (RTM) operator

With the development of computational power, reverse-time migration (RTM) have now been widely adopted as the standard algorithm for imaging complex subsurface structures. Mathematically, the RTM operator is the adjoint of the Born modeling operator defined as Equation 4.10 and can be further simplified as

$$\Delta P(\mathbf{x}_i, \mathbf{x}_s; \omega) = Lr(\mathbf{x}_j) \tag{4.11}$$

and L is

$$L = \sum_{j=1}^{M} S(\omega)\omega^2 \frac{2}{m_0^2} G(\mathbf{x}_i, \mathbf{x}_j; m_0; \omega) G(\mathbf{x}_j, \mathbf{x}_s; m_0; \omega) \tag{4.12}$$

where the adjoint of L can be achieved by the making the complex conjugate of L, which lead to the adjoint operation on data ΔP:

$$r(\mathbf{x}_j) = \sum_{j=1}^{M} S(\omega)\omega^2 \frac{2}{m_0^2} G^*(\mathbf{x}_i, \mathbf{x}_j; m_0; \omega) G^*(\mathbf{x}_j, \mathbf{x}_s; m_0; \omega) \Delta P(\mathbf{x}_i, \mathbf{x}_s; \omega) \tag{4.13}$$

Equation 4.13 is the definition of RTM in the frequency domain, where we have two complex conjugates on the Green's function G, which can be interpreted as the correlation imaging condition in the time domain with reversed time input on the receiver side. In real world practice, the term m_0^2 is generally neglected. Practically, Equation 4.13 is generally implemented in the time domain. When transfer from frequency to time domain, multiplication of two function $A(\omega)B(\omega)^*$ becomes cross-correlation of their inverse Fourier transformed function $a(t) \bigotimes b(t)$. The source is forward propagated in time, which is mathematically equivalent to the term $S(\omega)\omega^2 \frac{2}{m_0^2} G(\mathbf{x}_i, \mathbf{x}_j; m_0; \omega)$. The term $G^*(\mathbf{x}_j, \mathbf{x}_s; m_0; \omega) \Delta P(\mathbf{x}_i, \mathbf{x}_s; \omega)$ is implemented as propagating the reverse-timed data from the receiver side.

4.3 Least-squares reverse-time migration with Hessian preconditioning

In the previous section, the Born modeling operator and its adjoint operator, the RTM operator, are given explicitly. The Born modeling process with reflectivity model $\mathbf{m}$ can be further simplified in operator form, and is given by

$$\mathbf{d} = \mathbf{L}\mathbf{m} \tag{4.14}$$

where $\mathbf{L}$ is the Born modeling operator. Equation 4.14 can be solved through least-squares inversion discussed in Chapter 3. In practice, the performance of least-squares inversion can be improved through different methods. Aside from the numerical optimization, e.g., stochastic optimization, preconditioning of the gradient can also improve the performance of the least-squares inversion. In this section, we briefly outline the background of Hessian preconditioning on the least-squares reverse-time migration.

4.3.1 The update gradient for LSRTM

The iterative update of reflectivity image by least-squares migration can be written as

$$\mathbf{m}_{i+1} = \mathbf{m}_i - \alpha \mathbf{g}_i \tag{4.15}$$

where i denotes the iteration number, $\mathbf{m}_i$ is current reflectivity image , $\mathbf{g}_i$ is the gradient, and α is the step length. The gradient $\mathbf{g}_i$ can be calculated through

Newton's method, Gauss-Newton's methods, and other Quasi-Newton's methods. In Newton's method, the gradient $\mathbf{g}$ is calculated by ignoring higher-order (> 2) terms in the Taylor expansion of the objective function

$$J(\mathbf{m} + \delta\mathbf{m}) = J(\mathbf{m}) + \delta\mathbf{m}^T\mathbf{g} + \frac{1}{2}\delta\mathbf{m}^T\mathbf{L}^T\mathbf{L}\delta\mathbf{m} + ... \tag{4.16}$$

Ignoring the higher order term, we can see the update can be explicitly written as

$$\delta\mathbf{m} = -\mathbf{H}^{-1}\mathbf{g} \tag{4.17}$$

where $\mathbf{H}^{-1} = \left(\mathbf{L}^T\mathbf{L}\right)^{-1}$ is the inverse Hessian matrix and can be written as

$$\mathbf{H}_{jk} = \frac{\partial^2 J}{\partial\mathbf{m}_j\partial\mathbf{m}_k} \tag{4.18}$$

where j and k denote the different grid point in the model. If the direct Hessian can be achieved, the update step length α is unity, and the inversion does not incur a line search. However, for models with large numbers of parameters, the computational cost to compute the Hessian can be large, and prevents the direct Newton's method to be used in real-world applications. For example, in order to invert a 2D model with 1000×1000 grids points, a total of 10^{12} forward modeling steps have to be calculated. However, it is possible to find the approximate Hessian to precondition the gradient $\mathbf{g}$, and the line search method to find the step length α.

4.3.2 Approximate Hessian

From Equation 4.16 and Equation 4.17, it is clear that the Hessian matrix can be used to precondition the gradient $\mathbf{g}$. With a proper Hessian preconditioning, the model perturbation $\delta\mathbf{m}$ can be very close to the true perturbation. The exact expression of the Hessian in the linear inversion case can be written in the Greens function format (Plessix and Mulder, 2004; Pratt et al., 1998)

$$H(\mathbf{x}_j, \mathbf{x}_k) = \sum_{\omega} \omega^4 S_{\omega}^2 \sum_{s} G^*(s, \mathbf{x}_j, \omega) G(s, \mathbf{x}_k, \omega) \sum_{r} G^*(r, \mathbf{x}_j, \omega) G(r, \mathbf{x}_k, \omega) \quad (4.19)$$

where f_{ω}^2 is the source forcing signature, s and r correspond to the source and receiver locations. Physically, the complete Hessian denotes the matrix of the second derivatives of the error functional with respect to the model parameters (Plessix and Mulder, 2004). Due to the high computational cost in calculation of the Hessian matrix, the off-diagonal terms are generally ignored. For example, we assume $G^*(s, \mathbf{x}_j, \omega)G(s, \mathbf{x}_k, \omega) = 0$, when $j \neq k$, and we only keep the diagonal terms in the Hessian matrix, which is given by

$$H(\mathbf{x}_j, \mathbf{x}_j) = \sum_{\omega} \omega^4 S_{\omega}^2 \sum_{s} G^*(s, \mathbf{x}_j, \omega) G(s, \mathbf{x}_j, \omega) \sum_{r} G^*(r, \mathbf{x}_j, \omega) G(r, \mathbf{x}_j, \omega) \quad (4.20)$$

In general, the source side component of the Hessian can be readily computed, which can be stored during the wave propagation process, and the receiver side Hessian component has to be computed separately. This is computationally unpractical for large numbers of receivers. For infinite receiver coverage, the receiver part of the Hessian can be regarded as a constant. In marine acquisition, most geometries can be regularized into a fixed spread, where receivers are symmetric around the source location. In general, the assumption that source and receiver data are collocated is acceptable to some extent, especially for tow-streamer acquisition. Under the reciprocity assumption, the source and receiver side Hessian have the relation

$$G^*(s, \mathbf{x}_j, \omega) G(s, \mathbf{x}_j, \omega) = G^*(r, \mathbf{x}_j, \omega) G(r, \mathbf{x}_j, \omega) \quad (4.21)$$

With this approximation, the gradient can be readily conditioned by the approximate Hessian. Also, we can notice that preconditioning the Hessian using the reciprocity is equivalent to applying a double illumination compensation to the gradient. To compensate for the fast change of the diagonal component of Hessian due to large velocity variations, we further impose a smoothing operator on the Hessian to improve the stability. This is given by

$$H(\mathbf{x}_j, \mathbf{x}_k)_{smooth} = < H(\mathbf{x}_j, \mathbf{x}_k) > \tag{4.22}$$

where $< . >$ stands for smoothing in the image space in the x, y and z directions. Equation 4.22 is very similar to the smoothing imaging condition proposed by (Guitton et al., 2006), where the smoothing is proved to be providing more balanced amplitude map. The smoothing of the Hessian is to achieve a more balanced distribution of Hessian when there is a strong scattering component in the velocity model, such as, a rugose salt boundary.

4.4 Synthetic examples

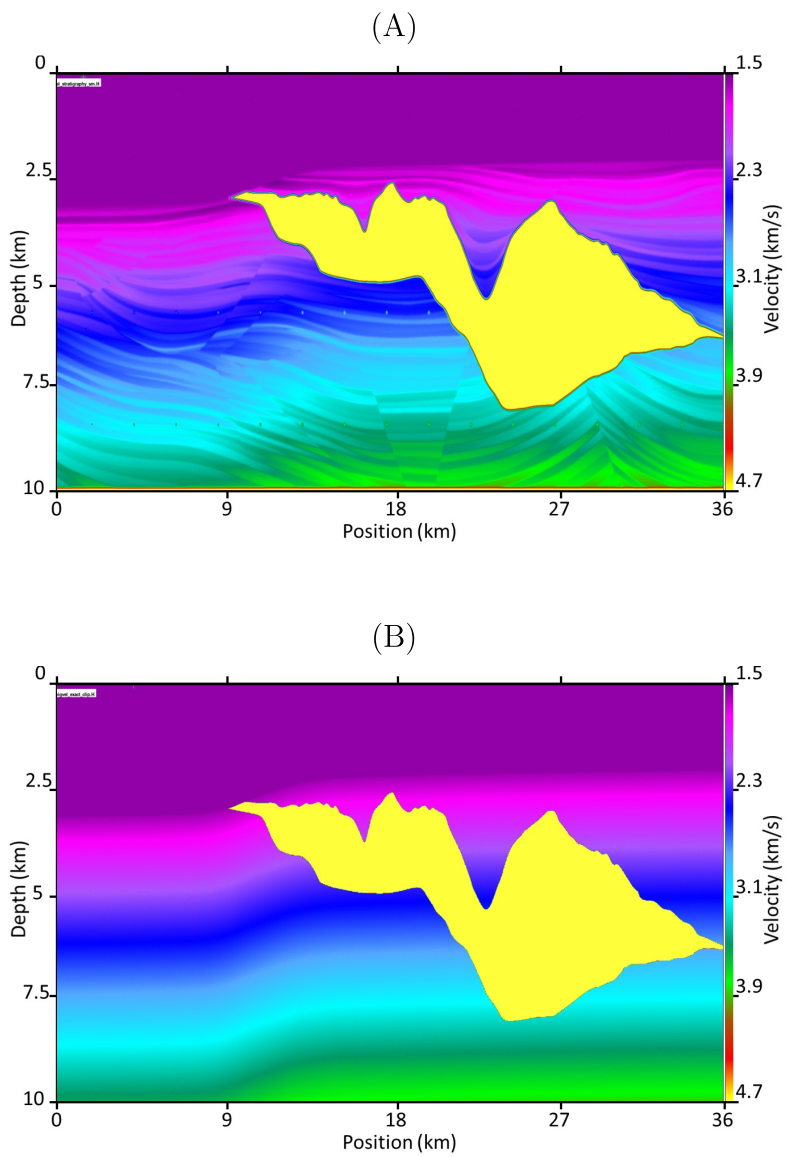
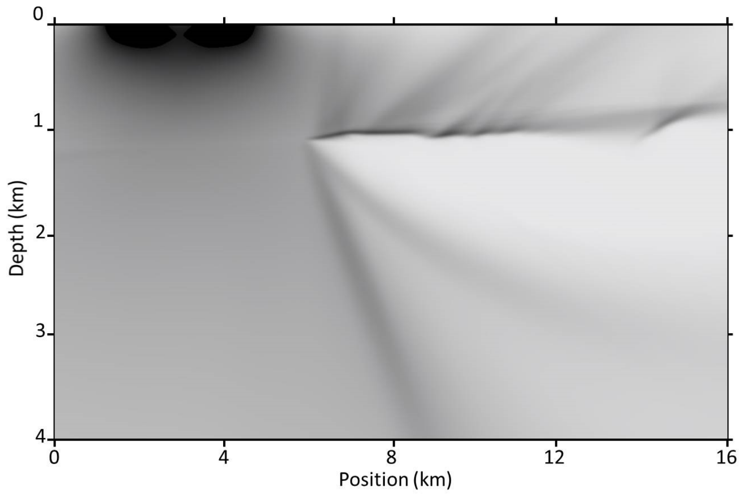


Figure 4.1: Velocity models for modeling and migration. (A) Sigsbee2B stratigraphic velocity used for forward modeling (B) Migration velocity for migration with the forward modeled data.

The synthetic simulation is performed using the Sigsbee2B model (Paffenholz et al., 2002). The seismic data is first forward modeled using a fine stratigraphic velocity model as shown in Figure 4.1(A). The migration velocity is shown in Figure 4.1(B), which is smoothed version of the original stratigraphic velocity. Source and receivers are put in the surface level to avoid the impact of ghost reflections. Acoustic forward modeling is used to generate the synthetic shots with an absorbing boundary to eliminate the surface related multiples. A total of 680 seismic shots were generated with maximum offset of 8000 m , similar to that seen for real-world acquisition with a receiver spacing of 12.5 m.

We first investigated the preconditioning scheme on the LSRTM gradient by comparing the gradients with and without Hessian preconditioning. This is shown by a two-shot migration covering both the sediment and subsalt region of Sigsbee2B. The low frequency backscatter is first removed though a Laplacian filter (Zhang and Sun, 2009). The comparison of the original Hessian and the smoothed one is presented in Figure 4.2 (A) and (B). The smoothing retains the main Hessian contribution while attenuates the strong variations in the illumination path. This stripy pattern comes from the destructive interference of the wave traveling over the rugose salt boundary. The gradients of the LSRTM with and without Hessian preconditioning are shown in Figure 4.3. It is clear that with Hessian preconditioning, the resulting gradient amplitudes are more balanced, especially in the deeper sessions of the image. The deeper reflectors are better illuminated due to the Hessian preconditioning.

(A)



(B)

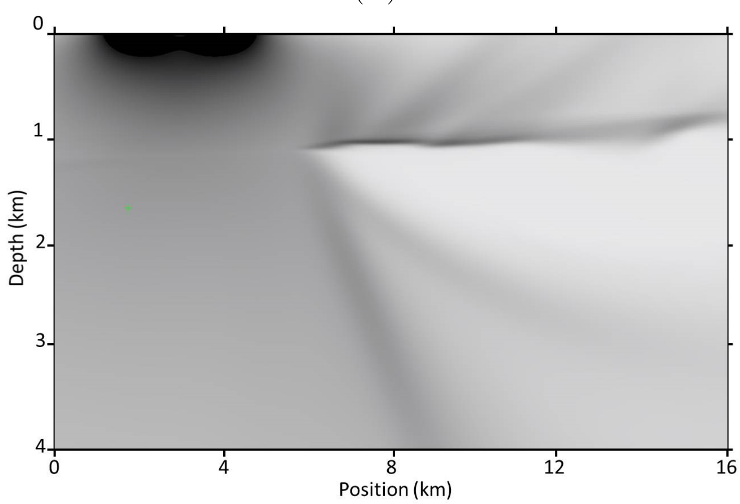


Figure 4.2: Hessian comparison between with and without smoothing operator. (A) Hessian without smoothing (B) Hessian with smoothing.

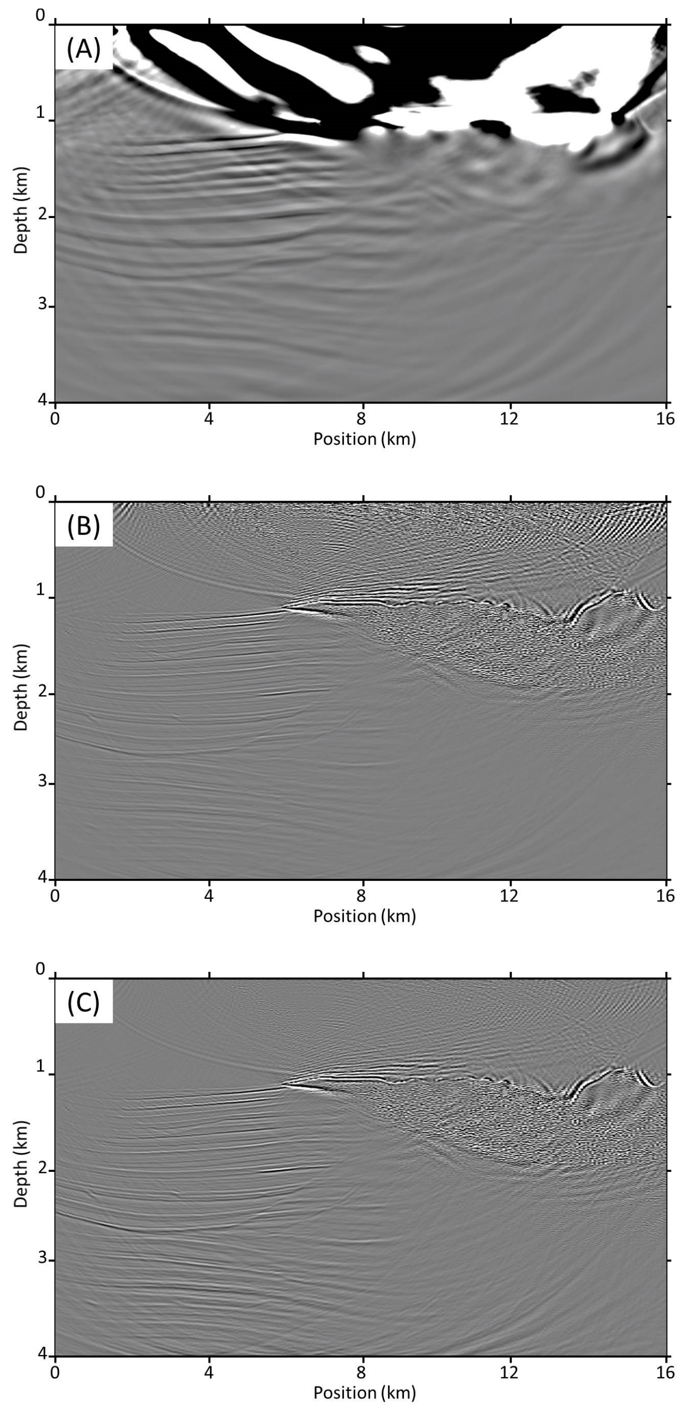


Figure 4.3: Comparison of gradients generated by different methods. (A) RAW RTM gradient (B) Laplacian-filtered RTM gradient (C) Hessian-preconditioned Laplacian-filtered RTM gradient.

The final migrated images using RTM, LSRTM and Hessian preconditioned LSRTM are shown in Figure 4.4. The LSRTM result of the sedimentary structure on the left is more sharper than the RTM results. It is prominent that the salt flank is more focused by the least-squares inversion process, as shown in the comparison of RTM and LSRTM image on the subsalt area. The sedimentary layers beneath the salt are better imaged in LSRTM with the Hessian preconditioning, when compared with the regular LSRTM. We also compare the convergence rates with and without Hessian preconditioning using a conjugate gradient method, where the normalized misfit over 15 iterations is shown in Figure 4.5. From the convergence comparison in Figure 4.5, it can be conclusive that better convergence is achieved through the simplified efficient Hessian preconditioning.

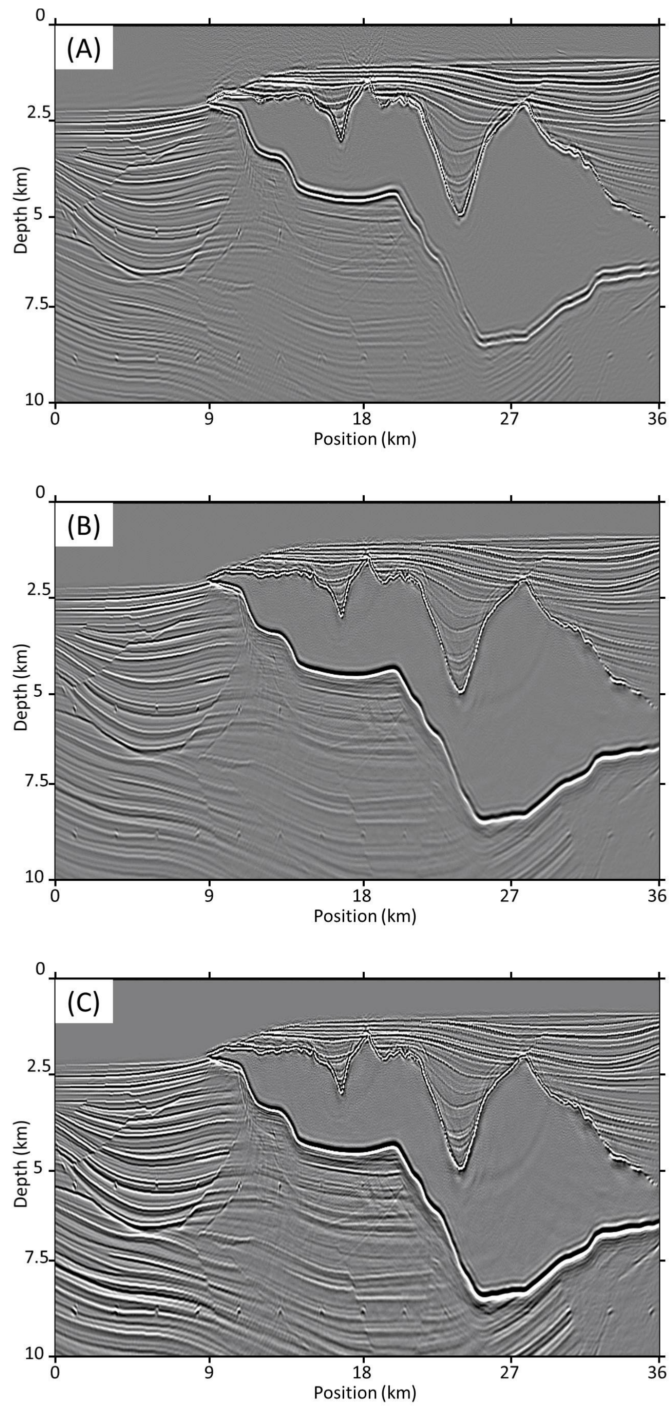


Figure 4.4: Comparison of images generated by different methods. (A) RTM image (B) LSRTM image without Hessian preconditioning (C) LSRTM image with Hessian preconditioning.

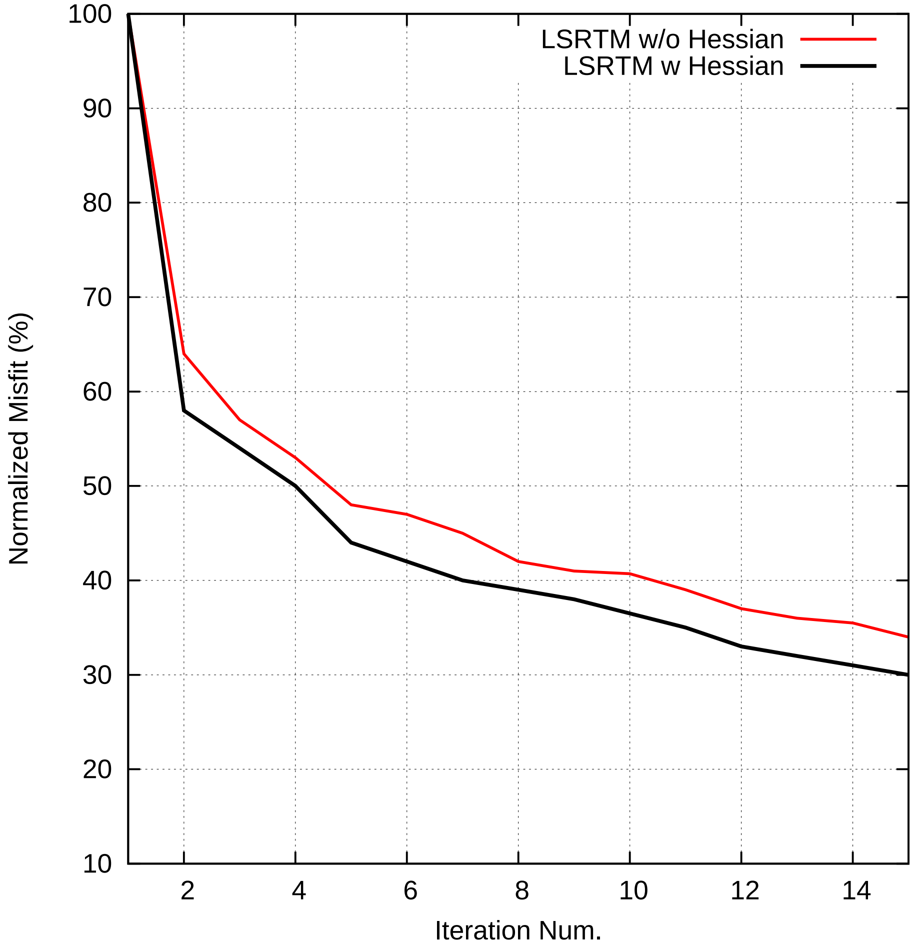


Figure 4.5: Convergence comparison between LSRTM and LSRTM with Hessian preconditioning.

4.5 Conclusion

LSRTM is an attractive method for high resolution imaging; however, the computational cost is extraordinary high due to multiple iterations of the migration and forward modeling process. The Hessian matrix in least-squares migration can be used as a preconditioning operator to compensate geometrical spreading effects and improve the performance of the least-squares inversion. The off-diagonal part of the Hessian is ignored and the receiver side Hessian is approximated by the source side Hessian, assuming reciprocity. In addition to approximating the receiver side Hessian, we further smoothed the Hessian to reduce the destructive contribution from rugose salt boundary to improve the amplitude balancing. In our synthetic test over the Sigsbee2B model, our proposed method can generate more balanced images than the RTM and traditional LSRTM.

Chapter 5

Locally normalized least-squares reverse-time migration: Field examples

Conventional least-squares migration uses acoustic modeling for a reflectivity image that best matches the amplitude of the observed data. However, real-world datasets are more challenging to match the amplitude than in the synthetic case. This can be attributed to several reasons, first, most migration algorithms are based on an acoustic approximation, which cannot account for the elastic effects in real data, such as amplitude attenuation. Second, amplitude information is more easily distorted by preprocessing and variation of actural seismic sources. In practice, we are more interested in the relative amplitude and location of the reflectors from a seismic image. To relax the requirement to match the signal amplitude, we proposed a localized amplitude correction for the objective function. Mathematically, the amplitude normalization scheme is equivalent to a time-domain phase inversion, where the inversion focuses on matching the phase of the data rather than the amplitude.

5.1 Introduction

Least-squares migration (LSM) has been shown to produce images with balanced amplitudes, higher resolution, and much less migration artifacts than conventional migration (Dai and Schuster, 2013; Nemeth et al., 1999; Zhang et al., 2013). The seismic image quality can be improved by matching the amplitude and phase of the observed data and modeled data under Born approximation, where the wave is scattered only once.

The standard implementation of LSM generally uses the L^2 norm of the direct difference between the observed and modeled data as the objective function. Image update is done by calculating the gradient of the L^2 norm objective function, and the optimal image is achieved by minimizing the L^2 norm of the difference. The implementation emphasizes the matching in both amplitude and phase of the observed and modeled data. However, real-world seismic data is more complicated than the synthetic data, it is challenging to match the amplitude directly (Zhang et al., 2013).

There are several reasons that prevent a successful matching of the seismic amplitude between observed and modeled data. First, most current seismic models are based on an acoustic approximation, while the real earth is visco-elastic. Generally, the estimation of an attenuation parameter Q is difficult (Zhang et al., 2012), which is computationally demanding in general. Also, reflection seismic data is determined by both the velocity and the density. However, the density is often assumed to be constant in most applications. Small scale density variations contributes to the seismic amplitude variation in most cases, and further increase the difficulty in matching the amplitude.

The amplitude information can be further distorted by pre-processing and variation of seismic source signatures. Real-world seismic data has undergone numerous pre-processing for final imaging, such as using different shaping filters and denoise operations. Variation of seismic source signatures can be due to cable sensitivity or natural variation of source guns, which is hard to predict and thus makes matching the amplitude more difficult.

In this study, the objective function for least-squares reverse-time migration is formulated as the difference between observed data and locally normalized modeled data. The new adjoint source is calculated according to the normalization scheme. Such an implementation is partially equivalent to the time-domain phase inversion method where the phase spectra of the observed data are matched with that of the modeled data (Luo and Schuster, 1991; Routh et al., 2011; Zhang et al., 2013). This method has more freedom to be implemented in different scales, such as different time window or different offset windows. Field test was done in this study to demonstrate the effectiveness of this method.

5.2 Theory

The detailed derivation of least-squares reverse-time migration has been discussed in Chapter 4, where the two-way wave equation is used as the modeling operator. The conventional objective for least-squares reverse-time migration for a given source s, and receiver g can be written as

$$J(\mathbf{m}) = \sum_{s=1}^{ns} \sum_{g=1}^{ng} \|\mathbf{d}_{mod,g,s} - \mathbf{d}_{obs,g,s}\| \tag{5.1}$$

where ns is the total source number, ng is the total receiver number, $\mathbf{d}_{mod,g,s}$ and $\mathbf{d}_{obs,g,s}$ are the data vector for modeled and observed data, respectively. To compensate the large amplitude difference, a result of improper amplitude simulation in the modeling process, the locally normalized objective function for least-squares reverse-time migration can be written as

$$J(\mathbf{m}) = \sum_{s=1}^{ns} \sum_{g=1}^{ng} \frac{1}{2} \left\| \frac{\|\mathbf{d}_{obs,g,s}\|}{\|\mathbf{d}_{mod,g,s}\|} \mathbf{d}_{mod,g,s} - \mathbf{d}_{obs,g,s} \right\| \tag{5.2}$$

The normalization is done on each trace. This can be further extended to window by window normalization. Apparently, the physical meaning of the normalization is to normalize the response for each source s and receiver g pair. By making derivative in relative to model operator $\mathbf{m}$ and the gradient can be written as

$$\frac{\partial J(\mathbf{m})}{\partial \mathbf{m}} = \frac{\partial \mathbf{d}^T_{mod,g,s}}{\partial \mathbf{m}} \frac{\|\mathbf{d}_{obs,g,s}\|}{\|\mathbf{d}_{mod,g,s}\|} \left[\frac{\mathbf{d}^T_{mod,g,s}\mathbf{d}_{obs,g,s}}{\|\mathbf{d}_{mod,g,s}\|^2} \mathbf{d}_{mod,g,s} - \mathbf{d}_{obs,g,s} \right] \tag{5.3}$$

The adjoint source $\mathbf{d}_{adj}$ (Plessix, 2006), which is to propagate in the receiver side, can be written as

$$\mathbf{d}_{adj} = \frac{\|\mathbf{d}_{obs,g,s}\|}{\|\mathbf{d}_{mod,g,s}\|} \left[\frac{\mathbf{d}^T_{mod,g,s}\mathbf{d}_{obs,g,s}}{\|\mathbf{d}_{mod,g,s}\|^2} \mathbf{d}_{mod,g,s} - \mathbf{d}_{obs,g,s} \right] \tag{5.4}$$

The detailed derivation of the gradient with locally normalized objective function is written at Appendix A. We can see that the gradient or update is calculated by following the steps. First, the cross-product of the observed data and the modeled data are calculated. It is then normalized by the norm of the modeled data. The resulting scalar is applied to the modeled data to scale the synthetic data. For example, if $\mathbf{d}_{obs}$ equals $\mathbf{d}_{mod}$, normalization will result in a zero value for the objective function. Similarly, if $\mathbf{d}_{obs}$ has the same amplitude distribution over the offset and time axis as $\mathbf{d}_{mod}$, only a global amplitude difference, Equation 5.3, will still result in a zero value. This suggests we already have a good reflectivity model. The normalization prescribed by Equation 5.3 is a general form for the data point to point normalization, and can be easily transformed into a shot by shot normalization, or a window by window normalization.

To compare the gradient of classic L^2 norm objective function with the proposed locally normalized objective function, the gradient of classic L^2 norm objective function is given as

$$\frac{\partial J(\mathbf{m})}{\partial \mathbf{m}} = \frac{\partial \mathbf{d}^T_{mod,g,s}}{\partial \mathbf{m}} \left(\mathbf{d}_{mod,g,s} - \mathbf{d}_{obs,g,s}\right) \tag{5.5}$$

where the adjoint source is defined as

$$\mathbf{d}_{adj} = \mathbf{d}_{mod,g,s} - \mathbf{d}_{obs,g,s} \tag{5.6}$$

The adjoint source of the regular direct difference denoted by Equation 5.6, is similar to that of normalized objective function, which is given by Equation 5.4. The difference between a classic objective function and the normalized objective function is that the modeled data is scaled by the ratio of a cross-product between the modeled and observed to the norm of observed data. In summary, the proposed normalized objective function is a more general definition and the classic objective function can be regarded as a special case of the normalized objective function.

Numerical implementation of the gradient $\mathbf{g}$ is done by taking the zero-lag cross-correlation of the forward propagated source wave-field and back-propagated adjoint source. In real-world seismic data, the true modeling operator is generally non-linear. The forward modeling and migration operators are not exactly adjoint. It is difficult to achieve adjoint by using a RTM operator, which may lead to an inaccurate step length by using the conjugate gradient method. For the field data application, the more practical quadratic line search was implemented. The iterative update can be written by

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha \mathbf{g}_k \tag{5.7}$$

where α is the optimal step length derived from a line search along the gradient direction using a parabolic fit (Nash, 1979), which is similar to the line-search used in full-waveform inversion, e.g., the method used by (Vigh and Starr, 2008).

5.3 Field examples

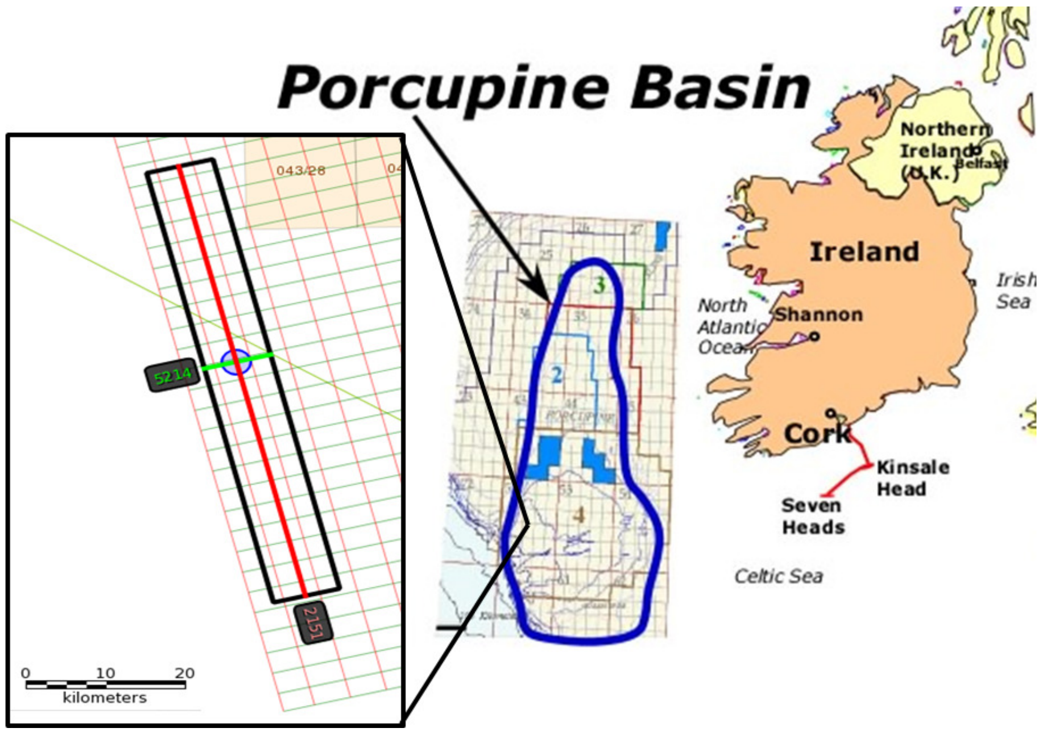


Figure 5.1: Field data location map. The field data are acquired in the Porcupine basin located west of Ireland.

To validate our approach, we applied our implementation on a dataset from Porcupine basin. The location of the dataset is shown in Figure 5.1. The Porcupine basin has proved to be productive since the 1970s. The field data is composed of two passes of sparse sources at a distance of 150 m and sparse crossline about 200 m. The survey has a lateral distance of 45 km and width about 4 km, which has about

3000 shots in total. The data has been processed only with preliminary denoise.

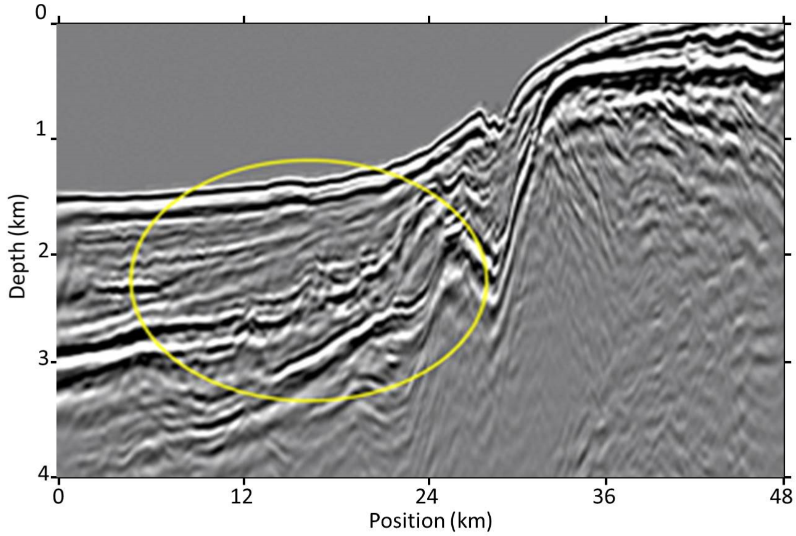
Due to the high computational cost by LSRTM, a total of 10 iterations of updates were performed. Figure 5.2 shows the inline images from the RTM and the implemented LSRTM with both normalization and Hessian preconditioning, where we can see the amplitude is more balanced inside the yellow circle. The reflection event in the sediment layer show enhanced details seen in Figure 5.2, suggesting a higher spatial resolution. Figure 5.3 shows the depth slices at 2500 m, and demonstrates the ability of LSRTM to generate better geological structures in comparison to conventional RTM. The depth slice clearly demonstrates the benefit of LSRTM, with a better fault delinearization and better horizontal structure seen at the center of the basin.

In order to have a more quantitative view of the amplitude variations of the seismic images, we extract the amplitude information by using a moving average absolute value for different image windows, which can be written as

$$I(x_i) = \sum_{j=-k,k}^{n} I(x_{i+j}) \tag{5.8}$$

$I(x_i)$ is the amplitude for specific imaging point and k is the size of the moving window. Practically, the amplitude information can be extracted by the amplitude gain control (AGC). The amplitude distributions extracted from RTM and LSRTM migration are shown in Figure 5.4. The yellow circle on the top image of Figure 5.4 clearly shows the amplitude variation stripes due to acquisition gap of shots, which is a result of the sparse sources; while in the amplitude map for LSRTM, which is the bottom of Figure 5.4, the acquisition footprint is partially removed due to the inversion process in LSRTM, which is prominent inside the yellow circle. Under the assumption that earth reflectivity is *white*, the amplitude of migrated image should be well balanced with much less horizontal variations due to acquisition or illumination. The amplitude comparison by Figure 5.4 justifies that the LSRTM can handle, at least partially, the acquisition footprint.

(A)



(B)

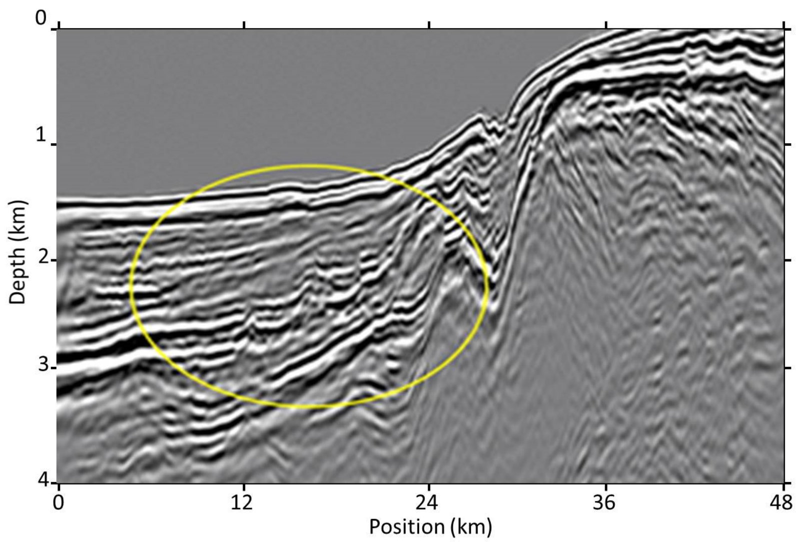
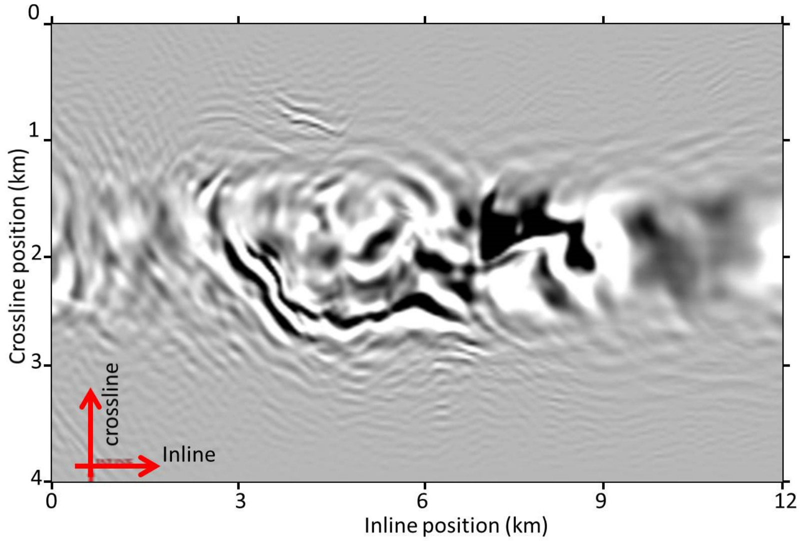


Figure 5.2: Image comparison between the RTM image and the LSRTM image in the inline direction. (A) RTM image of the field data in the inline direction (B) LSRTM image of the field data in the inline direction.

(A)



(B)

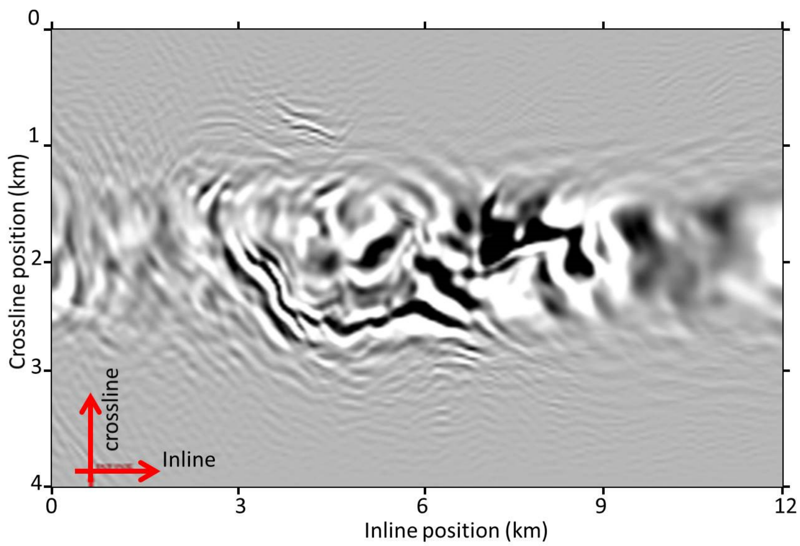


Figure 5.3: Image comparison between the RTM image and the LSRTM image in the depth slice. (A) Depth slice of the RTM image at 2500 m depth (B) Depth slice of the LSRTM image at 2500 m depth.

(A)

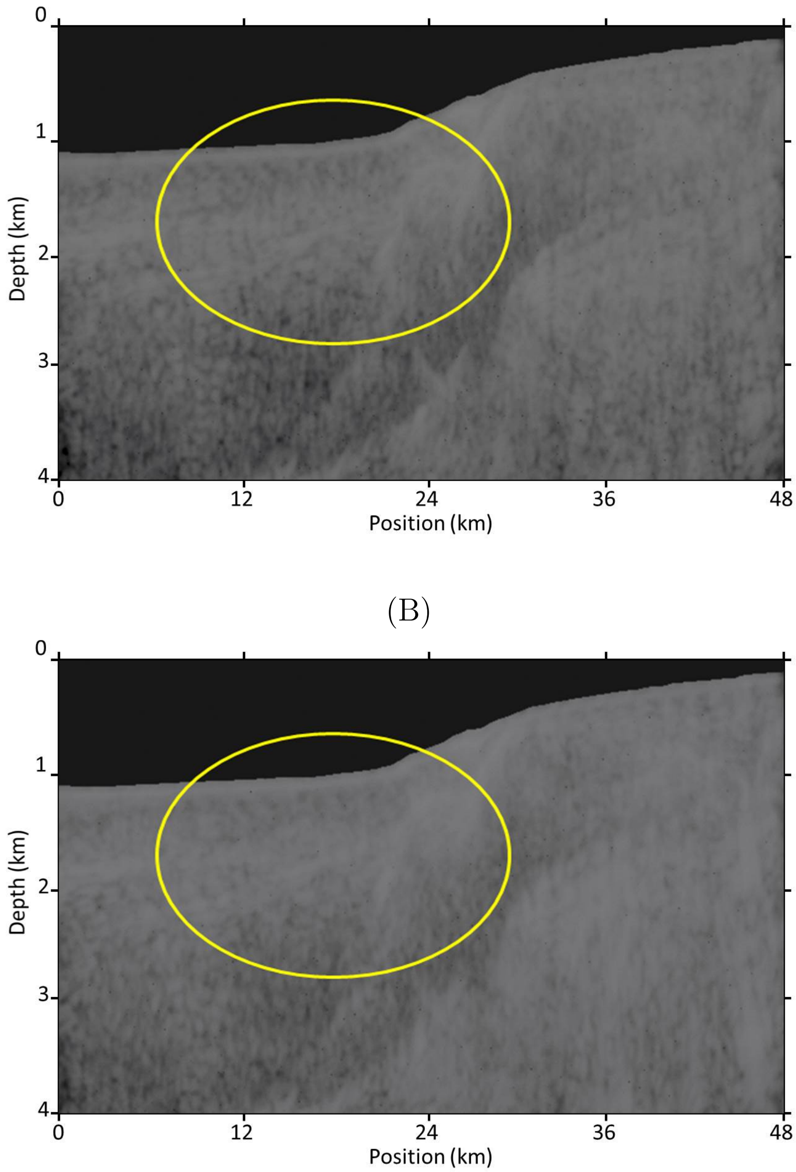


Figure 5.4: Amplitude comparison between the RTM image and the LSRTM image. (A) Amplitude map of the RTM image in the inline direction (B) Amplitude map of the LSRTM image in the inline direction.

5.4 Conclusion

Least-squares reverse-time migration with a locally-normalized objective function is presented in this study. The local-normalization emphasizes matching the phases of the observed data and modeled data. The normalization is to take into account the real-world amplitude variations due to elastic effects, as well as source signal variations. By comparing the classic misfit objective function, we concluded that the classic misfit objective is a special case prescribed by the normalized objective function. The proposed normalized objective function has more freedom to be implemented in different scales, such as time window based normalization for practical applications.

We apply our implemented LSRTM to a field dataset, where both Hessian preconditioning and locally normalized objective functions are applied. The resulted image from LSRTM is much better than the conventional RTM in terms of amplitude balancing and resolution. The observed strong acquisition footprints in RTM image, due to sparse sources, are partially removed by LSRTM.

Chapter 6

Conclusion

With the development of computational power, inversion-based imaging and model-building algorithms have received a lot of attention in recent years. These inversion-based techniques normally involve *modeling* and *back projection* processes, which will generate stability and efficiency issues. The focus of this study is to optimize the performance of least-square inversions, e.g. FWI and LSM, as well as making them applicable for real-world industrial-sized datasets.

The current established workflow for FWI in industrial application uses iterative minimization of the data residuals. The inversion generally progresses from low to high frequencies in order to mitigate the effects of nonlinearity and relies on the use of diving wave energy to achieve a low frequency model. In absence of an accurate initial model, FWI will fail in most real-world cases. Under classic FWI implementation, the velocity model is not decomposed into *background* and *reflectivity* component. I propose a multi-level temporal integration (MTI) for better performance of FWI that shows better results in the synthetic case . The MTI uses temporal integration to maximize the low-frequency content of the data difference, where we derive the relation between temporal integrated source wavefield and modeled wavefield. However, the MTI FWI shares the same migration kernel as that of RTM and can still be problematic when the velocity errors are large. Future development may include combining the MTI FWI with the reflection FWI to a better extraction of the background velocity. In real world seismic data, the low frequency data is contaminated by different noises, which adds additional difficulties in applying this method.

Migration is a complimentary component of velocity model building and in general provides a more direct view of the subsurface structure. Mathematically, the migration operator is an adjoint operator, which we use to approximate the inverse operator. However, the quality of migration can be limited due to the migration spreading effects, which blurs the image through point spread function. In order to achieve high quality image, least-squares inversion techniques are used to to achieve the exact inverse operator.

Although the least-squares migration (LSM) can generate high quality image data, there are a lot of challenges for LSM. The challenge of least squares migration lies in the fact that multiple migration and demigration processes have to be performed. LSM demands a much higher computational cost in comparison to the conventional migration, especially when wave-equation based imaging algorithms are used. In order to make the LSM less expensive for field data, I propose using stochastic conjugate gradient method and Hessian preconditioning to make the LSM more efficient. In the synthetic test, the stochastic conjugate gradient method generally perform much better, and can successfully generate better image. The stochastic conjugate method includes using different sample batches to update the model iteratively. The conjugate gradient directions are calculated over the same sample space for several iterations to make the convergence faster. A simplified Hessian preconditioning is also proposed in this study. The Hessian precondition is proved to be very efficient. The physical interpretation of Hessian preconditioning is that the receiver side illumination can be approximated as the source side. This assumption is generally valid for marine seismic surveys.

Application of the least-squares inversion on field data is more challenging than on the synthetic data. The *ground truth* for synthetic data is generally known, and there are more uncertainties in field data. In order to apply the least-squares reverse-time migration to field data, I proposed a normalization scheme in the objective function, which can partially remove the amplitude difference between field and modeled data. The new normalized objective function can preserve the phase and reflector location information. I applied the LSRTM incorporating the normalized objective function and the Hessian preconditioning to a porcupine field dataset. The result of LSRTM from the field data shows that the LSRTM outperforms the conventional RTM, which suggests LSRTM can be promising for future development. In the field example, the real-world forward wave propagator is far from the adjoint of the Born operator, which makes the standard line search more efficient than the conjugate gradient method in general. However, the convergence of real-world least-squares migration is extremely slow, suggesting future development on preconditioning the gradient or image using higher order of Hessian is needed.

Appendices

Appendix A

Normalized objective function for field data

We propose the following normalized objection function in the LSRTM for amplitude correction when we deal with real data

$$J(\mathbf{m}) = \frac{1}{2}\left|\left|\frac{\|\mathbf{d}_{obs}\|}{\|\mathbf{d}_{mod}\|}\mathbf{d}_{mod} - \mathbf{d}_{obs}\right|\right| \tag{A.1}$$

where $\mathbf{m}$ is the reflectivity model we want to invert, $\mathbf{d}_{mod}$ is the modeled data, and $\mathbf{d}_{obs}$ is the field data. The normalization can be done on synthetic data by shot, trace by trace, and window by window. A seismic trace can be treated as a data vector, where each sample is a function of the model space. Before we derive the Frechet derivative of $J(\mathbf{m})$, we make the derivation of the data vector in according to model vectors. Assuming we have function $g(x)$, which is a norm of a function vector, and can be described as

$$g(x) = \|f(x)\| \tag{A.2}$$

where $f(x)$ is is a function vector

$$f(x) = \sum_{i=1}^{n} f_i(x) \tag{A.3}$$

This makes $g(x) = (\sum_{i=1}^{n} f_i^2(x))^{\frac{1}{2}}$. By making derivative of $g(x)$ in relative to x, we have

$$\frac{\partial g(x)}{\partial x} = \frac{1}{2}\left(\sum_{i=1}^{n} f_i^2(x)\right)^{\frac{1}{2}} \left(2 f_i(x) \frac{\partial f_i(x)}{x}\right) \tag{A.4}$$

which can be compactly written in the vector form as

$$\frac{\partial g(x)}{\partial x} = \frac{1}{\|f(x)\|}\left(\frac{\partial f(x)}{\partial x}\right)^{T} f(x) \tag{A.5}$$

where ^{T} denote transpose operator. After reordering, the gradient of $J(\mathbf{m})$ in relative to $\mathbf{m}$ can be written as

$$\frac{\partial J(\mathbf{m})}{\partial \mathbf{m}} = \frac{\partial}{\partial \mathbf{m}}\left(\frac{\|\mathbf{d}_{obs}\|}{\|\mathbf{d}_{mod}\|}\mathbf{d}_{mod}^{T}\mathbf{d}_{obs} - \mathbf{d}_{obs}^{T}\frac{\|\mathbf{d}_{obs}\|}{\|\mathbf{d}_{mod}\|}\mathbf{d}_{mod}\right) \tag{A.6}$$

By using the derivation from Equation A.5, the gradient $\mathbf{g} = \frac{\partial J(\mathbf{m})}{\mathbf{m}}$ can be described as

$$\frac{\partial J(\mathbf{m})}{\partial \mathbf{m}} = \frac{\partial \mathbf{d}_{mod}^{T}}{\partial \mathbf{m}} \frac{\|\mathbf{d}_{obs}\|}{\|\mathbf{d}_{mod}\|}\left[\frac{\mathbf{d}_{mod}^{T}\mathbf{d}_{obs}}{\|\mathbf{d}_{mod}\|^2}\mathbf{d}_{mod} - \mathbf{d}_{obs}\right] \tag{A.7}$$

where the residual between the normalized synthetic data and the field data is the term

$$\mathbf{d}_{norm,residual} = \frac{\|\mathbf{d}_{obs}\|}{\|\mathbf{d}_{mod}\|}\left[\frac{\mathbf{d}_{mod}^{T}\mathbf{d}_{obs}}{\|\mathbf{d}_{mod}\|^2}\mathbf{d}_{mod} - \mathbf{d}_{obs}\right] \tag{A.8}$$

which can be physically explained by backpropagating the data residual, $d_{norm,residual}$, and forward propagating the source signature to calculate the gradient. After computing the gradient $\mathbf{g}$, the reflectivity update can be written as

$$\mathbf{m}_{n+1} = \mathbf{m}_n + \alpha\mathbf{g} \tag{A.9}$$

where α is the update step length, which can be achieved by a standard quadratic line search or a conjugate gradient method.

Bibliography

Albertin, U., P. Sava, and J. Etgen, 2006, Adjoint wave equation velocity analysis: Society of Exploration Geophysicists Annual Meeting.

Alkhalifah, T., and I. Tsvankin, 1995, Velocity analysis in transversely isotropic media: Geophysics, **60**, 1550–1566.

Backus, G., and F. Glbert, 1968, The resolving power of gross earth data: Geophysical Journal of the Royal Astronomical Society, **16**, 169–205.

Bamberger, A., G. Chavent, C. Hemon, and P. Lailly, 1982, Inversion of normal incidence seismograms: Geophysics, **47**, 757–770.

Baysal, E., D. Kosloff, and J. Sherwood, 1983, Reverse time migration: Geophysics, **56**, 897–906.

Berkhout, A., 1982, Imaging of acoustic energy by wave field extrapolation: Elsevier Scientific Publishing Company.

Berkhout, A., 1984, Seismic resolution: A quantitative analysis of resolving power of acoustical echo techniques: Geophyical Press.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Chen, J., and G. Schuster, 1999, Resolution limits of migrated images: Geophysics, **64**, 1046–1053.

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

Claerbout, J. F., 1985, Imaging the earth's interior: Stanford University.

Dai, W., and G. Schuster, 2013, Plane-wave least-squares reverse time migraiton: Geophysics, **78**, S165–S177.

Dutta, G., M. Sinha, and G. Schuster, 2013, A cross-correlation objective function for least-squares migration and visco-acoustic imaging: Society of Exploration Geophysicists Annual Meeting.

Farmer, P., Z. Zhou, and D. Jones, 2009, The role of reverse time migration in imaging and model estimation: The Leading Edge, **28**, 436–441.

Fichtner, A., and J. Trampert, 2011, Resolution analysis in full waveform inversion: Geophysical Journal International, **187**, 1604–1624.

Fletcher, R., and C. Reeves, 1964, Function minimization by conjugate gradients: The Computer Journal, **7**, 149–154.

Gao, F., A. Atle, and P. Williamson, 2010, Full waveform inversion using deterministic source encoding: Society of Exploration Geophysicists Annual Meeting.

Godwin, J., and P. Sava, 2010, Blended source imaging by amplitude encoding: Presented at the 80th Annual International Meeting, Society of Exploration Geophysicists, Expanded abstracts.

Guitton, A., A. Valenciano, B. D., and J. Claerbout, 2006, Smoothing imaging condition for shot-profile migration: Geophysics, **72**, S149–S154.

Ha, W., C. Shin, and T. Ha, 2012, Efficient laplace-domain modeling using an axis transformation technique: Society of Exploration Geophysicists Annual Meeting.

Hopkins, K., and B. Moore, 2007, Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information: Journal of the Acoustical Society of America, **122**, 1055–1068.

Huang, W., J. Kun, D. Vigh, J. Kapoor, K. Jiao, X. Cheng, and D. Sun, 2012, Application of full waveform inversion for salt sediment inclusion inversion: Society of Exploration Geophysicists Annual Meeting.

Huang, W., H. Ma, D. Vigh, J. Kapoor, K. Jiao, X. Cheng, and D. Sun, 2013, Velocity model building with long-offset and full-azimuth data: A case history for full-waveform inversion: Society of Exploration Geophysicists Annual Meeting.

Huang, W., and H. Zhou, 2014, Stochastic conjugate gradient method for least-square seismic inversion problems: Society of Exploration Geophysicists Annual Meeting.

Huang, W., and H. Zhou, 2015, Least-squares seismic inversion with stochastic conjugate gradient method: Journal of Earth Science, **26**, 473–470.

Jackson, D. D., 1972, Interpretation of inaccurate, insufficient, and inconsitent data: Geophysical Journal of the Royal Astronomical Society, **28**, 97–109.

Jiang, H., and P. Wilford, 2013, A stochastic conjugate gradient method for the approximation of functions: Computional and Applied Mathematics, **236**, 2529–2544.

Krebs, J., J. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M. Lacasse, 2009, Fast full-wavefield seismic inversion using encoded sources:

Geophysics, **74**, WCC177–WCC188.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on inverse scattering, theory and application, 206–220.

Luo, J., and R. S. Wu, 2013, Envelope inversion – some application issues: Society of Exploration Geophysicists Annual Meeting.

Luo, Y., and G. T. Schuster, 1990, Wave-equation traveltime + waveform inversion: Society of Exploration Geophysicists Annual Meeting.

Luo, Y., and G. T. Schuster, 1991, Wave equation traveltime inversion: Geophysics, **56**, 645–653.

Mackay, S., and R. Abma, 1992, Imaging and velocity estimation with depth focusing analysis: Geophysics, **57**, 1608–1622.

Moore, B., and C. T. Tan, 2003, Perceived naturalness of spectrally distorted speech and music: Journal of the Acoustical Society of America, **1142**, 403–409.

Mothi, S., and R. Kumar, 2014, Detecting and estimating anisotropy errors using full waveform inversion and ray-based tomography: A case study using long-offset acquisition in the gulf of mexico: Society of Exploration Geophysicists Annual Meeting.

Mulder, W. A., and R. E. Plessix, 2008, Exploring some issues in acoustic full waveform inversion: Geophysical Prospecting, **56**, 827–841.

Nash, J. C., 1979, Compact numerical methods form computers: Linear algebra and functions minimization: Pitman Press.

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

Paffenholz, J., B. Mclain, J. Zaske, and P. Keliher, 2002, Subsalt multiple attenuation and imaging: Observations from the sigsbee2b synthetic data set: Society of Exploration Geophysicists Annual Meeting.

Plessix, R. E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503.

Plessix, R. E., and W. Mulder, 2004, Frequency-domain finite-difference amplitude-preserving migration: Geophysical Journal International, **157**, 975–987.

Pratt, G., C. Shin, and G. J. Hick, 1998, Gauss–newton and full newton methods in frequency–space seismic waveform inversion: Geophysical Journal International, **133**, 341–362.

Routh, P., J. Krebs, S. Lazaratos, A. Baumstein, I. Chikichev, S. Lee, N. Downey, D. Hinkley, and J. Anderson, 2011, Full-wavefield inversion of marine streamer data with the encoded simultaneous souce method: Presented at the 73rd, European Association of Geoscientists and Engineers, Expanded Abstracts.

Saad, Y., 2014, Iterative methods for sparse linear systems: Presented at the Society for Industrial and Applied Mathematics.

Safar, M. H., 1985, On the lateral resolution achieved by kirchhoff migration: Geophysics, **50**, 1091–1099.

Sava, P., 2004, Migration and velocity analysis by wavefield extrapolation: PhD thesis, Stanford University.

Sava, P., and B. Biondi, 2004, Wave-equation migration velocity analysis - I: Theory: Geophysical Prospecting, **52**, 593–606.

Schraudolph, N. N., and T. Graepel, 2003, Combining conjugate direction methods with stochastic approximation of gradients: 9th Intl. Workshop Artificial Intelligence and Statistics (AIstats), society for Artificial Intelligence and Statistic, Expanded Abstracts, 7–13.

Shen, P., 2004, Wave-equation migration velocity analysis by differential sembalance optimization: PhD thesis, Rice University.

Shin, C., S. Jang, and M. D., 2001, Improved amplitude preservation for prestack depth migration by inverse scattering theory: Geophysical Prospecting, **49**, 592–606.

Shin, J., W. Ha, H. Jun, D.-J. Min, and C. Shin, 2014, 3D laplace-domain full waveform inversion using a single GPU card: Computers & Geosciences, **67**, 1–13.

Stork, C., 1992, Reflection tomography in the postmigration domain: Geophysics, **57**, 680–692.

Symes, W. W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790.

Tang, Y., 2009, Target-oriented wave-equation least-squares migration with phase encoded hessian: Geophysics, **74**, WCA95–WCA107.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

Tarantola, A., 1987, Inverse problem theory: methods for data fitting and model parameter estimation: Elsevier Scientific Publishing Company.

Thomsen, L., 1986, Weak elastic anisotropy: Geophysics, **51**, 1954–1966.

Tibshirani, R., 1996, Regression shrinkage and selection via the LASSO: J. Royal

Statistical Society, **8**, 267–288.

Van Leeuwan, T., A. Aravkin, and F. Herrmann, 2011, Seismic wave inversion by stochastic optimization: International Journal of Geophysics.

Van Leeuwan, T., and W. A. Mulder, 2010, A correlation-based misfit criterion for wave-equation traveltime tomography: Geophysical Journal International, **182**, 1383–1394.

Vigh, D., and W. Starr, 2008, 3D prestack plane-wave full-waveform inversion: Geophysics, **73**, VE135–VE144.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC127–WCC152.

Wang, C., D. Yingst, R. Bloor, and J. Leveille, 2012, VTI waveform inversion with practical strategies: Application to 3D real data: Society of Exploration Geophysicists Annual Meeting.

Wang, C., D. Yingst, J. Brittan, P. Farmer, and J. Leveille, 2014, Fast multi-parameter anisotropic full waveform inversion with irregular shot sampling: Society of Exploration Geophysicists Annual Meeting.

Warner, M., and L. Guasch, 2014, Adaptive waveform inversion: Theory: Society of Exploration Geophysicists Annual Meeting.

Woodward, M. J., D. Nicols, O. Zdraveva, P. Whitfield, and T. Johns, 2008, A decade of tomography: Geophysics, **73**, VE5–VE11.

Yang, T., and P. Sava, 2011, Image-domain waveform tomography with two-way wave-equation: Presented at the 81st Annual International Meeting, Society of Exploration Geophysicists, Expanded abstracts.

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

Zhang, Y., L. Duan, and Y. Xie, 2013, A stable and practical implementation of least-squares reverse time migration: Society of Exploration Geophysicists Annual Meeting.

Zhang, Y., and J. Sun, 2009, Practical issues in reverse time migration: true amplitude gathers, noise removal and harmonic source encoding: First Break, **26**, 29–35.

Zhang, Y., P. Zhang, and H. Zhang, 2012, Compenstation for visco-acoustic effects with reverse time migration: Society of Exploration Geophysicists Annual Meeting.

Zhou, H., 2003, Multiscale traveltime tomography: Geophysics, **68**, 1639–1649.

Zhou, H., L. Amundsen, and G. Zhang, 2012, Fundamental issues in full-waveform inversion: Society of Exploration Geophysicists Annual Meeting.