UNIVERSITY OF CALGARY

Comparisons and Implementations of Least-squares Reverse Time Migration and Full Waveform Inversion in Acoustic Media

by

Lei Yang

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN GEOLOGY AND GEOPHYSICS

CALGARY, ALBERTA

AUGUST, 2019

Abstract

Least-squares reverse time migration (LSRTM) and full waveform inversion (FWI) can both provide an image of the subsurface. In this thesis, I investigate the connections and differences between LSRTM and FWI, both in time and frequency domains in acoustic media. Generally, LSRTM can be treated as the inner loop of the FWI algorithm and solved in the image domain. Different from FWI, LSRTM uses reflection data and requires an accurate operator since linearization implies the operator (Born modeling) is independent of the model (reflectivity). FWI uses diving waves and reflection data and is more robust to velocity errors in the initial model because the operator is itself and changed during the optimization. Comparing implementations domains, LSRTM in the frequency domain is relatively easier to formulate than in the time domain. It has the advantage that the wavefield is solved simultaneously for many shots. Unfortunately, the frequency domain formulation becomes too expensive in 3D surveys in terms of memory requirements. Modern implementations of time domain LSRTM require less computer memory and are more efficient than the frequency domain for the 3D case. A similar situation happens for the FWI problem. One advantage of the frequency domain FWI is the easy formulation for the multigrid method. For this approach, updates of the velocity model start from the low frequency components, which can correct the model in the first several iterations and make the result more accurate.

Acknowledgments

I would like to thank my supervisor, Dr. Daniel Trad for giving me this chance to work with him in this project. From the past two and half years, his inspiration and encouragement support me to finish this project and overcome all the setbacks I encountered. I learned how to be an earnest, diligent and meticulous researcher from him.

I would like to thank to my committee members, Dr. Larry Lines and Dr. Hersh Gilbert for their help and useful comments. I also want to express my appreciation to Dr. Wenyong Pan for the initial ideas and the help in this project. Thanks to Dr. Pengliang Yang for his programs in Madagascar.

I am grateful for the help and valuable discussions from Dr. Yu Geng, Dr. Shahpoor Moradi, Dr. Junxiao Li, Dr. Jian Sun, Scott Keating and all the staff in CREWES.

I also want to thank Huaizhen Chen, Tianze Zhang, Shang Huang, Siming Lv, Xi Liu, Shuai Li and Shuoyu Yao for our friendship and other students in CREWES for making my studies an enjoyable experience. Special thanks to Zhan Niu for our friendship and jokes.

Finally, I would like to thank to my father and grandmother for their love and support during the time I studied abroad. Special thanks to my life partner, Qian Wang. Your love and encouragement help me become what I want to be.

To all the time I've wasted...

Table of Contents

List of Figures and Illustrations

List of Tables

List of Symbols, Abbreviations and Nomenclature

Symbol or abbreviation	Definition
LSRTM	Least-squares reverse time migration
RTM	Reverse time migration
FWI	Full waveform inversion
TD	Time domain
FD	Frequency domain
ρ	Density
κ	Bulk modulus
u	Wavefield
x_s	Shot coordinate
∇	First order spatial derivative
f	Source
v	Velocity
∇^2	Second order spatial derivative
Δx	Horizontal space interval
Δz	Vertical space interval
n_b	Index of the thickness of the sponge boundary
f_m	Dominant frequency
v_0	Smooth part of the velocity model
δv	Singular part of the velocity model
u_0	Background wavefield
δu	Perturbative wavefield
L	Linear Born modeling operator
ω	Angular frequency
$\delta(x)$	Dirac delta function
G	Green's function
n_x	Points in the horizontal direction
n_z	Points in the vertical direction
$\mathbf{A}$	Impedance matrix
$\mathbf{U}$	Wavefield matrix
$\mathbf{f}$	Source matrix
n_s	Number of shots

n_f	Number of frequencies
m_{mig}	Migration image
Δd	Data residual
f_N	Nyquist frequency
$J(m)$	Objective function
$\Phi(m)$	Objective function
d_{obs}	Observed data
d_{syn}	Synthetic data
R	Extraction operator
n_r	Number of receivers
g	Gradient of the objective function
$\mathbf{Re}$	Real part
$\mathbf{Im}$	Imaginary part
$\dagger$	Conjugate transpose
$*$	Complex conjugate
$\bigstar$	Convolution
T	Transpose
H	Hessian matrix
α	Step length
Δm	Searching direction
p	Wavefield
$\star$	Cross-correlation
$R(x,z)$	Reflectivity
$\ddot{s}(x,z,t)$	Second time derivative to s

Chapter 1

Introduction

The seismic methods in exploration geophysics often consist of two parts: the forward problem and the inverse problem. The forward problem is to obtain the seismic wavefields or data from the wave equation and the earth model parameters. Conversely, getting the geological structures or the earth model based on the seismic data and wave equations is the inverse problem. The success of seismic interpretation depends on whether or not we can approximate the true structure from a physical model and effectively solve the inverse problem. Figure 1.1 shows the relationships between model and data. In the following sections, I will introduce the forward and inverse problems and commonly used methods for these problems.

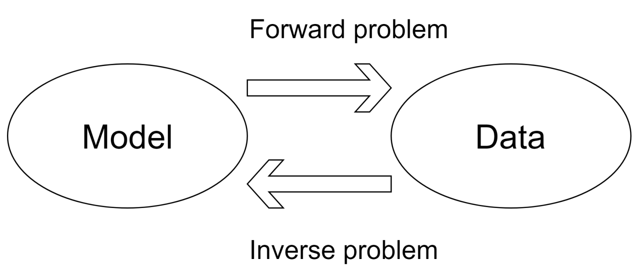


Figure 1.1: Forward problem vs. Inverse problem

1.1 The forward problem

Seismic waves are waves of energy which travel through the earth's layers and the wavefields can be recorded by geophones. The wave equation describes how seismic waves travel through the earth when one has the sources incited on the surface. The seismic waves can be classified as surface waves, which travel at or near the earth's surface and body waves, which travel through the earth. According to the particle motion, there are two types of body waves: primary waves (P-waves) and secondary waves (S-waves). The waves which are mainly discussed in the thesis are P-waves, where the particles oscillate in the same direction as the waves propagate. The forward problem in exploration seismology, as mentioned before, converts the earth model parameters into seismic data by the wave equation in exploration seismology. There are many solutions to wave propagation problems and two important techniques which are used in forward modeling are the finite-difference method and the finite-element method (Kosloff and Baysal, 1982).

In this thesis, the finite-difference methods are used to solve the forward problem, both in time and frequency domains. The solutions to wave propagation problems by finite-difference methods received considerable attention in 1970s (Boore, 1970) and the finite-difference equation formulation for the equations of elasticity was presented and applied to the problem of a layered half-space with a buried point source emitting a compressional pulse (Alterman and Karal Jr, 1968). Alford et al. (1974) discussed the accuracy of finite-difference modeling of the acoustic wave equation and Virieux developed a simulation of P-SV wave modes (Virieux, 1986), followed by a fourth-order accurate space, second-order accurate time, two-dimensional P-SV finite-difference scheme (Levander, 1988). However, although finite-difference method is widely used in seismic wavefield simulation, the accuracy and stability should be considered carefully. A recipe for stability of finite-difference wave equation computations is derived in Lines et al. (1999), which provides proper space and time sampling for the finite-difference computations. Another problem that finite-difference solutions have is artificial boundary reflections. An absorbing boundary condition is developed by Clayton and Engquist (1977), which addresses the wave propagation in a limited area and described physical behavior in an unbounded domain. Because of their efficiency, coding simplicity and easy parallelization in decomposition techniques, these schemes are often developed as time domain methods. Compared with time domain methods, frequency domain finite-difference methods for inversion in seismic exploration started to develop in the early 1990's (Pratt, 1990) but met many difficulties with solving linear systems. For 2D problems, LU decomposition is an effective method to solve the Helmholtz equation with multiple sources. Ajo-Franklin provided a detailed formulation of scalar Helmholtz equation in 2005 (Ajo-Franklin, 2005). In this thesis, I use finite-difference methods to solve the wave equation for seismic inversion and migration, both in the time and frequency domains.

1.2 The inverse problem

The final goal of seismic surveys is to obtain structures or physical parameters of the subsurface consistent with acquired data. Obtaining the structures of subsurface without any physical parameters is seismic migration and reconstructing earth properties, i.e., wave velocity or density, from seismic data is seismic inversion. The inverse problem has been developed during the last decades by many applied mathematicians and geophysicists. Claerbout (1971) proposed a scheme of mapping the reflectors from a single formula involving up and downgoing waves. Lailly (1983) and Tarantola (1984) recognized that the inverse problem could be solved iteratively by calculating a gradient obtained by back-propagating the data residuals and correlating the result with forward-propagated wavefields. This is in fact, equivalent to migration of the data residuals at each iteration. Seismic waveform inversion in the frequency domain was started in the 1990s by Pratt (1999). These techniques that seek to invert seismic data utilizing both phase and amplitude of the seismic traces are known as full-waveform inversion (FWI). In FWI, a full-wave equation modeling is required at each iteration and all wave types can be involved (Virieux and Operto, 2009), although often efficiency and robustness require a simplification of the data and modelling operators. FWI uses steps that are commonly applied in seismic processing/inversion. It can be viewed as an iterative scheme involving forward modeling, pre-stack migration, impedance inversion and velocity updating in each iteration (Margrave et al., 2010).

Seismic migration can be divided into two classes, ray based migration and wave equation based migration. As a wave equation based two-way migration method, reverse time migration (RTM) was proposed in 1980s, but because of its high computational cost it did not become practical until after 2000. Because it is a two-way migration method, it represents a significant improvement over computationally cheaper one way migration algorithms. This characteristic gives RTM higher accuracy than other methods for imaging of steep dips. Because of the progress of computational technology, RTM has become nowadays in one of the most important migration algorithms, in particular for complex structured environments.

In general, migration algorithms can be seen as an approximated inverse for data modelling. In fact, they usually are very similar to what is known as the adjoint operator for the modelling operator (Claerbout, 1992). During seismic processing and inversion, often linear operators are used in geophysical modelling calculations to predict data from models. A common task is to find the inverse of these calculations, i.e., given the observed data, find a model that can predict these data. For linear operators, it is possible to use iterative algorithms which require forward and adjoint operators to solve the inverse problem (Claerbout, 1992). Although in general these adjoint operators are different from the inverse operators, often they approximate the inverse quite well and are more stable and faster.

However, because migration methods are only approximations to inverses, they do not necessarily produce accurate reflectivities, in particular for amplitudes. To compensate these errors, many complicated weighted schemes are often applied. Alternatively, many migration algorithms like RTM can be solved by applying a least-squares scheme known as least-squares migration (LSM). When Born modelling by finite differences is the main engine driving the forward modelling, LSM becomes least-squares RTM (LSRTM).

Least-squares migration (LSM) is an alternative to migration that can, potentially, reduce migration artifacts and improve lateral resolution. It was first proposed by Schuster (1993), and applied for the first time to a real data set by Nemeth et al. (1999). Since then, much research has been done on this topic although its application to real data by industry remains elusive. LSM is often formulated by considering migration as a linear problem, which is a common approximation in seismic processing if multiple reflections are not considered. This approach involves a conceptual separation of the velocity model into a background velocity (keep unchanged during LSM) and a detail velociy variation (reflectivity) which is modified during the optimization. During iterations, LSM seeks to match the observed data by predicted them from the reflectivity which is updated at each iteration.

With this somewhat artificial separation of velocities on background and detail, LSM methods are posed as linear problems. However, strictly speaking, observed data depend non-linearly on earth parameters. FWI does not use this separation so behaves as a nonlinear method. Just like LSM, uses all information in the seismogram to get the earth model but because of its non-linear nature involves complications not seen for LSM. Boths, LSM and FWI involve the minimization of square objective function for the observed and synthetic data but their model definitions are different. Both require to find the minimum of an objective function but they do so with different assumptions. LSM assumes the operator that predicts the data has no errors and tries to decrease the prediction error (cost function) by changing the model (reflectivity). FWI assumes the operator that predicts the data has to be changed to decrease the cost function by modifying the velocities in full. In both cases, model perturbations (reflectivities and velocities respectively) are calculated by some optimization algorithm. The interesting link between these two approaches is that the direction on which corrections have to be applied (gradient) are calculated by migration. In this thesis, I review the theories of LSM for the particular case of LSRTM and FWI and compare them to find the relationship and use numerical examples to illustrate the merits and weakness of each method.

1.3 Thesis overview

This thesis can be summarized in one figure (Figure 1.2) which will become clear after the discussions on the different chapters. I compare and implement LSRTM and FWI in both time and frequency domains. The cross-correlation (I) and deconvolution (I_d) imaging conditions are applied to each RTM scheme with least-squares solutions provided. After this, I implement FWI in the frequency and time domain respectively.

In chapter 2, I described the finite-difference solution of the wave equation in the time domain and compare RTM vs LSRTM images using cross-correlation imaging condition. A conjugate gradient algorithm with forward modelling by time domain Born approximation and adjoint by time domain RTM is used to obtain the LSRTM results.

In chapter 3, I illustrated the formulations of RTM and LSRTM in the frequency domain. The impedance matrix can be solved by the lower-upper (LU) decomposition. I derived the cross-correlating imaging condition in the frequency domain and used the Marmousi model to test the result.

In chapter 4, I connected LSRTM with FWI in the frequency domain and showed the equivalence between the gradient of the FWI objective function and the cross-correlating imaging condition in RTM. Also, I solved LSRTM in the imaging domain based on FWI.

In chapter 5, I gave an example of time domain FWI and compare the two methods of FWI in different domains using the same experiment setup.

In chapter 6, I derived the cross-correlating, illumination compensation and deconvolution imaging condition and compared these different imaging conditions in different domains respectively.

Inverse Problem			
TD LSRTM	**FD LSRTM**	**FD FWI &LSRTM**	**TD FWI**
$R = s \star r$	$R = -\omega^2 RS^*$		
$R_d = \frac{s \star r}{s \star s}$	$R_d = \frac{RS^*}{SS^*}$	$J = \left\|\left\|d_{obs} - d_{syn}\right\|\right\|^2$	$J = \left\|\left\|d_{obs} - d_{syn}\right\|\right\|^2$
$d = Lm$	$d = Lm$	$g = -Hm = I$	$\Delta m = -H^{-1} g$
$J = \left\|\left\|d - Lm\right\|\right\|^2$	$J = \left\|\left\|d - Lm\right\|\right\|^2$	$m = -H^{-1} g$	$m_{n+1} = m_n + \alpha \Delta m$
$m = (L^T L)^{-1} L^T d$	$m = (L^T L)^{-1} L^T d$		

Figure 1.2: The roadmap of the thesis.

Chapter 2

Reverse time migration in the time domain

2.1 Overview

Reverse time migration (RTM) was introduced in 1980s and is now widely used in seismic exploration for subsalt imaging of reflectors. As a two-way migration method, it has more accuracy on reflector amplitudes and imaging of steep flanks. RTM can be implemented both in the time and frequency domains. In this chapter, I will illustrate the solutions to the acoustic wave equation in the time domain and discuss the function of the adjoint operator in migration. The least-squares solution to the RTM in the time domain is shown in the final section of the chapter.

2.2 Acoustic wave equation and solution

2.2.1 Finite-difference method

Acoustic waves are a special case of seismic waves, which travel with the speed of sound in the medium. As a kind of longitudinal waves, the particles oscillate along the same direction of the propagation. The acoustic wave equation with variable density $\rho(\mathbf{x})$ and bulk modulus $\kappa(\mathbf{x})$ in the time domain is

$$\nabla \cdot (\frac{1}{\rho(\mathbf{x})}\nabla u(\mathbf{x}, t; \mathbf{x_s})) - \frac{1}{\kappa(\mathbf{x})}\frac{\partial^2 u(\mathbf{x}, t; \mathbf{x_s})}{\partial t^2} = -f(\mathbf{x}, t; \mathbf{x_s}) \tag{2.1}$$

where $u(\mathbf{x}, t; \mathbf{x_s})$ is the pressure wavefield, with shot index $\mathbf{x_s}$, ∇ is the first order spatial derivative and $f(\mathbf{x}, t; \mathbf{x_s})$ is the source term. If we change the parameterization to wave velocity with

$$v(\mathbf{x}) = \sqrt{\frac{\kappa(\mathbf{x})}{\rho(\mathbf{x})}}, \tag{2.2}$$

and assume the density is constant, we obtain the 2D acoustic wave equation with constant density as

$$\nabla^2 u(\mathbf{x}, t; \mathbf{x_s}) - \frac{1}{v^2(\mathbf{x})}\frac{\partial^2 u(\mathbf{x}, t; \mathbf{x_s})}{\partial t^2} = -f(\mathbf{x}, t; \mathbf{x_s}) \tag{2.3}$$

where $\nabla^2 = \nabla \cdot \nabla = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}$ in 2D case with velocity term $v^2(\mathbf{x})$. Based on equation (2.3), when we know the velocity of the medium and the source function, we can solve for the pressure wavefield $u(\mathbf{x}, t; \mathbf{x_s})$. There are many solutions to wave propagation problems and two important techniques which are used in forward modeling are the finite-difference method and the finite-element method. Here I use the finite-difference method to solve for equation (2.3). If we write $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}$ in equation (2.3) with one shot

$$\frac{\partial^2 u(\mathbf{x}, t)}{\partial x^2} + \frac{\partial^2 u(\mathbf{x}, t)}{\partial z^2} - \frac{1}{v^2(\mathbf{x})}\frac{\partial^2 u(\mathbf{x}, t)}{\partial t^2} = -f(\mathbf{x}, t), \tag{2.4}$$

the two spatial derivatives and the time derivative can be discretized into

$$\frac{\partial^2 u(\mathbf{x}, t)}{\partial x^2} = \frac{u(x + \Delta x, z, t) - 2u(x, z, t) + u(x - \Delta x, z, t)}{\Delta x^2} + O(\Delta x^2) \tag{2.5}$$

$$\frac{\partial^2 u(\mathbf{x},t)}{\partial z^2} = \frac{u(x,z+\Delta z,t) - 2u(x,z,t) + u(x,z-\Delta z,t)}{\Delta z^2} + O(\Delta z^2) \tag{2.6}$$

$$\frac{\partial^2 u(\mathbf{x},t)}{\partial t^2} = \frac{u(x,z,t+\Delta t) - 2u(x,z,t) + u(x,z,t-\Delta t)}{\Delta t^2} + O(\Delta t^2) \tag{2.7}$$

using a 5-point scheme(Figure 2.1) so that the wavefield at the time $t+\Delta t$ can be computed by the wavefield at time t and $t-\Delta t$. For more accuracies, the spatial derivative can be also discretized by the 9-point scheme(Figure 2.2):

$$\begin{aligned}\frac{\partial^2 u(\mathbf{x},t)}{\partial x^2} &= \frac{1}{12\Delta x^2}[-u(x+2\Delta x,z,t) + 16u(x+\Delta x,z,t) - 30u(x,z,t) \\ &\quad + 16u(x-\Delta x,z,t) - u(x-2\Delta x,z,t)] + O(\Delta x^4)\end{aligned} \tag{2.8}$$

$$\begin{aligned}\frac{\partial^2 u(\mathbf{x},t)}{\partial z^2} &= \frac{1}{12\Delta z^2}[-u(x,z+2\Delta z,t) + 16u(x,z+\Delta z,t) - 30u(x,z,t) \\ &\quad + 16u(x,z-\Delta z,t) - u(x,z-2\Delta z,t)] + O(\Delta z^4).\end{aligned} \tag{2.9}$$

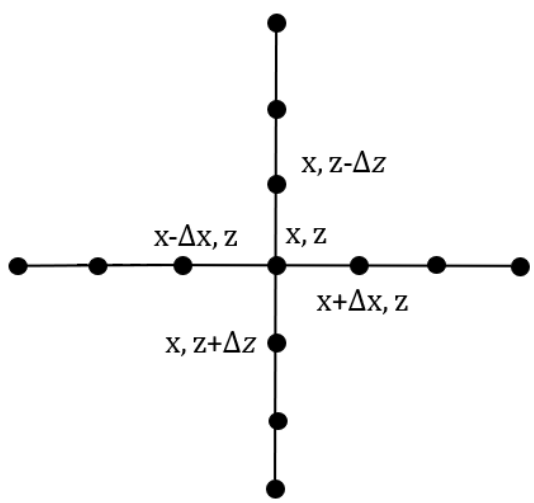


Figure 2.1: A stencil of 5-point finite-difference scheme

However, the finite-difference operators are only approximations to the partial derivatives and with the increasing of time, the errors accumulate and produce unstabilities resulting in amplitudes growing without bound. According to Lines et al. (1999), a criterion for stability

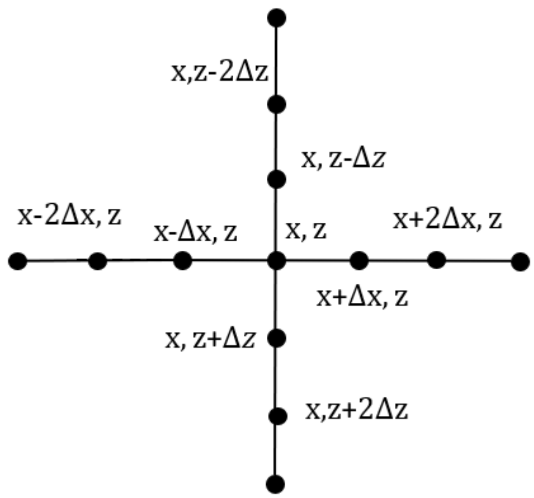


Figure 2.2: A stencil of 9-point finite-difference scheme

is given by the condition

$$\frac{v_{max}\Delta t}{h} \leqslant \frac{2}{\sqrt{a}} \tag{2.10}$$

where v_{max} is the maximum velocity of the model, a is the sum of the absolute values of the weights of the spatial derivatives. For simplicity, it is common to set $\Delta x = \Delta z = h$. With this assumption, the second-order partial derivative in the 2D case using a 5-point scheme $(1, -2, 1)$ has the following stability condition:

$$\frac{v_{max}\Delta t}{h} \leqslant \frac{2}{\sqrt{2(1+2+1)}} = \sqrt{\frac{1}{2}}, \tag{2.11}$$

For a 9-point scheme, we obtain:

$$\frac{v_{max}\Delta t}{h} \leqslant \frac{2}{\sqrt{2[\frac{1}{12}(1+16+30+16+1)]}} = \sqrt{\frac{3}{8}}. \tag{2.12}$$

Another phenomenon that is common to see in the finite-difference method is dispersion produced by coarse grid spacing. If the cell size is too large in comparison with the wavelength of the source, waves will disperse with increasing travel time. Usually, for the second order approximation, a minimum of 10 cells are needed for one wavelength, which can be computed by the wave velocity and the dominant frequency of the wavelet.

2.2.2 Boundary conditions

Boundary conditions are a set of additional constraints for a differential equation. We assume the Earth as a half space so there are no reflectors at the model boundaries. Therefore, it is necessary to develop boundary conditions which eliminate the boundary reflections. The absorbing boundary condition was proposed by Clayton and Engquist (1977) and Engquist and Majda (1977). Following Yang (2014), in the time domain wavefield simulation in this thesis I use the sponge absorbing boundary condition, which attenuates the reflections exponentially in the extended artificial boundary (Cerjan et al., 1985) using the factor

$$d = \exp(-[0.015(n_b - i)^2]) \tag{2.13}$$

where d is the decaying factor, n_b is the thickness of the sponge boundary and i is the index of the boundary grids. When $i = n_b$, the grid is in the inner side of the extended artificial boundary where there is no attenuation of the amplitude and when $i = 0$, the grid is in the outer side of the extended artificial boundary and the attenuation reaches maximum. Figure 2.3 shows a scenario with artificial boundaries. $A_1A_2A_3A_4$ is the original model size and $B_1B_2B_3B_4$ is the extended artificial boundary.

2.2.3 The source function and wavefield examples

A Ricker wavelet can be used as the source term :

$$f(t) = (1 - 2\pi^2 {f_m}^2 t^2) e^{-\pi^2 {f_m}^2 t^2} \tag{2.14}$$

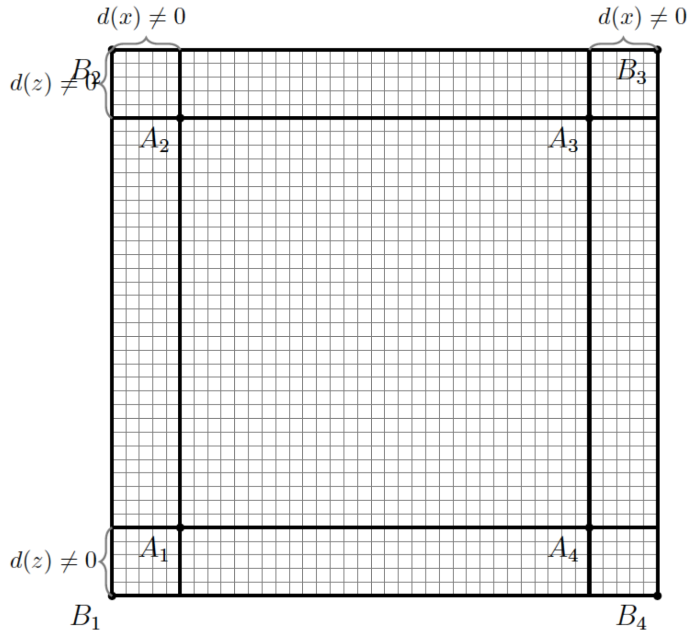


Figure 2.3: The sponge absorbing boundary condition (Yang, 2014)

where t is the time, which can be discretized as a time series and f_m is the dominant frequency of the wavelet. Figure 2.4 is an example of the Ricker wavelet with dominant frequency $f_m = 15$Hz.

Here is an example of wavefield simulation by the 5-point scheme at $t = 0.2s$ and $t = 0.4s$.

2.3 Born modeling and RTM

Under the Born approximation, the velocity model can be split into a smooth part v_0 and a singular part δv

$$v = v_0 + \delta v \tag{2.15}$$

and correspondingly, the acoustic wavefield can be split as

$$u = u_0 + \delta u. \tag{2.16}$$

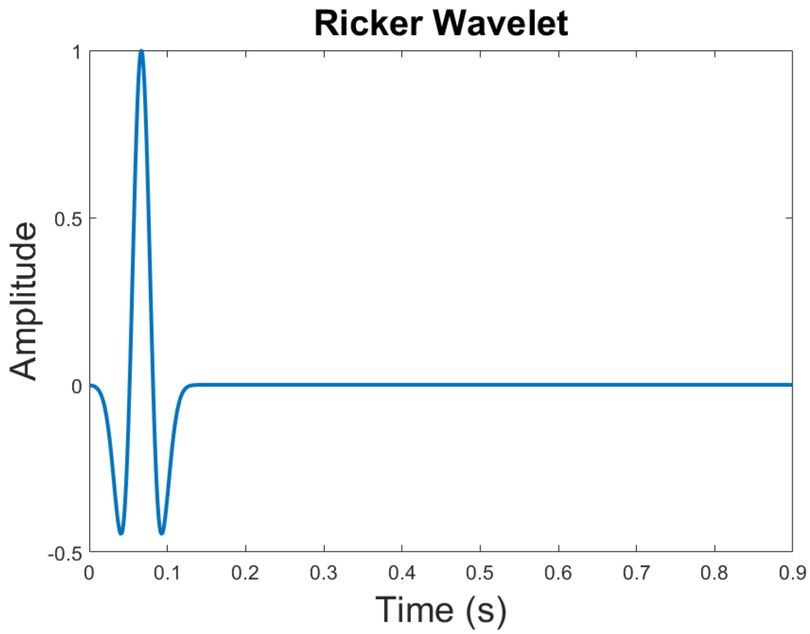


Figure 2.4: Ricker wavelet with $f_m = 15$Hz

v_0 represents the long-wavelength smooth background velocity model and δv is the short-wavelength velocity model, which contains the singular features. This short-wavelength component will produce reflections and contains high resolution features, which is the main object of LSRTM. By u I denote the wavefield generated by the true model v, and u_0 is the background wavefield generated by the long-wavelength components v_0.

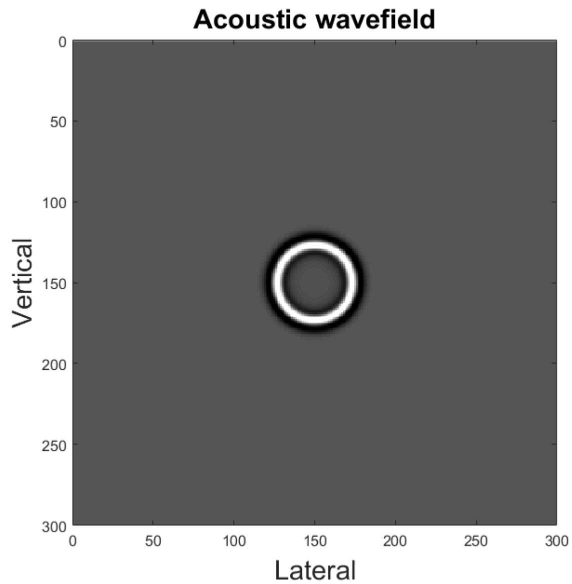
Using the Taylor's expansion, we have

$$\frac{1}{v^2(x)} = \frac{1}{[v_0(x) + \delta v(x)]^2} = \frac{1}{v_0^2(x)} - \frac{2\delta v(x)}{v_0^3(x)} + O(\delta v^2). \tag{2.17}$$

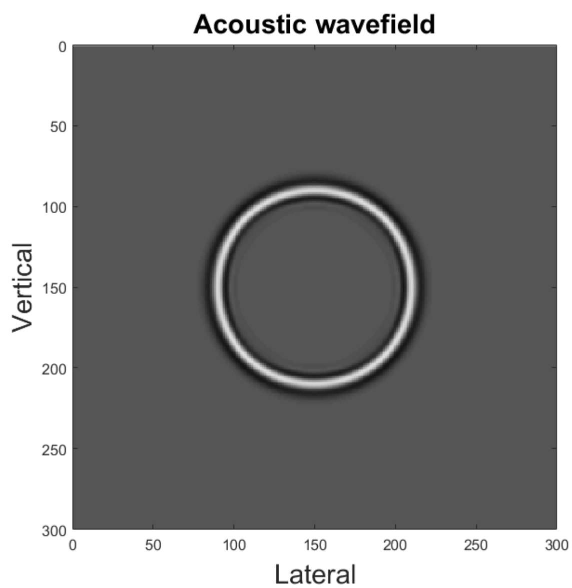
If we neglect the higher order term $O(\delta v^2)$, and substitute equations (2.16) and (2.17) into equation (2.3), we have

$$\left[(\frac{1}{v_0^2(\mathbf{x})} - \frac{2\delta v(x)}{v_0^3(\mathbf{x})}) \frac{\partial^2}{\partial t^2} - \nabla^2 \right] [u_0(\mathbf{x}, t; \mathbf{x_s}) + \delta u(\mathbf{x}, t; \mathbf{x_s})] = f(\mathbf{x}, t; \mathbf{x_s}) \tag{2.18}$$

Rearranging equation (2.18) and considering that $\delta u(\mathbf{x}, t; \mathbf{x_s})$ is weak, we have the incident



(a) The wavefield at $t = 0.2s$



(b) The wavefield at $t = 0.4s$

Figure 2.5: An example of wavefields by 5-point finite-difference scheme

wavefield

$$\left[\frac{1}{v_0^2(\mathbf{x})}\frac{\partial^2}{\partial t^2} - \nabla^2\right] u_0(\mathbf{x}, t) = f(\mathbf{x}, t; \mathbf{x_s}) \tag{2.19}$$

and scattered wavefield

$$\left[\frac{1}{v_0^2(\mathbf{x})}\frac{\partial^2}{\partial t^2} - \nabla^2\right]\delta u(\mathbf{x},t) = \frac{2\delta v(\mathbf{x})}{v_0^3(\mathbf{x})}\frac{\partial^2}{\partial t^2}[u_0(\mathbf{x},t;\mathbf{x_s}) + \delta u(\mathbf{x},t;\mathbf{x_s})] \tag{2.20}$$

where $\frac{2\delta v(\mathbf{x})}{v_0^3(\mathbf{x})}$ represents reflectivity scaled by the background velocity. We can see in this equation how the scattered wavefield can be obtained by the action of the second time derivative of the incident wavefield acting on the reflectivity. If we assume δv is small, then the perturbative wavefield δu generated by δv is also small and the term $\frac{2\delta v(\mathbf{x})}{v_0^3(\mathbf{x})}\frac{\partial^2}{\partial t^2}\delta u(\mathbf{x},t;\mathbf{x_s})$ on the source side of equation (2.20) is much smaller than $\frac{2\delta v(\mathbf{x})}{v_0^3(\mathbf{x})}\frac{\partial^2}{\partial t^2}u_0(\mathbf{x},t;\mathbf{x_s})$. Under the Born approximation, the scattered wavefield is considered very weak. Therefore,

$$\left[\frac{1}{v_0^2(\mathbf{x})}\frac{\partial^2}{\partial t^2} - \nabla^2\right]\delta u(\mathbf{x},t) \approx \frac{2\delta v(\mathbf{x})}{v_0^3(\mathbf{x})}\frac{\partial^2}{\partial t^2}u_0(\mathbf{x},t;\mathbf{x_s}). \tag{2.21}$$

For the incident wavefield u_0, the wavefield is generated by the source $f(\mathbf{x},t;\mathbf{x_s})$ in the background model $v_0(\mathbf{x})$ while the source changes to $\frac{2\delta v(\mathbf{x})}{v_0^3(\mathbf{x})}\frac{\partial^2}{\partial t^2}u_0(\mathbf{x},t;\mathbf{x_s})$ for the scattered wavefield in the same background model. This process of calculating the perturbative wavefield δu is Born modeling.

Now let us look at the RTM process. According to Claerbout (1971), reflectors exist at the points where the first arrival of the downgoing waves is the same as the upgoing waves. In other words, on a reflector downgoing and upgoing waves are coincident.This is called the imaging condition. RTM migrates the data by cross-correlating two wavefields at the same depth level at a given time, one is in the forward recursion and the other one in backward recursion (Symes, 2007). Therefore, a commonly used imaging condition is

$$m_{mig}(\mathbf{x}) = \sum_{n_s}\sum_{n_t} d(\mathbf{x},t)u(\mathbf{x},T-t), \tag{2.22}$$

where n_s is the number of shots, n_t is the number of time series, $d(\mathbf{x},t)$ is the downgoing wavefield and $u(\mathbf{x},T-t)$ is the upgoing wavefield. Equation (2.22) represents the cross-correlation imaging condition. We will illustrate this using the 2D velocity model in Figure 2.6. The model is discretized using 300 points in the vertical direction and 1000 points in the lateral. The grid spacing is 8m for both directions. The velocity of the upper layer is 3000m/s and the lower layer is 4000m/s and we assume the density is constant.

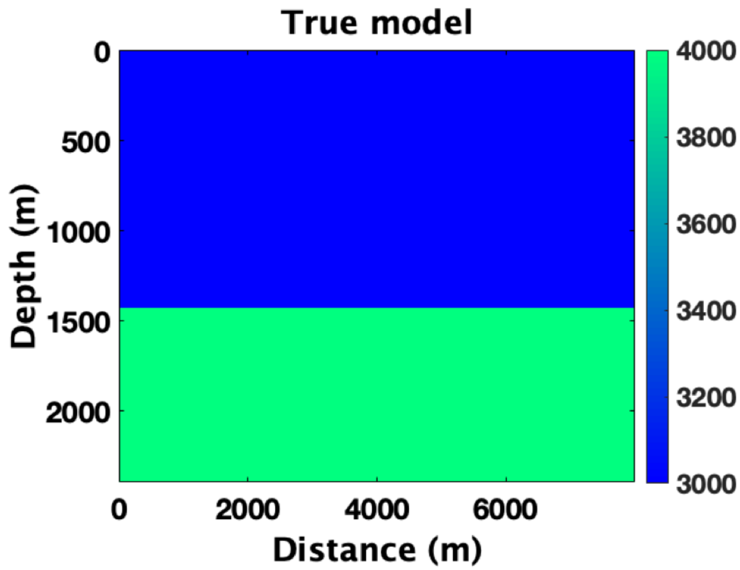


Figure 2.6: The 2-layer model

We set 334 receives with spacing 24m and 10 shots with spacing 800m on the surface. A Ricker wavelet with a dominant frequency $f_m = 20$Hz is used as the source. The RTM result is shown in Figure 2.7. The reflector is constructed by the cross-correlation of the 2 wavefields, downgoing from the source side and backpropagated upgoing from the receiver side.

The image suffers from some low frequency artifacts but this can be solved, among other techniques, by applying a Laplacian filter $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}$ to the result. This process is described as

$$I(x, z) = \frac{\partial^2 I(x, z)}{\partial x^2} + \frac{\partial^2 I(x, z)}{\partial z^2}, \tag{2.23}$$

where $I(x, z)$ is the image to which we want to apply filtering. The filtered result is shown in Figure 2.8. We can see that low frequency artifacts are supressed although certainly some

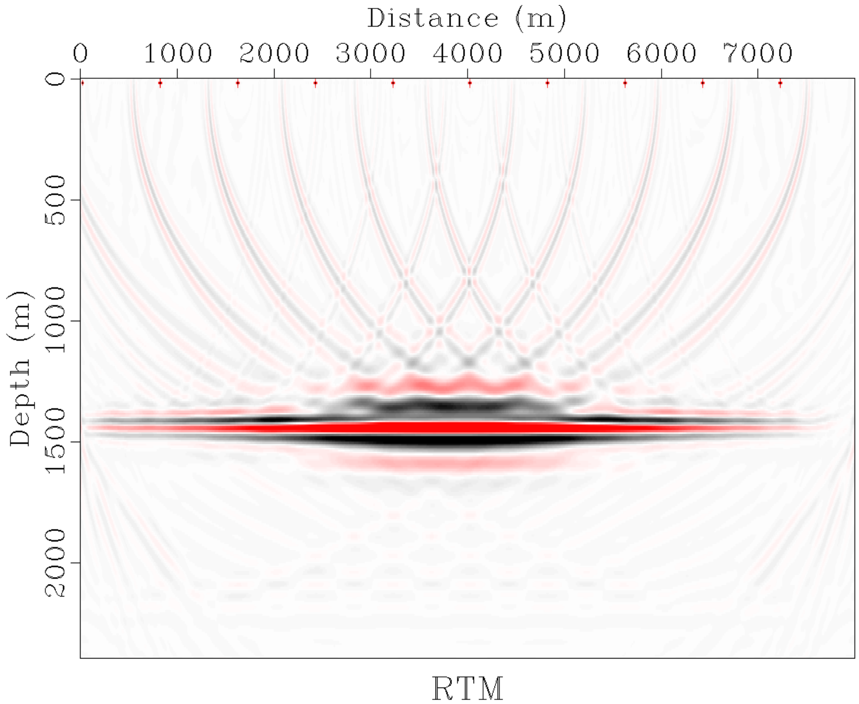


Figure 2.7: RTM result of 2-layer model

low frequency signal will also be lost. That is why much research is ongoing on the field of seismic imaging to develop other imaging conditions that can attenuate the low frequency artifacts without signal loss.

2.4 The least-squares process

The goal of the least-squares scheme is to find a reflectivity model $m(\mathbf{x})$ that can predict the data with minimum error. In the case of least-squares we use an l_2 norm to measure the size of the prediction error. This error is quantized in a cost function for which we try to find a minimum.

Before discussing the algorithm we use to find the minimum of the cost function, I will introduce the adjoint operator which is required. In LSM we use a linear operator to predict the data from a reflectivity model. The operation adjoint to multiplying by a matrix is multiplying by the transposed matrix. If a matrix has complex element, the conjugate

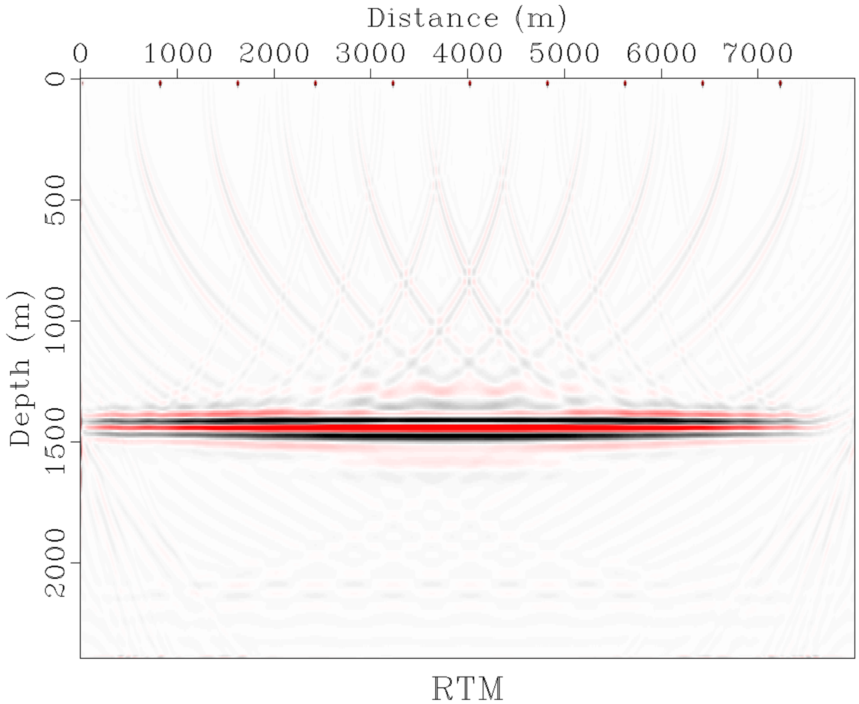


Figure 2.8: RTM result of 2-layer model after Laplacian

transpose is needed (Claerbout, 1992). In mathematical form, an adjoint operation could be:

$$y = Bx \tag{2.24}$$

$$\tilde{x} = B'y \tag{2.25}$$

where B' is the conjugate transpose of B.

We could get the model by solving the inverse operator of this process. However, in general the inverse of the operator can be difficult to obtain. Fortunately, we could use adjoint operator to get approximate models. The objective function of LSRTM is

$$\phi(m) = ||d - Lm||^2 \tag{2.26}$$

where L is the linear Born modeling operator, which is adjoint to the migration operator. The workflow for the least-squares scheme is shown in Figure 2.9. The observed data would generally be obtained by acquisition but in this example are simulated by finite differences from the true model. They can be migrated by the RTM operator L^T to get the reflectivity model Δm. Using the Born modeling operator L, the reflection data is computed. By iterative minimizing the data residual and stacking all images for every iteration, we produce the least-squares image. In Figure 2.9, the background model m_0, which is usually a smoothed model, is used to generate forward and backward propagating wavefields during each iteration. The loop is marked as the red box with a dash line. Table 2.1 shows the first 3 iterations of the least-squares process.

Table 2.1: The least-squares process of the first 3 iteration

$m_1 = L^T d = L^T d(I - LL^T)^0$
$d_1 = Lm_1 = LL^T d$
$r_1 = d - d_1 = (I - LL^T)d$
$m_2 = L^T r_1 = L^T d(I - LL^T)^1$
$d_2 = Lm_2 = LL^T d(I - LL^T)$
$r_2 = r_1 - d_2 = (I - LL^T)^2 d$
$m_3 = L^T r_2 = L^T d(I - LL^T)^2$
$\vdots$

The Neumann series is $\sum_{k=0}^{\infty} A^k$, where A is a bounded linear operator in the space X (Schuster, 2017). The Neumann series has the property:

$$(I - A)^{-1} = \sum_{k=0}^{\infty} A^k \tag{2.27}$$

where I is the identity operator in X. The result of the least-squares algorithm is

$$m = m_1 + m_2 + m_3 + \cdots = L^T d \sum_{k=0}^{\infty} (I - L^T L)^k. \tag{2.28}$$

According to the property of Neumann series, this process can be treated as $m = (L^T L)^{-1} L^T d$, which can provide a better result of reflectivity m.

I use the Marmousi model (Versteeg, 1994) and the implementation by Trad (2018) written in the open source software Madagascar (Fomel et al., 2012), as an example to show the time domain LSRTM scheme. Figure 2.10 shows the true velocity model. There are

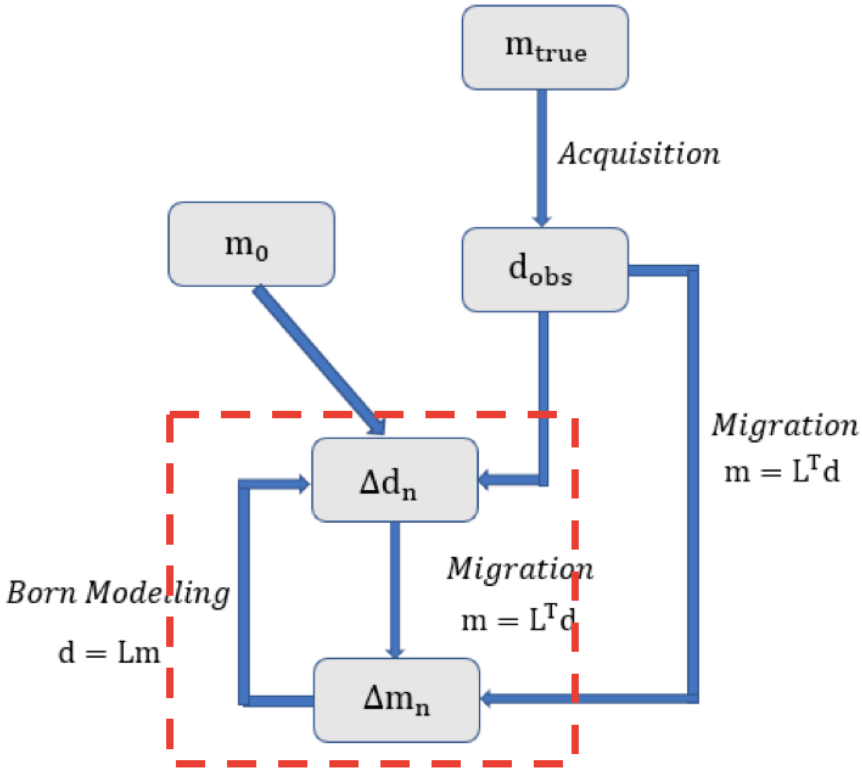


Figure 2.9: LSRTM work flow in the time domain

351 points in the vertical and 1151 points in the horizontal direction and the grid spacing is 8m. The velocity ranges from 1500m/s to 5500m/s. I set 384 receivers and 12 shots on the surface of the ground. The filtered result of RTM is showed in Figure 2.11 and LSRTM result is in Figure 2.12. After 6 iterations, the resolution is improved in the LSRTM image and the reflector in more continuous and accurate compared with the RTM result.

2.5 Conclusions

In chapter 2, I show the solution to the wave equation in an acoustic medium with constant density using the finite-difference method. This is the basic process of implementing the time domain RTM. After this, I describe the Born modeling and RTM process. RTM uses data residual as the source and back propagates the data residual and cross-correlates with the forward propagated source wavefield to get reflectivity. Born modeling uses reflectivity

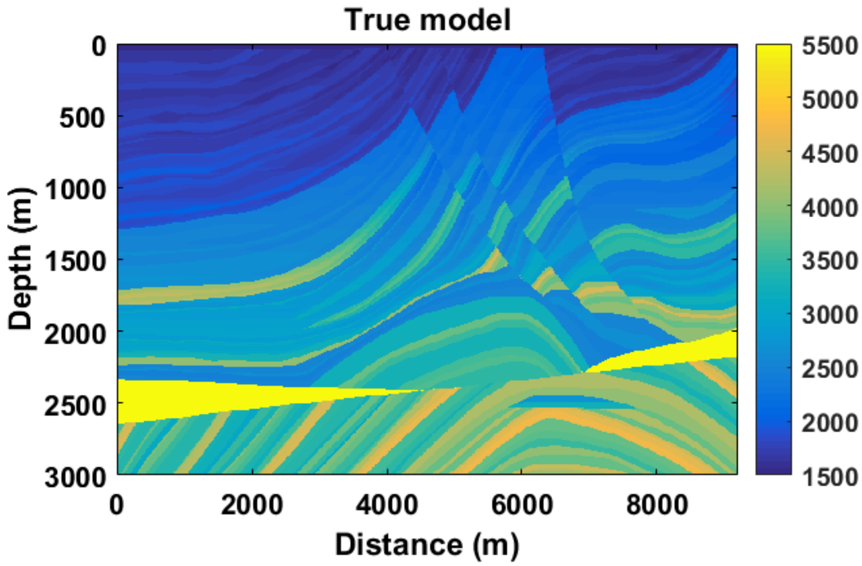


Figure 2.10: The marmousi model

to cross-correlate with the source wavefield such that the reflection data can be obtained at the surface to calculate data residual. I use a flowchart to illustrate this iterative process and use Marmousi model as the example of LSRTM. The image after the least-square scheme is improved compared with the RTM result.

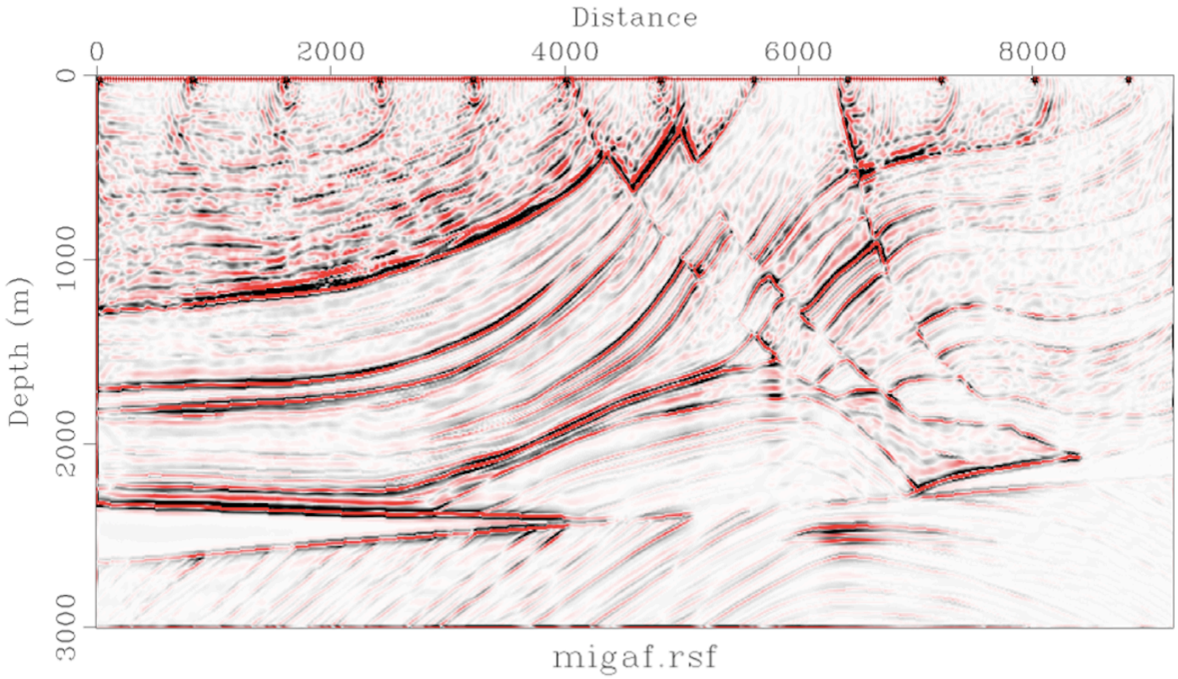


Figure 2.11: The RTM result of Marmousi model

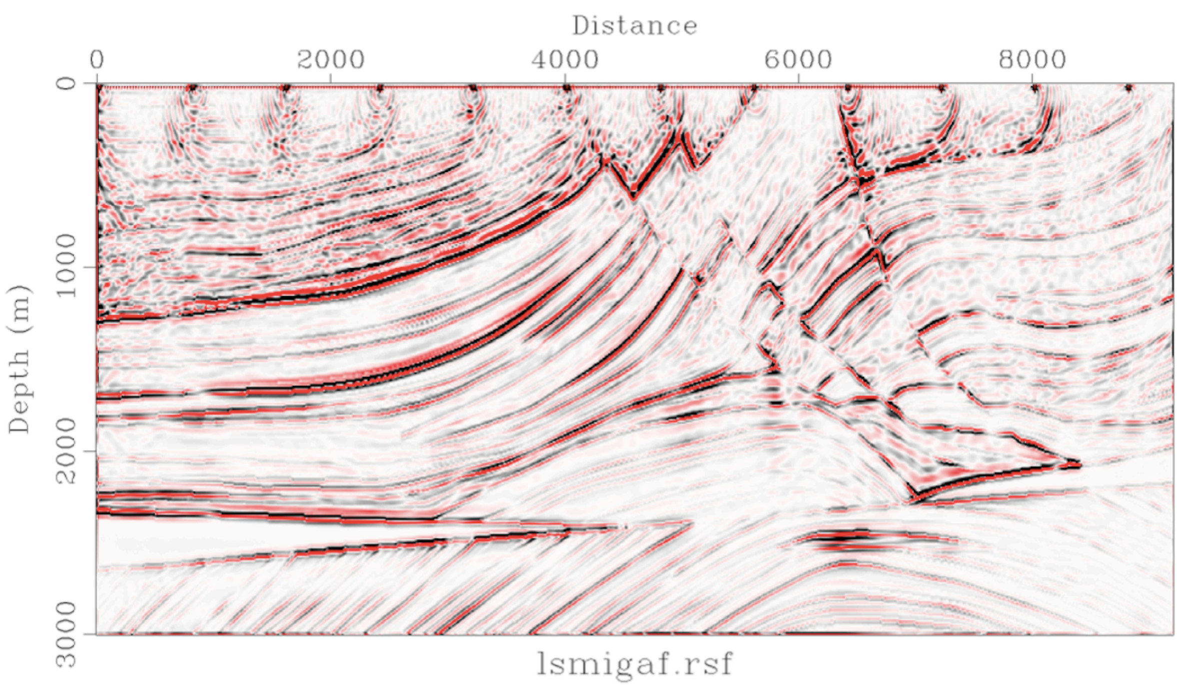


Figure 2.12: The LSRTM result of Marmousi model

Chapter 3

Reverse time migration in the frequency domain

3.1 Helmholtz equation and Green's function

In the frequency domain, the Helmholtz equation is used to solve the forward modeling problem. For the 2D acoustic case, the wave equation with constant density is:

$$\left(\frac{\omega^2}{v^2(\mathbf{x})} + \nabla^2\right) u(\mathbf{x}, \omega; \mathbf{x_s}) = -f(\mathbf{x}, \omega; \mathbf{x_s}), \tag{3.1}$$

where ω is the angular frequency, ∇^2 is the Laplacian operator, $v(\mathbf{x})$ is the velocity, $u(\mathbf{x}, \omega; \mathbf{x_s})$ is the wavefield and $f(\mathbf{x}, \omega; \mathbf{x_s})$ represents the source. When the source is the Dirac delta function $\delta(\mathbf{x}, \omega; \mathbf{x_s})$, equation (3.1) becomes:

$$(\frac{\omega^2}{v^2(\mathbf{x})} + \nabla^2) G(\mathbf{x}, \omega; \mathbf{x_s}) = -\delta(\mathbf{x}, \omega; \mathbf{x_s}), \tag{3.2}$$

where $G(\mathbf{x}, \omega; \mathbf{x_s})$ is the Green's function. The Green's function is the response of a medium due to a point source with specified initial conditions or boundaries conditions. The Green's function can be obtained by:

$$G(\mathbf{x}, \omega; \mathbf{x_s}) = -(\frac{\omega^2}{v^2(\mathbf{x})} + \nabla^2)^{-1}\delta(\mathbf{x}, \omega; \mathbf{x_s}). \tag{3.3}$$

This is usually the first step of computing the wavefield $u(\mathbf{x}, \omega; \mathbf{x_s})$. The term $(\frac{\omega^2}{v^2(\mathbf{x})} + \nabla^2)$ contains the information of velocity and density as well as the difference operator and is called the impedance matrix. Because of the large size of the impedance matrix, the inverse of this matrix requires large computer resources. The Green's function can be calculated by the inverse of the impedance matrix times the point source matrix. The wavefield can be obtained as

$$u(\mathbf{x}, \omega; \mathbf{x_s}) = G(\mathbf{x}, \omega; \mathbf{x_s})f(\mathbf{x}, \omega; \mathbf{x_s}). \tag{3.4}$$

In my tests, I use the Ricker wavelet as the source term $f(\mathbf{x}, \omega; \mathbf{x_s})$.

3.2 The solution to the Helmholtz equation

To solve equation (3.1), I use frequency domain finite-difference modeling to discretize the Helmholtz equation with a 5-point 2D square mesh with constant grid spacing on both x and z direction. i and j represent the horizontal and vertical direction respectively. By discretizing equation (3.1), we have

$$\frac{\omega^2}{v^2}u_{i,j} + \left(\frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{\Delta x^2} + \frac{u_{i,j-1} - 2u_{i,j} + u_{i,j+1}}{\Delta z^2}\right) = -f_{i,j} \tag{3.5}$$

where Δx and Δz are the grid spacing respectively. FIG. 3.1 shows the element locations for the 5-point 2D finite-difference stencil.

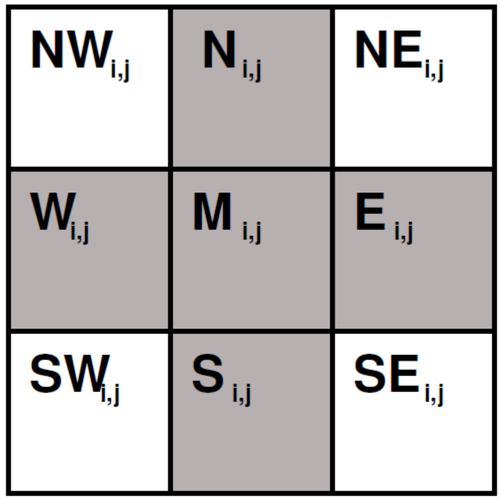


Figure 3.1: Symbolic abbreviations for element locations on a 5-point 2D FD stencil (Ajo-Franklin, 2005)

The coefficients of the 5-point FD method are (Ajo-Franklin, 2005):

$$\begin{aligned} M_{i,j} &= \frac{\omega^2}{v_{i,j}^2} - 2\left(\frac{1}{\Delta x^2} + \frac{1}{\Delta z^2}\right) \\ E_{i,j} &= \frac{1}{\Delta x^2} \\ W_{i,j} &= \frac{1}{\Delta x^2} \\ N_{i,j} &= \frac{1}{\Delta z^2} \\ S_{i,j} &= \frac{1}{\Delta z^2} \end{aligned} \tag{3.6}$$

To avoid boundary reflections, I use the absorbing boundary condition by Engquist and Majda (1977)

$$\frac{\partial u}{\partial n} - i\frac{\omega}{v}u = 0 \tag{3.7}$$

where n is the direction normal to the boundary. Figure 3.2 shows the boundaries for a 5×5 model. The red dots are the grid of artificial boundaries, and blue dots are the grids inside the boundaries, whose coefficients are illustrated in equation (3.6).

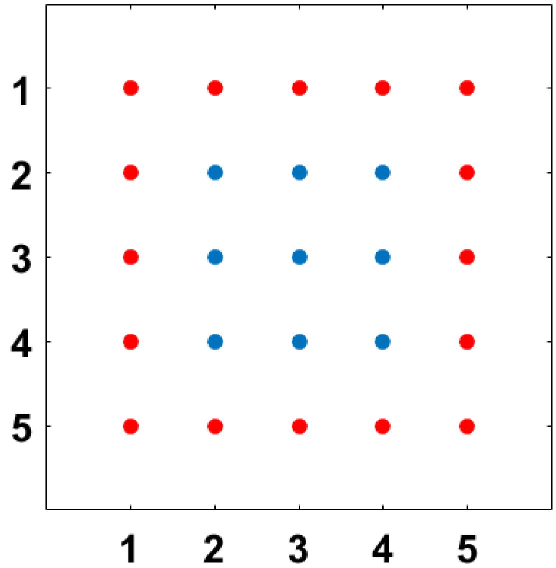


Figure 3.2: The artificial boundaries for a 5×5 discrete model

To solve the first order derivative in equation (3.7), I use the form

$$\frac{\partial u}{\partial x} = \frac{u_{i+1,j} - u_{i,j}}{\Delta x} \tag{3.8}$$

for the horizontal direction and treat the vertical direction similarly. According to equation (3.7) and (3.8), the discrete form of the absorbing boundary condition for top ($j = 1$) and left boundaries ($i = 1$) are

$$\frac{u_{i,2} - u_{i,1}}{\Delta z} + i\frac{\omega}{v_{i,1}}u_{i,1} = 0 \tag{3.9}$$

and

$$\frac{u_{2,j} - u_{1,j}}{\Delta x} + i\frac{\omega}{v_{1,j}}u_{1,j} = 0. \tag{3.10}$$

Therefore, the coefficients at the left boundary (without corner points) are

$$
\begin{aligned}
M_{i,j} &= -i\frac{\omega}{v_{i,j}} - \frac{1}{\Delta x} \\
E_{i,j} &= \frac{1}{\Delta x} \\
W_{i,j} &= 0 \\
N_{i,j} &= 0 \\
S_{i,j} &= 0
\end{aligned}
\tag{3.11}
$$

and the top boundary (without corner points) coefficients are

$$
\begin{aligned}
M_{i,j} &= -i\frac{\omega}{v_{i,j}} - \frac{1}{\Delta z} \\
E_{i,j} &= 0 \\
W_{i,j} &= 0 \\
N_{i,j} &= 0 \\
S_{i,j} &= \frac{1}{\Delta z}
\end{aligned}
\tag{3.12}
$$

For the corner points, I sum the two boundaries together to get the coefficients as followed (take the top left corner for example):

$$
\begin{aligned}
M_{i,j} &= -2i\frac{\omega}{v_{i,j}} - \frac{1}{\Delta x} - \frac{1}{\Delta z} \\
E_{i,j} &= \frac{1}{\Delta x} \\
W_{i,j} &= 0 \\
N_{i,j} &= 0 \\
S_{i,j} &= \frac{1}{\Delta z}
\end{aligned}
\tag{3.13}
$$

Figure 3.3 shows the impedance matrix for a 5×5 discrete model and the imaginary entries are just located in the main diagonal of the impedance matrix. These imaginary entries are introduced by the absorbing boundary conditions. If the model dimensions are

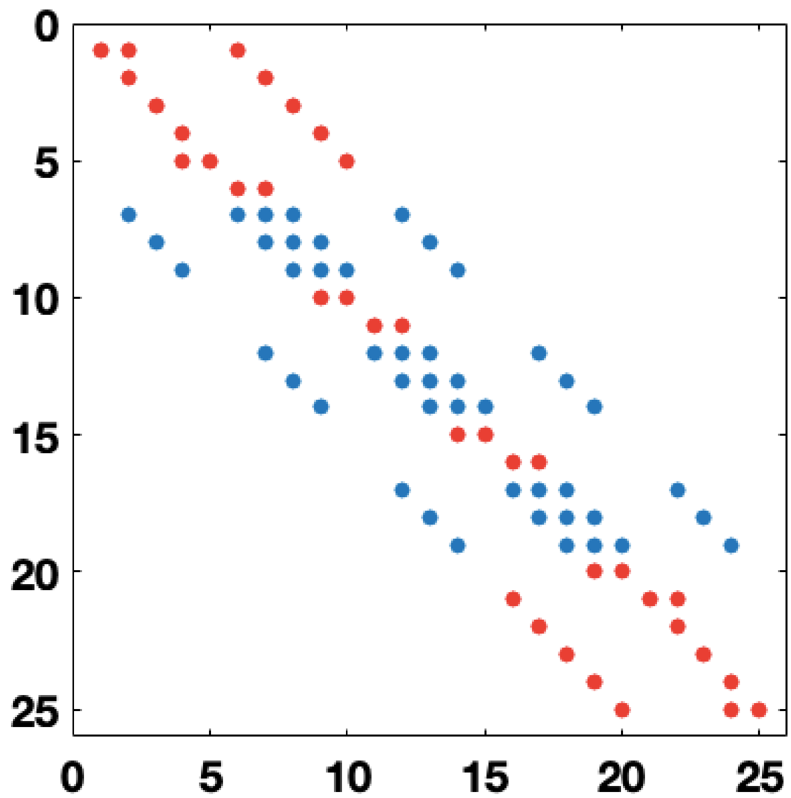


Figure 3.3: The impedance matrix for a 5×5 discrete model. The red dots are the points at the boundaries and blue points represents the inner points in the model.

$n_x \times n_z$, then the dimension of the impedance matrix is $(n_x \times n_z) \times (n_x \times n_z)$ because each row in the impedance matrix will multiply the wavefield to get the source, which is the wave equation. The rows of the red dots are the points at the boundaries, and the blue dots represent the points inside the boundaries. If we use the matrix form to illustrate equation (3.1),

$$\mathbf{A}\mathbf{u} = \mathbf{f} \tag{3.14}$$

the forward modeling can be described as equation (3.14), where $\mathbf{A}$ is the impedance matrix with dimension $(n_x \times n_z) \times (n_x \times n_z)$. $\mathbf{A}$ is a square matrix and has five diagonals because I use the 5-point finite-difference. The complex elements are only located in the main diagonal of the matrix, and they are introduced by the absorbing boundary conditions. $\mathbf{u}$ is the wavefield matrix with dimension $(n_x \times n_z) \times n_s$. n_s is the number of shots of the survey. To complete the data modelling and be able to subtract from the observations, we also need an operator to extract the data from the wavefield $\mathbf{u}$ that I do not show in the equations. On the right hand side, $\mathbf{f}$ is the source matrix with the same dimension as the wavefield. By setting all sources in one source matrix, it is possible to model all sources simultaneously. This makes possible to calculate many sources (shots) simultaneously in the frequency domain. This is an advantage of frequency domain multisource compared to time domain. However, still the modelling has to be done for each frequency independently. These matrices are all for single frequency, and when the forward modeling is completed, the wavefield is a 3D matrix with dimension $(n_x \times n_z) \times n_s \times n_f$, where n_f is the number of frequencies.

Figure 3.4 is the amplitude spectrum of a Ricker wavelet in the frequency domain with a dominant frequency of 15Hz. This wavelet is used as the source in Figure 3.5, which is an example of wavefields in the frequency domain. The model has 300×300 grids with constant velocity and density. The spacing is 8m for both lateral and vertical directions. A Ricker wavelet is injected in the center of the model. The two figures are the real parts of the complex wavefields.

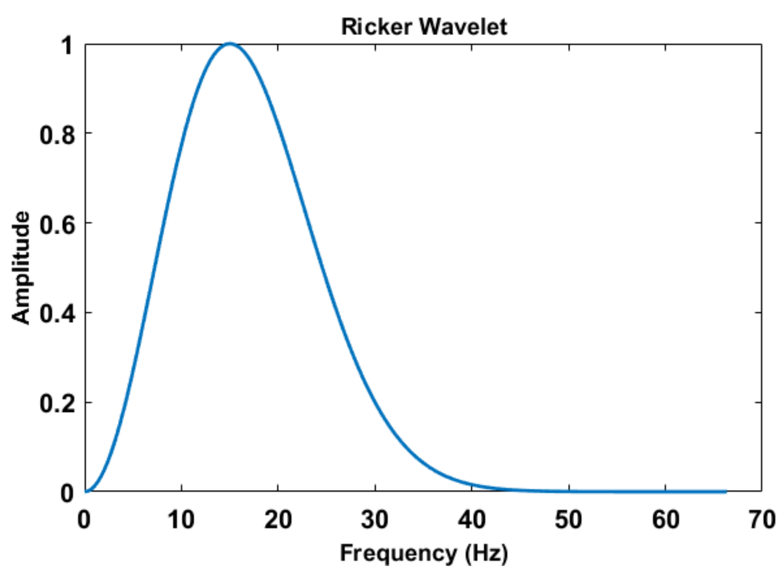
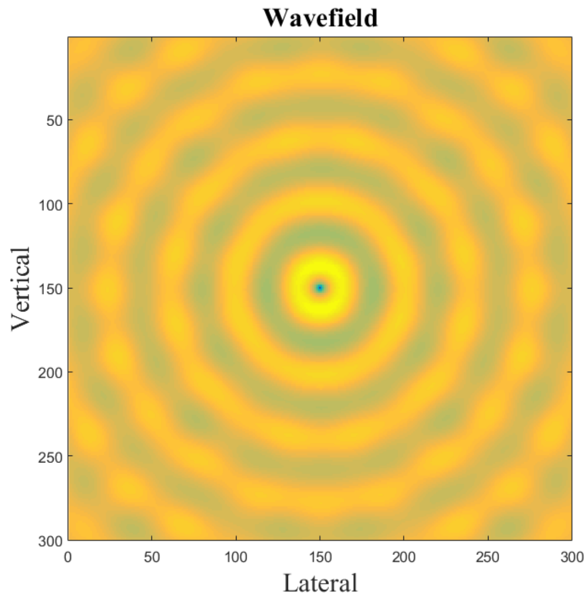
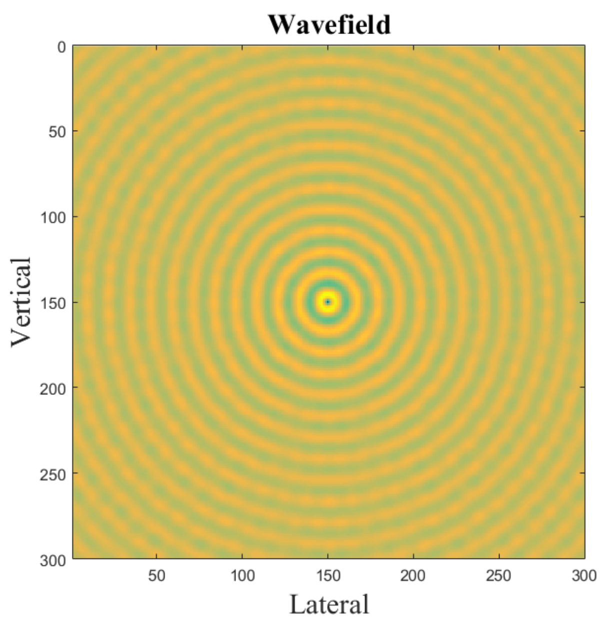


Figure 3.4: The amplitude spectrum of the Ricker wavelet in the frequency domain. $f_{dom} =$ 15Hz



(a) The wavefield at $f = 10$Hz



(b) The wavefield at $f = 30$Hz

Figure 3.5: An example of wavefields in the frequency domain.

3.3 RTM in the frequency domain

In the time domain, the process of RTM involves cross-correlating two wavefields, one from the source side and the other one from the receiver side. The wavefield from the receiver side is time-reversed. The imaging condition (Claerbout, 1971) that I use in this chapter's examples is

$$\begin{aligned} m_{mig}(\mathbf{x}) &= -\mathbf{Re}\left(\sum_{\omega}\sum_{\mathbf{x_s}} \omega^2 D(\mathbf{x},\mathbf{x_s},\omega) U^*(\mathbf{x},\mathbf{x_s},\omega)\right) \\ &= -\mathbf{Re}\left(\sum_{\omega}\sum_{\mathbf{x_s}} \omega^2 f(\mathbf{x},\mathbf{x_s},\omega) G(\mathbf{x},\mathbf{x_s},\omega) G(\mathbf{x},\mathbf{x_r},\omega) \Delta d^*(\mathbf{x},\mathbf{x_s},\omega)\right). \end{aligned} \tag{3.15}$$

In this imaging condition, $D(\mathbf{x},\mathbf{x_s},\omega)$ is the down-going wavefield and $U^*(\mathbf{x},\mathbf{x_s},\omega)$ is the up-going wavefield, where $*$ means complex conjugate. The down-going wavefield can be obtained by the multiplication of source function $f(\mathbf{x},\mathbf{x_s},\omega)$ and source side Green's function $G(\mathbf{x},\mathbf{x_s},\omega)$ while the up-going wavefield is the result of complex conjugate of the data residual $\Delta d^*(\mathbf{x},\mathbf{x_s},\omega)$ multiplied by the receiver side Green's function $G(\mathbf{x},\mathbf{x_r},\omega)$.

Computational complexity

The computational complexity is the amount of resources required for running the algorithm. This varies between the time and frequency domain methods and also 2D and 3D cases. We assume the model's dimension is n^2 in the 2D case and n^3 in the 3D case. In the frequency domain, because of the lower-upper (LU) decomposition of the impedance matrix, the complexity of forward modeling is $O(n_f n^3)$ in 2D (Mulder and Plessix, 2004) and $O(n_f n^6)$ in 3D (Knibbe et al., 2014). For the time domain cases, the complexity becomes $O(n_s n_t n^2)$ in 2D and $O(n_s n_t n^3)$ in 3D, respectively. Table 3.1 summarizes the complexity of each situations. When n_s is large, the frequency domains methods are preferred in 2D surveys.

However, for the 3D surveys, the computation cost is as large as $n_f n^6$ in the frequency domain, which makes it expensive to achieve, while the cost in the time domain is still affordable, $n_s n_t n^3$. This has been always a major obstacle for frequency domain use in 3D migrations and FWI.

Table 3.1: Computational complexity of the time and frequency methods in different dimensions

Method	2D	3D
Frequency domain	$O(n_f n^3)$	$O(n_f n^6)$
Time domain	$O(n_s n_t n^2)$	$O(n_s n_t n^3)$

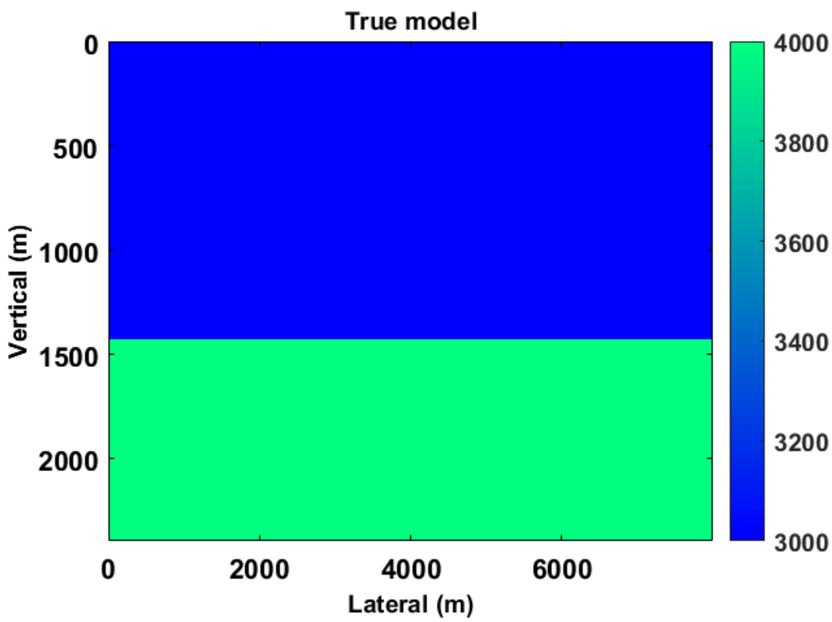
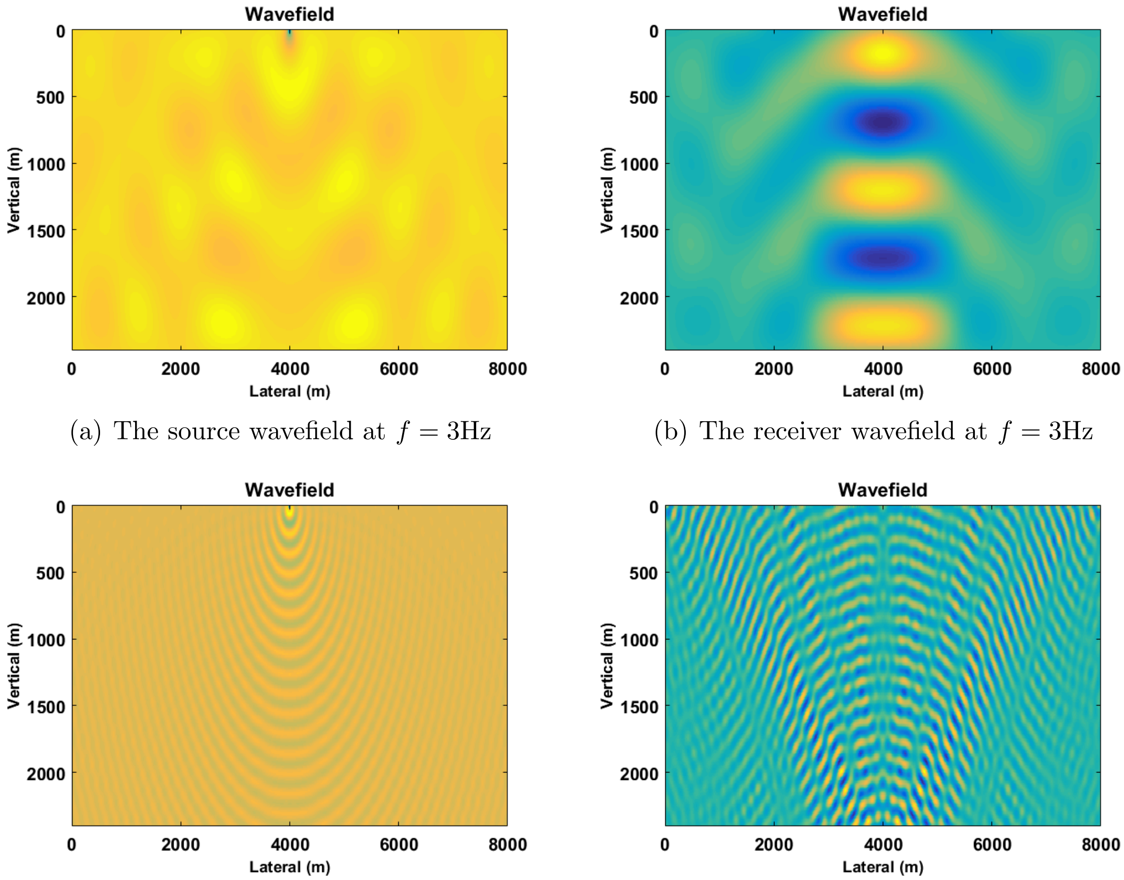


Figure 3.6: The 2-layer model used in the test

Figure 3.6 is an example of the model used in migration. The dimensions of the model are 1000×300 and grid spacing is 8m. The velocity of the upper layer is 3000m/s and the lower layer is 4000m/s in the true model. Figure 3.7 is the example of source and receiver wavefields at different frequencies for the layered model with one source and 334 receivers at the surface of the model. By cross-correlating (a) with (b) and (c) with (d), the RTM images are obtained for $f = 3$Hz and $f = 20$Hz.



(a) The source wavefield at $f = 3$Hz

(b) The receiver wavefield at $f = 3$Hz

(c) The source wavefield at $f = 20$Hz

(d) The receiver wavefield at $f = 20$Hz

Figure 3.7: The example of source and receiver wavefields at different frequencies for a horizontal layered model.

Figure 3.8 is the RTM result of the horizontal reflector model with 10 shots and a Ricker wavelet with $f_{dom} = 15$Hz as the source. Because of the band-limited frequency source, there is ringing Gibb's artifact in the image, especially under the horizontal reflector.

Figure 3.9 shows a section of the Marmousi model. Figure 3.9 (b) is the background model which is generated by applying a Gaussian smoother to the true model. The RTM result is shown in Figure 3.10 (a). Compared to the time domain result, the reflectors are more continuous but less sensitive to the small impedance contrast.

Figure 3.11 is the RTM result of the Marmousi model with different frequency intervals. The quality of the images is improved by reducing the frequency intervals, and the best

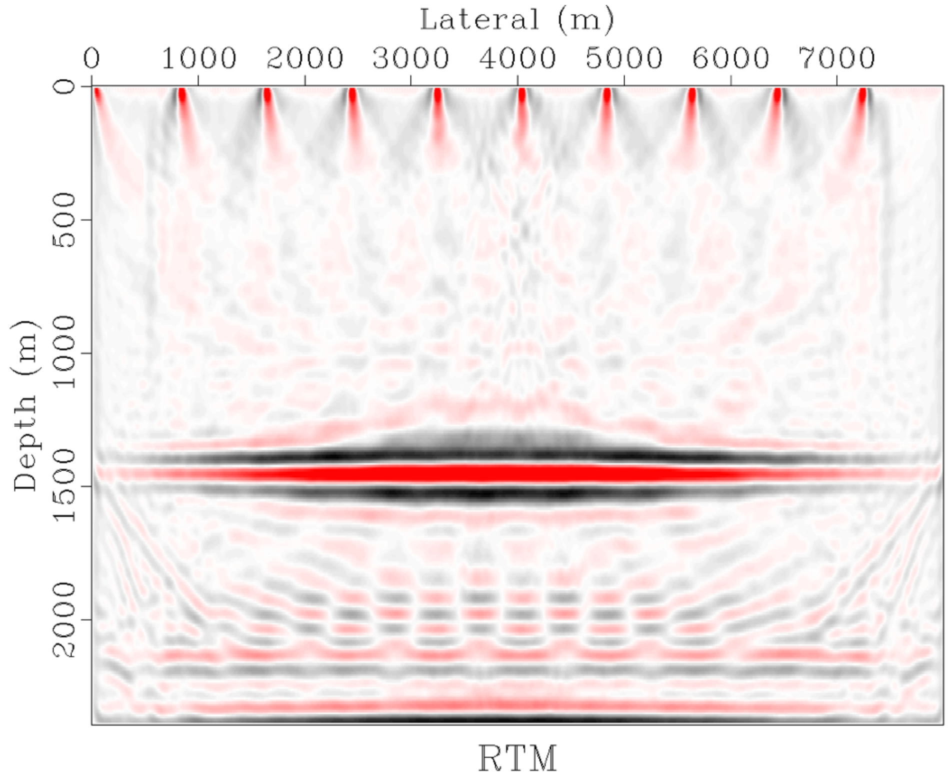


Figure 3.8: The RTM result of the horizontal reflector model.

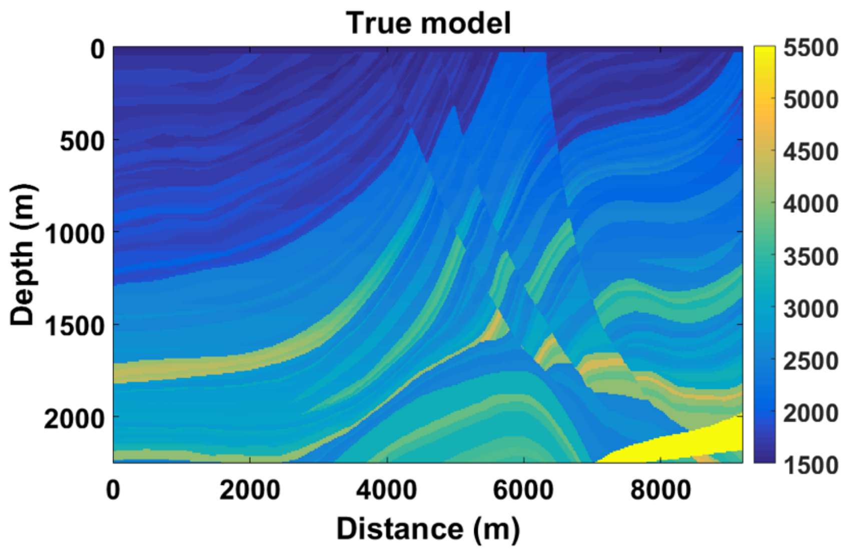
result is $\Delta f = 0.1$Hz. When reducing the frequency interval, we are applying a zero-padding to the signal in the time domain. To understand this, we start from the equation of Nyquist frequency:

$$f_N = \frac{1}{2\Delta t}, \tag{3.16}$$

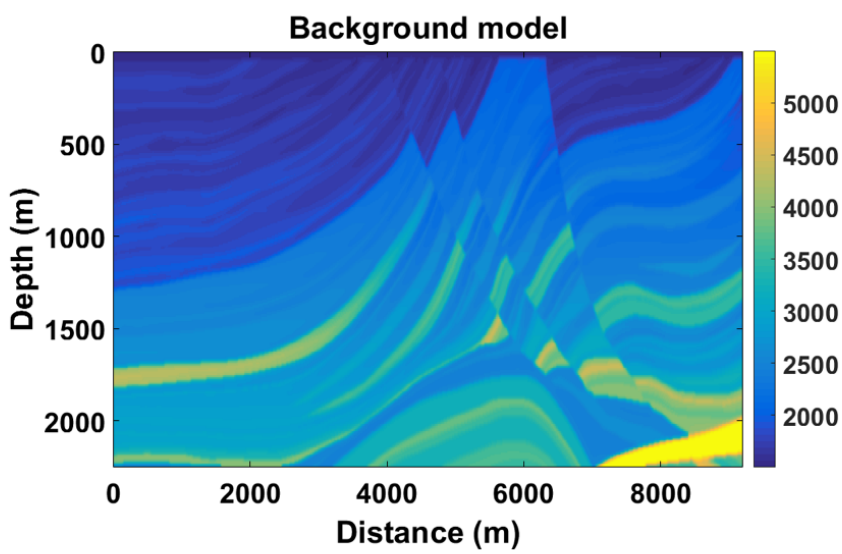
where f_N is Nyquist frequency of the signal and Δt is the sampling interval in the time domain. The frequency interval after discrete Fourier transform is:

$$\Delta f = \frac{f_N}{N/2} = \frac{1}{N\Delta t}, \tag{3.17}$$

where N is the number of sampling points in the time domain. When reducing Δf, N increases, which is equivalent to applying zero-padding in the time domain. Since the cross-correlation process is applied in the frequency domain (circular cross-correlation), extra padding is required compared to the time domain. This translates in another disadvantage of the frequency domain RTM, since the computational cost increases significantly when the finer frequency slices is set. The computation time is linear to the number of frequency slices, e.g. $\Delta f = 0.1$Hz is 10 times longer than $\Delta f = 1$Hz.

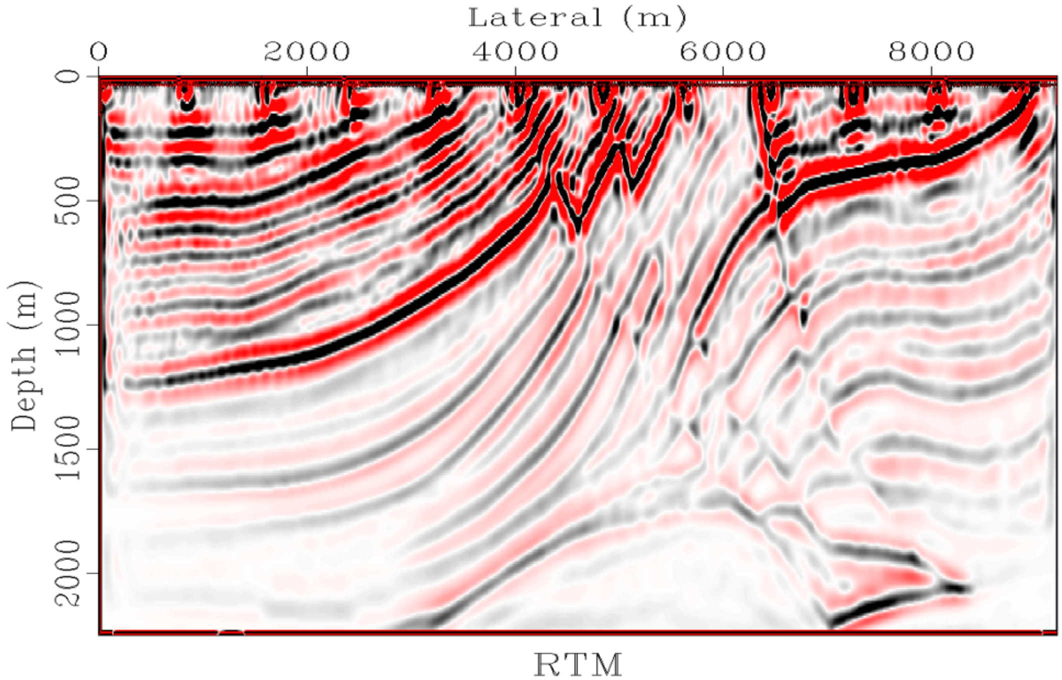


(a) The true Marmousi model

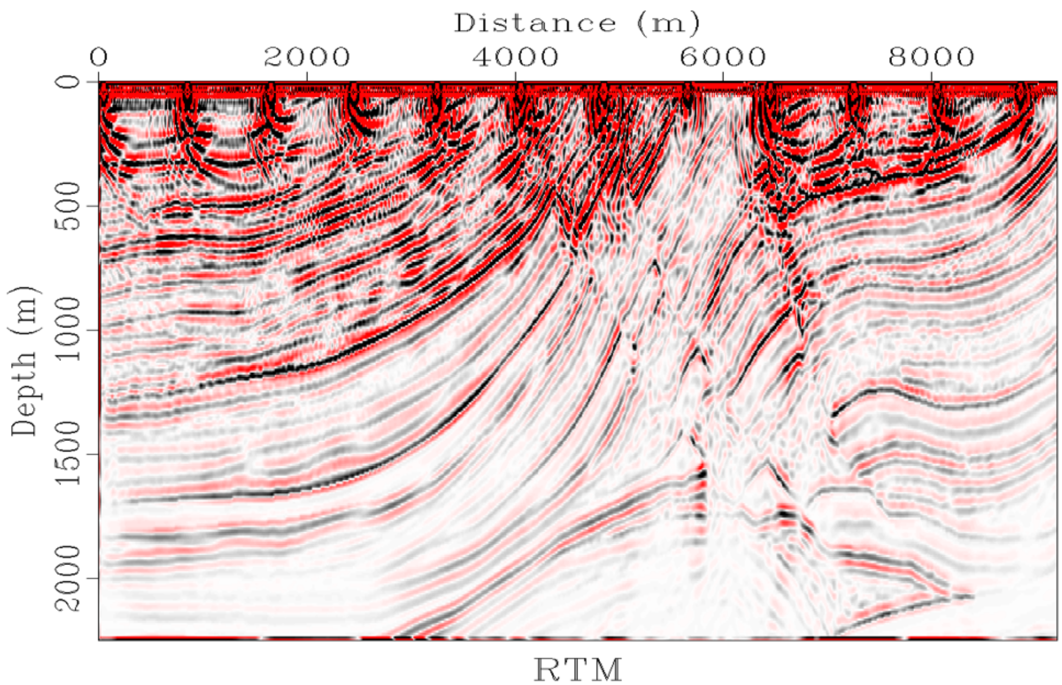


(b) The background Marmousi model

Figure 3.9: The true and background model of Marmousi.

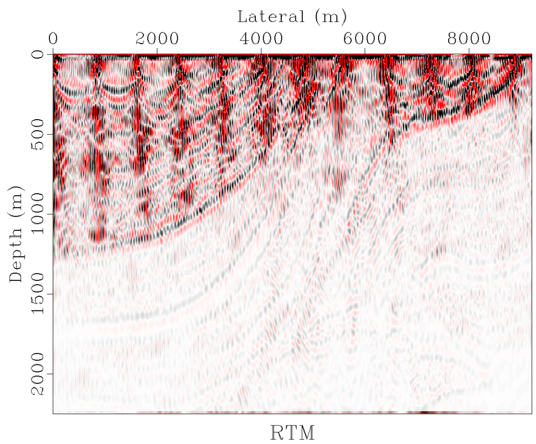


(a) RTM result in frequency domain

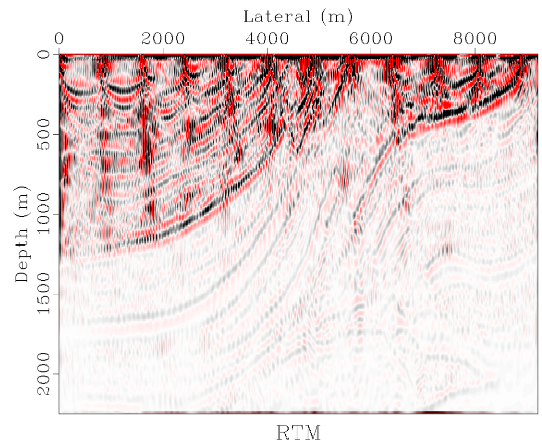


(b) RTM result in time domain

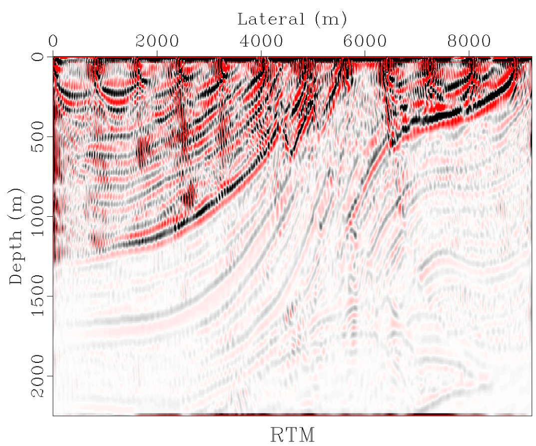
Figure 3.10: The RTM result of the Marmousi model.



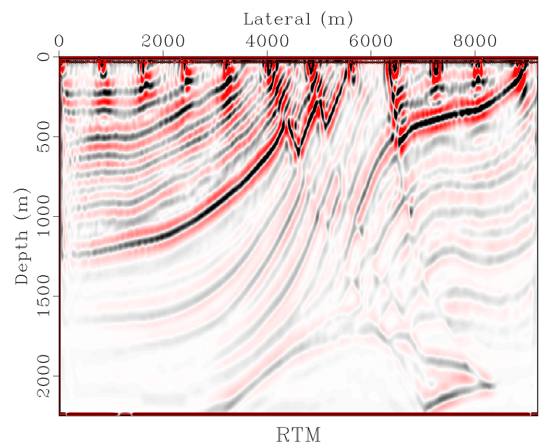
(a) RTM result of Marmousi with $\Delta f = 1$Hz



(b) RTM result of Marmousi with $\Delta f = 0.5$Hz



(c) RTM result of Marmousi with $\Delta f = 0.33$Hz



(d) RTM result of Marmousi with $\Delta f = 0.1$Hz

Figure 3.11: The RTM result of the Marmousi model with different frequency intervals.

3.4 Conclusions

In this chapter, I illustrate the process of RTM in the frequency domain. The finite-difference method is used to solve for the Helmholtz equation in the frequency domain, multiple sources can be simultaneously migrated in the matrix form. The wavefield for each shot can be obtained by the inverse of the impedance matrix times the source matrix. This makes the modeling in the frequency domain more efficient than that in the time domain when there are many sources. RTM in the frequency domain is very similar to the time domain. The cross-correlating imaging condition changes to matrix multiplication in the frequency domain.

However, the frequency method suffers from a considerable computation cost and requires large memory in 3D surveys because of the need to invert a large size of the impedance matrix. Also, using finer frequency intervals is required to achieve equivalent image quality compared to the time domain, but this increases the computational cost significantly.

Chapter 4

Frequency domain full waveform inversion and least-squares reverse time migration

4.1 Introduction

In LSRTM, linear operators are used in geophysical modeling calculations that predict data from models. A common task is to find the inverse of these calculations, i.e., given the observed data, find a model that can predict these data. For linear operators, it is possible to use iterative algorithms which require forward and adjoint operators to solve the inverse problem (Claerbout, 1992). Similar to LSRTM, FWI also involves minimizing the misfit function between observed data and synthetic data (Virieux and Operto, 2009). Usually, FWI problems can be solved by a two-loop algorithm: the inner loop is to iteratively solve for the model perturbation and the outer loop is to update the current model and compute the synthetic data to get the new residual. The inner loop can be treated as the LSRTM problem (Chen and Sacchi, 2018). In fact, the gradient of the objective function can be proved to be equivalent to the image condition of RTM. Therefore, the optimization of the gradient is a way to implement LSRTM. By the Gauss-Newton approximation, using truncated Newtons method is a good way to solve this problem. The Hessian-vector product is calculated in each iteration in a matrix-free form and the gradient of the objective function is solved by a conjugate gradient algorithm (Pan et al., 2017). The examples shown in this chapter were simulated with modified Matlab software originally developed by Wenyong Pan (Pan, 2017).

4.2 The formulation of FWI

I will start from the objective function of frequency domain FWI:

$$J(m) = \frac{1}{2} \sum_{n_\omega} \sum_{n_s} \| d_{obs}(x_s, \omega) - d_{syn}(m, x_s, \omega) \|^2 = \frac{1}{2} \sum_{n_\omega} \sum_{n_s} \| \delta d(m, x_s, \omega) \|^2, \tag{4.1}$$

where d_{obs} is the observed data, $d_{syn} = Ru$ is the synthetic data in which R is the extraction operator and u is the wavefield. The dimensions of the data are $nr \times ns \times nf$ where nr and ns are the number of receivers and sources respectively, and nf is the number of frequency slices.

4.2.1 The gradient

The squared data residual in equation (4.1) can be expanded as

$$\begin{aligned} \|\delta d\|^2 &= (d_{obs} - d_{syn})^T (d_{obs} - d_{syn})^* \\ &= d_{obs}^T d_{obs}^* - d_{obs}^T d_{syn}^* - d_{syn}^T d_{obs}^* + d_{syn}^T d_{syn}^*, \end{aligned} \tag{4.2}$$

where $*$ means complex conjugate and T means transpose. Inserting equation (4.2) into equation (4.1) and taking derivative of equation (4.1) with respect to m, we have

$$
\begin{aligned}
g &= \frac{\partial J(m)}{\partial m} \\
&= -\frac{1}{2}(-d_{obs}^T \frac{\partial d_{syn}^*}{\partial m} - d_{obs}^* \frac{\partial d_{syn}^T}{\partial m} + d_{syn}^* \frac{\partial d_{syn}^T}{\partial m} + d_{syn}^T \frac{\partial d_{syn}^*}{\partial m}) \\
&= -\frac{1}{2}\left((d_{syn} - d_{obs})^T \frac{\partial d_{syn}^*}{\partial m} + (d_{syn} - d_{obs})^* \frac{\partial d_{syn}^T}{\partial m}\right),
\end{aligned}
\tag{4.3}
$$

Because

$$
Z + Z^* = \mathbf{Re}Z + i\mathbf{Im}Z + \mathbf{Re}Z - i\mathbf{Im}Z = 2\mathbf{Re}Z, \tag{4.4}
$$

we can rewrite the gradient as:

$$
\begin{aligned}
g &= -\sum_{n_\omega} \mathbf{Re}\left(\frac{\partial d_{syn}^T}{\partial m}\delta d^*\right) \\
&= -\sum_{n_\omega} \mathbf{Re}\left(\frac{\partial d_{syn}}{\partial m}\delta d^\dagger\right)
\end{aligned}
\tag{4.5}
$$

The term $\frac{\partial d_{syn}^T}{\partial m}$ represents the Fréchet derivative and $\delta d^\dagger$ is the conjugate transpose of the data residual. To derive the complete form of the Fréchet derivative, we write the acoustic wave equation in the matrix form:

$$
A(m, \omega)u(m, x_s, \omega) = f(x_s, \omega) \tag{4.6}
$$

$$
A(m, \omega) = \left(\omega^2 m(x) + \nabla^2\right) \tag{4.7}
$$

where $A(m, \omega)$ is the impedance matrix.

Taking the derivative on each side of the equation (4.6) with respect to the model parameter m and putting one term in the other side of the equation, we have

$$
A(m, \omega)\frac{\partial u(m, x_s, \omega)}{\partial m} = -\frac{\partial A(m, \omega)}{\partial m}u(m, x_s, \omega). \tag{4.8}
$$

This equation shows that the Fréchet derivative can be obtained by solving equation (4.8) with a virtual source term

$$f_v = -\frac{\partial A(m,\omega)}{\partial m} u(m, x_s, \omega) \tag{4.9}$$

Therefore, the Fréchet derivative can be written as

$$\begin{aligned}\frac{\partial u(m, x_s, \omega)}{\partial m} &= -A^{-1}(m,\omega)\frac{\partial A(m,\omega)}{\partial m} u(m, x_s, \omega) \\ &= -G(x,\omega)\omega^2 G(x,\omega) f(x_s,\omega),\end{aligned} \tag{4.10}$$

where $f(x_s, \omega)$ is the source term in equation (4.6), and $\frac{\partial A(m,\omega)}{\partial m} = \omega^2$. Because $d_{syn} = Ru$ and R is a real-valued operator which is independent of the model m, the gradient of the FWI objective function is

$$\begin{aligned} g &= -\mathbf{Re}\left(\sum_{n_\omega}\sum_{n_s} \frac{\partial d_{syn}^T}{\partial m}\delta d^*\right) \\ &= -\mathbf{Re}\left(\sum_{n_\omega}\sum_{n_s} R^T \frac{\partial u^T(m, x_s, \omega)}{\partial m}\delta d^*\right) \\ &= \mathbf{Re}\left(\sum_{n_\omega}\sum_{n_s} \omega^2 R^T G^T(x,\omega) G^T(x,\omega) f^T(x_s,\omega)\delta d^*\right). \end{aligned} \tag{4.11}$$

If we take the derivative of the transpose of the wavefield $u(m, x_s, \omega)$ with respect to m and substitute the result into equation (4.5), the matrix form of equation (4.11) can be formed as:

$$\begin{aligned} g &= -\mathbf{Re}\left(\sum_{n_\omega}\sum_{n_s} R^T \left(-A^{-1}(m,\omega)\frac{\partial A(m,\omega)}{\partial m} u(m, x_s, \omega)\right)^T \delta d^*\right) \\ &= \mathbf{Re}\left(\sum_{n_\omega}\sum_{n_s} \left(u(m, x_s, \omega)^T \left(\frac{\partial A(m,\omega)}{\partial m}\right)^T (A^{-1}(m,\omega))^T R^T \delta d^*\right)\right). \end{aligned} \tag{4.12}$$

The conjugate of the data residual in the frequency domain is equal to the data time reversed in time domain and the data residual is back projected to the whole space using the operator R^T, before it is propagated back to the subsurface by the term $(A^{-1}(m,\omega))^T$. Since only the real part is considered in the gradient computation, by using the adjoint impedance matrix, we have

$$g = \mathbf{Re}\left(\sum_{n_\omega}\sum_{n_s}\left(u(m,x_s,\omega)^\dagger\left(\frac{\partial A(m,\omega)}{\partial m}\right)^\dagger(A^{-1}(m,\omega))^\dagger R^\dagger \delta d\right)\right), \tag{4.13}$$

where we keep $\left(\frac{\partial A(m,\omega)}{\partial m}\right)^\dagger$ for different parameterization. Therefore, we have the matrix form of gradient computation.

Now with the gradient equation, we are able to use the steepest decent method to update the initial model in a FWI problem. However, the gradient-based method only accounts for the first-order scattered data, which will cause severe faults when the data contains second-order scattered data . In this situation, the second-ordered derivative of the objective function (Hessian) can alleviate this problem (Geng et al., 2017). As we will see in the next section, this requires the Newton-based method to be applied in the model perturbation computation.

4.2.2 The Hessian

The expanded the objective function of FWI in Taylor series is

$$J(m+\delta m) = J(m) + \delta m^T g + \frac{1}{2}\delta m^T H \delta m + ... \tag{4.14}$$

where $H = \frac{\partial J^2(m)}{\partial m^2}$ represents the second derivative of the objective function. When reaching the minimum of the misfit function, and neglecting the higher order terms,

$$J(m+\delta m) - J(m) \approx \delta m^T g + \frac{1}{2}\delta m^T H \delta m = 0. \tag{4.15}$$

Because the Hessian is an extremely large and dense matrix for large scale inverse problem and has a large computation cost, many approximations have been developed over the years.

If the model dimension is $M \times N$, the Hessian will be a $MN \times MN$ symmetric square matrix (MN means $M \times N$). In the following section, I will derive the full Hessian by taking the second derivative of the FWI's objective function.

Recall the objective function of FWI:

$$\begin{aligned} J(m) &= \frac{1}{2} \sum_{n_\omega} \sum_{n_s} \|d_{obs}(x_s, \omega) - d_{syn}(m, x_s, \omega)\|^2 \\ &= \frac{1}{2} \sum_{n_\omega} \sum_{n_s} (d_{obs} - Ru)(d_{obs} - Ru)^\dagger. \end{aligned} \tag{4.16}$$

Take the first order partial derivative with m of the objective function

$$\frac{\partial J(m)}{\partial m} = \frac{1}{2} \sum_{n_\omega} \sum_{n_s} \left(-R \frac{\partial u}{\partial m} (d_{obs} - Ru)^\dagger + (d_{obs} - Ru)(-R^\dagger \frac{\partial u^\dagger}{\partial m}) \right) \tag{4.17}$$

and the second order derivative is

$$\begin{aligned} \frac{\partial J^2(m)}{\partial m^2} = \frac{1}{2} \sum_{n_\omega} \sum_{n_s} (&-R \frac{\partial^2 u}{\partial m^2} (d_{obs} - Ru)^\dagger + \frac{\partial u}{\partial m} R R^\dagger \frac{\partial u^\dagger}{\partial m} \\ &+ \frac{\partial u}{\partial m} R R^\dagger \frac{\partial u^\dagger}{\partial m} - R^\dagger \frac{\partial^2 u^\dagger}{\partial m^2} (d_{obs} - Ru)) \end{aligned} \tag{4.18}$$

Take the real part of equation (4.18), the full Hessian is

$$H(m) = Re \left(\frac{\partial J^2(m)}{\partial m^2} \right) = B(m) + C(m) \tag{4.19}$$

where

$$B(m) = \mathbf{Re} \left(\sum_{n_\omega} \sum_{n_s} \frac{\partial u}{\partial m} R R^\dagger \frac{\partial u^\dagger}{\partial m} \right), \tag{4.20}$$

$$C(m) = \mathbf{Re} \left(\sum_{n_\omega} \sum_{n_s} R^\dagger (Ru - d_{obs}) \frac{\partial^2 u}{\partial m^2} \right). \tag{4.21}$$

$C(m)$ is the multiplication of the data residuals and the second-order derivative of the wavefield, which works in suppressing the second-order scattering effect in the gradient (Métivier et al., 2013). This term is small when the initial model is close to the true model. The matrix $B(m)$ is known as the Gauss-Newton approximation of the Hessian operator when the expression $C(m)$ is neglected.

4.2.3 Optimization methods

There are many ways of solving equation (4.1). All these optimizing methods are iterative, which start with an initial guess of the model m, get the model updates, update the model until the objective function reaches its minimum. The most common methods are the steepest descent method, Newton's methods and the conjugate gradient method.

Steepest descent method

In the steepest descent method, the inverse of the Hessian matrix H^{-1} is treated as an identity matrix so that the searching direction is equal to the negative gradient:

$$\Delta m = -g. \tag{4.22}$$

The model can be updated as:

$$m_{k+1} = m_k + \alpha \Delta m_k \tag{4.23}$$

where α is the step length. We can use the Wolfe conditions or the Goldstein conditions to decide the step length (Nocedal and Wright, 2006). This method is computationally cheaper than the Newton's method but suffers from slow convergence. We may wish to start iterating with steepest descent.

Newton's methods

Different from the steepest gradient method, the searching direction in Newton's methods is:

$$\Delta m = -H^{-1} g. \tag{4.24}$$

For the full-Newton's method, both components of the Hessian in equations (4.20) and (4.21) are included. For the Gauss-Newton's method, only the single scattering term in equation (4.20) is included. Compared with the steepest descent method, Newton's methods use the second derivative (the Hessian) to improve the convergence speed. However, for a large-scale problem, the computation cost is considerable in calculating the inverse of the Hessian matrix. An alternative to the Newton's method, which uses much less memory is the truncated Newton's method which uses Hessian vector products to solve for the search direction. This method reduces the computational cost largely. For nonlinear problems, we may choose to iterate with a hybrid method that uses a combination of steepest descent and Gauss-Newton method.

Conjugate gradient method

The linear conjugate gradient (CG) method was proposed in 1950s (Hestenes and Stiefel, 1952) to solve a linear system $Ax = b$. In the FWI problem, the CG method can be applied to the linear problem $H\Delta m = -g$ to obtain the search direction Δm. This can be the inner-loop of the FWI algorithm. The advantage of the CG method compared to the steepest descent explained earlier is that, with a very minor addition of memory and computation time, it converges faster when the cost functions have narrow valleys. At difference of the steepest descent method that follows directly the residual direction on each step (greedy algorithm), the CG method uses a gradient that is a combination of previous residuals steps.

4.2.4 Workflow and examples

Figure 4.1 shows the workflow for a common FWI problem. The FWI problem starts with an initial guess of the model. By using the finite-difference method, the synthetic data are calculated and then the data residual is obtained. From the data residual, we can compute the gradient of the objective function. The searching direction varies with different methods used in the optimization process. This is usually the internal loop of the algorithm. After the searching direction is computed, we can calculate the step length to decide how much the model goes from this direction and update the model. This is the outer loop of the algorithm.

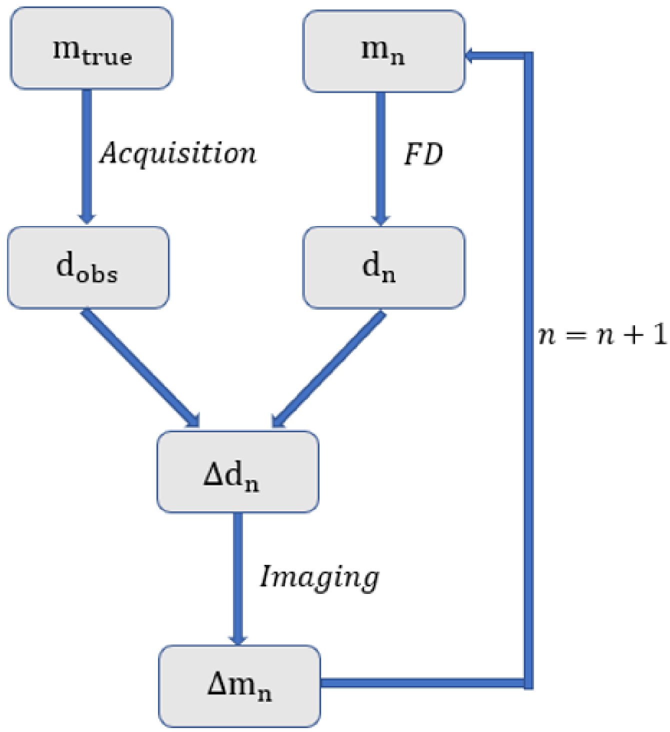


Figure 4.1: The workflow for a common FWI problem.

Figure 4.2 shows the resampled Marmousi. A smoothed model is obtained by a Gaussian smoother, which is shown in Figure 4.3. Figure 4.4 is the result after 10 iterations.

4.3 FWI-based least-squares reverse time migration

In this section, I will illustrate a method of FWI-based least-squares reverse time migration in the 2D acoustic constant density case. I will show the equivalence between the gradient of FWI objective function and the cross-correlation imaging condition. Usually, FWI is a two-loop algorithm. The outer loop is to calculate the step length and uses the searching direction to update the model iteratively. The inner loop is to calculate the searching direction based

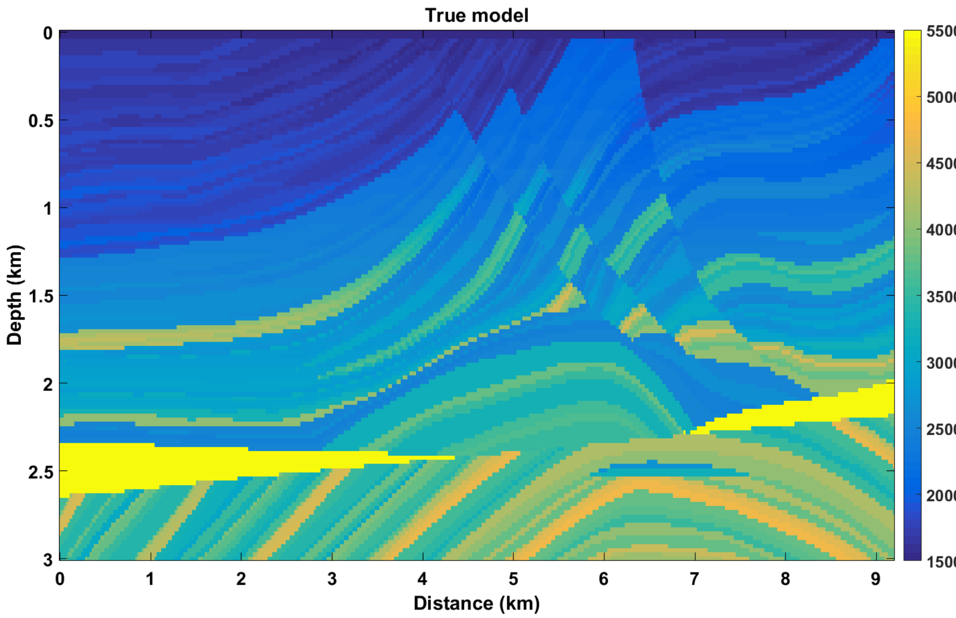


Figure 4.2: The resampled Marmousi model of the FWI problem.

on the gradient. The truncated Newton's method solves for a search direction, using a linear relationship between the gradient g, the Hessian H and the searching direction Δm. We can use the conjugate gradient method to solve for the searching direction iteratively, which is the result of FWI-based LSRTM in the frequency domain.

4.3.1 Theory

The gradient of the FWI objective function is:

$$g = \mathbf{Re}\left(\sum_{n_\omega}\sum_{n_s}\omega^2 G(x,\omega)G(x,\omega)f(x_s,\omega)\delta d^*\right), \tag{4.25}$$

which is equal to the cross-correlation imaging condition (3.15) in the frequency domain. Recall the cross-correlation imaging condition in the time domain:

$$m_{mig}(\mathbf{x}) = \sum_{n_s}\sum_{n_t} d(\mathbf{x},t)u(\mathbf{x},T-t). \tag{4.26}$$

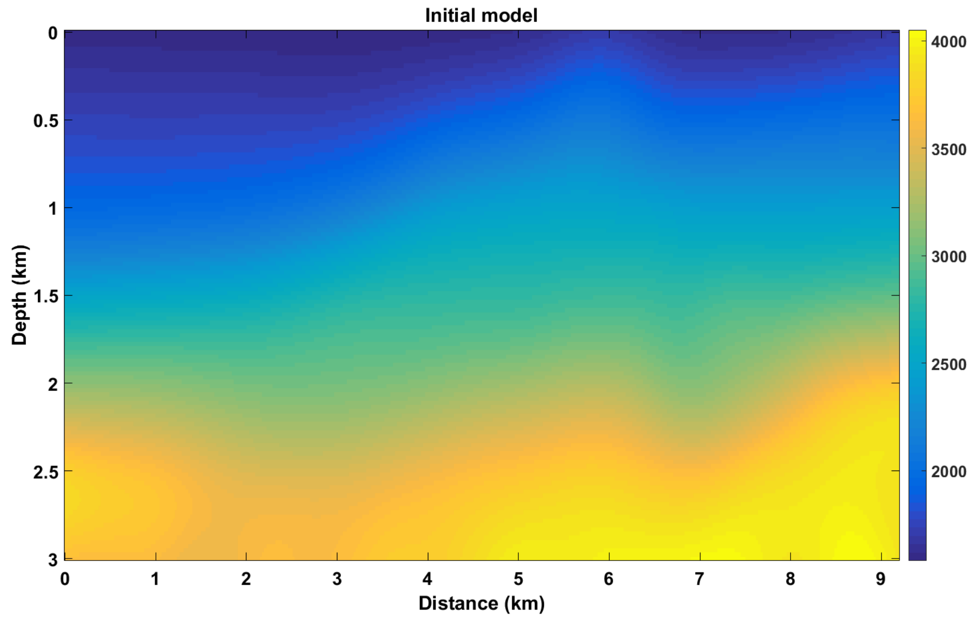


Figure 4.3: The initial Marmousi model of the FWI problem.

The source wavefield is the element-wise multiplication between the source side Green's function and the source function while the receiver side wavefield is the conjugate data residual multiply the Green's function in the receiver side. Except for the frequency weights in the gradient, the two imaging conditions are equivalent.

When the truncated Newton's method is used, a linear equation connects the gradient, Hessian and model perturbation. Starting from this equation, we formulate the linear relationship:

$$H\Delta m = -g. \tag{4.27}$$

Based on this, the objective function of FWI-based LSRTM in the image domain is

$$\Phi(\Delta m) = \frac{1}{2}\sum_{n_\omega}\sum_{n_s}\|g + H\Delta m\|^2. \tag{4.28}$$

Substituting equation (4.10) into (4.20), the Hessian vector product using the Gauss-Newton

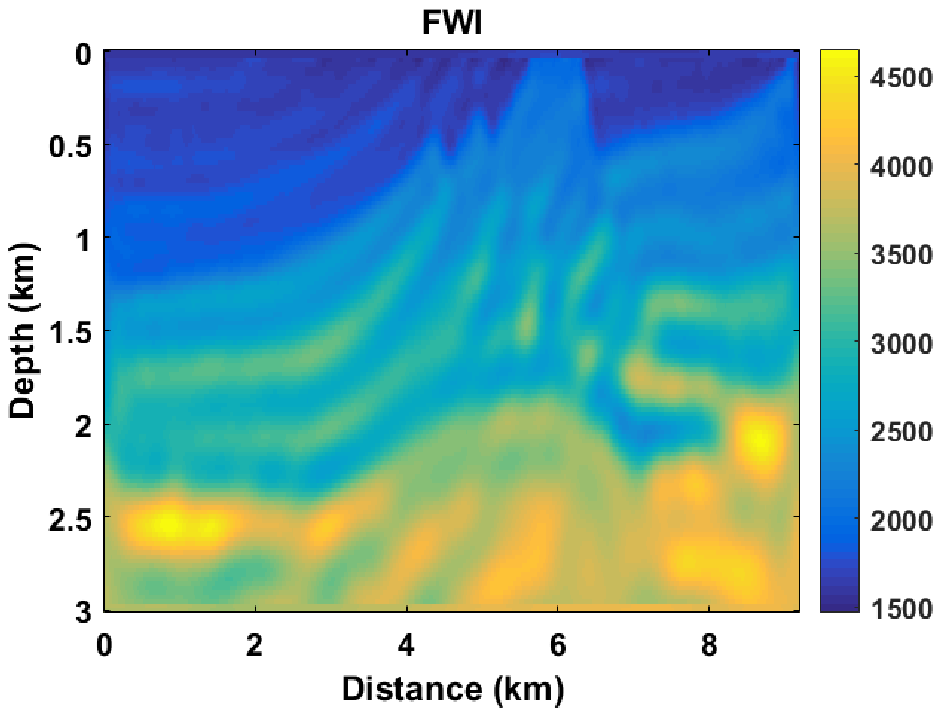


Figure 4.4: The result of FWI after 10 iterations.

approximation is

$$H\Delta m = u\frac{\partial A(m,\omega)}{\partial m}A^{-1}(m,\omega)RR^{\dagger}(A^{-1}(m,\omega))^{\dagger}\left(\frac{\partial A(m,\omega)}{\partial m}\right)^{\dagger}u^{\dagger}\Delta m \tag{4.29}$$

Note that there is a derivative term $\frac{\partial A(m,\omega)}{\partial m}$ in the Hessian and this term changes with different parameterizations. Recall the equation of the impedance matrix $A(m,\omega)$:

$$A(m,\omega) = \left(\omega^2 m(x) + \nabla^2\right), \tag{4.30}$$

where $m(x) = 1/v^2(x)$ when the density is constant in the acoustic medium. If we take the derivative with respect to the variable $m(x)$, then the derivative is ω^2. However, in FWI, if we try to solve the searching direction to update the velocity model $v(x)$, then the chain rule should be applied to the derivative equation. In this case,

$$\frac{\partial A(m,\omega)}{\partial v} = \frac{\partial A(m,\omega)}{\partial m}\frac{\partial m}{\partial v} = -\frac{2}{v^3(x)}\frac{\partial A(m,\omega)}{\partial m}, \tag{4.31}$$

which means that the result would contain the information of the initial model. This explains the significant improvement in quality compared to the previous results in the time domain, as we will see in the numerical examples. However, it also means that this method will contain significant information that was put in through the velocity model. In practice, the smoother the velocity model, the less difference would be with the other methods.

Because of the large computational cost to solve for H^{-1}, we use the conjugate gradient method to get Δm iteratively. A typical conjugate gradient algorithm to solve linear system $Ax = b$ is in Table 4.1:

Table 4.1: The algorithm of the conjugate gradient method

$r_0 = b - Ax_0$
$p_0 = r_0$
if r_0 is sufficiently small, then return x_0 as the result
$k = 0$

repeat
$\alpha_k = \frac{r_k^T r_k}{p_k^T A p_k}$
$x_{k+1} = x_k + \alpha_k p_k$
$r_{k+1} = r_k - \alpha_k A p_k$
if r_{k+1} is sufficiently small, then exit loop
$\beta_k = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$
$p_{k+1} = r_{k+1} + \beta_k p_k$
$k = k + 1$
end repeat

return x_{k+1} as the result

Figure 4.5 is the workflow of the FWI-based LSRTM. Similar to the algorithm of FWI, it starts with the computation of the forward modeling and the data residual. By using the truncated Newton's method, the searching direction Δm is solved by the conjugate gradient method. Different from FWI, in LSRTM the initial model is not updated.

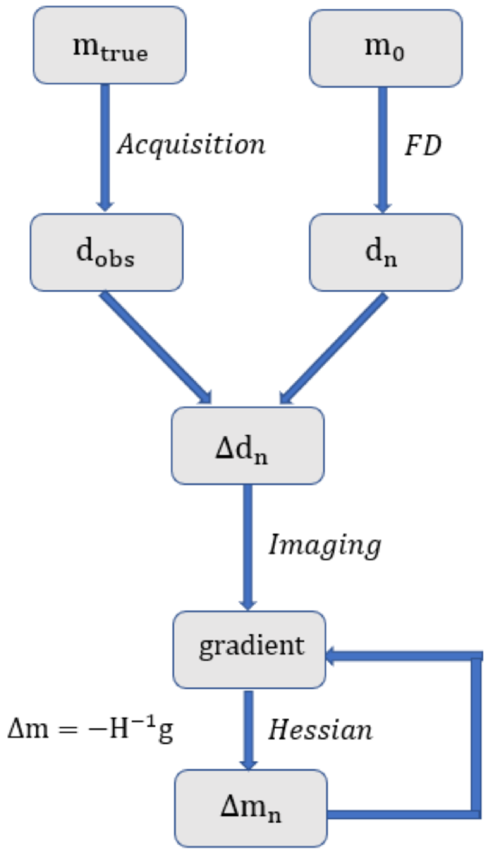


Figure 4.5: The workflow of the FWI-based LSRTM.

4.3.2 Numerical examples

Figure 4.6 and 4.7 show a similar example of the Marmousi model previously shown in Chapter 3. The reflectors are better imaged in the strong impedance contrast parts of the model. After applied least-squares scheme,, the image is more balanced than the pure RTM, but the improvement is limited.

4.4 Conclusions

In this chapter, I derived the gradient and the Hessian of the objective function in a common FWI problem and prove the equivalence between the gradient and the cross-correlation imaging condition in LSRTM. The inner loop of FWI to solve for the searching directions

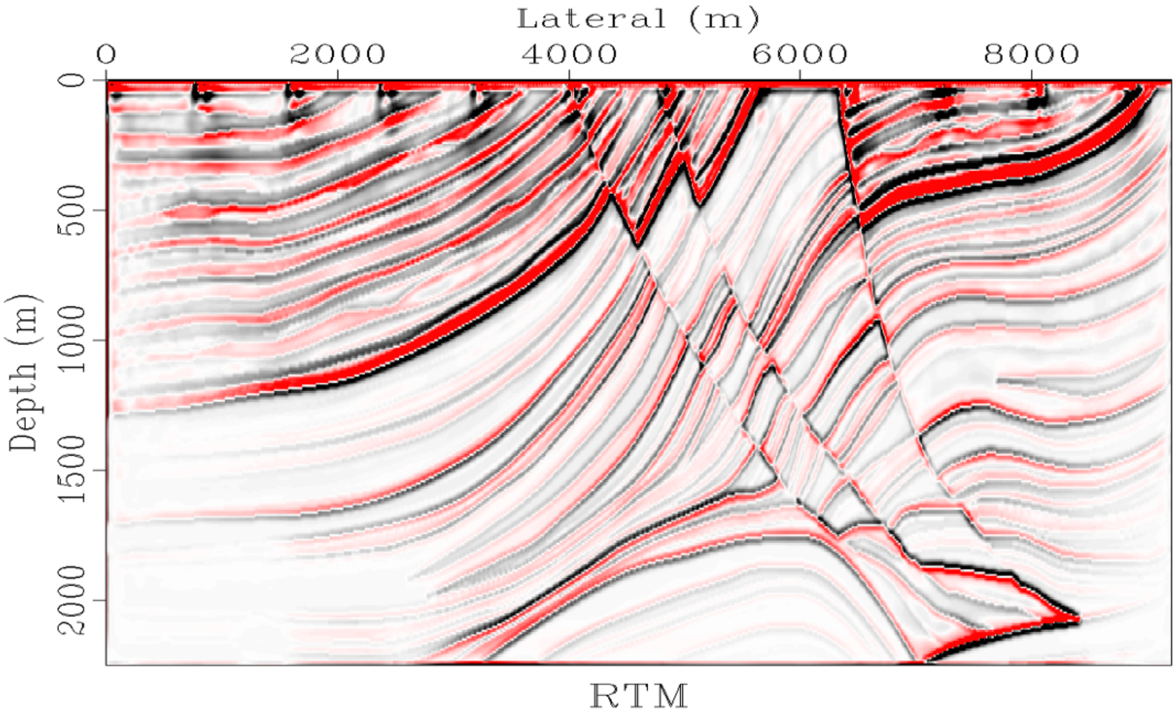


Figure 4.6: RTM for the Marmousi model. Results are better than previous methods, but as explained in the text, the imaging condition contains information about the velocity model.

can be treated as a LSRTM problem. By using the truncated Newton's method, the linear relationship between the gradient, the Hessian and the searching direction can be solved by the conjugate gradient method.

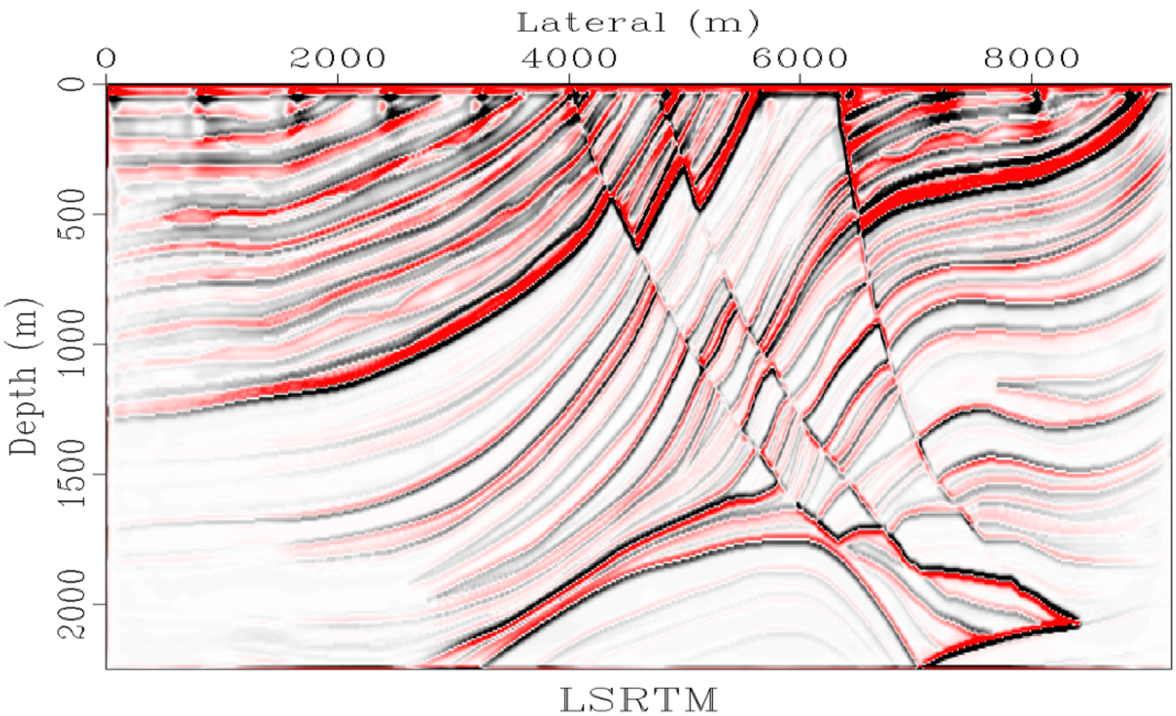


Figure 4.7: FWI-based LSRTM for the Marmousi model. Similar comments as for the previous figure apply here.

Chapter 5

Full waveform inversion in the time domain

In the previous sections, I illustrated the formulations of time domain RTM, frequency domain RTM and frequency domain FWI. To make this thesis complete, I will briefly discuss the time domain FWI in this chapter. For the examples and discussions, I follow Yang (2014) and his implementation in the open source software Madagascar (Fomel et al., 2012).

5.1 Theory

FWI in the time domain involves the following steps (Yang et al., 2015): (1) generate synthetic wavefields via finite difference method and calculate the residual data, (2) reconstruct the source wavefields and back propagate the data residual to calculate the gradient, (3) calculate the searching direction using the conjugate gradient method and the step length, update the initial model and (4) redo (1)-(3) until the updated model is accepted.

The wavefield construction can be achieved by the finite-difference scheme. The forward wavefield at time step $n+1$ is

$$p^{n+1} = 2p^n - p^{n-1} + \Delta t^2 v^2 L p^n + \Delta t^2 f^n \tag{5.1}$$

and the backward wavefield at time step $n-1$ is

$$p^{n-1} = 2p^n - p^{n+1} + \Delta t^2 v^2 L p^n + \Delta t^2 f^n. \tag{5.2}$$

L is the Laplacian operator and f^n is the source. The source term changes according whether the wavefield is forward or backward propagated. When the wavefield is propagated forward, f^n can be set as a wavelet. Conversely, f^n represents the data residual when the wavefield is propagated backward. Strictly speaking from the physical point of view, the source should be injected (added) to the wavefield, while the data residuals should be equated to the wavefield. The reason is that the source wavefield is energy added, while the residual data are observed points. In reality, there are many different arguments and implementations that blurs this distinction. For example, when implementing the adjoint operator, the residual wavefield is added instead of equated.

Same as the FWI in the frequency domain, time domain FWI minimizes the difference between observed data and synthetic data:

$$\Phi(m) = \frac{1}{2} \sum_{n_g} \sum_{n_s} \sum_{n_t} ||p_{syn}(m, x_s, x_g, t) - p_{obs}(x_s, x_g, t)||^2. \tag{5.3}$$

The Taylor's expansion of the objective function is

$$\Phi(m + \alpha \Delta m) = \Phi(m) + \alpha \Delta m^T g + \frac{1}{2} \alpha^2 \Delta m^T H \Delta m + ... \tag{5.4}$$

where Δm is the model update during each iteration, g and H are the first and second order derivatives of the objective function respectively. For the gradient based method, the model is updated by

$$m_{k+1} = m_k + \alpha \Delta m_k. \tag{5.5}$$

For the descent direction Δm, we set

$$\Delta m_k = \begin{cases} -g & k = 0 \\ -g + \beta_k \Delta m_{k-1} & k \geq 1 \end{cases} \tag{5.6}$$

There are many different ways to calculate β_k, here I list four best known formulas which are named after their developers:

- Hestenes-Stiefel:

$$\beta_k^{HS} = -\frac{g^T \left(g_k - g_{k-1}\right)}{\Delta m_{k-1}^T \left(g_k - g_{k-1}\right)} \tag{5.7}$$

- FletcherReeves:

$$\beta_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} \tag{5.8}$$

- PolakRibire:

$$\beta_k^{PR} = \frac{g_k^T \left(g_k - g_{k-1}\right)}{g_{k-1}^T g_{k-1}} \tag{5.9}$$

- DaiYuan:

$$\beta_k^{DY} = -\frac{g_k^T g_k}{\Delta m_{k-1}^T \left(g_k - g_{k-1}\right)} \tag{5.10}$$

The first derivative of the objective function is

$$\begin{aligned} g &= \mathbf{Re}\left(\sum_{n_g}\sum_{n_s}\sum_{n_t}\left(\frac{\partial p_{syn}(m, x_s, x_g, t)}{\partial m}\right)^* \Delta p(x_s, x_g, t)\right) \\ &= \mathbf{Re}\left(\sum_{n_g}\sum_{n_s}\sum_{n_t}\left(\frac{\partial^2 p_{syn}(m, x_s, x_g, t)}{\partial t^2}\frac{2}{v^3(x)}\right)^* \left(G\left(x_r, x, -t\right) \bigstar \Delta p(x_s, x_g, t)\right)\right) \end{aligned} \tag{5.11}$$

where $*$ represents conjugation and $\bigstar$ is convolution .

Comparing the gradient equation in the frequency domain,

$$g = \mathbf{Re}\left(\sum_{n_\omega}\sum_{n_s}\left(u(m,x_s,\omega)^\dagger\left(\frac{\partial A(m,\omega)}{\partial m}\right)^\dagger(A^{-1}(m,\omega))^\dagger R^\dagger \delta d\right)\right) \tag{5.12}$$

the second derivative with respect to time is $-\omega^2$ in the frequency domain which is contained in the term $\frac{\partial A(m,\omega)}{\partial m}$. Also, if we write this term in velocity $\frac{\partial A(m,\omega)}{\partial v}$, the term $\frac{2}{v^3(x)}$ will also appear in the frequency equation. In equation (5.11), $G\left(x_r, x, -t\right) \bigstar \Delta p(x_s, x_g, t)$ is the time-reversed wavefield using the data residual as the source. In equation (5.12), this term becomes the multiplication of the inverse of the impedance matrix $A^{-1}(m,\omega)$ (the Green's function) and the data residual δd. Therefore, the gradient of the objective function in the time and frequency domain are equal.

Cycle skipping artifacts

FWI minimizes the differences between the observed data and synthetic data. When the initial model is not accurate, FWI will suffer from cycle skipping artifacts because of the oscillations of the seismic data (Warner and Guasch, 2014). Figure 5.1 (Virieux and Operto, 2009) is the schematic of cycle skipping artifacts in FWI. When the time difference is larger than $T/2$, the n_{th} period of the synthetic data (the dash curve above) will fit the $(n-1)_{th}$ period of the observed data (the soild curve in the middle), which causes erroneous result. The dash curve below illustrates the scenario when the misfit is less than $T/2$. Cycle skipping is more likely to happen in time domain FWI because of the more difficult implementation required to apply the multigrid approach. With the multigrid method, the long-wavelength components of the model are updated first, moving to the short-wavelength components gradually as the current model becomes closer to the global minimum (true model). For the frequency domain FWI this approach arises naturally, while for the time domain it requires to bandpass the data and re-start the inversion again for each frequency band chosen.

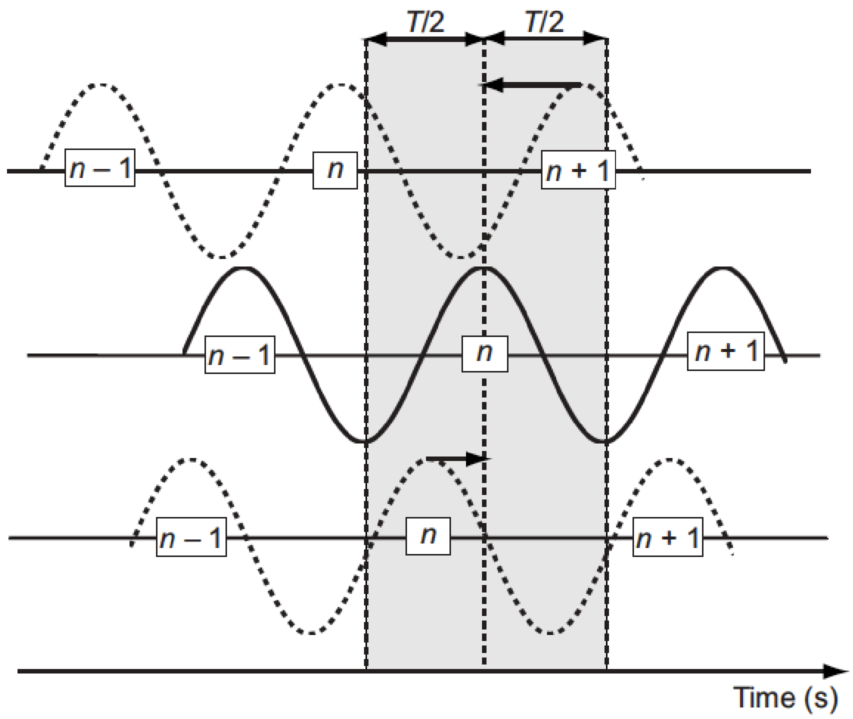


Figure 5.1: The cycle skipping artifacts in FWI problem. This is Figure 7 of Virieux and Operto (2009)

5.2 Numerical examples

In this section, I will show an example of time domain FWI using nonlinear conjugate gradient method to calculate the searching direction and update the model iteratively. Similarly to the previous examples, I use a resampled version of the Marmousi model as the true model, which shown in Figure 5.2. Figure 5.3 shows the initial model used in the time domain FWI. The model is obtained by applying a triangle filter in both horizontal and vertical directions. I set seven shots on the surface of the model of which four are shown in Figure 5.4. FWI results after 5, 10, 30 and 50 iterations are shown in Figure 5.5. Results clearly improve with iterations but after 50 iterations this improvement becomes too small to notice. This is also reflected in the misfit curve (Figure 5.6). Figure 5.7 is a comparison between time and frequency domain FWI using the same parameters. The frequency domain result contains more information in the deeper sections, because the velocity model can be updated from long wavelength to short wavelength components and the cycle skipping artifacts can be mitigated.

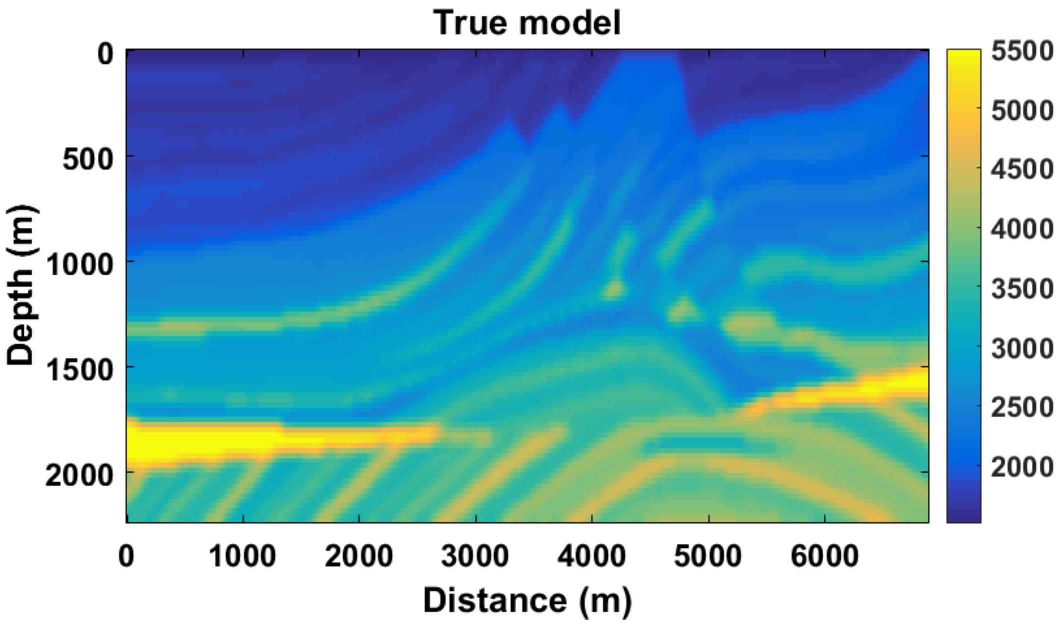


Figure 5.2: The true Marmousi model used in the time domain FWI. The dimensions are 288×94. The true model is smoothed slightly by a triangle filter.

5.3 Conclusions

In this chapter, I illustrate the algorithm of the time domain FWI using nonlinear conjugate gradient method to calculate the searching direction. I show that the gradient computation in the time and frequency domain are equivalent. Different from the frequency domain method, the gradient in the time domain is calculated shot by shot. Because the wavefields are not saved for each shot, the memory required in the time domain is usually less than that in the frequency domain. However, the calculation time is linear with the number of shots, differently from the frequency domain where there is a computation saving when solving simultaneously for several shots. In practice, FWI in the frequency domain is more efficient than in the time domain only for 2D surveys. In 3D data, the cost of the solution of Helmholtz equation is too large for its use in industrial settings.

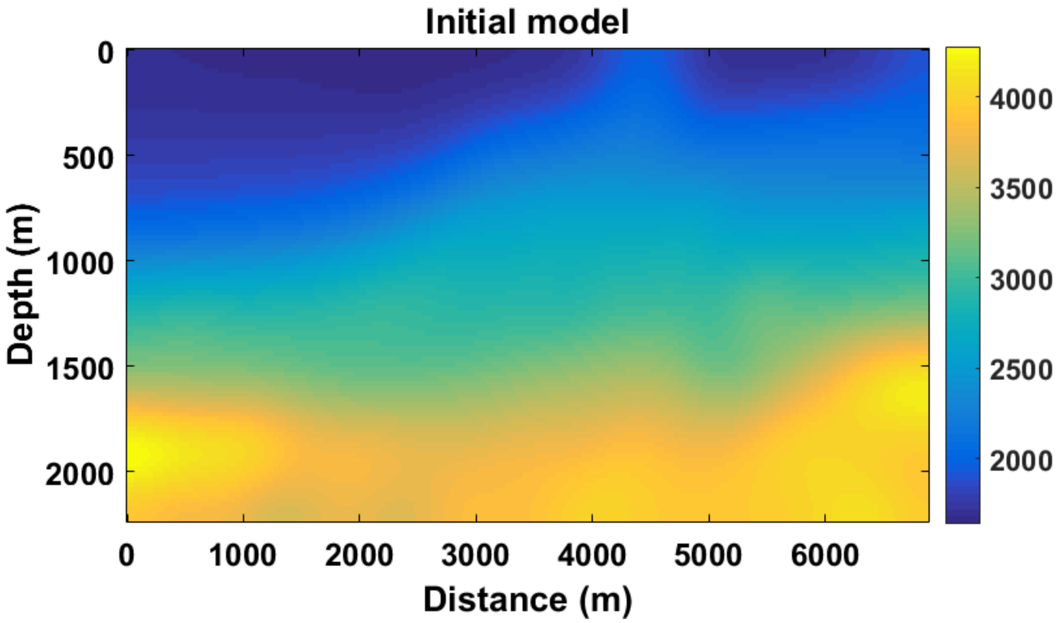


Figure 5.3: The initial Marmousi model used in the time domain FWI. The model is obtained by applying the triangle filtering in both horizontal and vertical directions.

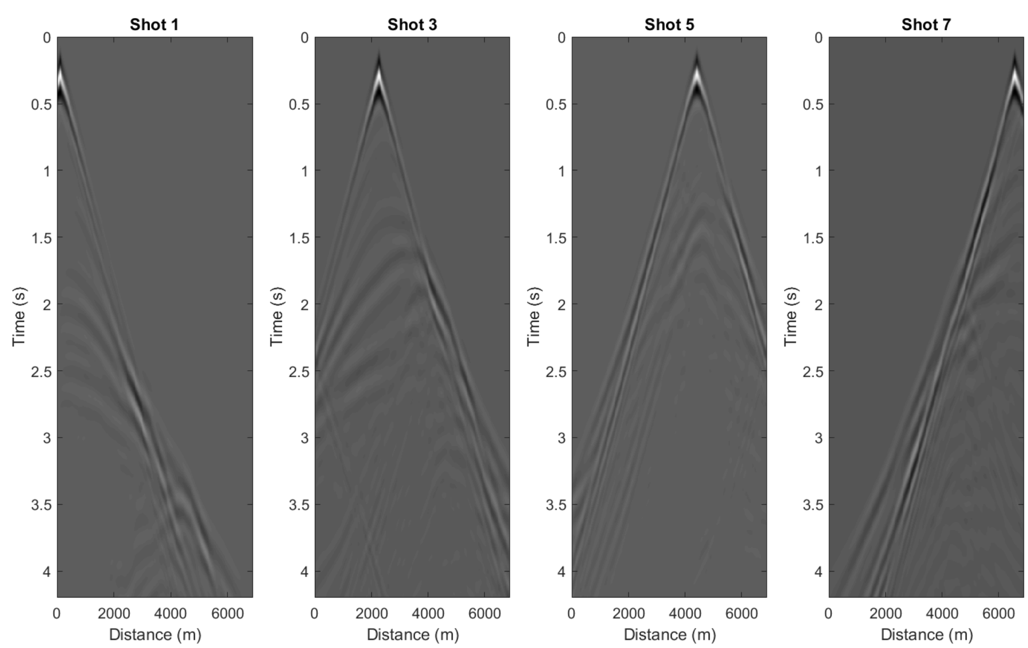


Figure 5.4: Four of the seven shots records used for the FWI example (arbitrarily chosen).

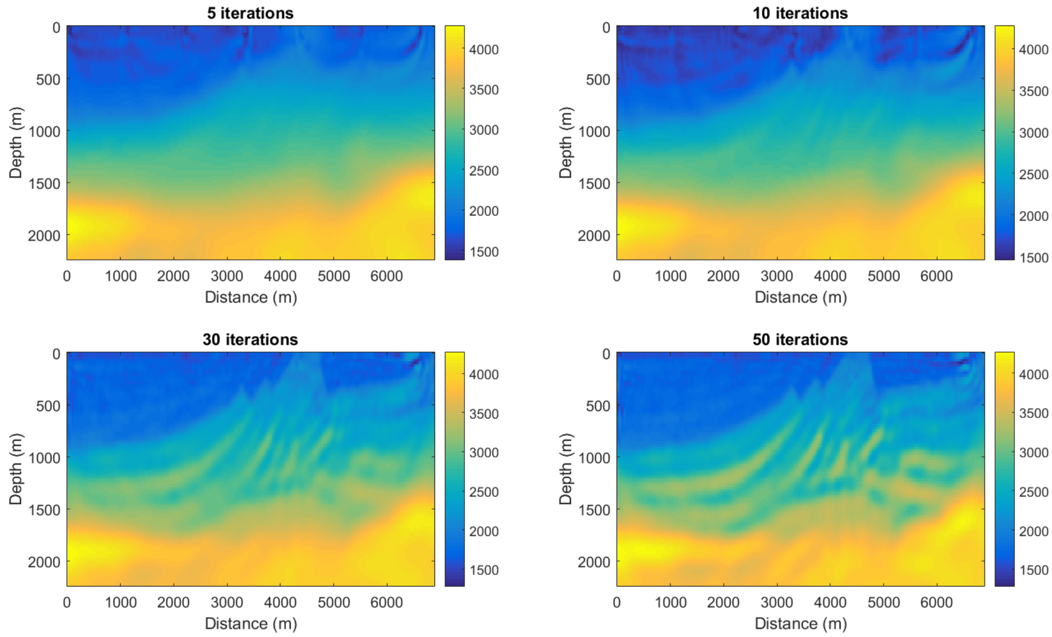


Figure 5.5: The velocity model updates at 5, 10, 30 and 50 iterations.

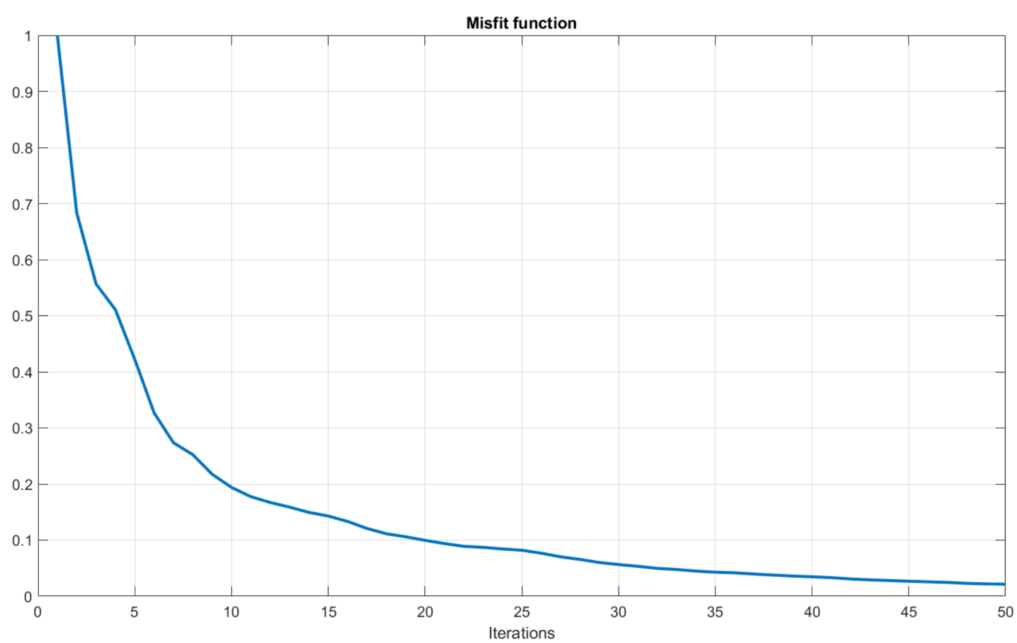


Figure 5.6: The misfit curve of the objective function.

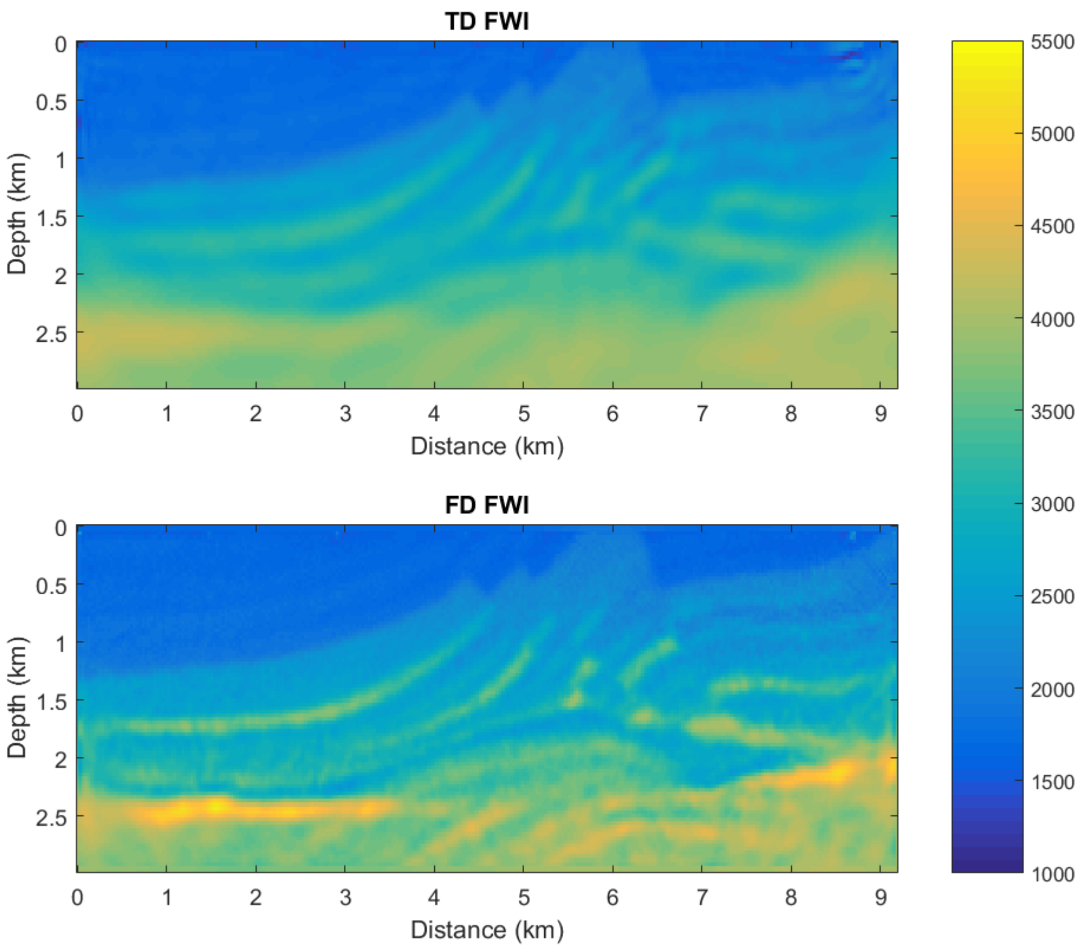


Figure 5.7: The comparison of time and frequency domain FWI using the same parameters. The frequency domain result contains more information of the deeper layers.

Chapter 6

The comparisons of different imaging conditions

6.1 Theory

The goal of migration is to image the reflectors' locations as deduced from seismic data. Imaging condition is the connection between the wavefields and the reflectors.

The imaging of the reflectors was first proposed by Hagedoorn (1954). Claerbout used deconvolution between downgoing and upgoing wavefields in the one dimensional case. This is because reflectors exist at points in the ground where the first arrival of the downgoing wave is time coincident with an up-going wave (Claerbout, 1971). This can be expressed as:

$$R(x, z) = \sum_{n_s} \frac{g(x, z = f(t))}{s(x, z = f(t))}, \tag{6.1}$$

where x and z represent horizontal and vertical coordinates respectively, $R(x, z)$ is the reflectivity, $g(x, z = f(t))$ is the upgoing wavefield and $s(x, z = f(t))$ is the downgoing wavefield. The time coordinate appears in the wavefield expressions just as an argument for the depth. That is because the reflectors at different depths are reached through propagation in time. Since this distinction adds cluttering to the equations, I will in general refer to the wavefields as $s(x, z, t)$ and $g(s, z, t)$.

This division is clearly unstable for positions in the subsurface where the downgoing wavefield is small, that is for deeper reflectors where illumination is weak. The image is usually obtained by the zero-lag cross-correlation imaging condition with summation over shots and time:

$$R(x, z) = \sum_{n_s} g(x, z, t) \star s(x, z, t), \tag{6.2}$$

where $\star$ represents the cross-correlation. For LSRTM in the time domain, a different imaging condition formulation, which resembles the FWI gradient, is more convenient (Trad, 2018). Because the Born modeling formulation has a term $\frac{\partial^2 u_0}{\partial t^2}$ as a source (equation 2.21), to make the operator adjoint, we also use the second derivative in the source wavefield of the imaging condition, which is

$$R(x, z) = \sum_{n_s} g(x, z, t) \star \ddot{s}(x, z, t). \tag{6.3}$$

In the frequency domain, the cross-correlation becomes a multiplication:

$$R(x, z) = \sum_{n_s} \sum_{n_\omega} S^*(x, z, \omega) G(x, z, \omega). \tag{6.4}$$

$*$ denotes the complex conjugate.

In LSRTM, the inverse of the Hessian matrix ($L^T L$) is solved by an iterative scheme (Figure 2.9). Because the Hessian contains illumination information, the image resolution is usually improved when a least-squares scheme is used to calculate the reflectivity. However, as discussed before, the large computational cost of LSRTM often prevents its application. A simpler and computationally cheaper way of approximating the inverse Hessian is to use the source auto-correlation in the denominator to compensate for differences of source energy at different locations (Trad, 2018). This gives a deconvolution imaging condition:

$$R_d(x,z) = \sum_{n_\omega} \frac{\sum_{n_s} S^*(x,z,\omega)G(x,z,\omega)}{\sum_{n_s} S^*(x,z,\omega)S(x,z,\omega)}. \tag{6.5}$$

To make this deconvolution imaging condition stable for frequencies where the source is weak, we can add a small value ϵ in the denominator (also known as Tikhonov's regularization):

$$R_d(x,z) = \sum_{n_\omega} \frac{\sum_{n_s} S^*(x,z,\omega)G(x,z,\omega)}{\sum_{n_s} S^*(x,z,\omega)S(x,z,\omega) + \epsilon}. \tag{6.6}$$

In the time domain, an illumination compensation imaging condition which resembles deconvolution imaging condition is:

$$R(x,z) = \sum_{n_s} \frac{g(x,z,t) \star s(x,z,t)}{\sum_{n_s} s(x,z,t) \star s(x,z,t)}. \tag{6.7}$$

Although the deconvolution imaging condition is important for illumination compensation and improves resolution, cross-correlation imaging condition is commonly used instead because amplitudes have many other sources of error, like attenuation and scattering which are rarely corrected for. Also, the computation cost is reduced using cross-correlation. In the following section, I will show examples to compare the cross-correlation and deconvolution imaging conditions, both in the time and frequency domain.

6.2 Numerical examples

Figure 6.1 and 6.2 show the comparison of cross-correlation and illumination compensation imaging conditions in the time domain with and without illumination compensation, for the Marmousi model. To illustrate the benefits of the illumination compensation imaging condition, I will use very few shots covering only half of the model on the horizontal direction. In Figure 6.1, the energy is weak on the right side of the model because of the limited shots coverage. This problem is mitigated in Figure 6.2.

Figure 6.3 and 6.4 are the comparisons of cross-correlation and deconvolution imaging condition in the frequency domain. In the deconvolution imaging condition, the energy in the deeper layers is stronger than the cross-correlation imaging condition, and the image is more balanced, but the quality of the image is not as good as time domain. In the frequency domain, because the denominator term is restricted to each frequency individually, there is a larger chance of instability and regularization becomes important.

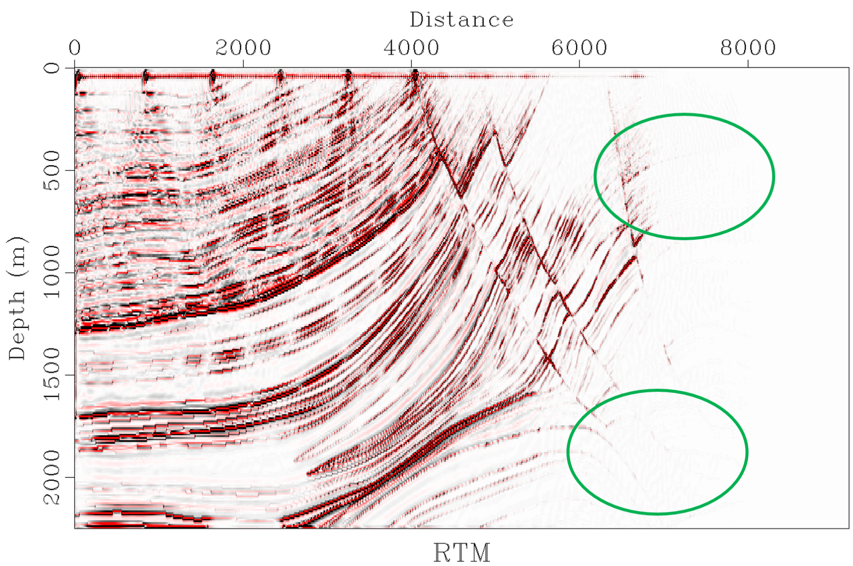


Figure 6.1: The result of cross-correlation imaging condition of RTM in the time domain. Because of the limited shots coverage, the energy in the green circle is very weak.

6.3 Conclusions

In this chapter, I compared the cross-correlation, illumination compensation and deconvolution imaging condition in different domains. The application of deconvolution imaging condition can provide good illumination compensation when there are recording deficiencies, which is an simpler way of improving resolution compared to least-squares scheme. LSRTM needs more computer resources and requires the operators to be accurate and exactly adjoint. Also it is often too expensive computationally for its routine application. In general, there is a trade-off between efficiency and image quality. If there are enough shots coverage, cross-correlation is an efficient choice and can provide reasonable improvements over de-

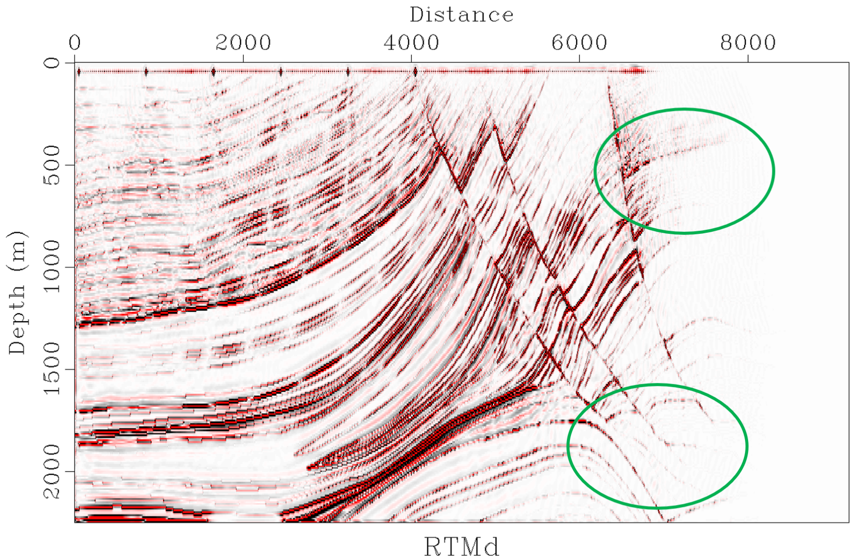


Figure 6.2: The result of illumination imaging condition of RTM. The weak energy caused by the limited shots coverage is mitigated in the green circle.

convolution. For the limited shot coverage and computer resources, deconvolution imaging condition is a relatively inexpensive method to compensate for differences in illumination, but will require regularization.

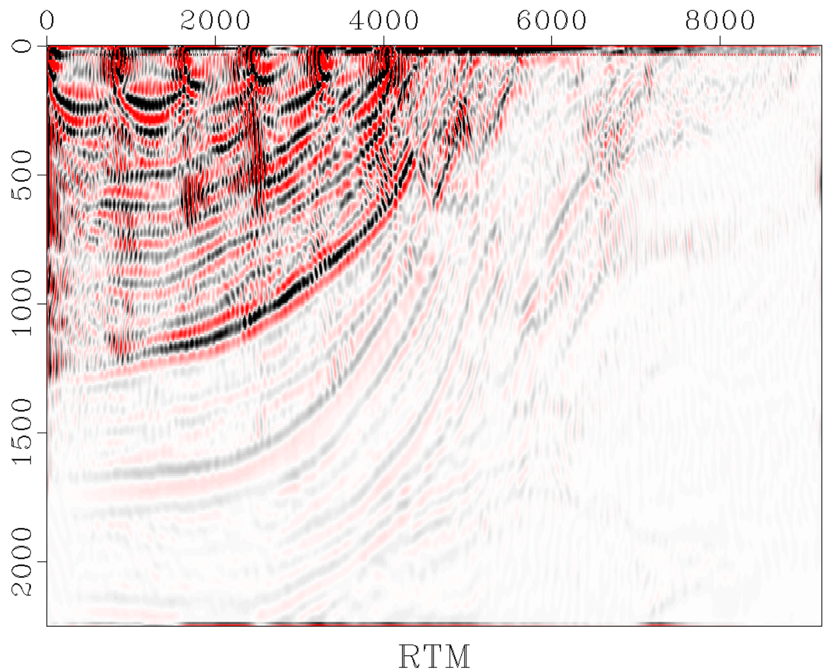


Figure 6.3: The result of cross-correlation imaging condition of RTM in the frequency domain.

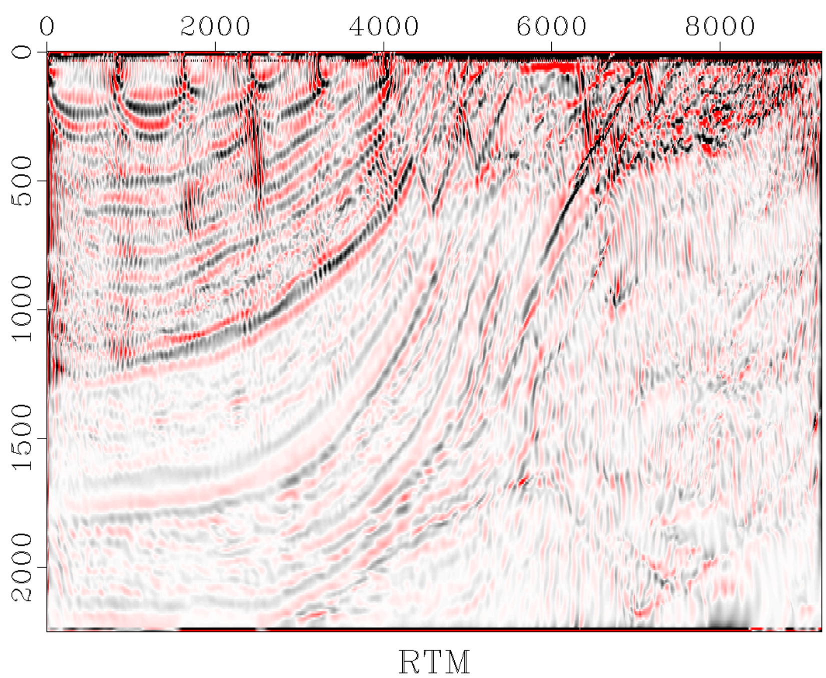


Figure 6.4: The result of deconvolution imaging condition of RTM in the frequency domain.

Chapter 7

Conclusions

Least-squares reverse time migration (LSRTM) and full waveform inversion (FWI) are both practical methods in seismic inversion. Although their goals are somewhat different, their principles and implementations are quite similar so that it is difficult to establish a clear separation between them. In this thesis, I investigated LSRTM and FWI both in the time and frequency domains in the acoustic media, with the goal of understanding their connections, similarities and differences.

The goal of each method is different, since for LSRTM the model is the reflectivity while for FWI the model is the velocity. More generally, FWI can use as a model different physical properties (e.g. velocity, density, Q, etc). These properties are also the goal of reflectivity inversion. Therefore, we see that both LSRTM+reflection inversion and FWI have at the end a similar goal, that is reservoir characterization.

This connection becomes actually much stronger as we investigate their formulations. LSRTM can be referred to as a linearized FWI. Both use a similar cost function, the energy of the differences between observed data and predictions, but different methods to predict the data. Because their definition of the model is different, their operators are different.

Generally speaking, LSRTM can be treated as the inner loop of the general FWI algorithm and solved in the image domain. Compared to FWI, LSRTM uses reflection data and requires an accurate operator to connect the model and data space. The result is very sensitive to the accuracy of the operator. FWI is more robust to velocity errors in the initial model, because the operator changes with the model updates. Prediction errors due to velocity uncertainties can not be fixed by changing the model (reflectivity) in LSRTM, but they can in FWI (velocity). On the other hand, because LSRTM is a linear problem and FWI is a non-linear problem, FWI fails to converge if the initial model is not close to the final model. LSRTM will always converge to some solution, but this solution will be affected by the operator errors.

The features of LSRTM and FWI in different domains are also different. In the frequency domain, LSRTM has the advantage of easy formulation and simultaneous calculation of multiple shots. The computation cost does not increase linearly with the number of shots. However, the memory requirement increases significantly. This makes frequency domain LSRTM and FWI not practical for 3D surveys. Time domain LSRTM in general is cleaner and more robust since the information for all frequencies constraint the problem simultaneously. The frequency domain brings some economy of computation for FWI but not for LSRTM. For FWI, information across frequencies is redundant, and it is possible to reduce the number of calculated frequencies. Conversely, for LSRTM, missing frequencies in the data result in missing frequencies in the reflector. Therefore, in principle, all frequencies need to be calculated at least once.

FWI in the frequency domain is advantageous because the multigrid approach allows us to avoid cycle skipping. Low frequencies are less susceptible to velocity errors and less non-linear, using only low frequency during initial iterations, and permits one to get a velocity model closer to the true model. When high frequencies are calculated, they benefit from the corrections achieved in lower frequencies. For LSRTM this mechanism does not apply because the operator is not changing during each iteration. However, it is possible that a similar idea could be used to make LSRTM more robust to velocity errors.

From the operator point of view, by comparing the formulations of RTM, LSRTM, and

FWI in the time and frequency domains, we can see some trade-off between dividing the data on shots or frequencies. The time domain implementations seem more stable and produce clean results. With proper implementation, the algorithms are efficient and feasible even for 3D surveys. However, to achieve this efficiency significant programming work is required. For example, forward wavefields should be stored only at boundaries and reconstructed. Boundary conditions should be carefully implemented. In the case of LSRTM, the migration operator should be built as a perfect adjoint of the Born modelling operator. The frequency domain versions have some flexibility that permits one to apply multigrid approaches for example with an easier implementation than in the time domain. The deconvolution image condition in the frequency domain can be used to eliminate, at least in principle, some of the band limited nature of the wavelet.

From the optimization point of view, there are differences between linear inversion in LSRTM and non-linear inversion in FWI. However, we can see that at the centre of FWI, the gradient is obtained by RTM and the inverse of the Hessian by LSRTM in the image space. Solving the linear relationship between the gradient, Hessian and searching direction is a LSRTM problem in the image space.

Bibliography

Ajo-Franklin, J. (2005). Frequency-domain modeling techniques for the scalar wave equation: An introduction.

Alford, R. M., Kelly, K. R., and Boore, D. M. (1974). Accuracy of finite-difference modeling of the acoustic wave equation. *Geophysics*, 39(6):834–842.

Alterman, Z. and Karal Jr, F. (1968). Propagation of elastic waves in layered media by finite difference methods. *Bulletin of the Seismological Society of America*, 58(1):367–398.

Boore, D. M. (1970). *Finite difference solutions to the equations of elastic wave-propagation, with application to Love waves over dipping interfaces.* PhD thesis, Massachusetts Institute of Technology.

Cerjan, C., Kosloff, D., Kosloff, R., and Reshef, M. (1985). A nonreflecting boundary condition for discrete acoustic and elastic wave equations. *Geophysics*, 50(4):705–708.

Chen, K. and Sacchi, M. D. (2018). Time-domain elastic gauss-newton full-waveform inversion via matrix-free adjoint-state method.

Claerbout, J. F. (1971). Toward a unified theory of reflector mapping. *Geophysics*, 36(3):467–481.

Claerbout, J. F. (1992). *Earth soundings analysis: Processing versus inversion*, volume 6. Blackwell Scientific Publications London.

Clayton, R. and Engquist, B. (1977). Absorbing boundary conditions for acoustic and elastic wave equations. *Bulletin of the seismological society of America*, 67(6):1529–1540.

Engquist, B. and Majda, A. (1977). Absorbing boundary conditions for numerical simulation of waves. *Proceedings of the National Academy of Sciences*, 74(5):1765–1766.

Fomel, S., Sava, P., Vlad, I., Liu, Y., Jennings, J., Browaeys, J., Bashkardin, V., Godwin, J., Song, X., and Hennenfent, G. (2012). Madagascar software package and reproducible research.

Geng, Y., Innanen, K. A., and Pan, W. (2017). Nonlinear multiparameter full waveform inversion based on truncated newton method. *CREWES Research Report*, 29.

Hagedoorn, J. G. (1954). A process of seismic reflection interpretation. *Geophysical Prospecting*, 2(2):85–127.

Hestenes, M. R. and Stiefel, E. (1952). *Methods of conjugate gradients for solving linear systems*, volume 49. NBS Washington, DC.

Knibbe, H., Mulder, W., Oosterlee, C., and Vuik, C. (2014). Closing the performance gap between an iterative frequency-domain solver and an explicit time-domain scheme for 3d migration on parallel architectures. *Geophysics*, 79(2):S47–S61.

Kosloff, D. D. and Baysal, E. (1982). Forward modeling by a fourier method. *Geophysics*, 47(10):1402–1412.

Lailly, P. (1983). The seismic inverse problem as a sequence of before stack migrations. In *Conference on inverse scattering: theory and application*, pages 206–220. Siam Philadelphia, PA.

Levander, A. R. (1988). Fourth-order finite-difference p-sv seismograms. *Geophysics*, 53(11):1425–1436.

Lines, L. R., Slawinski, R., and Bording, R. P. (1999). A recipe for stability of finite-difference wave-equation computations. *Geophysics*, 64(3):967–969.

Margrave, G. F., Ferguson, R. J., and Hogan, C. M. (2010). Full waveform inversion with wave equation migration and well control. *the 22nd Annual Research Report of the CREWES Project*.

Métivier, L., Brossier, R., Virieux, J., and Operto, S. (2013). Full waveform inversion and the truncated newton method. *SIAM Journal on Scientific Computing*, 35(2):B401–B437.

Mulder, W. A. and Plessix, R.-E. (2004). A comparison between one-way and two-way wave-equation migration. *Geophysics*, 69(6):1491–1504.

Nemeth, T., Wu, C., and Schuster, G. T. (1999). Least-squares migration of incomplete reflection data. *Geophysics*, 64(1):208–221.

Nocedal, J. and Wright, S. (2006). *Numerical optimization*. Springer Science & Business Media.

Pan, W. (2017). *Waveform inversion for estimating subsurface properties: phase-encoding strategies, optimization methods, interparameter tradeoffs quantification and reduction*. PhD thesis, University of Calgary.

Pan, W., Innanen, K. A., and Liao, W. (2017). Accelerating hessian-free gauss-newton full-waveform inversion via l-bfgs preconditioned conjugate-gradient algorithm. *Geophysics*, 82(2):R49–R64.

Pratt, R. G. (1990). Frequency-domain elastic wave modeling by finite differences: A tool for crosshole seismic imaging. *Geophysics*, 55(5):626–632.

Pratt, R. G. (1999). Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model. *Geophysics*, 64(3):888–901.

Schuster, G. T. (1993). Least-squares cross-well migration. In *SEG Technical Program Expanded Abstracts 1993*, pages 110–113. Society of Exploration Geophysicists.

Schuster, G. T. (2017). *Seismic inversion*. Society of Exploration Geophysicists.

Symes, W. W. (2007). Reverse time migration with optimal checkpointing. *Geophysics*, 72(5):SM213–SM221.

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266.

Trad, D. (2018). The importance of modeling operators for least-squares migration. In *SEG Technical Program Expanded Abstracts 2018*, pages 4246–4250. Society of Exploration Geophysicists.

Versteeg, R. (1994). The marmousi experience: Velocity model determination on a synthetic complex data set. *The Leading Edge*, 13(9):927–936.

Virieux, J. (1986). P-sv wave propagation in heterogeneous media: Velocity-stress finite-difference method. *Geophysics*, 51(4):889–901.

Virieux, J. and Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26.

Warner, M. and Guasch, L. (2014). Adaptive waveform inversion-fwi without cycle skipping-theory. In *76th EAGE Conference and Exhibition 2014*.

Yang, P. (2014). A numerical tour of wave propagation. Technical report, Technical report, Xian Jiaotong University.

Yang, P., Gao, J., and Wang, B. (2015). A graphics processing unit implementation of time-domain full-waveform inversion. *Geophysics*, 80(3):F31–F39.