# University of Alberta

Total variation and adjoint state methods for seismic wavefield imaging

by

Amsalu Y. Anagaw

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Geophysics

Department of Physics

© Amsalu Y. Anagaw
Fall 2009
Edmonton, Alberta

**Examining Committee**

Dr. Mauricio D. Sacchi (Supervisor), Geophysics

Dr. Mathieu Dumberry, Geophysics

Dr. Jeffrey L. Kavanaugh, Earth & Atmospheric Sciences

Dr. Massimo Boninsegni (Chair), Physics

# Abstract

Many geophysical inverse problems are ill-posed and have to be regularized. The most often used solution methods for solving ill-posed problems are based on the use of quadratic regularization that results in smooth solutions. Solutions of this type are not to be suitable when the model parameter is piecewise continuous blocky and edges are desired in the regularized solution. To avoid the smoothing of edges, which are very important attributes of an image, an edge-preserving regularization (non-quadratic regularization) term has to be employed. Total Variation (TV) regularization is one of the most effective regularization techniques for allowing sharp edges and the existence of discontinuities in the solutions.

The edge-preserving regularization based on the TV method for geophysical inverse problems to the problem of estimating the velocity perturbation is studied. The acoustic velocity perturbation is assumed to be piecewise continuous and blocky. The problem is based on linearization acoustic modeling using the framework of the single-scattering Born approximation from a known constant background medium. To solve this non-linear and ill-posed problem, an iterative scheme based on the conjugate gradient method is employed. The TV regularization method provides us with the opportunity to recover more useful information of velocity profiles from the measured seismic data. Though it requires more effort in implementing the TV term to control the smoothing and regularization parameter, the algorithm possesses the strong ability of marking the discontinuities and ensures their preservation from over-smoothing.

# Acknowledgements

# Contents

# Appendices

# List of Figures

# CHAPTER 1

# Introduction

## 1.1 Background

Seismic data contains information regarding the geological structure and physical properties of the subsurface. In geophysics, seismic migration is the most common and powerful tool for imaging the subsurface structures.

Seismic migration or seismic imaging is the process of transforming the near surface information - ground movement - recorded by seismometers into images of subsurface structures. In a geophysical seismic experiment, seismic energy propagates from a source to the subsurface. As seismic waves are impelled forward within the subsurface, these waves are reflected, diffracted or refracted at the interface between rock layers due to the variation in rock properties, that is, seismic velocities or densities. These reflected seismic waves from the subsurface are recorded at the surface by seismometers, allowing for conclusions to be drawn on the structures and physical properties of the subsurface. Measuring both the traveltimes from the source to the receiver and the amplitude at the receiver provides information about the structural specific velocities in the subsurface.

### 1.1.1 Seismic migration and wave-equation modeling/migration

Seismic migration, traditionally, comprised of an imaging process, is a data-processing technique that generates an image of Earths structure from observed seismic data. But theoretically, a seismic migration process can be restructured in two steps: forward modeling and imaging. In forward modeling, seismic data are generated by propagating the waves from a

source to scatterer to receivers. Migration attempts to locate acoustic reflection boundaries in the subsurface interior from the recorded seismic data at the surface, and produces an image of the subsurface. The challenge in seismic migration is to accurately locate and identify the positions where the wave energy is reflected and refracted.

Over the past decades, several migration techniques have been developed to perform this task. These techniques can be differentiated by geometrical methods or by methods based on the solution of the wave-equation, which in turn are based either on the ray and scattering theory or the wave-equation migration, which directly solves the wave-equation. Among these methods are the Kirchhoff migration (Schneider, 1978), Finite difference (FD) migration (Claerbout, 1985; Mufti et al., 1996), Fourier finite difference migration (Ristow and Rühl, 1994), Fourier migration (Stolt, 1978) and the phase shift migration or frequency wave number migration (Gazdag, 1978, 1984). Regardless of the methods employed to ascertain the solution to the wave-equation, all of these methods propagate the recorded seismic data back either in time or space to the proper subsurface location.

The Kirchhoff Migration is based on a ray and scattering theory which uses asymptotic solutions valid only for high-frequency approximation (Zhdanov, 2002). Multi-scattering effects are ignored (Schneider, 1978). Due to high frequency approximation, the wave propagation distances between scatterers and sources or receivers are limited to large - more than a few - wavelengths. Therefore, the high frequency approximation and any effects of multi-arrivals limit the accuracy of this migration.

The other method that directly solves the wave-equation is wave-equation migration (Claerbout, 1985; Mufti et al., 1996). Hypothetically, wave-equation migration methods are also based on scattering theory. Depending on what wave-equation migration method is used, assumptions are made just like in Kirchhoff, namely the various approximations to the Greens functions. By and large, wave-equation migration is more accurate than the Kirchhoff Migration as there is no high-frequency approximation involved, and multi arrivals are automatically considered. Most wave-equation migrations are based on the numerical approximation of finite difference methods. Reverse-time migration is an example that uses the finite difference wave-equation modeling as a means of migrating seismic data in the space-time domain. It solves the full (two-way) wave-equation, for example acoustic or elastic waves, by extrapolation in time, allowing waves to propagate in all directions. Though, the methods propagate the wavefields, but the propagation of source and receiver side wavefields, must be considered in conjunction with an imaging condition. FD solutions to the wave-equation are used to perform the migration as a backward time marching scheme (McMechan, 1983; Baysal et al., 1983). The two-way wave-equation finite difference modeling has no dip limitations, and produces all the events associated with the wave-equation such as multiple reflections, head waves and (when the elastic wave-equation

is used) anisotropic effects and mode conversions. Finite difference modeling is therefore an ideal way to obtain realistic seismic data from a model.

The purpose of my study is to investigate the possibility of using non-linear inversion of seismic waveforms on the study of known geophysical inversion algorithms, utilizing the reverse-time migration algorithm. Particular attention is given on the use of finite-difference modeling using the full acoustic wave-equation with constant density (Kelly et al., 1976). How this method can be applied for non-linear inversion of seismic waveform will form the next section of this thesis.

### 1.1.2 Adjoint state method

In practical applications, seismic migration based on the least-squares principle is most widely used for determining the structural profile of the subsurface (Aster et al., 2005). The inversion problem is methodically devised as an optimization problem that requires minimization of the least-squares function between observed and computed data. The full waveform inversion is, in general, a non-quadratic optimization problem. Seismic migration can be regarded as the minimization of the cost function. The optimization of cost function method allows for the determination of the location and the amplitude of the reflector from seismic reflected data measured at the surface (Tarantola, 1984). The locations of the reflectors can be obtained from seismic imaging principle (Clearbout, 1985). Seismic imaging condition is one piece of the migration operator. The migration operator is the first iteration of an inverse (i.e. it is the adjoint operator). This may not be true for all methods of computing an inverse. Lailly (1983) demonstrated that the waveform and the prestack reverse time migration share the same numerical algorithm and the prestack-stack reverse time migration can be regarded as the first iteration. Though the optimization is a non-linear, reverse time migration can also be the solution to the corresponding linearized optimization problem. Theoretically, in reverse time migration, one has to solve numerically the forward ( or from experiment it is the data recorded by a seismic reflection survey ) and backpropagate wavefields in different method. The finite difference approximation scheme is among many other methods for computing wavefield (Mora, 1987).

The backpropagation technique of reverse time migration, waveform inversion based on reverse time of solution to the corresponding linearized optimization problem, introduced by Lailly (1983) and Tarantola (1984) can be used for seismic waveform inversion. In this method, the first step is to compute the cost function that measures the difference between observed seismic data and computed data from the forward modeling. This residual data set indicates how accurate the current model with respect to the original data actually is. The backpropagated wavefield is, then, computed using the reverse-time technique. Next

the sources forward-propagated wavefield is cross-correlated with the backward-propagated wavefield of the data residual at each time step and summed over all the time steps to produce the total gradient volume. Finally, at each special point, the amplitude of the gradient is checked to be directly proportional to the velocity field change. This method has been used for large-scale geological model in frequency domain waveform inversion (Stekl and Pratt, 1998) Pratt et al. (1998), in the time-domain traveltime tomographic inversion and waveform inversion (Gauthier et al., 1986; Zhou et al., 1995).

In seismic waveform inversion, in order to minimize the cost function, several forward modeling and residual backpropagations are required to gradually update the velocity field. The gradient of the cost function is related to the velocity field and can be computed by using the Fréchet derivative, the perturbation theory or the adjoint state method. However, with Fréchet derivatives, as the size of the problem increases, a large linear system is created and requires excess computer memory. The adjoint state method enables us to compute the gradient of the misfit function without computing the Fréchet derivatives (Plessix, 2006; Symes, 2007). The result obtained by using the adjoint state method gives the same output as the result obained based on the perturbation approach for acoustic wave-equation. In the adjoint state method, a state of variables, whose physical meaning defines with the backpropagation of waves (see Chapter 4), is defined to be the solution of a linear system of equations.

## 1.2   Inverse problem and regularization

Wavefield inversion, such as velocity, density, etc., from the measured data, is, however, much more complicated than the migration problem. This is due to complexity of the physical quantity properties profile inside Earth with the intervals of such quantities not well determined. Estimation of such parameters remains a fundamental problem in geophysical research areas although there have been many efforts to address it.

The most common challenge in inverse geophysical problems is that problems are mathematically ill-posed. In other words, their solutions are not unique or highly sensitive to changes in the data because they operate with insufficient data and most observations are subject to noise (Jackson, 1972). In order to overcome this kind of problem, stabilization is needed. Supplemental information must be introduced to least-squares inversion of wavefield. For this reason, most inverse calculations involve weights and must be regularized (Tikhonov, 1963a). Least-squares inversion is the basis of many parameter estimations and data fitting procedures. Additionally, the linear least-squares methods usually involve singular matrices or matrices that are numerically singular and invertible. By regularizing the least-squares, these singularities can be managed.

The method of regularization has become an indispensable part of the inverse-problem theory particularly for ill-posed inverse problems. This method of regularization has also found many applications in geophysical problems: travel time tomography (Bube and Langan, 2008; Clapp et al., 2004), migration velocity analysis, high-resolution Radon transform (Trad et al., 2002, 2003), to name a few.

However, regularization of the solution of inverse problems in geophysics depends primarily on the choice of the regularization method used. This choice depends on the problem under consideration in which the solution becomes correct and guarantees the stability of the solution of the inverse problem. For example, the common regularization practiced was based on Tikhonov's idea (Tikhonov, 1963a,b) , which enforces spatial smoothness uniformly on the output image. This regulation term is a quadratic regularization that results in linear solutions computationally easily to solve for linear inverse problems. A consideration, however, is that these regularization terms apply homogeneous smoothing to the recovered image so consequently they will tend to blur sharp material properties boundaries. If the desire is to preserve sharpness of edges in the given observation, then non-quadratic regularizations function must be found (Charbonnier et al., 1997).

## 1.2.1 Edge-preserving regularization

Regularization of the inverse problem provides an opportunity to rediscover the origin of the models parameter by performing a reverse operation. If the model parameter or the image to be recovered is piecewise continuous and blocky, quadratic regularization methods tends to strongly penalize discontinuities, smoothing the object edges and producing a blurry image around each piecewise discontinuous edge. In order to preserve such edges, preserve discontinuities in reconstructed profiles or piecewise function, a non-linear or non-quadratic regularization method must be employed with strong edge-preserving properties.

The application and the performance of edge-preserving regularization has been studied in different areas of astronomy (Geman and Yang, 1995; Molina et al., 2001), medical imagining (Charbonnier et al., 1997) and in geophysics (Farquharson and Oldenburg, 1998; Portniaguine and Zhdanov, 1999), spectral decomposition, edge-preserving imaging (Youzwishen and Sacchi, 2006), velocity estimation (Valenciano et al., 2004). In essence, the quadratic penalty term is replaced by another non-quadratic function allowing for the presence of discontinuities. A Total Variation (TV) based edge-preserving regularization method is among such edge-preserving regularization methods that have the ability to reconstruct or restore images by preserving edge information of discontinuous model parameter without over-smoothing the reconstructed images.

## 1.2.2   Total variation regularization

Most regularization techniques, unlike the edge-preserving regularization (Valenciano et al., 2004; Youzwishen and Sacchi, 2006) are based on the assumption that the model parameter to be reconstructed is smooth and continuous. The total variation is independent of these assumptions and it preserves the edge information in the reconstructed image. If the image to be restored is blocky or continues piecewise, then the total variation only measures the sum of the magnitudes of the jumps and oscillatory around the edges in the image.

The TV leads to an edge-preserving regularization functional that encourages smoothing in the direction tangential to the edges and not in the direction orthogonal to the edges. This preserves edge information of the discontinuous model parameter and allows for sharper reconstruction.

The total variation regularization was first introduced by Rudin-Osher-Fetemi (ROF) for image de-noising or de-blurring (Rudin et al., 1992). In total variation de-noising, one attempts to remove noise from a signal or image by solving a nonlinear minimization problem involving a total variation criterion. This method relies on the minimization of TV norm, the $L_1$ norm. Minimizing the TV norm of the estimated solution de-noises images subject to constraints which relate the solution to the measured image and the estimated image. This minimization scheme offers the best combination of noise removal and feature preservation.

Several approaches based on this idea have recently been shown to be very effective, particularly for de-noising functions with discontinuities. It has also evolved into a more general tool for solving a variety of image restoration problems, such as for deconvolution, inpainting (Vogel and Oman, 1995; Chan and Kang, 2006) and in medical imaging denoising (Wang and Zhou, 2006; Christiansen et al., 2007). Recently this approach has also been used for full waveform inversions for seismic velocity and anelastic losses in heterogeneous structures (Askan et al., 2007).

Though the total variation regularization method is widely used for de-noising, its application in seismic inversion has not been commonly used. However, the application of edge-preserving regularization to the problem of estimating velocity perturbation using the framework of modified Cauchy prior distribution based on linearizing the single-scattering Born approximation from a known reference medium has been studied (Youzwishen and Sacchi, 2006). Their result encourages further exploration into the same geometrical experiment with different potential edge-persevering regularization term such as total variation regularization.

## 1.3   Summary of thesis work

This thesis focuses on two objectives: the application of total variation regularization in geophysical problems using inverse theory and subsequently the adjoint state based optimization algorithm for non-linear waveform inversion. Accomplishing these objectives is organized as follows.

Chapter 2 - Review of the basic assumptions of the Born scattering method in seismic modeling and the adjoint operators. The Born approximation method for linearizing the Earth's response from a source at the surface is briefly discussed and used to exemplify the concept of migration and inversion. The derivation of the Born approximation and assumption for seismic modeling/migration starts from the basic scattering theory, and the inverse of scattering problem using the framework of Born approximation is reviewed.

Chapter 3 - Discussion of the application of edge-preserving regularization to the problem based on the total variation method for estimating acoustic velocity perturbation from multi-source, multi-receiver geometrical experiment. The optimization problem is solved by the conjugate gradient algorithm in an iterative way. In doing, a routine based on Iterative Re-weighted Least-Squares (IRLS) (Scales and Smith, 1994) is employed to solve the sparsely regularized problem.

Chapter 4 - Review of the application of the adjoint state method for non-linear waveform inversion. The method offers to compute efficiently the gradient of the cost function. The two-way Helmholtz wave-equation in time domain is used for forward and backpropagation of wave. Solutions to the Helmholtz wave-equation are computed using an explicit time-marching method. An adjoint state method is used to compute the backpropagated waves of the residual data. The image is then obtained by the cross correlation between the forward and backward propagation waves (Claerbout, 1971).

Chapter 5 - Review of the summary of this thesis and further research.

# CHAPTER 2

# Linear scattering theory

## 2.1 Introduction

Non-linear wavefield inversion aimed at obtaining the structual image of the subsurface geology and the estimation of the subsurface physical properties from measured data at the surface are, in general, computationally expensive due to the highly non-linear properties of these parameters (such as velocity and density) and the measured data. The non-linear inversion algorithms usually require repetitive solutions of the forward problem defined by wave propagation, which is followed by the update of the inverse problem solution. The most common approach to overcoming the difficulties posed by the non-linearity is to linearize the inverse problem using the Born approximation and solve the resulting linear problem to reconstruct the model parameters. Linearization is important in geophysical problems so that an approximate inversion method can be employed, which is computationally cheaper than the non-linear inversion.

In this paper, we review the application of scattering problems in geophysics using the Born approximation. The main motivation behind the development the Born approximation is that the wavefield can be linearized and expressed linearly in terms of the acoustic velocity perturbation. This approximation is based on the fact that scattering is weak and assumes a small perturbation on the background acoustic potential, which are assumed to be homogeneous. The homogeneous backgrounds allow for analytical solutions of the forward problem for any arbitrary geometry. Analytical solutions for the linearized inverse problem can be solved computationally as well. However, the linearized forward scattering equation can be derived for any arbitrary background velocity and geometry. For the aforementioned reasons, the Born approximation is often employed because of its simplicity of implementation

and computational advantage (Beylkin, 1985; Beylkin and Burridge, 1990).

The inverse scattering problem for an acoustic medium based on the Born approximation is formulated by using the constant velocity background. A constant acoustic velocity medium is probed by a point source, and the scattered field is observed along a curved receiver array located outside the region where the medium velocity is different from the assumed background velocity. In the Born approximation, the model parameter that we want to reconstruct is viewed as a small perturbation about an assumed background velocity model, and the scattered wavefield is expressed linearly in terms of this perturbation. Though multiples due to the background model are included in the scattered field, in the Born approximation multiple scattered waves due to the velocity perturbations are neglected. In order to relate the scattering wavefield to the acoustic velocity potential, a high frequency single scattering asymptotic approximation and an additional approximation using a generalized Radon transform introduced by Beylkin are used (Beylkin, 1985).

In this chapter, the basic theory of scattering problems, the Born approximation and the derivation of linearization of scattering waves, which are often utilized in the geophysics community, are reviewed (Beylkin, 1985; Beylkin and Burridge, 1990; Youzwishen, 2001). The objective is to apply this linearization method to the regularized least-square inversion in Chapter 3.

## 2.2    Linearizing the scattering problem

In order to derive the scattering wavefield and compute the forward modelling, we first define the wave propagation of an incident field resulting from the point source in the constant background velocity medium. For acoustic waves in a medium with constant density, the wave-equation is governed by the Helmholtz wave-equation

$$\left(\nabla^2 + \frac{\omega^2}{c(\mathbf{x})^2}\right) u(\mathbf{x}, \mathbf{x}_s, \omega) = -f(\omega)\delta(\mathbf{x} - \mathbf{x}_s), \tag{2.1}$$

where $u(\mathbf{x}, \mathbf{x}_s, \omega)$ is the complex wavefield in the subsurface point $\mathbf{x}$ due to the $s^{th}$ excitation source located at $\mathbf{x}_s$, $f(\omega)$ is the source function, $\omega$ is the angular frequency and $c(\mathbf{x})$ is the wave velocity of the background medium from which we can solve the wave propagation problem. $\mathbf{x}$ is a vector that represents a coordinate $(x, y, z)$. The energy content and the shape of the waveform that results from such a wave-equation depends on the type of source we chose.

For the sake of simplicity, let us rewrite the above Helmholtz wave-equation in a simpler

form as

$$Lu = f, \tag{2.2}$$

where $L$ is the operator given by

$$L = \nabla^2 + \frac{\omega^2}{c(\mathbf{x})^2}. \tag{2.3}$$

A small perturbation in the background medium results in a small perturbation in the wavefield. Hence the operator $L$ can be decomposed into two parts, one results from the constant background (unperturbed) medium and the other comes from the result of small perturbation

$$L = L_o + \epsilon L_1, \tag{2.4}$$

where $\epsilon$ is a small perturbation parameter. $L_o$ is the unperturbed operator which is equivalent to

$$L_o = \nabla^2 + \frac{\omega^2}{c_o(\mathbf{x})^2}, \tag{2.5}$$

where $c_o(\mathbf{x})^2$ is the constant background velocity field. In order to get $L_1$, let us define $\alpha(\mathbf{x})$ to be the acoustic velocity potential given by

$$\alpha(\mathbf{x}) = \frac{1}{c(\mathbf{x})^2} - \frac{1}{c_o(\mathbf{x})^2}. \tag{2.6}$$

Equations [2.3 ] - [2.6 ] give the explicit value of $L_1$ as

$$L_1 = \frac{\omega^2 \alpha(\mathbf{x})}{\epsilon}. \tag{2.7}$$

Therefore, the change in the wavefield due to small perturbations in the background velocity can be expanded into a series of terms as

$$u = u_o + \epsilon u_1 + \epsilon^2 u_2 + \epsilon^3 u_3 + .... \tag{2.8}$$

which will be used to derive the single-scattering Born approximation. $u_o$ is the incident wavefield which is obtained from the constant background medium.

Plugging equation [2.4] and [2.8] into equation [2.2], and collecting terms that are equal powers in the perturbation $\epsilon$, leads to the following recursive equation:

$$\begin{aligned} L_o u_o &= f & \text{for } n = 0 \\ L_o u_n &= -L_1 u_{n-1} & \text{for } n \geq 1 \ . \end{aligned} \tag{2.9}$$

As we see from the above recursive equations, all of these equations are the form $L_o u = $ *a source function* and can be solved using the Green's function for the constant velocity medium

$$
\begin{aligned}
\left(\nabla^2 + \frac{\omega^2}{c_o^2\,(\mathbf{x})}\right) G(\mathbf{x}, \mathbf{x}_s, \omega) &= -\delta(\mathbf{x} - \mathbf{x}_s) \\
L_o G(\mathbf{x}, \mathbf{x}_s, \omega) &= -\delta(\mathbf{x} - \mathbf{x}_s),
\end{aligned}
\tag{2.10}
$$

where $G(\mathbf{x}, \mathbf{x}_s, \omega)$ is the point source Green's function for a source located at $\mathbf{x}_s$.

Making use of the above Green's function for a point source, the solution to equation [2.9] becomes the following recursive solution

$$
\begin{aligned}
u_o &= Gf & \text{for } n = 0 \\
u_n &= -GL_1 u_{n-1} & \text{for } n \geq 1 \ .
\end{aligned}
\tag{2.11}
$$

In other words, all higher order terms of the series depends on the incident wavefield. Finally, the solution to the perturbed wavefield can be solved recursively from equation [2.11] and then inserting the solution into equation [2.8]. Therefore, the total wavefield becomes

$$
u \ = \ \underbrace{u_o}_{\text{incident wave}} \ - \ \underbrace{\epsilon G L_1 u_o}_{\text{single scattered wave}} \ + \ \underbrace{\epsilon^2 G L_1 G L_1 u_o}_{\text{Double scattered wave}} \ + ....
\tag{2.12}
$$

The total wavefield consists of the incident waves, waves scattered once due to the heterogeneity of the medium ( $L_1$), waves that scattered twice due to the perturbation and all other higher order scattered waves. In other words, the total wavefield is the sum of two wavefields: the background and scattered wavefields. This series is called the Neumann series or at times the Born series (Snieder and Lomax, 1996). Since our intention is to linearize the scattering wavefield, we assume the scattering is weak. Hence we can apply the Born approximation, valid only when the scattered field is much smaller than the incident field. This implies that the heterogeneities are weak. As a result higher orders of scattering wavefields can be excluded.

The Born approximation consists of truncating the multiple scattering after a single scattered wave

$$
u = u_o - \epsilon G L_1 u_o.
\tag{2.13}
$$

The above equation is also known as Lippmann-Schwinger Equation (Zhdanov, 2002). One of the advantages of the Born approximation is that the scattered waves $-\epsilon G L_1 u_o$ depend linearly on the perturbation of the medium which we are interested in. In other words, the scattering wave linearly depends on the acoustic velocity potential because $L_1 = \epsilon \omega^2 \alpha(\mathbf{x})$. By linearzing the wavefield, we solve the linear inverse problem instead of the non-linear

inverse problem which is computationally expensive and, in most cases, hard to implement. Linearized inversion methods rely on the assumption that the forward problem is linear in the vicinity of the reference model. Therefore, the Born approximation is used to establish a linear relation between the model perturbation and the data residuals. However, one of the disadvantages of the Born approximation is that it ignores multiple scattering effects. When such effects are present, one should use extreme caution using the Born approximation.

The incident wavefield in a constant background velocity can be solved directly using the Green's function; see equation [2.11]. Let $G_o(\mathbf{x}_s, \mathbf{x}, \omega)$ be the Green's function for constant background velocity.

For the point source with a spectrum $f(\omega)$ located at $\mathbf{x}_s$, the incident wave is expressed by

$$u_o(\mathbf{x}_s, \mathbf{x}, \omega) = G_o(\mathbf{x}_s, \mathbf{x}, \omega)f(\omega). \tag{2.14}$$

In the Born approximation, the scattering wave $-\epsilon G L_1 u_o$ at the receiver position $\mathbf{x}_r$ is given by

$$u_{sc}(\mathbf{x}_r, \mathbf{x}_s, \omega) = \omega^2 f(\omega) \int G_o(\mathbf{x}_s, \mathbf{x}, \omega)\alpha(\mathbf{x})G_o(\mathbf{x}_r, \mathbf{x}, \omega)d\mathbf{x}^3 \tag{2.15}$$

where $L_1 = \epsilon \omega^2 \alpha(\mathbf{x})$. Here, it should pointed out that the above scattering wavefield can be applied for any arbitrary background velocity. When solved using a constant background velocity, it is called the Born approximation. Otherwise, it will be a distored Born approximation.

In three-dimensional space, the solution to the Green's function for the homogeous function is given by

$$G_o(\mathbf{x}_s, \mathbf{x}, \omega) = \frac{1}{|\mathbf{x} - \mathbf{x}_s|}\exp{(i\omega\frac{|\mathbf{x} - \mathbf{x}_s|}{c_o})}. \tag{2.16}$$

In two-dimensions, its solution is of the form

$$G_o(\mathbf{x}_s, \mathbf{x}, \omega) = \frac{i}{4}H_o^1\left(\omega\frac{|\mathbf{x} - \mathbf{x}_s|}{c_o}\right), \tag{2.17}$$

where $H_o^1$ is the zeroth order Hankel function of the first kind (Abramowitz and Stegun, 1972). Note that, in heterogeneous media, the Greens function cannot be solved analytically and the forward problem must be solved numerically.

Since the Green's function is a wave propagation, for any two arbitrary points, it established as

$$G_o(\mathbf{x}, \mathbf{x}', \omega) = A(\mathbf{x}, \mathbf{x}')e^{i\omega\tau(\mathbf{x}, \mathbf{x}')}, \tag{2.18}$$

where $\tau$ is the travel time of the wave taken to propagate from $\mathbf{x}$ to $\mathbf{x}'$ and $A(\mathbf{x}, \mathbf{x}')$ is the amplitude of the ray that travels from $\mathbf{x}$ to $\mathbf{x}'$. The amplitude satisfies the transport equation

$$2\nabla\tau(\mathbf{x}, \mathbf{x}') \cdot \nabla A(\mathbf{x}, \mathbf{x}') + A\nabla^2\tau(\mathbf{x}, \mathbf{x}') = 0. \tag{2.19}$$

The travel time should also satisfy the eikonal equation (Zhdanov, 2002)

$$|\nabla\tau(\mathbf{x}, \mathbf{x}')|^2 - \frac{1}{c^2(\mathbf{x})} = 0. \tag{2.20}$$

Making use of the above facts, the scattering wave in [2.15] for each frequency is written as

$$u_{sc}(\mathbf{x}_r, \mathbf{x}_s, \omega) = \omega^2 f(\omega) \int A(\mathbf{x}_s, \mathbf{x})e^{i\omega\tau(\mathbf{x}, \mathbf{x}_s)}\alpha(\mathbf{x})A(\mathbf{x}_r, \mathbf{x})e^{i\omega\tau(\mathbf{x}, \mathbf{x}_r)}d\mathbf{x}^3. \tag{2.21}$$

Let us define $A(\mathbf{x}_s, \mathbf{x}, \mathbf{x}_r)$ and $\tau(\mathbf{x}_s, \mathbf{x}, \mathbf{x}_r)$ to be the new total amplitude and travel time functions and given by:

$$A(\mathbf{x}_s, \mathbf{x}, \mathbf{x}_r) = A(\mathbf{x}, \mathbf{x}_s)A(\mathbf{x}_r, \mathbf{x}) \tag{2.22}$$

$$\tau(\mathbf{x}_s, \mathbf{x}, \mathbf{x}_r) = \tau(\mathbf{x}_s, \mathbf{x}) + \tau(\mathbf{x}, \mathbf{x}_r). \tag{2.23}$$

Now equation [2.21] is reduced to a simple form

$$u_{sc}(\mathbf{x}_r, \mathbf{x}_s, \omega) = \omega^2 f(\omega) \int A(\mathbf{x}_s, \mathbf{x}, \mathbf{x}_s)e^{i\omega\tau(\mathbf{x}_s, \mathbf{x}, \mathbf{x}_r)}\alpha(\mathbf{x})d\mathbf{x}^3. \tag{2.24}$$

In the time domain, the inverse Fourier transform, equation [2.24], is expressed as

$$u_{sc}(\mathbf{x}_r, \mathbf{x}_s, t) = \left[ -\frac{\partial^2}{\partial t^2} \int A(\mathbf{x}_s, \mathbf{x}, \mathbf{x}_r)\delta(t - \tau(\mathbf{x}_s, \mathbf{x}, \mathbf{x}_r))\alpha(\mathbf{x})d\mathbf{x}^3 \right] * f(t). \tag{2.25}$$

where the symbol $*$ represents convolution. Due to the convolution, the second derivative in time can be shifted onto the wavelet (Tarantola, 1984)

$$u_{sc}(\mathbf{x}_r, \mathbf{x}_s, t) = \left[ -\int A(\mathbf{x}_s, \mathbf{x}, \mathbf{x}_r)\delta(t - \tau(\mathbf{x}_s, \mathbf{x}, \mathbf{x}_r))\alpha(\mathbf{x})d\mathbf{x}^3 \right] * \frac{\partial^2 f(t)}{\partial t^2}. \tag{2.26}$$

Note that the delta function in the above integeral is a zero value except at points that satisfy

$$\delta(t - \tau(\mathbf{x}_s, \mathbf{x}, \mathbf{x}_r)) = \begin{cases} 1 & \text{for } t = (\mathbf{x}_s, \mathbf{x}, \mathbf{x}_r) = \tau(\mathbf{x}_s, \mathbf{x}) + \tau(\mathbf{x}, \mathbf{x}_r) \\ 0 & \text{for } t \neq (\mathbf{x}_s, \mathbf{x}, \mathbf{x}_r). \end{cases} \tag{2.27}$$

## 2.2.1   Wave-equation in two-dimensions

As stated earlier, the Green's function for any arbitrary background velocity can be expressed using the first order Hankel function given for given geometrical configurations (source-receiver geometry) (Miller et al., 1987; Beylkin, 1985). However, here our objective is to establish its solution in two-dimensional space using a constant acoustic velocity. Using the far-field expression for the zeroth order Hankel function of the first kind function (Miller et al., 1987), one finds that the Green's function for two arbitrary points in the far field $(\omega\tau(\mathbf{x},\mathbf{x}') \gg 1)$ is given by

$$G_o(\mathbf{x},\mathbf{x}',\omega) = -\left(\frac{c_o}{8\pi\omega|\mathbf{x}-\mathbf{x}'|}\right)^{1/2} e^{i(\omega\tau(\mathbf{x},\mathbf{x}')+\frac{\pi}{4})}, \tag{2.28}$$

which is equivalent to

$$G_o(\mathbf{x},\mathbf{x}',\omega) = (-i\omega)^{-1/2} A(\mathbf{x},\mathbf{x}') e^{i(\omega\tau(\mathbf{x},\mathbf{x}')+\frac{\pi}{4})}, \tag{2.29}$$

where

$$A(\mathbf{x},\mathbf{x}') = \left(\frac{c_o}{8\pi|\mathbf{x}-\mathbf{x}'|}\right)^{1/2}. \tag{2.30}$$

Making use of the solution to the Green's function and the amplitude factor in two-dimensions, the scattering potential reduces to the form

$$u_{sc}(\mathbf{x}_r,\mathbf{x}_s,\omega) = i\omega f(\omega) \int A(\mathbf{x}_s,\mathbf{x},\mathbf{x}_r) e^{i\omega\tau(\mathbf{x}_s,\mathbf{x},\mathbf{x}_r)} \alpha(\mathbf{x}) d\mathbf{x}^2, \tag{2.31}$$

where $A(\mathbf{x}_s,\mathbf{x},\mathbf{x}_r)$ in this case represents the new total 2D amplitude term

$$A(\mathbf{x}_r,\mathbf{x},\mathbf{x}_s) = \frac{c_o}{8\pi}\left(\frac{1}{|\mathbf{x}-\mathbf{x}_s||\mathbf{x}-\mathbf{x}_r|}\right)^{1/2}, \tag{2.32}$$

where $\tau(\mathbf{x}_s,\mathbf{x},\mathbf{x}_r)$ is travel time functions.

In the time domain, equation[2.31] is expressed as

$$u_{sc}(\mathbf{x}_r,\mathbf{x}_s,t) = \left[\frac{\partial}{\partial t}\int A(\mathbf{x}_s,\mathbf{x},\mathbf{x}_r)\delta(t-\tau(\mathbf{x}_s,\mathbf{x},\mathbf{x}_r))\alpha(\mathbf{x})d\mathbf{x}^2\right] * f(t), \tag{2.33}$$

which is equivalent to

$$u_{sc}(\mathbf{x}_r,\mathbf{x}_s,t) = \left[\int A(\mathbf{x}_s,\mathbf{x},\mathbf{x}_r)\delta(t-\tau(\mathbf{x}_s,\mathbf{x},\mathbf{x}_r))\alpha(\mathbf{x})d\mathbf{x}^2\right] * \frac{\partial f(t)}{\partial t}. \tag{2.34}$$

With this, we solve the propagation of waves numerically. For a detail of the calculations see Miller et al. (1987), Beylkin (1985), and Beylkin and Burridge (1990).

Equation[2.34] is referred as the generalized Radon transform (GRT) because the integral data can be viewed as a migration weighted projection of the perturbation and these weighted projections are curved integration surfaces (Beylkin, 1985). As a result, the scattering wavefield can be used as projection of data to correlate each point in the scattered waveform to the model parameter (acoustic potential).

## 2.3 Forward and adjoint modeling

The forward problem is obtained by discretizing equation [2.34], thus solving the single-scattering approximation from the known constant reference medium. This method attempts to transform the model to data form. The Green's function computed using an asymptotic approximation in constant background velocity. In a simple notion, the linearized discrete form of the forward problem is offen expressed as

$$\mathbf{d} = \mathbf{Gm}, \tag{2.35}$$

where $\mathbf{d}$ is the vector that represents the observed scattered wavefield $u_{sc}(\mathbf{x}_r, \mathbf{x}_s, t)$ after discretization of equation [2.34], $\mathbf{G}$ is the forward operator or demigration operator and $\mathbf{m}$ is the discretize acoustic potential.

The seismic inverse scattering problem is formulated via the Generalized Radon Transform (GRT). The GRT is often used for estimating the distorted solution of the inverse scattering problem because it is difficult to find the true inverse solution of the forward problem to retrieve the acoustic potential. However, to retrieve the distored invese solution, it does have an adjoint expression. The inversion requires the intoduction of the inverse integral operator to equation [2.31]. This adjoint partially recovers the distorted version of the acoustic potential. Applying the weighting function to the adjoint of the GRT creates the solution to the asymptotic inverse of the scattering problem. An analytical expression of the weighting function for the common source-receiver geometries is then used to recover the original model (Miller et al., 1987). The action of forward modeling and the adjoint migration operator can be interpreted in terms of integral geometry as projection and backprojection operator respectively (Miller et al., 1987). From Miller et.al. results it becomes clear that $\alpha(x)$ can be regarded as an approximation to the inverse problem.

An adjoint modeling, also called migration, is computed from a linearized forward propagation model using GRT. The transpose of the forward operator $\mathbf{G}$ propagates data observa-

tion points back to the medium perturbations. The adjoint problem solves the approximate acoustic potential $\mathbf{m}$ as

$$\mathbf{m}^\dagger = \mathbf{G}^T \mathbf{d}, \tag{2.36}$$

where $\mathbf{G}^T$ is the adjoint of the forward operator or demigration operator. $\mathbf{m}^\dagger$ is not the true acoustic potential, but an approximation to the inverse problem. It is based on the assumption that $[\mathbf{G}^T\mathbf{G}]^{-1} \approx \mathbf{I}$. For a more accurate result for the original model parameters, one has to solve the inverse of the demigration operation should the operator be a square matrix and its inverse exists. Then the inverse problem can be solved using a least-squares method (note $\mathbf{G}^T \neq \mathbf{G}^{-1}$). However, the adjoint solution gives the structure of the subsurface model, where the scatterers are located.

The method is computationally cheaper than the direct inversion method for estimating the acoustic potential, but the image is often smeared and the amplitude information is distorted. A schematic representation in Figure 2.1 shows how the results obtained using the inverse and adjoint operator method appear. The output obtained using the adjoint operator method is usually only suitable for structural studies. However, based on the frame work of the adjoint modeling, we can develop an inverse scheme to fit the seismic data and provide a more accurate solution, which will be discussed in the next chapter.

## 2.4   Synthetic data example

A 2D velocity model shown in Figure 2.2 is tested to show the effectiveness of the adjoint method in estimating the structure of the subsurface. A constant background velocity model set to 2200 m/s is used to approximate the Green's function. A small velocity contrast is used as the perturbation so that its wave propagation can be linearized using the Born scattering approximation. The perturbation region can be easily seen from the figure, a different areas with velocity perturbation. A uniform grid spacing of $\triangle x = \triangle z = 8$ m with a time sampling rate of $\triangle t = 1$ ms are used. A Ricker wavelet of 25 Hz is chosen. For forward modelling of wave propagation, a data set consisting of 5 shots (200 m spacing ) and 25 receivers (32 m spacing) was used at the surface.

Then, the migration image is obtained by back-propagating the observation data to the subsurface. Imaging methods require an estimation of the reference potential. The migrated image shows the exact location of where the scatterer are located. Though it predicts the location of the reflector, the adjoint method (unlike the inverse problem gives) a blurry and fuzzy image. This method, however, hardly predicts the correct amplitude of the deep reflector, which is a common problem in most migration algorithms. But, the method gives basic information where reflectors are located.
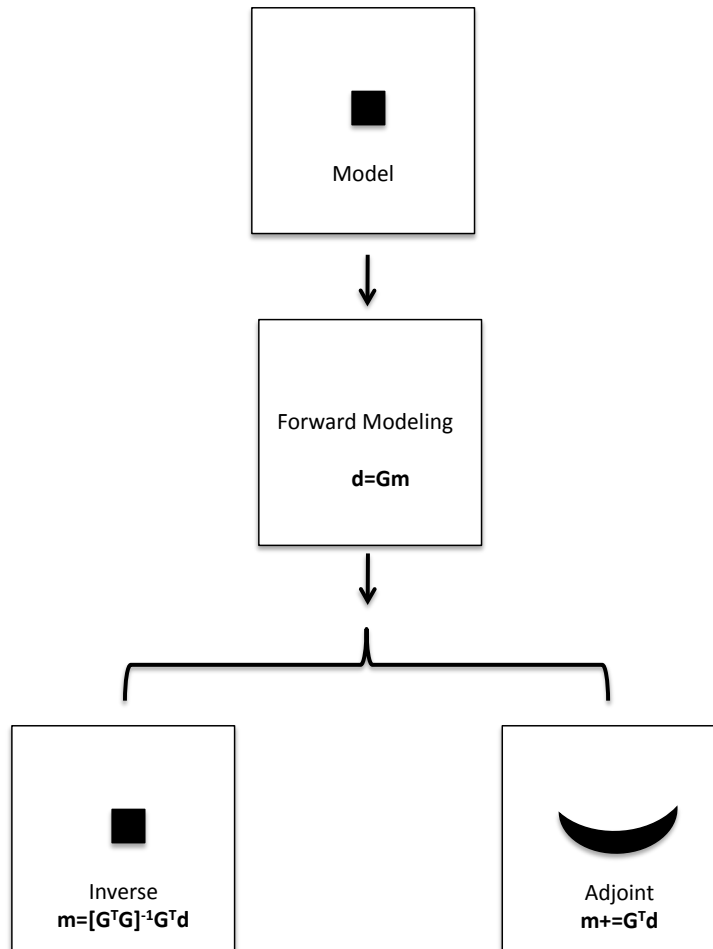
Figure 2.1: A schematic representation of modeling geophysical problems. The adjoint operator, unlike the inverse problem, gives a blurry and fuzzy image.
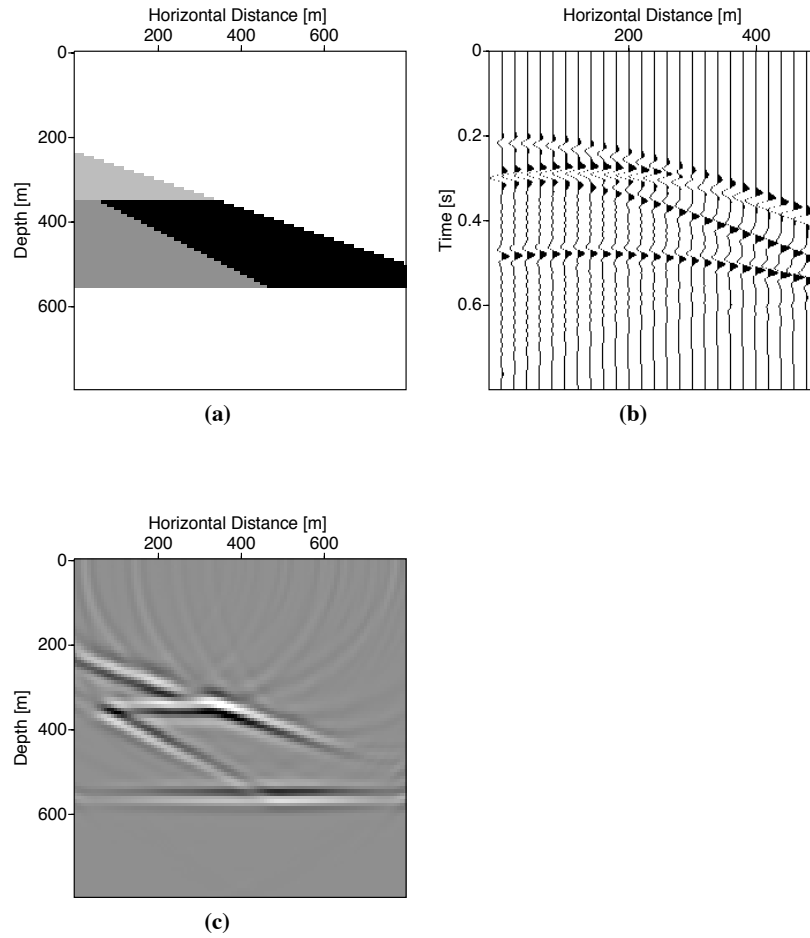
Figure 2.2:    (a) Velocity model. (b) Synthetic data generated by forward modeling with a source located at 200 m at the surface. (c) The acoustic velocity potential recovered by the adjoint operator.

## 2.5   Summary

In this chapter, we reviewed the linearization wavefield based on the Born approximation using scattering theory. The Born approximation and an asymptotic form for the incident wavefield is used to establish the linearization of the forward problem with respect to the model parameter, acoustic perturbation. The incident wavefield is determined by a constant reference model, which is assumed to be known a priori. The method, however, can be extended to variable reference model. The Born approximation is only valid for small perturbations of the velocity from assumed background velocities. This approximation is a single scattering approximation. Multiple scattering due to the heterogeneity of the medium is ignored.

The forward and the adjoint problems can be easily solved by discretization of the linear scattering waves. For an approximate solution of inverse problems, the method of GRT is used. Linearization of the forward problem is essential in geophysical problems. In the next chapter, we employ this method to compute the least-square inversion for estimating the acoustic potential profiles of the medium.

# CHAPTER 3

# Edge-preserving Regularization

## 3.1 Total variation regularization

A great number of geophysical-estimation problems are mathematically ill-posed because they operate with insufficient data (Jackson, 1972). The most common problem in waveform inversion of seismic reflection data based on the least-squares inversion is that the solution is not unique and computationally unstable. It is very unlikely to get an exact solution from any numerical computation. However, the solution we get from the computation is approximate, and it may be closer to the actual model parameter if the numerical method is accurate. One of the reasons for inaccuracy of the solution is that the actual measurements of geophysical quantities we acquire from the observation have noise and experimental errors. It is very challenging to fit and predict the result with the theoretical modeling. The best solution to this kind of problem is to be able to search and find the model and algorithm that produces and fits best to the measurement data.

Since inverse wave propagations are often ill-posed, the method of regularization in the least-squares optimization is indispensable. The goal of regularization is to impose additional constraints on the estimated model parameters to recover unique and stable solutions. The idea is to make the estimation problems well-posed (where a unique solution exists and depends continuosly on the data) by adding indirect constraints on the estimated model (Aster et al., 2005). The regularized inverse problems is found in many applications in geophysics problems: travel time tomography (Bube and Langan, 2008; Clapp et al., 2004), migration velocity analysis, high-resolution Radon transform (Trad et al., 2002, 2003), spectral decomposition, edge-preserving imaging (Youzwishen and Sacchi, 2006), and velocity estimation using edge-preserving regularization (Valenciano et al., 2004).

In this chapter, total variation (TV) regularization, a non-quadratic regularization method for imposing constraints to the least-square problem, is employed. The objective of this regularization method is straightforward. If the physical quantity that is going to be estimated is piecewise constant, has sharp edge boundaries and blocky, then this regularization term looks to reconstruct piecewise continuous profiles by enforcing a sparseness constraint on the model gradient without much smoothing (Youzwishen and Sacchi, 2006). Consequently it preserves its edge and discontinuities while suppressing artifacts due to noises.

The focus of the work is on the application of edge-preserving regularization (EPR) based on the total variation. The application of the TV is to the problem of estimating acoustic velocity perturbations from a multi-source and receiver geophysical experiment as described in Chapter one. This work is an extension of Youzwishen and Sacchi (2006). In the case of Youzwishen and Sacchi (2006), the work is based on the modified Cauchy prior distribution; here I am using total variation. The problem is linearized using the single scattering Born approximation about a known background reference medium (Beylkin, 1985; Beylkin and Burridge, 1990; Miller et al., 1987). The assumption is based on the fact that the velocity model profile is small with respect to the background reference model and that it can be modeled by blocky velocity perturbations.

## 3.2  The damped least-squares solution

Depending on the number of known data and unknown model parameters, say $n$ and $m$ respectively, a discrete inverse problem is classified as over-determined, under-determined and even-determined problems. If it is over-determined problem, then there exist more known observation data than unknown model parameters ($n > m$). For example, fitting more than two data points, which are not collinear or don't lie along a line, to a straight line is an overdetermined problem. The least-squares approach solves best fit line by minimizing the error function of an over-determined problem. An under-determined problem has fewer observation data points compared to the number of unknown model parameters ($n < m$). For example, fitting a straight line through one datum point is an under-determined problem. In this case, the solution to the problem is infinite and can not be solved with the least-squares approach alone. In order to solve the problem, one has to impose a constraint; the problem requires regularization. Depending on the nature of the problem, different kinds of regularization can be incorporated. In the case of even-determined, the number of known observation data points are exactly equal to the number of unknown model parameters ($n = m$). For example, fitting a straight line through two data points is an even-determined and yields a unique solution. Here, it should be pointed out that this kind of situation rarely occur in geophysical inverse problems. Even though in such cases we have equal number

of known data points and unknown parameters, it is still impossible to gather a perfect noise-free data to get all the valuable information of the Earth's subsurface properties. In general, most geophysical problem lie in the category of ill-posed (i.e. under-determined) problem due to aforementioned reasons.

As described from the first chapter, we focus on the solution of the discrete linear inverse problem that arises in the multi-source and receiver seismic acoustic probe in the framework of the Born approximation. If the data are represented by vector $\mathbf{d}$ and the model parameters by vector $\mathbf{m}$, (which in our case is the acoustic potential), their functional relationship is defined by the forward modeling operator or demigration operator $\mathbf{G}$ as

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

where $\mathbf{d}$ is the measured scattered wavefield at the surface of the earth and $\mathbf{n}$ is the additive uncorrelated noise. The reason for this additive noise is because experimental measurements are usually contaminated by some random noise and are thus inaccurate.

The cost function for the ill-posed inverse problem is written in terms of the acoustic perturbation $\mathbf{m}$, and the least-squares optimization amounts to minimizing the least-squares norm of the residual difference

$$J(\mathbf{m}) = ||\mathbf{Gm} - \mathbf{d}||_2^2, \tag{3.2}$$

where $|| \cdot ||_2$ represents an $L_2$ norm, and all quantities written in bold represent vectors.

One way to find the solution to the linear inverse problem without any regularization is to solve the least-squares minimization

$$\mathbf{m}_{ls} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d}. \tag{3.3}$$

The solution is obtained from the above equation is a naive solution, because it is done straightforward and can lead to a solution that is not useful. Second, in most geophysical problems, the solution obtained from least-square is not unique and is ill-posed as described previously. For this reason, in order to get a stable and unique solution, one can solve equation [3.2] with a quadratic, Tiknovon, regularization

$$J(\mathbf{m}) = ||\mathbf{Gm} - \mathbf{d}||_2^2 + \mu||\mathbf{Dm}||_2^2, \tag{3.4}$$

The first term is the least-square norm and represents the data misfit portion of the cost function. The second term represents the regularization term. $\mu$ is a positive value regularization parameter that determines the goodness of fit to the measured data and the

amount of regularization that has to be done to the measured data. The importance of the regularization parameter ($\mu$) can be seen from the cost function into two ways. For example, if the regularization parameter, ($\mu$), is large, then it means that more weight is given to minimizing the model norm over the misfit function. On the other hand if ($\mu$) is small, then the misfit is the main (importance) term in the minimization. In this case the model norm becomes less important. Therefore, in order to come up with the best value of the regularization parameter, one has to try simultaneously reducing the norm (stability) and misfit to honor the observation. $\mathbf{D}$ is the regularization operator matrix, and in most cases it is either first or second order derivatives. It is also referred to as the weighting matrix for the model parameter.

In Tikhonov's regularization approach, we look for the model parameter $\mathbf{m}$ that minimizes the least-squares norm of the compound vector $||\mathbf{Gm} - \mathbf{d}||_2^2 + \mu||\mathbf{Dm}||_2^2$. The least-square minimum solution is given by

$$\mathbf{m}_{mn} = (\mathbf{G}^T\mathbf{G} + \mu\mathbf{D}^T\mathbf{D})^{-1}\mathbf{G}^T\mathbf{d}, \tag{3.5}$$

where $\mathbf{G}^T$ and $\mathbf{D}^T$ is the transpose of $\mathbf{G}$ and $\mathbf{D}$, respectively. If $\mathbf{D} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix (i.e. no first or second order derivative operator are applied), then the solution is called the damped least-square solution (Clearbout, 1992). The term $\mathbf{G}^T\mathbf{G}$ in most cases is invertible for ill-posed problems (Strang, 1986 ). The introduction of the regularization prevents the singularity associated with inverting zero terms.

In practice, most of the regularization of inverse problems suffers from a trade-off between the size of the regularized solution and the quantity of the fit that it provides to the given data. Different regularization techniques differ on the basis on how they minimize this trade-off. This can be controlled by the selection of a proper regularization parameter ( $\mu$).

The above regularization not only tries to fit the regularized solution to the exact solution, but it also penalizes large norms. The solution gives a smooth solution, which is desirable for some applications. But, if the desired solution to be recovered is discontinuous and piecewise, then one approach is to replace the regularization term, which will be done in the next section.

## 3.3   Edge-preserving using total variation regularization

In this section we focus on total variation regularization method which is used to preserve edges and does take into consideration the information that the data set is blocky and discontinuous. The regularization term in equation [3.2] is replaced by a non-linear regularization term as

$$J(\mathbf{m}) = ||\mathbf{Gm} - \mathbf{d}||_2^2 + \mu||\overrightarrow{\bigtriangledown}\mathbf{m}||_1, \tag{3.6}$$

where $||\overrightarrow{\bigtriangledown}\mathbf{m}||_1$ is the total variation of $\mathbf{m}$. The operator $\overrightarrow{\bigtriangledown}$ is

$$\overrightarrow{\bigtriangledown} = \hat{x}\frac{\partial}{\partial x} + \hat{z}\frac{\partial}{\partial z}. \tag{3.7}$$

In a more clear way the above equation [3.6] can be written as

$$J(\mathbf{m}) = ||\mathbf{Gm} - \mathbf{d}||_2^2 + \mu||\sqrt{(D_x\mathbf{m})^2 + (D_z\mathbf{m})^2}||_1 \tag{3.8}$$

where $D_x$ and $D_y$ are the horizontal and the vertical discrete first order derivative operators with respect to $x$ and $z$ respectively. $\mu$ is the regularization parameter.

### 3.3.1   Minimization Problem

In this section, the minimization of the cost function with a non-linear total variation regularization term will be shown. For the sake of simplicity, let us define

$$\begin{aligned}
\mathbf{m}_x &= D_x\mathbf{m} \\
\mathbf{m}_y &= D_z\mathbf{m},
\end{aligned} \tag{3.9}$$

where $\mathbf{m}_x$ and $\mathbf{m}_z$ are the derivatives of the model parameter with respect to $x$ and $z$ respectively.

In order to minimize equation [3.6], we take the derivative of the cost function with respect to the model parameter $\mathbf{m}$. When minimizing it, the total variation regularization term is not differentiable at $\mathbf{0}$, as the derivative of $||\overrightarrow{\bigtriangledown}\mathbf{m}||_1$ at $\mathbf{0}$ leads to a singularity. To avoid this in numerical implementations a small constant value, $\alpha$, is added to give

$$||\overrightarrow{\bigtriangledown}\mathbf{m}||_1 = ||\sqrt{\mathbf{m}_x^2 + \mathbf{m}_z^2 + \alpha^2}||_1 = \sum_{i,j}|\sqrt{m_{x,i,j}^2 + m_{z,i,j}^2 + \alpha^2}|, \tag{3.10}$$

where $0 < \alpha < 1$. This is a smoothing parameter and has a significant effect on the solutions. The meaning and application of $\alpha$ will be discussed in detail in the next section.

The TV regularization method leads to the minimization of the functional

$$J(\mathbf{m}) = ||\mathbf{Gm} - \mathbf{d}||_2^2 + \mu||\sqrt{\mathbf{m}_x^2 + \mathbf{m}_z^2 + \alpha^2}||_1. \tag{3.11}$$

Minimizing equation [3.11] is a non-linear least-squares problem because the mifit term is quadratic in **m** whereas the second term is non-quadratic in **m** .

The gradient of equation [3.11] with respect to **m** gives

$$\nabla_{\mathbf{m}}J = \mathbf{G^T}(\mathbf{Gm} - \mathbf{d}) + \mu\overrightarrow{\nabla}.\frac{\overrightarrow{\nabla}\mathbf{m}}{||\overrightarrow{\nabla}\mathbf{m}||_1}. \tag{3.12}$$

Note that $||\overrightarrow{\nabla}\mathbf{m}||_1$ is known as the gradient magnitude; this term provides us with the information about the discontinuities in the image. The second term in the right hand side of the above equation is computed using finite difference approximation. Therefore, the second term in the right hand side of equation [3.12] is expanded as

$$\overrightarrow{\nabla} \cdot \frac{\overrightarrow{\nabla}\mathbf{m}}{||\overrightarrow{\nabla}\mathbf{m}||_1} = \frac{m_{xx}(m_z^2 + \alpha^2) - 2m_x m_z m_{xz} + m_{zz}(m_x^2 + \alpha^2)}{(m_x^2 + m_z^2 + \alpha^2)^{\frac{3}{2}}} \tag{3.13}$$

The solution of the derivatives of each term of $m$ in the above expression obtained by finite difference approximation are

$$m_{xx} = \frac{m_{i+1,j} - 2m_{i,j} + m_{i-1,j}}{\triangle x^2} \tag{3.14}$$

$$m_{zz} = \frac{m_{i,j-1} - 2m_{i,j} + m_{i,j-1}}{\triangle z^2} \tag{3.15}$$

$$m_{xz} = \frac{m_{i+1,j+1} + m_{i-1,j-1} - m_{i+1,j-1} - m_{i-1,j+1}}{4\triangle x\triangle z} \tag{3.16}$$

$$m_x = \frac{m_{i+1,j} - m_{i,j}}{\triangle x} \tag{3.17}$$

$$m_z = \frac{m_{i,j-1} - m_{i,j}}{\triangle z} \tag{3.18}$$

where $i$ and $j$ are indexes. For a uniform grid spacing, we take $\triangle x = \triangle z$ in numerical implementations.

## 3.3.2    The numerical scheme

The solution to equation [3.12] can be computed using different techniques. Among them are the steepest descent method using line search, Newton's method (Chan et al., 1995; Osher and Fatemi, 1992), the Prime-Dual Newton method (Vogel, 2002; Chan et al., 1998), and the lagged diffusivity fixed point iterative method (Vogel and Oman, 1998; Chan and Mulet, 1999). In this thesis the lagged diffusivity fixed point iterative method (Vogel and

Oman, 1996; Shi et al., 2007) is used in conjunction with the conjugate gradient method. One reason the lagged diffusivity method was chosen is because it is straightforward to implement ; it does however, requires a careful choice of $\alpha$ to avoid numerical instabilities for type of solutions we seek. A fixed-point lagged diffusivity iteration method has been applied in the regressive area to employ linearization of the non-linear differential equation (Dibos and Koepfler, 1999; Vogel and Oman, 1998; Vogel, 1997) and in total variation discretization (Moisan, 2007).

Following Dibos and Koepfler (1999) and Vogel and Oman (1998), here we utilized the lagged diffusivity fixed point method for solving equation [3.12]. This is a scheme where the regularization term is lagged by one step and iterated until it converges. The idea of the fixed-point lagged diffusivity iteration method is that it employs the linearization of a non-linear differential term by lagging the diffusion coefficient behind one iteration. For details see (Vogel, 2002; Chan and Mulet, 1999). This method has been used for obtaining total variation regularization and applied to an image in such way that the non-linear differential term in the right hand side of equation [3.12] is linearized in the form as (for details see Appendix)

$$\nabla_m J = \mathbf{G}^T(\mathbf{Gm} - \mathbf{d}) + \mu R(\mathbf{m})\mathbf{m}. \tag{3.19}$$

So, we seek the solution that minimize the above equation

$$\nabla_m J = 0, \tag{3.20}$$

which in turn gives the solution

$$[\mathbf{G}^T\mathbf{G} + \mu R(\mathbf{m})]\mathbf{m} = \mathbf{G}^T\mathbf{d}. \tag{3.21}$$

The above equation is a non-linear problem and can be solved iteratively method using the method of conjugate gradients. This method is also known as the iteratively reweighted least-squares (IRLS) method. The weighted term is $R(\mathbf{m})$. First we set $\mathbf{m}_k$ be the solution at the $k^{th}$ iteration, then the solution at the $k + 1$ iteration is solved from the following equation

$$[\mathbf{G}^T\mathbf{G} + \mu R(\mathbf{m}_k)]\mathbf{m}_{k+1} = \mathbf{G}^T\mathbf{d}. \tag{3.22}$$

A straightforward approach to solve the above equation is to tune the regularization parameter $\mu$ until a desired image is recovered, which is then solved iteratively. Tuning the

parameter every time, however, is time consuming and requires a lot of effort to get the solution that we seek.

Below is an algorithm to compute the iteratively reweighted least-square using the conjugate gradient (CG) algorithm. The method proceeds by generating vector sequences of iterates (i.e., successive approximations to the solution), and update the iterate solutions and residuals. The iteratively reweighted regularization term is embedded in an iterative-optimization scheme of the conjugate-gradient algorithm. At each step, the lagged diffusivity term (iterative reweighted term) is computed and improved, and used for the next CG minimization until a desirable solution is obtained. The implementation of the fixed-point iterative method of total variation regularization is done in Matlab.

*main loop*

*while* $||\mathbf{J}(\mathbf{m}_{k+1}) - \mathbf{J}(\mathbf{m}_k)|| \leq$ *tolerance ( user defined )*

*Obtain the Total varation regularization term*

$R(\mathbf{m}_k)$

*solve the following withing CG method*

$\left[\mathbf{G}^T\mathbf{G} + \mu R(\mathbf{m}_k)\right]\mathbf{m}_{k+1} = \mathbf{G}^T\mathbf{d}$

*assign* $\mathbf{G} \longleftarrow \left[\mathbf{G}^T\mathbf{G} + \mu R(\mathbf{m}_k)\right]$

*Start CG optimization*

$$\mathbf{r}_{(0)} : \quad = \quad \mathbf{b} - \mathbf{G}\mathbf{m}_{(0)},$$
$$\mathbf{P}_{(0)} : \quad = \quad \mathbf{r}_{(0)}$$
$$i := 0$$

$$\mathbf{p}_{(0)} : \quad = \quad \mathbf{r}_{(0)} = \mathbf{b} - \mathbf{G}\mathbf{m}_{(0)},$$
$$\alpha_{(i)} : \quad = \quad \frac{\mathbf{r}_{(i)}^T\mathbf{r}_{(i)}}{\mathbf{p}_{(i)}^T\mathbf{G}\mathbf{p}_{(i)}},$$
$$\mathbf{m}_{(i+1)} : \quad = \quad \mathbf{m}_{(i)} + \alpha_{(i)}\mathbf{p}_{(i)},$$
$$\mathbf{r}_{(i+1)} : \quad = \quad \mathbf{r}_{(i)} - \alpha_{(i)}\mathbf{G}\mathbf{p}_{(i)},$$

*if* $\mathbf{r}_{k+1}$ *is sufficiently small* $\rightarrow$ *exit*

$$\beta_{(i+1)} : \quad = \quad \frac{\mathbf{r}_{(i+1)}^T\mathbf{r}_{(i+1)}}{\mathbf{r}_{(i)}^T\mathbf{r}_{(i)}},$$
$$\mathbf{p}_{(i+1)} : \quad = \quad \mathbf{r}_{(i+1)} + \beta_{(i+1)}\mathbf{p}_{(i)}.$$
$$i : \quad = \quad i+1$$

*end of CG*

*update the model paramter*

*end of main loop*

## 3.4 Results and discussions

As a reference, first a simple calculation is tested to varify the validity and accuracy of the algorithm. The result of the calculations will provide some perspective on the quality of the result that can be obtained by modeling. A simple velocity layer medium shown in Figure 3.2 [a] is considered. The velocity model consists of different layers with different velocities. The vertical velocity profile is piecewise continuous, thus it contains sharp edges. A uniform grid spacing $\triangle x = \triangle z = 10$ m with a time sampling rate $\triangle t = 1$ ms are used. A Ricker wavelet of 30 Hz is chosen. For forward modeling of the wave propagation, a data set consisting of 3 shots and 30 receivers at the surface, $(x, 0)$, were used. The source-receiver
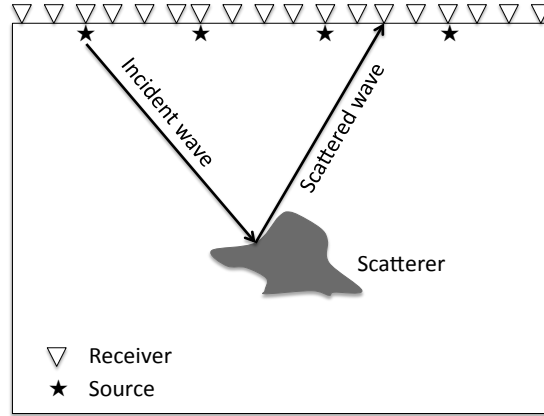
Figure 3.1:  Source-receiver geometry.

geometry is depicted in Figure 3.1. The kernel **G**, in this scheme, is computed based on the Born approximation of single scattering theory, as described in Chapter 2.

First, a synthetic seismic data are generated with the true model from a known reference medium set to be 1600 m/s. This synthetic data are assumed to be the observed data. A small noise, signal to noise ratio (SNR)=60, is added to the synthetic data. Figure 3.2 [b] is the generated synthetic seismic data for a single shot located at $(20m, 0m)$. The velocity inversion is then carried out using least-square data fitting with and without the total variation regularization, Figure 3.2 [c-d]. Figure 3.2 [c] is the recovered solution using the Tikhonov regularization. This method results in a stable and smooth solution. However, its solution has too many pumps or wiggles and fuzzy boundaries. Using the EPR method, such wiggles are mostly eliminated, Figure 3.2 [d]. The utilization of the total variation regularization is to make sure that the inversion preserves its edges without too much smoothing. In order to come up with such a solution, the discontinuities and sharp edges are controlled by two parameters, one by $\mu$ and the other by $\alpha$.

The parameter $\alpha$, which is introduced in the total variation and ensures that the TV functional is continuously differentiable, plays another role as well. It can have a great influence on the behavior of the regularized function, which can be very nonlinear. Appropriate choices of these parameters are very important criteria that should be carefully chosen for a good regularized output. During the optimization of the gradient, the initial model is set to be homogenous (set to zero) in the computations; no edges are detected in the first iteration. In the second iteration, and so forth, the edges and the discontinuities are controlled by trial and error (the choice of $\alpha$). The procedure starts from two extreme points and is refined until a solution with desirable features is obtained. If $\alpha$ is very large, it will
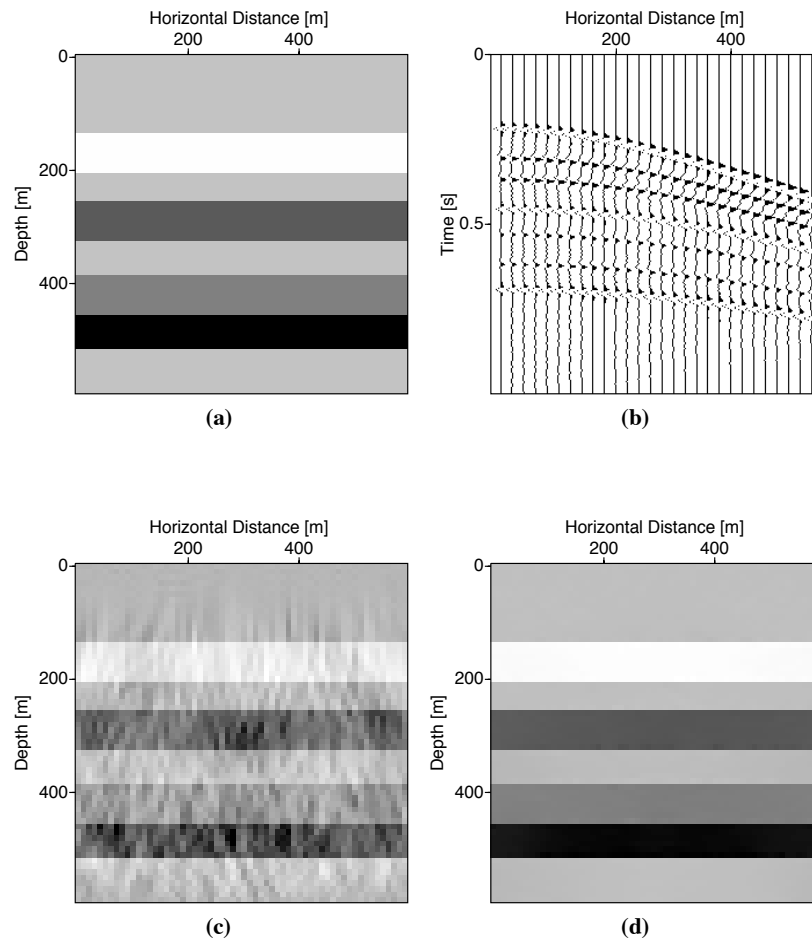
Figure 3.2:   The true layer velocity model (A), (B) synthetics data obtained from the single-scattering Born approximation computation, shot location at $20m$ from origin (0 m, 0 m). (C) reconstructed solution using least-squares damped method and solution using edge-preserving regularization method (D). Unlike the Tikhonov regularization, the EPR solution is clean and has no wiggles.

dominate the TV term and, as a result, the TV regularization behaves as a quadratic form. The regularized term loses its non-linearity property, TV $\approx ||\alpha||_1$. As a result, none of the discontinuities quantifies as edges and thus the gradient becomes more well-behaved. A full smoothing is applied, which is not ideal if sharp edges are to be recovered. The TV term becomes a constant independent of the model parameter $\mathbf{m}$, which means that the solution to the minimization problem will be the unregularized or naive least-squares solution. Therefore, the non-linearity behavior of the gradient is controlled by the TV term with an appropriate choice of $\alpha$. As a matter of fact the edge-preserving solutions goes to small $\alpha$, the weighting function magnifies all edges and discontinuities through the derivatives (see Equation [3.13]), preserves sparseness of $|\bigtriangledown \mathbf{m}|$. Contrary, large $\alpha$ smooth out $R(\mathbf{m})$ and the output solution as well. In other words, $\alpha$ controls the smoothness of the solution. The weighting function $R(\mathbf{m})$ locates the presences of edges and discontinuities.

For example, Figure 3.3[b] and[c] show the results obtained using improper choice of $\alpha$ and $\mu$. These results are computed using wrong values the regularization parameters; small deviations from the refined regularization parameters were taken. The optimization of the algorithms for each iteratively reweighed norm was run for 200 iterations, with $\alpha = 10\overline{\mathbf{m}}$, $\mu = 1e^{-1}$, and producing the a relatively smooth solution and the TV method doesnt exactly produce the edges.

One of the difficulties in this algorithm is that there is no successful heuristic way to determine the value for $\alpha$ through the iterations. On top this, finding the best combination of $\alpha$ with $\mu$ involves a lot of effort. While keeping $\mu$, $\alpha$ is chosen until large-scale edges are marked and preserved. Next $\mu$ is adjusted until a desirable solution is obtained. In most of these case studies, $\alpha$ is chosen carefully by multiplying small values (in most cases $10^{-2} < \alpha < 1$) with the mean of the corresponding iterated solution of the acoustic perturbation.

As for the computational cost the algorithm is very expensive as compared to the least-square method, in which solutions are obtained by one linear inverse problem. The EPR method requires more iterations (in most cases 5-10) of a weighted linear solution as described in the previous section. Typically, in our model, each minimization of a newly iteratively reweighted solution is achieved within 200-300 iterations of the conjugate gradient algorithm depending on the type and complexity of the model.

In order to see the effectiveness of the algorithm, we compared the result obtained from the total variation regularization technique with the one recovered using a zero order quadratic regularization. To see the efficiency of EPR with TV on edges, we plot a certain portion of the velocity model. Figure 3.4 depicted the acoustic velocity profiles for the true model [a], [b] is the solution obtained with damped least-square; and [c] is the solution obtained with TV method. Looking at these figures, it is clear to say that the EPR inversion method produces better resolution, a more piecewise continuous solution of the original model pa-

Figure 3.3: The true layer velocity model (A), (B) synthetics data obtained from the single-scattering Born approximation computation, shot location at $800m$ from origin. (C) reconstructed solution using least-squares damped method and solution using edge-preserving regularization method (D). Result based on wrong choice of the regularization parameters; $\alpha$ and $\mu$.
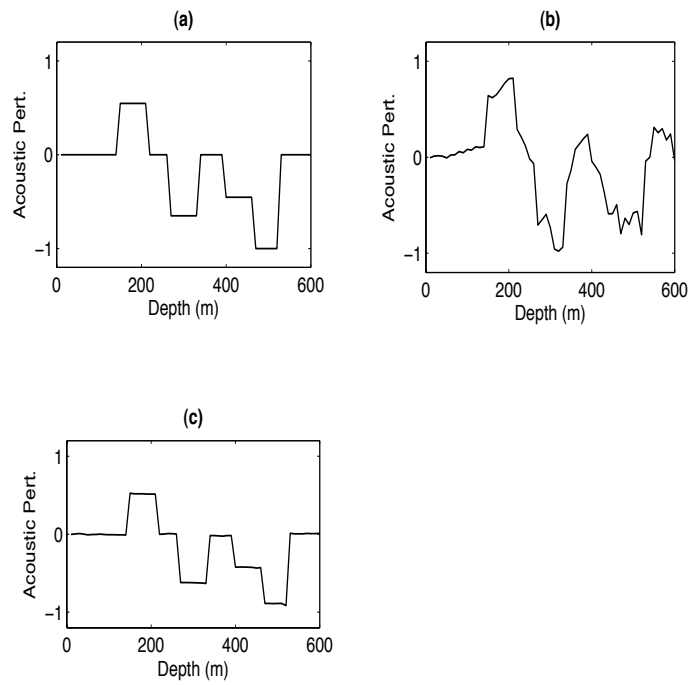
Figure 3.4: The depth acoustic velocity perturbation profile at $(300m, 0m)$. True depth acoustic velocity perturbation (a), solution using damped least-squares (b) and solution using EPR based on TV (c). The vertical indicates the normalized acoustic pertubation.

rameter while still being able to smooth the noise. The weighting function $R(\mathbf{m})$, during the reweighted iteratively linear solution, identifies the position of the edges and discontinuities and turns off smoothing at these positions. In Figure 3.4 [c], the total variation regularization functional eliminated highly oscillatory components observed in Figure 3.4 [b] while preserving discontinuities and levels off sharp edges piecewise constants without penalizing discontinuities.

The next tested model is relatively complex comapared to the previous layer velocity model (Figure 3.5 [a]). The same computational procedure is used, except for the regularized constraint parameters and the amount of noise added to the synthetic data. In this model, the velocity model profiles range from 2300 to 2800 m/s. A constant background velocity model set to 2500 m/s is used to approximate the Green's function. For forward modelling of wave propagation, a data set consisting of 5 shots (400 m spacing) and 50 receivers (32 m spacing ) was used at the surface. Figure 3.5 [b] is the generated synthetic seismic data for a single shot located at $(800m, 0m)$. The optimization of the algorithms for each iteratively reweighed norm was run for 300 iterations, with $\alpha = 0.06\overline{\mathbf{m}}$, $\mu = 1e^{-2}$, and producing the edge map and restored image shown in Figure 3.5 [d]. Figure 3.6 [b-c] is the vertical acoustic velocity profiles at $(300m, 0m)$ . Results obtained in this case still produce much better solutions than the damped least-square solutions ( see Figure 3.5 [c] and  3.6 [c] ) in terms of resolution and sharpness of the edges.

For the last test image, we present a SEG salt model Figure 3.7 [a]. In this model, the velocity model profiles range from 2400 to 2700 m/s. Here it should pointed out that the velocity of SEG salt is adjusted in such a way that the velocity profile structure departure from the reference medium is small enough that the Born approximation is applicable. The idea here is to test the velocity profile of model using TV regularization as it has a wide range of magnitude of discontinuities. However, the true velocity profile of the SEG salt model used in most case studies is not the same as the one used in this study. A constant background velocity model set to 2200 m/s is used to approximate the Green's function. For forward modelling of wave propagation, a data set consisting of 8 shots (400 m spacing ) and 108 receivers (32 m spacing) were used at the surface. The optimization of the algorithms for each iteratively reweighed was run for 300 iterations, with $\alpha = 0.02\overline{\mathbf{m}}$, $\mu = 1$, and producing the edge map and restored image shown in Figure 3.7 [c]. Figure 3.8 [b-c] is the vertical acoustic velocity profiles at $(10.5km, 0km)$. Results obtained in this case once again produce better solutions in terms of resolution and sharpness of the edges than the damped least-square solutions ( see Figure 3.7 [c] and  3.8 [c]). However, the total variation regularization method is not prefect to recover all the spikes observed in the original model (Figure 3.8 [a]). In this method, it should be pointed out that smooth gradient is not estimated.

The solutions of the EPR method based on TV is smooth and almost free of noise (see
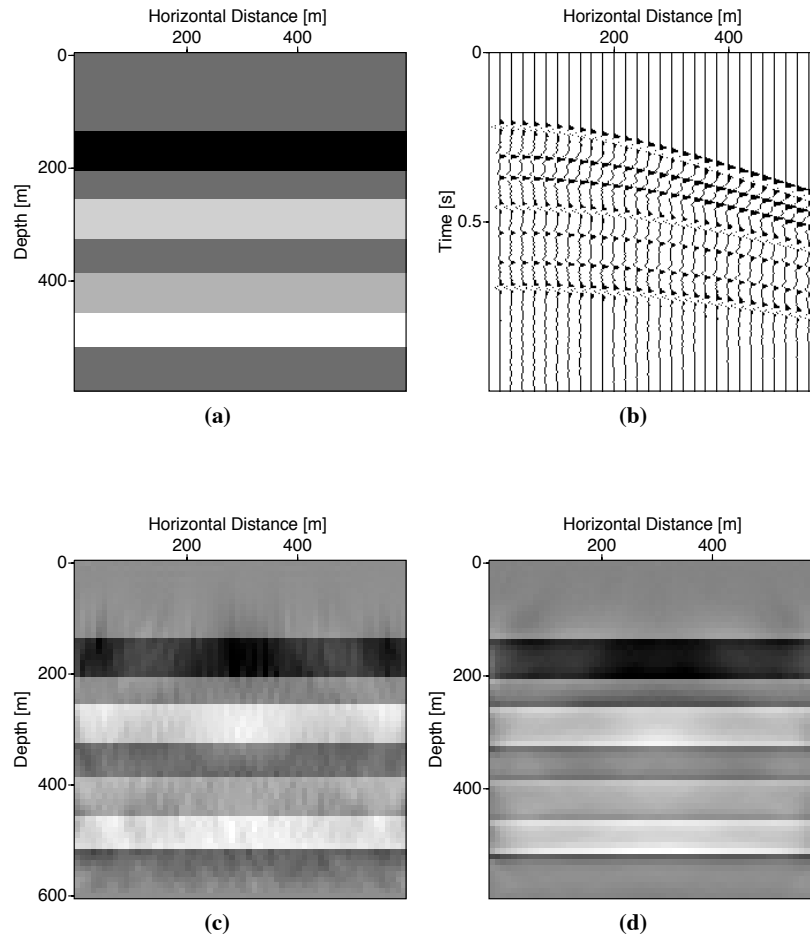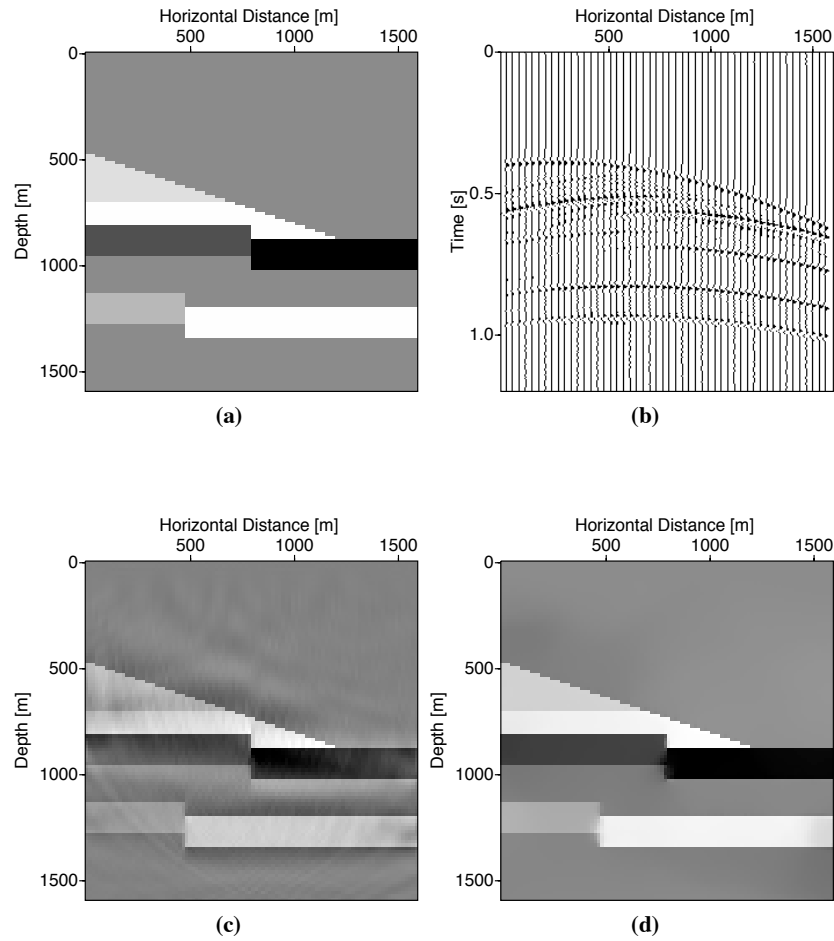
Figure 3.5: The true layer velocity model (A), (B) synthetics data obtained from the single-scattering Born approximation computation, shot location at $800m$ from origin. (C) reconstructed solution using least-squares damped method and solution using edge-preserving regularization method (D). Unlike the Tikhonov regularization, the EPR solution is clean and has no wiggles.
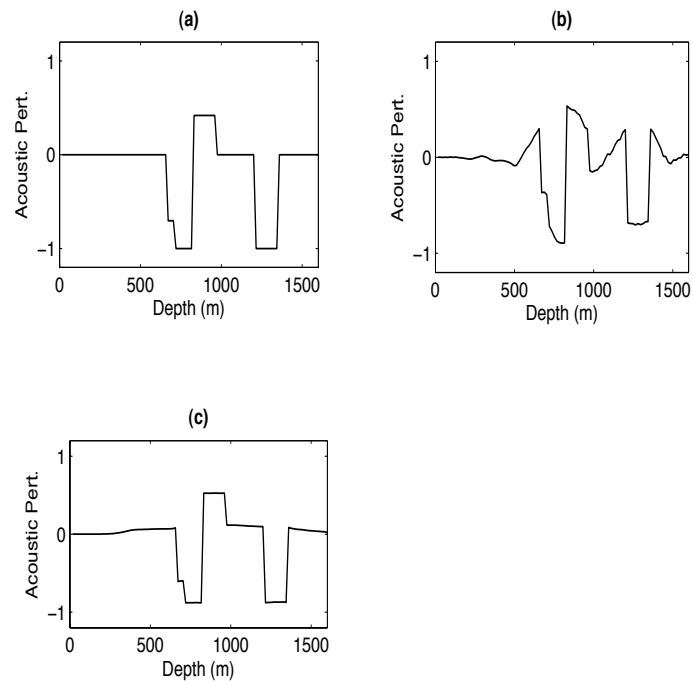
Figure 3.6: The depth acoustic velocity perturbation profile at $(560m, 0m)$. True depth acoustic velocity perturbation (a), solution using damped least-squares (b) and solution using EPR based on TV (c). The vertical indicates the normalized acoustic pertubation.
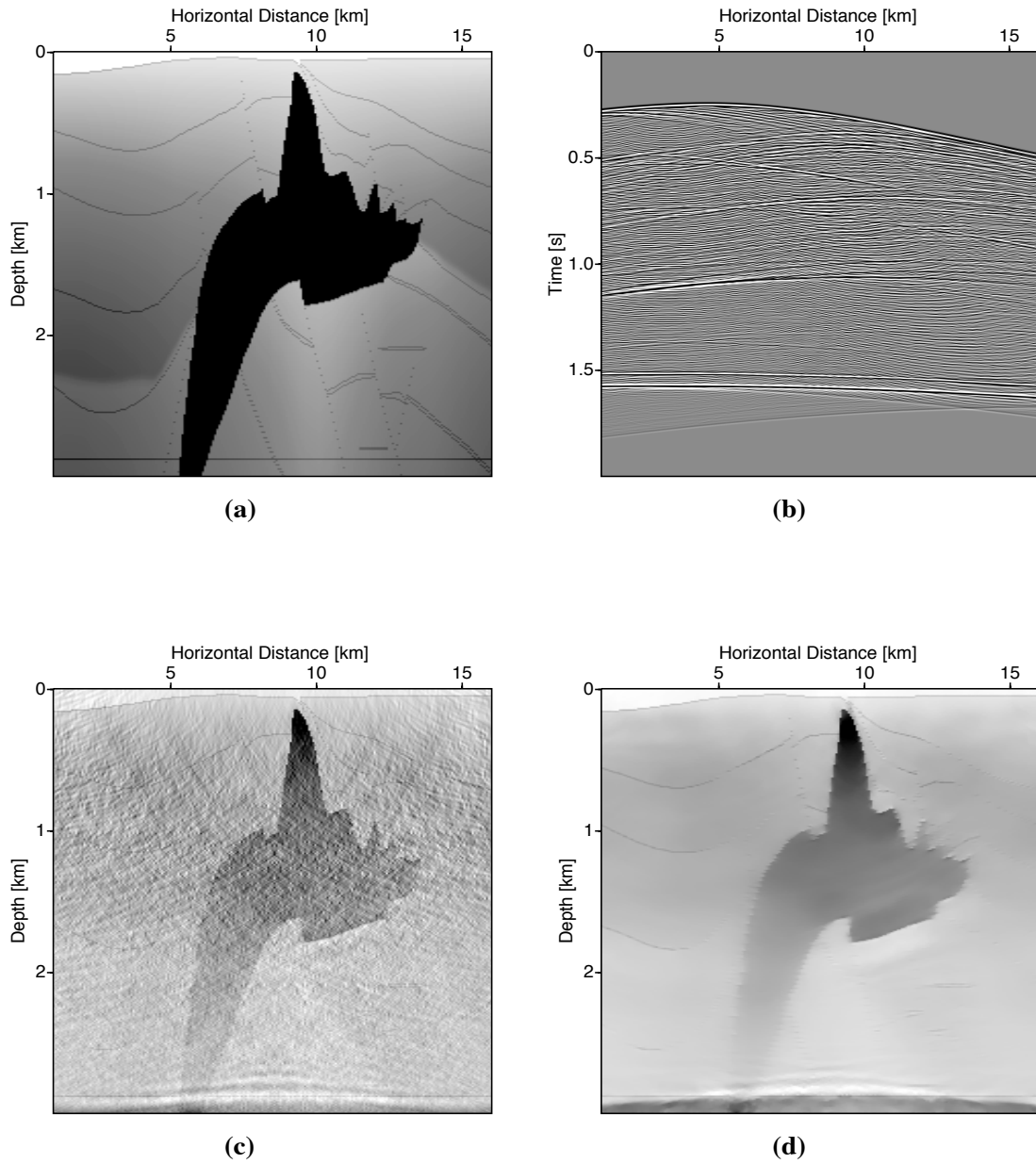
Figure 3.7:   The true layer velocity the SEG SALT model (A), (B) synthetics data obtained from the single-scattering Born approximation computation, shot location at $4.62km$ from origin. (C) reconstructed solution using least-squares damped method and solution using edge-preserving regularization method (D). Unlike the Tikhonov regularization, the EPR solution is clean and has no wiggles.

Figure 3.8:   The depth acoustic velocity perturbation profile at $(10.5km, 0km)$.  True depth acoustic velocity perturbation (a), solution using damped least-squares (b) and solution using EPR based on TV (c). The vertical indicates the normalized acoustic pertubation.

Figure 3.7 [d]) compared to the least-squares solution in Figure 3.7 [cb]. Even though the EPR method recovers better solutions; it doesnt produce sharp edges of the depth location and horizontal, and it als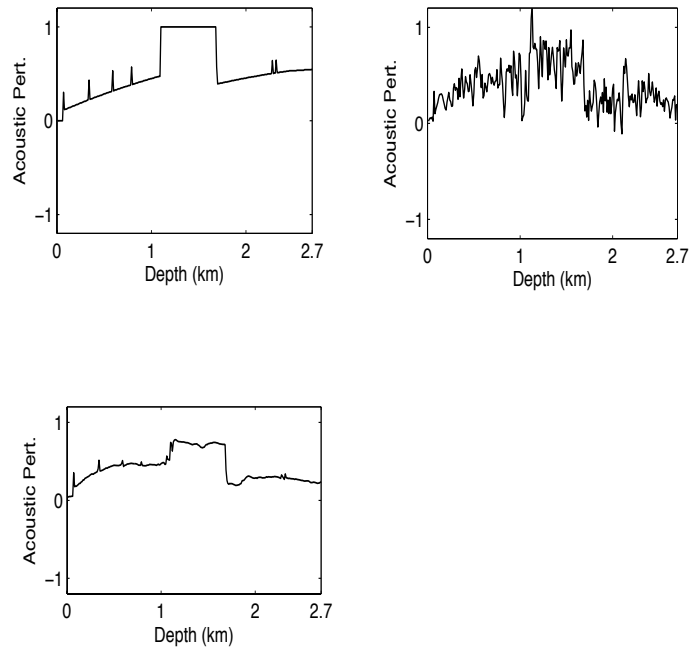o smooths some of the features of the true velocity profile of the original model due to insufficient number of sources and receivers used in the algorithm. The Born approximation, however, lacks the ability to produce the correct amplitude of the wave of the deep location. On top of that, adjusting the correct combination of the two parameters to minimize the noise while keeping the discontinuities is very difficult. Overall, the algorithm with reasonable choices of these parameters recover better results and performs simultaneous turning off of the noise while keeping the edges, discontinuity, and small details compared to the damped least-square solution.

## 3.5   Summary

In this chapter, the edge-preserving regularization based on the total variation method for small scale geophysical inverse problems is studied. The method can be very useful if edges of model or the reconstructed velocity, which in our case is the acoustic velocity perturbation, is piecewise continuous and blocky. The TV regularization method provides us with the opportunity to recover more useful information of velocity profiles from the available seismic data. Unlike in quadratic regularization, total variation regularization, a priori knowledge on the details to be preserved in the restored model is taken into account. Though it requires more effort in implementing the TV term to control the smoothing and regularization parameters, the algorithm possesses strong convergence properties, recovers the piecewise constants, reduces oscillations and, more importantly, the algorithm is practically efficient. Results obtained using TV are better compared to damped least-squares method.

The total variation method, however, has some drawbacks. For example, a very small change in the velocity profiles in the model parameter hardly preserves fine structures. Controlling the regularized parameters is also not easy. The two variable parameters have to be chosen carefully to create constraints to get a desirable solution. Further improvement in the solution may be gained by modifying the TV regularization, applying it in conjunction with other edge-preserving functions (Chan et al., 1998), using prime-dual in combination or other ways of discretizing the total variation term.

# CHAPTER 4

# Wave-equation migration using adjoint state method

## 4.1 Wave-equation migration using adjoint state method

Waveform inversion based on the least-squares principle is the most common method used in geophysical applications for determining the structure of the subsurface of Earth. The main goal of seismic data inversion is to the Earth model that best fits the observed seismic data. This chapter reviews the application of the adjoint state method in non-linear geophysical inverse problems. In order to reconstruct the model parameter from measured data in geophysical problems, the minimization of the misfit function is required. This is an inverse problem, which may be either linear or non-linear. If non-linear, the adjoint state method comes into play for minimizing the misfit function without computing the Fréchet derivatives or Jacobian matrix, which can be expensive to compute because a large linear system is created and requires excess memory as the size of the problem increases (Plessix, 2006; Symes, 2007). The use of the adjoint state method is especially important for wave-equation velocity analysis (waveform inversion). In this method a state of variables, whose physical meaning will be defined later, are the solutions of a linear system of equations.

The interest of this work is to explore the potential application of the adjoint state method in minimizing the gradient function. In addition to waveform inversions, the method is also used for seismic migration because seismic migration is an instance of inverse problems, because the gradient of the misfit functional is related to the migrated image. The optimization algorithm allows us to determine the location and amplitude of the reflector from the seismic data measured at the surface (Tarantola, 1984).

This chapter focuses on the basic methodology and underlying geophysical principles of seismic migration. Due to the time constraints, the waveform inversion has not been done. The focus of the future direction is to explore and extend the application of the method for velocity inversion. Here we restrict the method to cases of constant density and an instance of the inverse problem for migration for simplicity only. This method, however, can be generalized to variable density and can be extended for waveform inversion. The reason why we chose the adjoint state method instead of the reversal time migration is due to its importance in finding the gradient with less computational time, should one need to perform waveform inversion.

In seismic migration, the key is to retrieve the locations and the amplitudes of the reflectors, thereby constructing an image of the subsurface. Full (two-way) wave-equation migration using the finite difference (FD) method is among many other techniques in determining the structure of the subsurface of the Earth. The application of the finite difference method in seismology has been used extensively (Moczo et al., 2007). The FD method is a general method that transforms an ordinary partial differential equation (PDE) into a difference equation that can be solved numerically. The FD approximation to the wave-equation is used extensively in the seismic exploration community. One example is the Helmholtz wave-equation, which is a hyperpolic PDE in nature. This is because the fact that environment of interest is, generally, characterized by very complex environmental geometry and strong anisotropic material properties. The Helmholtz wave-equation is often used to approximately model wave propagation in inhomogeneous media.

The standard FD migration has relied on one-way wave-equations, which allow energy to propagate in one direction (generally downward). Although the method gives a reliable solution, this method may not give accurate migration when the structure has strong lateral changes in velocity or steep dips. At the same time, one-way wave-equations are also incapable of producing correct amplitudes and ignore multiple scattering along the depth coordinate (Berkhout, 1982).

Two of the reasons for using one-way wave-equation methods are the relative ease of implementation and the computational efficiency they afford. For example the Gazdag phase shift migration, which assumes a laterally-constant velocity model, is the simplest and cheapest among a few to mention (Gazdag, 1978, 1984). On the other hand, the two-way wave-equation migration, while easy to implement, is not nearly so computationally friendly should inversion be required. Two of the advantages of two-way migration is that it has no dip limitations and the propagation effects before and after reflection are taken into account. For this reason, here presented is a full wave-equation or two-way wave-equations migration in two-dimensional space using adjoint state method. An explicit time-stepping scheme is used.

## 4.2   Theory

The wave propagation is modelled by a constant-density acoustic wave-equation, the Helmholtz wave-equation. In a 2D isotropic medium the Helmholtz equation is

$$\left(\nabla^2 - \frac{1}{c(x,z)^2}\frac{\partial^2}{\partial t^2}\right)u(x,z;x_s,z_s,t) = -f(t)\delta(x-x_s)\delta(z-z_s), \tag{4.1}$$

where $u(x,z;x_s,z_s,t)$ is the seismic wavefield at $(x,z)$ due to the $s^{th}$ excitation source located at $(x_s,z_s)$, $c(x,z)$ is the wave velocity and $f(t)$ is the seismic source function. The left hand side of the above equation, the Dirac delta function, satisfies

$$f(t)\delta(x-x_s)\delta(z-z_s) = \begin{cases} f(t) & \text{for } x = x_s \text{ and } z = z_s \\ 0 & \text{for } x \neq x_s \text{ or } z \neq z_s. \end{cases} \tag{4.2}$$

In order to solve the above Helmholtz equation [4.1], we discretize and solve it using an explicit finite difference scheme based on the $2^{nd}$ order centered finite difference operator in space and time. The discretized form of the wave-equation becomes

$$\left(\frac{u^n_{i_x-1,i_z} - 2u^n_{i_x,i_z} + u^n_{i_x+1,i_z}}{\triangle x^2}\right) + \left(\frac{u^n_{i_x,i_z-1} - 2u^n_{i_x,i_z} + u^n_{i_x,i_z+1}}{\triangle z^2}\right) \tag{4.3}$$

$$+\frac{1}{c^2_{i_x,i_z}}\left(\frac{u^{n+1}_{i_x,i_z} - 2u^n_{i_x,i_z} + u^{n-1}_{i_x,i_z}}{\triangle t^2}\right) = -f(t)\delta(x_{i_x} - x_s)\delta(z_{i_z} - z_s), \tag{4.4}$$

where $\triangle t$ is the time step, and $\triangle x$ and $\triangle z$ are the grid spacing in the $x$ and $z$ direction respectively. The value of $u$ at n means its value at the $n^{th}$ time.

### 4.2.1   Boundary condition

The common problem in reflective or scattering of wave studies is its solution of the wave-equation outside the domain. If its domain is unbounded, the spurious reflections arising from the boundary will be problematic for the numerical accuracy to the solution. In order to correct for such problems, appropriate boundary conditions can be used. In this method, the $1^{st}$ order approximation of the Engquist Majda absorbing boundary conditions are used (Engquist and Majda, 1977). The first order absorbing boundary condition is expressed as

$$\left(\frac{\partial}{\partial\widehat{n}} - \frac{1}{c(x,z)}\frac{\partial}{\partial t}\right)u = 0, \tag{4.5}$$

where $\widehat{n}$ is the normal direction with respect to the appropriate domain boundary. For example, the discretization of the equation [4.5] at the boundary of the domain along the $x$ axis and $z = 0$ (that means the case for $i_z = 1$ and $2 <= i_x <= n_x - 1$) is

$$\frac{u_{i_x,2}^n - u_{i_x,1}^n}{\triangle z} - \frac{1}{c_{i_x,1}} \frac{u_{i_x,1}^n - u_{i_x,1}^{n-1}}{\triangle t} = 0. \tag{4.6}$$

## 4.2.2   Stability Condition

Solutions of a system of linear equations, when its elements are constructed using the FD method are highly susceptible to numerical dispersion and instability. In order to overcome computational errors that arise from grid dispersion effects, a fine grid spacing must be used; the numerical accuracy of the solution is controlled by number of grid point per wavelength (Lin et al., 2008). Second order finite difference implementation of the Helmholtz operator will require a minimum of 10-12 samples per wavelength. However, the Nyquist criteria requires at least two grid points per wavelength to avoid an aliasing effect (Liu, 1997).

On top of that, a time step ($\triangle t$ ) size restriction has to be imposed for stability. Information cannot be propagated across the grid faster than the grid velocity $v_g$,

$$v_g = \frac{\sqrt{\triangle x^2 + \triangle z^2}}{\triangle t}. \tag{4.7}$$

In other words, to properly sample the wavefield, the size of the grid cells and the time step have to be carefully sampled. If so, the explicit FD approximation of the solution of the Helmholtz wave-equation is conditionally stable. In addition to that, the size of the time step has to decrease as the size of the grid cells decreases.

Note that the implementation of FD method can be an explicit or implicit method. It is just a different technique for numerically solving differential equations. The explicit FD technique computes the value at time $n+1$ as a function of values at time $n$. The numerical calculation is solved recursively from the initial time to the final simulation time. On the other hand, the implicit FD method computes the value of the function at time $n$ as a function of time $n + 1$. In this case, the implicit method requires solving systems of linear equations to develop calculations from time $n$ to time $n + 1$. This method is widely used in frequency-domain versions of the acoustic wave-equation. The accuracy of the numerical solution using FD methods can be stable, unstable, conditionally stable or unconditionally stable. The unstable FD method gives a large change in the function value for small change of the initial conditions, hence it propagates large numerical errors and may not converge to the solution of the partial differential equation. The conditionally stable FD method, on the

other hand, propagates a small change in value for a small change of the initial conditions and converges to the solution of the partial differential equation, but with numerical errors. The Helmholtz wave-equation implemented in our case using the FD method is conditionally stable. For a 2D model, second order finite difference method, the stability should be stable for time steps (Kosloff and Baysal, 1982)

$$\triangle t \leq \frac{2}{\pi v_{max}} \left( \frac{1}{\triangle x^2} + \frac{1}{\triangle z^2} \right)^{-\frac{1}{2}}, \tag{4.8}$$

where $v_{max}$ is the maximum velocity of the medium. Note that the grid velocity is always greater than the maximum velocity allowed by the stability condition. For stability, the frequency content of the source must satisfy (Kosloff and Baysal, 1982)

$$f_{max} \leq \frac{1}{2} \frac{v_{min}}{max \left( \triangle x, \triangle z \right)}, \tag{4.9}$$

where $v_{min}$ is the minimum velocity of the medium.


## 4.3  Forward problem

The forward modeling is obtained by solving the explicit centered finite difference scheme of equation [4.3] - [4.4] with the following initial boundary conditions

$$u(x, z; x_s, z_s, 0) = 0 \tag{4.10}$$

$$\frac{\partial u(x, z; x_s, z_s, 0)}{\partial t} = 0. \tag{4.11}$$

The pressure field is then computed by explicit time marching scheme

$$u_{i_x,i_z}^{n+1} = 2u_{i_x,i_z}^n - u_{i_x,i_z}^{n-1} + \frac{c_{i_x,i_z}^2}{\triangle t^2} \nabla^2 u_{i_x,i_z}^n + f(t)\delta(x_{i_x} - x_s)\delta(z_{i_z} - z_s), \tag{4.12}$$

where $\nabla^2$ is the Laplace operator. Once the pressure field is obtained over the model domain, we then pick those points at which the receivers are located

$$d^{pred}(x_r, z_r, t) = R(x, z, x_r, z_r)u(x, z; x_s, z_s, t), \tag{4.13}$$

where $R(x, z, x_r, z_r)$ is a matrix operator on $u$ that contains the receiver positions $(x_r, z_r)$. All entry values except those at receiver positions are zero.

## 4.4   Migration and inversion using adjoint state method

For the sake of simplicity, let us write the wave-equation in the frequency domain

$$A(\omega, m)u(x, z; x_s, z_s, \omega) = -f(\omega)\delta(x - x_s)\delta(z - z_s), \tag{4.14}$$

where $A(\omega, m)$ is the discretized Helmholtz operator of $\left(\nabla^2 + \frac{w^2}{c(x,z)^2}\right)$, $u(x, z; x_s, z_s, \omega)$ is a complex pressure field and $m$ is the model parameter. The model parameter, $m$, is the square of the slowness $\frac{1}{c(x,z)^2}$.

The synthetic data at the receiver position are

$$d^{pred}(x_r, z_r, \omega) = R(x, z, x_r, z_r)u(x, z; x_s, z_s, \omega), \tag{4.15}$$

where $R(x, z, x_r, z_r)$ is a matrix operator on $u$ and contains the receiver positions. This matrix is exactly the same as the matrix stated in the forward modelling problem. The inversion then requires the minimization of the misfit or the least-square functional

$$J(m) = \frac{1}{2}\sum_\omega\sum_{s,r} \parallel R_{s,r}u_s(\omega, m) - d_{s,r}(\omega) \parallel^2 = \frac{1}{2} \parallel Ru - d \parallel^2 . \tag{4.16}$$

Since all the numerical calculations were done in time domain, in order to have a clear view, the above equation is equivalent to minimizing the least-square functional in the time domain as

$$J(m) = \frac{1}{2}\sum_{s,r}\int_0^T (R_{s,r}u_s(t) - d_{s,r}(t))dt, \tag{4.17}$$

where $d_{s,r}$ are the recorded data at the receiver positions. Note that the dependencies of spatial positions are not written explicitly. This problem is a non-linear inverse problem. The misfit function depends on the model parameter of the complex pressure field. The idea here is to minimize the function with respect to the model parameters. However, the model parameter, which in our case is the inverse of the slowness, is not written explicitly. Hence the cost function is implicitly dependent on the model parameters through the pressure field. For this reason, we look for other means of minimizing the cost function. One is to apply a constraint. This method is commonly used for least-squares data fitting methods. To get the expression of $\nabla_m J$ we apply constrains. The constrained optimization problem may be reformulated as an equivalent problem of searching for the optimum point of the associated Lagrangian:

$$\mathcal{L}(m, u_s, \lambda_s) = J(u_s) - \lambda_s(Au_s - f_s), \tag{4.18}$$

where $\lambda'_s s$ are the Lagrange multipliers or adjoint variables. In other words, we minimize the misfit function by forcing $Au_s - f_s = 0$. Note that,$A$ is the discretized Helmholtz operator. The direct and adjoint state equations may now be stated as stationary point conditions of the Lagrangian. Now let us take the derivative of $\mathcal{L}$ with respect to $\lambda$, $m$ and $u_s$, and compute for $\nabla_m J$. The derivative of $\mathcal{L}$ with respect to $\lambda$ and setting it to zeros gives

$$\nabla_\lambda \mathcal{L}(m, u_s, \lambda_s) = 0. \tag{4.19}$$

This gives the direct equations or the forward problem; equivalent to solving equation [4.1] in time domain

$$Au_s - f_s = 0. \tag{4.20}$$

In a similar way setting the derivative of $\mathcal{L}$ with respect to $u_s$ to zero gives

$$\nabla_u \mathcal{L}(m, u_s, \lambda_s) = 0. \tag{4.21}$$

The above equation gives the adjoint equations

$$A^* \lambda_s = R^T(Ru_s - d), \tag{4.22}$$

where $A^*$ is the adjoint operator. It propagates backward the residual wavefield between the observed and predicted data and it assumes this residual as its source term. Bear in mind that, for each of the $s$ shots, the source or residual term in the right hand side is summed over all receivers; $\lambda_{\mathbf{s}}$ is then the backpropagation of the residual field. For sufficiently smooth background velocity, $\lambda_{\mathbf{s}}$ is simply equal to the observed back-propagated wavefield of the shot gather because $Ru_s = 0$. This is equivalent to time reversal migration. Note that calculation is performed for each source. In the time domain the above equation is equivalent to solving:

$$\left(\nabla^2 - \frac{1}{c(x,z)^2}\frac{\partial^2}{\partial t^2}\right)\lambda(x, z; x_s, z_s, t) = \sum_r R^T_{s,r}(R_{s,r}u_s(T-t) - d_{s,r}(T-t)), \tag{4.23}$$

$$\lambda(x, z; x_s, z_s, 0) = 0, \tag{4.24}$$

$$\frac{\partial \lambda(x, z; x_s, z_s, 0)}{\partial t} = 0, \tag{4.25}$$

where $\lambda(x, z; x_s, z_s, t)$ back-propagates the residual (right hand side of equation [4.23]) into the model domain starting from the final times (reverse in time). It is called back-propagated field of the residual. This back-propagated wave is solved in the same procedure as the forward problem. The following procedures determines how to solve gradients of $J(m)$. Let us first assume $u = u(m)$ be the exact solution to the direct equation. The Lagrangian then reduces to

$$\mathcal{L}(m, u(m), \lambda(m)) = J(m) \tag{4.26}$$

Using equation [4.18] and [4.26], and the chain rule, we will have the following expression

$$\frac{d\mathcal{L}}{dm_i} = \frac{\partial J}{\partial m_i} = \frac{\partial u^T}{\partial m_i} \nabla_u \mathcal{L} + \frac{\partial \lambda^T}{\partial m_i} \nabla_\lambda \mathcal{L} + \frac{\partial \mathcal{L}}{\partial m_i} = \frac{\partial \mathcal{L}}{\partial m_i}. \tag{4.27}$$

Finally equations [4.18] and [4.27] give the gradient

$$\frac{\partial J}{\partial m} = \left\langle \lambda_s, \frac{\partial A}{\partial m} u_s \right\rangle_{\mathbf{x}}, \tag{4.28}$$

where $\langle . \rangle_{\mathbf{x}}$ stands for the dot product in $\mathbf{x}$. The above equation gives the gradient for one source and single frequency. To get the total gradient, we sum over all sources and frequencies

$$\frac{\partial J}{\partial m} = \sum_\omega \sum_s \left\langle \lambda_s, \frac{\partial A}{\partial m} u_s \right\rangle_{\mathbf{x}}. \tag{4.29}$$

The gradient $J$ is a vector. Outside the boundary conditions, we will have

$$\frac{\partial A}{\partial m} = \omega^2. \tag{4.30}$$

Finally at the discretization point $x$ over the range of the domain, the gradient becomes

$$\frac{\partial J}{\partial m} = w^2 Re \left( \sum_\omega \sum_s \lambda(x, z; x_s, z_s, w) u(x, z; x_s, z_s, \omega) \right). \tag{4.31}$$

In the time domain, the above equation is equivalent to

$$\frac{\partial J}{\partial m} = \sum_s \int_0^T \lambda(x, z; x_s, z_s, T - t) \frac{\partial^2 u(x, z; x_s, z_s, t)}{\partial t^2} dt. \tag{4.32}$$

The gradient of $J(m)$ is then becomes

$$\frac{\partial J}{\partial m}(\mathbf{x}) = \sum_s \int_0^T \lambda_s(\mathbf{x}, T - t) \frac{\partial^2 u_s(\mathbf{x}, t)}{\partial t^2} dt, \tag{4.33}$$

The gradient is similar to a migrated image; Claerbouts imaging principle (Claerbout, 1971). According to Lailly (1983) and Tarantola (1984), the migration corresponds to the negative of the gradient and multiplied by a positive constant, which corresponds to migration weight (equivalent to the diagonal of inverse of the Hessian matrix) (Pratt, 1990). The constant multiplication can also be obtained from the perturbation theory. For details see Pratt (1990). Equation [4.33] is summed over all shots. The quantity within the summation is the gradient per shot. For each shot gather, the migration is obtained using the gradient in equation [4.33] (the quantity within the summation).

Below is the extended approach if one is able to find the velocity model from the above expression. Since the gradient of the least-squares equation is non-linear with the model parameter, the model parameters can be optimized using the gradient descent method that minimizes $J(\mathbf{m})$ by updating the model parameter in the opposite direction of the gradient of $J(\mathbf{m})$ iteratively. By this method,

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \alpha \nabla_{\mathbf{m}} J(\mathbf{m}), \tag{4.34}$$

where the subscript indicates the iteration number and $\alpha$ is the step length. With this methodology, one can solve the velocity model without explicitly solving the Fréchet derivative, which is very time consuming. This is one of the advantages of using the adjoint state method. In this thesis, as stated early the inversion of velocity model has not been done. In the future, I intend to extend and explore the non-linear inversion method further in the frame work of the adjoint state method.

## 4.5   Result and discussions

The validity and accuracy of the finite difference method is tested using a simple dipping interface with a dip angle $18.5^o$, upper velocity 2000 m/s and lower velocity 2500 m/s; Figure 4.1. A shot gather technique is used in which a source is placed at a location $x_s = 12$ m (x-axis in a 2D) and depth of $z_s = 12$ m in a 2D space domain, see Figure 4.1. A
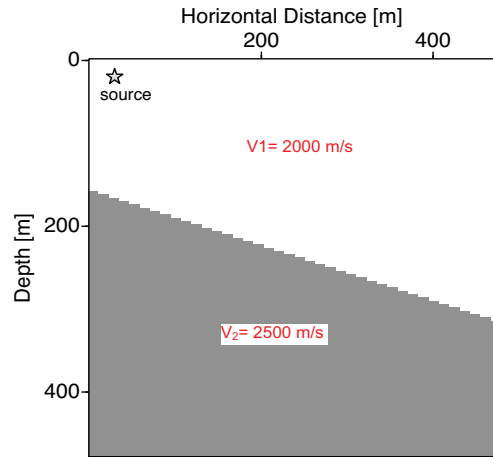
Figure 4.1: The velocity model with dipping inteface used to test the accuracy of the numerical computations. The velocity of the upper medium is 2000 m/s whereas the lower is 2500 m/s.
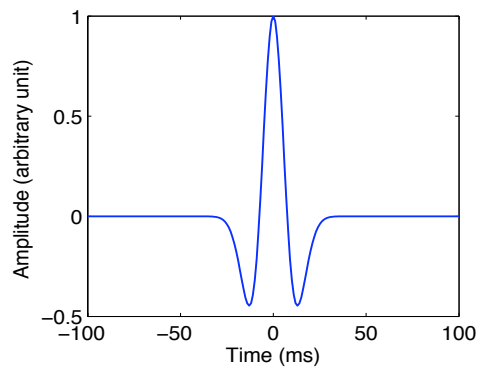


Figure 4.2: Ricker wavelet used for source excitation.

continuous equally spaced, $\triangle x = 4$ m, receivers were placed along the depth $z = 12$ m. A Ricker wavelet with center frequency of 30 Hz is used as an excitation source. A sampling rate of 1 ms is used. The location of first receiver coincides with that of the source, whereas the rest is separated from the source by $n_r \triangle x$, where $n_r$ is the $r^{th}$ receiver. Its accuracy of the calculation is then verified by calculating the travel times of the dipping layer, T,

$$T^2 = \left(\frac{2z_s}{v_1}\right)^2 + \left(\frac{x}{v_1}\right)^2 + \left(\frac{4xz_s sin(\theta)}{v_1}\right)^2, \tag{4.35}$$

where, $z_s$ is the depth of the source from the surface, $\theta$ is the dipping angle of the reflector, $v_1$ is the velocity in the first medium and $x$ is the source-receiver distance for a fixed source. Figure 4.3 shows the two travel times calculated by using the FD method of the acoustic wave-equation and the above equation. The plot is to show the comparison of the first arrival (P-waves) in these two methods. The top curve just above the seismogram is the first arrival of reflected waves obtained from equation [4.35], whereas the seismogram are synthetic reflected waves obtained from FD wave-equation. A close look at the results at $x_s = 12$ m and nearby, there is a discrepancy between the results obtained by FD approximation and the one obtained by equation [4.35]. This can be attributed to the accuracy of the numerical FD calculations (artifacts) near the boundary since the source is close to the left wall domain. Generally there will be spurious reflections arising from the boundary that is problematic for the numerical accuracy to the solution. However, when the source position is changed to $x_s = 240$ m and $z_s = 12$ m, this effect is not seen; see Figure 4.4. The first arrival of travel times (P-waves) obtained using equation [4.35] agrees well with the synthetic numerical results obtained from the FD wave-equation.

Next a relative complex model is used to better understand the quality and efficiency of the adjoint state method. The velocity model used is depicted in Figure 4.5. The velocities from the top to bottom, looking in the middle, are 2000 m/s, 2500 m/s, 3000 m/s, 3500 m/s and 4000 m/s. This method is chosen in such a way that it has steep dip, multiple layers and large velocity contrast so that it represents a practical geophysical model. A uniform grid spacing $\triangle x = \triangle z = 6$ m with a time sampling rate of $\triangle t = 1$ ms were used and a Ricker wavelet of 30 Hz was chosen. For forward modelling of wave propagation, a data set consisting of 21 shots and 155 receivers were used. The source-receiver geometry is depicted in Figure 4.6. Each source has 155 receivers. Many receivers were used in order to increase the data coverage of the medium (since there is a small numbers of sources). The receivers are located between 6 m and 1854 m. The shots are located between 12 m and 1812 m with a 90 m interval. Once again, a shot gather technique is used. For better understanding on how each shot gather estimate the location of reflectors, data from each synthetic shot gather was plotted; see Figure 4.7 - Figure 4.12. The reason for this is to show the migrated images of each source as described in the theory section and the total
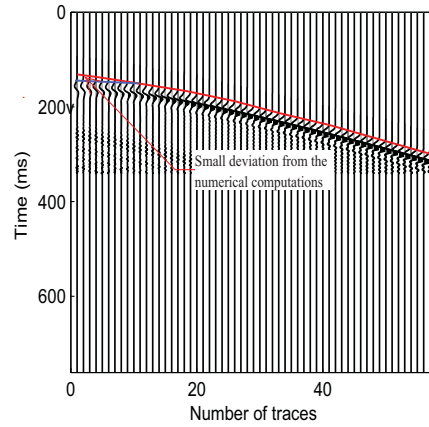
Figure 4.3: Synthetics data obtained from the finite difference computation when the source is placed at $(12m, 12m)$ and the top curve just above the seismogram is the first arrival of reflected waves obtained from equation [4.35] calculating the travel times of the dipping layer.
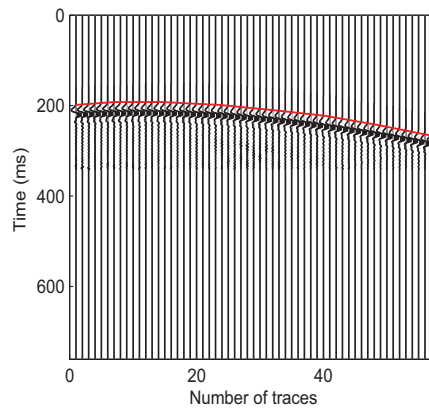


Figure 4.4: Synthetics data obtained from the finite difference computation when the source is placed at $(250m, 12m)$ and the top curve just above the seismogram is the first arrival of reflected waves obtained from equation [4.35] calculating the travel times of the dipping layer.
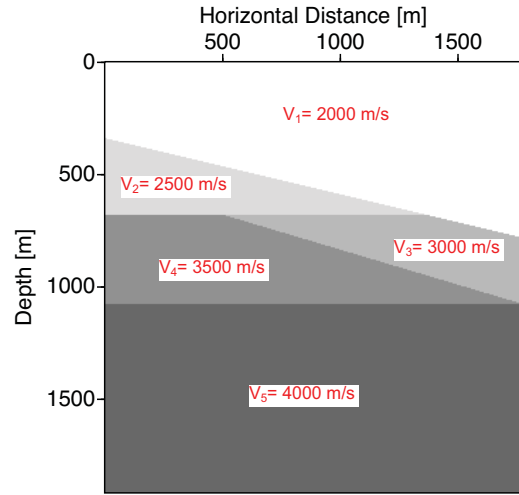
Figure 4.5: The velocity model. Looking in downward in the middle, the velocities of the medium from the top to bottom are 2000 m/s, 2500 m/s, 3000 m/s, 3500 m/s and 4000 m/s.

migrated image. These migrated images show how each source gather migration predicts the location of the reflector.

Figures 4.7 - 4.12 show the shot gather synthetic reflected seismic data obtained from the forward problem. In each case the first direct arrivals show a move out away from the source location due to the non-linearity dependency of velocity. For each shot gather, the migration is then obtained using equation [4.33]. Figures 4.13 - 4.18 show the corresponding migrated images for each shot gather. Comparing each of the shot migration images, we see the effects of source position in predicting the locations of the reflector. For example, the $1^{st}$ shot hardly predicts the location of the reflector from the right side of the medium due to most of the energy of the source propagated is reflected from the reflector on the left side of the medium. In the same way shot 11 hardly predicts the location of the reflector on the left side of medium. Note that time stepping FD method creates some artifacts. The numerical artifacts are especially visible in the migrated images.
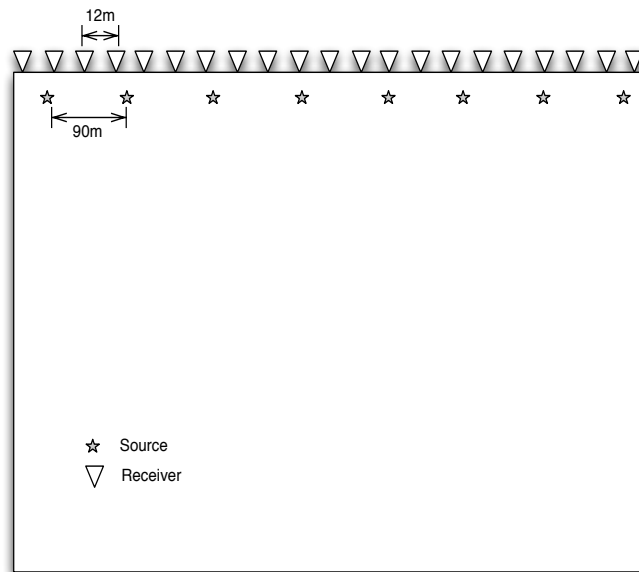
Figure 4.6: Source-receiver geometry.

The amplitude of the wave can also be improved using the normalized imaging condition rather than the conventional cross-correlation imaging condition (Claerbout, 1971) which does not consider the normalization of the amplitudes of downgoing waves. However, the FD wave-equation migration method or the reverse-time migration is still potentially the most accurate method in the sense of faithfully honoring the wave-equation. All other methods eventually fall short of it in their approximations. However, FD migration methods have the potential of generating unwanted artifacts, for example, internal multiples from locations of sharp impedance contrasts.
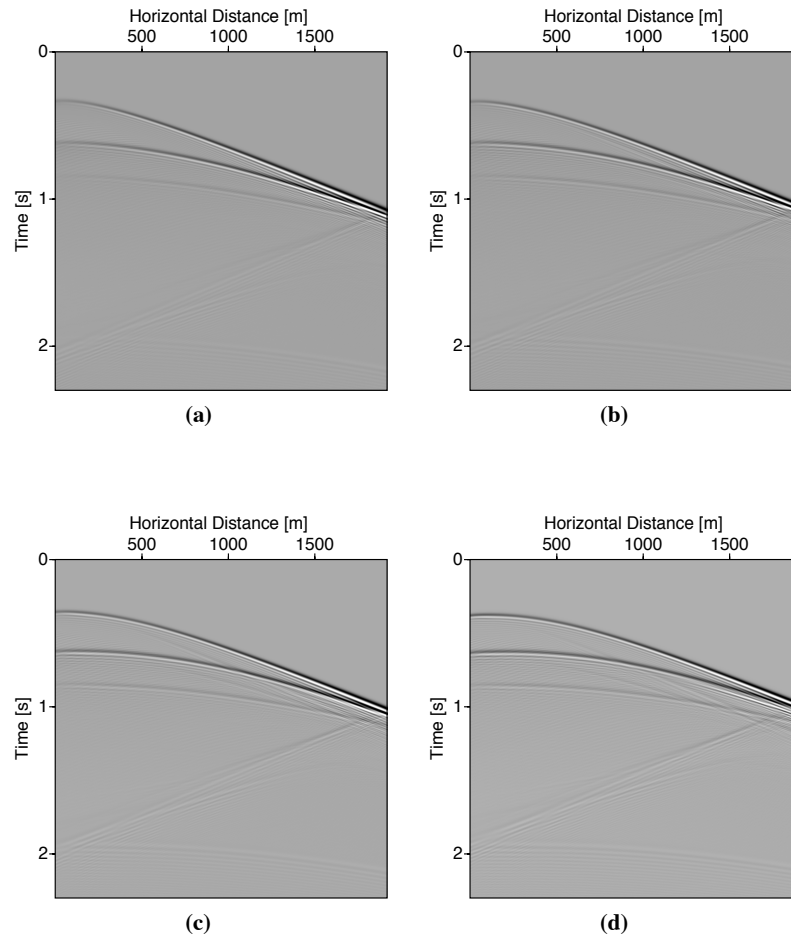
Figure 4.7: Synthetic shot gather for the model in Figure 4.5 for the first four consecutive shots; shot located at (a) $(12m, 0m)$, (b) $(102m, 0m)$ , (c) $(192m, 0m)$ and (d) $(282m, 0m)$.
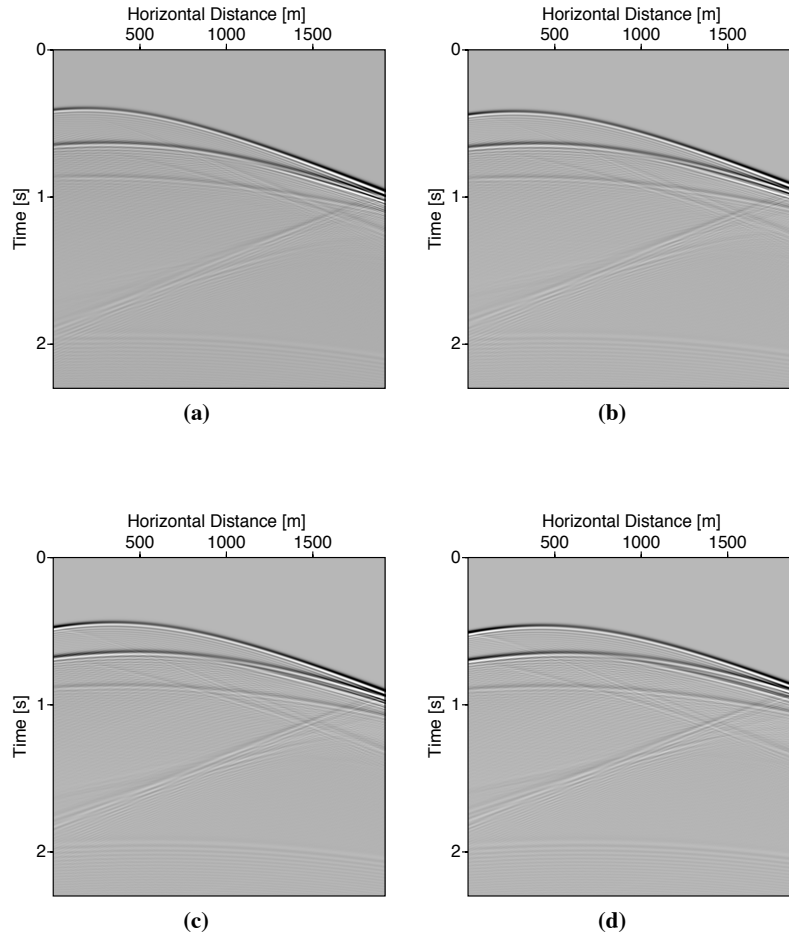
Figure 4.8: Synthetic shot gather for the model in Figure 4.5 for shot located at (a) $(372m, 0m)$, (b) $(462m, 0m)$ , (c) $(552m, 0m)$ and (d) $(642m, 0m)$ .

Figure 4.9: Synthetic shot gather for the model in Figure 4.5 for shot located at (a) $(732m, 0m)$, (b) $(822m, 0m)$ , (c) $(912m, 0m)$ and (d) $(1002, 0m)$ .

Figure 4.10: Synthetic shot gather for the model in Figure 4.5 for shot located at (a) $(1092, 0m)$, (b) $(1182, 0m)$ , (c) $(1272, 0m)$ and (d) $(1362, 0m)$ .

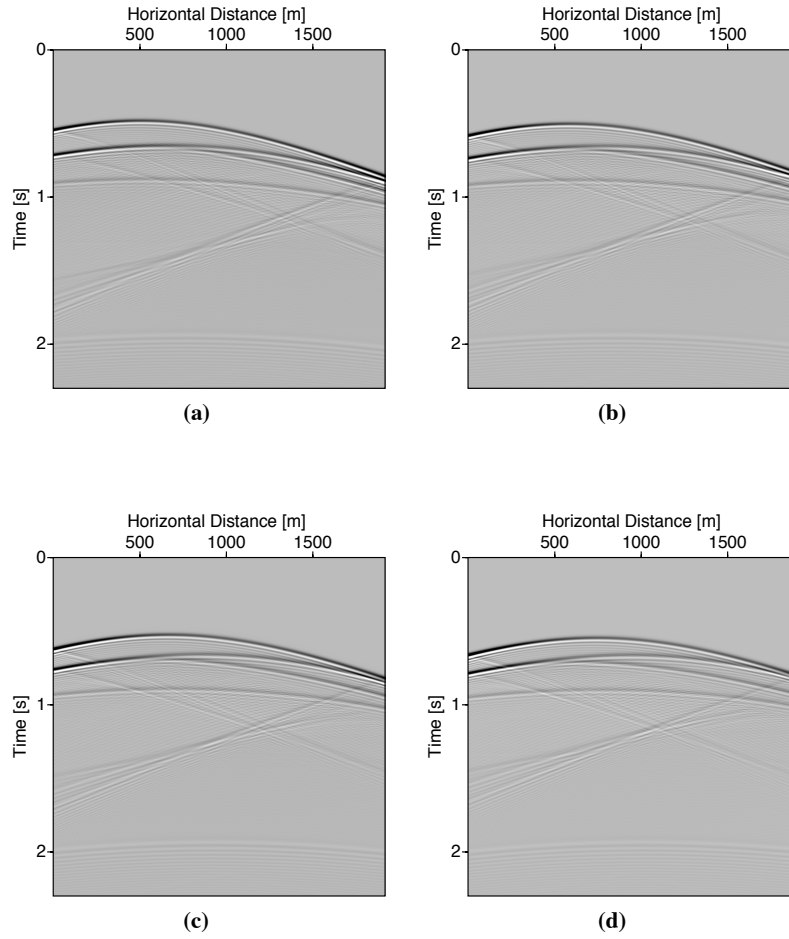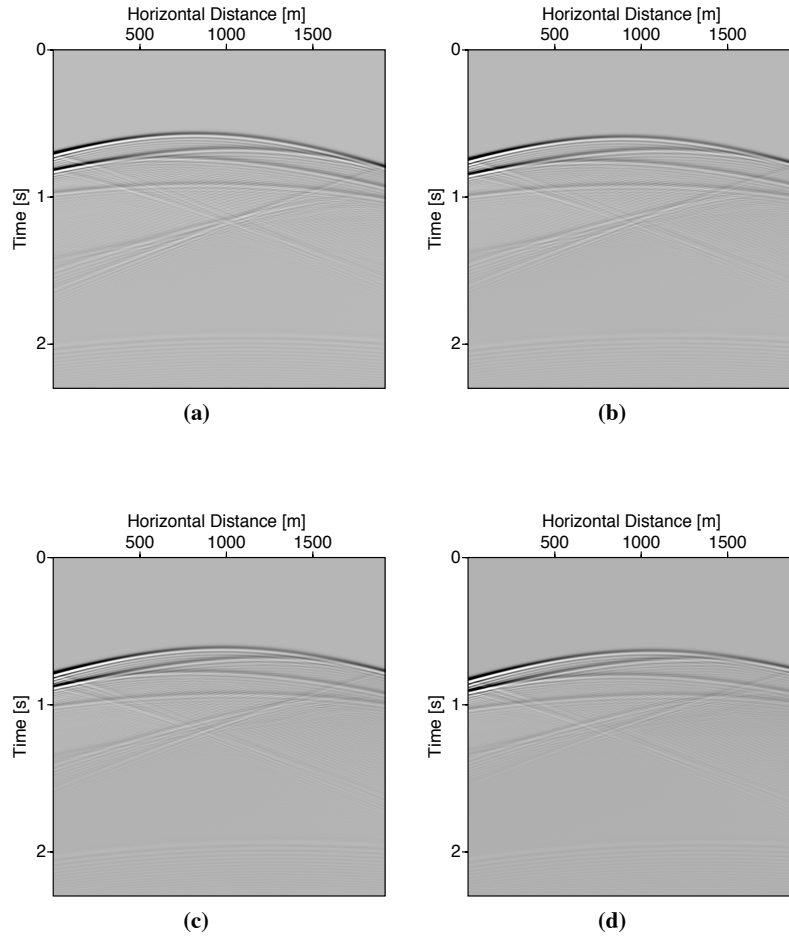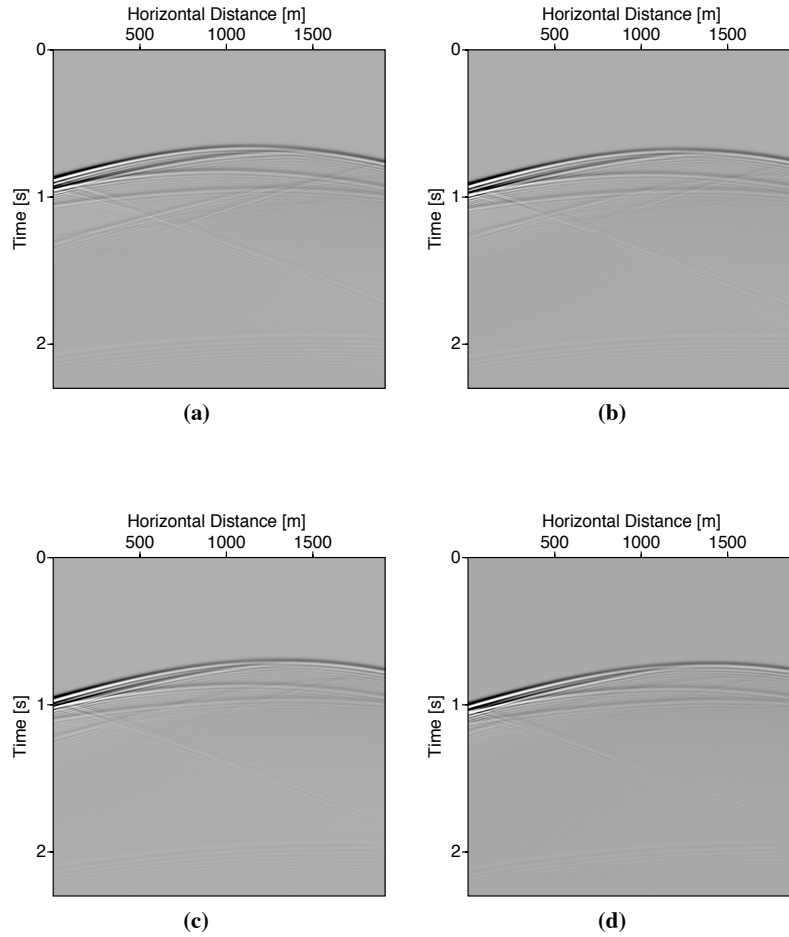Figure 4.11: Synthetic shot gather for the model in Figure 4.5 for shot located at (a) $(1452, 0m)$, (b) $(1542, 0m)$ , (c) $(1632, 0m)$ and (d) $(1722, 0m)$.
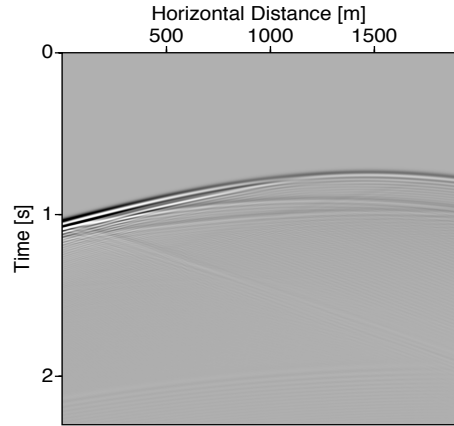
Figure 4.12: Synthetic shot gather for the model in Figure 4.5 for shot located at $(1812, 0m)$.

Each shot gather migration image is then stacked to produce the final migration image. Some of the artifacts appeared in each shot gather were suppressed in the final image. The final migration (Figures 4.19) shows the location of the reflector; in good agreement with the reflector in the velocity model, see Figures 4.5 and 4.19. However, the amplitude of the deep reflections is weak. This is seen in the figure. Even though the adjoint states method using the FD method predicts the location of the reflector; it still underestimates the amplitude of deep reflector. In order to boost and preserve somewhat the relative amplitude of deep reflection, the migration weight, which is equivalent to the diagonal of inverse of the Hessian matrix has to be taken into account (Plessix and Mulder, 2004). In our case, the migration weight is approximated using infinite receiver coverage, but the results do not indicate whether the amplitude is improved. The methodology presented here gives better resolution compared to the one computed with the Born approximation using an acoustic single-scattering if the velocity contrast in the medium is high. This is because the perturbation theory assumes roughness to be small in terms of both amplitude and slope. The Born approximation is limited to weak back-scattering features. The method given in this chapter, two-way wave-equation migration, has potential to address multiples, transmission loss and converted waves in other forms to some extent. However, it still requires detailed information about the velocity profile of the medium (Wapenaar and Berkhout, 1986). Small velocity errors can make the algorithm unstable. For this reason, two-way wave-equation migration is not so robust as the one-way wave equation migration.
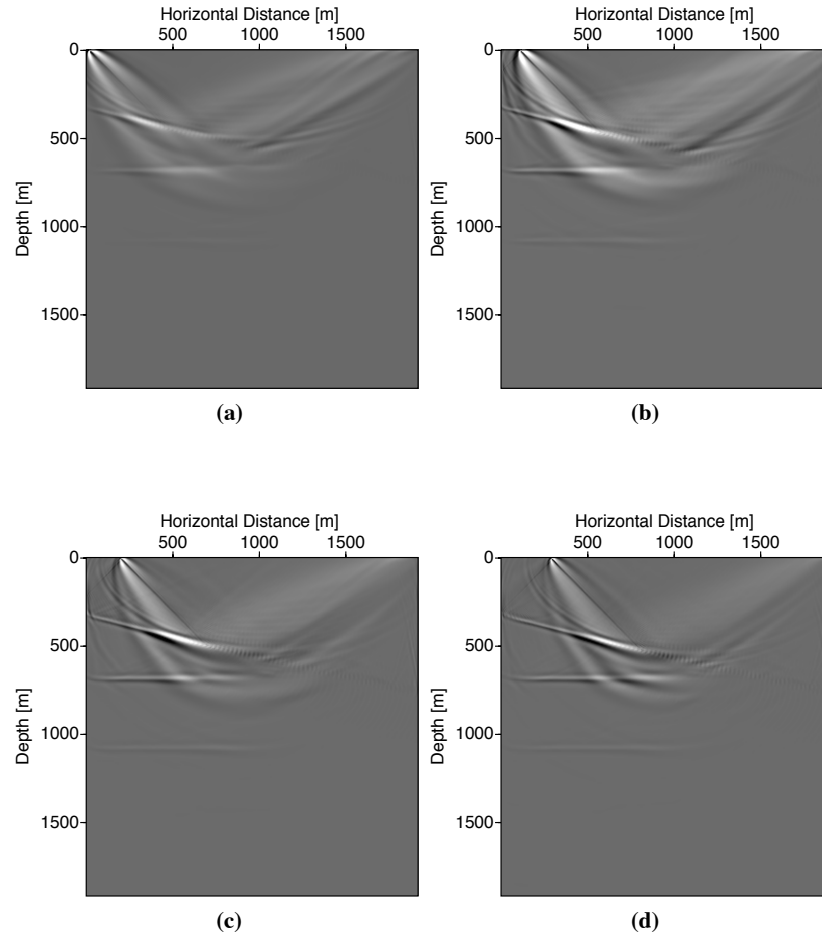
Figure 4.13: Common shot gather migration image for the $1^{st}$ shot (a), $2^{nd}$ shot (b), the $3^{rd}$ shot (c) and for the $4^{th}$ shot (d).
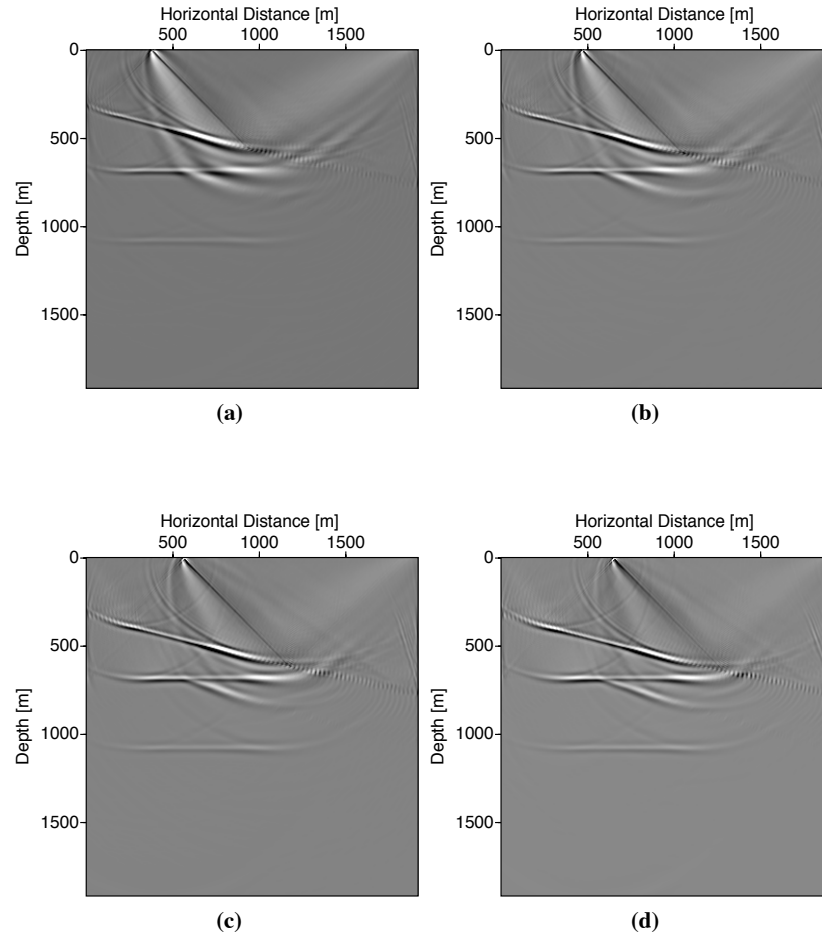
Figure 4.14: Common shot gather migration image for the $5^{th}$ shot (a), $6^{th}$ shot (b), the $7^{th}$ shot (c) and for the $8^{th}$ shot (d).
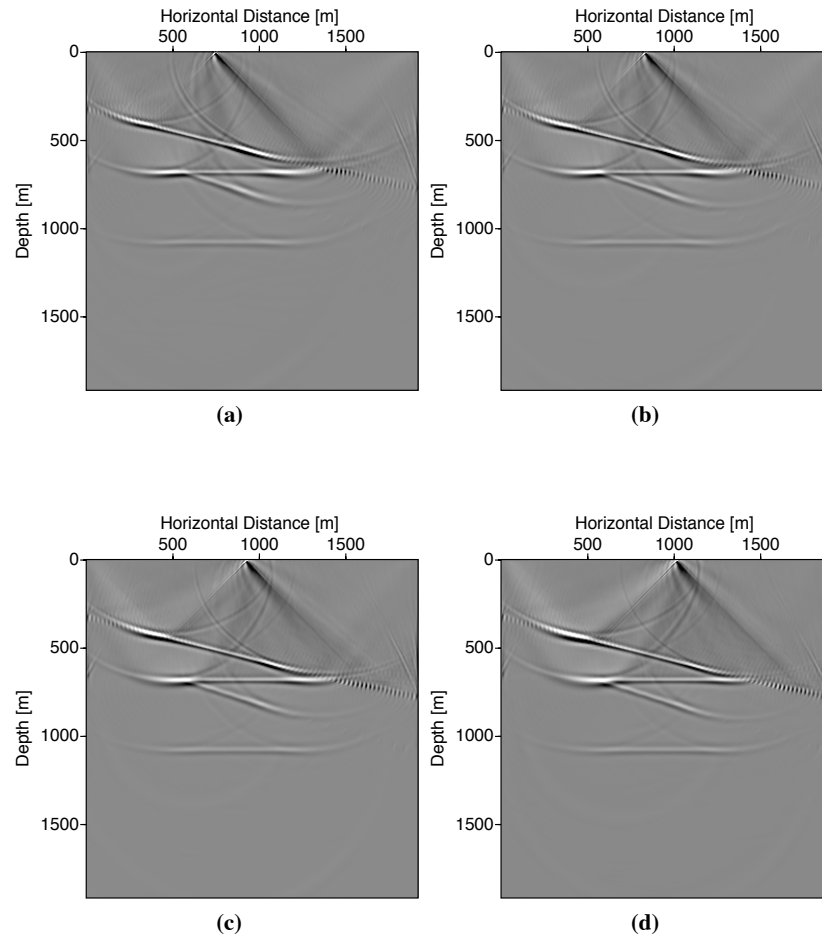
Figure 4.15: Common shot gather migration image for the $9^{th}$ shot (a), $10^{th}$ shot (b), the $11^{th}$ shot (c) and for the $12^{th}$ shot (d).
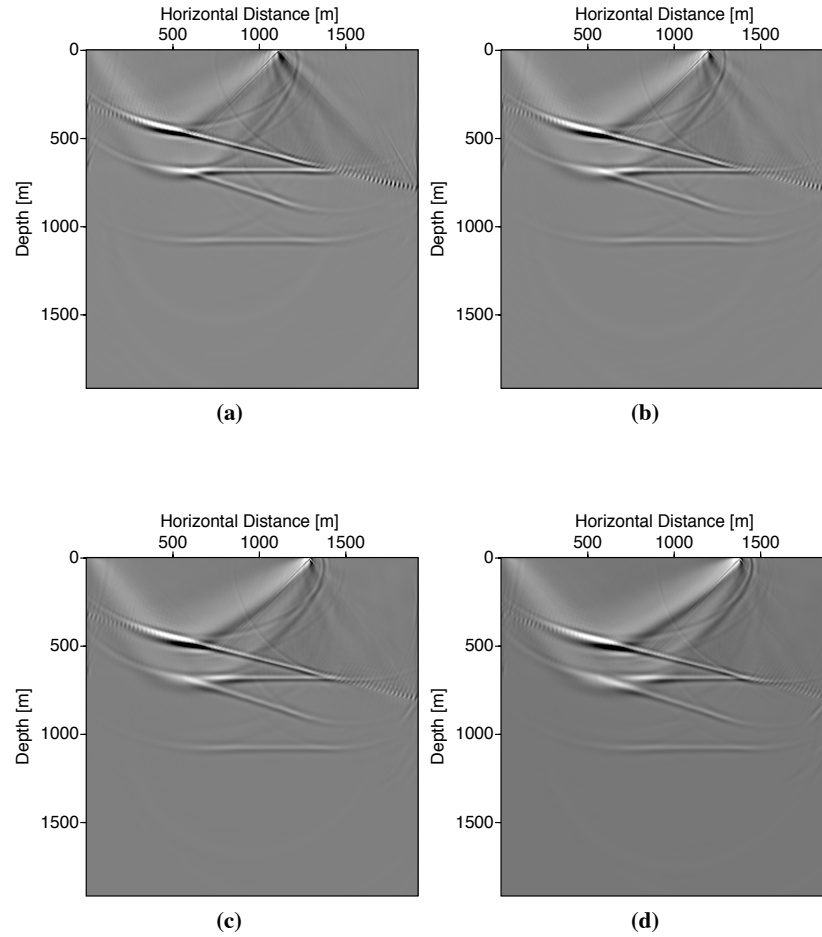
Figure 4.16: Common shot gather migration image for the $13^{th}$ shot (a), $14^{th}$ shot (b), the $15^{th}$ shot (c) and for the $16^{th}$ shot (d).

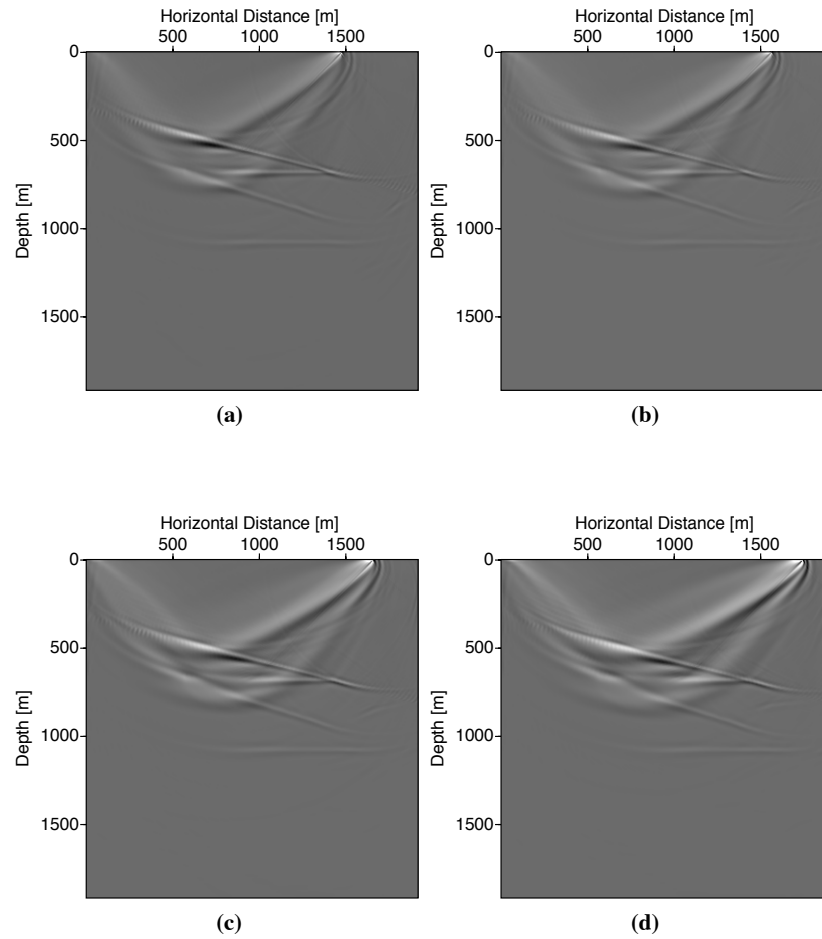Figure 4.17: Common shot gather migration image for the $17^{th}$ shot (a), $18^{th}$ shot (b), the $19^{th}$ shot (c) and for the $20^{th}$ shot (d).
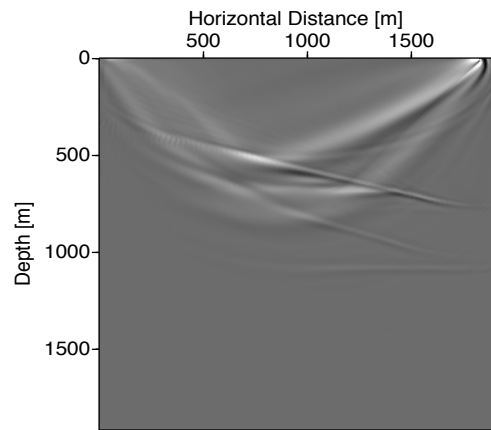
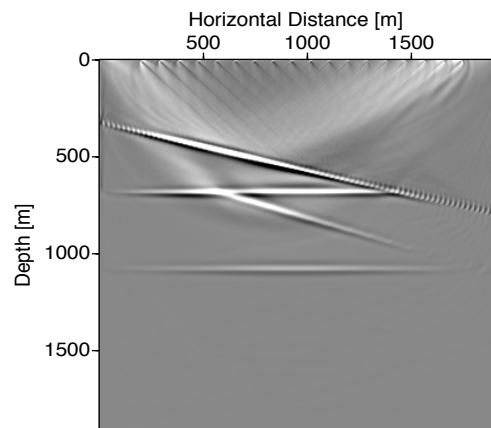Figure 4.18: Common shot gather migration image for the $21^{th}$ shot.



Figure 4.19: Final migration image which is the the sum of all the migrations.

## 4.6   Summary

In this chapter, a review of the two-way wave-equation migration in a constant density medium using the finite-difference in the time-domain was presented. The FD method is a robust numerical method applicable to structurally complex media. Due to its relative accuracy and computational efficiency, it is the dominant method in seismic modeling and is increasingly more important in the seismic industry and for structural modeling. The basic formulations and properties of the FD schemes were reviewed. The explicit marching or time-stepping schemes based on FD discretization of initial/boundary value problem was used. The use of explicit time-stepping schemes unlike the implicit method makes the numerical calculation fast. However, it requires a careful choice of step size and grid size for stability. The method is similar to the reverse-time migration algorithm implemented using a adjoint state method based on the FD scheme with second-order accuracy in time and space. The second order accuracy gives solutions closer to the actual one. Migration is applied to each shot gather, where recorded multi component data are reversed in time order and applied at the corresponding receiver positions as sources for reverse-time extrapolation. Migration based on the two-way method should be expected to produce better lateral resolution than the one-way wave-equation migration.

The use of the adjoint state method for optimizing the gradient of the misfit has a significant advantage should waveform inversion or model parameter estimation be required. In estimating the model parameters from measured data that generally consist of minimizing an error/cost functional requires the computation of the gradient. Calculating the gradient with respect to the model parameters is not simple. The adjoint-state method is computationally more efficient than optimization using the method of Fréchet derivatives. Results obtained using the adjoint-state method are non-linear with the model parameters. As a result, the optimization problem should be solved with a non-linear optimization method such as a quasi-Newton, non-linear conjugate gradient technique. The objective of this thesis was to extend the application of the method for velocity inversion. Although the non-inversion method using the adjoint state has not been done in this thesis, the work can be extended in the future.

# CHAPTER 5

# Conclusions

## 5.1   Conclusions

In this thesis, the basic theory of inverse problems has been reviewed. In the first two chapters, the linear scattering problem has been posed as an inverse problem. The linearization of the wavefield was based on the framework of the single-scattering Born approximation that ignores multiple scattering due to the heterogeneity of the medium. This approximation is only valid for small perturbation of the velocity from the assumed background velocity.

The forward and adjoint modeling problems can be computationally solved easily by discretization of the linearization of the scattering waves. For approximate solution of inverse problems to be implemented computationally, the theory of the GRT, which relates the scattering fields to velocity potential, is employed. This method is then employed in Chapter 3 to compute least-square inversion for estimating the acoustic potential profiles of the medium.

Many geophysical inverse problems are ill-posed and have to be regularized. The most often used solution methods for solving ill-posed problems are based on the use of quadratic regularization. However, this results in smooth solutions and are known not to be suitable when the model parameter is piecewise continuous, blocky and when edges are desired in the regularized solution. To avoid the smoothing of edges, which are very important attributes of the image, an edge-preserving regularization (non-quadratic regularization) term has to be employed.

The edge-preserving regularization based on the total variation method for small-scale geophysical inverse problems is studied. The total variation function allows discontinuities and

a sharp edges class of solutions. The method can be very useful if edges of the model or the reconstructed velocity, which in our case is the acoustic velocity perturbation, is piecewise, continuous and blocky. The TV regularization method provides an opportunity to recover more useful information of velocity profiles from the available seismic data. Unlike the quadratic regularization that results in smooth solutions, in total variation regularization, a priori knowledge on the details to be preserved in the restored model is taken into account. Although more effort in implementing the TV term to control the smoothing and regularization parameter is required, the algorithm possesses strong convergence properties, recovers the piecewise constants, reduces oscillations and, more importantly, is practically efficient.

The total variation method, however, has some drawbacks. For example, a very small change in the velocity profiles in the model parameter is tamed and hardly preserves fine structures. Controlling the model parameters is also very difficult. The two variable parameters have to be chosen carefully to create constraints to obtaining a desirable solution. Further improvement in the solution may be gained by modifying the TV regularization, applying it in conjunction with other edge-preserving functions (Chan et al., 1998), using prime-dual in combination or other ways of discretizing the total variation term.

The two-way wave-equation migration using the finite-difference in the time-domain is also reviewed. The finite-difference method is a robust numerical method applicable to structurally complex media. Due to its relative accuracy and computational efficiency, this is the dominant method in seismic modeling and is increasingly more important in the seismic industry and for structural modeling. The basic formulations and properties of the finite-difference schemes were reviewed. The explicit marching or time-stepping schemes based on finite difference discretization of initial/boundary value problem were used. The use of explicit time-stepping schemes, unlike the implicit method, allows for faster numerical calculations. However, it requires a careful choice of step size and grid size for stability. The method is similar to the reverse-time migration algorithm although it is implemented using an adjoint state method based on the finite difference scheme with second-order accuracy in time and space. Migration is applied to each shot gather, where recorded multi component data are reversed in time order and applied at the corresponding receiver positions as sources for reverse-time extrapolation. Migration based on the two-way method should be expected to produce better lateral resolution.

The use of the adjoint state method for optimizing the gradient of the cost function has a significant advantage should waveform inversion or model parameter estimation be required. In seismic waveform inversion, in order to minimize the cost function, several forward modeling and residual back propagations are required to gradually update the velocity field. The gradient of the cost functional is related to the velocity filed, and can be computed by using

Fréchet derivative, perturbation theory or the adjoint state method. However, with Fréchet derivatives, a large linear system is created and requires large memory machine as the size of the problem increases. Calculating the gradient with respect to the model parameters is not simple. The adjoint state method enables computation of the gradient of the cost function without computing the Fréchet derivatives (Plessix, 2006; Symes, 2007).

The adjoint state method is also computationally more efficient than optimization using the method of Frechet derivatives. Results obtained using the adjoint state method are non-linear with the model parameters. As a result, the optimization problem should be solved with a non-linear optimization method such as a quasi-Newton, non-linear conjugate gradient technique.

The objective of this thesis was to extend the application of the method for velocity inversion. However, due to the time constraints the non-inversion method using the adjoint state has not been done. The work can be extended in the future.

# Bibliography

Abramowitz, M. and Irene A. Stegun. "Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables." *U.S. Department of Commerce* (1972): 355–434.

Askan, Aysegul, Volkan Akcelik, Jacobo Bielak, and Omar Ghattas. "Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures." *Bulletin Of The Seismological Society Of America* 97 (DEC 2007): 1990–2008.

Aster, Richard, Brian Borchers, and Clifford Thurber. *Parameter Estimation and Inverse Problems (International Geophysics).* Har/cdr edition. Academic Press, January 2005.

Baysal, E., D. D. Kosloff, and J. W. C Sherwood. "Reverse Time Migration." *Geophysics* 48 (1983): 1514–1524.

Berkhout, A. L. " A theoretical aspects." *Elsevier Science Publ. co. Inc* 14A ( 1982): 365–381.

Beylkin, G. "Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform." *J. Math. Phys.* 26 (1985): 99–108.

Beylkin, G. and R. Burridge. "Linearized inverse scattering problems in acoustics and elasticity." *Wave motions* 12 (1990): 15–52.

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

Chan, T. F., G. H. Golub, and P. Mulet. "A nonlinear primal-dual method for total variation-based image restoration." *SIAM Journal on Scientific Computing* 20 (JUL 22 1998): 1964–1977.

Chan, Tony F. and Sung Ha Kang. "Error analysis for image inpainting." *Journal Of Mathematical Imaging And Vision* 26 (NOV 2006): 85–103.

Chan, Tony F. and Pep Mulet. "On the Convergence of the Lagged Diffusivity Fixed Point Method in Total Variation Image Restoration." *SIAM Journal on Numerical Analysis* 36 (1999): 354–367.

Chan, Tony F., H. M. Zhou, and Raymond H. Chan. "Continuation method for total variation denoising problems." *Advanced Signal Processing Algorithms* 2563 (1995): 314–325.

Charbonnier, P., L. BlancFeraud, G. Aubert, and M. Barlaud. "Deterministic edge-preserving regularization in computed imaging." *IEEE Transactions on Image Processing* 6 (FEB 1997): 298–311.

Christiansen, Oddvar, Tin-Man Lee, Johan Lie, Usha Sinha, and Tony F. Chan. "Total Variation Regularization of Matrix-Valued Images." *Int J Biomed Imaging* 27432 (2007): 1–6.

Claerbout, J. F. *Imaging the Earth's Interior.* Blackwell Science Inc., 1985.

Claerbout, Jon F. "Toward a unified theory of reflector maping." *Geophysics* 36 (1971): 467–481.

Clapp, Robert G., Biondo L. Biondi, and Jon F. Claerbout. "Incorporating geologic information into reflection tomography." *Geophysics* 69 (2004): 533–546.

Clearbout, J. F. " Imaging the Earth.s Interior." *Blackwell Scienti.c Publications, London* (1985).

Dibos, Francoise and Georges Koepfler. "Global total variation minimization." *SIAM Journal on Numerical Analysis* 37 (1999): 646–664.

Engquist, B and A Majda. "Absorbing boundary-conditions for numerical-simulation of waves." *Mathematics Of Computation* 31 (1977): 629–651.

Farquharson, Colin G. and Douglas W. Oldenburg. "Non-linear inversion using general measures of data misfit and model structure." *Geophysical Journal International* 134 (July 1998): 213–227(15).

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

Gazdag, J. "Wave equation migration with the phase shift method." *Geophysics* 43 (1978): 1342–1351.

Gazdag, J. "Migration of seismic data by phase shift plus interpolation." *Geophysics* 49 (1984): 124–131.

Geman, D and C. D. Yang. "Nonlinear image recovery with half-quadratic regularization." *IEEE Transactions on Image Processing* 4 (JUL 1995): 932–946.

Jackson, D. D. "Interpretation of inaccurate, insufficient and inconsistent data." *Geophysical Journal Of The Royal Astronomical Society* 28 (1972): 97–&.

Kelly, K. R., R. W. Ward, Sven Treitel, and R. M. Alford. "Synthetic Seismograms: Finite-Difference Approch." *Geophysics* 41 (1976): 2–27.

Kosloff, Dan D. and Edip Baysal. "Forward modeling by a Fourier method." *Geophysics* 47 (1982): 1402–1412.

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

Lin, Tim T.Y., Evgeniy Lebed, Yogi A. Erlangga, and Felix J. Herrmann. "Interpolating solutions of the helmholtz equation with compressed sensing." *SEG Technical Program Expanded Abstracts* 27 (2008): 2122–2126.

Liu, Q. H. "The PSTD algorithm: A time-domain method requiring only two cells per wavelength." *Microwave and Optical Technology Letters* 15 (1997): 158–165.

McMeChan, G. A. "Migration by extrapolation of time-dependent boundary-values." *Geophysical Prospecting* 31 (1983): 413–420.

Miller, D., M. L. Oristaglio, and G. Beylkin. "A new slant on seismic imaging: Migration and integral geometry." *Geophysics* 52 (1987): 943–964.

Moczo, P., J. O. A. Robertsson, and L. Eisner. "The finite-difference time-domain method for modeling of seismic wave propagation." *Advances in Geophysics* 48 (2007): 421–516.

Moisan, Lionel. "How to discretize the Total Variation of an image?." *PAMM* 7 (2007): 1041907–1041908.

Molina, R., J. Nunez, F. J. Cortijo, and J. Mateos. "Image restoration in astronomy - A Bayesian perspective." *IEEE Signal Processing Magazin* 18 (MAR 2001): 11–29.

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

Mufti, I. R., J. A. Pita, and R. W. Huntley. "Finite-difference depth migration of exploration-scale 3-D seismic data." *Geophysics* 61 (1996): 776–794.

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

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

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

Portniaguine, Oleg and Michael S. Zhdanov. "Focusing geophysical inversion images." *Geophysics* 64 (1999): 874–887.

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

Ristow, D. and T. Rühl. "Fourier finite-difference migration." *Geophysics* 59 (1994): 1882–1893.

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

Scales, J. A. and M. L. Smith. *Introductory Geophysical Inverse Theory*. Samizdat Press, 1994.

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

Shi, Yuying, Qianshun Chang, and Jing Xu. "Convergence of fixed point iteration for deblurring and denoising problem." *Applied Mathematics And Computation* 189 (JUN 15 2007): 1178–1185.

Snieder, R. and A. Lomax. "Wavefield smoothing and the effect of rough velocity perturbations on arrival times and amplitudes." *Geophysical Journal International* 125 (JUN 1996): 796–812.

Stekl, I. and R. Gerhard Pratt. "Accurate viscoelastic modeling by frequency-domain nite differences using rotated operators." *Geophysics* 63 (1998): 17791794.

Stolt, R. H. "Migration by Fourier Transform." *Geophysics* 43 (1978): 23–48.

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

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

Tikhonov, A. N. "Regularization of incorrectly posed problems." *Doklady Akademii Nauk SSSR* 153 (1963): 49–&.

Tikhonov, A. N. "Solution of incorrectly formulated problems and regularization method." *Doklady Akademii Nauk SSSR* 151 (1963): 501–&.

Trad, Daniel, Tadeusz Ulrych, and Mauricio Sacchi. "Latest views of the sparse Radon transform." *Geophysics* 68 (2003): 386–399.

Trad, Daniel O., Tadeusz J. Ulrych, and Mauricio D. Sacchi. "Accurate interpolation with high-resolution time-variant Radon transforms." *Geophysics* 67 (2002): 644–656.

Valenciano, Alejandro A., Morgan Brown, Antoine Guitton, and Mauricio D. Sacchi. "Interval velocity estimation using edge-preserving regularization." *SEG Technical Program Expanded Abstracts* 23 (2004): 2431–2434.

Vogel, C. R. and M. E. Oman. "Fast Numerical Methods for Total Variation Minimization in Image Reconstruction." (1995): 359–67.

Vogel, C. R. and M. E. Oman. "Iterative Methods for Total Variation Denoising." *SIAM Journal on Scientific Computing* 17 (1996): 227–238.

Vogel, C. R. and M. E. Oman. "Fast, robust total variation-based reconstruction of noisy, blurred images." *IEEE Transactions On Image Processing* 7 (JUN 1998): 813–824.

Vogel, Curtis R. "Nonsmooth Regularization." (1997): 1–11.

Vogel, Curtis R. "Computational Methods for Inverse Problems." (2002).

Wang, Yang and Haomin Zhou. "Total variation wavelet-based medical image denoising." *Int J Biomed Imaging* 89095 (2006): 1–6.

Wapenaar, C. P. A. and A. J. Berkhout. "Wave field extrapolation techniques for inhomogeneous media which include critical angle events. Part III: appications in modeling, migration and inversion." *Geophysical Prospecting* 34 (1986): 180–207.

Youzwishen, C. F. "Non-linear sparse and blocky constraints for seismic inverse problems." *MSc. thesis, the University of Alberta* (2001).

Youzwishen, C. F. and M. D. Sacchi. "Edge preserving imaging." *Journal Of Seismic Exploration* 15 (MAY 2006): 45–57.

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

Zhou, Changxi, Wenying Cai, Yi Luo, Gerard T. Schuster, and Sia Hassanzadeh. "Acoustic wave-equation traveltime and waveform inversion of crosshole seismic data." *Geophysics* 60 (1995): 765–773.

# APPENDIX A

# Discretization of total variation

## A.1 Discretization of total variation regularization operator

In this section, the discretization of the penality functional stated in Chapter 3 will be discussed. The linearization of the total variation is based on Vogel (Vogel, 2002). The discretize regularization operator of the total variation penalty function in two-dimensional is assumed of the form

$$J(f) = \frac{1}{2} \sum_{i_x=1}^{n_x} \sum_{i_z=1}^{n_z} \Psi((D_{i_x,i_z}^x f)^2 + (D_{i_x,i_z}^z f)^2) \triangle x \triangle z, \tag{A.1}$$

where $f$ is the model parameter and

$$D_{i_x,i_z}^x f = \frac{f_{i_x,i_z} - f_{i_x-1,i_z}}{\triangle x}, \tag{A.2}$$

$$D_{i_x,i_z}^z f = \frac{f_{i_x,i_z} - f_{i_x,i_z-1}}{\triangle z}, \tag{A.3}$$

and $\Psi(t)$ is of the form

$$\Psi(t) = 2\sqrt{t + \alpha^2}. \tag{A.4}$$

The gradiant function can be computed by taking the derivative of the above equation with respect to $f$. Here, the main intention is to linearize the gradient of the penality function. To minimize, let $v$ be a two-dimensional model parameter and use the following optimization

technique trick,

$$\frac{d}{d\tau}J(f + \tau v)|_{\tau=0} = \left(\frac{d}{df}J(f + \tau v) \cdot \frac{d}{d\tau}(f + \tau v)\right)|_{\tau=0}, \tag{A.5}$$

$$\frac{d}{d\tau}J(f + \tau v)|_{\tau=0} = \left(\frac{d}{df}J(f + \tau v) \cdot v\right)|_{\tau=0}, \tag{A.6}$$

$$\frac{d}{d\tau}J(f + \tau v)|_{\tau=0} = J'(f) \cdot v = \langle J'(f), v \rangle, \tag{A.7}$$

where $\langle .,. \rangle$ denotes the inner dot product between two functions and $grad\ J(f) = J'(f)$.

In order to linearize the discretizte of gradient of the penality function, we use the following mathematical formalism.

$$\frac{d}{d\tau}J(f + \tau v)|_{\tau=0} = \frac{1}{2}\sum_{i_x=1}^{n_x}\sum_{i_z=1}^{n_z}\Psi'_{i_x,i_z}\left[(D^x_{i_x,i_z}f)(D^x_{i_x,i_z}v) + (D^z_{i_x,i_z}f)(D^x_{i_x,i_z}v)\right], \tag{A.8}$$

where

$$\Psi'_{i_x,i_z} = \Psi'((D^x_{i_x,i_z}f)^2 + (D^z_{i_x,i_z}f)^2). \tag{A.9}$$

Here, the discretized form of $f$ and $v$ can be rearranged in a vector form; $\mathbf{f}$ be vector of $f$ and $\mathbf{v}$ be vector of $v$, corresponding lexicographic ordering of the two-dimensional array components. $D^x$ and $D^z$ is matrix operator corresponding to grid operators and $diag\Psi(\mathbf{f})$ is a $n_x n_z$ x $n_x n_z$ matrix whose diagonal elements are the $\Psi_{i_x,i_z}'s$. The resulting regularization operator is then expressed as

$$\frac{d}{d\tau}J(\mathbf{f} + \tau\mathbf{v})|_{\tau=0} = \left[(D_x\mathbf{v})^T diag(\Psi'(\mathbf{f}))(D_x\mathbf{f}) + (D_z\mathbf{v})^T diag(\Psi'(\mathbf{f}))(D_z\mathbf{f})\right]\triangle x\triangle z \tag{A.10}$$

which is equivalent to

$$\frac{d}{d\tau}J(\mathbf{f} + \tau\mathbf{v})|_{\tau=0} = \left\langle D_x^T diag(\Psi'(\mathbf{f}))(D_x\mathbf{f}), \mathbf{v}\right\rangle + \left\langle D_z^T diag(\Psi'(\mathbf{f}))(D_z\mathbf{f}), \mathbf{v}\right\rangle. \tag{A.11}$$

$\triangle x$ and $\triangle z$ are not included in the above equation since they can be absorbed in the regularization parameter. Note that $D^x = D_x$ and $D^z = D_z$. We know that

$$\frac{d}{d\tau}J(\mathbf{f}) + \tau\mathbf{v}))|_{\tau=0} = \langle J'(\mathbf{f}), \mathbf{v}\rangle, \tag{A.12}$$

Making use of equations [A.11] and [A.12], the gradient is then expressed as

$$gradJ(\mathbf{f}) = J'(\mathbf{f}) = L(\mathbf{f})\mathbf{f}. \tag{A.13}$$

The discretize of the gradient of the regulrization operator, then, turns out to be

$$L(\mathbf{f}) = D_x^T diag(\Psi'(\mathbf{f}))D_x + D_z^T diag(\Psi'(\mathbf{f}))D_z. \tag{A.14}$$

which is equivalent to

$$L(\mathbf{f}) = \begin{bmatrix} D_x^T & D_z^T \end{bmatrix} \begin{bmatrix} diag(\Psi'(\mathbf{f})) & 0 \\ 0 & diag(\Psi'(\mathbf{f})) \end{bmatrix} \begin{bmatrix} D_x \\ D_z \end{bmatrix}. \tag{A.15}$$

This represents the expression of the discretized regularization operator. The gradient of the non-linear regularization operator is linearized.