Full waveform inversion using simultaneous encoded sources based on first- and second-order optimization methods

by

Amsalu Y. Anagaw

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

Doctor of Philosophy

in
Geophysics

Department of Physics
University of Alberta

© Amsalu Y. Anagaw, 2014

Abstract

Full waveform inversion (FWI) is an emerging seismic technology engine for estimating subsurface model parameters such as velocity, density and attenuation through local minimization of an objective function. The minimization problem is often performed iteratively via gradient-based method that is based on the first-order derivatives of the objective function. One of the major obstacles of FWI that hinders its practical applications for estimating subsurface model parameters for large scale problems is its huge computational demand for performing many forward modeling for multiple sources. One way to reduce the overall cost of waveform inversion is by adopting a simultaneous source strategy. In other words, multiple sources are simultaneously fired to simulate super-shot gathers and thereby reduce the number of seismic modeling simulations that are performed during the inversion. However, the use of simultaneous sources introduces cross-talk artifacts that arise from the interference among the sources that constitute a super-shot.

In this thesis, we study and extend the practical applications of the second-order optimization methods in the framework of simultaneous sources. First, we examined the effect of model parameterizations on velocity model building using Newton-based optimization methods. Three model parameterizations for the acoustic FWI were investigated. These include velocity, slowness and slowness squared. We then analyze the influence of different simultaneous multi-frequency selection strategies on cross-talk artifacts. Our analysis focuses on a multiscale frequency-domain FWI algorithm that is implemented with simultaneous sources that are randomly encoded with a source-encoding function. In the multiscale conventional FWI inversion strategies, a finite set of discrete frequencies will be selected and the inver-

sion is carried out sequentially from low to high frequency data components. We examine six frequency selection strategies and test the performance of the algorithm with encoded datasets. Numerical tests show that high fidelity results can be attained by inverting partially overlapped groups of temporal frequencies. In order to mitigate cross-talk artifacts during the numerical inversion, a new encoding is generated at every iteration. We also found that high resolution images can be obtained by re-sampling new source positions and new encoding functions at every iteration of the FWI algorithm. Finally, we present a full Newton FWI algorithm and its application in the presence of high amplitude multi-reflected waves that are generated by strong velocity contrast and/or from free surface reflections.

To my beloved daughter
Maya Amsalu

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Mauricio D. Sacchi for his support and guidance throughout my research. Without his full support and encouragement, my graduate studies and dissertation would not have been possible. Next, my gratitude goes to all my colleagues at the Signal Analysis and Imaging Group (SAIG). I am very honored to be a member of such a great group. I've been benefited from many people while I was at the department of physics and would like to acknowledge them. In particular, I would like to thank Aaron Stanton, Dr. Emmanuel Bongajum, Dr. Nadia Kreimer and Wubseht Alemie (in alphabetical order) for their help and scientific discussions I had over the course of my graduate studies. Aaron and Emmanuel helped me with editing of my publication papers and my dissertation. I am also indebted to Dr. Sam T. Kaplan who has helped me in setting up all the necessary tools of my research, in particular, at the early stages of my graduate studies in Geophysics. Thank you all!

I would like to extend my heartfelt thanks to members of my supervisory committee Dr. Douglas R. Schmitt, Dr. Yu Jeffrey Gu, Dr. Richard Sydora and Dr. Peter D. Minev and as well as my external examiner Dr. Kristopher A. Innanen. The input and feedback they provided me while I was in PhD program was valuable components of my graduate success.

Finally yet importantly, my PhD program would have been a lot difficult without the unparalleled support with so much love and compassion over the years I had from my sister Yeshifan Yigzaw. Our parents including me are always proud of you for such unwavering encouragement you have given me throughout my graduate schools and also the incredible permanent support you have done for us. Thank you very much for being with us whenever we needed you.

I am very grateful to my wife Tegist Gezehegn for being so understanding and the compromises that you made during the course of my long journey. My daughter Maya, who came into my life exactly a year before I start writing my dissertation, deserves special thanks. Time has come to you to have my full attention.

Special heartfelt thanks go to my parents: Alebachew Embial, Zemetu Mengesha,Adna

Alebachew, Dr.Yigzaw Anagaw, Birtukan Yigzaw, Zemenay Yigzaw and Mulat Anagaw for their inseparable support from time to time while I have been in schools. Thank you for your abiding support and love you have on me.

I would also like to thank my friends who have been with me in encouraging and supporting me over the journal of my research studies: Azamed Gezahagn, Lidya, Ayesheshim Ayesheshim, Aregaw Fereja, Muluken Gezehegn, Messele Fentabil and also thank to all my friends who I couldn't mention your names but part of my successful completion of my dissertation. Finally, I also want to thank the Department of Physics at the University of Alberta and the sponsors of SAIG.

Contents

List of Tables

List of Figures

List of symbols

List of Abbreviations

FWI ..... Full waveform inversion
SSFWI ..... Simultaneous source full waveform inversion
CG ..... Conjugate gradient
PCG ..... Preconditioned conjugate gradient
NLCG ..... Nonlinear conjugate gradient
l-BFGS ..... Limited-memory Broyden-Fletcher-Goldfarb-Shanno
MR ..... Minimum residual
PrMR ..... Preconditioned Minimum residual
GN ..... Gauss-Newton
FN ..... Full Newton
SNR ..... Signal to noise ratio
LU ..... Lower and Upper
KKT ..... Karush-Kuhn-Tucker
FD ..... Finite-difference
PML ..... Perfectly matched layer

CHAPTER 1

Overview of Full Waveform Inversion (FWI)

1.1 Traveltime tomography and Full waveform inversion

The main objective of full waveform inversion (FWI) is to produce a high-fidelity model parameters that best fits the observed seismic data at various scales for improved seismic imaging (Lailly, 1983; Tarantola, 1984b, 1987; Stekl and Pratt, 1998; Pratt et al., 1998; Operto et al., 2007; Virieux and Operto, 2009; Sourbier et al., 2009; Hu et al., 2011; Liu and Gu, 2012; Warner et al., 2013). Since the initial development of waveform inversion in the mid 1980's by Lailly (1983) and Tarantola (1984b), the feasibility of FWI has been largely discarded by the oil and gas industry mainly due to complex and expensive processing associated with solving many forward modeling simulations. Due to the vast advancement of computer hardware and performance, the application of full waveform inversion is becoming more practical for high resolution imaging, in particular for velocity model building, and is a major research and development focus area (Virieux and Operto, 2009; Vigh et al., 2010; Warner et al., 2013).

An accurate velocity model is crucial for conventional seismic migration methods, such as reverse time migration to produce depth images with reflectors in correct lateral and depth positions, and also to achieve better focusing. A less accurate but more versatile and widely used technique for velocity estimation from measured seismic field data are traveltime tomography (Pratt and Worthington, 1990; Williamson, 1991). However, migrated data with the updated velocity model from FWI have shown that images with higher resolution, better reflector positioning and improved focusing of events (Vigh et al., 2010; Sirgue et al., 2011; Warner et al., 2013). The standard FWI procedure starts from a good initial velocity

model. The initial velocity models for FWI are often obtained from early arrival traveltime tomography that are formulated based on reflected or refracted waves that best describes the field data (White, 1989; Kubota et al., 2009; Wang et al., 2014). Traveltime tomography is the most widely used technique due to its efficiency and robustness for long wavelength structure estimation of the Earth's physical properties, primarily the velocity, from seismic field data. The method works by iteratively updating a velocity model, often using a gradient descent method, to produce a satisfactory misfit between the model and observed data. Then, with full waveform inversion, as the inversion continues iteratively, a wide range of wavenumber components of the velocity model structures are estimated. Unlike FWI, where the complete seismic records are used, traveltime tomography only uses picked traveltime events. Only certain portions of data (the early arrivals) are used after some preprocessing and it employs a high frequency approximation. Full waveform inversion accounts for the full wavefield and has no high frequency limitation.

The spatial resolution of traveltime tomography is limited to the bandwidth of the first-order Fresnel zone $\sim \sqrt{\lambda d_{off}}$, where λ is the wavelength of the wavefield and d_{off} is the maximum offset distance (Williamson, 1991), whereas the spatial resolution of full waveform inversion is on the order of the wavelength $\sim \lambda/2$ of the propagated wavefield (Wu and Toksöz, 1987; Sirgue and Pratt, 2004; Liu and Gu, 2012). There are many reasons why FWI is preferable over conventional velocity model building, where the conventional methods fail. For example, complex near surface velocity structure in shallow areas have the potential for large, rapid velocity variations causing conventional ray based tomography methods to fail (Vigh et al., 2010; Warner et al., 2013).

1.2 State of the art of full waveform inversion

The flow of FWI starts by predicting the modeling wavefield data at receiver locations from a forward modeling engine using a starting model, source-receiver geometry, source wavelet and other necessary input parameters. Once the modeling wavefield data are predicted at the receiver positions, the residual wavefield data are computed by subtracting the modeled data from the observed datasets trace by trace at each receiver location (in the case of an l_2 norm). We then back-propagate the residual wavefield from the receiver locations using the same forward modeling engine. The gradient is then computed by taking the zero-lag cross-correlation between the forward and back-propagation wavefields everywhere within the model space. The velocity model is then updated iteratively using a line search method by perturbing the starting model in the appropriate descent direction. The process of updating the velocity continues until the change in the residual wavefield is minimized.

The most widely used engine for updating the velocity model are based on gradient descent

via linearized least-squares. The inversion scheme mainly relies on transmitted arrivals (diving and head waves) after preprocessing of the observed data (Virieux and Operto, 2009; Stekl and Pratt, 1998; Vigh et al., 2010; Warner et al., 2013). Transmitted energy (head waves and diving waves) are key to FWI success. In the preprocessing of the data, most events that are not modeled by the forward modeling are removed. For example, in the case of the acoustic wave approximation all shear wave energy must be removed. Typically FWI requires basic preprocessing such as careful de-noising (in particular to have good single-to-noise ratio in the low frequency range), band-pass filter to preserve the low frequency range (removing high frequencies), removing of the surface multiples, removing of elastic arrivals, muting of data that are not used for inversion (FWI uses both reflection and refraction waves, maybe mute everything except transmitted waves), as well as testing the accuracy of the starting model. Most parts of the data are kept constant (Vigh and Kapoor, 2009; Vigh et al., 2011; Warner et al., 2013).

With a smooth starting model, FWI first tries to approximately match traveltimes of the diving waves from the modeling to those linear head wave events of the observed data over the limited-offset range. As the iterative process continues, FWI proceeds to provide more details of velocity gradient updates as it fits the transmitted arrivals beyond the near-offset of diving waves where refractions exist. This actually holds if the modeled data are within half a cycle of the observed data throughout the inversion process, otherwise the inversion will not give a satisfactory result.

Understanding the properties of refracted and transmitted waves on the model are vital for high resolution velocity model building because this energy contains detailed information about the target area. FWI will not yield satisfactory results unless transmitted energy passed through the target area. The penetration depth of the transmitted energy depends on surface survey offsets; i.e. long offset acquisition is a key. Updating a specific depth also depends on a given geological setting and gradient in the velocity model.

Though much progress has been made in FWI in recent years, many problems remain that hinder its practical application to large scale 3D datasets. Unlike traveltime tomography, full waveform inversion is not robust. The waveform objective function is highly nonlinear, which can result in poor convergence properties and has the propensity to get trapped into one of many local minima. Even though FWI has the potential to produce accurate and high-resolution models from observed field data, the success of FWI depends on many factors. Below are the most common challenges faced in FWI

- FWI highly depends on the quality of low frequencies available in real data. In order to avoid solutions that are trapped in local minima, one needs to have data with high-quality low frequency components or provide a good initial velocity model (Brenders

and Pratt, 2007; Virieux and Operto, 2009). This is because at the lowest usable frequency, there is a high chance that the modeling and observed data match within half a cycle.

- FWI depends on the accuracy of the starting model. This is the most crucial requirement to have any success in model building. If the starting model does not produce the modeled wavefield data within half a wave-cycle of the observed data, FWI will tend to converge to one of the many local minima due to the wrapping around nature of phase in the frequency-domain or the cyclic-skipping problem in the time-domain (Bunks et al., 1995; Sirgue and Pratt, 2004). This makes the technique very sensitive to the starting velocity model especially when 3D data are considered. An accurate starting model is essential to avoid the cycle-skipping problem.

- FWI is also hampered by the lack of an efficient optimization method. FWI based upon least-squares is commonly solved through local optimization methods utilizing conjugate gradients or steepest descent. However, gradient optimization methods are generally considered slower and can have difficulties scaling the reconstructed model parameters in comparison to methods that require second-order information of the objective function, such as the Newton method.

- Even with the current drastic computer hardware and performance improvements, the application of full waveform inversion is still computationally expensive and requires considerable amount of computing memory for multiple sources and receivers, in particular, for large scale 3D problems.

- FWI also suffers from problems related to the high nonlinearity introduced by near surface effects due to presence of complex surface topography and propagation of surface waves/rayeligh waves in the strongly heterogeneous near surface (in particular a problem for land data). In addition, land datasets usually have a low signal to noise ratio in particular at low frequencies, incoherence of source level, incomplete illumination etc, which works against the requirements of FWI (Brenders and Pratt, 2007).

- For better efficiency and accuracy of velocity model building, FWI requires specific acquisition geometry: in particular large source to receiver offsets. The presence of large offset seismic data are important for FWI because it provides low wavenumber information of the subsurface and contains the deepest penetration of the transmitted energy (diving and head waves). It is at the longer offsets where these energies contains the most valuable information of deep events, and therefore produces better velocity updates. It is important to note that the achievable maximum depth in FWI to update the velocity model depends on the information of deep events contained in

the transmitted energy. FWI without long offset dataset, that does not pass through the desired target depth (especially without diving and refraction waves), will not be able to update the macro velocity model properly. Practically, due to the lack of good quality low frequency and long offset dataset, the existing FWI applications are often limited to near surface velocity model building: applicable primarily for updating shallower anomalies (Virieux and Operto, 2009; Vigh et al., 2010; Warner et al., 2013). However, the longer the offset present in the survey, the better the resolution and inversion results. Some recent applications of FWI performed to real data based on the acoustic wave approximation are very promising, though they require careful data preprocessing (Pratt and Shipp, 1999; Operto et al., 2006; Brenders and Pratt, 2007; Bleibinhaus et al., 2007; Warner et al., 2013)

– The accuracy of the estimated velocity model depends on the method of Physics used to describe the data we are inverting. Since FWI is based on data comparison between the modeled and observed data, we need to analyze our data such that all events in the observed data are also present in the modeling; otherwise the inversion will end up with a wrong result. Most available FWI algorithms are based on the acoustic wave-equation approximation and do not deal properly with elastic effects, shear velocity, density and any other anisotropy parameters. Elastic effects and any unwanted events present in the real data should be removed before the waveform inversion if the modeling is based on the acoustic wave approximation. Otherwise, other physical properties that affect the inversion, if possible, should be included in the modeling. One can also incorporate density, attenuation and other anisotropy parameters in the forward modeling while fixing these parameters during the velocity model building. For FWI to utilize all parts of the waveform, the full anisotropic 3D visco-elastic Physics of the Earth should be used as a forward modeling engine. However, at the current stage, due to the complexity nature of the physics used and the computational burden associated with inverting the multi-component parameters, trade-off between model parameters, the feasibility of applying a 3D visco-elastic FWI is far from practical application and its practical feasibility has yet to be seen.

– FWI is an ill-posed inverse problem, and requires regularization and appropriate data preconditioning to stabilize the inversion. Inverse problems are called ill-posed if their solutions are not unique or if the solutions are unstable. If it is unstable, a small perturbation of the data may produce large perturbations of the inverse problem solution. Conversely, if it has a unique and stable solution, then it is called well-posed problem. In the ill-posed inverse problem, for example in the case of FWI, the forward modeling has a unique solution, whereas the inverse problem may not be unique as there would be many models that could fit the observed data or provide very small data

residuals. In order to overcome having an ill-posed problem, supplemental information and stabilization are needed. Prior information about the model parameters, data preconditioning and regularization methods are key for successful FWI (Tarantola, 1987; Virieux and Operto, 2009).

– FWI requires an accurate source signature for the forward modeling. The source signature at the lowest frequency range should be accurate enough that the modeled data matches the observed data at the lowest usable frequencies within half a wave-cycle. The accuracies of polarity and phase of the source wavelet are essential for major events to have any success in velocity model building by FWI.

The aforementioned problems are among the most common problems in FWI. Addressing these problems, in particular, the accuracy of the starting model, the quality of the low frequencies in the data and the presence of long offset data are keys for any successful FWI. Inversion results obtained by FWI also varies from one problem to another. Depending upon the geological area of interest, each problem should be treated and dealt with independently.

1.3 Time-domain vs frequency-domain

Historically, Lailly (1983) and Tarantola (1984b, 1987) introduced the theoretical framework of FWI in the time-domain and later the method was extended to the frequency-domain by Pratt (1999), Laplace-domain by Shin and Cha (2008), and/or the combination of the two methods using a hierarchical hybrid approach (Cha and Shin, 2009; Brenders et al., 2012). Each of those methods has its own distinct advantage. From the standpoint of computationally efficiency, the frequency-domain implementation is preferable over a time-domain implementation because of its ability to reconstruct model parameters from limited discrete frequencies by taking advantage of the wavenumber redundancy in wide-aperture seismic data (Sirgue and Pratt, 2004). The frequency-domain allows us to perform efficient hierarchical multiscale strategies to mitigate the nonlinearity problem of FWI by performing the inversion from low to high frequencies. In the frequency-domain, a finite set of discrete frequencies is selected and the inversion is carried out in a sequential approach starting from low frequency to high frequency data components (Sirgue and Pratt, 2004). The inversion can be carried out sequentially with single or multiple frequencies. The model parameters obtained from the inversion of the low frequency data components are used as starting model parameters for the higher frequency data components. First, the long wavelength components of model parameters will be recovered from the low frequency data, and then more details and features will be extracted as the inversion proceeds to the higher frequency data. The multiscale inversion strategy helps to avoid the cycle-skipping problem of FWI

because the low frequencies are less sensitive to cyclic-skipping than the high frequencies (Bunks et al., 1995). The technique avoids getting trapped to local minima when the starting model is far from the actual velocity model. If all frequencies are inverted, then the frequency-domain method is equivalent to the time-domain method. In 2D FWI, from the standpoint of computationally efficiency, the frequency-domain implementation is preferable over a time-domain implementation because of its ability to reconstruct model parameters from limited discrete frequencies by taking advantage of the wavenumber redundancy in wide-aperture seismic data (Sirgue and Pratt, 2004) and the forward modeling for multiple sources can be performed efficiently using an LU decomposition of the Helmholtz wave-equation. In this thesis, we adopt the frequency-domain FWI method. However, from the practical point of view and the simplicity of implementation, for large scale 3D problems, the time-domain is the most preferable and widely used method for FWI (Vigh and Kapoor, 2009; Sirgue et al., 2011; Liu and Gu, 2012; Warner et al., 2013). Time-domain based FWI allows a more flexible way for preconditioning or processing a certain portion of data by time-windowing and is also robust with noise compared to the frequency-domain (Warner et al., 2013). The multiscale inversion strategy can also be implemented in time-domain by band-pass filtering the time-domain seismogram based on different bandwidths (Bunks et al., 1995). In the hybrid approach, the forward modeling is performed in the time-domain whereas the inversion is carried out in the frequency-domain and this technique provides a good trade-off between efficiency and flexibility (Brenders et al., 2012).

In the case of the Laplace-domain, the inversion provides a long-wavelength (smooth) velocity model without having a low frequency data component even if the starting model for inversion is significantly different from the actual velocity model. This method has been shown to be less sensitive to intermediate wavelengths contained in the velocity model but it fails to provide details of the model structure at higher frequencies (Shin and Ho Cha, 2009). FWI in the Laplace-domain is viewed as a zero frequency component of damped seismograms in the time-domain (Shin and Ho Cha, 2009). Equivalently, in frequency-domain inversion, it can also be viewed as using complex-valued frequencies in which the time-damping of the wavefield is controlled by the complex part while the real part is taken as zero. One of the advantages of the Laplace-domain waveform inversion is that it can be applied to smaller frequencies than the lowest available frequency of the source bandwidth. However, Shin and Cha (2008) have shown that the penetration depth of Laplace-domain waveform inversion is very sensitive to the choice of Laplace-domain damping constants and the maximum offset of the data. However, the smooth velocity model obtained from Laplace-domain inversion can be used as the starting velocity model for frequency-domain inversion; a technique known as Laplace-Fourier inversion (Shin and Ho Cha, 2009).

1.4 Optimization engine

FWI is an iterative process that attempts to build high resolution model parameters of Earth's subsurface structure. Most existing full waveform inversion numerical algorithm schemes are based on a local iterative method using a steepest decent or nonlinear conjugate gradient method. This method only takes into account the first-order derivatives of the objective function and neglects the second-order derivatives of the objective function (Hessian matrix). The gradient or steepest descent method often poorly scales the model perturbation update and takes many iterations to converge to the global minimum. In order to increase the convergence rate, higher-order optimization methods, for example, like the Newtonian methods such as Limited-memory Broyden-Fletcher-Goldfarb-Shanno (l-BFGS) (Nocedal, 1980), Gauss-Newton or full Newton methods that take into account the information of the second-order derivatives of the objective function should be implemented.

The Newtonian methods generally converge faster than the gradient based nonlinear conjugate gradient method, but this is at the expense of solving a denser system of linear equations (Hessian matrix) at each iteration which is computationally demanding. Of the Newtonian methods, the full Newton method is known to converge faster than the Gauss-Newton or l-BFGS methods. Incorporating the inverse of second-order derivatives of the objective function, the inverse of Hessian matrix, in the numerical inversion scheme not only speeds up the converge rate but also take into account the effect of high amplitude energy multi-reflected wavefields present in the data. In addition, the use of the inverse of the Hessian matrix in the numerical inversion better scales amplitude energies on poorly illuminated subsurface areas due to uneven source-receiver coverage, and removes unwanted artifacts that arise in the gradient due to multiple reflected wavefields (Pratt et al., 1998). Theoretically, incorporating higher-order optimization method such as second-order derivatives of the objective function (Hessian matrix) is possible and become a current research topic.

Apart from mono-parameter inversion (for example velocity estimation, the focus of this thesis), Hessian matrix is also plays a significant role in multiparameter FWI. Multiparameter full waveform inversion is a very difficult problem due to many unknown parameters appears in the same wave-equation. It also appears in the gradient of the objective function in the iterative numerical inversion. Incorporating the inverse of Hessian matrix in the numerical inversion plays a key role in multiparameter full waveform inversion. It mitigates the effects of cross-talk or trade-off between different classes of parameter. Recent work by Innanen (2014) shows the importance of incorporating the Hessian matrix in the multiparameter seismic inverse problem. The inverse of the Hessian matrix combats parameter cross-talk and carries out an essential form of amplitude-variation-with-offset (AVO)

analysis on pre-critical reflection modes (Innanen, 2014).

At the present time, the Hessian matrix can be implemented in an efficient way in 2D problems. However, the computational and memory demand for computing the Hessian matrix and its inverse for larger scale 3D problems is enormous. For this reason, in almost all the current 3D FWI algorithm, the Hessian matrix information are often ignored and the steepest descent method is the most preferable method in the oil and gas industries. In generally, the efficiency of the FWI algorithm and accuracy of the reconstructed model parameters depend on the choice of model parameterization, the type of optimization engine employed, and a proper preconditioner of the gradient of the objective function are among a few to mention (Pratt et al., 1998; Virieux and Operto, 2009; Guitton and Daz, 2012).

1.5 Simultaneous sources

Though much progress has been made in FWI during the last couple of years, one of the many existing problems associated with FWI that hinders its practical application to large scale 3D datasets is its huge computational time demand for simulating a number of forward modeling in the iterative process. The computational cost of full waveform inversion is proportional to the number of sources in the experiment. One way to efficiently perform full waveform inversion at larger scales and on 3D datasets is to perform a computationally inexpensive waveform inversion by simulating multiple sources at one time, a technique referred to as simultaneous sources full waveform inversion. The use of simultaneous sources was initially employed for seismic migration by Morton and Ober (1998) and Romero et al. (2000). Since the development of simultaneous sources by Morton and Ober (1998) and Romero et al. (2000), its use for seismic migration and full waveform inversion (Krebs et al., 2009; Tang, 2009; Berkhout and Blacquiére, 2011; Ben-Hadj-Ali et al., 2011) have been developed and extensively used to reduce the computational cost of simulating the forward modeling. The basic idea of the simultaneous sources technique is to create super-shots through the summation of individual sources. One of the main advantages of the simultaneous source technique in seismic migration or waveform inversion is that it reduces the computational cost by a factor proportional to the number sources stacked within a super-shot for every forward modeling performed in the inversion scheme.

On the other hand, simultaneous sources introduce cross-talk artifact noises in the final image due to interference among the individual sources within a super-shot. A good inversion algorithm scheme should able to balance the quality of the reconstructed model parameters and its computational speed. Morton and Ober (1998) and Romero et al. (2000) introduce phase-encoding strategies of individual shots to create super-shots to reduce cross-talk artifacts. Since then, in addition to the phase-encoding technique, different strategies for the

encoding of sources during forward modeling have been devolved for full waveform inversion such as time-shift, phase-shift or plane-wave encoding methods (Krebs et al., 2009; Berkhout and Blacquiére, 2011; Tao and Sen, 2013) are among a few to mention. The other problem with simultaneous source technique is that it can only be suited for land data acquisitions, were all the receiver spreads are fixed for every shot within a super-shot. But in the case of marine acquisition where receiver arrays move for every shot, there will be an aperture mismatch between the predicted data and the observed data and as a result minimizing the conventional least-squares objective function is no longer applicable. A different or perhaps a new objective function is required to address the aperture mismatch (Krebs et al., 2009).

1.6 Objectives and structure of the thesis

The main objective of this thesis is on the practical application of the first-order and second-order optimization methods for 2D acoustic full waveform inversion that is based on conventional shot gather and simultaneous source methods. One of the main objective in this thesis to find an efficient work flow full waveform inversion for cross-talk artifact reductions in the simultaneous source inversion techniques. The thesis studies the effects of multi-frequency simultaneous frequency inversion on cross-talk artifacts and studies the efficiency of the use of simultaneous sources for higher-order of optimization methods. The influence of model parameterization on the Newton-based acoustic full waveform inversion is also studied. My thesis is primarily based on 2D acoustic wave approximation of the wave propagation.

The thesis is structured as follows:

Chapter 2 reviews the state of the art in full waveform inversion and presents the theoretical framework of 2D acoustic full waveform inversion in the frequency-domain. First I describe the mathematical formulations and implementations of the first-order and the second-order optimization methods of full waveform inversion. The mathematical formulations present both for conventional shot gather and simultaneous source techniques. Different source encoding techniques are considered. The gradient-based and Newton-based optimization method is briefly described, and in this chapter I describe the key ingredients of 2D full waveform inversion algorithms that are used through the rest of the chapters in the thesis. Preliminar synthetic examples are also presented to assess the influences of density and attenuation to the velocity model building in full waveform inversion.

Chapter 3 presents the studies of model parameterization strategies on velocity model building using full waveform inversion based on the acoustic wave-equation approximation. Particular emphasis is given to Newton-based optimization methods, where it requires the reconstruction of the Hessian matrix directly or indirectly with which the choice of model

parameterization greatly influences the characteristics of the Hessian matrix. The study presents an eigenvalue decomposition to understand and analyze the influences of model parameterization on the Hessian matrix and convergence rate of the optimization engine. Three types of model parameterizations are considered. These include velocity, slowness and slowness squared. Numerous numerical inversions are carried out using gradient-based and Newton-based methods to assess convergence rates and the quality of the retrieved model that can be obtained using the three different model parameterizations.

Chapter 4 details the work I have done in developing full waveform inversion for reduction of cross-talk artifacts when simultaneous sources are used and it presents a detailed comparison studies of multi-frequency selection strategies for simultaneous-source full waveform inversion using the Gauss-Newton method. Various mutli-frequency selection strategies are studied to evaluate their performance in reducing cross-talk artifacts and to define an inversion strategy that provides a good balance between quality of the reconstructed model and computational time. A series of numerical experiments are carried out to draw conclusions.

Chapter 5 presents the application of the second-order optimization method especially the full Newton method for full waveform inversion and the type of iterative method that should be used for solving the Newton equations. The objective of this chapter is to find the most suitable and efficient iterative solver for the Newton equations at relatively cheaper computational time. The preconditioned conjugate gradient and minimum residual iterative methods are considered and compared for solving the Newton equations.

Chapter 6 presents the conclusions of the main objectives of the thesis findings, identifies the contributions and summarizes the dissertation research with recommendations of future work.

CHAPTER 2

Full Waveform Inversion

2.1 Introduction

This chapter describes the theoretical formulation of full waveform inversion in the frequency-domain. The formulation is based on the two-dimensional acoustic wave approximation. The most common choice of objective function for FWI is the least-squares norm, l_2, of misfit between the observed and predicted data. This method has been shown to be less sensitive and provides acceptable results for high signal-to-noise ratio (SNR) datasets. I describe a detailed formulation of the first- and second-order derivatives of the objective function based on an l_2 norm.

2.2 Theory

We begin by considering the forward problem in the frequency-domain after Fourier transforming the time-domain acoustic wave-equation. The acoustic wave-equation in two dimensions is governed by the following equation (Hu et al., 2009)

$$\frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial}{\partial x}p(\omega)\right)+\frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial}{\partial z}p(\omega)\right)+\frac{\omega^2}{\kappa}p(\omega)=src(\omega), \tag{2.1}$$

where $\kappa=\rho v^2$ is the bulk modulus, ρ is the density, v is the velocity, $p(\omega)$ is the pressure field, ω is the angular frequency and $src(\omega)$ is the source signature value at a given frequency ω. For the sake of simplicity, the dependencies of spatial positions of the pressure field are not written explicitly. In order to avoid boundary reflections during forward modeling, we apply a perfectly matched layer absorbing boundary condition (Berenger, 1994).

After discretization of equation 2.1 by using an optimal 9-point stencil staggered-grid finite-difference approximation (Jo et al., 1996; Hustedt et al., 2004), the acoustic wave-equation for a single source and frequency can be written in a compact form as

$$A(\mathbf{m},\omega)\mathbf{p}(\mathbf{x},\mathbf{x}_s,\omega) = \mathbf{f}(\omega)\delta(\mathbf{x}-\mathbf{x}_s), \tag{2.2}$$

where $A(\mathbf{m},\omega)$ is the discretized stiffness operator, $\mathbf{m}$ is the model parameter (for example velocity), $\mathbf{f}(\omega)$ is the matrix form of complex source function ($src(\omega)$) and $\mathbf{p}(\mathbf{x},\mathbf{x}_s,\omega)$ is the complex pressure field everywhere inside the physical medium $\mathbf{x}$ for a shot located at position $\mathbf{x}_s$. Here, $\mathbf{x}_s$ is a vector where its elements are zeros everywhere except 1 at the source location. Details of the discretization of equation 2.1 by using an optimal 9-point stencil staggered-grid finite-difference approximation is described in Appendix-A. This operator is quite sparse and, therefore, storable in memory. The system of linear equations, equation 2.2, is solved with a direct solver based on a multi-frontal LU decomposition of the finite-difference Helmholtz operator $\mathbf{A}$ into a lower and upper triangular decomposition scheme (Amestoy et al., 2001; Schenk and Gartner, 2004; Amestoy et al., 2006; Schenk et al., 2007). The main advantage of this method is that once the decomposition is performed and available for a given angular frequency ω and background velocity, the complex pressure field is efficiently solved for multiple sources using the forward and backward substitutions. A more efficient factorization scheme like Cholesky factorization can also be used to decompose the matrix $\mathbf{A}$ (Chabert, 1999). However, the Cholesky factorization requires a Hermitian, positive definite matrix. In our case, the matrix $\mathbf{A}$ is not Hermitian inside the perfectly matched layer region (see Appendix-A for more detail). During the modeling of the wavefield, we avoid boundary reflections by applying the perfectly matched layer absorbing boundary condition (Berenger, 1994) for absorbing outgoing waves at the boundaries. However, this still does not completely absorb all outgoing waves incident at all angles.

The forward problem for a single frequency is obtained by solving the system of linear equations in equation 2.2 and then calculating the observation via

$$\mathbf{d}^{cal}(\mathbf{x}_r,\mathbf{x}_s,\omega) = s(\omega)\mathbf{r}(\mathbf{x}_r,\mathbf{x})\mathbf{p}(\mathbf{x},\mathbf{x}_s,\omega), \tag{2.3}$$

where $\mathbf{r}(\mathbf{x}_r,\mathbf{x})$ is the sampling operator that maps the wavefield to the observation space and $s(\omega)$ is a scaling parameter for the source wavelet signature which is estimated using a least-squares method, see equation 2.8. From here onwards, the subscripts s and r represent for the source and receiver, respectively. When the spatial dependencies are not explicitly written in the equation, for example, we represent the pressure wavefield for a given source $\mathbf{p}(\mathbf{x},\mathbf{x}_s,\omega)$ by $\mathbf{p}_s(\omega)$. Bold letters in equations represent either a vector or matrix.

Waveform inversion entails minimizing the l_2 norm of residual between the observed data

$\mathbf{d}^{obs}$ and the model data $\mathbf{d}^{cal}$

$$J(\mathbf{m}) = \frac{1}{2} \sum_{\omega_j \in \Omega_i}^{N_w} W_j \sum_{s,r}^{N_s,N_r} (\mathbf{d}^{obs}_{s,r}(\omega_j) - \mathbf{d}^{cal}_{s,r}(\omega_j))^{\dagger} (\mathbf{d}^{obs}_{s,r}(\omega_j) - \mathbf{d}^{cal}_{s,r}(\omega_j)) + \mu R(\mathbf{m}), \qquad (2.4)$$

where $\dagger$ is the complex conjugate transpose, $R(\mathbf{m})$ is the regularization term (assumed the same for each frequency group), μ is the regularization parameter, N_s and N_r represent the number of sources and receivers respectively, and W_j is the data weight for the simultaneous inversion of multiple frequencies (frequencies in a group). The parameter N_w represents the number of frequencies within a frequency group. The variable Ω_i represents a group of frequencies. We stress that the inversion of the first group (Ω_1) is initialized with a smooth model; the inversion with data from group Ω_{i+1} is initialized with the solution found by inverting group Ω_i. In this thesis, unless otherwise stated, for the regularization term, we used damped regularization term

$$R(\mathbf{m}) = ||\nabla^2(\mathbf{m}_k - \mathbf{m}_o)||^2, \qquad (2.5)$$

where $\mathbf{m}_o$ is the starting velocity model at every frequency group and $\mathbf{m}_k$ is the velocity model at the k^{th} iteration.

2.2.1 Lagrangian formulation

The minimization of the objective function is a nonlinear problem. The objective function depends explicitly on the complex pressure field and implicitly on the model parameters through the pressure field, which depends on the properties of the medium. The goal is to reconstruct the model parameters. In this thesis our model parameters are subsurface compressional velocities.

There two types of optimization methods for nonlinear inverse problems: local and global methods. Global optimization techniques aim to search the model space in order to find the best solution or global minima. Such techniques allow us to solve strongly nonlinear problems but are typically slower as the search continues to avoid becoming trapped at local minima and use a stochastic search in order to find a global minimum. They often work best with a small number of parameters. Such methods include Monte Carlo, simulated annealing and Genetic algorithms. Whereas the local optimization methods are based on local information about the gradient of the objective function that defines, for example in the case of l_2 norm, the data mismatch between the model and observed data. Local optimization methods in general require a good search direction to locally optimize in an iterative sequence. Hence, they require a good starting model to converge to the local minimum. For geophysical exploration applications, FWI is a local optimization method

and depends strongly on the choices of starting model. This thesis exclusively focuses on local minimization based on gradient and Newton-based methods.

In order to formulate the gradient-based and Newton-based optimization methods, we adopt the second-order adjoint state method based on the Lagrangian formulation of the objective function (Akcelik et al., 2002; Plessix, 2006; Fichtner and Trampert, 2011). The objective function is minimized using a constrained optimization method

$$\begin{aligned} \underset{\mathbf{m}}{\arg\min} \quad & J(\mathbf{m}) \\ \text{subject to} \quad & \mathbf{A}(\mathbf{m},\omega)\mathbf{p}_s(\omega) = \mathbf{f}_s(\omega), \end{aligned} \tag{2.6}$$

where $\mathbf{A}(\mathbf{m},\omega) \in \mathbb{C}^{N\times N}$, $\mathbf{p}_s(\omega)$ and $\mathbf{f}_s(\omega) \in \mathbb{C}^N$, $\mathbf{m} \in \mathbb{R}^N$, and $J(\mathbf{m}) : \mathbb{R}^N \to \mathbb{R}$. The parameter N represents the number of model parameters. Note that for simplicity the spatial dependencies are not written explicitly and we omit the kronical delta function. The constrained optimization problem can be reformulated as an equivalent problem where we minimize the associated Lagrangian $\mathcal{L}$

$$\begin{aligned} \mathcal{L}(\tilde{\mathbf{p}},\tilde{\mathbf{m}},\tilde{\mathbf{\Lambda}}) = \frac{1}{2}\sum_{\omega_j\in\Omega_i}^{N_w} W_i \sum_{s,r}^{N_s,N_r} \parallel \mathbf{d}^{obs}_{s,r}(\omega_j) - s(\omega_j)\mathbf{r}\tilde{\mathbf{p}}_s(\omega_j) \parallel^2 + \\ \Re e\big(\sum_{\omega_j,s}^{N_w,N_s} \langle\tilde{\mathbf{\Lambda}}_s(\omega_j), \mathbf{A}(\tilde{\mathbf{m}},\omega_j)\tilde{\mathbf{p}}_s(\omega_j) - \mathbf{f}_s(\omega_j)\rangle_{\mathbf{x}}\big) + \mu R(\mathbf{m}), \end{aligned} \tag{2.7}$$

where $\langle,\rangle_{\mathbf{x}}$ is the scalar product in $\mathbf{x}$, $\mathbf{r} \in \mathbb{R}^{N_r\times N}$, $\mathbf{\Lambda}_s(\omega_j) \in \mathbb{C}^N$ is a vector of Lagrangian multiplier, which its physical meaning become clear later, $\mathbf{d}^{obs}_s(\omega_j) \in \mathbb{C}^{N_r}$ and $R(\mathbf{m}) : \mathbb{R}^N \to \mathbb{R}$ and $s(\omega_j)$ is the scaling parameter for source wavelet signature and in the frequency-domain FWI for single frequency (ω_j) (Pratt et al., 1998)

$$s(\omega_j) = \frac{\mathbf{d}^{obs}(\omega_j)^\dagger \mathbf{d}^{cal}(\omega_j)}{\mathbf{d}^{cal}(\omega_j)^\dagger \mathbf{d}^{cal}(\omega_j)}. \tag{2.8}$$

The source needs to be estimated because in exploration seismology it is not possible to measure the seismic source wavelet (Ulrych and Sacchi, 2005).

The first-order necessary conditions, also known as the Karush-Kuhn-Tucker (KKT) conditions (Kuhn and Tucker, 1951), for the optimum solution satisfies

$$\nabla_\xi \mathcal{L}(\tilde{\mathbf{p}},\tilde{\mathbf{m}},\tilde{\mathbf{\Lambda}}) = \mathbf{0}, \qquad \xi \in \{\tilde{\mathbf{p}},\tilde{\mathbf{m}},\tilde{\mathbf{\Lambda}}\}. \tag{2.9}$$

This results in

$$\mathcal{L}_{\mathbf{p}} = \mathbf{A}^*(\mathbf{m},\omega)\mathbf{\Lambda}_s(\omega) - \mathbf{r}^\dagger \triangle \mathbf{d}(\omega) = \mathbf{0} \tag{2.10}$$

$$\mathcal{L}_{\mathbf{m}} = \nabla_{\mathbf{m}}\mathbf{A}(\mathbf{m},\omega)\mathbf{p}_s(\omega) \odot \mathbf{\Lambda}_s^*(\omega) + \mu\nabla_{\mathbf{m}}R(\mathbf{m}) = \mathbf{0} \tag{2.11}$$

$$\mathcal{L}_{\mathbf{\Lambda}} = \mathbf{A}(\mathbf{m},\omega)\mathbf{p}_s(\omega) - \mathbf{f}_s(\omega) = \mathbf{0}, \tag{2.12}$$

where " $*$ " indicates the complex conjugate operator, $\odot$ represents the Hadamard product (see in the Appendix equation A.34 for the definition of the Hadamard product) and

$$\mathcal{L}_\xi = \frac{\partial \mathcal{L}}{\partial \xi}, \qquad \xi \in \{\mathbf{p}, \mathbf{m}, \mathbf{\Lambda}\} \tag{2.13}$$

$$\nabla_{\mathbf{m}}\mathbf{A}(\mathbf{m},\omega) = \frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}} \tag{2.14}$$

$$\triangle \mathbf{d}(\omega) = \mathbf{d}^{obs}(\omega) - \mathbf{d}^{cal}(\omega). \tag{2.15}$$

Equation 2.10 is the adjoint state equation that computes the adjoint wavefield $\mathbf{\Lambda}_s(\omega)$ by back-propagating the residual wavefields at the receiver locations (Plessix, 2006). Equation 2.11 is the gradient of the cost function that one wants to minimize $\mathbf{g}$ ($\mathbf{g} = \mathcal{L}_{\mathbf{m}}$). Equation 2.12 is the forward modeling operator.

From the first-order optimality conditions, setting the derivative of equation 2.7 with respect to the model parameters $\mathbf{m}$ to zero gives an expression for the gradient, $\mathbf{g}$,

$$\mathbf{g} = \Re e((\mathbf{A}_{\mathbf{m}}(\mathbf{m},\omega)\mathbf{p}_s(\omega))\mathbf{\Lambda}_s^\dagger(\omega)) + \mu\nabla_{\mathbf{m}}R(\mathbf{m}). \tag{2.16}$$

The optimality conditions of the KKT system in equation 2.10-2.12 define a system of nonlinear equations. If the problem is solved via Newton's method, the parameter perturbation is obtained by solving the following second-order KKT system of linear equations (linearizing equation 2.9)

$$\begin{pmatrix} \mathbf{r}^\dagger\mathbf{r} & \mathbf{B} & \mathbf{A}^\dagger \\ \mathbf{B}^\dagger & \mathbf{K} + \mu\nabla^2_{\mathbf{m}}R & \mathbf{C}^\dagger \\ \mathbf{A} & \mathbf{C} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \triangle\mathbf{p} \\ \triangle\mathbf{m} \\ \triangle\mathbf{\Lambda} \end{pmatrix} = - \begin{pmatrix} \mathcal{L}_{\mathbf{p}} \\ \mathcal{L}_{\mathbf{m}} \\ \mathcal{L}_{\mathbf{\Lambda}} \end{pmatrix}, \tag{2.17}$$

where

$$\mathbf{K} = (\nabla^{\mathbf{2}}_{\mathbf{m}}\mathbf{A})\mathbf{p}_s(\omega)\mathbf{\Lambda}^{\dagger}_s(\omega) \tag{2.18}$$

$$\mathbf{B} = (\nabla_{\mathbf{m}}\mathbf{A}^{\dagger})\mathbf{\Lambda}_s(\omega) \tag{2.19}$$

$$\mathbf{C} = (\nabla_{\mathbf{m}}\mathbf{A})\mathbf{p}_s(\omega). \tag{2.20}$$

The solution to the Newton's step is obtained by solving for $\triangle\mathbf{m}$, after eliminating the perturbation of the state and the adjoint wavefields ($\triangle\mathbf{p}$ and $\triangle\mathbf{\Lambda}$), respectively (Askan et al., 2007; Anagaw and Sacchi, 2012a; Métivier et al., 2013). The resulting equation leads to the following linear system of equation

$$\Re e[\mathbf{H}]\triangle\mathbf{m} = -\Re e[\mathbf{g}], \tag{2.21}$$

where $\Re e$ represents for real part and $\mathbf{H}$ is the full Hessian matrix given by

$$\begin{aligned}\mathbf{H} =&((\nabla_{\mathbf{m}}\mathbf{A})\mathbf{p_s})^{\dagger}(\mathbf{A}^{\dagger})^{-1}\mathbf{r}^{\dagger}\mathbf{r}\mathbf{A}^{-1}(\nabla_{\mathbf{m}}\mathbf{A})\mathbf{p_s} + \mu\nabla^2_{\mathbf{m}}R \\ &+ \mathbf{G} - \mathbf{B}^{\dagger}\mathbf{A}^{-1}\mathbf{C} - \mathbf{C}^{\dagger}(\mathbf{A}^{\dagger})^{-1}\mathbf{B},\end{aligned} \tag{2.22}$$

Further details on the formulation and explanation of the KKT conditions can be found in any optimization books (Nocedal and Wright, 2006; Kuhn and Tucker, 1951; Akcelik et al., 2002). The first term on the right hand side of equation 2.22 can be expressed in terms of the Jacobian matrix, $\mathbf{J}$, of the wavefields; where $\mathbf{J} = \mathbf{A}^{-1}(\nabla_{\mathbf{m}}\mathbf{A})\mathbf{p_s}$. The Jacobian matrix is the first-order derivative of the wavefield with respect to the model parameters. The last three terms in equation (2.22) relate to the second-order derivatives of the wavefields, and are equivalent to $\nabla^{\mathbf{2}}_{\mathbf{m}}\mathbf{p_s}\mathbf{r}^{\dagger}\triangle\mathbf{d}(\omega)$. The physical interpretations of these terms are given by Pratt et. al. Stekl and Pratt (1998) and are re-stated in Chapter-3 and 5. Incorporating the inverse of the Hessian matrix in the FWI optimization scheme not only improves the convergence rate of FWI by improving the amplitude energy on poorly illuminated subsurface areas but also acts as a debluring operator (Stekl and Pratt, 1998). The inverse of the Hessian matrix removes artifacts that arise in the gradient vector due to multiple reflected wavefields. The second-order derivatives of the wavefield, the last three terms of equation 2.22, account for the double reflected wavefields from two different points in the subsurface. In the presence of high velocity contrast, the effect of the second-order derivatives of the wavefield is significant and it should not be ignored (Anagaw and Sacchi, 2012a).

In the Gauss-Newton approximation, the data residual between the modeled and observed data are assumed to be very small and the last three terms in equation 2.22 are often ignored. Consequently, the Hessian can be approximated by

$$\mathbf{H}_{GN} = ((\nabla_{\mathbf{m}}\mathbf{A})\mathbf{p_s})^{\dagger}(\mathbf{A}^{\dagger})^{-1}\mathbf{r}^{\dagger}\mathbf{r}\mathbf{A}^{-1}(\nabla_{\mathbf{m}}\mathbf{A})\mathbf{p_s} + \mu\nabla^2_{\mathbf{m}}, \tag{2.23}$$

where $\mathbf{H}_{GN}$ represents the approximate Gauss-Newton Hessian matrix. In the next section we will describe the optimization scheme of the objective function based on gradient-based and Newton-based methods.

2.2.2 Numerical optimization

The gradient-based minimization scheme, the first-order derivatives of the objective function, such as the steepest descent and conjugate gradient (Nocedal, 1980) methods ignores the Hessian matrix by assuming the Hessian matrix to be an identity matrix $\mathbf{I}$ ($\mathbf{H} = \mathbf{I}$). These methods require the solutions of two forward problems per shot to compute the gradient; one for the forward modeling $\mathbf{p}_s(\omega)$ and the other for the adjoint state $\mathbf{\Lambda}_s(\omega)$. The gradient is then computed by the product between the forward propagated wavefields and back-propagated residual wavefields in space for each frequency. We leave the reader to see the details of the steepest descent and conjugate gradient algorithm in Nocedal (1980) & Nocedal and Wright (2006). For gradient-based optimization, in this thesis, we used a nonlinear conjugate gradient (NCG) with a line search that satisfies the Wolfe conditions (Hestenes, 1969; Nocedal and Wright, 2006). Algorithm 1 is the pseudo-code of a nonlinear conjugate gradient method based on the Fletcher-Reeves method (Guanghui et al., 1995). This method only requires the gradient of the objective function.

Algorithm 1: Pseudo-code: nonlinear conjugate gradient algorithm (NLCG). Notations: W_{iw} :- data weight, nw :- number of frequencies within a group frequency nwg, $\mathbf{m}$ is the model parameter.

Input : $\mathbf{d}^{obs}(\omega)$, $\mathbf{m}_0$, N_w, max_iter
Output : $\arg\min_{\mathbf{m}} J(\mathbf{m})$
for $w_i \leftarrow 1$ **to** N_w **do**
  Starting model $\leftarrow \mathbf{m}_o$;
  Compute:
  forward wavefields $\mathbf{d}^{cal}(\omega)$;
  `residual wavefields` $\triangle\mathbf{d}(\omega)$;
  `back-propagate residual wavefields` $\Lambda(\omega)$;
  gradient $\nabla_{\mathbf{m}} J(\mathbf{m}_o) = \sum_{iw}^{nw} W_{iw} \nabla_{\mathbf{m}} J(\mathbf{m}_o)_{iw}$;
  Initialize parameters;
  $\mathbf{q}_o = \mathbf{r}_o = -\nabla_{\mathbf{m}} J(\mathbf{m}_o)$;
  for $k \leftarrow 1$ **to** max_iter **do**
    Update model:
    find $\alpha(k)$ *that minimizes* $J(\mathbf{m}_k + \alpha(k)\mathbf{q}_k)$
    $\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha(k)\mathbf{q}_k$
    $\mathbf{r}_{k+1} = -\nabla_{\mathbf{m}} J(\mathbf{m}_{k+1})$
    $\beta_{k+1} = \dfrac{\nabla_{\mathbf{m}} J(\mathbf{m}_{k+1})^T \nabla_{\mathbf{m}} J(\mathbf{m}_{k+1})}{\nabla_{\mathbf{m}} J(\mathbf{m}_k)^T \nabla_{\mathbf{m}} J(\mathbf{m}_k)}$ →Fletcher-Reeves;
    $\mathbf{q}_{k+1} = \mathbf{r}_{k+1} + \beta_{k+1}\mathbf{q}_k$;
  end
end

Gradient-based iterative methods have been shown to converge slowly (Shin et al., 2001). On the other hand, the full Newton and the Gauss-Newton methods are well known for their fast convergence properties for well-posed problems (Askan et al., 2007; Anagaw and Sacchi, 2012a; Métivier et al., 2013). However, both methods are computationally expensive for large scale inverse problems. Solving the model parameters with the Newton-based algorithms require the gradient, the Hessian and the inverse of the Hessian matrix, which is dense and large. In practice, one does not need to compute the inverse of the Hessian matrix by direct estimation of the model perturbation update in equation 2.21 but often solved via Krylov space iterative methods (Kuhn and Tucker, 1951; Akcelik et al., 2002). In this thesis, for Chapter 3 and 4 numerical results, we adopt the conjugate gradient method (CG) to estimate the model perturbation or to solve the Newton's direction. This technique is less reliable than sovling using the exact Newton method, but it is efficient and handles large problems. Algorithm 2 is the pseudo-code of a conjugate gradient method used to solve for the Newton's step. Once the Newton's step is computed, we then update the solution via a parabolic line search method

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha \triangle \mathbf{m}_k, \tag{2.24}$$

where the scalar parameter α is an optimal step length. The resulting Newton direction is not exact because the Newton step is estimated using an iterative technique. The Newton step does not always guarantee to point in the decrease of the objective function. In FWI, sometimes the quadratic approximation of the objective function is not accurate enough to point in the descent direction. In order to prevent the objective function going in the ascent direction, a line search algorithm is necessary.

In Chapter 6, we used preconditioned minimum residual method to solve for Newton's direction (Fischer et al., 1998).

Algorithm 2: Pseudo-code: conjugate gradient algorithm (CG).

Input : $\mathbf{A}$, $\mathbf{x}_0$, $\mathbf{b}$,max_iter,η, ε
Output : $\mathbf{x}$
Initialize parameters;
$\mathbf{r}_0 = \mathbf{b} - A\mathbf{x}_0$;
$\mathbf{p}_1 = \mathbf{r}_0$;
for $k = 1$ **to** max_iter **do**
  $\alpha_k = \frac{\mathbf{r}_{k-1}^T \mathbf{r}_{k-1}}{\mathbf{p}_k^T \mathbf{A} \mathbf{p}_k}$;
  $\mathbf{x}_k = \mathbf{x}_{k-1} + \alpha_k \mathbf{p}_k$;
  $\mathbf{r}_k = \mathbf{r}_{k-1} - \alpha_k \mathbf{A} \mathbf{p}_k$;
  $\beta_k = \frac{\mathbf{r}_k^T \mathbf{r}_k}{\mathbf{r}_{k-1}^T \mathbf{r}_{k-1}}$;
  $\mathbf{p}_{k+1} = \mathbf{r}_k + \beta_k \mathbf{p}_k$;
  if $||\mathbf{A}\mathbf{x}_k + \mathbf{b}|| < \eta ||\mathbf{b}||$ *then return* $\mathbf{x}_{k-1}$;
  or if $||\mathbf{x}^T{}_k \mathbf{A} \mathbf{x}_k|| < \varepsilon ||\mathbf{x}_k||^2$ *then return* $\mathbf{x}_{k-1}$;
end

2.3 Simultaneous sources and encoding functions

For large scale problems, the full waveform inversion is generally computationally intense. The computational cost of full waveform inversion that results from multi-source simulations can be reduced by randomly assembling sources into super-shots using encoding functions (Romero et al., 2000; Berkhout and Blacquiére, 2011). In the frequency-domain encoded sources scheme, the encoded source for a single super-shot is implemented via a simple linear relationship

$$\mathbf{S}(\omega)_k = \Gamma_k \odot \mathbf{D}_{\mathbf{R}k} src(\omega) = \mathbf{\Psi}_k src(\omega), \tag{2.25}$$

where $\odot$ represents the Hadamard product, $\mathbf{S}(\omega)_k$ is a collection of encoded sources (k is the index number of super-shots), $src(\omega)$ is a single source value at a given frequency ω, and

$\mathbf{D}_R \in \mathbb{R}^{N\times 1}$ is a source randomization operator whose elements contain a one where a source location is picked and zero otherwise. The number of nonzeros in the source randomization operator is equivalent to the total number individual sources that are randomly assembled within a single super-shot. The function $\Gamma_k \in \mathbb{R}^{N\times 1}$ is an encoding function with nonzero elements at the selected source locations. Note that, $\mathbf{\Psi}_k = \Gamma_k \odot \mathbf{D}_{\mathbf{R}k}$.

There are different kinds of source encoding techniques which encodes the sources and summed them together to form new super-shots (Ben-Hadj-Ali et al., 2011). Among the most common source encoding function are the time-shift (Berkhout and Blacquiére, 2011), phase-shift (Krebs et al., 2009), random phase-shift (Romero et al., 2000; Ben-Hadj-Ali et al., 2011) and plane-wave (Tao and Sen, 2013) encoding techniques.

In this thesis, we only employ the following three types sources encoding techniques:

- *Random phase-encoded simultaneous sources:* In the phase-encoding scheme, the phase-encoding function, Γ, is expressed in a matrix representation as

$$\Gamma(\phi)_{nonzeros} = [e^{-i\phi_1} \quad e^{-i\phi_2} \quad e^{-i\phi_3} \quad ... \quad e^{-i\phi_{ns}}]^T, \tag{2.26}$$

  where each ϕ_i is the phase that is randomly chosen in the interval [0, 2π] and ns is the number of shots in a single super-shot.

- *Phase-shift encoded simultaneous sources:* In a simple phase-shift encoding scheme, Γ is expressed as

$$\Gamma(\phi)_{nonzeros} = [\phi_1 \quad \phi_2 \quad \phi_3 \quad ... \quad \phi_{ns}]^T, \tag{2.27}$$

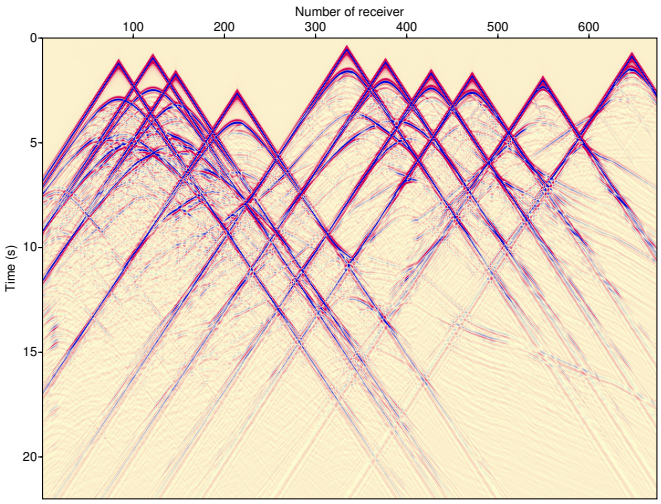
  where each ϕ_i is the phase that is randomly chosen {-1, 1}. This is equivalent to equation 2.26 in which ϕ_i are randomly chosen either 0 or π

- *Time-shift encoded simultaneous sources:* In the time-shift ending scheme, sources are randomly encoded with random time-delays (we denote them "blended sources"). In the frequency-domain blended sources scheme, the blended source for a single super-shot is implemented via a simple linear relationship

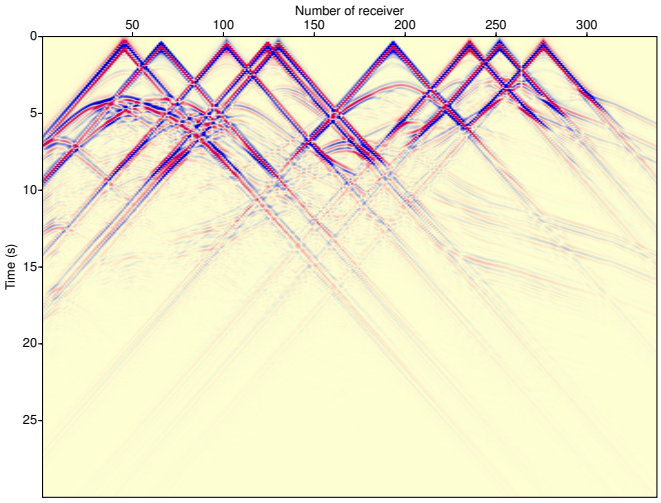
$$\Gamma(\omega,\tau)_{nonzeros} = [e^{-i\omega\tau_1} \quad e^{-i\omega\tau_2} \quad e^{-i\omega\tau_3} \quad ... \quad e^{-i\omega\tau_{ns}}]^T, \tag{2.28}$$

  where τ_i is a random time-delay randomly chosen in the interval $[0, t]$, and t is the maximum time-delay.

To understand how the source encoding functions change individual shots in a super-shot, see Figure 2.1a & 2.1b. These two super-shots are based on time-shift and phase-shift. In



(a) time-shift



(b) phase-shift

Figure 2.1: Simultaneous source also know as super-shot when individual sources are encoded with random time-shift (a) and phase-shift (b) techniques.

the case of time-shift, the source summation is based on time-delay whereas in the case of phase-shift, it is only the phases of sources that are changing during summation.

Forward problem for simultaneous sources:

Making use of the encoding function described before, the forward problem becomes

$$\mathbf{p}_k^{cal}(\omega) = \mathbf{A}(\mathbf{m},\omega)^{-1}\mathbf{\Psi} src(\omega). \tag{2.29}$$

Similarly, the objective function is modified as follows

$$J(\mathbf{m}) = \frac{1}{2}\sum_{\omega_j\in\Omega_i}^{N_w} W_j \sum_{k,r}^{N_{ss},N_r} ||\mathbf{\Psi d}(\omega_j)^{obs} - s(\omega_j)\mathbf{rp}(\omega_j)_{k,r}^{cal})||^2 + \mu R(\mathbf{m}). \tag{2.30}$$

The parameter N_{ss} represents the total number of super-shots. The numbers of super-shots for the numerical inversion are chosen manually depending up how many individual sources we want to have within a single super-shot. The gradient, the adjoint state equations and the Hessian formulation will be modified in a similar manner (see the previous sections for the optimization). The same optimization procedure as the conventional shot gather is applied.

Figure 2.2 is a schematic illustration of data reduction using simultaneous source techniques. On the right side, each line represents a single shot gather. The schematic diagram depict shot number as a function of receiver numbers. The solid red circles represent source positions that are randomly selected to form a single super-shot on the left hand side of the diagram. Once the source positions are chosen, we then use different encoding functions, Γ, to encode the selected shot gather data to form a super-shot. As shown in the Figure 2.2, for example, if the operator $\mathbf{D}_R = 1$ is selected for each conventional shot gather, a single super-shot can easily be formed at every iteration during the inversion process. For each super-shot, the operator $\mathbf{D}_R$ randomly selects source locations that will make a super-shot. Figure 2.3 illustrates how the source positions are selected, $\mathbf{D}_R$, randomly at a given iteration during the numerical inversion scheme.

2.4 The choice of starting model

FWI, in general, is formulated as an iterative local minimization of the misfit between recorded and modeled datasets. It only locates the local global minimum misfit if the starting model is close to the true model. One key assumption and criteria in the localized inversion is that the modeled and observed waveforms are within half a wave-cycle at the

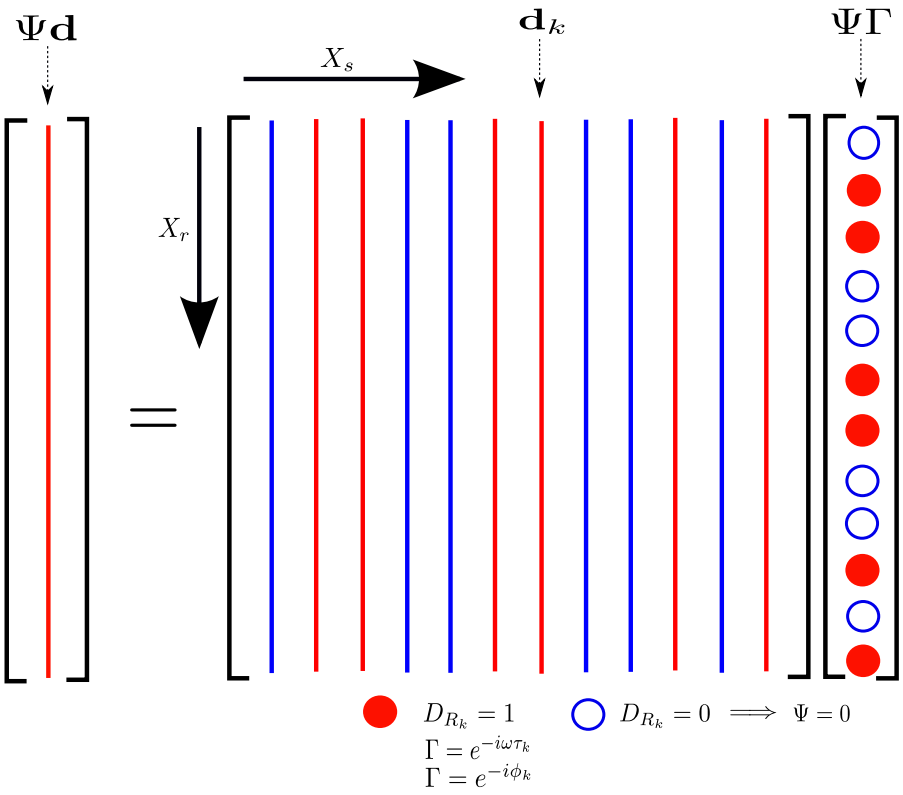


Figure 2.2: A schematic representation of data reduction using Simultaneous sources. The lines represent the shot-gather. The red lines represent the gathers that are randomly picked and summed together to form a super-shot gather on the left side.

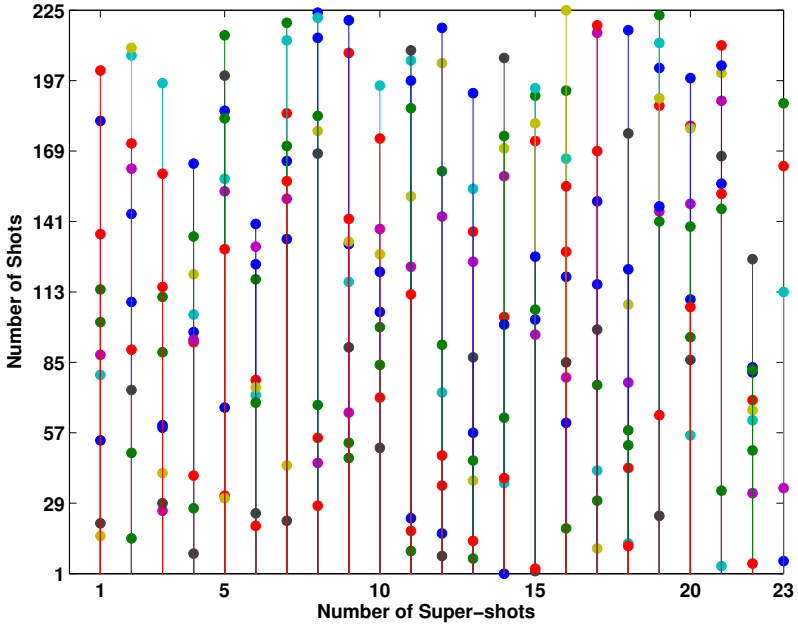


Figure 2.3: Source randomization: source location selections ($\mathbf{D}_R$) during source assembling to make super-shots. A total of 23 super-shots created from 225 individual sources.

lowest frequency to converge iteratively in the right direction. Otherwise the reduction in the misfit function can steer in the wrong direction or away from the observed data. Figure 2.4 is a schematic diagram that illustrates how the choices of starting model affect the final inversion results. For example if our starting velocity model is at position B or C, then the inversion result will end up with the strong or weak local minimum as shown in Figure 2.4. On the other hand if the starting velocity model is at A, then the inversion will end up at the local global minima. In other word, when the starting model data is cycle-skipped at the lowest usable frequency, conventional full waveform inversion converges to a local minimum. Next, we will see how cycle-skipping occurs with a toy example.

Before using FWI, we should first examine the accuracy of the starting model used for inversion by comparing the level of fitting between forward modeled of shots and real data. The starting velocity model used for FWI needs to be as accurate as possible to prevent cycle-skipping. If the starting model is not good enough, then different initial models should be fine-tuned or examined until a certain desired criterion is met; i.e. checking the traveltime or phase matches. It is not always possible to fine-tune the starting model in FWI, but the key is to test the degree of the modeling and observed data matching in the dominate wavelengths before running FWI. Figure 2.5 illustrates the cycle-skipping problem of FWI. If the initial model is not close enough, then the initial modeled data at the first iteration

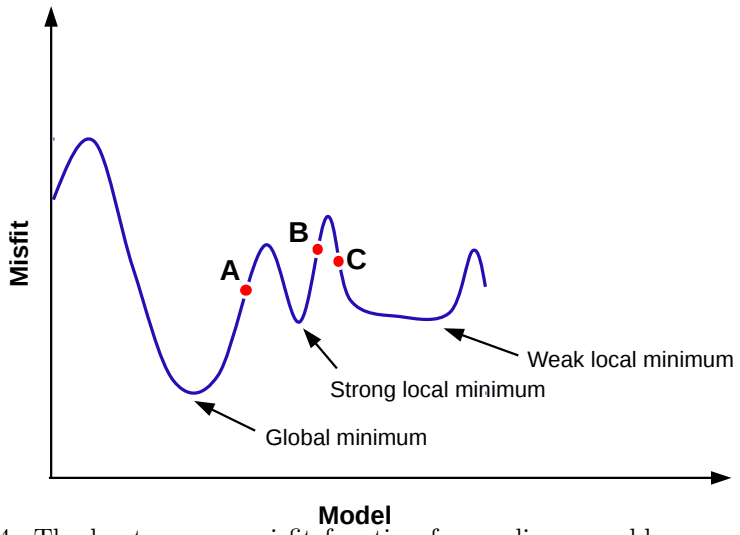


Figure 2.4: The least-squares misfit function for nonlinear problem as a function of model parameter. If the inversion starts with starting model at B or C then, the inversion will steer to the strong or weak local minimum. If the starting model starts at A, the inversion will end up with global mimimum.

will be apart by more than half a wave-cycle from the observed data. Thus, the inversion will be trapped in one of the many local minima (move in the wrong direction) and will be cyclic-skipped (see the top schematic representation of Figure 2.5). If the initial starting model is accurate enough, then initial modeled data will be within half a wave-cycle from the observed data. Consequently, the inversion will steer in the right direction and will end up at the local global minima (see the bottom schematic representation of Figure 2.5 and Figure 2.4).

The most widely used conventional l_2 norm for FWI takes both amplitude and phase information into account in the inversion. As we stressed in the previous paragraphs, for successful convergence of FWI, the modeled data must match the observed data at the lowest usable frequencies within half a wave-cycle. For this reason, at the earliest stage of the FWI scheme, the long wavelength or low-wavenumber components of the model must be recovered first before proceeding to higher frequencies; theoretically at low frequency, there is a high chance that the modeled and observed data match within half a wave-cycle. Therefore, at the earliest stage of the full waveform inversion, phase matching is more important than amplitude. It is, therefore, important to give more emphasize on phase matching at early iterations and followed by both phase and amplitude matching. An objective function that can accommodate this would be preferable (Shin and Ha, 2008; Alkhalifah and Choi, 2014). The presence of very good signals at low frequencies in the data substantial helps the cycle-skipping problem. Inversion of the low frequencies retrieves the long wavelength struc-

tures of the model, and as a result the largest changes to the model would be recovered first within early iterations. Inversions of higher frequency components of data add intermediate wavenumbers, high resolution, and details of fine scale structures and features. A careful selection and consideration of other parameters such as source-receiver positions, frequency bandwidth, starting model, damping parameters etc, are crucial. Note that cycle-skipping occurs when the modeled and observed waveforms are initially separated in time by more than half a wave-cycle at the lowest frequency available in the data (Beydoun and Tarantola, 1988). Therefore, the key to overcome the problem is either to have very low frequencies present in the observed data or to begin with an accurate starting velocity model (Brenders and Pratt, 2007; Virieux and Operto, 2009).

2.5 Effects of density and attenuation on velocity model building

In this section I will show with numerical examples on how attenuation and density affect the reconstruction of the compressional velocity model. Most acoustic wave-equation based FWI algorithms assume homogeneous density and ignore attenuation. We stress that the primary objective in this is velocity model building. FWI is a data fitting technique, which means the waveform inversion matches both amplitude and phase of the numerical modeled data with observed data. The amplitude of reflected waves is influenced by acoustic impedance. Acoustic impedance is the product of compressional velocity and density. However, most FWI schemes based on the acoustic approximation ignore the density by assuming only one variable, compressional velocity, and incorrectly fits the amplitude. One of main reasons for this approximation is that density is the challenging parameter to reconstruct along with velocity (Jeong and Min, 2012; Prieux et al., 2013). The same is also true for attenuation (Kamei and Pratt, 2013). One has to incorporate the second-order derivatives of the objective function in the numerical inversion scheme to mitigate their trade-off effects. Multiparameter inversion of density, attenuation and velocity is a challenging problem of FWI and has not been extensively studied in acoustic FWI (with only a few recent works (Jeong and Min, 2012; Kamei and Pratt, 2013; Prieux et al., 2013)).

In order to start understanding the role and effects of density and attenuation on velocity model building, we start with the true Marmousi velocity model as shown in the Figure 2.6a. Figure 2.6b is the smooth velocity model used to initiate FWI.

Figure 2.7a & 2.7b are density and attenuation models derived from the true Marmousi velocity model, respectively. The density was estimated using the empirical formula of Gardner relationship for rock velocities, $\rho \approx 310v^{0.25}$ (Gardner et al., 1974). The attenuation

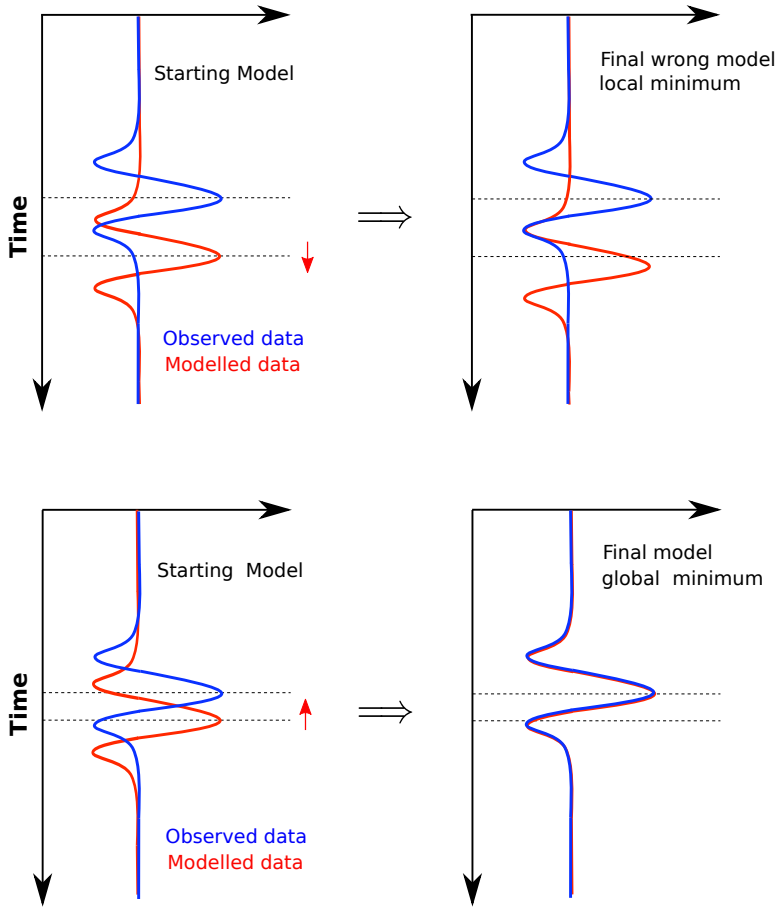


Figure 2.5: Schematic illustration of the cycle-skipping problem in FWI. If the initial modeled data and observed data have a lag that is more than half a cycle, then the local inversion moves in the wrong direction (top). If the lag is within half a cycle, then the local inversion steers in the right direction (bottom).

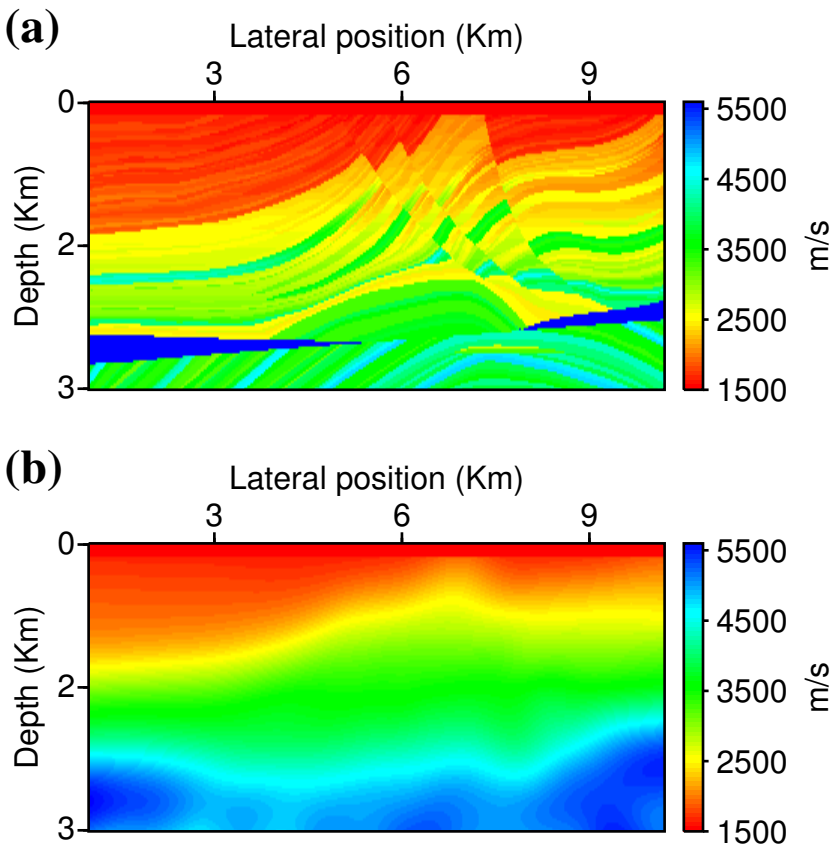


Figure 2.6: Marmousi velocity model (a) and smooth velocity model used as a starting velocity model (b).

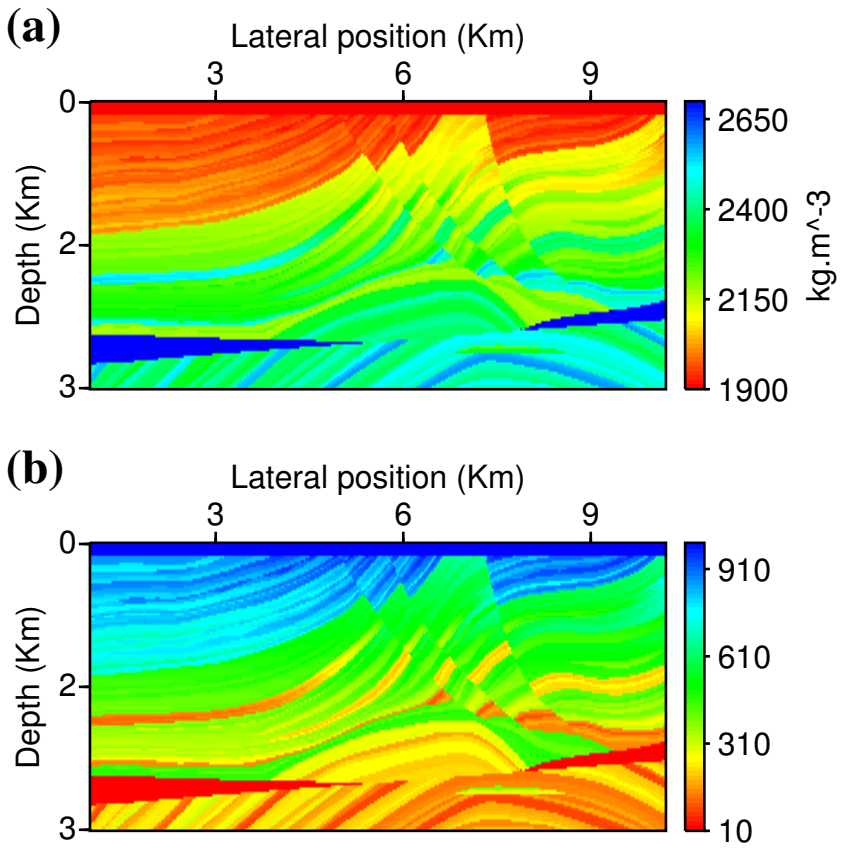
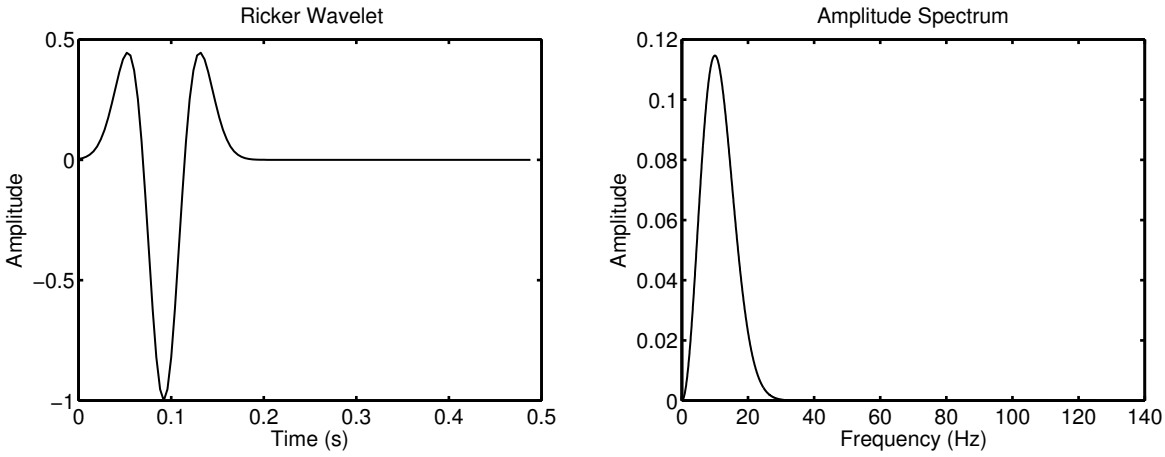


Figure 2.7: Marmousi density (a) and attenuation (b) models.

is also empirically computed from the velocity model. For FWI, we consider and compare the following three scenarios

- with pure compressional velocity
- with and without density
- with and without attenuation

First we generate a synthetic data from 96 sources and 192 receivers that are equally spaced at the surface. For the source function, we used a Ricker wavelet with a 10 Hz central frequency. Unless otherwise stated, in this thesis I used a Ricker wavelet with 10 Hz central frequency as shown in the Figure 2.8. In frequency domain, complex sources can also be



(a) (b)

Figure 2.8: Ricker wavelet with 10 Hz central frequency (left) used as a source signature and its amplitude spectrum (right).

implemented in a straight forward way. On the right hand side of Figure 2.8 is its amplitude spectrum.

For each case, first synthetic seismogram data were generated as follows: in the first case a synthetic data was generated from pure compressional velocity, in the second case the data was generated from the true density and compressional velocity, and in the third case the data was generated from the true compressional velocity and true attenuation. For the waveform inversion, in each cases we used acoustic wave approximation algorithm. The main goal here is to access effects of attenuation and density on velocity model building. For the inversion tests, 10 discrete frequencies were selected between 3.5 Hz and 18.0 Hz and the inversion is carried out in a multiscale hierarchical strategy sequentially from low to high frequency data. For the optimization engine, a nonlinear conjugate gradient method is used and each frequency runs for a maximum of 60 iterations.

Figure 2.9a is the reconstructed velocity model when the density and attenuation are assumed constant both in the data and the inversion. Figure 2.9b is the result when the data were generated from the true density and velocity models whereas the inversion assumed constant density. As we see from the reconstructed models, the lack of incorporating density in the forward modeling engine for the waveform inversion affects the amplitude of the estimated velocity model, in particular the deeper parts of the velocity model seem weaker (see Figure 2.9a & b). When attenuation is considered in the data, the amplitude of velocity estimation is really hammered by loses of the propagated amplitude energy through the media and also the waveform broadening due to physical dispersion (see Figure 2.9c).

From these numerical inversion results, we infer that in the acoustic wave approximation full waveform inversion that ignores both density and attenuation, velocity estimation is more affected by attenuation than density. Note that in frequency-domain it is numerically easy to implement both density and attenuation in the forward modeling engine if the velocity and attenuation are known. Practically, none of these physical quantities are known. Even though there are some progresses are made in retrieving velocity and attenuation or density in the multiparameter FWI, the trade-off effects between velocity and attenuation or velocity and density makes the multiparameter FWI very challenging and difficult (Jeong and Min, 2012; Prieux et al., 2013; Kamei and Pratt, 2013; Kurzmann et al., 2013). In this thesis, our forward modeling engine can incorporate both attenuation and density, however, the inversion process is restricted to the forward modeling engine that takes only pure compressional velocity model and constant density. Hence, elastic effects are ignored. In examples of BP/EAGE and Marmosui-II velocity models where densities are known, the synthetic data are computed with the true velocity and density models whereas the waveform inversion is carried out by assuming constant density (Billette and Brandsberg-Dahl, 2005; Bourgeois, A., M. Bourget, P. Lailly, M. Poulet, P. Ricarte, and R. Versteeg, 1991). This would create a mismatch in the Physics between the observed and the modeled data. Similarly, attenuation in all examples is ignored ($\mathbf{Q}$ is assumed infinity).

2.6 Summary

In this chapter, we discussed the theoretical formulations of full waveform inversion in frequency-domain that is based on the least-squares norm, l_2, of misfit between the observed data and predicted data is detail. The mathematical formulations include the simultaneous source techniques. The use of simultaneous sources that randomly assembles individual sources into super-shots is to gain computational speed up. For the FWI optimization engine, both the gradient-based and Newton-based optimization methods were formulated using the second-order adjoint state method. The importance of the starting velocity model and quality of availability of the low frequency data for efficient full waveform inversion were described. In the next three chapters, we will extensively use the application of full waveform inversion based on the mathematical formulation described in this chapter.

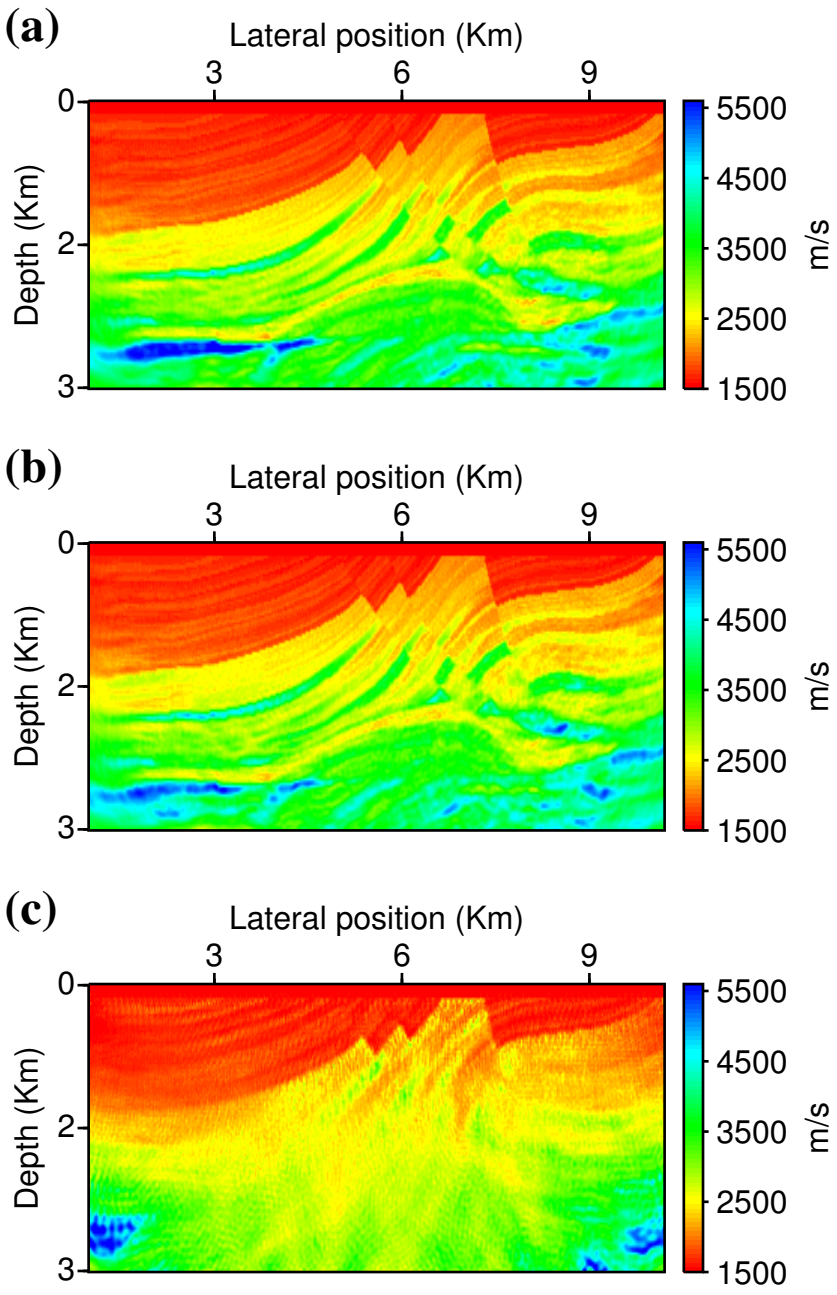


Figure 2.9: Inverted Marmousi velocity model. When density and attenuation are assumed constant (a), when the modeling is performed using the true density while the inversion is performed using constant density (b) and attenuation (b) models.

CHAPTER 3

Model parameterization strategies for Newton-based acoustic full waveform inversion[1]

3.1 Introduction

The goal of Full Waveform Inversion (FWI) is to estimate high resolution images of subsurface structures by minimizing the misfit between the observed seismograms and modeled data through local iterative optimization methods (Lailly, 1983; Tarantola, 1984b, 1987; Stekl and Pratt, 1998; Pratt et al., 1998; Operto et al., 2007; Virieux and Operto, 2009). Even though full waveform inversion can provide high resolution images, many challenges impede its practical implementation for large scale problems. One of the main problems is the lack of efficient optimization schemes to cope with the nonlinear optimization problem associated to the retrieval of a velocity model from observed wavefields. Solutions are often trapped in local minima and as a result one needs to provide a good initial velocity model and/or data with high-quality low frequency components (Brenders and Pratt, 2007; Virieux and Operto, 2009).

Various iterative optimization methods based either on gradient or Newton methods such as steepest-descent, nonlinear conjugate gradient (NCG) (Hestenes and Stiefel, 1952), quasi-Newton limited-memory BFGS (l-BFGS) (Nocedal, 1980), Gauss-Newton (GN) and full Newton (FN) methods have been employed as an optimization engine for FWI (Pratt et al., 1998; Brossier et al., 2010; Hu et al., 2011; Herrmann et al., 2011; Anagaw and Sacchi, 2012a). For large scale problems, given the computational cost of simulating multiple sources on every iteration, gradient-based methods are the most practical and popular optimization

[1] A version of this chapter has been submitted to IOP for publication in Inverse Problems

engine for FWI. However, gradient-based optimization methods are generally very slow and have difficulties to properly scale the reconstructed model parameter perturbations in comparison to Newton's methods. The most common method to improve the rate of convergence in the gradient optimization algorithm is to scale the gradient using the diagonal components of a pseudo-Hessian matrix (Shin et al., 2001).

On the other hand, Newton-based methods that require the Hessian matrix have a fast convergence rate compared to gradient-based methods. However, the improved convergence rate comes at the expense of having to solve the inverse of a large dense Hessian matrix. Quasi-Newton methods provide cost effective alternatives to the full inversion of the Hessian matrix. For instance, the limited-memory BFGS (l-BFGS) method has become a popular tool for solving the FWI problem. The l-BFGS method approximates the inverse of the Hessian matrix without having to explicitly compute or store in memory the Hessian matrix. Instead, the l-BFGS method approximates the inverse of a positive definite full Hessian matrix from a limited number of gradients and model perturbation updates (Nocedal, 1980).

Even though the Newton-based methods have a faster convergence rate, the choice of model parameterization influences the behaviour of the Hessian matrix and thereby affects the convergence rate and the quality of the reconstructed model parameters. In acoustic full waveform inversion, the earth model estimation is often carried out via velocity model parameterization. However, other types of model parameterizations can also be employed such as slowness and slowness squared. These three model parameterizations do not change the topology of the cost function but they affect the way the gradient is scaled, and consequently, the convergence rate and the accuracy of the final solution.

In this chapter, we study the influence of the three model parameterizations on Newton-based full waveform inversion. The acoustic wave approximation is used in the inversion process. For our numerical algorithm, we adopt a multiscale approach in the frequency-domain, where a finite number of discrete frequencies are selected and the inversion is carried out sequentially starting from low to high frequency data components (Sirgue and Pratt, 2004). First, long wavelength components of the model parameters are retrieved from low frequency data and then more detailed structures and features are extracted as the inversion proceeds to higher frequencies. In the multiscale inversion technique, the model parameters obtained from the inversion of the low frequency data components are used as a starting model parameter for the higher frequency data components (Bunks et al., 1995).

This chapter is organized as follows. First, we present the formulation and implementation of the Newton-based method for FWI. In Newton's method, we used a matrix-free conjugate gradient (CG) iterative technique to approximate solutions of the Newton equations. The matrix-free Newton-based method is formulated using the second-order Lagrangian formulation. The gradient of the objective function can be directly obtained from the

first-order adjoint state formulation (Plessix, 2006). Then, we investigate the influence of model parametrization on the convergence rate and on the final solution via an eigenvalue decomposition analysis of the Hessian matrix. Finally, the Gauss-Newton (GN) and the full Newton (FN) methods are used for full waveform inversion using the three aforementioned model parameterizations. Our numerical tests use the Marmousi model (Billette and Brandsberg-Dahl, 2005).

3.2 Theory

We begin by start considering the forward problem from equation 2.2

$$A(\mathbf{m},\omega)\mathbf{p}(\omega)_s = \mathbf{f}(\omega)_s. \tag{3.1}$$

The discrete Helmholtz operator $\mathbf{A}(\mathbf{m},\omega)$ depends explicitly on the model parameters $\mathbf{m}$. The model parameters $\mathbf{m}$ can be written in terms of velocity (v), slowness (v^{-1}) or slowness squared (v^{-2}). We then follow the same constrained optimization method strategies as described in Chapter 2. Here, we only reiterate the solutions to the Newton's step

$$\Re e[\mathbf{H}]\triangle\mathbf{m} = -\Re e[\mathbf{g}]. \tag{3.2}$$

And the full Hessian ($\mathbf{H}$) and the approximate Hessian matrix, $\mathbf{H}_{GN}$ (approximation Gauss-Newton Hessian matrix), are given by

$$\begin{aligned}\mathbf{H} =&((\nabla_{\mathbf{m}}\mathbf{A})\mathbf{p_s})^{\dagger}(\mathbf{A}^{\dagger})^{-1}\mathbf{r}^{\dagger}\mathbf{r}\mathbf{A}^{-1}(\nabla_{\mathbf{m}}\mathbf{A})\mathbf{p_s} + \mu\nabla^2_{\mathbf{m}}R\\ &+\mathbf{K}-\mathbf{B}^{\dagger}\mathbf{A}^{-1}\mathbf{C}-\mathbf{C}^{\dagger}(\mathbf{A}^{\dagger})^{-1}\mathbf{B}\end{aligned} \tag{3.3}$$

and

$$\mathbf{H}_{GN} = ((\nabla_{\mathbf{m}}\mathbf{A})\mathbf{p_s})^{\dagger}(\mathbf{A}^{\dagger})^{-1}\mathbf{r}^{\dagger}\mathbf{r}\mathbf{A}^{-1}(\nabla_{\mathbf{m}}\mathbf{A})\mathbf{p_s} + \mu\nabla^2_{\mathbf{m}}. \tag{3.4}$$

For the regularization term, $R(\mathbf{m})$, we have adopted a zero-order quadratic regularization term (Zhdanov, 2002).

In the next section we will describe and see how the three types of model parameterizations changes the properties of the Hessian matrices in equations 3.3 & 3.4, and the efficiency of optimization scheme based on gradient-based and Newton-based methods.

3.2.1 Numerical optimization

As stated in the previous chapter, the first-order optimization methods such as the steepest descent and conjugate gradient methods require only the gradient of the objective function. To compute the gradient of the objective function, it requires the solutions of two forward problems; one for the forward modeling $\mathbf{p}_s(\omega)$ and the other for the adjoint state $\mathbf{\Lambda}_s(\omega)$. The gradient, for each frequency, is then computed by taking the product between the forward-propagated wavefields and back-propagated residual wavefields in space, see equations 3.11 & 3.12. For gradient-based optimization, we used a nonlinear conjugate gradient (NCG) with a line search that satisfies the Wolfe conditions (Hestenes, 1969; Nocedal and Wright, 2006) (see Algorithm 1). We leave the reader to see the details of the steepest descent and conjugate gradient algorithm in any optimization books (Nocedal, 1980; Nocedal and Wright, 2006).

Shin et al. (2001) and Plessix and Mulder (2004) have shown that the gradient-based iterative method for geophysical inverse problem converges slowly and the final model result obtained is poorly scaled. Instead, in order to speed up the numerical inversion and scale the model parameters, they propose a preconditioned nonlinear conjugate gradient method, where the preconditoner is the diagonal of elements the pseudo-Hessian matrix. The diagonal of elements of pseudo-Hessian matrix is efficiently computed by making use of the source-receiver reciprocity theory

$$\mathbf{H}_{xx} = \Re e\Big[\sum_{\omega}\omega^4\sum_{s}|f(\mathbf{x}_s,\omega)|^2|\mathbf{G}(\mathbf{x}_s,\mathbf{x},\omega)|^2\sum_{r}|\mathbf{G}(\mathbf{x},\mathbf{x}_r,\omega)|^2\Big], \tag{3.5}$$

where $\mathbf{G}(\mathbf{x}_s,\mathbf{x},\omega)$ is the Green's function for a given shot computed via

$$A(\mathbf{m},\omega)\mathbf{G}(\mathbf{x}_s,\mathbf{x},\omega) = \delta(\mathbf{x}-\mathbf{x_s}). \tag{3.6}$$

In the same way the Green's function of the receiver, $\mathbf{G}(\mathbf{x},\mathbf{x}_r,\omega)$, can also be computed.

On the other hand, a faster convergence rate can be achieved with the full Newton and Gauss-Newton methods (Bube and Langan, 2008; Anagaw and Sacchi, 2012a). However, these methods are computationally expensive for large scale inverse problems. In order to solve the Newton's step, the Newton-based algorithms require the gradient, and inverse of the Hessian matrix. As described in the previous chapter, the Hessian matrix is very dense and large, and is impractical to solve the inverse of the Hessian for large scale problems. Practically, the Newton's direction that computes the model perturbation update is solved using an iterative method (Kuhn and Tucker, 1951; Akcelik et al., 2002). In this chapter, we adopt the conjugate gradient method (CG) to estimate the Newton's direction. The

CG scheme inside the Algorithm 3 requires only the action of the Hessian on the model perturbation, i.e a matrix-vector product. The matrix-vector product can be implemented in a matrix-free manner. For a single frequency, as described in Chapter 2, once the matrix $\mathbf{A}$ is factorized using LU decomposition, a set of solutions for multiple shots can be computed by forward and backward substitutions at a relatively low computational cost. To compute the action of the Hessian on a vector inside the CG, we only need to solve two forward problems and the product is computed on the fly at every CG iteration.

The product of the full Hessian matrix on a vector $\mathbf{x}$, $\mathbf{Hx}$, is computed as follows:

$$\begin{aligned}\mathbf{Hx} =&[((\nabla_{\mathbf{m}}\mathbf{A})\mathbf{p_s})^{\dagger}(\mathbf{A}^{\dagger})^{-1}\mathbf{r}^{\dagger}\mathbf{r}\mathbf{A}^{-1}(\nabla_{\mathbf{m}}\mathbf{A})\mathbf{p_s} + \mu\nabla^2_{\mathbf{m}}R \\ &+ \mathbf{K} - \mathbf{B}^{\dagger}\mathbf{A}^{-1}\mathbf{C} - \mathbf{C}^{\dagger}(\mathbf{A}^{\dagger})^{-1}\mathbf{B}]\mathbf{x}.\end{aligned} \tag{3.7}$$

Note that

$$\mathbf{G} = (\nabla^{\mathbf{2}}_{\mathbf{m}}\mathbf{A})\mathbf{p}_s(\omega)\mathbf{\Lambda}^{\dagger}_s(\omega) \tag{3.8}$$

$$\mathbf{B} = (\nabla_{\mathbf{m}}\mathbf{A}^{\dagger})\mathbf{\Lambda}_s(\omega) \tag{3.9}$$

$$\mathbf{C} = (\nabla_{\mathbf{m}}\mathbf{A})\mathbf{p}_s(\omega). \tag{3.10}$$

- First, we perform two forward modeling to compute the gradient

$$\mathbf{p}_s(\omega) = \mathbf{A}^{-1}\mathbf{S}(\omega)_{nsp} \tag{3.11}$$

$$\mathbf{\Lambda}_s(\omega) = \mathbf{A}^{-1}\mathbf{r}^{\dagger}W \triangle \mathbf{d}(\omega)^*. \tag{3.12}$$

  The term $\nabla^2_{\mathbf{m}}R\mathbf{x}$ and $\mathbf{Gx}$ in equation 3.7 are easily computed, for example

$$\mathbf{Kx} = (\nabla^{\mathbf{2}}_{\mathbf{m}}\mathbf{Ax})\mathbf{p}_s(\omega)\mathbf{\Lambda}^{\dagger}_s(\omega). \tag{3.13}$$

- We then pass both $\mathbf{p}_s(\omega)$ & $\mathbf{\Lambda}_s(\omega)$ to the CG iterative algorithm, and we compute for

$$\mathbf{y} = (\nabla_{\mathbf{m}}\mathbf{Ax})\mathbf{p}_s(\omega). \tag{3.14}$$

- Inside the CG, we then perform two more forward modeling to compute the action of the Hessian on a vector

$$\mathbf{z} = \mathbf{A}^{-1}\mathbf{y} \tag{3.15}$$

$$\mathbf{q} = \mathbf{A}^{-1}[\mathbf{r}^{\dagger}\mathbf{r}\mathbf{z} - (\nabla_{\mathbf{m}}\mathbf{A}\mathbf{x})\mathbf{\Lambda}_s^{\dagger}]. \tag{3.16}$$

Finally, we get all the terms in the right hand side of equation 3.7

$$[\mathbf{B}^{\dagger}\mathbf{A}^{-1}\mathbf{C}]\mathbf{x} = (\nabla_{\mathbf{m}}\mathbf{A}\mathbf{\Lambda}_s^{\dagger})\mathbf{y} \tag{3.17}$$

$$[((\nabla_{\mathbf{m}}\mathbf{A})\mathbf{p_s})^{\dagger}(\mathbf{A}^{\dagger})^{-1}\mathbf{r}^{\dagger}\mathbf{r}\mathbf{A}^{-1}(\nabla_{\mathbf{m}}\mathbf{A})\mathbf{p_s} - \mathbf{C}^{\dagger}(\mathbf{A}^{\dagger})^{-1}\mathbf{B}]\mathbf{x} = ((\nabla_{\mathbf{m}}\mathbf{A}\mathbf{p}_s)^{\dagger})\mathbf{q}. \tag{3.18}$$

This results in

$$\mathbf{H}\mathbf{x} = ((\nabla_{\mathbf{m}}\mathbf{A}\mathbf{p}_s)^{\dagger})\mathbf{q} - (\nabla_{\mathbf{m}}\mathbf{A}\mathbf{\Lambda}_s^{\dagger})\mathbf{y} + \mathbf{K}\mathbf{x} + \nabla_{\mathbf{m}}^2 R\mathbf{x}. \tag{3.19}$$

In the Gauss-Newton approximation, the Hessian matrix is a positive definite matrix (see equation 3.4) whereas in the full Newton method, the full Hessian is an indefinite symmetric matrix. Therefore the standard CG method is no longer suitable for such a system of linear equations. As a result, we adopt the stopping criterion proposed by Eisenstat et al. (1994) in the inner loop of our algorithm (see Algorithm 3) to avoid solving for the negative curvature direction of Newton equations otherwise CG would fail or yields wrong result. Once the Newton's direction is solved, the model is updated iteratively using line search method

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha \triangle \mathbf{m}_k. \tag{3.20}$$

The above iterative technique to solve Newton's equations is called the inexact or truncated Newton method. Algorithm 3 is the pseudo-code of an inexact full Newton and Gauss-Newton methods; the stopping criterion is also used to avoid over solving the system of Newton's equations. Hence, we terminate the CG iterative early or if negative curvature is detected. Then, the model is updated using descent direction.

The other economical cheap method for solving the approximate solution of the inverse of the Hessian matrix is the l-BFGS method (Nocedal, 1980). Algorithm 4 is the pseudo-code of the l-BFGS method. The l-BFGS method approximates the inverse of the Hessian matrix without forming the Hessian matrix. To approximate the inverse of the Hessian matrix, the l-BFGS method only requires the history of the gradient and model updates from previous iterations.

Algorithm 3: Pseudo-code: Multiscale Newton's method

Input : $\mathbf{d}^{obs}(\omega)$, $\mathbf{m}_0$, N_w, max_iter
Output: $\arg\min_{\mathbf{m}} J(\mathbf{m})$
Initialize parameters;
for $w_i \leftarrow 1$ **to** N_w **do**
  Starting model $\leftarrow \mathbf{m}_k$;
  for $k \leftarrow 1$ **to** max_iter **do**
    Compute:
    forward wavefields $\mathbf{d}^{cal}(\omega)$;
    `residual wavefields` $\triangle\mathbf{d}(\omega)$;
    `back-propagate residual wavefields` $\Lambda(\omega)$;
    gradient $\nabla_{\mathbf{m}} J(\mathbf{m}_k)$;
    Solve:
    $\triangle\mathbf{m}_k = -\Big[\Re e(\mathbf{H})\Big]^{-1}\Re e(\nabla_{\mathbf{m}} J(\mathbf{m}_k))$;
    solve $\triangle\mathbf{m}_k$ using CG method: call Algorithm 2;
    Two forward modeling operators are required to compute;
    the action of the Hessian matrix on $\triangle\mathbf{m}$ inside the CG algorithm;
    Compute:
    α using line search method;
    Update model:
    $\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha\triangle\mathbf{m}_k$;
  end
end

Algorithm 4: Pseudo-code: Multiscale FWI using limited-memory quasi-Newton l-BFGS method

Input : $\mathbf{d}^{obs}(\omega)$, $\mathbf{m}_0$, n, ngw, max_iter
Output: $\arg\min_{\mathbf{m}} J(\mathbf{m})$
Initialize parameters;
for $w_i \leftarrow 1$ **to** N_w **do**
  Starting model $\leftarrow \mathbf{m}_k$;
  for $k \leftarrow 1$ **to** max_iter **do**
    Compute:
    gradient $\mathbf{g}_{\mathbf{m}_k} \leftarrow \nabla_{\mathbf{m}} J(\mathbf{m}_k)$;
    Diag. Hessian $\mathbf{H}_k$;
    $\mathbf{G}_{\mathbf{m}_k} \leftarrow \mathbf{H}_k^{-1}\mathbf{g}_{\mathbf{m}_k}$;
    if $(k > n)$ **then**
      Discard $\mathbf{s}_{k-n-1}$;
      Discard $\mathbf{y}_{k-n-1}$;
    end
    if $(n > 1)$ **then**
      store: $\mathbf{s}_{k-1} = \mathbf{m}_k - \mathbf{m}_{k-1}$;
      store: $\mathbf{y}_{k-1} = \mathbf{G}_{\mathbf{m}_k} - \mathbf{G}_{\mathbf{m}_{k-1}}$;
    end
    $\mathbf{q} \leftarrow \mathbf{G}_{\mathbf{m}_k}$;
    for $i = k-1, k-2, \ldots$ **to** $k-n$ **do**
      $\rho_i \leftarrow \frac{1}{\mathbf{y}_i^T \mathbf{s}_i}$;
      $\alpha_i \leftarrow \rho_i \mathbf{s}_i^T \mathbf{q}$;
      $\mathbf{q} \leftarrow \mathbf{q} - \alpha_i \mathbf{y}_i$;
    end
    $\gamma_n \frac{\mathbf{s}_{n-1}^T \mathbf{y}_{n-1}}{\mathbf{y}_{n-1}^T \mathbf{y}_{n-1}}$;
    $\mathbf{r} \leftarrow \gamma_n \mathbf{q}$;
    for $i = k-n, k-n+1, \ldots$ **to** k **do**
      $\beta \leftarrow \rho_i \mathbf{y}_i^T \mathbf{r}$;
      $\mathbf{r} \leftarrow \mathbf{r} + \mathbf{s}_i(\alpha_i - \beta)$;
    end
    $\triangle\mathbf{m} \leftarrow -\mathbf{r}$;
    Update model:
    $\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha\triangle\mathbf{m}$;
    α using line search method;
  end
end

3.2.2 Model parametrization

In this section we describe the three types of model parameterization that we have studied for 2D acoustic full waveform inversion. Even though our main goal is to estimate velocity, one can parametrize the model parameter $\mathbf{m}$ in terms of velocity (v), slowness (v^{-1}) or slowness squared (v^{-2}). In this study, we only consider those three types of model parameterizations. The shape of the misfit function in equation 2.4 does not change by the type of the model parameterization adopted by the numerical optimization scheme. However, the choice of model parameterization influences the eigenvalue spectrum of the Hessian matrix and the direction of the gradient. Consequently, this also influences the convergence rate of the algorithm. The change in the direction of the gradient in equation 2.16 comes from the scaling term of the partial derivatives of the Helmholtz operator, $\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}}$. Similarly, the characteristic nature of the eigenvectors of the Hessian matrix in equation 3.3 are influenced by the factors $\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}}$ and $\frac{\partial^2 \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}^2}$. The matrix $\mathbf{A}(\mathbf{m},\omega)$ is the discretized form of the 2D Helmholtz wave-equation

$$A = \frac{\partial}{\partial x}\frac{1}{\rho}\frac{\partial}{\partial x} + \frac{\partial}{\partial z}\frac{1}{\rho}\frac{\partial}{\partial z} + \frac{\omega^2}{\rho v^2} \tag{3.21}$$

The Helmoltz operator in equation 3.21 is a large sparse matrix with nonzero diagonal elements. Computing derivatives of the above matrix with respect to the model parameter is trivial. Outside the boundary condition or PML region, the first or second derivatives have nonzero elements only along the matrix diagonal corresponding to the m_k model parameter. Table 3.1 summarizes derivatives of the matrix $\mathbf{A}(\mathbf{m},\omega)$ with respect to the three model parameterizations.

Table 3.1: Model parametrization $\mathbf{m}$. Note that derivatives of the matrix $\mathbf{A}$ lie on the diagonal.

Model Parametrization		$\partial_{m_j}\mathbf{A}_{ik}$	$\partial^2_{m_{jl}}\mathbf{A}_{ik}$
Velocity	v	$-3\omega^2\rho^{-1}v^{-3}\delta_{ij}\delta_{jk}$	$6\omega^2\rho^{-1}v^{-4}\delta_{ij}\delta_{jk}\delta_{il}\delta_{lk}$
Slowness	v^{-1}	$\omega^2\rho^{-1}v^{-1}\delta_{ij}\delta_{jk}$	$\omega^2\rho^{-1}\delta_{ij}\delta_{jk}\delta_{il}\delta_{lk}$
Slowness squared	v^{-2}	$\omega^2\rho^{-1}\delta_{ij}\delta_{jk}$	0

In the next section, we compare numerical solutions of the estimated model parameters from the Newton-based numerical scheme that incorporates the Hessian matrix associated to these model parameterizations.

3.3 Results and Discussions

Based on various numerical optimization schemes for FWI, the practical performances of the three types of model parameterizations considered in this thesis are compared. This is achieved through a series of numerical experiments using the Marmousi velocity model. In order to evaluate and assess the behaviour of the Hessian matrix on the choice of model parameterization, we initially consider a portion of Marmousi velocity model. Figure 3.1 shows the portion of the true Marmousi velocity model (Figure 3.5a), which ranges approximately from 3.9 km to 4.8 km. The analysis focuses on the eigenvalues and eigenvectors obtained from the eigenvalue decomposition of the Hessian matrix. The model is discretized on a 41×61 grid with a uniform grid spacing size of 24 m$\times$24 m in the vertical and horizontal directions, respectively. For this example, we used a single source located in the middle of the survey and one grid point size below the model surface. We also considered a total of 36 receivers at the surface.

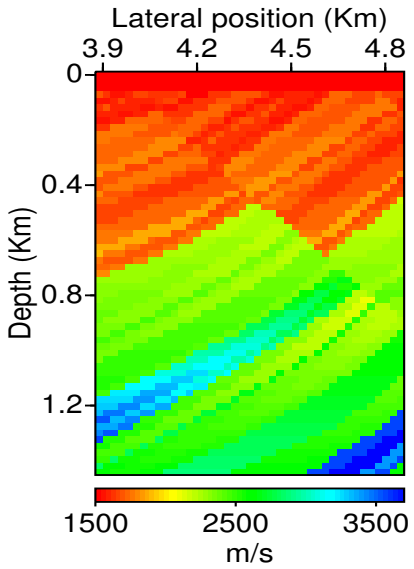


Figure 3.1: Portion of Marmousi velocity Model (Figure 3.5a) from 3.9Km to 4.8Km.

For the sake of simplicity, instead of computing the full Hessian matrix, we compute the approximate Hessian matrix in the Gauss-Newton approximation to demonstrate our claims. Figures 3.2a-c show the approximate Hessian matrix of the model (Figure 3.1) at 12Hz using velocity, slowness, and slowness squared model parameterizations, respectively. To compute the approximate Hessian matrix, the model in Figure 3.1 is smoothed 2D Hanning window over a small window length (5 m $\times$ 5 m). The approximate Hessian matrix for a single source is computed using equation 3.4 after ignoring the last two terms of the full Hessian matrix. Only the Hessian matrix in the Gauss-Newton (GN) approximation is considered. In the GN approximation, each element of the Hessian matrix represents the correlation between the partial derivatives of the wavefield at points $\mathbf{m}_j$ and $\mathbf{m}_i$ (Stekl and Pratt, 1998). According to equation 3.4, each element in the Hessian matrix is the zero-lag correlation between the recorded partial derivatives of the wavefields at receiver locations. In the high frequency limit, the partial derivatives of the wavefields with respect to two different model parameters would be uncorrelated with each other and the Hessian matrix would be mainly diagonal and banded due to auto-correlation of the partial derivatives of wavefield with itself. Often, in the high frequency approximation, the Hessian matrix can be approximated by its diagonal. However, the nonzero elements of the Hessian matrix observed off the main diagonal are due to the band limited nature of seismic data, limited acquisition geometry and heterogeneity of the model. Due to geometrical spreading of energy, the amplitude of the correlation of the partial derivative wavefields decay with depth along the main diagonal (see Figure 3.2a)

(Stekl and Pratt, 1998). In the next section, we analyze the eigenvectors and eigenvalues of the Hessian matrix in order to evaluate the influence of the three model parameterizations on the amplitude of the Hessian matrix elements.

3.3.1 Eigenvectors and Eigenvalues

As shown in Figure 3.2, the amplitude of the Hessian matrix elements decay along the main diagonal more rapidly in the velocity model parameterization than using the other two model parameterizations. In the same way, the amplitude of main diagonal decays faster when using slowness than the one obtained from the slowness squared parameterization. To better understand which model parameterization one should use, we analyze the eigenvalues and eigenvectors of the Hessian. The approximate Hessian matrix that arises from the minimization of equation 2.4 using a second-order adjoint formulation in the Gauss-Newton approximation, equation 3.4, can be factorized to

$$\mathbf{H}_{GN} = \mathbf{U}\Xi\mathbf{U}^T, \tag{3.22}$$

where Ξ and $\mathbf{U}$ are the eigenvalue and eigenvector matrices of the approximate Hessian matrix $\mathbf{H}_{GN}$, respectively. Plugging equation 3.22 into equation 2.21

$$\mathbf{g} = -\sum_{k=1}^{N} \sigma_k \mathbf{u}_k \mathbf{u}_k^T \triangle\mathbf{m}, \tag{3.23}$$

where $\mathbf{u}_k$ is the k^{th} column of the eigenvector matrix $\mathbf{U}$ and σ_k is the corresponding eigenvalue in Ξ. The matrix $\sum_{k=1}^{N} \sigma_k \mathbf{u}_k \mathbf{u}_k^T$ is called the model resolution matrix (Pratt et al., 1998). From the relationship between the gradient vector and model perturbation, in equation 3.23, we see that the gradient vector is the sum of the projected model perturbation along the eigenvectors. The direction of the eigenvectors, $\mathbf{u}_k$, corresponding to the largest and smallest eigenvalues, σ_k, point in the direction of the greatest and smallest curvature of the objective function and provide the best and worst retrieved parts of the model parameter, respectively. The eigenvector, $\mathbf{u}_k$, corresponding to the largest eigenvalue, σ_k, provides the parts of the model parameters that are most easily retrieved by FWI.

Figure 3.3 shows the largest 60 eigenvalues of the Hessian matrix for the three types of model parameterization. Figure 3.4 shows the first three eigenvectors in depth corresponding to the largest three eigenvalues. The resulting eigenvectors were extracted at the middle part of the velocity model. Plots from bottom to top shown in Figure 3.4 are the 1^{st}, 2^{nd} and 3^{rd} largest eigenvectors in depth associated with the largest three eigenvalues, respectively. The red, green and blue colors represent eigenvectors from velocity, slowness and slowness squared model parameterization, respectively. As we see from the plots, unlike

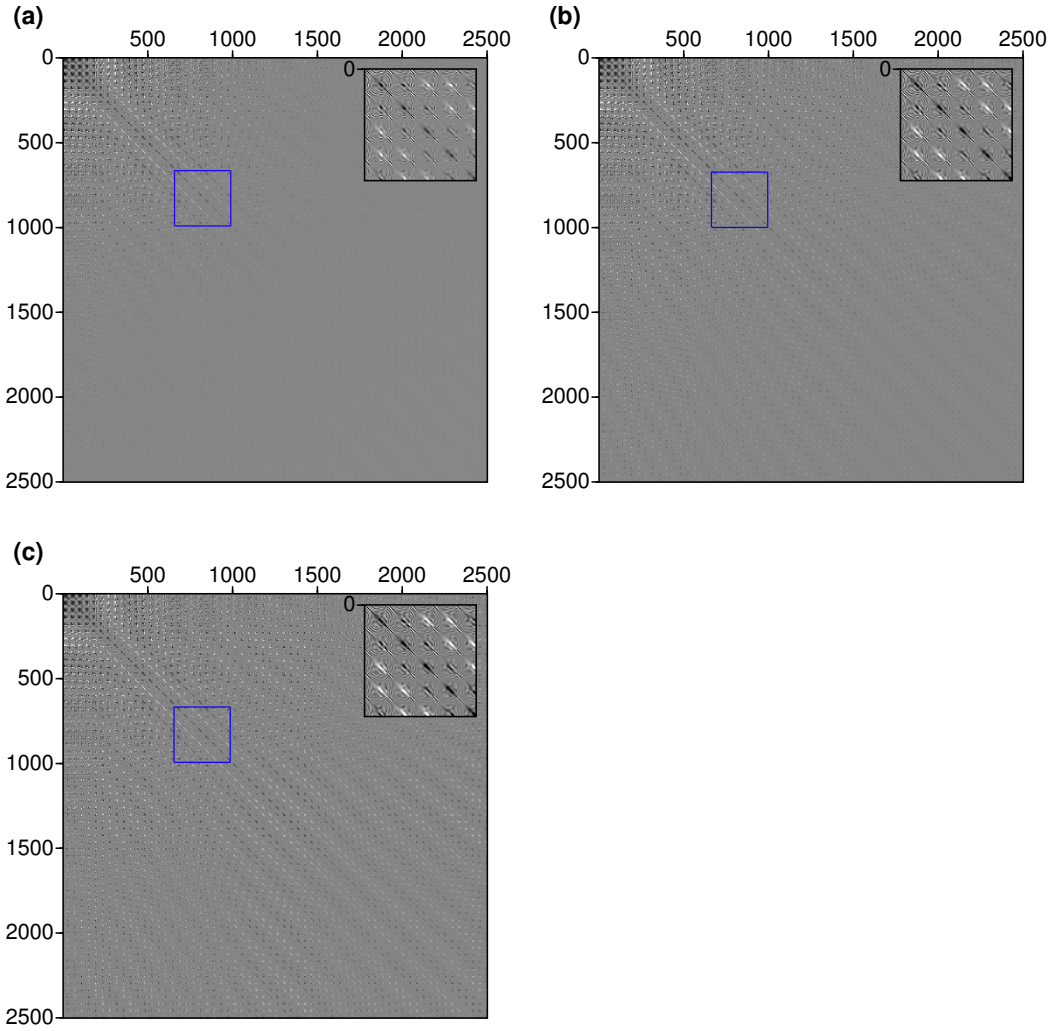


Figure 3.2: The approximate Hessian matrix for the three choices of model parameterizations from a single source at 12 Hz. The model parameterizations are: (a) velocity (v), (b) slowness (v^{-1}) and (c) square of slowness (v^{-2}).

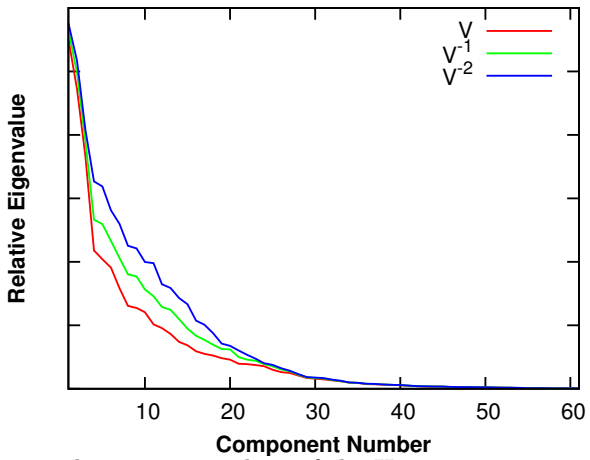


Figure 3.3: The first 60 largest eigenvalues of the Hessian matrix in Figure 3.2 for the three types of model parameterizations; velocity (red), slowness (green) and slowness squared (blue).

the eigenvectors that arise from slowness or slowness squared model parameterization, the magnitude of the eigenvectors from the velocity model parametrization decreases with depth. As a result, when the velocity model parametrization is used by our FWI code, the velocity model of very shallow parts would be better retrieved than the deeper part of the model where the velocities are generally the highest. As we see in Figure 3.4, in the case of a slowness squared parameterization, the eigenvectors are almost depth independent and are stronger than the eigenvectors from the slowness or velocity parameterization. Hence, the deeper part of the model will be properly reconstructed. The resulting gradient vectors from velocity, slowness and slowness squared model parameterizations are scaled by a magnitude proportional to $\sim 3\omega^2\rho^{-1}v^{-3}$, $\omega^2\rho^{-1}v^{-1}$ and $\omega^2\rho^{-1}$, respectively. The gradient associated to high velocities in the deeper part of model are damped by a factor $\sim 3\omega^2 v^{-3}$ in the case of the velocity model parameterization than the other two cases because the model perturbation update is proportional to gradient. The Hessian matrix will also be changed accordingly (see equation 3.3 and Table 3.1). In the same manner, the amplitude of the Hessian matrix decays the most along the main diagonal when the velocity model parameterization is adopted. As a result, there would be unbalanced amplitudes between the shallow and deeper parts of the model during the inversion process. The resulting weight in the model perturbation update would be low in the deeper parts and large in shallow parts. It is clear that the effect of the unbalanced amplitude scaling observed in the velocity parameterization case can be mitigated by the use of a slowness or slowness squared parameterization. The use of slowness or slowness squared results in eigenvectors of the Hessian matrix that are stronger in amplitude with depth than eigenvectors results from velocity parametrization. As a

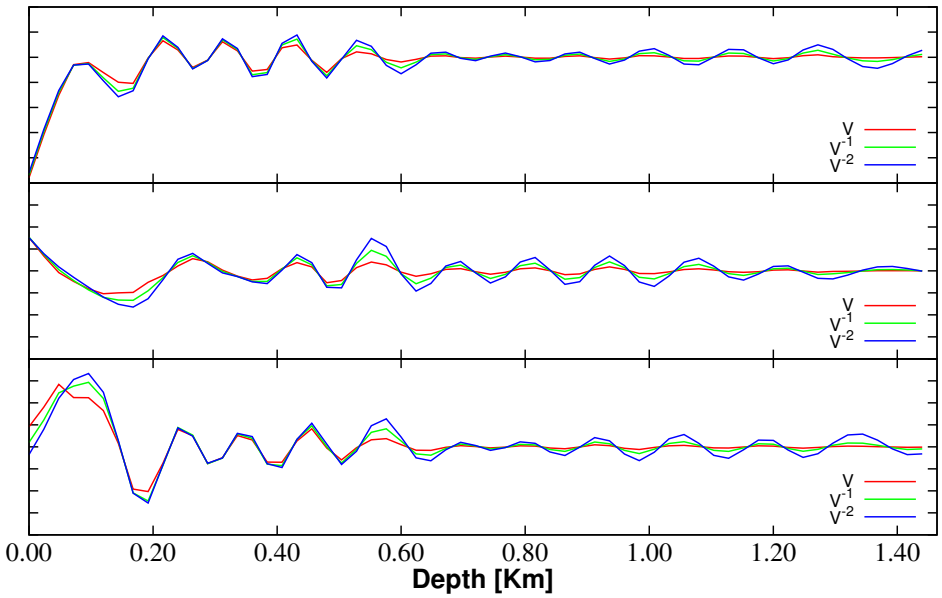


Figure 3.4: The first three eigenvectors of the approximate Hessian matrix against depth corresponding to the largest eigenvalues of the Hessian matrix using velocity (red), slowness (green) and slowness squared (blue) parametrizations. The three plots from bottom to top represent the eigenvectors to the 1^{st}, 2^{nd} and 3^{rd} largest eigenvalues, respectively.

result the deeper parts of the model would be rescaled properly and resolved well by FWI. Also, we see from the eigenvectors of the Hessian matrix in Figure 3.4 that the effect of high velocities in the deeper part of the model are less important in the case of a slowness squared parameterization than in the a case of slowness parameterization. Consequently, a more accurate reconstruction of the velocity model can be achieved by the use of slowness squared. In the next section, we demonstrate the effectiveness of the three types of model parameterizations using gradient and Newton-based optimization methods.

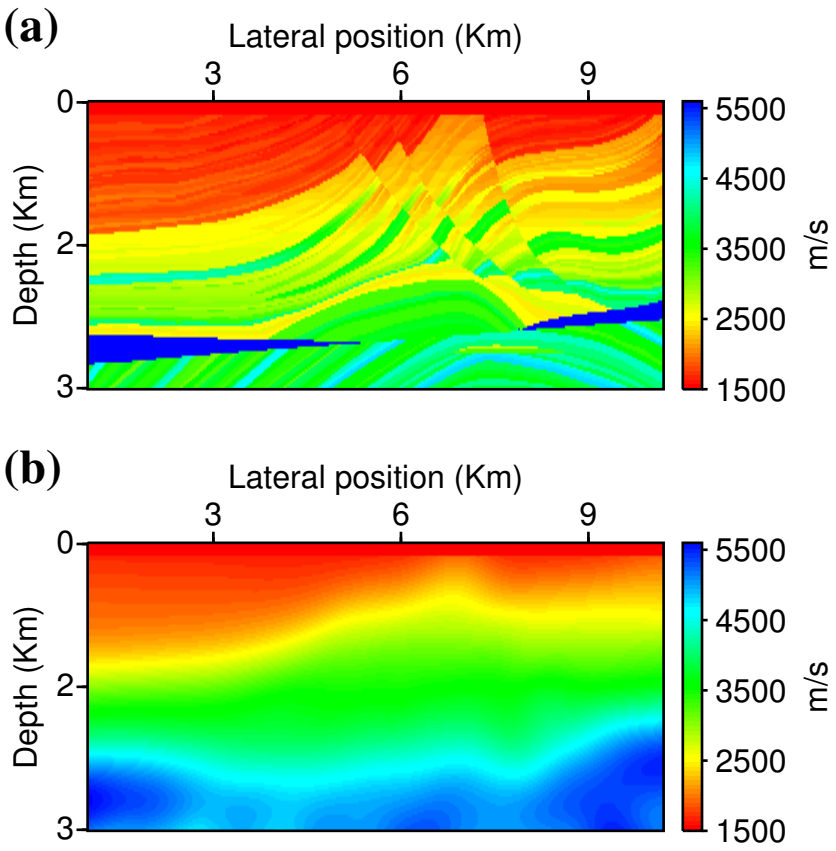


Figure 3.5: True Marmousi velocity model (a) and smooth velocity model used as starting model for inversion (b).

3.3.2 Synthetic examples on Marmousi velocity model

We first present FWI results that were obtained via gradient and Newton-based schemes on the Marmousi velocity model shown in Figure 3.5a. The 2D Marmousi velocity model is discretized by 126×384 grid points with a 25×25 m regular spacing. Figure 3.5b is a starting velocity model used to run all the inversion algorithm presented in this chapter. The starting model is obtained by smoothing the true velocity model with a 2D Hanning window. A good starting model for full waveform inversion is crucial to ensure convergence to a minimum.

3.3.3 Numerical experimental setup

Synthetic data were initially generated from 95 sources with an interval of 100 m and 192 receivers with an interval of 25 m. The sources are placed 25 m (one grid point) below the surface and receivers are placed at the surface of earth. To simulate the data, we used a source signature modeled by a Ricker wavelet with a 10 Hz central frequency. A set of 8 discrete frequencies were selected between 3.1 Hz and 21 Hz based on the strategy proposed by Sirgue et. al (Sirgue and Pratt, 2004). These frequencies were utilized by our inversion code in a multiscale fashion described by Sirgue et. al (Sirgue and Pratt, 2004). This set of frequencies include (3.15, 4.70, 6.22, 7.90, 9.52, 13.00, 15.87, 20.02 Hz). The numerical inversion is then carried out in a sequential approach starting from low to high frequencies. The final updated model of the lower frequency is used as a starting model for the inversion of the next higher frequency data component. During inversion, we recalibrate the source wavelet signature at every iteration using equation 2.8.

We consider the following numerical optimization methods: nonlinear conjugate gradient (NCG), preconditioned nonlinear conjugate gradient (PNCG), the quasi-Newton limited memory BFGS (l-BFGS), Gauss-Newton (GN) and full Newton (FN) methods. Results from the gradient based methods presented here are merely for comparison with the Newton-based methods. For the preconditioned NCG, the diagonal of the Hessian matrix in the GN approximation is taken as a preconditioner. In order to fairly compare solutions obtained from these optimization methods, all numerical inversion presented in this chapter were designed to approximately have the same number of forward modeling operations performed for each frequency. For each frequency in the FN and GN methods, we run up to a maximum of 12 GN or FN iterations and/or stop iterating whenever $||\mathbf{m}_k - \mathbf{m}_o||/||\mathbf{m_o}|| < \delta$ is satisfied; where $\mathbf{m}_k$ and $\mathbf{m_o}$ are model parameters at the k^{th} iteration and the reference model at the first iteration respectively, and $\delta = 10^{-3}$. We also used a stopping criteria proposed by Eisenstat et. al (Eisenstat et al., 1994) for solving the inner GN and FN system of equations and the average number of forward modeling performed for GN or FN methods is approximately ≈ 2.6, which leads to approximately the same computational speed. With the same stopping criteria, a maximum of 25 iterations per frequency is used for the NCG, PNCG and l-BFGS methods.

3.3.4 Numerical results

First, we analyze the influence of the three model parameterizations in the FWI numerical algorithms on the solutions of the velocity model that would be retrieved from FWI using Newton-based methods. Figure 3.6a-e are the reconstructed velocity models after the end

of the 1^{st} frequency data component using the GN and FN methods. From top to bottom are the reconstructed velocity models using velocity (v), slowness (v^{-1}) and slowness squared (v^{-2}) parameterizations, respectively. On the right from top to bottom (a), (b) and (c) are results obtained using the GN method and on the left (d), (e) and (f) are results obtained using the FN method, respectively. We also compare these results with the quasi-Newton l-BFGS method. Figure 3.7a-c are the corresponding reconstructed velocity models using l-BFGS after the 1^{st} frequency inversion using velocity, slowness and slowness squared parameterizations, respectively. As we see from these reconstructed velocity models, long wavelength structures of the velocity model are well resolved by the use of slowness or slowness squared parameterization. The deep parts of the model, where velocities are the highest, are better determined in the case of slowness and slowness squared parameterizations (see the relative reconstructed image quality measurements in Table 3.2). Thus, the solutions from slowness and slowness squared parameterizations have a more balanced amplitude in the shallow and deep parts of the model. The results confirm our previous analysis from the eigenvalue decomposition of the Hessian matrix.The results shown in Figures 3.6 and 3.7 confirm that the GN and FN outperform the l-BFGS method. Note that, to have a fair comparison between the Newton-based methods, all the inversion scheme were implemented to have approximately the same number of forward modeling operations. We do not see many differences between the GN and FN results. This is due to the lack of strong amplitude multi-reflected wavefields in the data as there are no strong velocity contrast structures in the Marmousi model (see Chapter 5 for other cases). As a result, we would expect both methods to provide similar results.

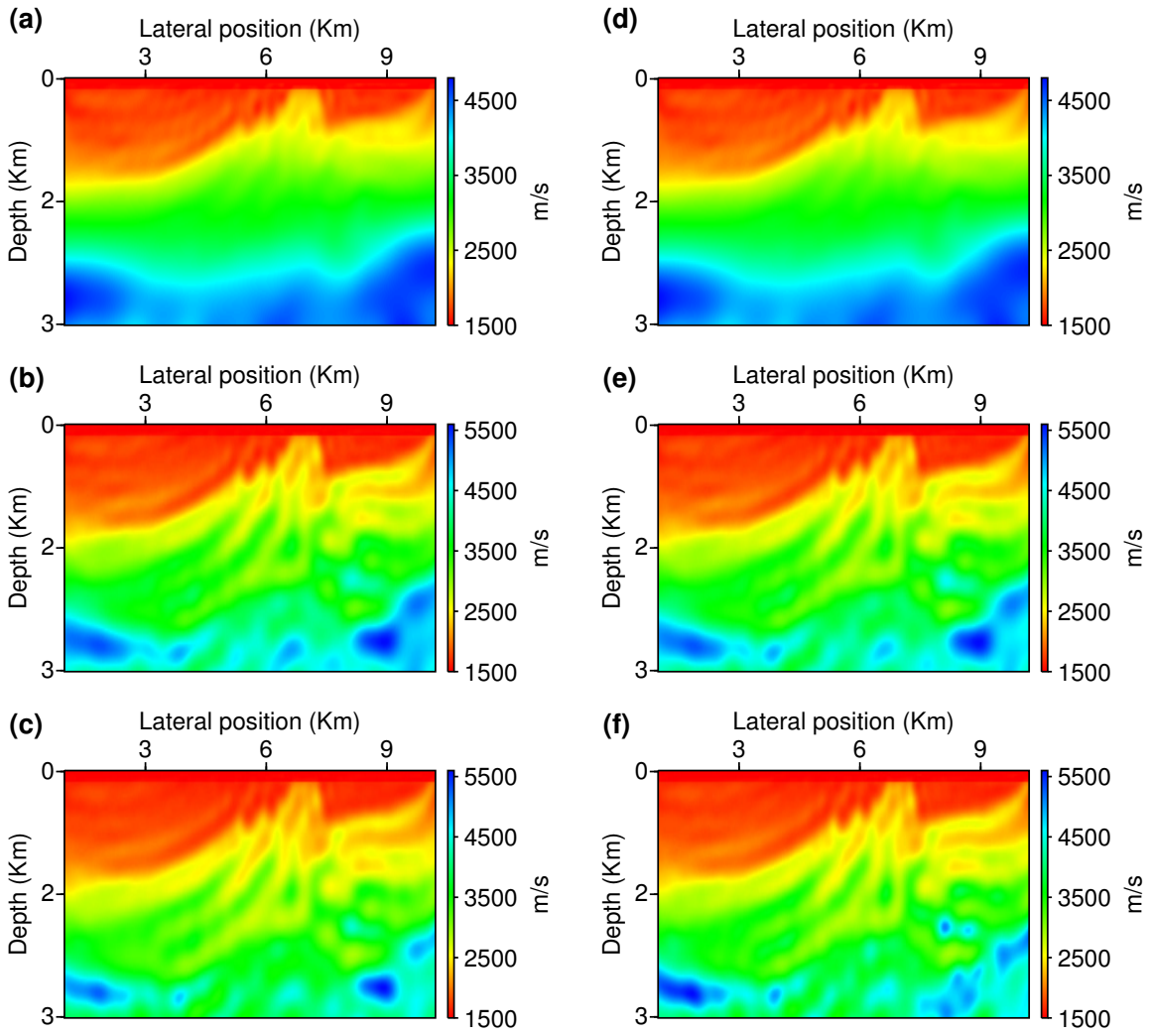


Figure 3.6: Gauss-Newton and full Newton FWI. Reconstructed velocity models after the end of the 1^{st} frequency data component inversion. From top to bottom are the reconstructed velocity models using velocity (v), slowness (v^{-1}) and slowness squared (v^{-2}) parameterization, respectively. On the right from top to bottom (a), (b) and (c) are results obtained using Gauss-Newton and on the left (d), (e) and (f) are results obtained using full Newton method.

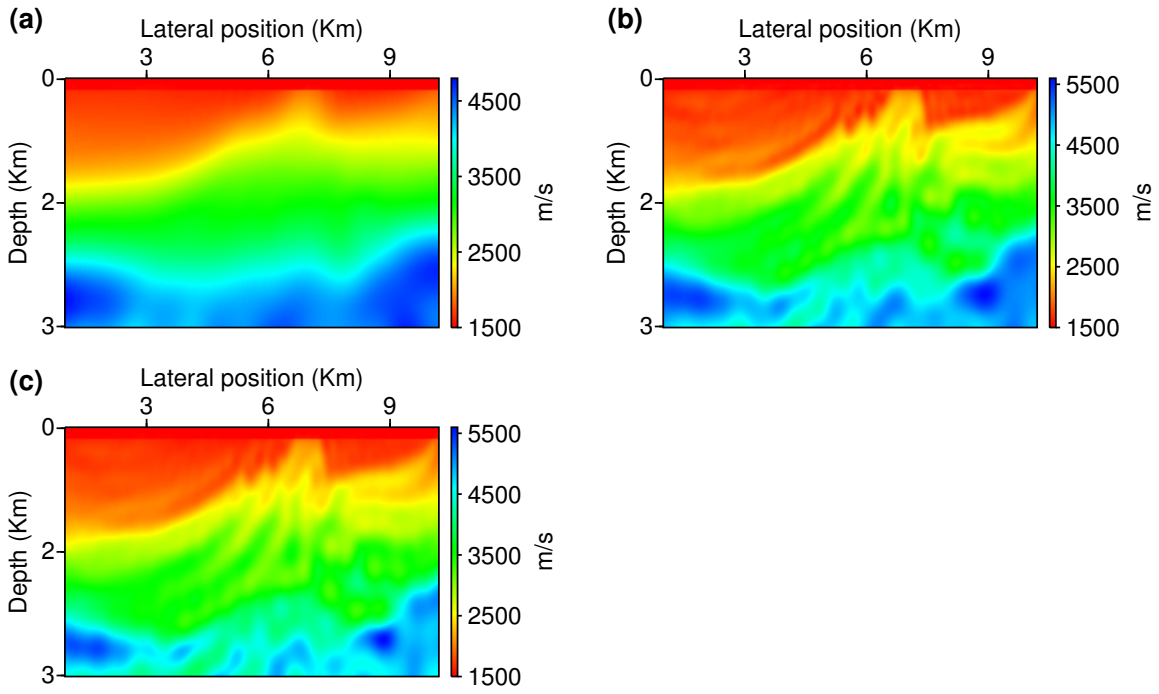


Figure 3.7: Full waveform inversion using the *l*-BFGS method. Reconstructed velocity models after the end of the 1^{st} frequency data component inversion with velocity (v) (a), slowness (v^{-1}) (b) and slowness squared (v^{-2}) (c) parameterization, respectively.

Figure 3.8, 3.9 and 3.10 show reconstructed velocity model results from various full waveform inversion optimization engines using velocity, slowness and slowness squared parameterization, respectively. Each figure contains reconstructed models estimated by the NCG, PNCG, l-BFGS, GN and FN methods. The results show that all the optimization methods, except PNCG, are struggling to retrieve the deep part of the model when FWI is parametrized in terms of velocity. The main reason for better results in the deeper parts of the model in the case of PNCG FWI is that its gradient is scaled by the inverse of the diagonal of the Hessian matrix. As a result, the shallow and the deep parts of the model are well retrieved by the algorithm. The shallow parts of the model are well retrieved by the NCG, GN and FN methods, and by PNCG as well. The l-BFGS algorithm for this numerical setup with velocity parametrization fails to converge. On the other hand, when a slowness or slowness squared parametrization is employed, the deeper parts and all features of the original Marmousi velocity model are properly reconstructed. The reconstructed velocity model obtained by using a slowness squared parametrization provides a more accurate velocity model than the other two model parameterizations. All features of the original Marmousi velocity model are reconstructed in the correct position (see Figure 3.10).

In order to further examine the influence of the three model parameterizations, we compare the vertical velocity profile of the results obtained from all numerical optimization methods investigated in this thesis. Figures 3.11, 3.12 and 3.13 show the velocity profiles extracted at (2.25 km, 0 km) and (3.5 km, 0 km). To make the results more visible, we plotted the vertical velocity profiles obtained by the l-BFGS, GN and FN methods with the three types of model parameterizations on separate figures. Figures 3.14, 3.15 and 3.16 depict the corresponding vertical velocity profiles obtained by the l-BFGS, GN and FN methods, respectively. These results highlight once again that the velocity profiles reconstructed both on the shallow and deep parts of the model are fairly well resolved when slowness or slowness squared parameterizations are employed. In the case of velocity model parameterization, almost all methods fail to properly recover the deep parts of the model. Hence, these observations suggest that the presence of the unbalanced amplitude between the shallow and deep parts of the model in the velocity parameterization of FWI is mitigated by the use of a slowness or slowness squared parametrization.

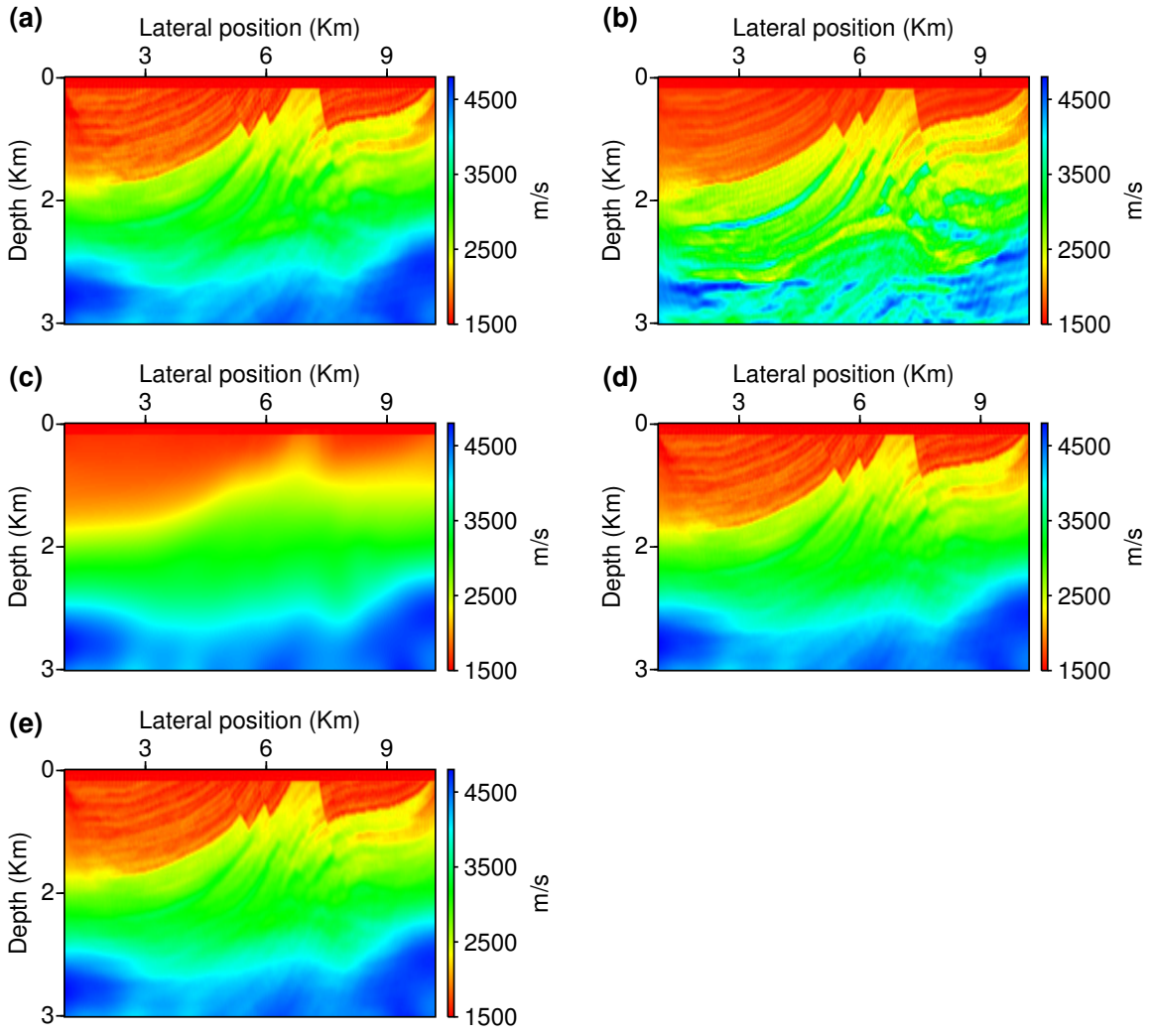


Figure 3.8: Reconstructed velocity models after the end of all frequency data components using velocity (v) model parameterization: (a) nonlinear conjugate gradient (NCG), (b) preconditioned nonlinear conjugate gradient (PNCG), (c) quasi-Newton l-BFGS, (d) Gauss-Newton (FN) and (e) full Newton (FN) methods. Note that the l-BFGS fails to converge, as result the final model looks like the initial model.

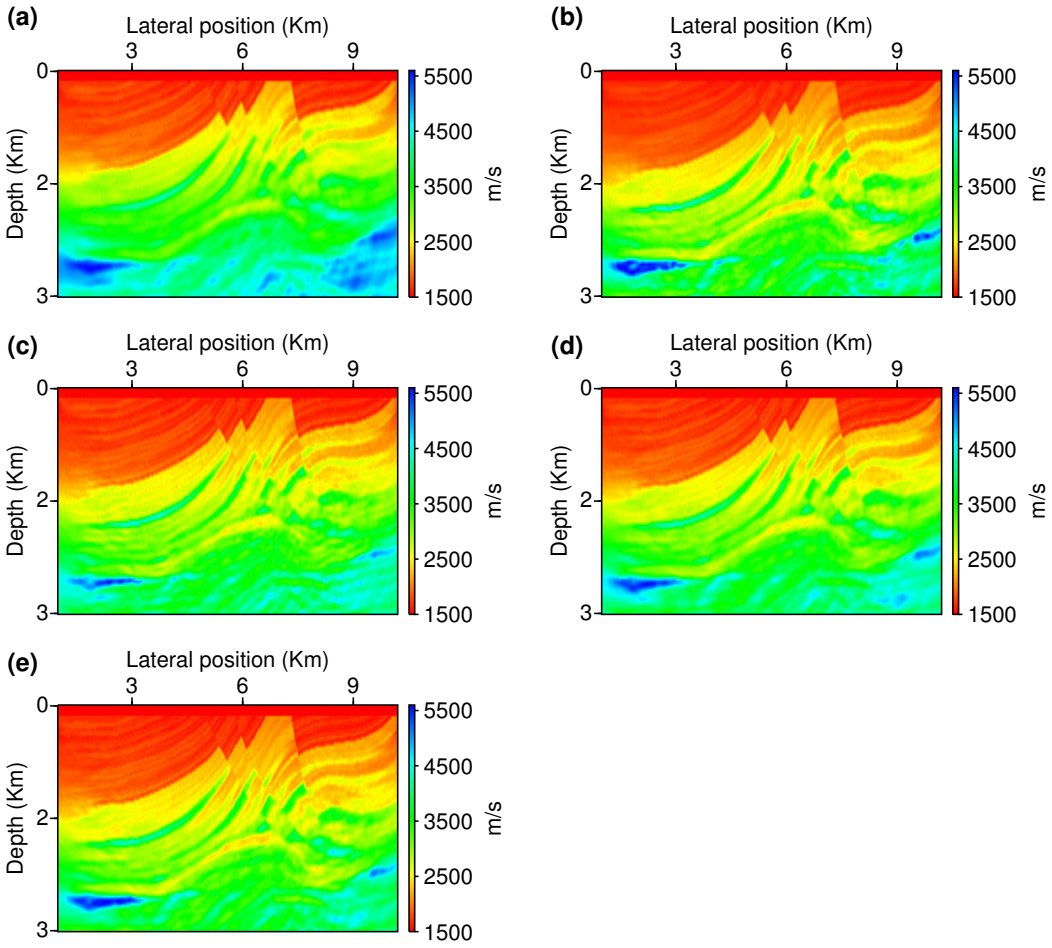


Figure 3.9: Reconstructed velocity models after the end of all frequency data components using slowness (v^{-1}) parameterization: (a) nonlinear conjugate gradient (NCG), (b) preconditioned nonlinear conjugate gradient (PNCG), (c) quasi-Newton l-BFGS, (d) Gauss-Newton (FN) and (e) full Newton (FN) methods.

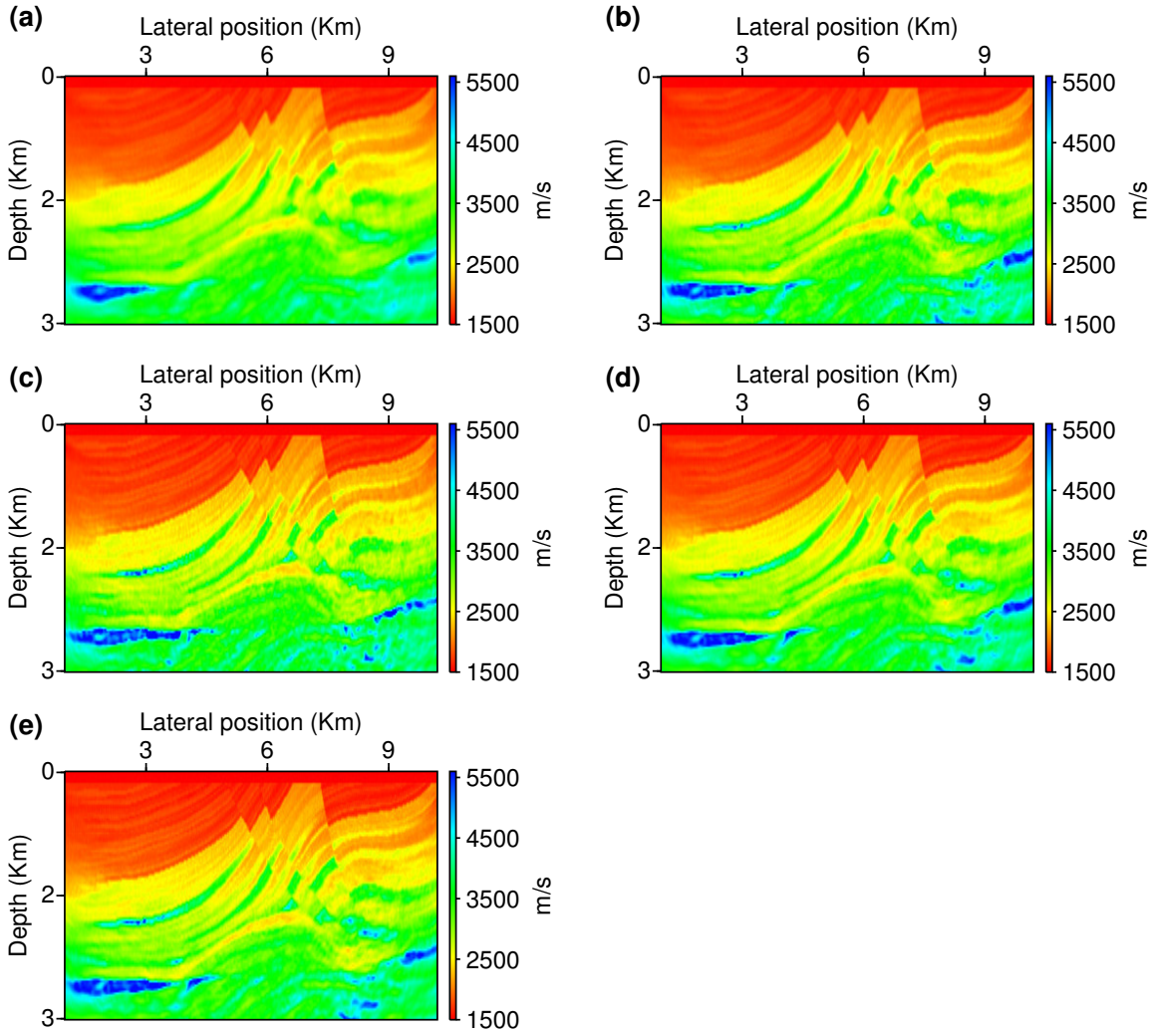


Figure 3.10: Reconstructed velocity models after the end of all frequency data components using slowness squared (v^{-2}) parameterization: (a) nonlinear conjugate gradient (NCG), (b) preconditioned nonlinear conjugate gradient (PNCG), (c) quasi-Newton l-BFGS, (d) Gauss-Newton (FN) and (e) full Newton (FN) methods.

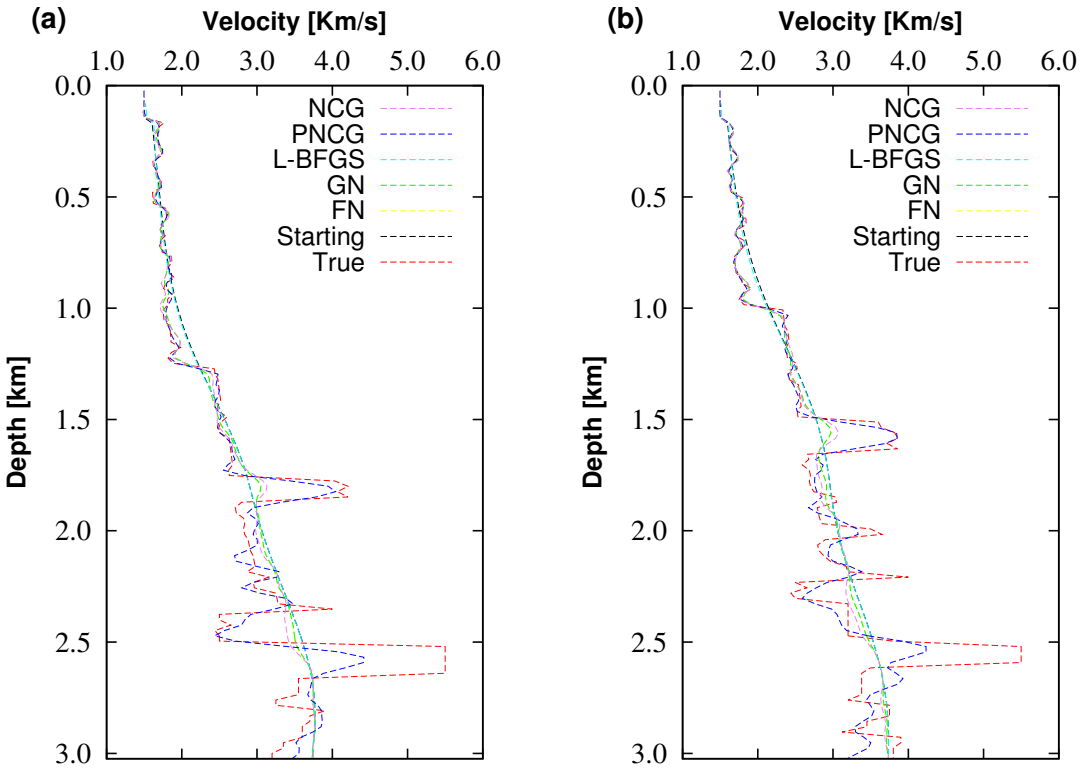


Figure 3.11: Comparison of vertical velocity profiles of the reconstructed velocity models obtained from different optimization methods using velocity model parametrization. The depth velocity profiles are extracted from Figure 3.8 at (2.25 km,0 km) (a) and (3.50 km,0 km) (b).

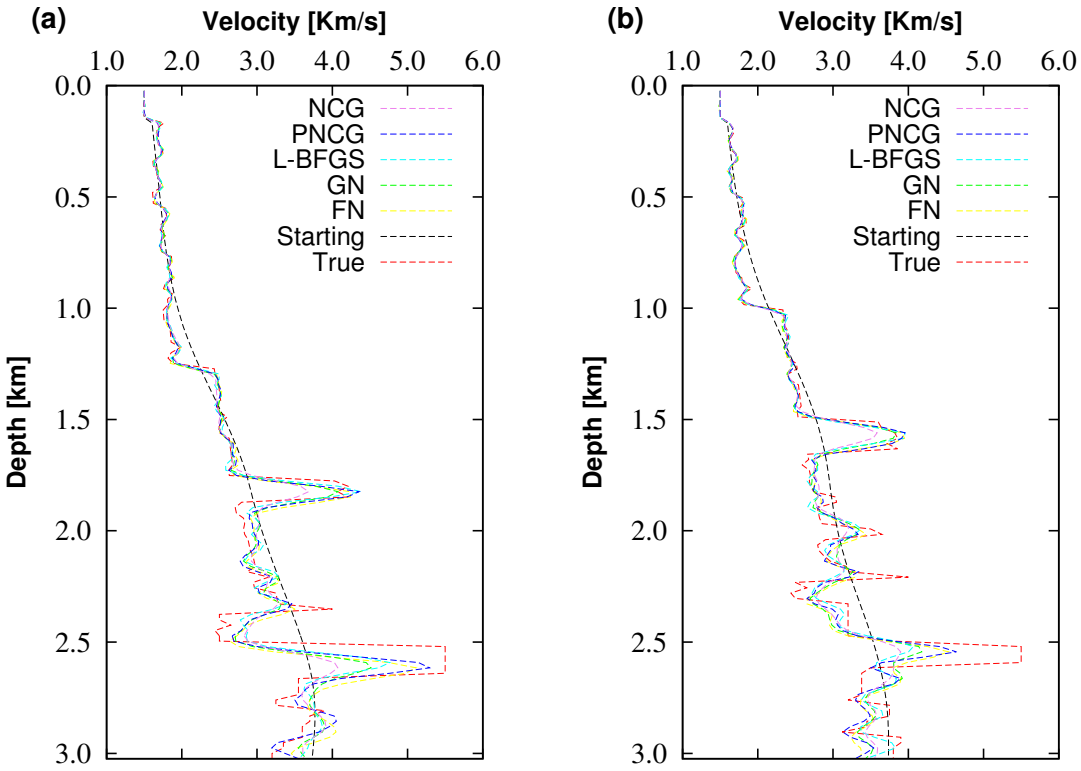


Figure 3.12: Comparison of vertical velocity profiles of the reconstructed velocity models obtained from different optimization methods using slowness parametrization. The depth velocity profiles are extracted from Figure 3.9 at (2.25 km,0 km) (a) and (3.50 km,0 km) (b).

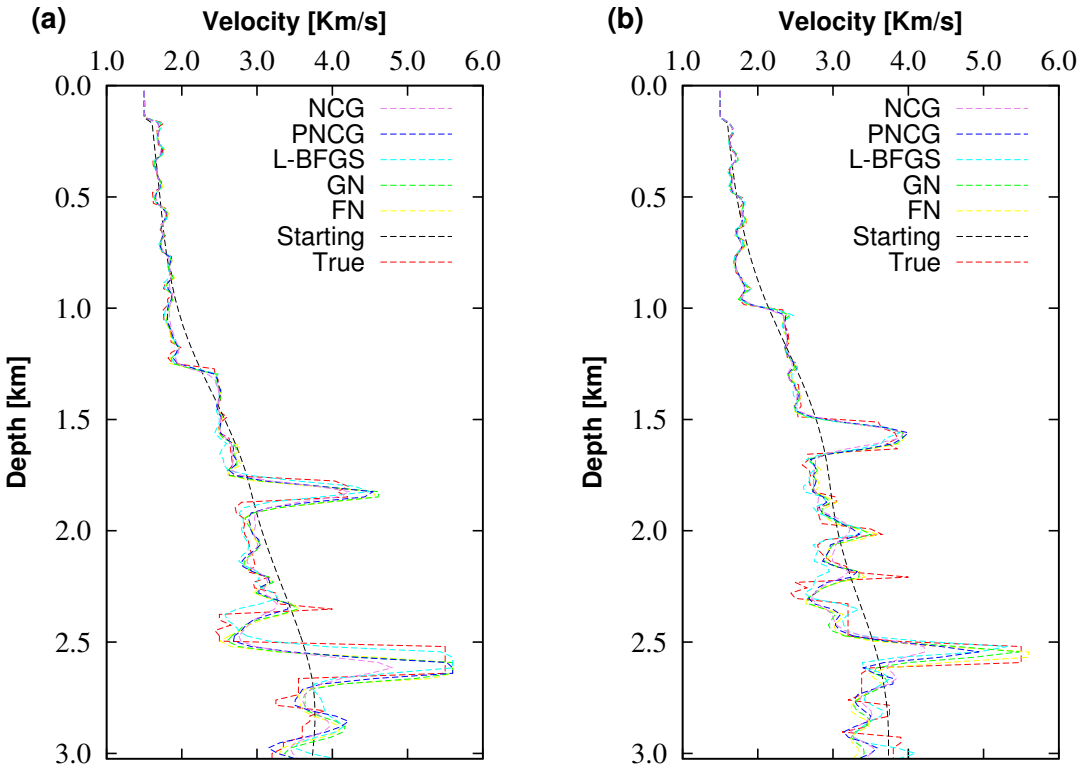


Figure 3.13: Comparison of vertical velocity profiles of the reconstructed velocity models obtained from different optimization methods using slowness squared parametrization. The depth velocity profiles are extracted from Figure 3.10 at (2.25 km,0 km) (a) and (3.50 km,0 km) (b).

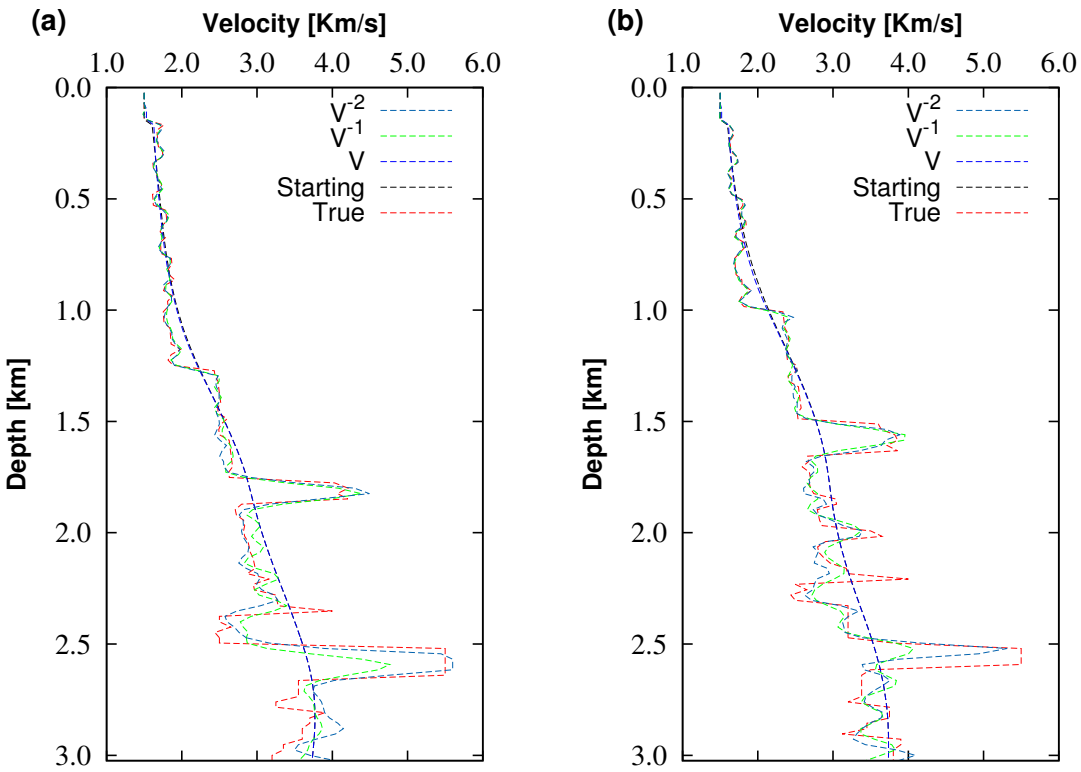


Figure 3.14: Vertical velocity profiles of the reconstructed velocity model obtained using the three model parametrizations by l-BFGS method. The depth velocity profiles are extracted at (2.25 km,0 km) (a) and (3.5 km,0 km) (b).

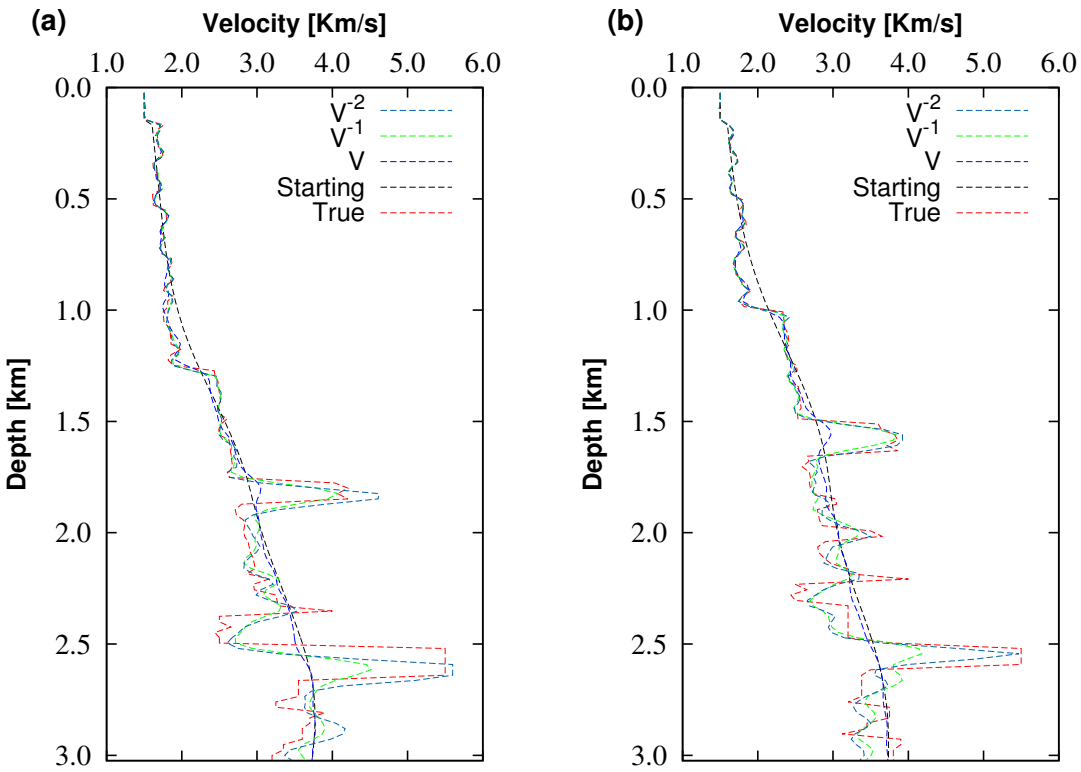


Figure 3.15: Vertical velocity profiles of the reconstructed velocity model obtained using the three model parametrizations by Gauss-Newton method. The depth velocity profiles are extracted at (2.25 km,0 km) (a) and (3.5 km,0 km) (b).

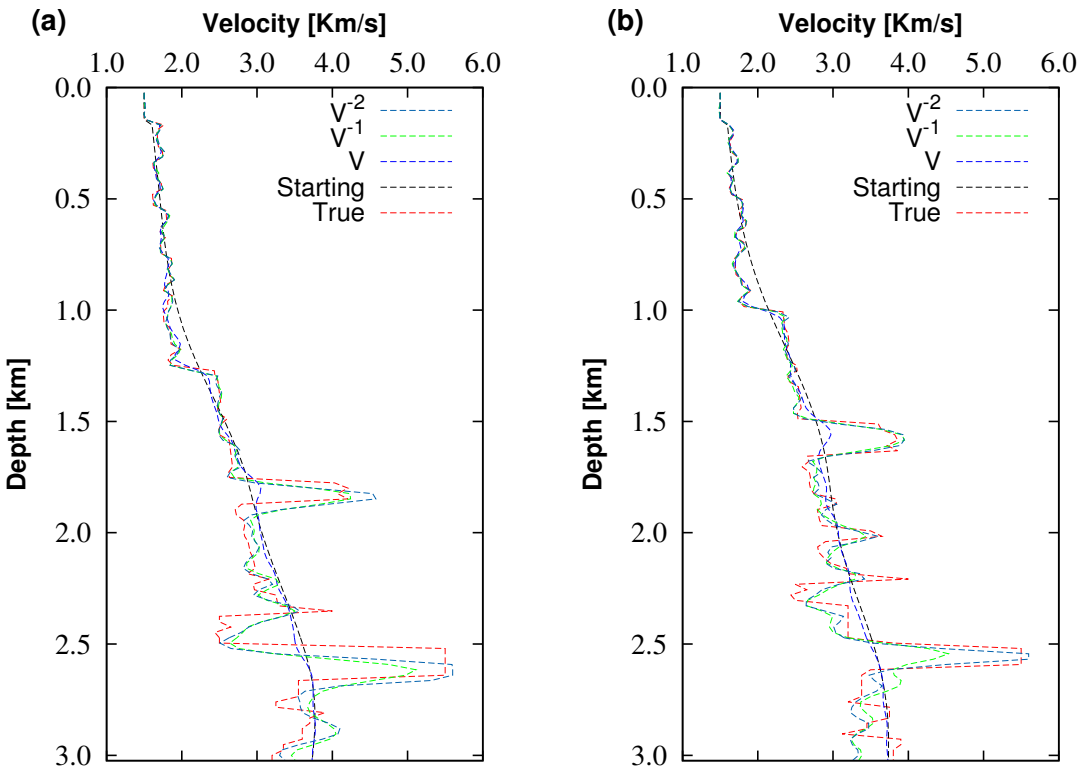


Figure 3.16: Vertical velocity profiles of the reconstructed velocity model obtained using the three model parametrizations by full Newton method.The depth velocity profiles are extracted at (2.25 km,0 km) (a) and (3.5 km,0 km) (b).

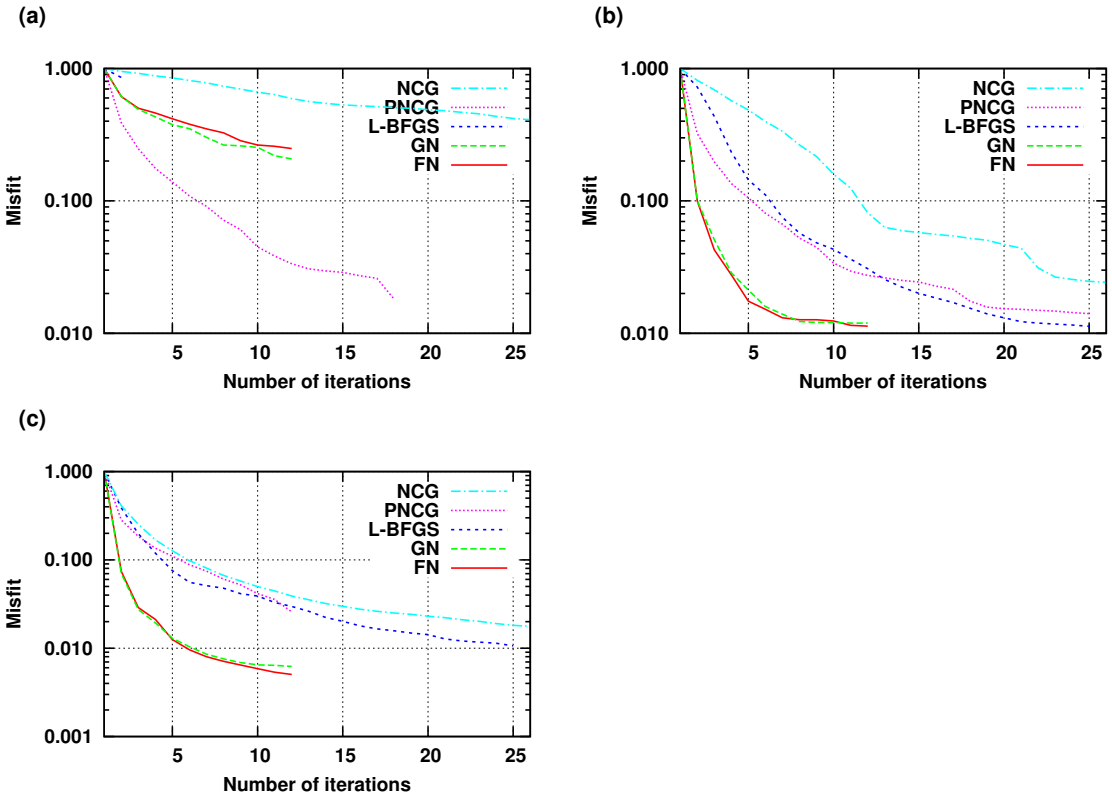


Figure 3.17: Relative data misfit reduction of the Marmousi velocity model for different optimization methods using three types of model parametrizations: velocity (v) (a), slowness (v^{-1}) (b) and the square of slowness (v^{-2}) (c) at 4.70 Hz.

3.3.5 Convergence properties

An efficient optimization engine with a good line search method is a key for any practical applications of full waveform inversion for large scale geophysical inverse problems. In this chapter, we compare the convergence properties of all optimization methods presented based on the three types of model parameterizations employed in our numerical inversion scheme. Figures 3.17 and 3.18 show the convergence rate of the nonlinear conjugate gradient, preconditioned nonlinear conjugate gradient, limited memory BFGS, Gauss-Newton and full Newton methods at 4.70 Hz and 20.02 Hz, respectively. In the plots, each misfit function is normalized by the misfit function value at the first iteration for each frequency.

The convergence rate of the gradient and Newton-based methods, in the case of velocity parameterization show that the PNCG method converges faster than the other methods because it updates the deeper parts of the model and scales the amplitude of the model

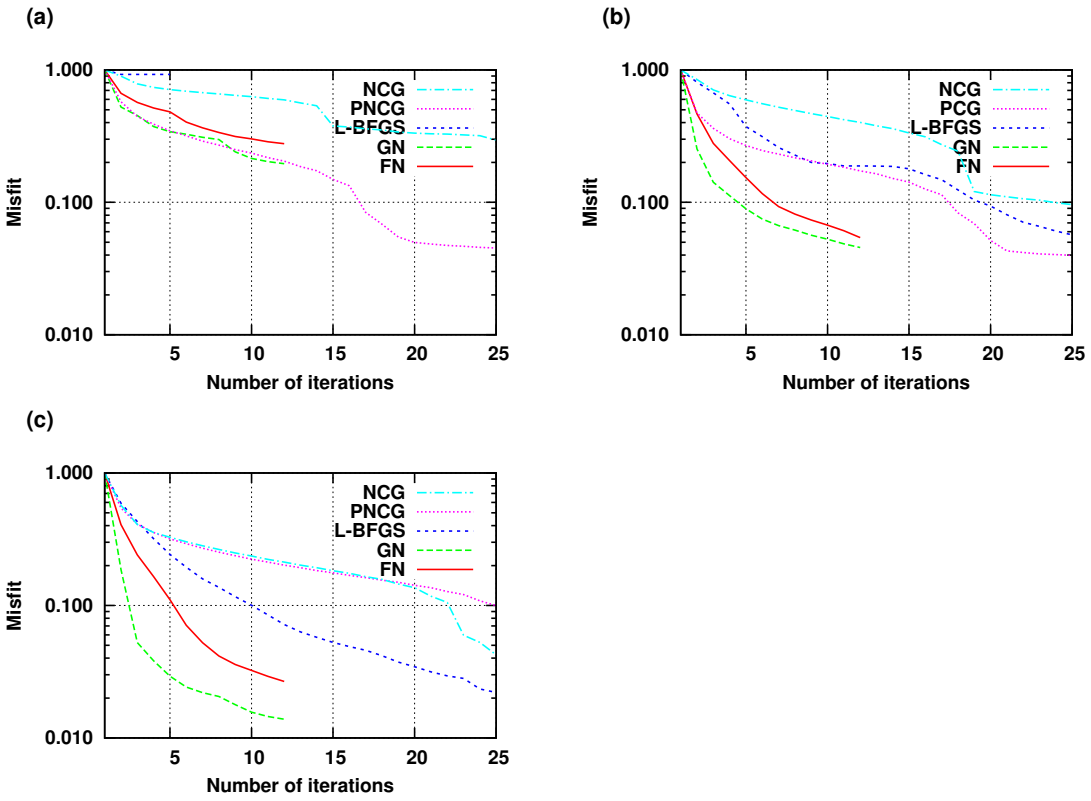


Figure 3.18: Relative data misfit reduction of the Marmousi velocity model for different optimization methods using three types model parametrizations: velocity (v) (a), slowness (v^{-1}) (b) and the square of slowness (v^{-2}) (c) at 20.02 Hz.

perturbation in the deeper parts. Hence, the optimization scheme properly updates the deep and shallow parts, and as a result the inversion proceeds in the right direction faster. The convergence rate of NCG is the slowest in all three types of model parameterizations, and hence requires many iterations to achieve a desired solution.

As a result of the lack of proper scaling in either the shallow or deeper parts of the model and the failure in the line search direction for updating the model parameter, the optimization scheme does not provide a satisfactory inversion result (see Figures 3.8c). However, a much improved rate of convergence can be achieved by the use of a slowness or slowness squared parameterization as these model parameterizations provide proper scaling of the descent direction both in the deep and shallow parts of the model. Compared to the velocity parameterization method, the performance of the GN and FN methods shown in Figures 3.17 and Figures 3.18 are superior in the case of slowness or slowness squared model parameterization. To have a clear understanding of the influence of the three types of model parameterization on convergence rates of the Newton-based optimization methods such as the GN and FN methods, we plotted them in Figure 3.19. In the left, from top to bottom are misfit functions for 4.70 Hz and 20.02 Hz datasets using Gauss-Newton method, respectively. To the right from top to bottom are misfit functions using full Newton optimization method. A faster reduction in the misfit function using either GN or FN method is achieved using slowness squared than a slowness or velocity type model parameterization. From the eigenvalue and eigenvector analysis, we see that the slowness squared parameterization performs better in scaling the deeper part of the model than slowness and velocity parameterizations. Consequently, better updates to the model occur in the deep parts during early iterations, and leads to the rapid reduction in the misfit function. One can also find different model parametrizations. However, the change in the model paramterizations should be chosen in such a way that the elements of the Hessian matrix are changed so as to provide good balance amplitude between the deeper and shallower parts of the model.

From our numerical results, we conclude that, velocity estimation methods that employ FWI based on the GN and FN optimization methods should be carried out using slowness or slowness squared parameterization. Superior convergence rates of Gauss-Newton and full Newton methods can be achieved by the use of a slowness squared model parameterization. In the next section, we address the noise sensitivity of the three types of model parameterizations in the Gauss-Newton and full Newton methods.

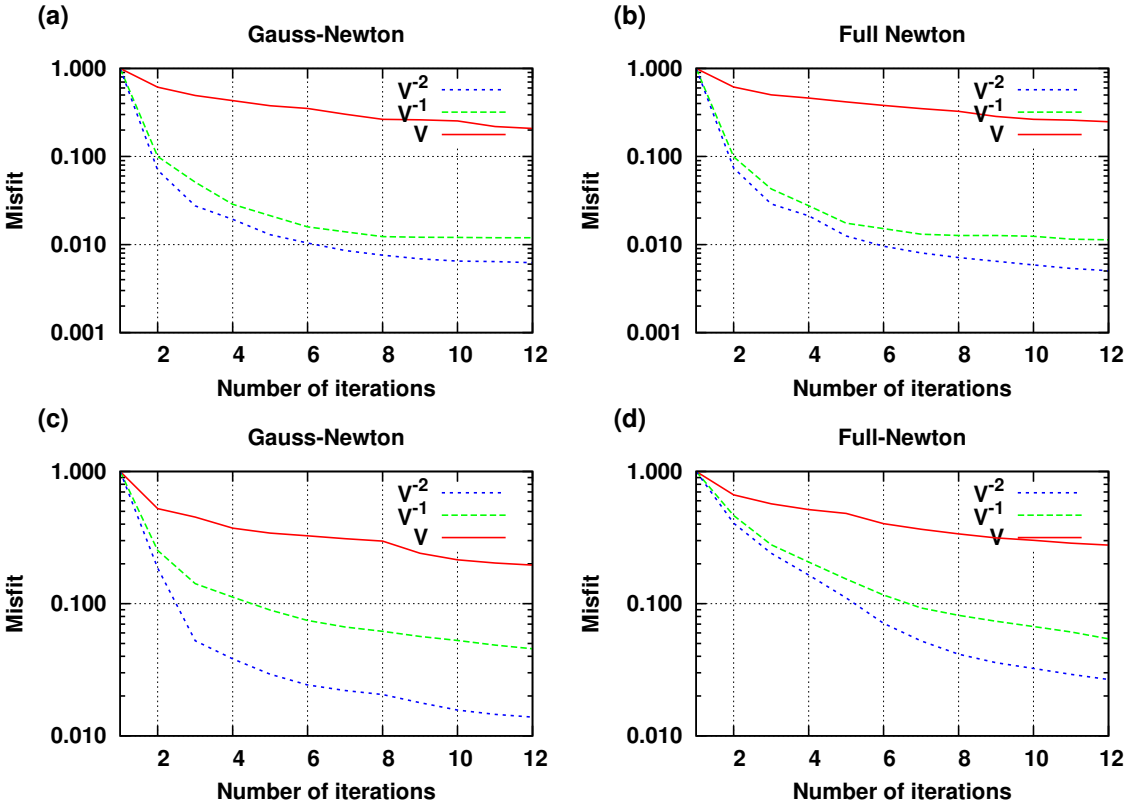


Figure 3.19: Data misfit reduction of the Marmousi velocity model for three types of model parametrizations: velocity (v), slowness (v^{-1}) and slowness squared (v^{-2}). On the left, from top to bottom are misfit functions for 4.70 Hz and 20.02 Hz datasets using Gauss-Newton method, and on the right from top to bottom are misfit functions using full Newton optimization method.

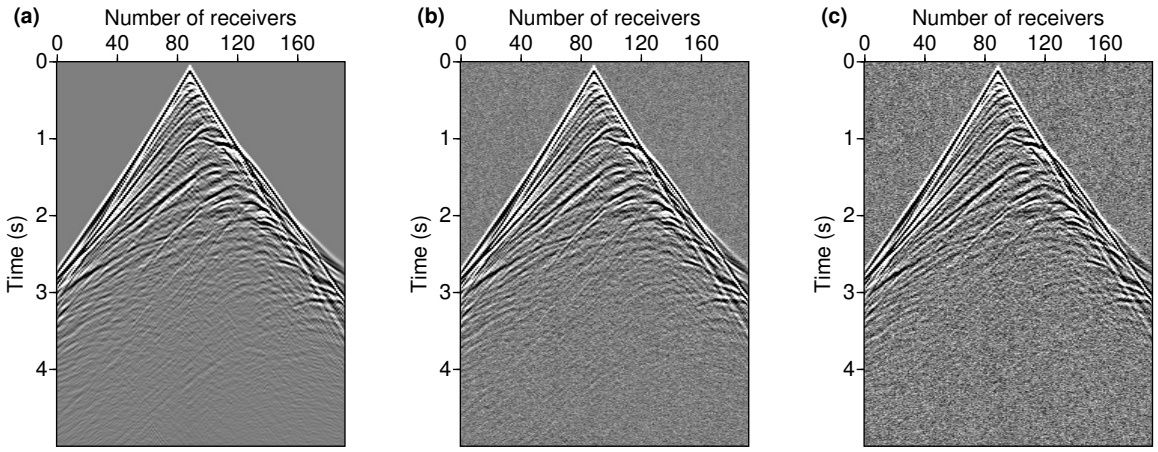


Figure 3.20: Synthetic seismogram data from Marmousi velocity model obtained by time-domain forward modeling from a single source located at (4.5 km,0.25 km). Clean data (a), noisy data with SNR=10 (b) and 5 (c).

3.3.6 Noisy dataset

Note that there would be lots of factors that might affect the efficiencies of the model parameterizations used for inversion such as elastic effects, AVO signature, surface waves, and incomplete physics are among a few to mention. Here we restrict ourselves to the case where the data is contaminated with noise. In this section, we investigate the effect of noisy data on velocity estimation using the Gauss-Newton and full Newton minimization methods for the three types model parameterizations. The noisy data is a time-domain synthetic dataset, which is contaminated by random noise with SNR=10 and SNR=5 as shown in Figure 3.20. Figure 3.20a, b and c are clean data, noisy data with SNR=10 and SNR=5, respectively. Figures 3.21 and 3.22 depict the reconstructed velocity models obtained by GN and FN methods from noisy data with SNR=10 and SNR=5, respectively. Figures 3.21a, b and c are reconstructed velocity models using GN method by velocity, slowness and slowness squared parameterization, respectively. On the other hand, Figure 3.21d, e and f are reconstructed velocity models using FN method by velocity, slowness and slowness squared parameterization, respectively; and the same goes for Figure 3.22.

Results in Figures 3.21 and 3.22 show that the numerical inversion by GN and FN methods reproduced models that are comparable to the original velocity model. As the noise in the data increases from SNR=10 to SNR=5, more artifacts are present in the deep parts of the model, in particular when the slowness squared parameterization is employed. Less artifacts are observed when the slowness parameterization is employed. This demonstrates that the slowness squared parameterization is more sensitive to noise than the slowness parameteriza-

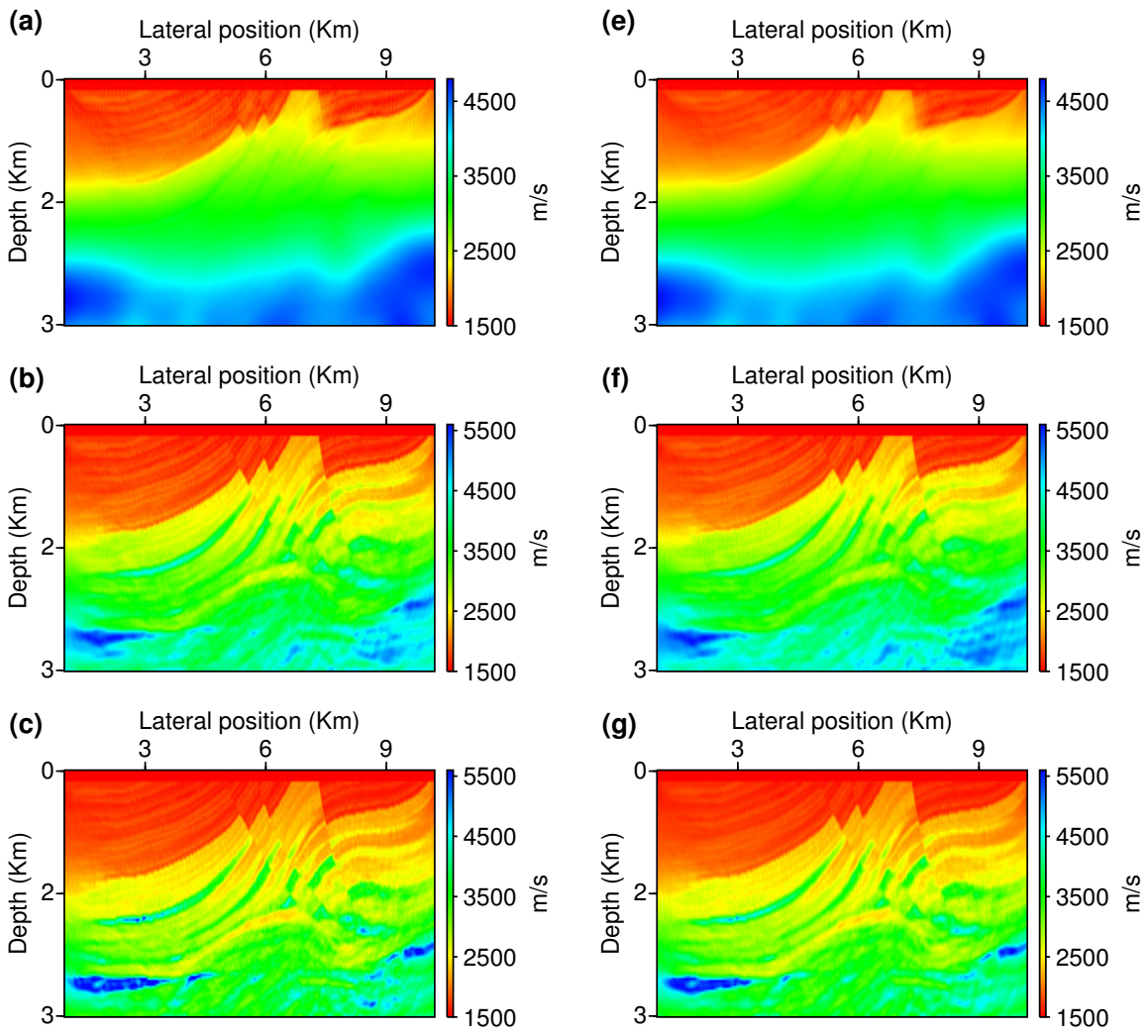


Figure 3.21: Gauss-Newton and full Newton FWI. Reconstructed velocity models after the end of all frequency components. From top to bottom are the reconstructed velocity models using velocity (v), slowness (v^{-1}) and slowness squared (v^{-2}) parameterizations, respectively. On the right from top to bottom (a), (b) and (c) are results obtained using the Gauss-Newton and on the left (d), (e) and (f) are results obtained using the full Newton method obtained from noisy data with SNR=10.

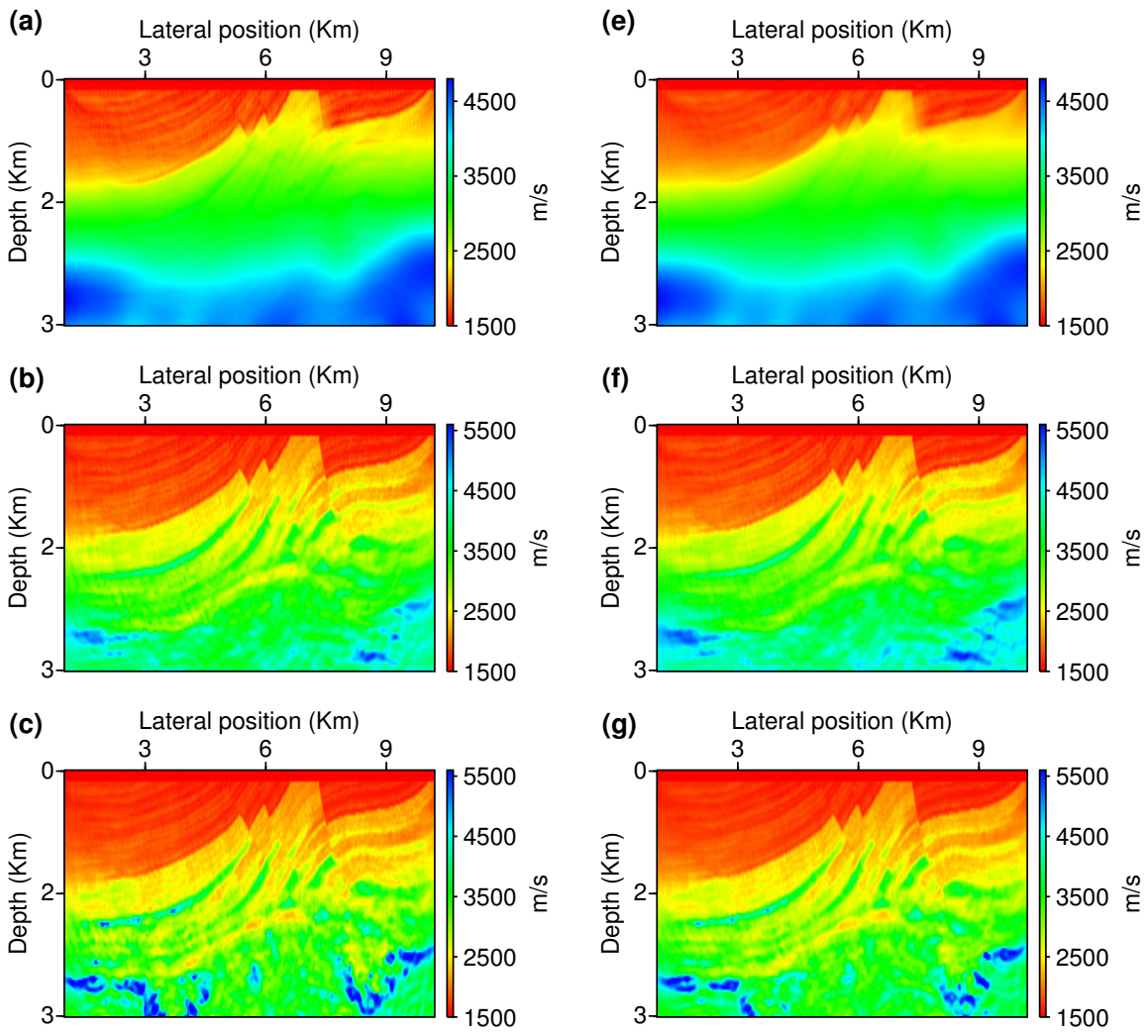


Figure 3.22: Gauss-Newton and full Newton FWI. Reconstructed velocity models after the end of all frequency components. From top to bottom are the reconstructed velocity models using velocity (v), slowness (v^{-1}) and slowness squared (v^{-2}) parameterizations, respectively. On the right from top to bottom (a), (b) and (c) are results obtained using the Gauss-Newton method and on the left (d), (e) and (f) are results obtained using the full Newton method obtained from noisy data with SNR=5.

tion. The eigenvalues of the Hessian matrix in the deep parts of the model are very sensitive to noise in the slowness squared parametrization than the velocity parametrization, where its eigenvectors are less important in the deep parts. For very noisy datasets, the velocity estimation that employs FWI is less sensitive to slowness parameterization than slowness squared parameterization. Better result in terms of the resolution and the quality of the inverted model can be achieved by using the slowness parameterization. More artifacts in the retrieved model are observed with slowness squared parameterization than the others. To quantify our analysis, we compute Q, a quality of the reconstruction, as our relative comparison metric

$$Q = 10\log_{10}[||\mathbf{m}_o||^2/||\mathbf{m}_o - \mathbf{m}||^2], \tag{3.24}$$

where $\mathbf{m}_o$ and $\mathbf{m}$ are the true velocity model and reconstructed velocity model, respectively. A high Q value corresponds to a more accurate solution (see Table 3.2). This image quality measurement is merely used to assess the relative quality of the reconstructed velocity models.

Table 3.2: Quality of the reconstructed Marmousi velocity model obtained from the three type model parameterization from a noisy dataset using Gauss-Newton (GN) and full Newton (FN) methods.

Model Parameter	SNR =10		SNR =5	
	GN	FN	GN	FN
Velocity	36.21	36.18	36.27	36.20
Slowness	39.48	38.61	38.53	38.36
Slowness squared	40.25	40.10	35.23	37.25

The relative image quality values in Table 3.2 demonstrate that for a high noisy dataset, the Newton-based full waveform inversion should be carried out using a slowness model parameterization. In the case of clean data, which is unlikely practise, a slowness squared parameterization is the best model parameterization to employ in the FWI optimization scheme in terms of quality and efficiency. Eigenvalues of the Hessian matrix are very sensitive to noise and a high signal to noise ratio is needed to retrieve reliable model parameters.

3.4 Summary

In this chapter, we examined the influence of model parameterization on velocity model building using gradient-based and Newton-based optimization methods. Three model parameterizations for the acoustic FWI were investigated. These include velocity, slowness and slowness squared. Numerical results demonstrate that the quality of the reconstructed ve-

locity model and convergence rate of, in particular, Newton methods are greatly affected by the type of model parameterization that one adopts. From the numerical results conducted with Marmousi velocity model, we found that the best convergence rate can be achieved by using the slowness squared as model parameterization. The latter is particularly true for datasets containing a moderate level of noise. However, for datesets contaminated by important levels of noise, we have found that our best results were found by parametrizing the problem with slowness.

In this thesis, unless otherwise stated, all the numerical results were obtained using slowness parametrization.

CHAPTER 4

Comparison of multi-frequency selection strategies for simultaneous-source full waveform inversion[1]

4.1 Introduction

Full Waveform inversion (FWI) is becoming part of our arsenal of methods for determining subsurface velocity models (Lailly, 1983; Tarantola, 1984b, 1987; Pratt et al., 1998; Virieux and Operto, 2009; Plessix et al., 2010; Vigh et al., 2011; Warner et al., 2013). The main objective of seismic waveform inversion is to retrieve the Earth model that best fits the observed seismic data. Although waveform inversion can provide accurate and high resolution geological models, multiple shortcomings confront its practical implementation for large scale problems. One of the main problems in FWI is the computational cost of the inversion for multiple sources and receivers. In addition, to avoid solutions that are trapped in local minima, one needs to provide a good initial velocity model or data with high-quality low frequency components (Brenders and Pratt, 2007; Virieux and Operto, 2009). FWI is also mathematically ill-posed and waveform inverse problems require regularization techniques in order to retrieve a unique solution by stabilizing the inverse problem (Tarantola, 1987; Virieux and Operto, 2009).

The computational cost of full waveform inversion is proportional to the number of sources in the experiment. This computational cost can be reduced by the use of simultaneous shooting techniques (Herrmann et al., 2009; Krebs et al., 2009; Ben-Hadj-Ali et al., 2011). The basic idea of simultaneous source methods is to generate super-shots by summing individual sources with a random encoding function (Romero et al., 2000). Recent work by van

[1]A version of this chapter has been published in GEOPHYSICS: Amsalu Y. Anagaw and Mauricio D. Sacchi (2014). Geophysics 79(5), R165-R181. doi: 10.1190/geo2013-0263.1

Leeuwen and Herrmann (2013) also show a reduction in computational cost based on a batch sampling method. This method uses different, randomly selected, sources at each iteration allowing a gradual increasing batch of sources. In simultaneous source full waveform inversion (SSFWI), multiple sources can be randomly encoded with a phase encoding function (Romero et al., 2000; Berkhout and Blacquiére, 2011) or with a time-shift function, where multiple sources are randomly encoded with random time-delays (Berkhout and Blacquiére, 2011). It is well known that simultaneous source techniques introduce random cross-talk that arises from the correlation between shots in the super-shot. One way to reduce the cross-talk noise is by generating new randomly encoded super-shots at every iteration (Krebs et al., 2009) or by smoothing the gradient along predefined directions (Guitton and Daz, 2012). In this thesis we extend the method by re-sampling source positions in every iteration during the minimization process to achieve a faster reduction of cross-talk artifacts. However, generating new random source positions and new blending operators in every iteration makes the inversion unstable because FWI becomes more sensitive to cross-talk noise (Bertsekas and Tsitsiklis, 1996). This, however, depends on the type of algorithm scheme with which FWI is carried out. Last, we point out that another major challenge of seismic imaging and inversion via source-encoding techniques is the sensitivity of the method to noisy data (van Leeuwen et al., 2011).

Full waveform inversion can be performed in either the time-domain (Lailly, 1983; Tarantola, 1984a; Gauthier et al., 1986; Mora, 1987) or the frequency-domain (Pratt and Worthington, 1990; Pratt, 1999; Operto et al., 2007; Stekl and Warner, 2007). From the standpoint of computationally efficiency, the frequency-domain implementation is preferable over a time-domain implementation because of its ability to reconstruct model parameters from limited discrete frequencies by taking advantage of the wavenumber redundancy in wide-aperture seismic data (Sirgue and Pratt, 2004). We adopt the frequency-domain FWI method. In the frequency-domain, a finite set of discrete frequencies is selected and the inversion is carried out in a sequential approach starting from low frequency to high frequency data components (Sirgue and Pratt, 2004). The model parameters obtained from the inversion of the low frequency data component are used as starting model parameters for the higher frequency data component.

The main scope of the chapter is to analyze the effect of different frequency selection strategies for FWI on datasets that are encoded using random time-delay shifts. A few discrete frequencies were selected based on the work by Sirgue and Pratt (2004) on wide-aperture seismic data. A common multiscale frequency-domain inversion technique is the sequential inversion of single frequencies. In addition to the sequential method, we created a subset of simultaneously inverted frequencies from the selected set of finite frequencies, which are called group frequencies. A total of five possible scenarios of group frequencies were chosen for our tests. The scenarios include the discrete frequency strategy in which a single

frequency dataset is inverted at time of inversion. For simultaneous frequencies, the first one is the individual group method in which selected temporal frequencies were grouped from low to high frequencies sequentially without overlapping frequencies across the groups (Kim et al., 2011). The second method is Bunks' method (Bunks et al., 1995). The Bunks' multiscale method was initially proposed for time-domain waveform inversion. This method is similar to the first grouping method (Kim et al., 2011), however, in this case there is an overlap of frequencies across groups. The number of frequencies per group increases, and each higher frequency group contains higher frequencies as well as all frequencies in the lower group. In time-domain waveform inversion, the inversion is carried out from low frequency bands using low-pass filtered data to higher frequency bands. The third method is the partial overlap method, where only a few frequencies are overlapped partially across the groups. We proposed and considered both one-frequency and two frequency overlaps. More overlapping of frequencies across the group can also be considered The inversion obtained from the first group of frequencies is used to initialize the inversion for the next group. The last method is the simultaneous frequency approach where all frequency data components are inverted simultaneously.

In addition to using a diversity of multi-frequency selection strategies, we propose a strategy for further reduction in cross-talk artifacts by regenerating new encoding and source positions at every iteration. We also examined the use of a second-order optimization algorithm that adopts the Gauss-Newton method and compare it to a first-order optimization method such as the conjugate gradient method with the source encoding strategy. We demonstrate that a faster convergence rate is achieved by using a second-order optimization method. In the regularized Gauss-Newton (GN) algorithm that is adopted in this chapter, we found that re-sampling both source positions and encoding functions at every iteration are key elements for obtaining high resolution images (faster cross-talk reduction) and for a faster convergence rate.

The chapter is outlined as follows. First we show the formulation and implementation of a matrix-free Gauss-Newton method for FWI that both uses conventional shot-gathers and simultaneous sources. In the present study, we build a matrix-free Gauss-Newton full waveform algorithm based on the acoustic wave-equation. The matrix-free Gauss-Newton method is formulated using the second-order Lagrangian formulation. The gradient of the objective function is computed efficiently via the adjoint state method (Plessix, 2006). At each iteration and frequency group, a new random time-shift encoding operator is generated to mitigate the effect of cross-talk artifacts in the blended image. We also demonstrate the efficiency of utilizing new source positions and encoding functions at every iteration of the inversion scheme. We then present numerical results with comparisons of different simultaneous multi-frequencies strategies and their ability to attenuate cross-talk. Though all frequency selection strategies proposed provide reasonable results, the partial overlapping

group method show better results in terms of resolution and quality of the inverted model. Numerical results from using the Marmousi and the BP/EAGE velocity models (Billette and Brandsberg-Dahl, 2005) are presented to illustrate the merits and shortcomings of the strategies for simultaneous shot techniques for full waveform inversion. We also examine the effects of different source density in the generation of the super-shots that are used by our FWI algorithm. Finally, we address the noise sensitivity problem.

4.2 Simultaneous-source full waveform inversion (SS-FWI)

For simultaneous sources full waveform inversion, we considered the weighted objective function

$$J(\mathbf{m}) = \frac{1}{2} \sum_{\omega_j \in \Omega_i}^{N_w} W_j \sum_{k,r}^{N_{ss},N_r} ||\mathbf{\Psi d}(\omega_j)^{obs} - s(\omega_j)\mathbf{rp}(\omega_j)_{k,r}^{cal})||^2 + \mu R(\mathbf{m}). \tag{4.1}$$

In this chapter, we restrict ourselves to an inversion scheme in which sources are randomly encoded with random time-delays (we denote them "blended sources"). The blended sources in frequency-domain is expressed in a simple linear relationship as

$$\mathbf{S}(\omega)_k = \Gamma(\omega, \tau) \odot \mathbf{D_R} src(\omega) = \mathbf{\Psi} src(\omega), \tag{4.2}$$

and the nonzeros elements of $\Gamma(\omega, \tau)$ are expressed in a vector form as

$$\Gamma(\omega, \tau)_{nonzeros} = [e^{-i\omega\tau_1} \quad e^{-i\omega\tau_2} \quad e^{-i\omega\tau_3} \quad ... \quad e^{-i\omega\tau_{ns}}]^T, \tag{4.3}$$

where $\tau_i \in [0, t]$ with a maximum time-delay t.

For better performance in simultaneous-source technique, which will be described later in detail in this chapter, we used the second-order optimization method: the Gauss-Newton method (see equations 3.2 & 3.4 in Chapter 3 or 2 for details). As I described in Chapter 2, the formulation of the Gauss-Newton optimization method for the simultaneous-source technique is similar to that of the conventional inversion method.

For example, the forward modeling is formulated as

$$\mathbf{p}_k^{cal}(\omega) = \mathbf{A}(\mathbf{m}, \omega)^{-1} \mathbf{\Psi} src(\omega), \tag{4.4}$$

and the gradient for single super-shot and frequency and is give by

$$\begin{aligned}\mathbf{g} =& \Re e\Big\{\Big\langle \mathbf{A_m}(\mathbf{m},\omega)\mathbf{A}(\mathbf{m},\omega)^{-1}\mathbf{\Psi} src(\omega), \mathbf{A}(\mathbf{m},\omega)^{-1}W_i\mathbf{r}^{\dagger}\sum_{r}^{N_r}(\mathbf{\Psi d}(\omega_j)^{obs} - s(\omega_j)\mathbf{rp}(\omega_j)_r^{cal})\Big\rangle_{\mathbf{x}}\Big\} \\ &+ \mu\nabla_{\mathbf{m}}R(\mathbf{m}).\end{aligned} \tag{4.5}$$

In the same way the Hessian matrix will be reconstructed.

4.2.1 Practical implementations of simultaneous-sources and encoding functions in FWI

Algorithm 5 shows the pseudo-code of an inexact or truncated Gauss-Newton method where the CG iterations are terminated early. Similarly to Chapter 3, the action of the Hessian matrix on a vector is computed on the fly at every conjugate gradient iteration. In the Gauss-Newton method, we adopt an early termination of the conjugate gradient iteration whenever a negative curvature corresponding to negative eigenvalues of the Hessian matrix is detected or if the condition $|\mathbf{m}^T{}_k\mathbf{H}\mathbf{m}_k| < \varepsilon||\mathbf{m}_k||^2$ is satisfied for small ε (Fasano and Lucidi, 2009). Note that, the choice of a stopping criteria has an important influence on the quality of the reconstructed model parameters and on the convergence rate of the algorithm (Eisenstat et al., 1994; Métivier et al., 2013). A detailed study about the influence of the stopping criteria can be found in Métivier et al. (2013).

In our algorithm we have three options to generate the blending operators and source positions. This algorithm allows us to select random source positions and blending operators simultaneously or separately at different stages. As shown in the pseudo-code of the GN method, the first option is to generate the random source positions and blending functions at the beginning, and keep them throughout the FWI iterations. The second option is to regenerate random blending and/or the source positions at every new frequency or frequency group. The third option is to generate both the blending and/or the source position operators at every iteration and frequency or frequency group. A combination of all these methods can also be used. For example, we can randomly choose and keep source positions at the beginning of the FWI process, and randomly blend the sources throughout the inversion at each iteration.

Algorithm 5: Pseudo-code: Multiscale Gauss-Newton method. Notations: W_{iw} :- data weight, nw :- number of frequencies within a group frequency nwg, $\mathbf{m}$ is the model parameter, Γ and D_R are the encoding and source position operators, respectively.

Input : $\mathbf{d}^{obs}(\omega)$, $\mathbf{m}_0$, N_w, max_iter
Output : $\arg\min_{\mathbf{m}} J(\mathbf{m})$
Initialize parameters;
Option 1: Generate random operators $\rightarrow \Gamma, D_R$
for $igr \leftarrow 1$ **to** $ngw \leftarrow$ *over group frequency* **do**
  Starting model $\leftarrow \mathbf{m}_k$;
  Option 2: Generate random operators $\rightarrow \Gamma, D_R$
  for $k \leftarrow 1$ **to** max_iter **do**
    Option 3: Generate random operators $\rightarrow \Gamma, D_R$
    Compute:
    forward wavefields $\mathbf{d}^{cal}(\omega)$;
    `residual wavefields` $\triangle\mathbf{d}(\omega)$;
    `back-propagate residual wavefields` $\lambda(\omega)$;
    gradient $\nabla_{\mathbf{m}} J(\mathbf{m}_k) = \sum_{iw}^{nw} W_{iw} \nabla_{\mathbf{m}} J(\mathbf{m}_k)_{iw}$;
    Solve:
    $\triangle\mathbf{m}_k = -\Big[\sum_{iw}^{nw} (\mathbf{H}_{GN})_{iw}\Big]^{-1} \nabla_{\mathbf{m}} J(\mathbf{m}_k)$;
    solve $\triangle\mathbf{m}_k$ using CG method : call Algorithm 2;
    Two forward modeling simulations are required to compute;
    the action of the Hessian matrix on $\triangle\mathbf{m}$ inside the CG algorithm;
    Compute:
    α using line search method;
    Update model:
    $\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha\triangle\mathbf{m}_k$;
  end
end

4.2.2 Frequency selection strategy

In this section, we briefly review a strategy for choosing optimal discrete frequencies for frequency-domain full waveform inversion proposed by (Clayton and Stolt, 1981; Sirgue and Pratt, 2004). The strategy for choosing optimal frequencies is based on continuous wavenumber coverage of vertical wavenumbers (see Figure 4.1). The strategy is based on homogeneous background with one flat reflector and wide-aperture geometry. The idea of frequency strategy for seismic imaging was initial proposed by Clayton and Stolt (1981). The key idea is to take advantage of the wavenumber redundancy provided by maximum offset data in the acquisition from which only few discrete frequencies are chosen, which in turn provides a reliable image (Clayton and Stolt, 1981; Sirgue and Pratt, 2004). The formulations of obtaining discrete frequencies is through the resolution analysis of the gradient

of the objective function by replacing the Green's function with plane wave approximation in reference homogeneous background velocity with one reflector, assuming that the amplitude effects may be ignored and we are in far field regions. The basic idea is that for a single frequency, the contribution of a single source-receiver pair has only a single wavenumber component, and a range of continues vertical wavenumber components of the velocity model can be modeled by using a range of source-receiver pairs. For a given range of offsets, a single frequency (f_n) span an area $[k_{zmin}, k_{zmax}]$ in the wavenumber domain. The minimum and maximum wavenumbers are defined as

$$k_{zmin}(f_n) = 4\pi f_n \frac{\alpha_{min}}{c_o} \tag{4.6}$$

$$k_{zmax}(f_n) = 4\pi f_n \frac{1}{c_o}, \tag{4.7}$$

where c_o is the constant background velocity. The parameter α_{min} depends on cosine reflection angle of a wavefield propagating with maximum incident angle and is given by

$$\alpha_{min} = \frac{1}{\sqrt{1 + (h_{max}/z_{max})^2}}, \tag{4.8}$$

where h_{max} is the maximum source-receiver half-offset and z_{max} is maximum depth target to be imaged.

In the multiscale hierarchical frequency inversion strategy, the frequency selection strategy must ensure a continuous wavenumber illumination in which the maximum wavenumber of the frequency used at the current frequency inversion must corresponds to the minimum wavenumber of the frequency for the next inversion. Figure 4.1 is the schematic diagram that illustrates the frequency-domain optimal frequencies selection strategies for full waveform inversion proposed by Sirgue and Pratt (2004). Therefore, from the wavenumber continuity, we have

$$k_{zmin}(f_{n+1}) = k_{zmax}(f_n) \tag{4.9}$$

To ensure continuity, the next frequency, f_{n+1}, to be selected is given

$$f_{n+1} = \frac{f_n}{\alpha_{min}} \tag{4.10}$$

As it illustrated on the graph (see Figure 4.1), the frequency increment ($\triangle f_{n+1}$, $\triangle f_{n+2}$, ...) across the wavenumber spectrum increases with frequency. Thus, it results in a significant

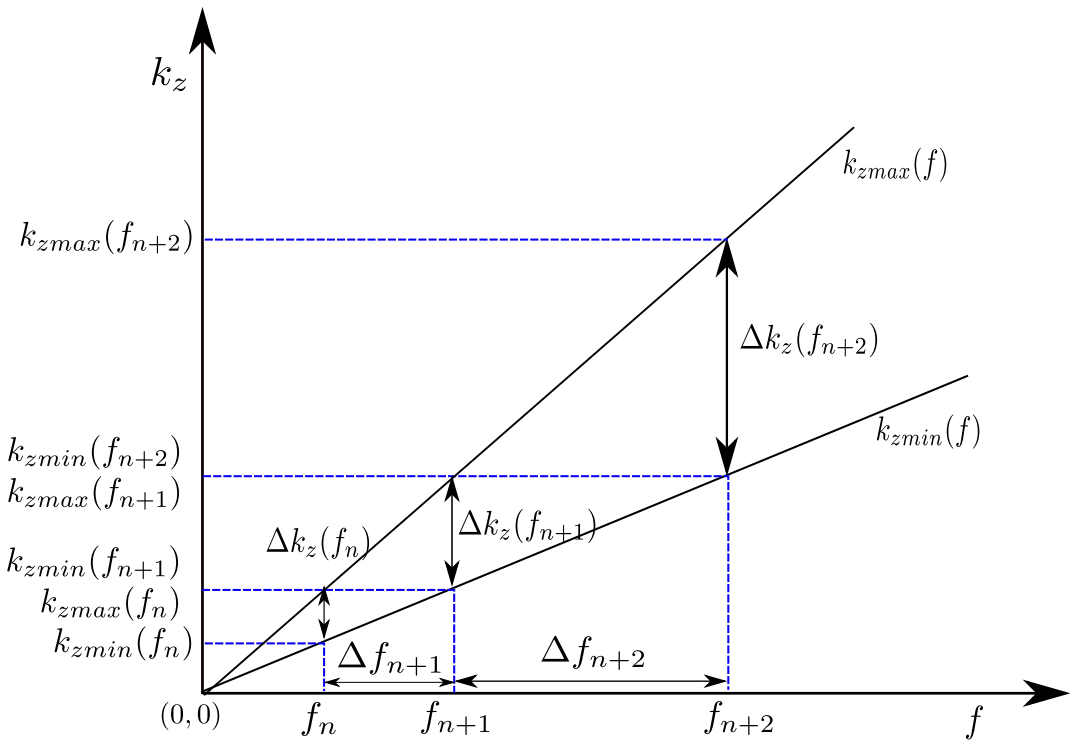


Figure 4.1: Frequency-domain optimal frequencies selection strategies for full waveform inversion proposed by Sirgue and Pratt (2004).

reduction in the computational time due to the reduction of the number of frequency used for inversion. The strategy only relies on 1D problem with full range offsets available but still a good approximation to extend for 2D or 3D problems.

4.3 Results and discussions

In the numerical inversion algorithm for our frequency selection strategies, we conducted a series of examples with the complex velocity models Marmousi, Marmousi-II (Bourgeois, A., M. Bourget, P. Lailly, M. Poulet, P. Ricarte, and R. Versteeg, 1991) and the BP velocity model (Billette and Brandsberg-Dahl, 2005).

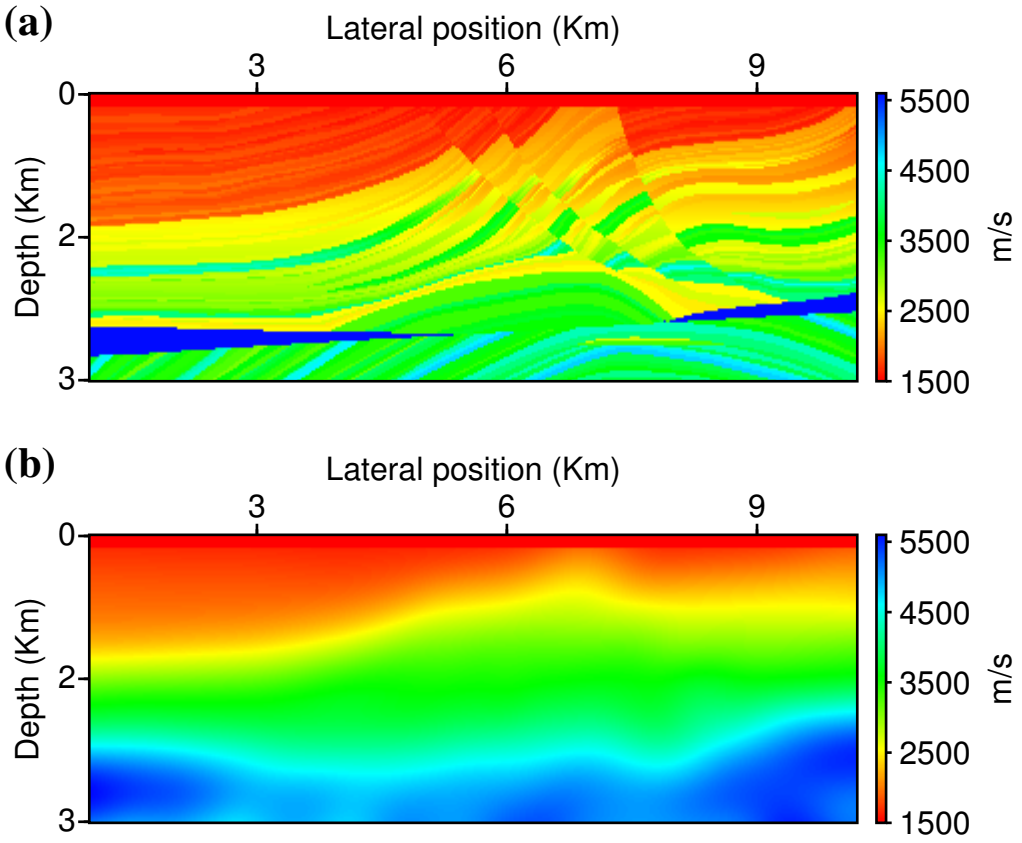


Figure 4.2: Marmousi velocity model (a) and smooth velocity model used as starting model for inversion (b).

4.3.1 Numerical examples on Marmousi velocity model

First, we consider the Marmousi velocity model, Figure 4.2a, discretized on a 126×384 grid with a grid spacing for both vertical and horizontal directions of 24 m. Figure 4.2b is a smooth velocity model used as a starting model to run the FWI algorithm. The starting velocity model is obtained by smoothing the true velocity model with a 2D Hanning window. Generally, a good starting velocity model is crucial to ensure convergence to a minimum near the global minimum.

First, synthetic data were generated from 96 sources and 192 receivers using the time-domain wave-equation. The sources are placed 24 m below the surface, and receivers are regularly placed at the surface. For the source function, we used a Ricker wavelet with a 10 Hz central frequency. We recalibrate the source wavelet signature at every iteration

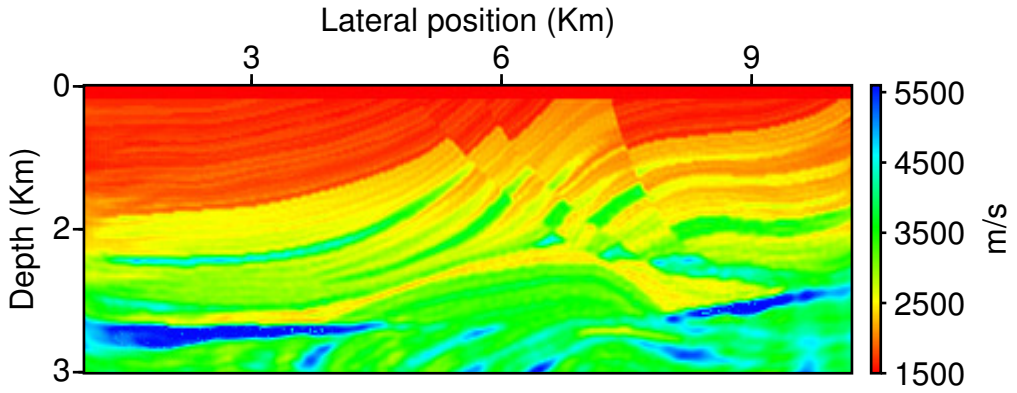


Figure 4.3: Reconstructed Marmousi velocity model obtained using the GN method with the conventional shot gather method (no super-shots). For this inversion, 5 partially overlapped frequency bands are used.

throughout the inversion (Pratt et al., 1998). For the inversion tests, a set of 12 discrete frequencies were selected between 3.0 Hz and 22 Hz based on the strategy proposed by Sirgue and Pratt (2004). The set of frequencies is (3.05, 3.54, 4.27, 5.13, 6.10, 7.20, 8.67, 10.25, 12.33, 14.65, 17.46, 20.87 Hz). Details of discrete frequency selection criteria for 1D and an approximation for 2D can be found in Sirgue and Pratt (2004). The most common method used in frequency-domain FWI with a finite set of frequencies is the discrete sequential frequency method, in which the waveform inversion is carried out sequentially from low to high frequency data. As the inversion is carried out each time for a single frequency, the final updated model of the lower frequency is used as the starting model for the inversion of the next higher frequency data (Sirgue and Pratt, 2004). The nonlinear relation between the model parameters and the data is mitigated by proceeding the inversion sequentially from low to high frequency data.

In order to test the full waveform inversion algorithm, we first consider conventional FWI using a smooth starting velocity model. Figure 4.3 is the result obtained using the Gauss-Newton method using 5 partially overlapping frequency bands with two frequency overlapped (see Table 4.1). For each frequency band, we run up to a maximum of 30 Gauss-Newton iterations with a stopping criteria $\varepsilon = 10^{-3}$ and a regularization parameter $\mu = 5.0\times10^{-4}$. As shown in Figure 4.3, the inverted velocity model obtained using the Gauss-Newton method provides an accurate velocity model. All features of the original Marmousi velocity model are reconstructed in the correct position.

For simultaneous-source inversion, we used different source density. Each super-shot is constructed from randomly encoded monochromatic sources with random time-shift delay.

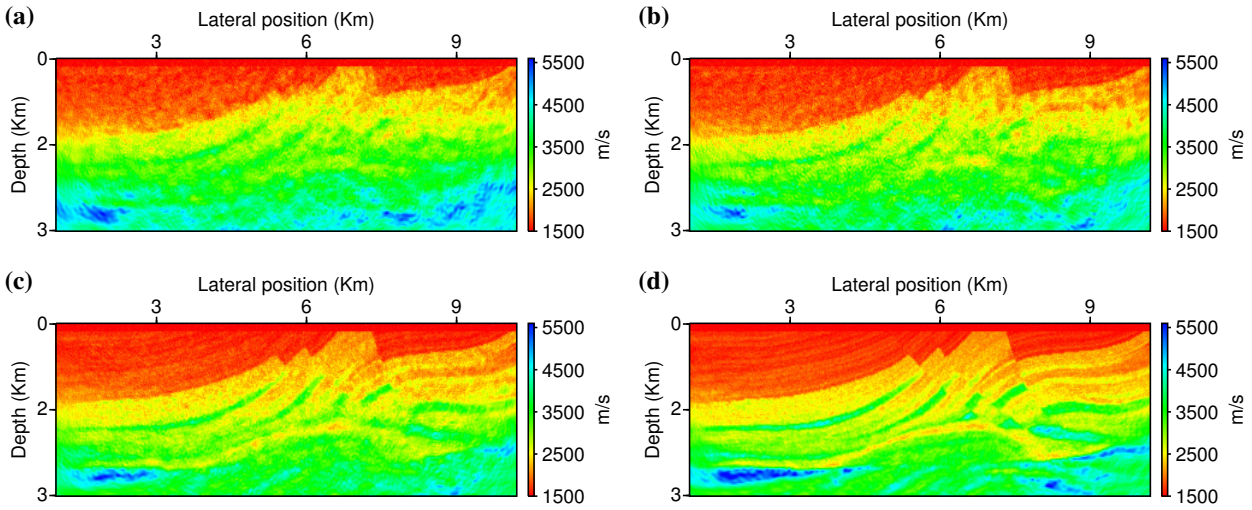


Figure 4.4: Reconstructed Marmousi velocity model using simultaneous sources that are encoded with a time-shift technique. Results showing a sequential single discrete frequencies method using 2 (a), 3 (b), 6 (c) and 12 (d) super-shots . Maximum time-delay is 2.0 sec.

We used 2, 3, 6 and 12 encoded super-shots, where each encoded source consisted of 48, 32, 16 and 8 individual monochromatic sources, respectively. We used a maximum time-delay of 2.0s.

One of the fundamental issues regarding the case of simultaneous shots for FWI is whether it can be possible to estimate an acceptable reconstructed velocity model given the level of cross-talk noise that arises from the interference of sources in the inverted model. In order to analyze results obtained from the simultaneous-shot FWI, first, the quality of the reconstructed velocity model is investigated for different number of sources within a single super-shot. Figure 4.4 depicts the reconstructed Marmousi velocity model using different numbers of encoded monochromatic sources from a smooth starting velocity model with a sequential single discrete frequency method. Figure 4.4 [a]-[d] shows the reconstructed velocity models for different encoded sources density. For inversion, for the 6 and 12 encoded super-shots, we compute up to a maximum of 40 Gauss-Newton iterations with $\varepsilon = 5.0{\times}10^{-3}$ and $\mu = 10^{-3}$, and for the 2 and 3 encoded super-shots we choose a maximum of 50 GN iterations with $\varepsilon = 10^{-2}$ and $\mu = 5.0{\times}10^{-2}$. As depicted in Figure 4.4a, there are high cross-talk artifacts in the reconstructed velocity model with two super-shots in which each super-shot was obtained by encoding 48 individual sources. This cross-talk artifacts arise from the correlation among the high number sources within the encoded sources (Berkhout et al., 2008; Berkhout and Blacquiére, 2011). When the number of sources within the

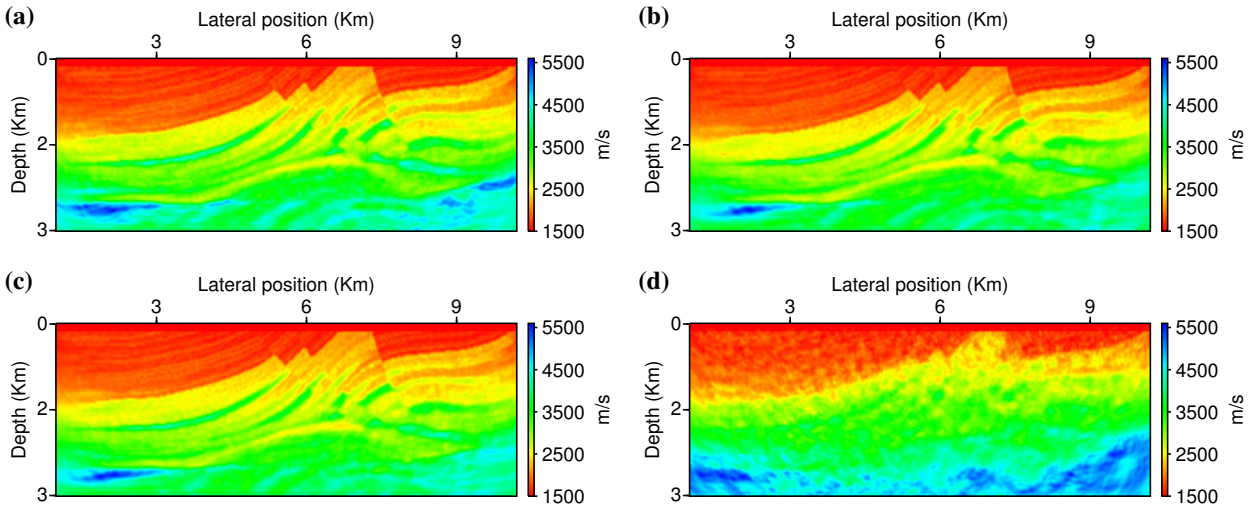


Figure 4.5: Reconstructed Marmousi velocity model using two super-shots where each super-shot contains 48 shots. Source sampling and time-shift encoding function are generated at each iteration (a), new random time-shifts encoding function are generated at each iteration while keeping the source sampling positions (b), new source sampling positions are generated at each iteration while keeping the time-shift encoding function kept constant throughout the inversion (c), and when both source sampling positions and time-shift encoding function generated are kept constant at each iteration (d).

super-shot decrease, the quality of the reconstructed image improves and cross-talk noise reduces drastically (see Figure 4.4a-Figure 4.4d). In particular, the result obtained using 8 individual shots per super-shots as shown in Figure 4.4d is more comparable to the result obtained using the conventional shot gathers but at a lower computational time. The higher the number of sources that is packed within a single super-shot, the higher the cross-talk noise and the lower the quality of the reconstructed image. In fact Krebs et al. (2009) have shown that in source encoding strategy, cross-talk noise present in the final retrieved model is proportional to the square of sources in a super-shot. One way to improve the quality of the reconstructed velocity model is to regenerate new blending encoding functions at every iteration. In addition, re-sampling source positions at every iteration further suppresses artifacts.

Figure 4.5 shows the reconstructed Marmousi velocity model using two super-shots (Figure 4.4a). Figure 4.5a is the reconstructed model when the source sampling and time-shift encoding functions are generated at every iteration, Figure 4.5b is the result when new random time-shifts encoding functions are generated at each iteration while keeping the source sampling positions, Figure 4.5c corresponds to the case where new source sampling

positions are generated at each iteration while keeping the time-shift encoding function kept constant through out the inversion, and finally Figure 4.5d corresponds to the case where both source sampling positions and time-shift encoding function are kept constant. All these images show the importance of generating new blending or encoding functions and source sampling positions operators at every iteration. As these results show, generating either source positions or encoding function improves the final retrieved velocity model.

Table 4.1: Frequency selection strategies for FWI.

Method	Frequency group
Discrete sequential[2]	f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9, f_{10}, f_{11}, f_{12}
Group	f_1, f_2, f_3, f_4 f_5, f_6, f_7, f_8 f_9, f_{10}, f_{11}, f_{12}
Partial overlap group (1)	f_1, f_2, f_3, f_4 f_4, f_5, f_6, f_7 f_7, f_8, f_9, f_{10} f_{10}, f_{11}, f_{12}
Partial overlap group (2)	f_1, f_2, f_3, f_4 f_3, f_4, f_5, f_6 f_5, f_6, f_7, f_8 f_7, f_8, f_9, f_{10} f_9, f_{10}, f_{11}, f_{12}
Bunks	f_1, f_2, f_3, f_4 f_1, f_2, f_3, f_4,f_5, f_6 f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8 f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9, f_{10} f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9, f_{10}, f_{11}, f_{12}
Simultaneous[3]	f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9, f_7, f_8, f_9, f_{10}, f_{11}, f_{12}

[2]One frequency at a time is inverted
[3]All frequencies are simultaneously inverted

We first chose a total of six possible scenarios of multi-frequency simultaneous frequency inversion strategies for FWI. These multi-frequency simultaneous frequency strategies are outlined in Table 4.1. Note that in multi-frequency simultaneous inversion, the frequency bands for each frequency group can be increased and a higher number of frequencies can be added. However, in this thesis, we are interested and focused on the strategies of simultaneous frequency inversion and their effect on cross-talk artifacts. In multiscale FWI, the velocity model obtained from the lower group frequency is used as a starting velocity

model for the next higher frequency group. All frequencies within one group are inverted simultaneously. Data weight has also been implemented within a group. The data weight during the simultaneous frequency group inversion is chosen in such a way that slightly more weight is given to the lower frequencies

$$W_i = exp(-\frac{\omega_i - \omega_o}{\omega_{max} - \omega_o}\beta), \tag{4.11}$$

where ω_o and ω_{max} are the lowest and highest frequencies within the multi-frequency group. The parameter β is a positive value chosen by the user. In this chapter the value of β ranges from 2 to 5. Note that a different weighting function can also be chosen. The key is to give a relatively less weight to the higher frequencies so that the higher frequencies do not dominate the lower frequencies at the early stages of the numerical inversion.

When several frequencies are inverted simultaneously, the lower frequency data components were given more weight than higher frequency data components. This is to guarantee that higher frequency data components do not dominate the lower frequency. Otherwise, the inversion will end up with artifacts and sometimes it will also fail to converge to the global minimum.

To test the multi-frequency selection strategies, first, we consider few number of individual sources within a super-shot. A total of 12 super-shots are used, with each super-shot composed of 8 individual shots. For each frequency group, we compute up to a maximum of 40 Gauss-Newton iterations. We also use a stopping criteria, $\varepsilon = 5.0\times10^{-3}$, for solving the approximate solutions of the Gauss-Newton equation and a regularization parameter $\mu = 10^{-3}$.

Figure 4.6 depicts the reconstructed velocity models from a noisy free dataset using six possible different multi-frequency selection methods. It is clear from our results that the numerical inversion was able to reproduce models that are comparable to the original velocity model. All the features and structures of the Marmousi model are properly reconstructed. Some artifacts that arise due to the cross-talk artifacts in the discrete sequential frequencies method as shown in Figure 4.6a have been attenuated when several data frequencies are inverted simultaneously (see Figure 4.6b-4.6f). The deep part of the velocity model is determined to be properly reconstructed when a simultaneous multi-frequency strategy is employed. Though all frequency selection strategies provide reasonable results, the individual and partial overlap grouping methods demonstrate better results in terms of resolution and quality of the inverted model. In order to compare the quality and the robustness of the frequency selection strategies, we compute Q, a quality of the reconstruction, as our comparison metric,

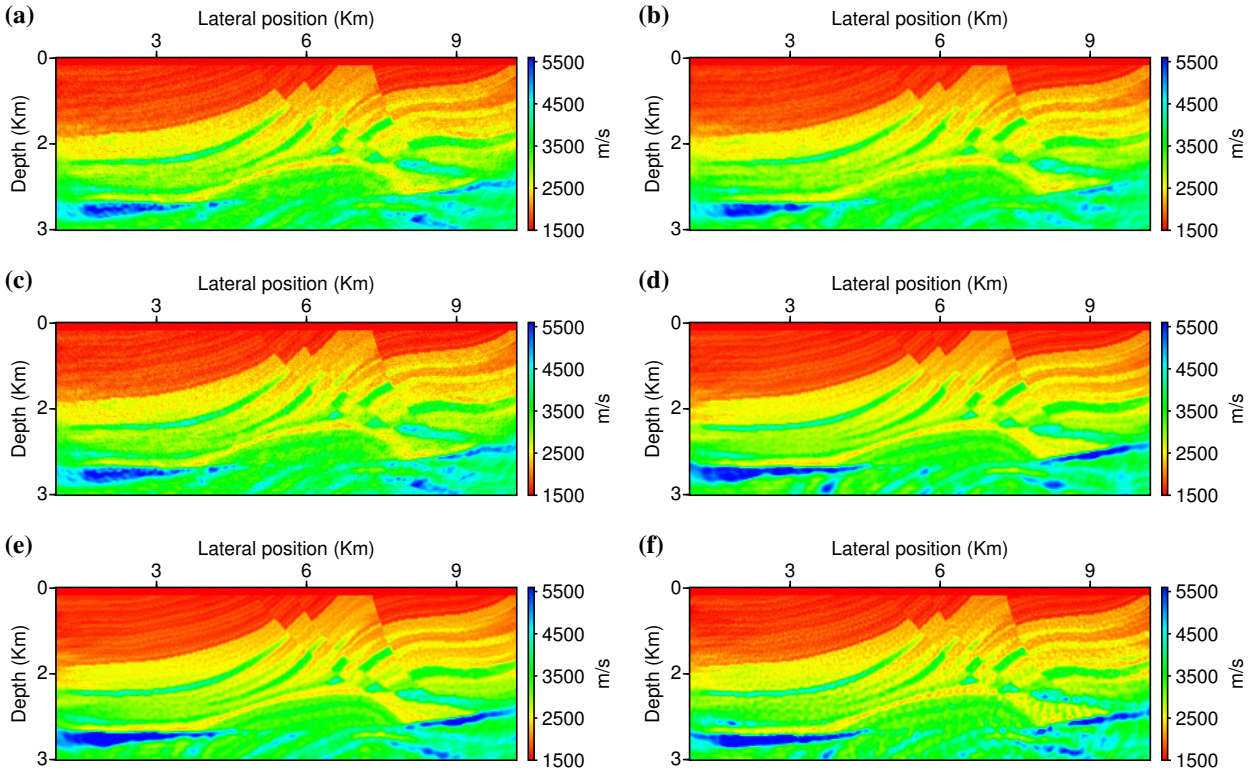


Figure 4.6: Simultaneous sources FWI using 12 super-shots with each contains 8 individual monochromatic sources. Reconstructed velocity model using sequential single discrete frequencies method (a), individual group (b), partial overlap group method with one frequency overlapped (c), partial overlap group method with two frequencies overlapped (d), the Bunks' method (e) and simultaneous method (f).

$$Q = 10\log_{10}[||\mathbf{m}_o||^2/||\mathbf{m}_o - \mathbf{m}||^2], \tag{4.12}$$

where $\mathbf{m}_o$ and $\mathbf{m}$ are the true velocity model and reconstructed velocity model, respectively. A high Q value corresponds to a more accurate solution. The image quality measurements of the reconstructed velocity models for a clean dataset in Figure 4.6 is shown in Table 4.2.

Table 4.2: Quality of the reconstructed Marmousi velocity model (Q) for clean data.

Method	Clean Data
	Q
Discrete sequential	40.09
Group	40.37
Partial overlap group (1)	40.97
Partial overlap group (2)	42.73
Bunks	42.36
Simultaneous	40.36

4.3.2 Numerical example: Noisy dataset

Next we use a noisy dataset that is contaminated with SNR=10. First a time-domain seismogram was generated and noise with SNR=10 was added (see Figure 4.7). Unlike the previous example, here we use 3 super-shots with each super-shot containing 32 sources. For inversion, for each frequency group, we run up to a maximum of 60 Gauss-Newton iterations. New source sampling position and encoding functions are generated at every iteration.

The quality of the final reconstructed models and the computational costs are assessed by adapting and monitoring the change in the image quality measurement for each frequency group such that all the final models are equivalent in quality. We define Q_k as an average image quality at k^{th} iteration for the $k-2$, $k-1$ and k^{th} iterations for each frequency group. We should emphasis that the final image quality measurement is influenced by the cross-talk noise. Due to the influence of the cross-talk noise in the final model, two different models could have similar image quality measurements. As a result, for a fair comparison among the final models obtained by different multi-frequency selection strategies, we take the average of the last three iteration image quality measurements. For each frequency group, we run the inversion until $Q_k < 38.5$ is satisfied. For solving the approximated solution of the Gauss-Newton equation, we used a stopping criteria $\varepsilon = 0.05$ or stop iterating when a negative curvature is detected and regularization parameter $\mu = 0.1$. Often, the inner loop runs up to a maximum of 15 iterations.

Figure 4.8 depicts the reconstructed velocity models from noisy dataset with three super-shots. With some degree of low quality, our numerical inversion still reproduced models

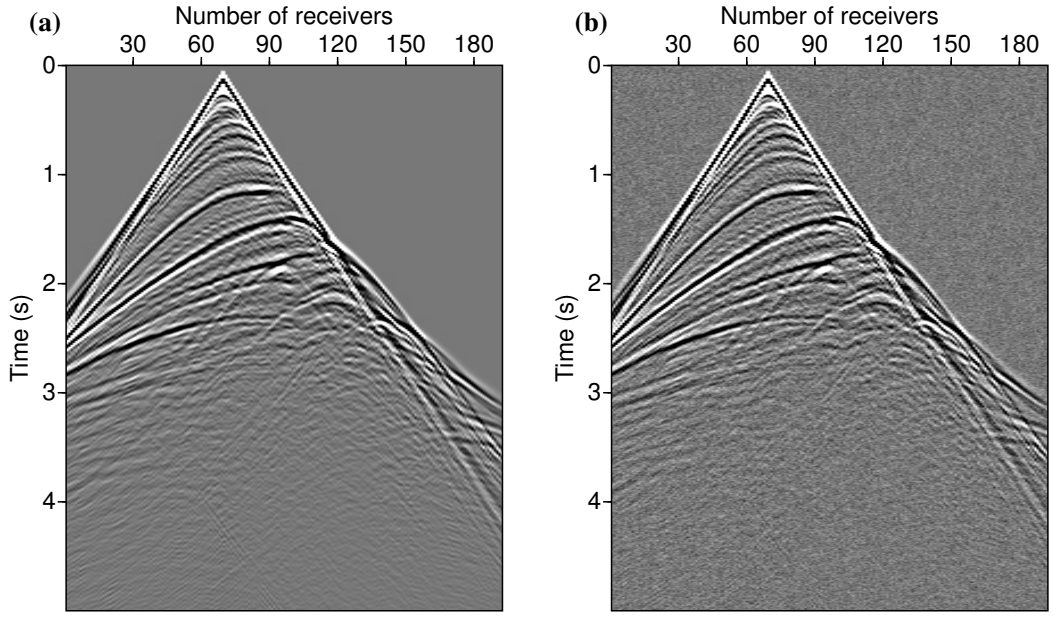


Figure 4.7: Seismogram obtained from true Marmousi model at 3.36Km (a) and noisy data contaminated with noise; SNR =10 (b).

that are as acceptable as to the one obtained with the clean dataset and 12 super-shots. However, in the noisy dataset, we can see that the Bunks and partial overlap with two frequencies grouping methods outperform the other methods in terms of resolution and cross-talk attenuation; see Figures 4.8d and 4.8e. This can be confirmed by computing the model residual between the true model and the results obtained with different multi-frequency inversion methods.

Figure 4.9 shows the difference between the true Marmousi model and the reconstructed velocity models in Figure 4.8. The simultaneous inversion method is least correct and is the poorest method for attenuating the cross-talk noise. On the other hand, we see there is more residual left over when the discrete, individual group and one-frequency partial overlap group methods are used. The residual for the top 2km, in the Bunks' and two frequencies partial overlap methods are almost zero. These methods once again demonstrate better results in terms of resolution and quality of the inverted model would be achieved by using the Bunks' and two frequencies partial overlap grouping methods. Table 4.3 shows the total number of forward modeling performed to achieve the same average image quality at each frequency group for the noisy dataset for different multi-frequency strategies. In terms computational cost, the Bunks' method is the most time consuming method. In Bunks' method, the number of forward modeling simulations that are required to achieve the final

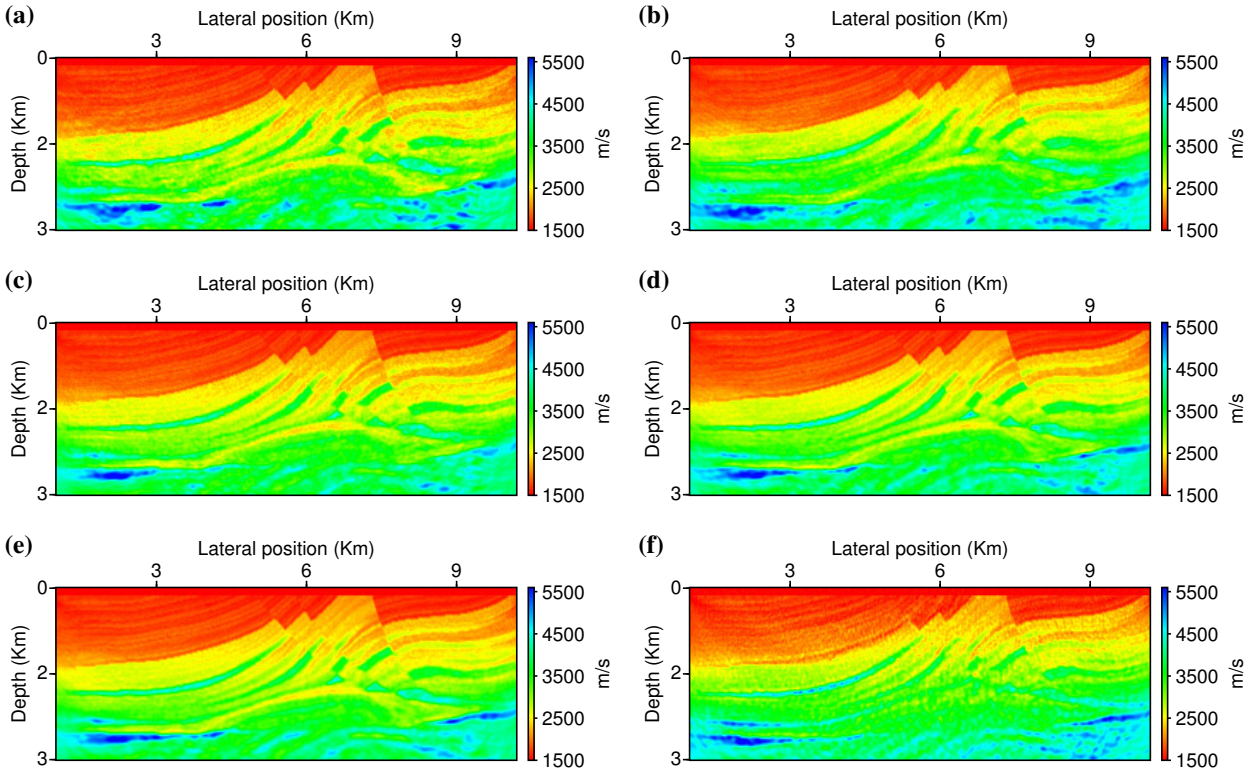


Figure 4.8: Simultaneous sources FWI using 3 super-shots with each contains 48 individual monochromatic sources; from noisy data with SNR =10. Reconstructed velocity model using sequential single discrete frequencies method (a), individual group (b), partial overlap group method with one frequency overlapped (c), partial overlap group method with two frequencies overlapped (d), the Bunks' method (e) and simultaneous method (f).

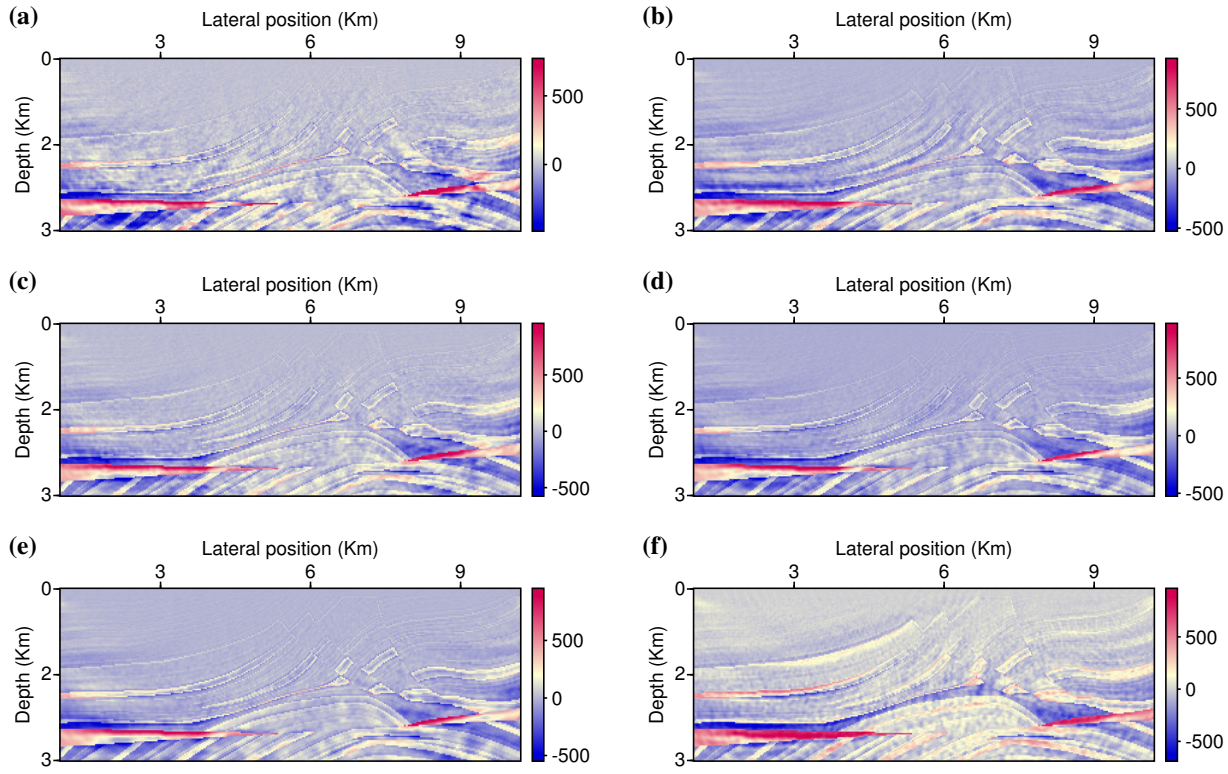


Figure 4.9: Difference between the true Marmousi model and the reconstructed velocity models in Figure 4.8; residual from the velocity model using sequential single discrete frequencies method (a), individual group (b), partial overlap group method with one frequency overlapped (c), partial overlap group method with two frequencies overlapped (d), the Bunks' method (e) and simultaneous method (f).

model is almost twice as the total number of simulations for the other methods with the exception of the simultaneous frequency inversion strategy. The partial overlap method with two frequencies overlapped across the frequency group provides an improved image with a visible attenuation of cross-talk artifacts. For this example, we also found that the convergence rate of the partial overlap method is faster than the rate of convergence of Bunks' method.

4.3.3 Marmousi II velocity model

In the second numerical example, we consider the portion of the Marmousi II velocity model shown in Figure 4.10a. Figure 4.10b is the starting velocity model used for inversion, which is obtained by smoothing the true velocity model with a 2D Hanning window. The original

model is 16km long and 3.5km deep. We decimated the original size onto a coarser grid size 25 m by 25 m and took the right side of the model up to 12.2km and 3.2km depth, which leads to a model of size 126 x 481. First, we generated time-domain seismograms from 120 sources and 240 receivers using a time-domain finite-difference code, and noise with SNR=10 is added. To generate the time-domain seismograms, we used both density and velocity models where as for inversion we only used the velocity model. The sources are placed 25 m below the surface, and receivers are placed at the surface, equidistantly. For the velocity model building, once again a set of 12 discrete frequencies were selected between 3.0 Hz and 21 Hz based on the strategy proposed by Sirgue and Pratt (2004). The set of frequencies is (2.93, 3.58, 4.23, 5.05, 6.02, 7.16, 8.46, 10.09, 11.88, 14.16, 16.93, 20.02Hz). The frequency dataset can be easily selected by using the Discrete Fourier transform.

For our inversion tests, we used the same number of super-shots as in the previous example, 3 super-shots with each super-shot contains 40 individual monochromatic sources that are encoded with random time-shift. For the GN full waveform inversion, for each frequency group, we run the inversion until $Q_k < 50.0$ is satisfied. For solving the approximated solution of the Gauss-Newton equation, we used a stopping criteria $\varepsilon = 0.05$ or stop solving the Gauss-Newton equation when a negative curvature is detected. For the regularization parameter $\mu = 0.01$. The selected 12 frequency dataset were sampled and grouped according to the outlined multi-frequency strategies in Table 4.1.

Figure 4.11 depicts the reconstructed velocity models from noisy dataset using 3 super-shots. The numerical inversion of all methods except the simultaneous method reproduced models that are less cross-talk artifacts. All the features and structures of the Marmousi-II model are properly reconstructed. The simultaneous method performs poorly. Some artifacts that arise in the lower deep parts when using the discrete, individual group and one frequency partial overlap methods (Figure 4.11a), have been attenuated when Bunks' method and the two frequency overlapping methods are used. The deep part on the velocity model is determined to be properly reconstructed when a simultaneous multi-frequency strategy is

Table 4.3: Total number of forward modeling performed to reach within approximately the same quality of the reconstructed of the reconstructed Marmousi velocity model at each frequency group inversion for noisy data with SNR =10.

Method	Noisy Data: SNR =10
	Number of forward modeling
Discrete sequential	4102
Group	3786
Partial overlap group (1)	3890
Partial overlap group (2)	4004
Bunks	8600
Simultaneous	7680

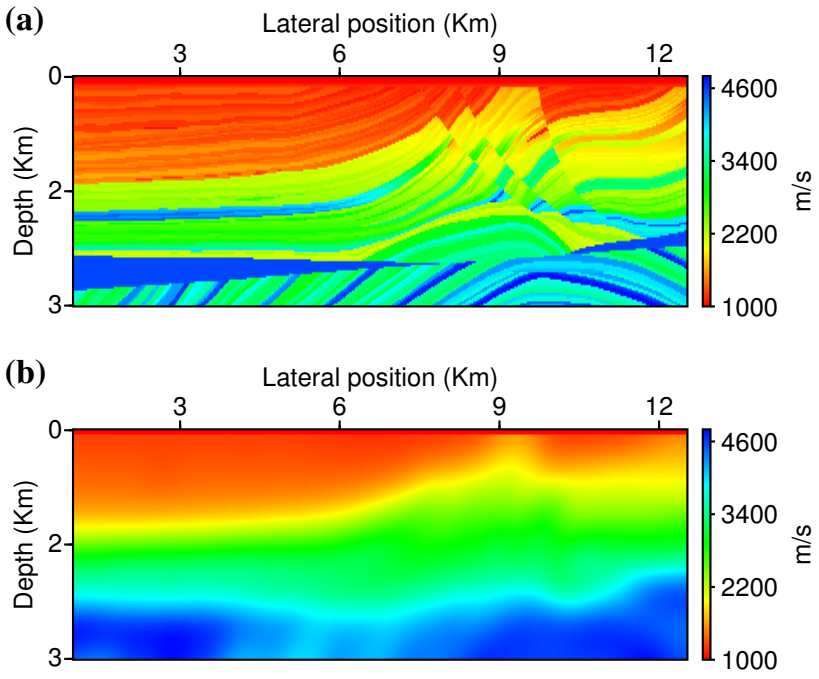


Figure 4.10: Portion Marmousi-II velocity model (a) and smooth velocity model used as starting model for inversion (b).

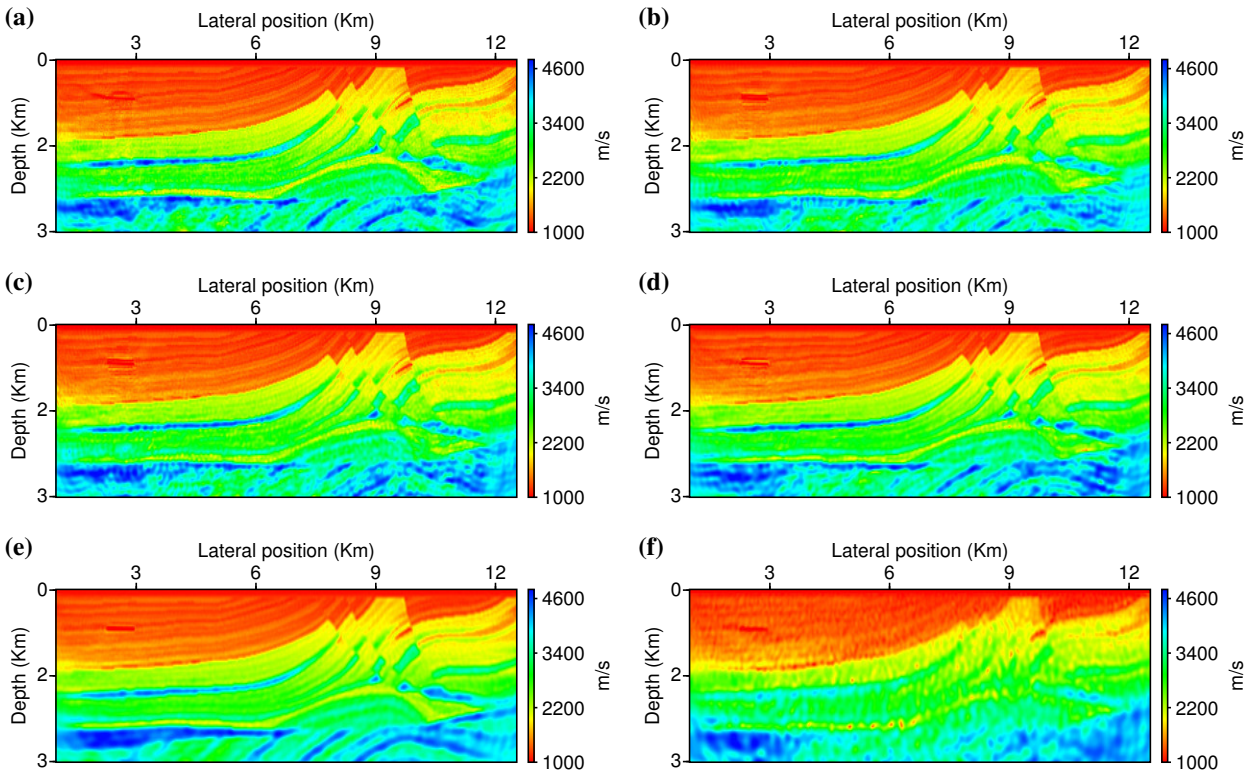


Figure 4.11: Simultaneous sources FWI using 3 super-shots with each contains 40 individual monochromatic sources. Reconstructed velocity model using sequential single discrete frequencies method (a), individual group (b), partial overlap group method with one frequency overlapped (c), partial overlap group method with two frequencies overlapped (d), the Bunks' method (e) and simultaneous method (f).

employed. This demonstrates, once again, that the importance of using Bunks' method and partial overlap groupings to obtain high resolution and high quality models. For this particular example, it is clear that these methods are superior in terms of attenuating cross-talk artifacts.

Figures 4.12a-4.12f depict the difference between the true Marmousi-II model and the reconstructed velocity models in Figure 4.11. Model residuals confirm that all methods except the partial overlap grouping (when two frequencies are overlapped) Figure 4.12d and Bunks' method (Figure 4.12f) have artifacts left over that are not attenuated. Table 4.4 shows the total number of forward modeling performed to achieve the same average image quality at each frequency group for the noisy dataset for different multi-frequency strategies. In terms computational cost, once again, the Bunks' method is the most time consuming method. In Bunks' method, the number of forward modeling required to achieve the final model is more than twice as the other methods required except the simultaneous method. The partial overlap method with two frequencies overlapped across the frequency group provides a high resolution image and attenuates cross-talk artifacts at a faster convergence rate than Bunks' method.

Table 4.4: Total number of forward modeling performed to reach within approximately the same quality of the reconstructed of the reconstructed Marmousi-II velocity model at each frequency group inversion for noisy data with SNR =10.

Method	Noisy Data: SNR =10
	Number of forward modeling
Discrete sequential	5038
Group	4937
Partial overlap group (1)	5356
Partial overlap group (2)	5832
Bunks	13918
Simultaneous	8201

4.3.4 Sampling frequencies

Frequencies selection strategies are important, in particular, if the inversion is carried out from a very few number of frequencies. Because each frequency has a limited and finite band contribution to the reconstructed model parameters, a continuity in wavenumber coverage is crucial for accurate model parameters building (Sirgue and Pratt, 2004). In order to better understand the choices of frequency selection, we choose a set of 7 discrete frequencies from approximately 3.0 Hz to 21 Hz; (a) using Sirgue and Pratt (2004) approach with little modification; say choice-1 (2.93, 4.07, 5.70, 7.81, 10.74, 14.81, 20.35 Hz) and (b) using equally spaced interval with 2.91 Hz ; say choice-2 (2.93, 5.84, 8.75, 11.66, 14.57, 17.48,

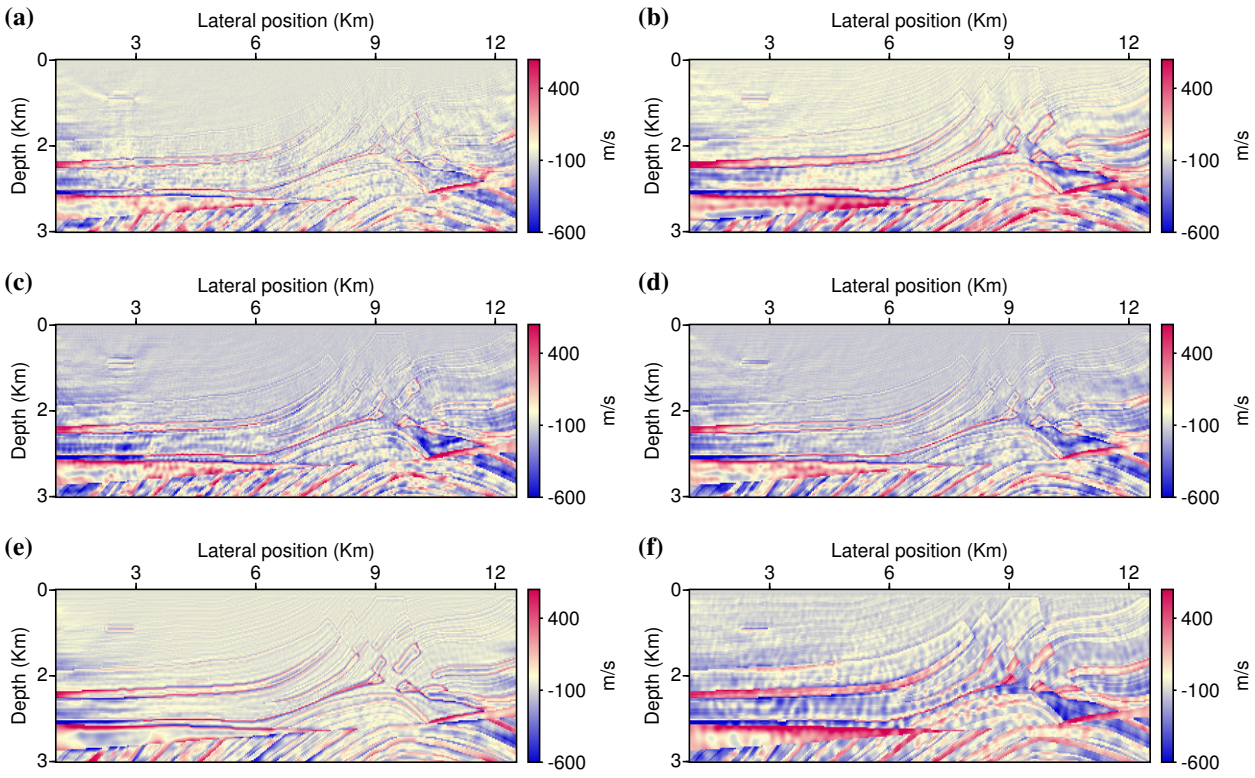


Figure 4.12: Difference between the true Marmousi-2 model and the reconstructed velocity models in Figure 4.11; residual from the velocity model using sequential single discrete frequencies method (a), individual group (b), partial overlap group method with one frequency overlapped (c), partial overlap group method with two frequencies overlapped (d), the Bunks' method (e) and simultaneous method (f).

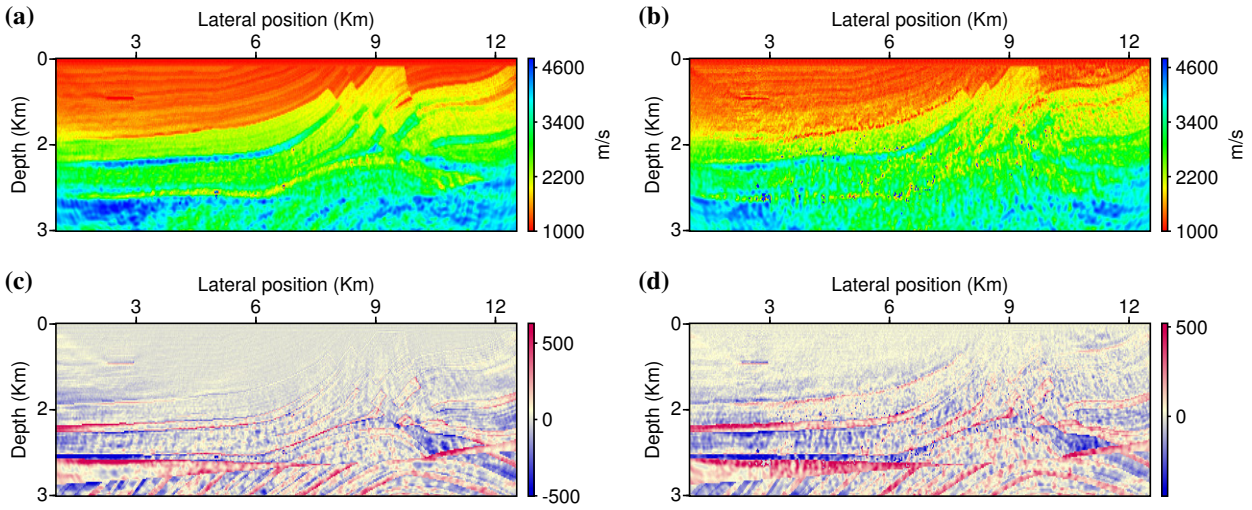


Figure 4.13: Reconstructed Marmousi-II velocity model using the selected frequencies (a) and equally spaced frequencies (b). Residual model between (a) and the true model (c), and (d) is the difference between (b) and the true model.

20.39Hz). Unlike the previous cases in which 12 frequencies are used, here we restrict ourselves to few number of frequencies with a reason which will be clear in a moment. For our inversion tests, we use the noisy dataset. We then apply the same strategies and procedures that were adopted in previous examples. We employed the partial overlapping technique with two overlap frequencies method.

Figure 4.13a and 4.13b are the reconstructed models using the selected frequencies of choice-1 and equally spaced frequencies of choice-2, respectively. Figure 4.13c and 4.13d are the difference between the true model and the model obtained using choice-1 and choice-2 frequencies, respectively. As we see from these images, the equally spaced frequencies provide very poor quality image. This due to the lack of continuity in the wavenumber coverage, which is a key in order to recover the velocity model accurately (Sirgue and Pratt, 2004). Whereas the selections of choice-1 type frequencies are based on the approximation of the continuity of the wavenumbers (Sirgue and Pratt, 2004), and as a result better and more accurate result would be achieved. On the other hand, we do not see a big difference in the results between the two choices if a large number of frequencies are selected. For example, from our frequency choices, we see that more than 12 frequencies lead to a good continuity of the wavenumber. However, as a large number of frequencies are selected, there will be a high chance of data redundancies, which contributes to a higher computational cost. It is interesting to note that even the conventional FWI without source encoding will result with similar wraparound artifacts when the temporal frequencies are improperly sampled.

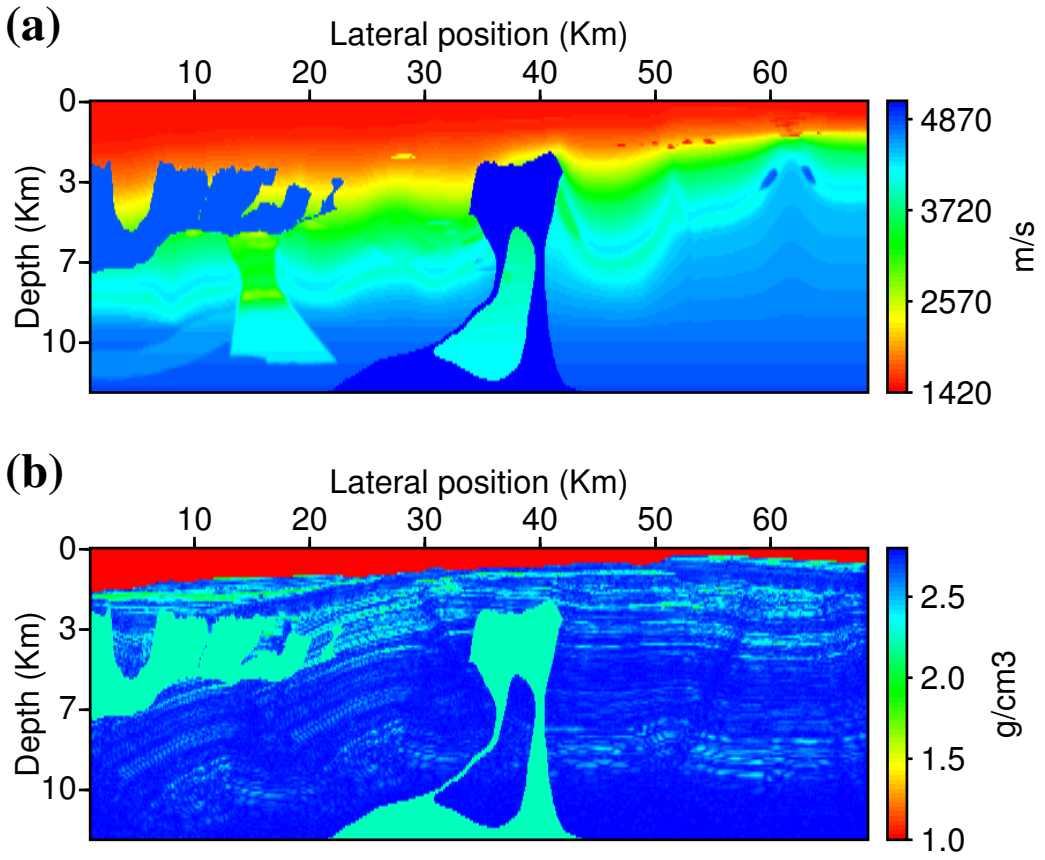


Figure 4.14: BP/EAGE velocity model (a) and density (b).

4.3.5 BP/EAGE velocity model

In the last numerical example, we consider the original BP/EAGE velocity model (Billette and Brandsberg-Dahl, 2005). The original BP/EAGE velocity model is 67km long and 12km deep. It was built on a 6.25m x 6.25m grid size with 5395 by 1911 model parameters. The BP velocity model was then regridded onto a coarser grid size 50 m by 50 m, which in turn leads to a model of size 140 x 1350. Therefore there are 140x1350 unknown parameters for the inversion. The original velocity model contains salt domes of high velocity contrast, subsalt low velocity zones under salt and anomalies. Figure 4.14a shows the true BP/EAGE velocity model and Figure 4.14b is density model, respectively.

A synthetic time-domain seismogram from 113 sources and 338 receivers using a time-domain finite-difference code was generated and a noise with SNR=10 is added. We used

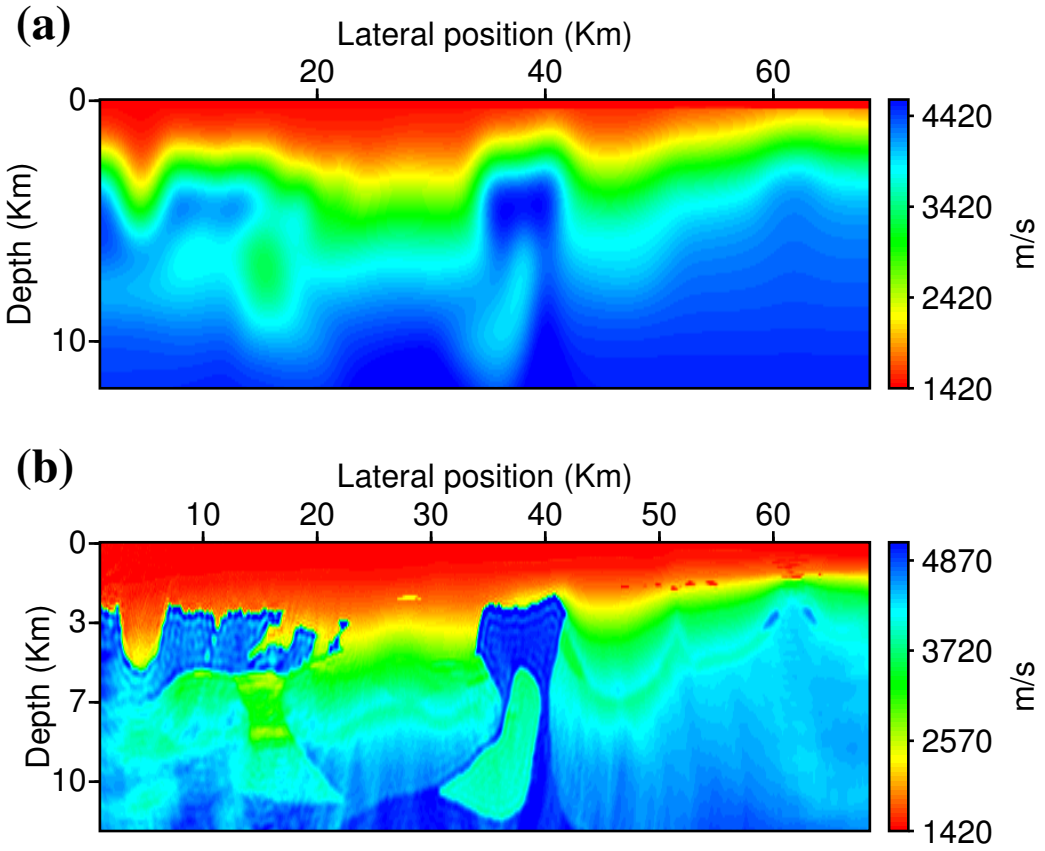


Figure 4.15: BP/EAGE smooth velocity model used as starting model for inversion (a) and reconstructed velocity model using the conventional shot-gather dataset (b).

both the density and velocity model to generate the seismogram. The sources and receivers are placed at the surface equidistantly. For inversion, a set of 12 discrete frequencies were selected from approximately 1.0Hz to 15.0Hz; (1.05, 1.41, 1.95, 2.48, 3.08, 3.81, 4.73, 5.85, 7.25, 9.00, 11.15, 13.82Hz). For this model due to the coarser grid size used for the inversion, in order to recover the long wavelength components of the velocity model, one has to start the inversion with very low frequency data. The low frequency data components for BP velocity model inversion have also been used by Shin and Ho Cha (2009) and Hu et al. (2011).

For the FWI, we first consider the conventional multiscale discrete frequency sequential approach using the individual shot-gather technique. Figure 4.15a is the smooth velocity

model used for initiating FWI and Figure 4.15b is the reconstructed velocity model. The numerical inversion reproduces a model that is comparable to the original velocity model. The salt dome body structures, the low velocity zones under the salt and the shallow anomalies are properly reconstructed. We then examine and explore the effect of different frequency selection strategies to the quality of the final reconstructed model parameter and their performance in terms of attenuating the cross-talk artifacts. A total of three super-shots are used with each super-shot contains a maximum of 40 individual monochromatic sources that are encoded with random time-shift with a maximum time-delay of 2.0 seconds. In our numerical tests, for each frequency group, we run the inversion until $Q_k < 55.0$ is satisfied at the k^{th} iteration. For the inner loop of the GN algorithm, we used a stopping criteria with $\varepsilon = 0.01$ or we stop solving the Gauss-Newton equation whenever a negative curvature is detected. The regularization parameter is $\mu = 0.05$. The selected 12 frequency datasets were sampled and grouped as shown in Table 4.1.

Figures 4.16a-4.16f are the reconstructed velocity models using simultaneous encoded source techniques based on different frequency selection methods. As depicted in the figure, all the frequency selection strategies, except the simultaneous frequency inversion, provide acceptable and higher resolution images with respect to the starting model. Among the different frequency selection strategies utilized in our study, the Bunks' method and the partial overlap grouping method with two frequencies yield the best final results in terms of resolution and quality of the inverted model.

Table 4.5 shows the total number of forward modeling performed to achieve the same average image quality at each frequency group for the noisy dataset for different multi-frequency strategies. Similar pattern is observed in terms of the total numbers of forward modeling performed to achieve the same Q_k at each frequency group. However, a careful visual examination of the quality and resolution of the final reconstructed velocity models is what it matters (see Figures 4.16a-4.16f). The partial overlap method with two frequencies overlapped across the frequency group provides better resolution image and attenuate cross-talk artifacts at faster convergence rate compared to the Bunks' method.

Figure 4.17a- 4.17f depict the difference between the true BP salt model and the reconstructed velocity models in Figure 4.16. Once again the Bunks' and the partial overlap grouping methods have least artifacts left over, see Figure 4.17d and 4.17e.

The inversion utilizing encoded-sources provides the model which was in closest agreement to the one obtained using the conventional shot gather inversion method, see Figure 4.15b and 4.16a-4.16f. While results in Figure 4.16a-4.16e are encouraging, the example with a simultaneous inversion of all frequency method at a time is struggling to recover the velocity model. This method was also having difficulties with the other two models investigated in this chapter. Overall, all other frequency strategy techniques for FWI in frequency-domain

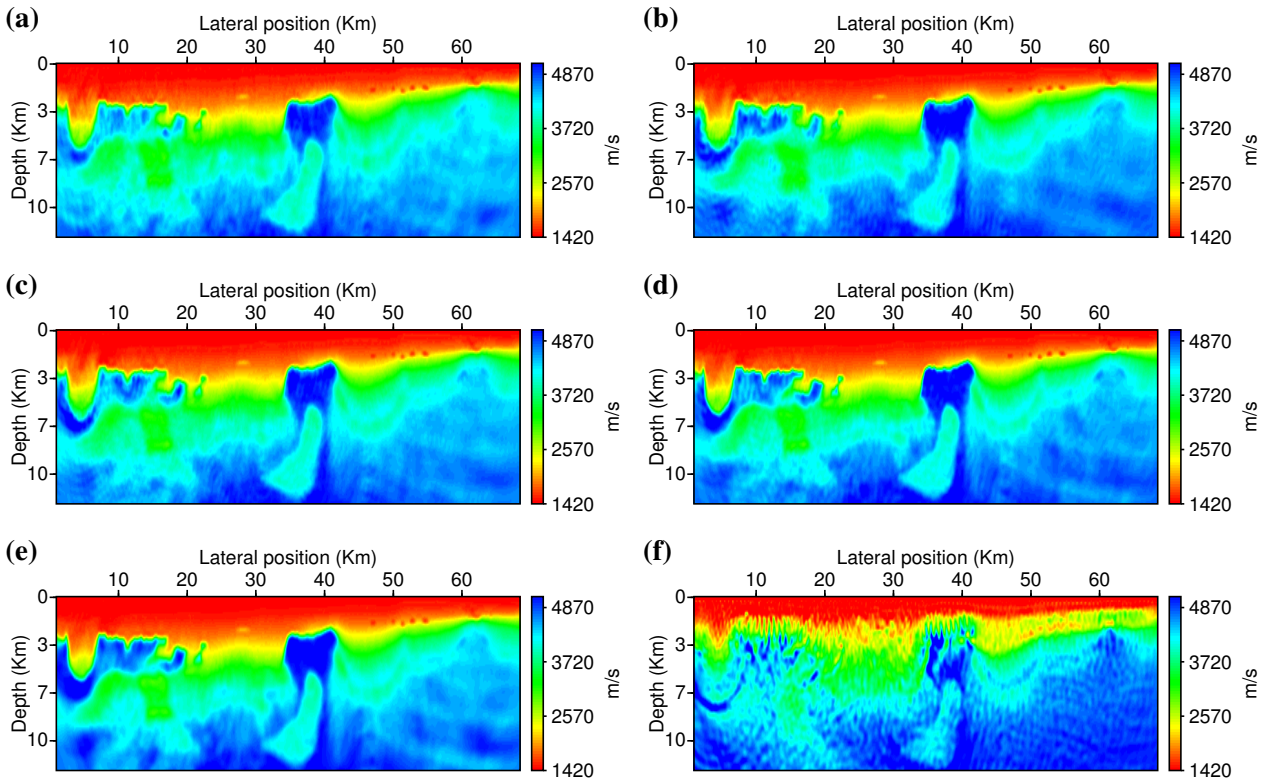


Figure 4.16: Simultaneous sources FWI using 3 super-shots with each contains {40,40,33} individual monochromatic sources. Reconstructed velocity model using sequential single discrete frequencies method (a), individual group (b), partial overlap group method with one frequency overlapped (c), partial overlap group method with two frequencies overlapped (d), the Bunks' method (e) and simultaneous method (f).

Table 4.5: Total number of forward modeling performed to reach within approximately the same quality of the reconstructed BP salt model at each frequency group inversion for noisy data with SNR =10.

Method	Noisy Data: SNR =10
	Number of forward modeling
Discrete sequential	4020
Group	4091
Partial overlap group (1)	4180
Partial overlap group (2)	6483
Bunks	9391
Simultaneous	10245

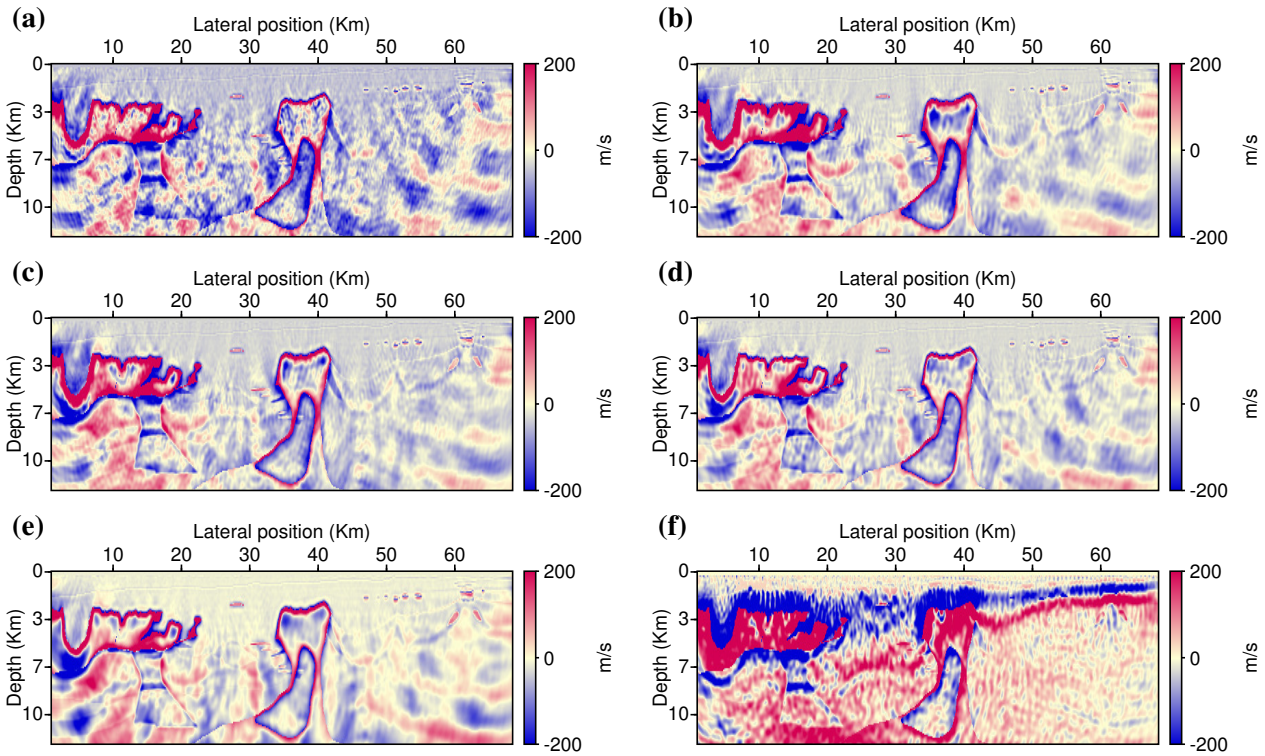


Figure 4.17: Difference between the true BP velocity model and the reconstructed velocity models in Figure 4.16; residual from the velocity model using sequential single discrete frequencies method (a), individual group (b), partial overlap group method with one frequency overlapped (c), partial overlap group method with two frequencies overlapped (d), the Bunks' method (e) and simultaneous method (f).

reproduce images that are comparable to the original velocity model. The cross-talk artifacts are better attenuated when multiple frequencies are inverted simultaneously. However, high resolution images can be obtained by using the Bunks' method or by partially overlapping frequencies across the frequency groups; the more the overlap frequencies across the group, the better is the result. However, more frequency overlapping results in more computational time as it requires more number of forward modeling to be performed.

4.4 Computational cost

The computational cost of the simultaneous multi-frequency FWI is expensive compared to the sequential single discrete frequency method. This is because in the multi-frequency inversion for every new frequency the matrix $\mathbf{A}$ must be factored using LU decomposition. Much of the computational cost is associated with the LU decomposition or factorization of the matrix $\mathbf{A}$. If the same number of iterations is needed for FWI, it is clear that Bunks' method is the most time consuming method compared to the rest of the methods. Note that both Bunks' and partial overlapping frequency methods provide similar image quality values. However, Bunks' method requires approximately twice as many forward modeling simulations as the partial overlapping method in order to achieve the same quality of reconstruction. Note that, we stress that in this chapter and next one we take full advantage of the direct solver to solve for multiple sources at a time by forward and backward substitutions after the Helmholtz operator is factorized. However, the simultaneous source techniques are more relevant to a problem where no direct solvers are used for the forward modeling .

Next, we assess the computational efficiency of the Gauss-Newton method in the framework of simultaneous encoded sources and compare it with the conjugate gradient method. For inversion, we used the Marmousi velocity model (Figure 4.2) and we run the inversion using the conventional FWI method and with simultaneous sources using 3 super-shots, where each super-shot contains 32 individual sources that are encoded with random time-shift. At every iteration, new source sampling positions and encoding functions are generated. For the numerical inversion, a sequential single discrete frequencies method is assumed and we run the inversion until $Q_k < 39$ is satisfied for each frequency. The computational speed up is then compared using

$$S = (1 - F_{sp}/F_o) \times 100, \tag{4.13}$$

where F_{sp} is the total number of forward modeling simulations performed when using simultaneous sources and F_o is the total number of forward modeling simulations performed using the conventional method (Castellanos-Lopez et al., 2013).

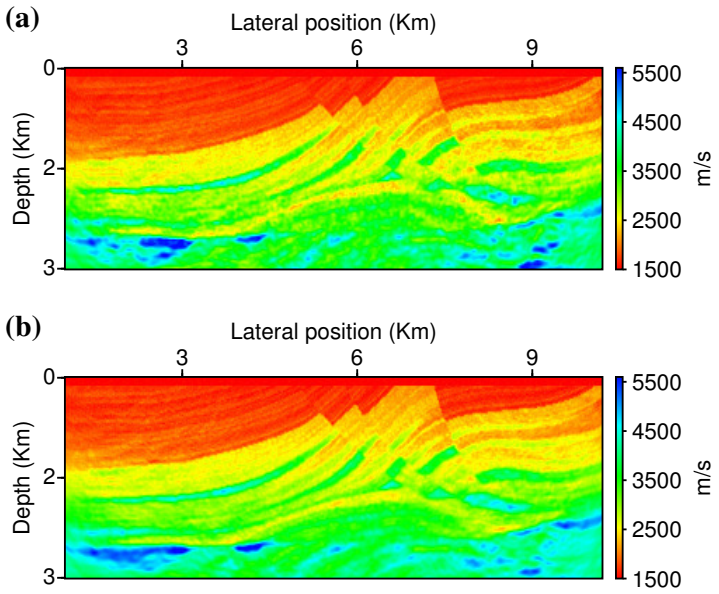


Figure 4.18: Reconstructed Marmousi velocity model obtained using conjugate gradient method (a) and using Gauss-Newton method (b) using sequential single discrete frequencies method.

Figures 4.18a & 4.18b are the reconstructed velocity models from the three super-shots using conjugate gradient and Gauss-Newton methods, respectively. The total forward modeling solved for the conjugate gradient method and Gauss-Newton method to achieve the same average image quality value at each frequency group are 5383 and 4120, respectively. For the conventional FWI, we found a total of 8615 number of forward problems are performed. As a result, a faster convergence rate is achieved by using the Gauss-Newton method with the generation of new source positions and new encoding functions at every iteration. The speed up for conjugate gradient and Gauss-Newton methods to achieve similar quality models are 38% and 52%, respectively. The speed up in the convergence rate by the Gauss-Newton can be confirmed from the reduction in the misfit function as shown in Figure 4.19. Similar pattern was also observed by Castellanos-Lopez et al. (2013).

4.5 Summary

Using the advantage of simultaneous source techniques based on time-shift source encoding methods, we are able to show that simultaneous-source full waveform inversion provides inversion results that are comparable with those obtained from FWI using the conven-

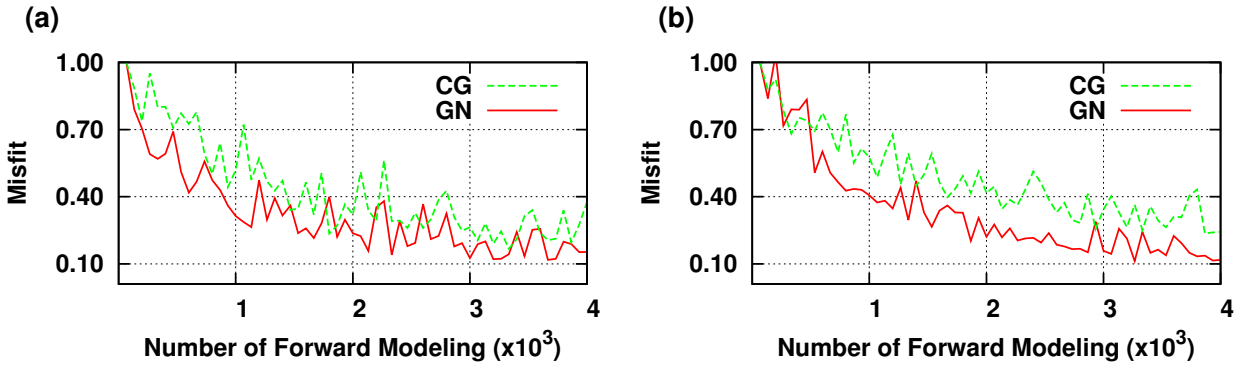


Figure 4.19: Reduction in misfit function vs number of forward modeling performed at 4.27Hz (a) and 12.33Hz (b).

tional individual shot-gather datasets. The full waveform inversion algorithm is formulated via a quadratic matrix-free Gauss-Newton method, where the objective function converges quadratically. Although the Gauss-Newton method is computationally expensive for large scale geophysical problems, the use of simultaneous shooting technique reduces the computational cost and makes the algorithm more suitable for practical applications. However, the reduction of computational cost comes at the cost of introducing cross-talk artifacts in the estimated image. These artifacts arise from interference among the sources of the super-shots. The Gauss-Newton method converges faster than the conjugate gradient method whenever a new encoding functions and source positions are generated at every iteration and hence a faster reduction in cross-talk artifact is achieved.

In this chapter, we examined and addressed a way to attenuate the cross-talk noise problem further by using frequency selection methods for frequency-domain FWI. In addition to a single sequential frequency inversion, we studied the effects of different simultaneous multi-frequency selection inversion strategies. We demonstrate that inverting multiple frequencies simultaneously is important to attenuate cross-talk artifacts. We back up our conclusions with a series of examples using the Marmousi, Marmousi-II and BP/EAGE velocity models with clean and noisy datasets. We were also able to show that higher resolution results can be attained with a partially overlapping simultaneous multi-frequency strategy. The larger the overlap of frequencies across the group, the better is the result. The latter permits us to define an inversion strategy with a good balance between quality and computational time in comparison to the other multi-frequency inversion strategies. We point out that for our simulations, the Bunks' method is also the best method in terms of its ability to recover model parameters. However, Bunk's method also required a large number of

forward modeling simulations making it computationally more demanding than the partial overlapping simultaneous multi-frequency strategy. Note that the performance of different frequency selection strategies depends on many of factors: the quality of the initial model, acquisition geometry, the level of noise present in the data, the band width of the data and the complexity of the unknown model are among a few to mention. We also point out that the presented simultaneous source encoding techniques are based on an L_2 objective function that requires a fixed receiver geometry. Moreover, we have assumed frequency selection strategies that are optimal for a wide-aperture seismic data. Further research is needed to examine the performance of different multi-frequency inversion strategies on data that is poorly represented at low frequencies and/or for receivers spread that are not stationary.

CHAPTER 5

The application of full Newton method for full waveform inversion

5.1 Introduction

FWI is usually formulated as an iterative local optimization problem for solving model parameters by minimizing the l_2-norm of the misfit function between the observed data and computed data (Tarantola, 1984b; Virieux and Operto, 2009; Hu et al., 2011). In order to estimate the model parameters from the measured seismic data, gradient based optimization methods such as steepest-descent and nonlinear conjugate gradients or Newtonian methods such as quasi-Newton such as l-BFGS (Nocedal, 1980), Gauss-Newton (GN) and full Newton (FN) can be used (Pratt et al., 1998). For large scale problems, due to the computational cost, the gradient-based methods are the most popular local optimization methods for FWI problems. However, the convergence is very slow and sometimes these methods fail to converge to the desired global local minimum. As described in the previous chapter, the Newton methods require the Hessian matrix: the second-order derivatives of the misfit function with respect to the model parameters.

Newtonian methods generally converge faster than the gradient-based nonlinear conjugate gradient (NCG) method, but this comes at the expense of solving a large system of equations that incorporates a dense Hessian matrix. In practice, the solutions to model parameters update in Newton method are often approximated by using an iterative method such as the conjugate gradient (CG) method (Navon et al., 1988). Of the Newtonian methods, the FN method is known to converge faster than the Gauss-Newton or the quasi-Newton methods like l-BFGS methods. However, proper care must be taken in solving or approximating the solutions when using FN method. This is because FN method converges poorly and/or

gives rise to wrong solutions when the Newton step is solved using CG. This is because, the Hessian matrix that arises from the second-order derivative of the misfit function is an indefinite matrix. Therefore the standard CG method is no longer suitable for such system of linear equations due to negative curvature directions of eigenvectors corresponding to negative eigenvalues of the Hessian matrix. Instead, we encompass the minimum residual method (MR) (Fischer et al., 1998; Lai et al., 1997), which can be applied to such indefinite system of linear equations. Most recent works of FWI in literatures that is based on the Newton methods are solved using an iterative conjugate gradient method (Hu et al., 2009; Métivier et al., 2013; Anagaw and Sacchi, 2014). The CG method is only applicable for solving a system of linear equation with a symmetric positive definite matrix. However, the CG method often breaks down when the matrix has both positive and negative eigenvalues; i.e. when the matrix is an indefinite matrix. Therefore, approximating the solutions of the indefinite matrix using the CG method is not appropriate. The minimal residual method, instead, can be applied to symmetric indefinite systems. However, for FWI applications, the MR method converges slowly. In order to speed up the convergence rate and its efficiency, we propose a preconditioned regularized minimum residual method (PrMR), where the preconditioner is the pseudo-Hessian matrix (Shin et al., 2001).

In addition to the computational aspect of the Newton method, Pratt et al. (1998) has shown the importance of incorporating the inverse of the Hessian matrix (Pratt et al., 1998). Incorporating the inverse of the Hessian matrix not only improves the convergence of the optimization algorithm but also it tries to correct and removes artifacts in the gradient that arise from the second-order multi-reflected wavefields such as multiples. In other words the use of the Hessian matrix in the waveform inversion process not only speeds up the convergence rate but can also act as a filtering operator (Pratt et al., 1998). The presence of high amplitude multiple reflections wavefields challenges the gradient-based optimization method to converge to the global local minima, which is often based on the first-order derivatives of the misfit function. Moreover, the Hessian in the numerical inversion scheme handles the defocusing that affects the gradient due to the incomplete and uneven illuminat ion subsurface areas. The Hessian matrix predicts the defocusing on the gradient and its inverse focuses the defocusing on the gradient. Thus, incorporating the inverse of the Hessian matrix not only properly scales the amplitude of the gradient but also sharps the gradient (Pratt et al., 1998). More can be found in Pratt et al. (1998) paper.

Given high computational cost of the forward modeling process, the practical application of the FN method for large scale geophysical problem has yet to be seen. Note that the computational cost associated with the forward modeling is proportional to the number of sources of the experiment and is an intrinsic problem of FWI that prevents the feasibility of FN method for large scale geophysical problem particularly for 3D problems. This is because at each iteration in the FWI scheme, forward modeling will be performed for all

sources present in the data. In this chapter, we should reiterate once again that we take advantage of the simultaneous source techniques.

In the previous Chapter 3, it is observed that results obtained from the full Newton and Gauss-Newton methods are similar. One of the reasons is that the Marmousi velocity model does not have strong velocity structure contrasts and the resulting data from the velocity model will not have strong amplitude multi-reflected wavefields. As a result, the Newton's methods provide similar results. If there are no strong amplitude multi-reflected wavefields present in the data, the FN method provides a comparable as the GN method. On the other hand, in this chapter, we demonstrate that the high velocity contrast in the top 6 Km of BP/EAGE velocity model between the salt dome and the layer of the low velocity sediments underlying the basalt generate high amplitude multiple reflections. Inversion results from BP velocity model highlight the potential applicability of the FN method for high resolution velocity model building. We would emphasize the advantage of full Newton method over the Gauss-Newton method for such types of models. Analysis of the FWI process also suggest that faster reduction in the misfit function is achieve when the Newton's equation is solved iteratively with the PrMR method than with the CG. We demonstrate our premises using simultaneous source phase-encoded technique. Numerical results based on the BP/EAGE velocity model (Billette and Brandsberg-Dahl, 2005) are presented to highlight the numerical efficiency of FN method of full waveform inversion with PrMR.

5.2 Methodology

In this chapter, we start from equation 4.1 (in Chapter 4) but we would like to give more stress on the regularization term. For the regularization term, $R(\mathbf{m})$, we used the Tikhonov regularization term (Anagaw and Sacchi, 2011).

$$R(\mathbf{m}) = ||\nabla^2(\mathbf{m} - \mathbf{m_o})||^2, \tag{5.1}$$

where ∇^2 is the second-order derivative in space. Here, we would like to emphasize the importance of incorporating a regularization term (equation 5.1) for solving the Newton's equation in the framework of source encoding strategies. In this chapter, the Tikhonov damping regularization is considered. The choices of the regularization term and regularization parameter μ greatly determine the quality of the reconstructed model and the convergence rate of the algorithm employed. The damping regularization term, however, may not be appropriate choices as it enforce the solution to remain close to the initial model. For this chapter, we considered Tikhonov smoothing regularization terms, where the second-order derivatives are penalized instead of distance from the prior model.

Instead of penalizing the model, other regularization terms can also be implemented like the Total variation (TV) method (Osher and Fatemi, 1992). The TV regularization tries to simultaneously preserve the edges and smooths the images along the direction adjacent to edges. Previous studies using the Total Variation regularization term (Anagaw and Sacchi, 2011) suggest that the use of a smoothing damped regularization term is a better option for our purpose as a stabilizer and penalizing the model than using TV. When the TV regularization term is adopted, two parameters have to be chosen simultaneously. Selecting these regularization parameters to help preserve both the sharpness and the smoothness of the model is very hectic and time consuming. One has to choose optimum parameters that balance the trade off between the model and data weight. Often the regularization parameters are chosen by trial and error. Consequently, the suitability of TV regularization term for large scale problems in FWI is less practical. On the other hand, the Tikhonov smoothing regularization is more practical and efficient to employ in the FWI scheme. Details about the importance of the Tikhonov and TV regularizations have been extensively studied in previous research works (Anagaw and Sacchi, 2011, 2012b; Anagaw, 2009).

5.2.1 Phase-encoding simultaneous sources

In this chapter, we focus on the simultaneous sources inversion scheme in which sources are randomly encoded with random phase-shifts. Following Chapter 4, we instead replace equation 4.3 by the source phase encoding vector

$$\Gamma(\phi) = [e^{-i\phi_1} \quad e^{-i\phi_2} \quad e^{-i\phi_3} \quad ... \quad e^{-i\phi_{ns}}]^T, \tag{5.2}$$

where $\phi_i \in [0.0, \pi]$.

A simple phase shift $\{1, -1\}$ or $\phi_i \in 0, \pi$ can be used. However, from our experiences with the simultaneous source methods, the difference in the final results from the source encoding functions such as phase-shift or time-delay are negligible. Both random phase-shift and time-delay are efficient encoding techniques for attenuating cross-talk noises than a simple phase shift $\{1, -1\}$ technique. However, detail investigations on the merits of source encoding function are beyond the scope of this thesis.

5.2.2 Workflow implementation

The theoretical formulation and implementation of a matrix-free GN and FN methods for FWI was discussed in Chapter 2 and Chapter 4. The idea of minimum residual iterative method is not new and have been covered on the literature for quite sometime. Details

about the formulations of MR algorithm can also be found in books on optimization. In this section, we would like to extend the application of the preconditioned MR algorithm for full waveform inversion when the full Newton method is employed in which its Hessian matrix is an indefinite symmetric matrix. The PrMR scheme, like the CG iterative technique, requires only the action of the Hessian matrix on the model perturbation: matrix vector product. Table 6, which calls Algorithm 7 is the pseudo-code of full Newton algorithm with the preconditioned minimum residual method. For others, see the previous chapters.

Algorithm 6: Pseudo-code: Multiscale full Newton method. Notations: nw :- number of frequencies within a group frequency nwg, $\mathbf{m}$ is the model parameter , Γ and D_R are the phase encoding and source position operators, respectively.

Input : $\mathbf{d}^{obs}(\omega)$, $\mathbf{m}_0$, N_w, max_iter
Output : $\arg\min_{\mathbf{m}} J(\mathbf{m})$
Initialize parameters;
Option 1: Generate random operators $\rightarrow \Gamma, D_R$
for $w_i \leftarrow 1$ **to** N_w **do**
  Starting model $\leftarrow \mathbf{m}_k$;
  Option 2: Generate random operators $\rightarrow \Gamma, D_R$
  for $k \leftarrow 1$ **to** max_iter **do**
    Option 3: Generate random operators $\rightarrow \Gamma, D_R$
    Compute:
    forward wavefields $\mathbf{d}^{cal}(\omega)$;
    `residual wavefields` $\triangle\mathbf{d}(\omega)$;
    `back-propagate residual wavefields` $\lambda(\omega)$;
    gradient $\nabla_{\mathbf{m}} J(\mathbf{m}_k)$;
    Solve:
    $\triangle\mathbf{m}_k = -\Big[\Re e(\mathbf{H})\Big]^{-1} \Re e(\nabla_{\mathbf{m}} J(\mathbf{m}_k))$;
    solve $\triangle\mathbf{m}_k$ using PrMR method call: Algorithm 7;
    Two more forward modeling operators are required to compute;
    the action of the Hessian matrix on $\triangle\mathbf{m}$ inside the PrMR algorithm;
    Compute:
    α using line search method;
    Update model:
    $\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha\triangle\mathbf{m}_k$;
  end
end

For preconditioner $\mathbf{M}$ in Algorithm 7, we used the diagonal elements of pseudo-Hessian matrix based on the source-receiver reciprocity theory (Shin et al., 2001)

$$\mathbf{M} = \Big[\Re e\Big\{\sum_{\omega}\omega^4\sum_{s}|f(\mathbf{x}_s,\omega)|^2|\mathbf{G}(\mathbf{x}_s,\mathbf{x},\omega)|^2\sum_{r}|\mathbf{G}(\mathbf{x},\mathbf{x}_r,\omega)|^2\Big\} + \beta\mathbf{I}\Big]^{-1}, \tag{5.3}$$

where β is a damping factor that controls the stability of the inversion, which is used to prevent a singularity problem.

For each frequency, the preconditioner is computed once at the beginning of the numerical calculation.

Algorithm 7: Pseudo-code: Precondition minimum residual (PrMR).

Input : max_{iter}, ε, $\mathbf{x}_0$, $\mathbf{g}$, $\mathbf{H}$ and $\mathbf{M}$
Output : $\mathbf{x}$
Initialize parameters;
$\mathbf{r}_0 = \mathbf{0}$ $\mathbf{r}_{-1} = \mathbf{0}$ $\mathbf{r}_{-2} = \mathbf{0}$
$\mathbf{r}_1 = \mathbf{g} - \mathbf{H}\mathbf{x}_0$ $\mathbf{z}_1 = \mathbf{M}\mathbf{r}_1$ $\beta_1 = \langle \mathbf{z}_1, \mathbf{r}_1 \rangle$ $\eta = \beta_1$
$\gamma_0 = 1$ $\gamma_1 = 1$ $\sigma_0 = 0$ $\sigma_1 = 0$
$\mathbf{y}_0 = \mathbf{0}$ $\mathbf{y}_{-1} = \mathbf{0}$ $\mathbf{y}_{-2} = \mathbf{0}$
for $k \leftarrow 1$ **to** *max_iter* **do**
  Calculate the lanczos vectors
  $\mathbf{r}_k = \mathbf{z}_k/\beta_k$; $\alpha = \mathbf{r}_k^T \mathbf{H}\mathbf{r}_k$
  $\mathbf{r}_{k+1} = \mathbf{H}_k\mathbf{r}_k - \alpha\mathbf{r}_k - \beta_k\mathbf{r}_{k-1}$
  $\mathbf{z}_{k+1} = \mathbf{M}\mathbf{r}_{k+1}$; $\beta_{k+1} = \langle \mathbf{z}_{k+1}, \mathbf{r}_{k+1} \rangle$
  Calculate QR factors
  $\sigma = \gamma_k\alpha_k - \gamma_{k-1}\sigma_k\beta_k$; $\rho_1 = \sqrt{(\sigma^2 + \beta_{k+1}^2)}$
  $\rho_2 = \sigma_k\alpha_k + \gamma_{k-1}\gamma_k\beta_k$; $\rho_3 = \sigma_{k-1}\beta_k$
  Update solution
  $\mathbf{y}_k = (\mathbf{r}_k - \rho_3\mathbf{y}_{k-2} - \rho_2\mathbf{y}_{k-1})/\rho_1$; $\mathbf{x}_k = \mathbf{x}_{k-1} + \gamma_{k+1}\eta\mathbf{y}_k$
  $\eta = \sigma_{k+1}\eta$
  if $\eta > \varepsilon$ return
  end do
end

5.3 Results and Discussions

First, we present a good representative model suitable for the application of full Newton method. Figure 5.1 is a very simple velocity model used for the FN FWI investigation. The background velocity is 1900 m/s and the small anomaly is 3000 m/s. Hence, there is a strong velocity contrast between the anomaly and the background. The presence of the high velocity contrast produces high reflection amplitudes and if a free surface acquisition survey is assumed, there would also be strong multi-reflected waves. For the inversion, we assumed a total of 900 unknown model parameters, i.e. the model grid size is 30 x 30 with an isotropic grid spacing of 10 m.

Due to presence of such strong reflection amplitudes and multi-reflected event, the optimization of the first-order approximation of the objective function would struggle to retrieve the

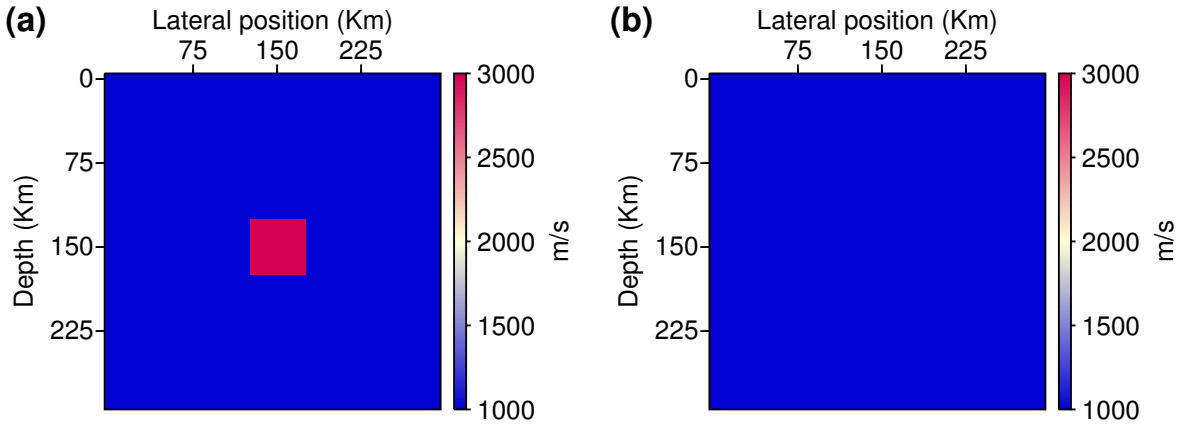


Figure 5.1: True model with small anomaly (a) and constant velocity model used as a starting model for inversion (b).

model parameters in FWI. Moreover the inversion process also generates artifacts in the final image. Often the iterative process runs for a large number of iterations. However, incorporating the second-order approximation of the objective function in the optimization methods enables to balance the amplitudes of model updates and reduces artifacts in the gradient vector. Consequently, high resolution images would be achieved.

In order to validate our premise, we used the descent, Gauss-Newton and full Newton method to test the performance of the optimization methods. Figures 5.2a-c are results obtained after the 1^{st} iteration from simultaneous multiple frequency inversion with the descent, Gauss-Newton and full Newton methods, respectively. A set of 20 frequencies were selected and all data frequencies were inverted simultaneous. As we see from the results, the descent method based on the first-order derivatives of the objective function produces artifacts in the final image. On the other hand a more focused and less noisy image is achieved by employing the second-order optimization method using GN and FN methods. The full Newton method produces a high resolution image and the amplitude of the velocity anomaly is well retrieved. All artifacts observed in the final model by the descent method are removed by the full Newton method. Note that a less noisy image is observed in the case of full Newton method than the Gauss-Newton method. This is due to the fact that the last two terms of the Hessian matrix predict artifacts in the gradient vector due to multiple-reflected events and the contribution to the inverse of the Hessian matrix corrects the artifacts. Details of the meaning of the Hessian matrix and its inverse are described in Chapter-1 and 2. In this chapter we only demonstrate its applications.

The next example is the 2004 BP/EAGE salt model (see Figure 4.14). This model is a good

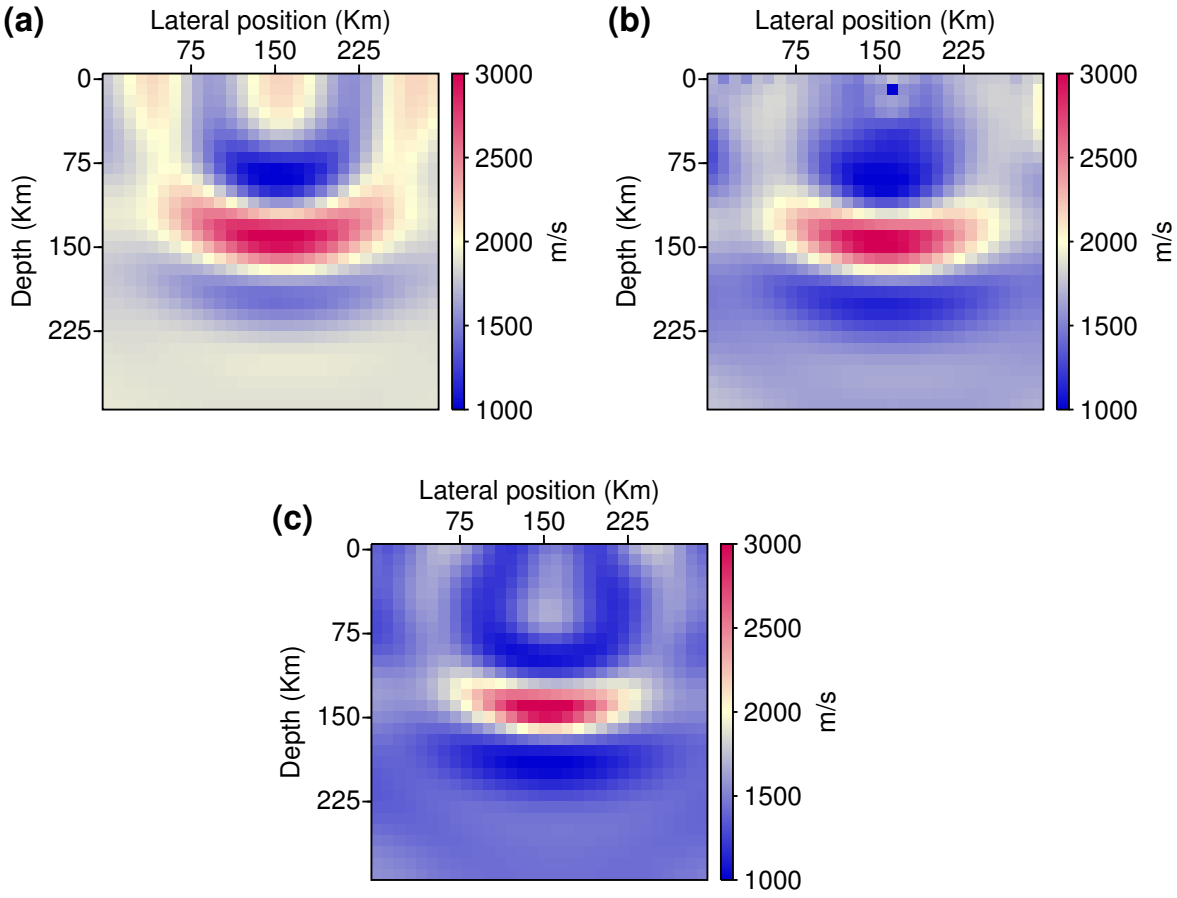


Figure 5.2: Retrieved model after the 1^{st} iteration from simultaneous frequency inversion using the descent (a), Gauss-Newton (b) and full Newton methods (c).

representative candidate to use full Newton method. Similar to Chapter-4, the original BP/EAGE velocity model is re-gridded onto a 50 m x 50 m spatial grid. Figure 4.14a depicts the true BP/EAGE velocity model and Figure 4.15a is the smooth model used as a starting model. The BP/EAGE salt model contains salt domes with high velocity and subsalt slow velocity anomalies.

The high velocity contrast between the top of the salt domes, subsalt velocity anomalies and the water layer is responsible for generating strong amplitude multi-reflections. Free surface acquisition geometry will also produce high amplitude multi-reflections and also strong reflection amplitudes in the base of the salt. The main challenges in this model for FWI are to retrieve salt body and the subsalt velocity anomalies properly. This is because high reflection amplitude on the base of the salt and the low subsalt velocity anomalies is a challenge for any imaging technique, including FWI, to appropriately reconstruct the velocity model and to focus reflectors correctly.

With free surface acquisition geometry, we generate a synthetic seismogram data from 113 sources that are spread equally at surface, and a total of 337 receivers are equally spread one grid size below the surface. All receivers are active for each source. A Ricker wavelet with 10 Hz central frequency is used to initiate the source. A set of thirteen discrete frequencies were selected for the full waveform inversion. The synthetic data is then contaminated by noise with SNR = 10. In order to speed up the numerical inversion, a simultaneous-source technique is used. A total of 11 super-shots are used where each super-shot contains approximately 10 individual monochromatic sources that are randomly assembled with phase-encoding technique. Since the number of sources within a single super-shot is relatively small as compared to the super-shots used in the previous chapter, we expect fewer cross-talk artifacts in the final images. For this reason, instead of using the partially overlapping frequencies group technique, which is efficient in attenuating cross-talk artifacts, we restrict ourselves to a sequential single frequency inversion strategy. Because the main goal in this chapter is to demonstrate the potential applicability of the full Newton method in the BP/EAGE salt velocity model.

As described in the previous chapter, generating new source encoding functions and source sampling positions at every iteration are crucial for faster reduction in cross-talk artifacts and faster convergence rate. Similar strategies are also adopted here. However, the l-BFGS method that requires the histories of the gradient and the model perturbation requires a special treatment. Changing the source encoding functions and sampling positions at every iteration also changes the histories of the gradient and the model perturbation. This makes the l-BFGS method unstable and often fails to converge. Consequently, in cases where simultaneous source encoding strategies were used, we only compared FWI results from the gradient-based method such as the PNCG, and the Newton-based methods such as the GN

and FN method.

In order to emphasize our premise about the effectiveness of the PrMR algorithm in finding the solutions of the Newton directions, a comparison between the results obtained from the PrMR and the CG methods are examined. First, the solutions to the Newton's equation, the GN and the FN methods are iteratively solved using CG. For all preconditioned iterative methods, the preconditoner is the diagonal of the pseudo-Hessian matrix. For each frequency in the numerical inversion, we run the inversion until the model quality value $Q_k < 60.00$ is satisfied at the k^{th} iteration (see the definition of the image quality measurement in Chapter 4) . When the Newton's direction is solved using the CG method, the same inversion strategy is adopted as the strategy outlined for GN method in Chapter 4. For the inner loop of the GN and FN algorithm, we used a maximum of 15 Newton iterations with a stopping criteria with $\varepsilon = 0.005$ or we force to stop solving the Newton's equation whenever a negative curvature is detected. For stability purpose, the regularization parameter for both Newton methods (GN and FN) is set to 0.008. From a number of trial and error tests, we found that the choice of this regularization parameter provides a good trade off balance between the misfit function and regularization weight. In the next section, the influence of regularization parameters on the final reconstructed model is presented. Note that the choice of stopping criteria in the Newton's iteration also affects the overall computational cost, convergence rate and quality of the reconstructed model parameters. However, there is no silver bullet in finding the best stopping criteria. In this thesis as described in the previous chapter, we adopted the Eisenstat et al. (1994) condition. An extensive study on the choice of stopping criteria is beyond the scope the thesis. Detailed investigations on the subject can be obtained in Eisenstat et al. (1994) and Métivier et al. (2013). Similar stopping criteria are also used for PrMR.

Figures 5.3a-5.3c are the reconstructed velocity model using PNCG, GN and FN methods, respectively. These numerical inversions reproduce models that are comparable to the original velocity model. All methods retrieved the velocity model properly; the top salt body structures, shallow anomalies and the left side of the velocity model are all well retrieved. However, in the case of PNCG method, beneath the salt body to the left of BP model, the subsalt velocity anomalies are poorly reconstructed. This is due to the high velocity contrast present in the model and the PNCG method struggles to properly delineate the presence of strong amplitude multi-reflections in the region. On the other hand, the use of GN and FN methods provide more trustworthy results by properly retrieving the information of the strong multi-reflected amplitudes. However, the subsalt velocity anomalies are by far better reconstructed, and the image is clearer in the case of FN method than the GN method. The FN method provides a more detail structures of the velocity model.

If we change the CG method for solving the Newton's direction by the PrMR, we would

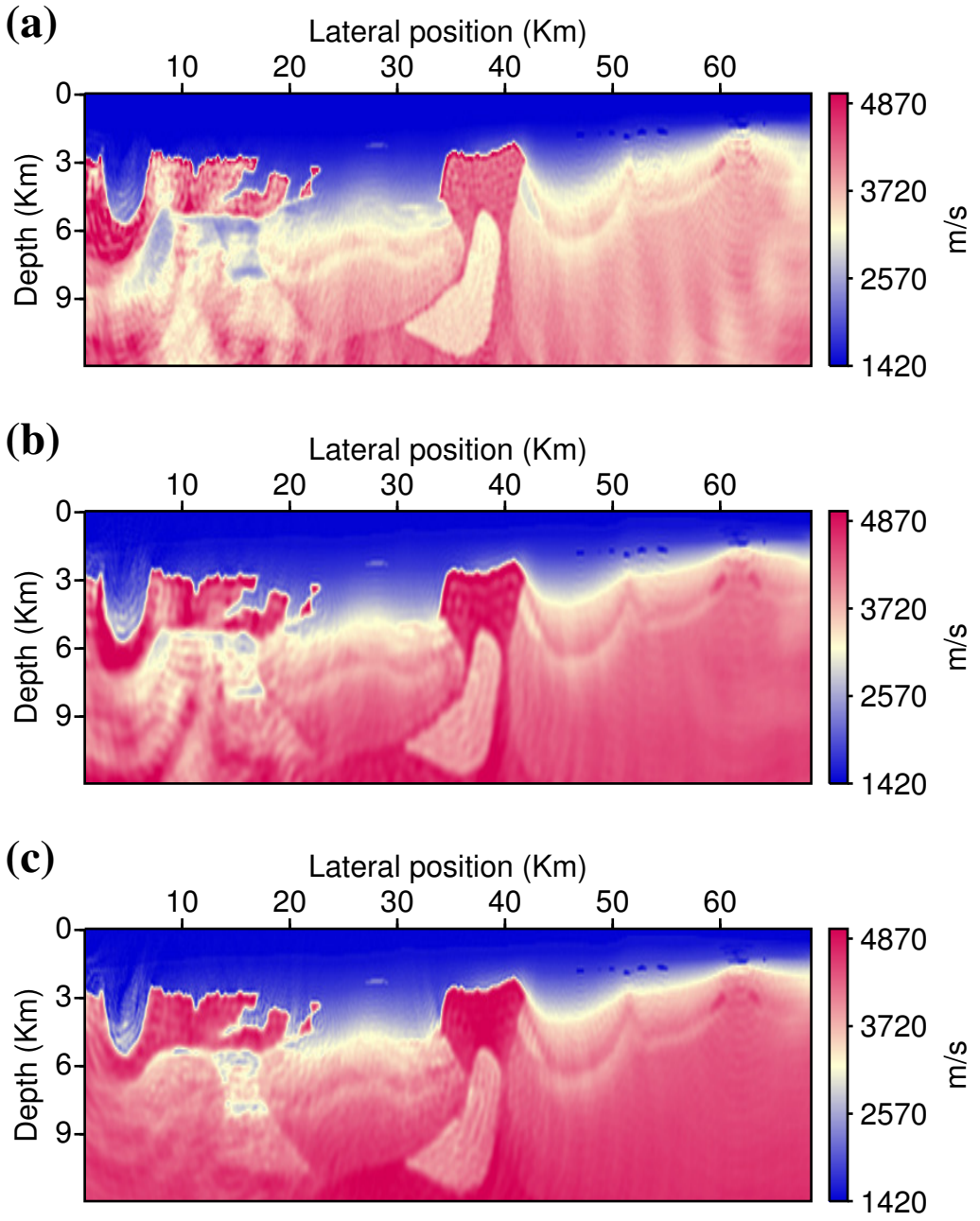


Figure 5.3: Reconstructed BP/EAGE salt velocity model using PNCG (a), GN (b) and FN (c) methods. The Newton's equation is iteratively solved using conjugate gradient method.

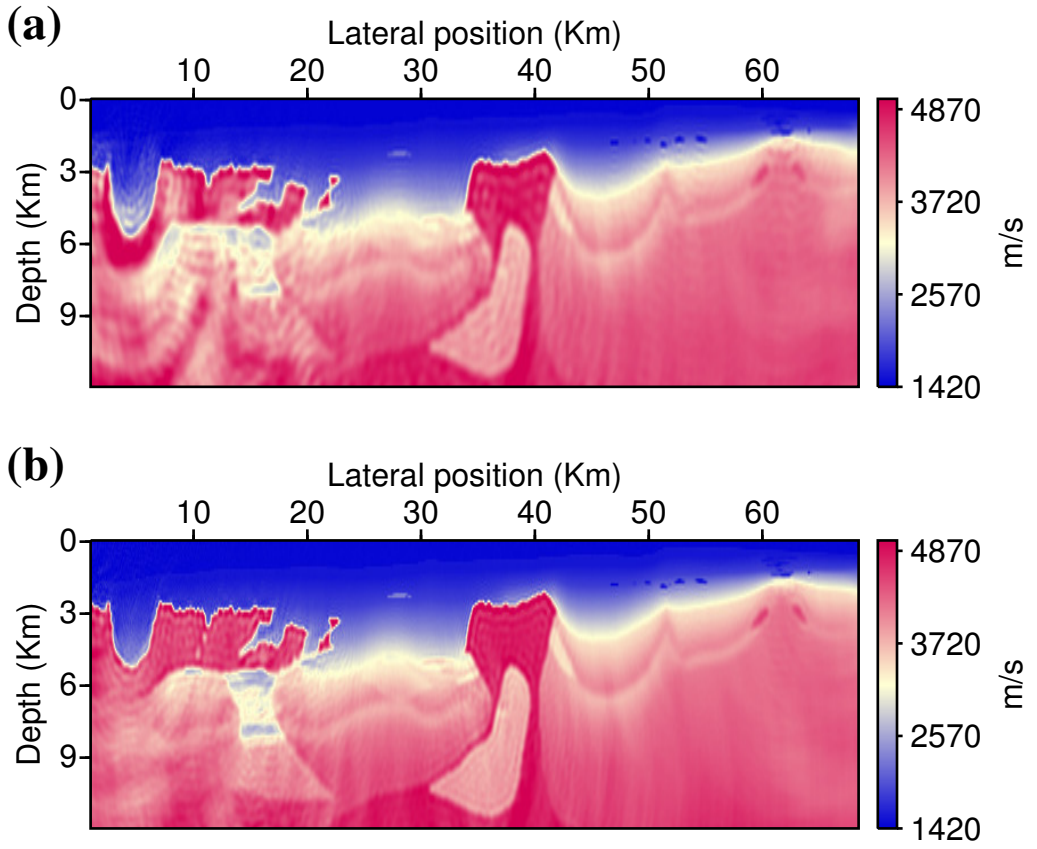


Figure 5.4: Reconstructed BP/EAGE salt velocity model using GN (a) and FN (b) methods. The Newton's equation in the FN method is iteratively solved using the PrMR method.

able to get a better result in terms of resolution and quality of the final retrieved model. In addition, the use of PrMR for solving the FN directions delineates both the shallow and subsalt velocity anomalies, compare Figure 5.3c & 5.4b. Figure 5.4b shows the estimated velocity model using FN method when the PrMR solver is employed. Figure 5.4a is the result obtained from GN. Comparing the subsalt velocity anomalies beneath the salt body clearly shows the advantages of the FN over the GN method. The numerical inversion result from FN method is comparable to the original velocity model. This is due to the fact that the inverse Hessian matrix is properly computed by using PrMR than CG. PrMR has a natural way of handling the negative curvature of the Hessian matrix. In the case of conjugate gradient, on the other hand, we stopped iterating the CG whenever a negative curvature is detected. In the Gauss-Newton method, the approximate Hessian matrix is positive definite matrix, hence the CG method is suitable for solving the corresponding Newton's system of equation. In case of negative curvature detection that results from numerical simulation and approximation, we stop the iterative process of solving the Newton's equation. Consequently, for the GN method when the CG or the PrMR is employed as an iterative solver, we will pretty much get similar results. The major difference in the final reconstructed models comes when the FN algorithm is solved with CG and PrMR methods.

Figures 5.5a-5.5c show the difference between the true BP salt velocity model and the reconstructed velocity models in Figure 5.3a-5.3c using the PNCG, GN and FN methods, respectively. As we see from the model residuals, the GN and FN methods have fewer artifacts leftover than the one obtained using PNCG method. Figure 5.6b is the model residual obtained when FN numerical method is solved iteratively using PrMR. For comparison purpose, the difference model between the result from the GN algorithm and the true model is also sown (see Figure 5.6a). If we compare Figure 5.5c & Figure 5.6b, we see that the FN algorithm with PrMR yields the least artifacts. In the residual model, the small velocity anomalies or small heterogeneities located from 0 Km to 3 Km become negligible or close to zero when the FN method is solved using PrMR than the CG method (see Figure 5.5c, 5.6a & 5.6b).

5.3.1 Effect of regularization

In this section, we demonstrate the effect of regularization on the quality of velocity model building in the framework of simultaneous source encoding strategy. For comparison, the inversion is performed using PrMR method with the same strategy as described before except the regularization term. Figure 5.7a is the estimated velocity model without regularization (or when $\mu = 0.0$). Figure 5.7b & 5.7c are the estimated velocity models with regularization when the regularization parameter 0.0001 and 0.008 are chosen for the inversion, respectively. The resulting velocity model from this example demonstrates that how the quality of the

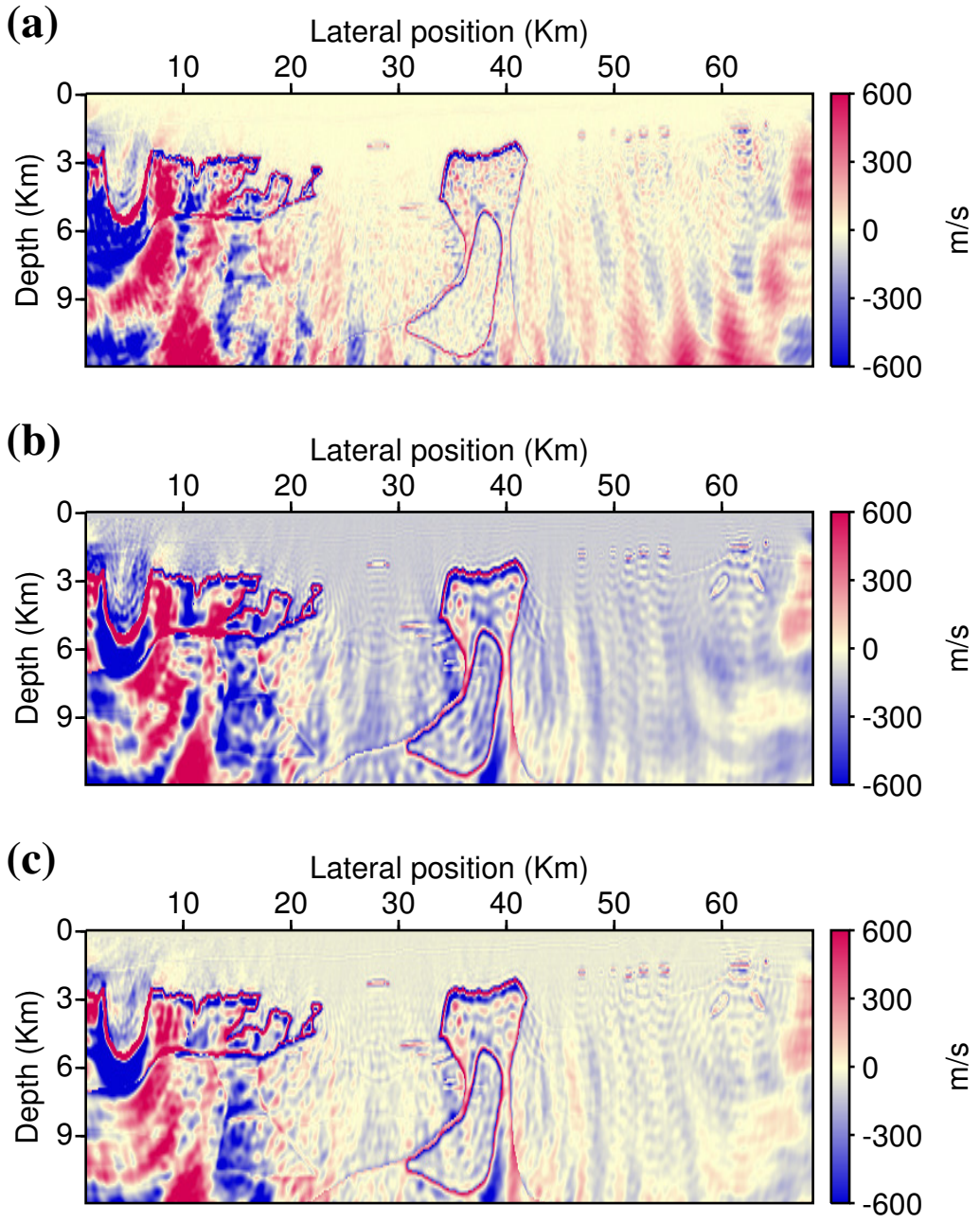


Figure 5.5: Model residual between the true BP velocity model and the reconstructed velocity models in Figure 5.3; residual velocity from the PNCG (a), GN (b) and FN (c) methods when the Newton's equation is solved using CG method.

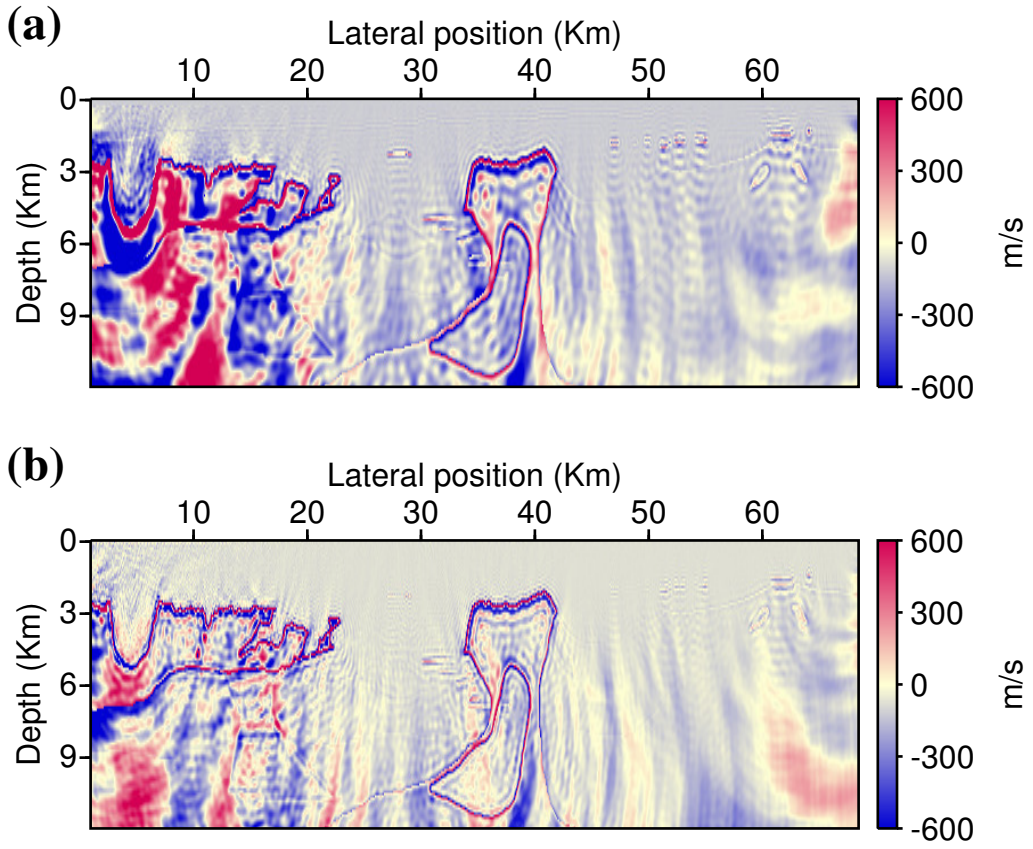


Figure 5.6: Model residual between the true BP velocity model and the reconstructed velocity models in Figure 5.4; residual velocity from the GN (a) and the FN (b) methods. The Newton's equation is solved using PrMR method.

estimated velocity method is affected by the regularization parameter. This confirms the importance of regularization in simultaneous source FWI.

5.3.2 Convergence properties

Next, we see the convergence behavior of the CG and MR methods for the indefinite matrix. For further insight on the performance of FN method using PrMR, we present the convergence rates obtained from preconditioned nonlinear conjugate gradient, Gauss-Newton method, and full Newton method using conjugate gradient and full Newton with preconditioned minimum residual method. Figure 5.8a-5.8d show the reduction in misfit function against the number of forward modeling performed at different frequency stages during the inversion process. The use of PrMR for solving the full Newton method provides a significant enhancement in speed at the early stage of the inversion. However, as we also see from the figure, the performance of the FN method with the CG method is sometimes very poor especially at low frequencies (for example at 1.62Hz). This is due to the fact that CG method fails to handle the negative curvature detection of the Hessian matrix. We often address the negative curvature problem by forcing the algorithm to stop the iterative process; otherwise the inversion would fail or if iterative process is not stopped, we tend to over solve the conjugate gradient, which in turn provides a wrong result. From these numerical tests, for high resolution and better performance, we thus conclude that the full Newton method for full waveform inversion should be solved with PrMR method (Fischer et al., 1998; Lai et al., 1997). It can also be performed with any iterative technique that handles negative curvature of the operator matrix with a good preconditioner.

5.4 Summary

In this chapter, we present a full Newton full waveform inversion algorithm and its application in the presence of high amplitude multi-reflected waves that are generated by strong velocity contrast and/or from free surface reflections. Both the preconditioned conjugate gradient and the minimum residual methods were performed to solve the Newton's equation. In each case, the diagonal of the pseudo-Hessian matrix is used as a preconditioner. Through tests with the BP/EAGE velocity model, we demonstrate that the full Newton FWI algorithm with preconditioned minimum residual method outperforms the one with the conjugate gradient method or Gauss-Newton method in terms of the accuracy and resolution of the results, as well as convergence rate.

In order to reduce the computational cost, we have also incorporated the simultaneous sources technique. The simultaneous source technique presented in this chapter is merely to

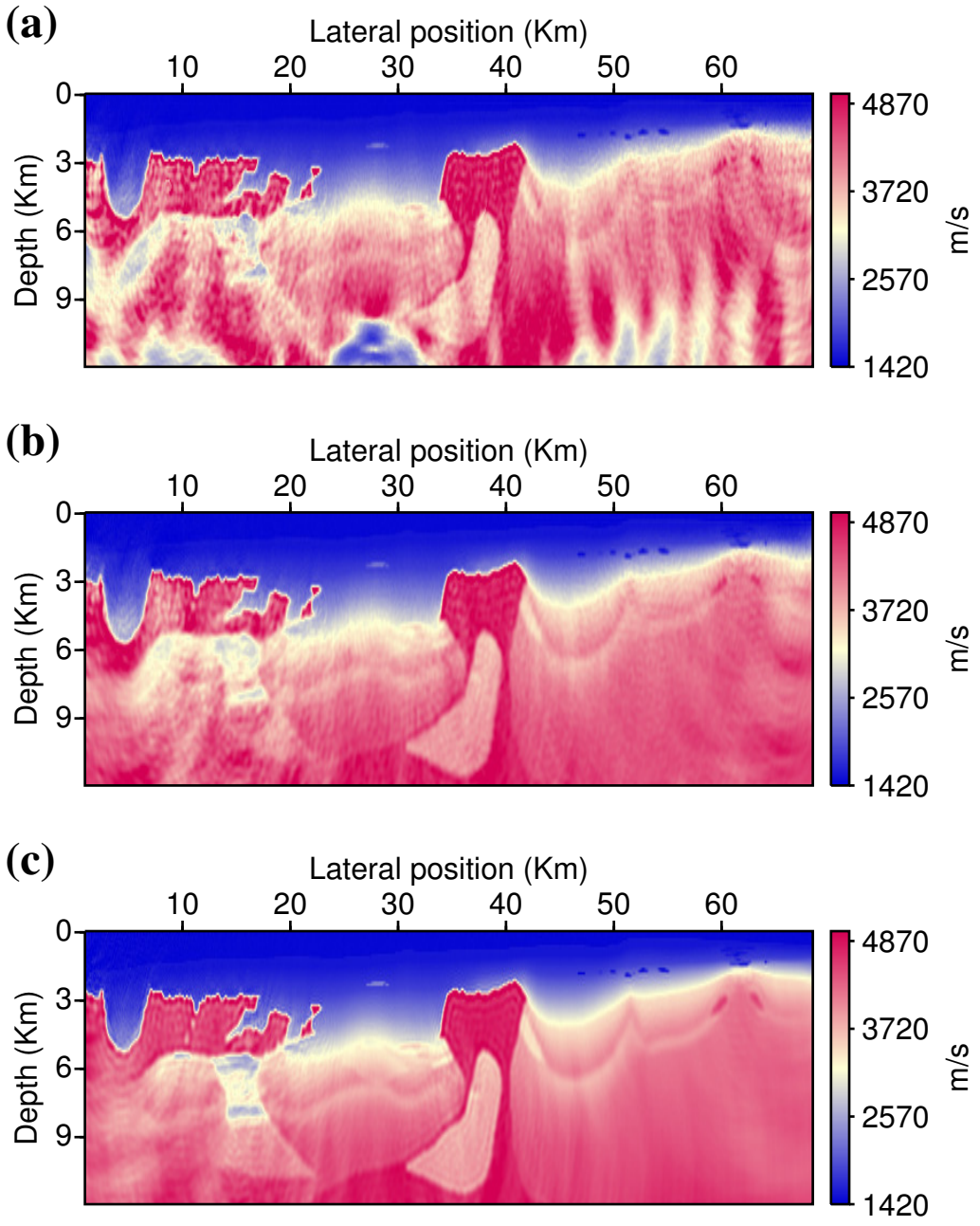


Figure 5.7: Effect of regularization on the reconstructed velocity model in the framework of simultaneous source method: (a) no damping, with regularization parameter $\mu = 0.0001$ (b) and $\mu = 0.008$ (c).

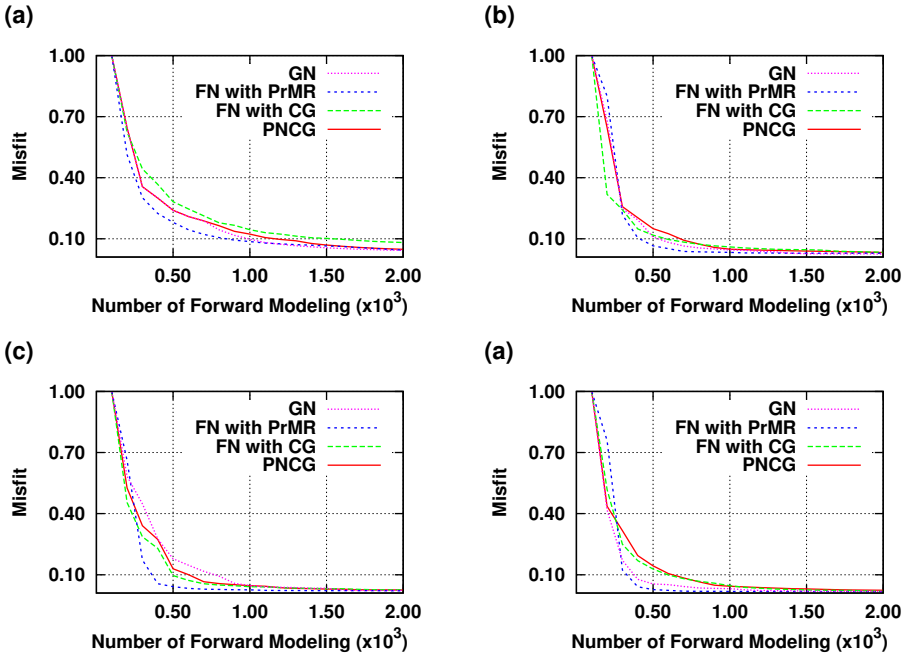


Figure 5.8: Relative reduction in misfit function vs number of forward modeling performed at 1.62Hz (a), 3.07Hz (b), 5.85Hz (c) and 11.15Hz (d).

speed up the computational time. Unlike the previous chapter where each super-shot was constructed from many sources, here a less dense sources in the super-shot is used. For this reason, the inversion will not produce significant cross-talk artifacts. For further cross-talk artifacts reduction, new source encoding and sources sampling positions are generated at each iteration.

From the numerical tests presented in this chapter, we suggest that the application full Newton method for FWI should be performed with preconditoned minimum residual method in order to achieve a high performance of the algorithm and high resolution images. The minimum residual method has been in the literature for decades. To best of our knowledge, the application of the method to geophysical inverse problem, in particular, to large scale implementation of the full Newton method for FWI has not been explored. If any, often the full Newton method of waveform inversion is carried out by conjugate gradient method with a sophisticated stopping criteria. Here, instead, we propose that the solution to the Newton's system of linear equations in the full Newton method should be solved iteratively with preconditioned minimum residual method, when the Hessian matrix is no longer a positive definite matrix.

CHAPTER 6

Summary

6.1 Conclusions

In this thesis, we presented an efficient work flow full waveform inversion based on the first-order and second-order optimization methods. We presented the theoretical formulation of FWI that is based on the conventional shot-gather and simultaneous-sources techniques. Full waveform inversion is still not robust and requires lots of effort to make it work. Among the most important factors are initial starting model and the quality of the low-frequency

From the synthetic numerical examples presented in this thesis, the conclusions can be summarized as follows:

In chapter 3, we examined the effect of model parameterization on 2D acoustic full waveform inversion using gradient and Newton-based optimization methods. Velocity, slowness and slowness squared model parameterizations were considered. In the Newton's based algorithm, the full waveform inversion is formulated using a matrix-free method. For each frequency, the action of the Hessian matrix on the model perturbation is computed on the fly within a CG algorithm after the Helmoltz matrix is factorized using an LU decomposition. The use of the Hessian matrix in FWI numerical algorithms plays a crucial role in improving resolution and the amplitude of deep parts of the model by re-calibrating the gradients in the shallow and deeper parts of the model. We analyzed the influence the three model parameterizations on the behaviour of the Hessian matrix using eigenvalue decomposition of the Hessian matrix. Analyzing the eigenvectors of the Hessian matrix gives some insights into the influence of the model parameterization in FWI. Eigenvectors analysis of the Hessian matrix show that slowness or slowness squared model parameterization makes the eigenvectors almost depth independent. As a result the deep parts of the model would

be properly reconstructed, and fast convergence rates can be achieved. The high convergence rate is due to the fact that slowness or slowness squared parameterization provide a proper scaling of the descent direction both in the deep and shallow parts of the model. In the Newton-based optimization methods with GN and FN methods, faster convergence rate can be achieved by using slowness or slowness squared parameterization than with velocity parameterization.

From a series of numerical experiments carried out in this chapter, we propose that velocity estimation methods that employ FWI based on the Gauss-Newton and full Newton optimization methods should be carried out using slowness or slowness squared model parameterization. For noise free dataset or data with a moderate amount of noise, a superior convergence rate of Gauss-Newton and full Newton methods can be achieved by the use of slowness squared parameterization. On the other hand, the slowness squared model parameterization is sensitive to noise. In such cases, the slowness model parameterization is a suitable parameterization for velocity estimation, which provides a good trade-off between resolution of the solution and the convergence rate in comparison to slowness squared parameterization.

In Chapter 4, we studied the effects of different simultaneous multi-frequency selection inversion strategies for FWI in the framework of source encoding strategies. We propose a strategy to attenuate cross-talk noise problem that arises from interferences of among the sources of the super-shots. From a series of numerical calculations, we found that in simultaneous sources full waveform inversion, partially overlapping frequencies across a simultaneous multi-frequency group inversion along with regeneration of new source encoding functions and positions are key to achieve a high resolution image and to suppress cross-talk artifacts faster. We found that, whenever a new encoding functions and source positions are generated at every iteration, a faster reduction in cross-talk artifacts would also be achieved by using second-order optimization method than with the first-order optimization method.

Among the overlapping frequencies selection strategies studied in this thesis, the Bunk's method provides the highest resolution image but this method computational expensive. However, by partially overlapping frequencies across the group frequencies, we find a good balance between quality of the resulting image and computational time in comparison to other multi-frequency inversion strategies. This method can also be applied to time-domain FWI, where each frequency group bands are selected by band passing the source and time-domain data. Numerical examples demonstrate that, the larger the overlap of frequencies across the group, the better is the result. Note that the performance of different frequency selection strategies depends on many of factors: the quality of the initial model, acquisition geometry, the level of noise present in the data, the band width of the data, the complexity of the unknown model.

In Chapter 5, utilizing the advantage of a direct solver based on the LU decomposition of the forward modeling operator, we present an inexact full Newton full waveform inversion algorithm that incorporates the simultaneous sources technique. In this chapter, we presented application of full Newton FWI algorithm to the presence of high amplitude multi-reflected waves that are generated by strong velocity contrasts. If there are no strong amplitude multi-reflected waves present in the data, then the full Newton method would not help that much. Sometime, the results would be worse than the other optimization methods. Since the Hessian matrix in the FN method is an indefinite matrix, we suggest that for high resolution image results, the FN method should be performed with preconditoned minimum residual method.

6.1.1 Future work

– There are still many obstacles that hinders the robustness of the FWI algorithm. Primarily, FWI still requires good quality of low-frequency data and also depends on the accuracy of the starting model. These requirements, at the current stage, still remain the main issues in FWI apart from its huge computational demand. The optimization algorithms such as Gauss-Newton and full Newton methods presented in this thesis are still not robust. They are all derived based on local methods, and a robust and less expensive optimization engine is required for any feasible practical application of FWI to large scale 3D problems.

  Detail investigation is required to improve the performance of FWI using a different but proper data functionals, instead of an l_2 norm, that accounts for both the data and model space and accounts for different acquisition geometry. We need to have robust regularization and/or good prior information's to feed the iterative process to make it work. These are also crucial if one wants to do multiparameter inversion. We need also to improve the performances of FWI to higher frequencies.

– The present simultaneous source encoding techniques are based on an l_2 objective function that requires a fixed receiver geometry. One of the limitations of simultaneous sources is that it requires specific acquisition geometry. The method is only applicable to land acquisition survey in which the receivers are spread over for every source. Moreover, in the multi-frequency selection strategies, we have assumed frequency selection strategies that are optimal for a wide-aperture seismic data. Further research is needed to examine the performance of different multi-frequency inversion strategies on data that is poorly represented at low frequencies and/or for receivers spread that are not stationary; maybe by reformulating a new objective function other than an l_2 objective function.

- Extend the application of Gauss-Newton and full Newton to multiparameter FWI inversions in elastic full waveform inversion.

- In Chapter 3, we examined three types of model parameterizations. For better performance and convergence rate, one can consider different types of model parameterization that changes the characteristic nature of the gradient and the Hessian matrix. For example, a combination of density and velocity, a function that dependent on velocity, etc

- Extend the application of Newton-based methods to multiparameter full waveform inversion. Because incorporating the inverse of the Hessian matrix in the numerical inversion mitigates the cross-talk among different classes of parameter. Study the applications and roles of different model parameterizations in multiparameter FWI to 2D/3D elastic FWI. However, it requires thorough studies of sensitivity analysis of FWI with respect to the choice of the model parameterization. One has to find an efficient work flow strategy of model parameterization in multiparameter inversion to account for the different parameter. This also requires a careful analysis of the choice of model parameterization used on the Hessian matrix.

Bibliography

Akcelik, V., G. Biros, and O. Ghattas. "Parallel Multiscale Gauss-Newton-Krylov Methods for Inverse Wave Propagation." *Proceedings of the 2002 ACM/IEEE Conference on Supercomputing*. Los Alamitos, CA, USA: IEEE Computer Society Press, 2002, 1–15.

Alkhalifah, T. and Y. Choi. "From tomography to full-waveform inversion with a single objective function." *Geophysics* 79 (2014): R55–R61.

Amestoy, P., I. Duff, J. L'Excellent, and J. Koster. "A fully asynchronous multifrontal solver using distributed dynamic scheduling." *SIAM Journal on matrix analysis and applications* 23 (2001): 15–41.

Amestoy, P., A. Guermouche, J. L'Excellent, and S. Pralet. "Hybrid scheduling for the parallel solution of linear systems." 3^{rd} Workshop on Parallel Matrix Algorithms and Applications (PMAA 2004), CIRM, Luminy, France, OCT 20-22, 2004. *Parallel Computing* 32 (2006): 136–156.

Anagaw, A. and M. Sacchi. "Full waveform inversion with simultaneous sources using the full Newton Method." *SEG Technical Program Expanded Abstracts* 531 (2012): 1–5.

Anagaw, A. and M. Sacchi. "Comparison of multifrequency selection strategies for simultaneous-source full-waveform inversion." *Geophysics* 79 (2014): R165–R181.

Anagaw, A. and M. Sacchi. "Regularized 2D Acoustic Full Waveform Inversion." 73^{rd} *EAGE Conference & Exhibition, EAGE Expanded Abstracts* (2011): P027.

Anagaw, A. Y. "Total variation and adjoint state methods for seismic wavefield imaging." *MSc. thesis, the University of Alberta* (2009).

Anagaw, A. Y. and M. D. Sacchi. "Edge-preserving seismic imaging using the total variation method." *Journal of Geophysics and Engineering* 9 (2012): 138.

Askan, A., V. Akcelik, J. Bielak, and O. Ghattas. "Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures." *Bulletin Of The Seismological Society Of America* 97 (2007): 1990–2008.

Ben-Hadj-Ali, H., S. Operto, and J. Virieux. "An efficient frequency-domain full waveform inversion method using simultaneous encoded sources." *Geophysics* 76 (2011): R109–R124.

Berenger, J. "A perfectly matched layer for the absorption of electromagnetic-waves." *Journal Of Computational Physics* 114 (1994): 185–200.

Berkhout, A. and G. Blacquiére. "Blended acquisition with dispersed source arrays, the next step in seismic acquisition?." *SEG Technical Program Expanded Abstracts.* 2011, 16–19.

Berkhout, A. J. G., G. Blacquiere, and E. Verschuur. "From simultaneous shooting to blended acquisition." *SEG Technical Program Expanded Abstracts* 27 (2008): 2831–2838.

Bertsekas, D. P. and J. Tsitsiklis. *Neuro-Dynamic Programming.* Har/cdr edition. Athena Scientific, 1996.

Beydoun, W. B. and A. Tarantola. "First Born and Rytov approximations: Modeling and inversion conditions in a canonical example." *The Journal of the Acoustical Society of America* 83 (1988): 1045–1055.

Billette, F. and S. Brandsberg-Dahl. "The 2004 BP velocity benchmark.." 67^{th} *Annual Internat. Mtg., EAGE Expanded Abstracts.* 2005, B035.

Bleibinhaus, F., J. A. Hole, T. Ryberg, and G. S. Fuis. "Structure of the California Coast Ranges and San Andreas Fault at SAFOD from seismic waveform inversion and reflection imaging." *Journal of Geophysical Research: Solid Earth* 112 (2007).

Bohlen, T. and E. H. Saenger. "Accuracy of heterogeneous staggered-grid finite-difference modeling of Rayleigh waves." *Geophysics* 71 (2006): T109–T115.

Bourgeois, A., M. Bourget, P. Lailly, M. Poulet, P. Ricarte, and R. Versteeg. "The Marmousi Experience, in The Marmousi Experience, Proceedings of the 990 EAGE Workshop: Mamousi, model and data." *EAGE Expanded Abstracts* (1991): 5–16.

Brenders, A. J. and R. G. Pratt. "Efficient waveform tomography for lithospheric imaging: implications for realistic, two-dimensional acquisition geometries and low-frequency data." *Geophysical Journal International* 168 (2007): 152–170.

Brenders, A. J., U. Albertin, and J. Mika. "Comparison of 3D time- and frequency-domain waveform inversion: Benefits and insights of a broadband, discrete-frequency strategy." *SEG Technical Program Expanded Abstracts* (2012): 1–5.

Brossier, R., S. Operto, and J. Virieux. "Which data residual norm for robust elastic frequency-domain full waveform inversion?." *Geophysics* 75 (2010): R37–R46.

Bube, K. P. and R. T. Langan. "A continuation approach to regularization of ill-posed problems with application to crosswell-traveltime tomography." *Geophysics* 73 (2008): VE337–VE351.

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

Castellanos-Lopez, C., L. Métivier, S. Operto, and R. Brossier. "Fast Full waveform inversion with source encoding and second order optimization methods." *In* 75^{th} *EAGE Conference & Exhibition incorporating SPE EUROPEC* (2013): We1110.

Cha, Y. H. and C. Shin. "2D Laplace-Fourier-domain full waveform inversion for both velocity and density models: An experience of the 2004 BP velocity-analysis benchmark dataset." *SEG Technical Program Expanded Abstracts* (2009): 2258–2262.

Chabert, J.-L. "A history of algorithms : from the pebble to the microchip." Springer, Berlin, New York: Trad. de Histoire d'algorithmes : du caillou á la puce (1999).

Clayton, R. and R. Stolt. "A Born-WKBJ inversion method for acoustic reflection data." *GEOPHYSICS* 46 (1981): 1559–1567.

Eisenstat, S. C., H. F. Walker, S. C. Eisenstatt, Homer, and F. Walker. "Choosing the Forcing Terms in an Inexact Newton Method." *SIAM J. Sci. Comput* 17 (1994): 16–32.

Fasano, G. and S. Lucidi. "A nonmonotone truncated Newton-Krylov method exploiting negative curvature directions, for large scale unconstrained optimization." *Optimization Letters* 3 (2009): 521–535.

Fichtner, A. and J. Trampert. "Hessian kernels of seismic data functional based upon adjoint techniques." *Geophysical Journal International* 185 (2011): 775–798.

Fischer, B., A. Ramage, D. Silvester, and A. Wathen. "Minimum residual methods for augmented systems." *BIT Numerical Mathematics* 38 (1998): 527–543.

Gardner, G., L. Gardner, and A. Gregory. "Formation velocity and density -The diagnostic basic for stratigraphic traps." *Geophysics* 39 (1974): 770–780.

Gauthier, O., J. Virieux, and A. Tarantola. "Two-dimensional nonlinear inversion of seismic waveforms: Numerical results." *Geophysics* 51 (1986): 1387–1403.

Guanghui, L., H. Jiye, and Y. Hongxia. "Global convergence of the fletcher-reeves algorithm with inexact linesearch." *Applied Mathematics-A Journal of Chinese Universities* 10 (1995): 75–82.

Guitton, A. and E. Daz. "Attenuating crosstalk noise with simultaneous source full waveform inversion." *Geophysical Prospecting* 60 (2012): 759–768.

Herrmann, F. J., Y. A. Erlangga, and T. T. Y. Lin. "Compressive simultaneous full-waveform simulation." *Geophysics* 74 (2009): A35–A40.

Herrmann, F. J., X. Li, A. Y. Aravkin, and T. van Leeuwen. "A modified, sparsity promoting, Gauss-Newton algorithm for seismic waveform inversion." *Wavelets and Sparsity XIV*. Ed. Papadakis, M and VanDeVille, D and Goyal, VK Conference on Wavelets and Sparsity XIV, San Diego, CA, AUG 21-24, 2011, 2011.

Hestenes, M. R. and E. Stiefel. "Methods of conjugate gradients for solving linear systems." *Journal of Research of the National Bureau of Standards* 49 (1952): 409–436.

Hestenes, M. R. "Multiplier and gradient methods." *Journal of Optimization Theory and Applications* 4 (1969): 303–320.

Hu, W., A. Abubakar, T. M. Habashy, and J. Liu. "Preconditioned non-linear conjugate gradient method for frequency domain full-waveform seismic inversion." *Geophysical Prospecting* 59 (2011): 477–491.

Hu, W., A. Abubakar, and T. M. Habashy. "Application of the nearly perfectly matched layer in acoustic wave modeling." *Geophysics* 72 (2007): SM169–SM175.

Hu, W., A. Abubakar, and T. M. Habashy. "Simultaneous multifrequency inversion of full-waveform seismic data." *Geophysics* 74 (2009): R1–R14.

Hustedt, B., S. Operto, and J. Virieux. "Mixed-grid and staggered-grid finite-difference methods for frequency-domain acoustic wave modelling." *Geophysical Journal International* 157 (2004): 1269–1296.

Innanen, K. A. "Seismic AVO and the inverse Hessian in precritical reflection full waveform inversion." *Geophysical Journal International* 199 (2014): 717–734.

Jeong, W. and D.-J. Min. "Application of acoustic full waveform inversion for density estimation." *SEG Technical Program Expanded Abstracts* (2012): 1–5.

Jo, C., C. Shin, and J. Suh. "An optimal 9-point, finite-difference, frequencyspace, 2-D scalar wave extrapolator." *Geophysics* 61 (1996): 529–537.

Kamei, R. and R. G. Pratt. "Inversion strategies for visco-acoustic waveform inversion." *Geophysical Journal International* 194 (2013): 859–884.

Kim, Y., H. Cho, D.-J. Min, and C. Shin. "Comparison of Frequency-Selection Strategies for 2D Frequency-Domain Acoustic Waveform Inversion." *Pure and Applied Geophysics* 168 (2011): 1715–1727.

Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse. "Fast full-wavefield seismic inversion using encoded sources." *Geophysics* 74 (2009): WCC177–WCC188.

Kubota, R., E. Nishiyama, K. Murase, and J. Kasahara. "Traveltime estimation of first arrivals and later phases using the modified graph method for a crustal structure analysis." *Exploration Geophysics* 40 (2009): 105–113.

Kuhn, H. W. and A. W. Tucker. "Nonlinear Programming." *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability.* (1951): 481–492.

Kurzmann, A., A. Przebindowska, D. Köhn, and T. Bohlen. "Acoustic full waveform tomography in the presence of attenuation: a sensitivity analysis." *Geophysical Journal International* 195 (2013): 985–1000.

Lai, Y.-L., W.-W. Lin, and D. Pierce. "Conjugate gradient and minimal residual methods for solving symmetric indefinite systems." *Journal of Computational and Applied Mathematics* 84 (1997): 243 – 256.

Lailly. "The seismic inverse problem as a sequence of before-stack migrations." *Conference on Inverse Scattering: Theory and Application* (1983).

Liu, Q. and Y. Gu. "Seismic imaging: From classical to adjoint tomography." *Tectonophysics* 566-567 (2012): 31–66.

Métivier, L., R. Brossier, J. Virieux, and S. Operto. "Full Waveform Inversion and the Truncated Newton Method." *SIAM Journal on Scientific Computing* 35 (2013): B401–B437.

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

Morton, S. A. and C. C. Ober. "Faster shot-record depth migrations using phase encoding." *SEG Technical Program Expanded Abstracts* (1998): 1131–1134.

Navon, I. M., P. K. H. Phua, and M. Ramamurthy. "Vectorization of Conjugate-gradient Methods for Large-scale Minimization." *Proceedings of the 1988 ACM/IEEE Conference on Supercomputing*. Los Alamitos, CA, USA: IEEE Computer Society Press, 1988, 410–418.

Nocedal, J. and S. J. Wright. *Numerical Optimization*. 2nd edition. Springer, 2006.

Nocedal, J. "Updating Quasi-Newton Matrices with Limited Storage." *Mathematics of Computation* 35 (1980): pp. 773–782.

Operto, S., J. Virieux, J.-X. Dessa, and G. Pascal. "Crustal seismic imaging from multifold ocean bottom seismometer data by frequency domain full waveform tomography: Application to the eastern Nankai trough." *Journal of Geophysical Research: Solid Earth* 111 (2006).

Operto, S., J. Virieux, P. Amestoy, J.-Y. L'Excellent, L. Giraud, and H. B. H. Ali. "3D finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: A feasibility study." *Geophysics* 72 (2007): SM195–SM211.

Osher and E. Fatemi. "Nonlinear total variation based noise removal algorithms." *Physica D* 60 (1992): 259–268.

Plessix, R., G. Baeten, J. de Maag, M. Klaassen, Z. Rujie, and T. Zhifei. "Application of acoustic full waveform inversion to a lowfrequency largeoffset land data set." *SEG Technical Program Expanded Abstracts* (2010): 930–934.

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

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

Pratt, G., C. Shin, and Hicks. "Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion." *Geophysical Journal International* 133 (1998): 341362.

Pratt, R. G. "Frequency-domain elastic wave modeling by finite differences: A tool for crosshole seismic imaging." *Geophysics* 55 (1990): 626–632.

Pratt, R. G. "Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model." *Geophysics* 64 (1999): 888–901.

Pratt, R. G. and R. M. Shipp. "Seismic waveform inversion in the frequency domain; Part 2; Fault delineation in sediments using crosshole data." *Geophysics* 64 (1999): 902–914.

Pratt, R. G. and M. H. Worthington. "Inverse theory applied to multi-source cross-hole tomography." *Geophysical Prospecting* 38 (1990): 287–310.

Prieux, V., R. Brossier, S. Operto, and J. Virieux. "Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the Valhall field. Part 1: imaging compressional wave speed, density and attenuation." *Geophysical Journal International* 194 (2013): 1640–1664.

Ravaut, C., S. Operto, L. Improta, J. Virieux, A. Herrero, and P. Dell'Aversana. "Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: application to a thrust belt." *Geophysical Journal International* 159 (2004): 1032–1056.

Romero, L. A., D. C. Ghiglia, C. C. Ober, and S. A. Morton. "Phase encoding of shot records in prestack migration." *Geophysics* 65 (2000): 426–436.

Saenger, E. H., N. Gold, and S. A. Shapiro. "Modeling the propagation of elastic waves using a modified finite-difference grid." *Wave Motion* 31 (2000): 77 – 92.

Schenk, O. and K. Gartner. "Solving unsymmetric sparse systems of linear equations with PARDISO." International Conference on Computational Science (ICCS 2002), Amsterdam, Netherlands, APR 21-24, 2002. *Future Generation Computer Systems* 20 (2004): 475–487.

Schenk, O., A. Wächter, and M. Hagemann. "Matching-based Preprocessing Algorithms to the Solution of Saddle-Point Problems in Large-Scale Nonconvex Interior-Point Optimization." *Comput. Optim. Appl.* 36 (2007): 321–341.

Shin, C. and W. Ha. "A comparison between the behavior of objective functions for waveform inversion in the frequency and Laplace domains." *Geophysics* 73 (2008): VE119–VE133.

Shin, C. and Y. H. Cha. "Waveform inversion in the Laplace domain." *Geophysical Journal International* 173 (2008): 922–931.

Shin, C. and Y. Ho Cha. "Waveform inversion in the Laplace-Fourier domain." *Geophysical Journal International* 177 (2009): 1067–1079.

Shin, G., S. Jang, and D. Min. "Improved amplitude preservation for prestack depth migration by inverse scattering theory." *Geophysical Prospecting* 49 (2001): 592–606.

Sirgue, L. and R. Pratt. "Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies." *Geophysics* 69 (2004): 231–248.

Sirgue, L., B. Denel, and F. Gao. "Integrating 3D full waveform inversion into depth imaging projects." (2011): 2354–2358.

Sourbier, F., S. Operto, J. Virieux, P. Amestoy, and J.-Y. L'Excellent. "FWT2D: A massively parallel program for frequency-domain full-waveform tomography of wide-aperture seismic data-Part 2 Numerical examples and scalability analysis." *Computers & Geosciences* 35 (2009): 496–514.

Stekl, I. and R. G. Pratt. "Accurate viscoelastic modeling by frequency-domain nite differences using rotated operators." *Geophysics* 63 (1998): 1779–1794.

Stekl, I. and M. R. Warner. "3D Frequency Domain Waveform Inversion-Synthetic Shallow Channel Example.." 69^{th} *Conference & Exhibition, EAGE Expanded Abstracts* (2007): C026.

Tang, Y. "Target-oriented wave-equation least-squares migration/inversion with phase-encoded Hessian." *Geophysics* 74 (2009): WCA95–WCA107.

Tao, Y. and M. Sen. "Frequency-domain full waveform inversion with plane-wave data." *Geophysics* 78 (2013): R13–R23.

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

Tarantola, A. "Linearized inversion of seismic-reflection data." *Geophysical Prospecting* 32 (1984): 998–1015.

Tarantola, A. *Inverse problem theory. Methods for data fitting and model parameter estimation*. Amsterdam: Elsevier, 1987.

Ulrych, T. J. and M. D. Sacchi. *Information-based Inversion and Processing with Applications. Hardbound*. Elsevier, 2005.

Leeuwen, T.van , A. Y. Aravkin, and F. J. Herrmann. "Seismic waveform inversion by stochastic optimization." *International Journal of Geophysics* 2011 (2011).

Leeuwen, T.van and F. J. Herrmann. "Fast waveform inversion without source-encoding." *Geophysical Prospecting* 61 (2013): 10–19.

Vigh, D., J. Kapoor, and H. Li. "Fullwaveform inversion application in different geological settings." *SEG Technical Program Expanded Abstracts* (2011): 2374–2378.

Vigh, D., B. Starr, J. Kapoor, and H. Li. "3D Full waveform inversion on a Gulf of Mexico WAZ data set." *SEG Technical Program Expanded Abstracts* (2010): 957–961.

Vigh, E.,D.and Starr and J. Kapoor. "Developing Earth models with full waveform inversion." *The Leading Edge* 28 (2009): 432–435.

Virieux, J. and S. Operto. "An overview of full-waveform inversion in exploration Geophysics." *Geophysics* 74 (2009): WCC1–WCC26.

Wang, H., S. C. Singh, and H. Calandra. "Integrated inversion using combined wave-equation tomography and full waveform inversion." *Geophysical Journal International* 198 (2014): 430–446.

Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. tekl, L. Guasch, C. Win, G. Conroy, and A. Bertrand. "Anisotropic 3D full-waveform inversion." *Geophysics* 78 (2013): R59–R80.

White, D. J. "Two-Dimensional Seismic Refraction Tomography." *Geophysical Journal International* 97 (1989): 223–245.

Williamson, P. "A guide to the limits of resolution imposed by scattering in ray tomography." *Geophysics* 56 (1991): 202–207.

Wu, R. and M. Toksöz. "Diffraction tomography and multisource holography applied to seismic imaging." *Geophysics* 52 (1987): 11–25.

Zhdanov, M. S. "Geophysical Inverse Theory And Regularization Problems." *Methods in Geochemistry and Geophysics* 36 (2002): 406–415.

APPENDIX A

Finite-difference discretization of acoustic wave-equation

A.1 Finite-difference approximation

The second-order finite-difference(FD) based on staggered-grid discretization of acoustic wave-equation is accurate to 10 grid points per wavelength (Pratt and Worthington, 1990; Pratt, 1990; Bohlen and Saenger, 2006). The resulting discretization is a 5-point or 9-point stencil and is more dispersive than higher-order discretizations (Pratt and Worthington, 1990; Pratt, 1990). On the other hand, the so-called mixed-grid stencils that combines the second-order stencil with a 45^o rotated second-order stencil has shown that the mixed-grid stencil requires less grid points and is accurate to 4 grid points per shortest wavelength for seismic modeling (Jo et al., 1996; Hustedt et al., 2004). The accuracy of the mixed-grid approach is also comparable to that of the fourth-order approximation of the first derivative operator. For better efficiency and accuracy of the seismic wave modeling in frequency-domain, we adopted the 9-point mixed-grid stencil FD approximation. The discretization of the 2D acoustic wave-equation equation in this thesis is, primarily, based on the Jo et al. (1996) and Hustedt et al. (2004) papers. Here, we review the Jo et al. (1996) and Hustedt et al. (2004) works. I show the discretizations of acoustic wave-equation based on a mixed-grid staggered-grid FD approximation. First, we start from the first-order hyperbolic equation of the pressure p and particle velocity $\mathbf{v}$

$$\begin{aligned} \bigtriangledown p(t) &= \rho \frac{\partial \mathbf{v}(\mathbf{t})}{\partial t} \\ \frac{\partial p(t)}{\partial t} &= \kappa \bigtriangledown \cdot \mathbf{v}(\mathbf{t}) + src(t). \end{aligned} \tag{A.1}$$

For now, for simplicity, the dependencies of spatial positions are not written explicitly. If we eliminate the particle velocity in equation A.1, we get the wave-equation in time-domain

$$\frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial}{\partial x}p(t)\right)+\frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial}{\partial z}p(t)\right)-\frac{1}{\kappa}\frac{\partial^2 p(t)}{\partial t^2}=-s(t). \tag{A.2}$$

Following the Fourier transform of equation A.1, we get

$$\begin{aligned} \bigtriangledown p(\omega) &= -i\omega\rho\mathbf{v}(\omega) \\ -i\omega p(\omega) &= \kappa \bigtriangledown \cdot\mathbf{v}(\omega)+src(\omega). \end{aligned} \tag{A.3}$$

If we decompose the particle velocity vector into two components along horizontal and vertical components, equation A.3 becomes

$$\begin{aligned} \frac{-i\omega}{\kappa}p_x(\omega) &= \frac{\partial v_x(\omega)}{\partial x}+s_x(\omega) \\ \frac{-i\omega}{\kappa}p_z(\omega) &= \frac{\partial v_z(\omega)}{\partial x}+s_z(\omega) \\ \frac{\partial p(\omega)}{\partial x} &= -i\omega\rho v_x(\omega) \\ \frac{\partial p(\omega)}{\partial z} &= -i\omega\rho v_z(\omega). \end{aligned} \tag{A.4}$$

Note that $p(\omega)=p_x(\omega)+p_z(\omega)$, $s(\omega)=s_x(\omega)+s_z(\omega)$ and we replace $src(\omega)/\kappa$ by $s(\omega)$; the two components is used merely for the separation of the particle velocity components.

A.1.1 Absorbing boundary condition

To avoid artificial reflections at the boundaries, we implemented the perfectly matched layer (PML) absorbing boundary condition (Berenger, 1994; Hu et al., 2007) (see Figure A.1).

Inside the PML regions, equation A.4 becomes

$$\begin{aligned} \frac{-i\omega}{\kappa}p_x(\omega) &= \frac{\partial v_x(\omega)}{\partial \tilde{x}}+s_x(\omega) \\ \frac{-i\omega}{\kappa}p_z(\omega) &= \frac{\partial v_z(\omega)}{\partial \tilde{x}}+s_z(\omega) \\ \frac{\partial p(\omega)}{\partial \tilde{x}} &= -i\omega\rho v_x(\omega) \\ \frac{\partial p(\omega)}{\partial \tilde{z}} &= -i\omega\rho v_z(\omega), \end{aligned} \tag{A.5}$$

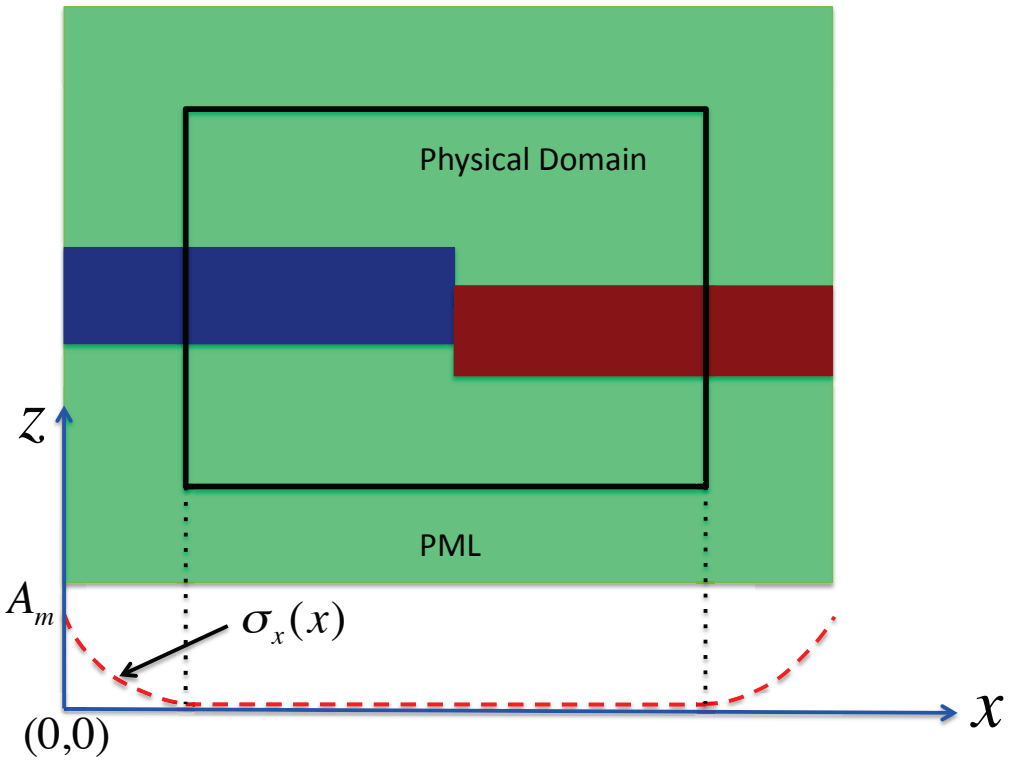


Figure A.1: Perfectly matched layer (PML) absorbing boundary condition.

where $\tilde{x}$ and $\tilde{z}$ are the stretched coordinates. They are defined as

$$\tilde{x}(x) = x + if(x) \quad \& \quad \tilde{z}(z) = z + if(z), \tag{A.6}$$

where $f(x)$ & $f(z)$ are damping functions inside the PML layer and these damping functions in the PML layer are expressed as

$$\tilde{x}(x) = x + \frac{i}{\omega}\int_0^x \sigma_x(\tau)d\tau \quad \& \quad \tilde{z}(z) = z + \frac{i}{\omega}\int_0^z \sigma_z(\tau)d\tau, \tag{A.7}$$

where the σ_x and σ_z are damping profiles inside the PML and are zeros ($\sigma_x = 0$ and $\sigma_z = 0$) inside the physical domain, see Figure A.1.

Upon differentiating equation A.7, we get the following change of variables

$$\frac{\partial}{\partial\tilde{x}} = \frac{1}{\xi_x}\frac{\partial}{\partial x} \quad \& \quad \frac{\partial}{\partial\tilde{z}} = \frac{1}{\xi_z}\frac{\partial}{\partial z}, \tag{A.8}$$

where

$$\begin{aligned} \xi_x &= 1 + i\frac{\sigma_x(x)}{\omega} \\ \xi_z &= 1 + i\frac{\sigma_z(x)}{\omega}. \end{aligned} \tag{A.9}$$

For the damping profile, for example, in the x direction, we used the following (Hu et al., 2007)

$$\sigma(x) = \begin{cases} A_m\left(\frac{L_{pml}-x}{L_{pml}}\right)^3 & 0 \leqslant x \leqslant L_{pml} \\ 0 & L_{pml} < x \leqslant L_{pml} + L_{nx} \\ A_m\left(\frac{x-L_{pml}-L_{nx}}{L_{pml}}\right)^3 & x > L_{pml} + L_{nx} \end{cases} \tag{A.10}$$

where A_m is the decay amplitude factor, L_{pml} and L_{nx} are the thickness of the PML layer and the physical medium along the x-direction (see Figure A.1).

Making use of the stretching coordinates in equation A.8 inside the PML, we reformulate equation A.5 as

$$\begin{aligned}
\frac{-i\omega\xi_x}{\kappa}p_x(\omega) &= \frac{\partial v_x(\omega)}{\partial x} + s_x(\omega) \\
\frac{-i\omega\xi_z}{\kappa}p_z(\omega) &= \frac{\partial v_z(\omega)}{\partial x} + s_z(\omega) \\
\frac{1}{\xi_x}\frac{\partial p(\omega)}{\partial x} &= -i\omega\rho v_x(\omega) \\
\frac{1}{\xi_z}\frac{\partial p(\omega)}{\partial z} &= -i\omega\rho v_z(\omega).
\end{aligned} \tag{A.11}$$

A.2 Staggered-grid finite-difference approximation

Figure A.2 is a schematic illustration of staggered-grid finite-difference of the pressure p, the bulk modulus κ and the particle velocities (v_x and v_z). In the staggered-grid method, the pressure p and the bulk modulus κ are defined at the grid points whereas, the density ρ and the two components of the particle velocities v_x and v_z are defined at half-grid points.

Therefore, using the second-order centered finite-difference approximation, the partial derivatives of the pressure p and particle velocities v_x and v_z are approximated as follows

$$\begin{aligned}
\left[\frac{\partial v_{x_{i,j}}}{\partial x}\right]_x &\approx \frac{v_{x_{i+1/2,j}} - v_x i-1/2,j}{\triangle x} \\
\left[\frac{\partial v_{z_{i,j}}}{\partial z}\right]_z &\approx \frac{v_{z_{i,j+1/2}} - v_{z_{i,j-1/2}}}{\triangle z},
\end{aligned} \tag{A.12}$$

and

$$\begin{aligned}
\left[\frac{\partial p_{i+1/2,j}}{\partial x}\right]_x &\approx \frac{p_{i+1,j} - p_{i,j}}{\triangle x} \\
\left[\frac{\partial p_{i-1/2,j}}{\partial z}\right]_z &\approx \frac{p_{i,j} - p_{i-1,j}}{\triangle z} \\
\left[\frac{\partial p_{i,j+1/2}}{\partial x}\right]_x &\approx \frac{p_{i,j+1} - p_{i,j}}{\triangle x} \\
\left[\frac{\partial p_{i,j-1/2}}{\partial z}\right]_z &\approx \frac{p_{i,j} - p_{i,j-1}}{\triangle z}.
\end{aligned} \tag{A.13}$$

Plugging the above equations A.12 & A.13 into equation A.11, we get the following system of equations

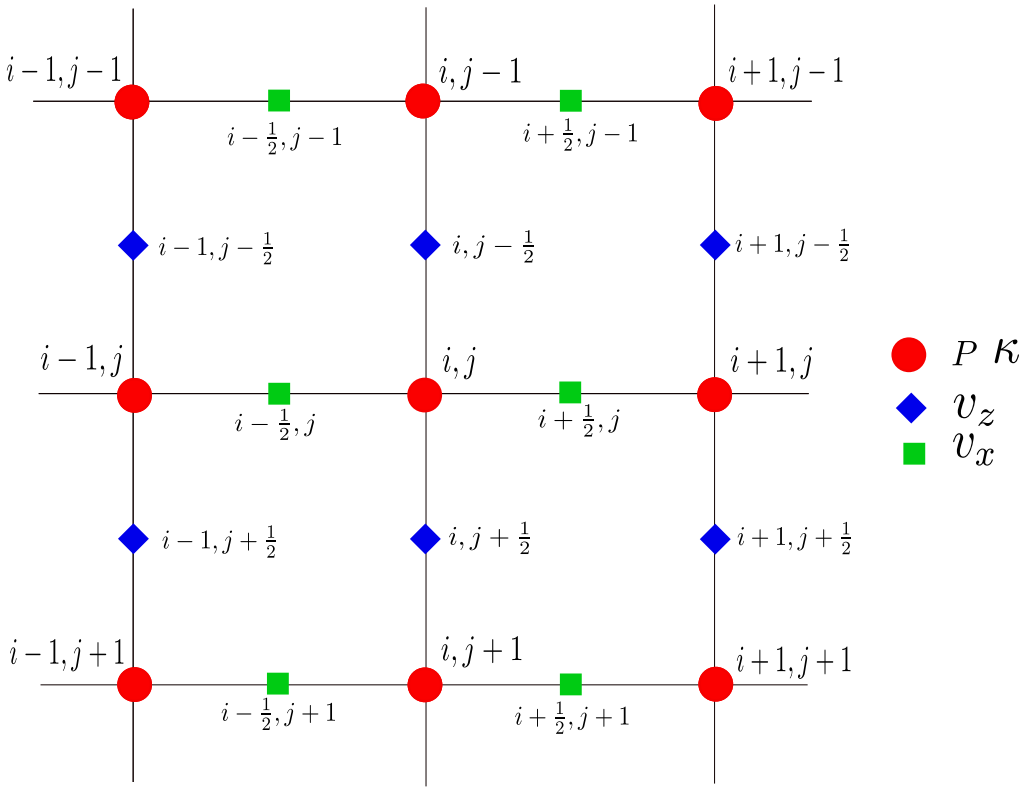


Figure A.2: A 2D schematic illustration of staggered-grid scheme of the pressure, bulk modules and particle velocities, where the spatial direction is along the horizontal and vertical axis. The pressure p and bulk modulus κ are defined at the grid points with a red circle symbol, while the particle velocities v_x and v_z and density (not shown here) are defined at the half-grid points.

$$\begin{aligned}
\frac{-i\omega\xi_{x_i}}{\kappa_{i,j}} p_{x_{i,j}} &= \frac{1}{\triangle x}(v_{x_{i+1/2,j}} - v_{z_{i-1/2,j}}) + s_{x_{i,j}} \\
\frac{-i\omega\xi_{z_j}}{\kappa_{i,j}} p_{z_{i,j}} &= \frac{1}{\triangle z}(v_{z_{i,j+1/2}} - v_{z_{i,j-1/2}}) + s_{z_{i,j}} \\
-i\omega v_{x_{i+1/2},j} &= \frac{1}{\rho_{i+1/2,j}\xi_{x_{i+1/2}}} \frac{1}{\triangle x}(p_{i+1,j} - p_{i,j}) \\
-i\omega v_{x_{i-1/2},j} &= \frac{1}{\rho_{i-1/2,j}\xi_{x_{i-1/2}}} \frac{1}{\triangle x}(p_{i,j} - p_{i-1,j}) \\
-i\omega v_{z_{i,j+1/2}} &= \frac{1}{\rho_{i,j+1/2}\xi_{z_{j+1/2}}} \frac{1}{\triangle z}(p_{i,j+1} - p_{i,j}) \\
-i\omega v_{z_{i,j-1/2}} &= \frac{1}{\rho_{i,j-1/2}\xi_{z_{j-1/2}}} \frac{1}{\triangle z}(p_{i,j} - p_{i,j-1}),
\end{aligned} \tag{A.14}$$

where

$$\begin{aligned}
\frac{1}{\rho_{i+1/2,j}} &= \frac{1}{2}\left(\frac{1}{\rho_{i+1,j}} + \frac{1}{\rho_{i,j}}\right) \\
\xi_{z_{j+1/2}} &= \frac{1}{2}\left(\xi_{z_{j+1}} + \xi_{z_j}\right).
\end{aligned} \tag{A.15}$$

If we eliminate the particle velocities and sum the horizontal and vertical pressure fields ($p = p_x + p_z$), we obtain the following second-order staggered-grid system of equations

$$\begin{aligned}
&\frac{1}{\triangle x^2}\frac{1}{\xi_{x_i}}\frac{1}{\xi_{x_{i+1/2}}\rho_{i+1/2,j}}(p_{i+1,j} - p_{i,j}) - \frac{1}{\triangle x^2}\frac{1}{\xi_{x_i}}\frac{1}{\xi_{x_{i-1/2}}\rho_{i-1/2,j}}(p_{i,j} - p_{i-1,j}) + \\
&\frac{1}{\triangle z^2}\frac{1}{\xi_{z_j}}\frac{1}{\xi_{z_{j+1/2}}\rho_{i,j+1/2}}(p_{i,j+1} - p_{i,j}) - \frac{1}{\triangle z^2}\frac{1}{\xi_{z_j}}\frac{1}{\xi_{z_{j-1/2}}\rho_{i,j-1/2}}(p_{i,j} - p_{i,j-1}) + \frac{\omega^2}{\kappa_{i,j}} p_{i,j} = -s_{i,j}.
\end{aligned} \tag{A.16}$$

The above equation is a 5-point stencil where the coefficients of the 5-point stencil are given by (see Figure A.3)

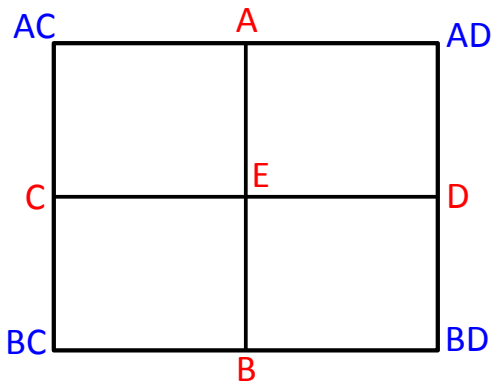


Figure A.3: A schematic illustration of the standard 9-point stencil for 2D cases

$$
\begin{aligned}
\mathbf{A} &= \frac{1}{\triangle z^2}\frac{1}{\xi_{z_j}}\frac{1}{\xi_{z_{j-1/2}}\rho_{i,j-1/2}} \\
\mathbf{B} &= \frac{1}{\triangle z^2}\frac{1}{\xi_{z_j}}\frac{1}{\xi_{z_{j+1/2}}\rho_{i,j+1/2}} \\
\mathbf{C} &= \frac{1}{\triangle x^2}\frac{1}{\xi_{x_i}}\frac{1}{\xi_{z_{i-1/2}}\rho_{i-1/2,j}} \\
\mathbf{D} &= \frac{1}{\triangle x^2}\frac{1}{\xi_{x_i}}\frac{1}{\xi_{z_{i+1/2}}\rho_{i+1/2,j}} \\
\mathbf{E} &= \frac{\omega^2}{\kappa_{i,j}} - \frac{1}{\triangle x^2}\frac{1}{\xi_{x_i}}\frac{1}{\xi_{z_{i+1/2}}\rho_{i+1/2,j}} - \frac{1}{\triangle x^2}\frac{1}{\xi_{x_i}}\frac{1}{\xi_{z_{i-1/2}}\rho_{i-1/2,j}} \\
&\quad - \frac{1}{\triangle z^2}\frac{1}{\xi_{z_j}}\frac{1}{\xi_{z_{j+1/2}}\rho_{i,j+1/2}} - \frac{1}{\triangle z^2}\frac{1}{\xi_{z_j}}\frac{1}{\xi_{z_{j-1/2}}\rho_{i,j-1/2}}.
\end{aligned}
\tag{A.17}
$$

Equation A.16 is called the parsimonions staggered-grid finite-difference stencil for wave-equation in frequency-domain because the formulation is based on the elimination of the particle velocities (Hustedt et al., 2004).

A.3 Rotated staggered-grid finite-difference approximation

The 9-point stencil discretization of the frequency-domain wave-equation proposed by Jo et al. (1996) was built on the discretization of the differential equation along the coordinate

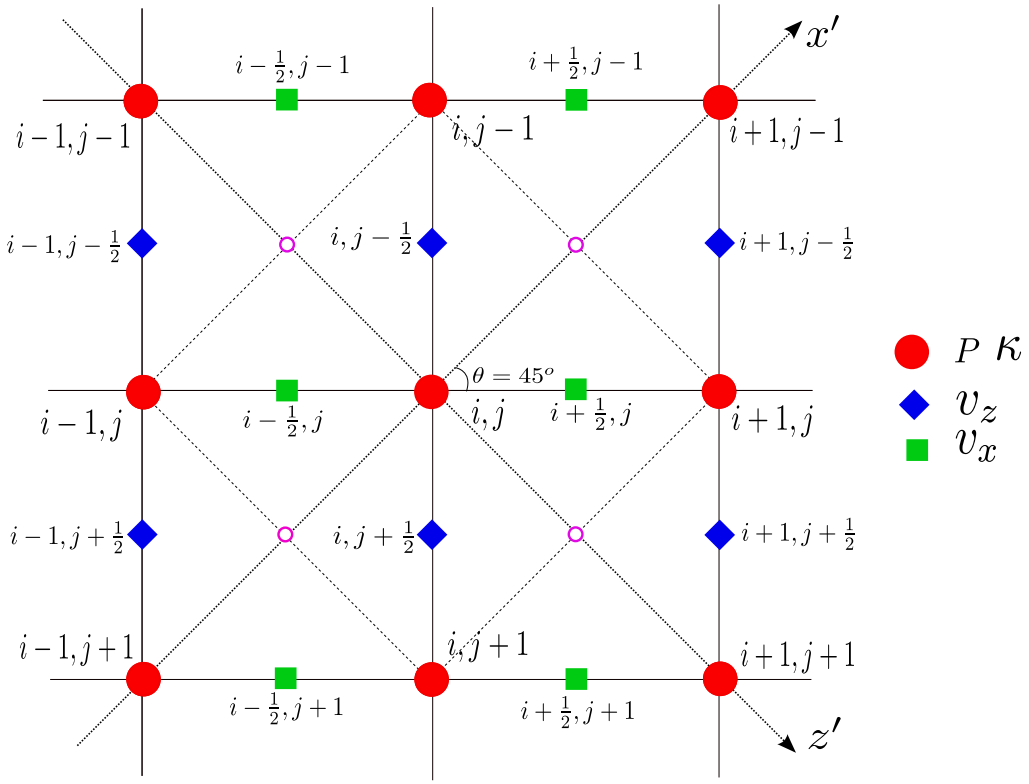


Figure A.4: A 45^o rotated staggered-grid illustration. The particle velocities v_x and v_z and density are moved to the new rotated coordinate system as shown small rose circle.

systems (x, y) through 45^o rotated coordinate systems (x', y') (Saenger et al., 2000; Hustedt et al., 2004). Figure A.4 illustrates the 45^o rotated staggered-grid scheme.

Thus, the differential operators in equation A.11 in the rotated coordinate systems (x', y') through 45^o are expressed as

$$\begin{aligned}
\frac{-i\omega\xi_x}{\kappa} p_x(\omega) &= \frac{\sqrt{2}}{2}\Big\{\frac{\partial v_x(\omega)}{\partial x'} + \frac{\partial v_x(\omega)}{\partial z'}\Big\} + s(\omega) \\
\frac{-i\omega\xi_z}{\kappa} p_z(\omega) &= \frac{\sqrt{2}}{2}\Big\{-\frac{\partial v_z(\omega)}{\partial x'} + \frac{\partial v_z(\omega)}{\partial z'}\Big\} \\
\frac{-i\omega\xi_x}{\rho} v_x(\omega) &= \frac{\sqrt{2}}{2}\Big\{\frac{\partial p(\omega)}{\partial x'} + \frac{\partial p(\omega)}{\partial z'}\Big\} \\
\frac{-i\omega\xi_x}{\rho} v_z(\omega) &= \frac{\sqrt{2}}{2}\Big\{-\frac{\partial p(\omega)}{\partial x'} + \frac{\partial p(\omega)}{\partial z'}\Big\}.
\end{aligned} \tag{A.18}$$

Using coordinate transformation under rotation of 45^o, in the rotated Cartesian coordinate system, the second-order centered staggered-grid FD approximations on the x'-y' plane for the particle velocities are expressed as

$$\begin{aligned}\left[\frac{\partial v_{x_{i,j}}}{\partial x'}\right]_x &\approx \frac{v_{x_{i+1/2,j-1/2}} - v_{x_{i-1/2,j+1/2}}}{\sqrt{2}\triangle x} \\ \left[\frac{\partial v_{z_{i,j}}}{\partial z'}\right]_z &\approx \frac{v_{z_{i+1/2,j+1/2}} - v_{z_{i-1/2,j-1/2}}}{\sqrt{2}\triangle z}.\end{aligned} \tag{A.19}$$

For the pressure and bulk modules, the second-order centered staggered-grid FD approximation is performed at their grid locations. Making use of equations A.19 & A.18, and eliminating the particle velocities, we get the following system of equations (9-point stencil)

$$\begin{aligned}
&\left[\frac{1}{4h^2}\left(\frac{1}{\rho_{i+1/2,j-1/2}}\left(\frac{1}{\xi_{x_i}\xi_{x_{i+1/2}}}+\frac{1}{\xi_{z_i}\xi_{z_{j-1/2}}}\right)+\frac{1}{\rho_{i-1/2,j+1/2}}\left(\frac{1}{\xi_{x_i}\xi_{x_{i-1/2}}}+\frac{1}{\xi_{z_i}\xi_{z_{j+1/2}}}\right)\right.\right.\\
&\left.\left.+\frac{1}{\rho_{i+1/2,j+1/2}}\left(\frac{1}{\xi_{x_i}\xi_{x_{i+1/2}}}+\frac{1}{\xi_{z_i}\xi_{z_{j+1/2}}}\right)+\frac{1}{\rho_{i-1/2,j-1/2}}\left(\frac{1}{\xi_{x_i}\xi_{x_{i-1/2}}}+\frac{1}{\xi_{z_i}\xi_{z_{j-1/2}}}\right)\right)-\frac{\omega}{\kappa_{i,j}}\right]p_{i,j}\\
&-\frac{1}{4h^2}\left[\frac{1}{\rho_{i+1/2,j-1/2}}\left(\frac{1}{\xi_{x_i}\xi_{x_{i+1/2}}}-\frac{1}{\xi_{z_i}\xi_{z_{i-1/2}}}\right)+\frac{1}{\rho_{i+1/2,j+1/2}}\left(\frac{1}{\xi_{x_i}\xi_{x_{i+1/2}}}-\frac{1}{\xi_{z_i}\xi_{z_{i+1/2}}}\right)\right]p_{i+1,j}\\
&-\frac{1}{4h^2}\left[\frac{1}{\rho_{i-1/2,j+1/2}}\left(\frac{1}{\xi_{x_i}\xi_{x_{i-1/2}}}-\frac{1}{\xi_{z_i}\xi_{z_{i+1/2}}}\right)+\frac{1}{\rho_{i-1/2,j+1/2}}\left(\frac{1}{\xi_{x_i}\xi_{x_{i-1/2}}}-\frac{1}{\xi_{z_i}\xi_{z_{i+1/2}}}\right)\right]p_{i-1,j}\\
&+\frac{1}{4h^2}\left[\frac{1}{\rho_{i-1/2,j+1/2}}\left(\frac{1}{\xi_{x_i}\xi_{x_{i-1/2}}}-\frac{1}{\xi_{z_i}\xi_{z_{i+1/2}}}\right)+\frac{1}{\rho_{i+1/2,j+1/2}}\left(\frac{1}{\xi_{x_i}\xi_{x_{i+1/2}}}-\frac{1}{\xi_{z_i}\xi_{z_{i+1/2}}}\right)\right]p_{i,j+1}\\
&+\frac{1}{4h^2}\left[\frac{1}{\rho_{i+1/2,j-1/2}}\left(\frac{1}{\xi_{x_i}\xi_{x_{i+1/2}}}-\frac{1}{\xi_{z_i}\xi_{z_{i-1/2}}}\right)+\frac{1}{\rho_{i-1/2,j-1/2}}\left(\frac{1}{\xi_{x_i}\xi_{x_{i-1/2}}}-\frac{1}{\xi_{z_i}\xi_{z_{i-1/2}}}\right)\right]p_{i,j-1}\\
&-\frac{1}{\rho_{i+1/2,j-1/2}}\frac{1}{4h^2}\left(\frac{1}{\xi_{x_i}\xi_{x_{i+1/2}}}+\frac{1}{\xi_{z_i}\xi_{z_{i-1/2}}}\right)p_{i+1,j-1}\\
&-\frac{1}{\rho_{i-1/2,j+1/2}}\frac{1}{4h^2}\left(\frac{1}{\xi_{x_i}\xi_{x_{i-1/2}}}+\frac{1}{\xi_{z_i}\xi_{z_{i+1/2}}}\right)p_{i-1,j+1}\\
&-\frac{1}{\rho_{i+1/2,j+1/2}}\frac{1}{4h^2}\left(\frac{1}{\xi_{x_i}\xi_{x_{i+1/2}}}+\frac{1}{\xi_{z_i}\xi_{z_{i+1/2}}}\right)p_{i+1,j+1}\\
&-\frac{1}{\rho_{i-1/2,j-1/2}}\frac{1}{4h^2}\left(\frac{1}{\xi_{x_i}\xi_{x_{i-1/2}}}+\frac{1}{\xi_{z_i}\xi_{z_{i-1/2}}}\right)p_{i-1,j-1}=-s(\omega)_{i,j}.
\end{aligned} \tag{A.20}$$

Note that, we assumed a uniform discretization where $\triangle x = \triangle z = h$. The coefficients of the rotated staggered-grid stencil will then be extracted from equation A.20. The above system of equation is a 9-point stencil inside the PML layers and but it's still a 5-point stencil inside the physical domain, where $\xi_x = \xi_z = 1$.

A.4 Mixed-grid stencil

In order to increase the accuracy of the numerical differential operator on the rotated staggered-grid scheme, Jo et al. (1996) proposed a mixed-grid approach by incorporating grid-points in the finite-difference stencil along several directions. The mixed-grid approach combines linearly the standard second-order centered staggered-grid stencils in the reference and the 45^o rotated second-order centered staggered-grid coordinate systems. The basic idea of Jo et al. (1996) work is as follow: First we start from the frequency-domain acoustic wave-equation

$$\frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial}{\partial x}p(\omega)\right) + \frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial}{\partial z}p(\omega)\right) + \frac{\omega^2}{\kappa}p(\omega) = -s(\omega). \tag{A.21}$$

which is equivalent to

$$\nabla^2 p(\omega) + \frac{\omega^2}{\kappa}p(\omega) = -s(\omega). \tag{A.22}$$

The discretization of the first two terms of the differential operators in equation A.21 on the reference Cartesian coordinate system (x, y) and on 45^o rotated coordinate system (x', y') are given by

$$\begin{aligned}
\left[\frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial}{\partial x}p\right)\right]_{i,j} &\approx \frac{1}{\triangle x^2}\left(\frac{1}{\rho_{i+1/2,j}}(p_{i+1,j} - p_{i,j}) - \frac{1}{\rho_{i-1/2,j}}(p_{i,j} - p_{i-1,j})\right) \\
\left[\frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial}{\partial z}p\right)\right]_{i,j} &\approx \frac{1}{\triangle z^2}\left(\frac{1}{\rho_{i,j+1/2}}(p_{i,j} - p_{i+1,j}) - \frac{1}{\rho_{i,j-1/2}}(p_{i,j} - p_{i,j-1})\right),
\end{aligned} \tag{A.23}$$

and

$$\left[\frac{\partial}{\partial x'}\left(\frac{1}{\rho}\frac{\partial}{\partial x'}p\right)\right]_{i,j} \approx \frac{1}{2\triangle x^2}\left(\frac{1}{\rho_{i+1/2,j-1/2}}(p_{i+1,j-1}-p_{i,j}) - \frac{1}{\rho_{i-1/2,j+1/2}}(p_{i,j}-p_{i-1,j+1})\right)$$
$$\left[\frac{\partial}{\partial z'}\left(\frac{1}{\rho}\frac{\partial}{\partial z'}p\right)\right]_{i,j} \approx \frac{1}{2\triangle z^2}\left(\frac{1}{\rho_{i+1/2,j+1/2}}(p_{i+1,j+1}-p_{i,j}) - \frac{1}{\rho_{i-1/,j-1/2}}(p_{i,j}-p_{i-1,j-1})\right). \tag{A.24}$$

Jo et al. (1996) proposed a mixed operator approach to the differential operators of the reference coordinate system (x, y) and the 45^o rotated coordinate system (x', y') with an averaging coefficient a as

$$\nabla^2 p(\omega) = a\nabla^2_{0^o} p(\omega) + (1-a)\nabla^2_{45^o} p(\omega). \tag{A.25}$$

Thus,

$$a\nabla^2_{0^o} p(\omega) + (1-a)\nabla^2_{45^o} p(\omega) + \frac{\omega^2}{\kappa}p(\omega) = -s(\omega). \tag{A.26}$$

Due to the coupling of grid points both in the reference and the rotated coordinate systems, Jo et al. (1996) has shown that the mixed-grid stencil is more dispersive than the second-order staggered-grid stencil, as a result they adopted the finite-element method called a lumped-matrix approach to the term $\frac{\omega^2}{\kappa}$ to approximate the solution of equation A.25 by taking its average values from the neighbouring points

$$\begin{aligned}\left[\frac{\omega^2}{\kappa}p(\omega)\right]_{i,j} \approx \frac{\omega^2}{\kappa}p(\omega)\,[cp_{i,j} + d(p_{i+1,j} + p_{i-1,j} + p_{i,j+1} + p_{i,j-1}) \\ + \frac{1-c-4d}{4}(p_{i+1,j+1} + p_{i-1,j+1} + p_{i+1,j-1} + p_{i-1,j-1})\Big],\end{aligned} \tag{A.27}$$

where the parameters c and d are additional averaging coefficients of the grid for the reference and the 45^o rotated coordinate systems. The optimum coefficient values were obtained from minimization of a function that depends on phase velocities (that provides the least numerical dispersion)(Jo et al., 1996). These value are: $a = 0.5461$, $c = 0.6248$ and $d = 0.09381$. Making use of the above equations with a uniform discretization along the two directions, we get the compact 9-point stencil staggered-grid finite-difference formulation as

$$
\begin{aligned}
&\left[c\frac{\omega^2}{\kappa_{i,j}} - \frac{a}{h^2}\left(\frac{1}{\rho_{i+1/2,j}} + \frac{1}{\rho_{i-1/2,j}} + \frac{1}{\rho_{i,j+1/2}} + \frac{1}{\rho_{i,j-1/2}}\right)\right.\\
&\left.+\frac{a-1}{2h^2}\left(\frac{1}{\rho_{i+1/2,j+1/2}} + \frac{1}{\rho_{i-1/2,j+1/2}} + \frac{1}{\rho_{i+1/2,j-1/2}} + \frac{1}{\rho_{i-1/2,j-1/2}}\right)\right] p_{i,j}\\
&+\left(c\frac{\omega^2}{\kappa_{i,j}} + \frac{a}{h^2}\frac{1}{\rho_{i+1/2,j}}\right) p_{i+1,j} + \left(c\frac{\omega^2}{\kappa_{i,j}} + \frac{a}{h^2}\frac{1}{\rho_{i-1/2,j}}\right) p_{i-1,j}\\
&+\left(c\frac{\omega^2}{\kappa_{i,j}} + \frac{a}{h^2}\frac{1}{\rho_{i,j+1/2}}\right) p_{i,j+1} + \left(c\frac{\omega^2}{\kappa_{i,j}}\frac{a}{h^2}\frac{1}{\rho_{i,j-1/2}}\right) p_{i,j-1}\\
&+\frac{1-a}{h^2}\frac{1}{\rho_{i+1/2,j+1/2}} p_{i+1,j+1} + \frac{1-a}{h^2}\frac{1}{\rho_{i-1/2,j+1/2}} p_{i-1,j+1}\\
&+\frac{1-a}{h^2}\frac{1}{\rho_{i+1/2,j-1/2}} p_{i+1,j-1} + \frac{1-a}{h^2}\frac{1}{\rho_{i-1/2,j-1/2}} p_{i-1,j-1} = -s(\omega)_{i,j}.
\end{aligned}
\tag{A.28}
$$

Consequently, the coefficients of the 9-point stencil are given by (see Figure A.3)

$$
\begin{aligned}
\mathbf{A} &= c\frac{\omega^2}{\kappa_{i,j}} + \frac{a}{h^2}\frac{1}{\rho_{i,j-1/2}}\\
\mathbf{B} &= c\frac{\omega^2}{\kappa_{i,j}} + \frac{a}{h^2}\frac{1}{\rho_{i,j+1/2}}\\
\mathbf{C} &= c\frac{\omega^2}{\kappa_{i,j}} + \frac{a}{h^2}\frac{1}{\rho_{i-1/2,j}}\\
\mathbf{D} &= c\frac{\omega^2}{\kappa_{i,j}} + \frac{a}{h^2}\frac{1}{\rho_{i+1/2,j}}\\
\mathbf{E} &= c\frac{\omega^2}{\kappa_{i,j}} - \frac{a}{h^2}\left(\frac{1}{\rho_{i+1/2,j}} + \frac{1}{\rho_{i-1/2,j}} + \frac{1}{\rho_{i,j+1/2}} + \frac{1}{\rho_{i,j-1/2}}\right)\\
&+\frac{a-1}{2h^2}\left(\frac{1}{\rho_{i+1/2,j+1/2}} + \frac{1}{\rho_{i-1/2,j+1/2}} + \frac{1}{\rho_{i+1/2,j-1/2}} + \frac{1}{\rho_{i-1/2,j-1/2}}\right)\\
\mathbf{AC} &= \frac{1-a}{h^2}\frac{1}{\rho_{i+1/2,j-1/2}}\\
\mathbf{AD} &= \frac{1-a}{h^2}\frac{1}{\rho_{i-1/2,j-1/2}}\\
\mathbf{BC} &= \frac{1-a}{h^2}\frac{1}{\rho_{i-1/2,j+1/2}}\\
\mathbf{BD} &= \frac{1-a}{h^2}\frac{1}{\rho_{i+1/2,j+1/2}}.
\end{aligned}
\tag{A.29}
$$

Finally, using a 9-point stencil representation, the discretized form of the wave-equation results in a sparse matrix $\mathbf{A}$ and the resulting the system of equation is written in compact form as

$$\mathbf{A}(\mathbf{x},\omega)\mathbf{p}(\mathbf{x},\omega) = \mathbf{s}(\mathbf{x},\omega) \tag{A.30}$$

A.5 Implementation of attenuation

One advantage of frequency-domain over time-doming forward modeling is that attenuation in frequency-domain can be easily incorporated in the forward modeling by using a complex velocities, for example (Ravaut et al., 2004)

$$\frac{\omega^2}{\kappa} \rightarrow \frac{\omega^2}{\rho}\left[\frac{1}{v}\left(1+\frac{i}{2Q}sign(\omega)\right)\right]^2, \tag{A.31}$$

where Q quality factor and *sign* is sign function.

Definition

Fourier transform

$$p(\omega) = \int_{-\infty}^{\infty} p(t)e^{-i\omega t}dt \tag{A.32}$$

$$p(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} p(\omega)e^{i\omega t}d\omega \tag{A.33}$$

Hadamard product

The Hadamard product is element by element product of two matrices. For example $\mathbf{A}$ and $\mathbf{B}$ be two $m \times n$ matrices. The Hadamard product is defined as

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \odot \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1n}b_{1n} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2n}b_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1}b_{m1} & a_{m2}b_{m2} & \cdots & a_{mn}b_{mn} \end{pmatrix} \tag{A.34}$$