NOTE TO USERS

This reproduction is the best copy available.

UMI®
RICE UNIVERSITY
Wave Equation Migration Velocity Analysis by Differential Semblance Optimization
by
Peng Shen
A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE DOCTOR OF PHILOSOPHY

APPROVED, THESIS COMMITTEE:

Dr. William W. Symes, Noah Harding Professor, Department of Computational and Applied Mathematics, Rice University

Dr. Alan Levander, Carey Croneis Professor, Department of Earth Science, Rice University

Dr. Scott Morton, Amerada Hess Corp.

Dr. Colin Zelt, Associate Professor, Department of Earth Science, Rice University

Dr. Adrian Lenardic, Assistant Professor, Department of Earth Science, Rice University

September, 2004
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI Microform 3168140
Copyright 2005 by ProQuest Information and Learning Company. All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346
ABSTRACT

Wave Equation Migration Velocity Analysis by Differential Semblance Optimization

by

Peng Shen

Differential semblance measures the deviation from flatness or focus of image gathers. The differential semblance objective function posed on the sub-surface offset domain responds smoothly to velocity changes. Therefore gradient descent methods are uniquely attractive for velocity updating by differential semblance optimization. Because of their kinematic fidelity, wave equation (depth extrapolation) migration methods are natural platforms for velocity analysis in complex structures. The gradient of the objective function with respect to velocity is formulated through the adjoint of differential migration. Limited memory BFGS algorithm is used for the velocity optimization. The method for wave equation velocity analysis developed in this thesis study is applied to both synthetic and real data examples.
Acknowledgments

This thesis is made possible through continuous support from Dr. William W. Symes, my thesis committee chair, and Dr. Scott Morton, my intern advisor at Amerada Hess Corporation. I would like to express my sincere gratitude towards Dr. Symes for his vision, inspiration and the guidance throughout my thesis work. His grasps of mathematical geophysics at its very fundamental level enlightened me enormously.

At the same time, it is difficult to overstate my gratitude to Dr. Morton. He has patiently walked me through the state of the art industrial migration algorithms, and has always been an immediate helper on many computational problems. His deep insight in physics offered me directions in many difficult times.

I am indebted to Dr. Alan Levander. He is one of the best mentors and teachers I could have wished for. I benefit a lot from the friendly and positive working atmosphere as well as numerous inspiring discussions with him.

I owe many thanks to Dr. Collin Zelt. He has been a part of my graduate education as an enthusiastic teacher and constructive advisor. He always has his students’ best
interest in mind.

It is also my sincere gratitude to Dr. Adrain Lenardic. He has been abundantly helpful in supporting me defending my thesis.

In the end, I want to thank my wife Cathy, an angel placed by my side, who has been a constant source of inspiration and encouragement, and witnessed my progress through the years.
Contents

Abstract ii
Acknowledgments iii
List of Figures x
List of Tables xvii

1 Thesis introduction 1

1.1 Thesis structure ............................................ 1

1.2 Separation of scales: a mathematical account of seismic inversion 2

1.3 Why velocity analysis is important .......................... 3

1.4 Historical development of velocity analysis ................ 4

1.4.1 Data domain methods .................................... 5

1.4.2 Image domain methods .................................... 7

1.4.3 Optimization criteria in common image gathers ........ 9

1.4.3.1 Differential semblance criterion ..................... 10

1.4.3.2 Minimum image perturbation criterion ............... 12

1.4.3.3 Relationship between differential semblance criterion and the minimum image perturbation criterion .... 15

1.4.4 Wave equation migration and Kirchhoff migration based velocity analysis ............................................. 16
1.5 Thesis chapter summary ........................................... 18

2 Wave equation migration .......................................... 20
  2.1 Forward propagation ............................................ 21
    2.1.1 Seismic wave propagation .................................. 21
    2.1.2 Forward Born modeling ..................................... 23
  2.2 Migration as the adjoint of Born modeling ................... 27
  2.3 Wave equation migration ........................................ 30
    2.3.1 One-way wave equation ...................................... 33
    2.3.2 Computation of single square root equation ............... 37
    2.3.3 Solution by implicit finite difference scheme .......... 42

3 Velocity analysis by differential semblance optimization .... 45
  3.1 Generalized Born modeling ....................................... 46
  3.2 Adjoint of generalized Born modeling ......................... 48
    3.2.1 Double-square-root wave equation ......................... 48
    3.2.2 Sub-surface offset image gathers .......................... 51
  3.3 Differential semblance criteria .................................. 54
    3.3.1 Offset domain DSO .......................................... 55
    3.3.2 Relationship between $I_h$ and $I_\theta$ .................... 56
    3.3.3 Angle domain DSO ........................................... 60
3.3.4 Relationship between $P_h$ and $P_q$ .......................... 62

3.4 Kinematics of the image in offset .................................. 68

3.4.1 Kinematics of image in offset parameterized by surface offset

$I_h(x, z, h; H)$ ........................................ 68

3.4.2 The envelope of one-parameter family of curves .......... 70

3.4.3 Offset range in relation to surface acquisition limit ....... 71

4 Adjoint state calculation and inversion 74

4.1 Gradient calculation and adjoint state analysis ............... 75

4.1.1 Gradient formulation by the Green's function representation . 75

4.1.2 Gradient formulation by wave equation ....................... 80

4.1.3 Gradient by discrete adjoint state calculation ............. 87

4.1.3.1 The chain rule ........................................ 87

4.1.3.2 Numerical adjoint code construction .................. 90

4.1.3.3 A discrete formulation for differential migration and
its adjoint ..................................................... 96

4.2 Inversion ...................................................... 101

4.2.1 Limited memory BFGS method ............................... 102

4.2.2 Smoothing properties of the gradient ...................... 104

4.2.3 Problem ................................................... 110
5 Data examples

5.1 Offset image gathers due to flat reflectors and constant velocity ... 114
  5.1.1 Synthetic data generation .................................. 114
  5.1.2 Preprocessing .............................................. 116
  5.1.3 Effects of acquisition geometry ............................ 122

5.2 Optimization of low velocity lens model ....................... 126
  5.2.1 Introduction ............................................... 126
  5.2.2 Synthetic data generation and migration .................. 126
  5.2.3 Velocity inversion ......................................... 135

5.3 Optimization of HAFB high resolution data .................... 145
  5.3.1 Introduction ............................................... 145
  5.3.2 Results from previous studies ............................. 146
  5.3.3 Results by differential semblance velocity analysis ...... 150

5.4 Marmousi data set ............................................ 155

5.5 Discussion: problems associated with the model roughness ..... 161

6 Conclusions ......................................................... 167

A PML absorbing boundary condition and its attenuation analysis 169

B Construction of asymptotic solutions .......................... 174
C Migration by an analysis of stationary phase

References
List of Figures

1.1 Left: at the correct velocity the position of the reflector is located at $x$ irrespective to the scattering angle $\theta$ between downward continued source wavefield $S$ and receiver wavefield $R$; right: the constructive interference occurs only at the zero subsurface offset $h$ if the velocity model is correct. ......................................................... 9

3.1 $I_h(x, h)$ is obtained by correlation of $S(x + h, x_s, \omega)$ and $R(x - h, x_s, \omega)$ at each point $x$ in the model. The point of correlation $x$ is not necessarily on the surface. ................................................................. 52

3.2 $\hat{n}$ is migration dip unit vector, and $\theta$ is the scattering angle. The wavenumber of $G^+(x, x_r, \omega)$ points downward. So the wavenumber of $\overline{G^+}(x, x_r, \omega)$ points upward. ......................................................... 59

3.3 Common image gathers in angle are flat with respect to $\theta$. .......... 61

3.4 Left: original signal, right: transformed signal ............................... 66

3.5 Left: original signal with sharp peaks, right: transform signal ........... 66
4.1 The ray path connecting from $x_s$ to $x + h$, $x - h$ to $y$ and $y$ to $x_r$, where $G^+(x, y, \omega)$ is the Green's function observed at $x$ emanated at $y$. The adjoint of differential migration picks up an image residual and smoothly distributed it along the corresponding ray paths.  

4.2 Inversion procedure. Forward project the B-spline model parameters to obtain the current velocity in the image space through $c = Bm$, then evaluate the objective function, calculate the gradient with respect to the velocity in the image space, project this gradient to the model space through the adjoint of the B-spline projection to obtain the gradient with respect to the B-spline model parameters. The search direction is constructed by the L-BFGS method in the model space.  

5.1 Synthetic data generated at flat reflectors and constant velocity.  

5.2 Effects of data preprocessing on images. (a). Images at the correct velocity correspond to preprocessed data using $\rho = 0.9$. (b). Images at the correct velocity correspond to preprocessed data using $\rho = 0.5$.  

5.3 Effects of data preprocessing on offset gathers. (a). Offset gathers at the correct velocity correspond to preprocessed data using $\rho = 0.9$. (b). Offset gathers at the correct velocity correspond to preprocessed data using $\rho = 0.5$.  

108  
112  
115  
118  
119
5.4 Preprocessed data using $\rho = 0.9$ and $\rho = 0.5$ projected back into the $(x, t)$ domain. Two source gathers are shown for sources located in the middle of the model. ........................................ 121

5.5 An “x” pattern in offset gathers obtained at the correct velocity. ... 122

5.6 Offset gathers due to migration of single source gathers at correct velocity. The reflector become a straight line in each offset gather. The phenomenon is closely related to relationship between offset domain image and angle domain image. ................................. 125

5.7 Velocity model of 0.8km by 4km with a low velocity lens embedded in the middle. Sources and receivers are evenly distributed on the surface at the same spatial interval. Six horizontal reflectors are located at equal distance from $z = 0.12$ to $z = 0.78$. ............................... 128

5.8 Source gathers obtained by acoustic simulation using velocity model equation (5.6) with refracting strength $\alpha = 0.4$. ......................... 129

5.9 Source gathers obtained by acoustic simulation using velocity model equation (5.6) with refracting strength $\alpha = 0.6$. Multipathing arrivals start to establish. .................................................. 130

5.10 Source gathers obtained by acoustic simulation using velocity model equation (5.6) with refracting strength $\alpha = 0.8$. Multipathing arrivals are evident. ...................................................... 131
5.11 Synthetic source gathers obtained by acoustic simulation using velocity model equation (5.6) with refracting strength $\alpha = 1.0$ which generates strong multipathing arrivals. Maximum velocity perturbation is 50%. 132

5.12 Images obtained at constant velocity of 2 km/sec. Large imaging errors are evident, in particular for $\alpha = 1$. 133

5.13 Offset gathers obtained at constant velocity of 2 km/sec. Nonzero offset amplitudes are significant near the lens region. 134

5.14 Decay of objective function value and gradient magnitude. Values of objective function and magnitude of gradient at the correct velocity are shown by the dashed lines. Red and blue correspond to the objective function and the magnitude of gradient, respectively. 138

5.15 Offset image gathers at the fifth iteration. Inverted offset gathers are concentrated at the zero offset. 139

5.16 Offset image gathers migrated using the correct velocities. The comparison with Figure (5.15) show little geometrical differences. 140

5.17 Images obtained at the fifth iteration. Reconstructed images are reasonably flattened. 141

5.18 Images migrated using the correct velocities. The comparison with Figure (5.17) shows little differences. 142
5.19 Inverted velocities at the fifth iteration for various lens models. The amplitude and the shape of the lens are well reconstructed.

5.20 The differences between inverted velocities and the correct velocities. The magnitude of the differences is in general within 5% of the background velocity.

5.21 The 3-D reflection seismic experiments made use of 624 Texan portable seismographs, and 2 Geometrix multi-channel seismographs. For sources we used 223 caliber single shot rifles, fired in 6 cm drill-holes. The Texan seismographs were deployed in 6 parallel lines with cross-line separations of 2.1m and with inline geophone spacing of 35 cm. Shots were fired in a rotated staggered brick pattern, with 120 shots/line. Forty-six seismic lines were occupied, producing a survey area of 94.5 by 36.05 m. The yellow line in the channel marks the VSP cross-section.

5.22 Results of previous studies. (a). Ray coverage of surface-bore-hole geometry. (b). Images of 3-D Kirchhoff migration interpolated in the VSP model are plotted on top of the velocity model obtained by surface-bore-hole diffraction tomography. The images are plotted in black.

5.23 Source gathers at $x_s = 12.6$ and $x_s = 15.4$. The data is surface wave dominated.
5.24 Comparison of initial and final images. The deeper structures are better characterized in the final image. .......................................................... 152

5.25 Comparison of initial and final velocity models. The final velocity model is obtained using surface data only. It agrees well with the inverted velocity through diffraction tomography which uses both the surface and the surface-bore-hole data. .............................................. 153

5.26 Offset gathers migrated at initial and final (4th iteration) velocity models. The offset range for each panel is \((h_{\min}, h_{\max}) = (-4m, 4m)\). Better focused offset gathers are obtained at the optimized velocity model. 154

5.27 (a). The Marmousi velocity structure. (b). Image obtained at the correct velocity. ................................................................. 157

5.28 The differential semblance objective function varies in a wide and smooth curve with the minimum centered at the true velocity. ........ 158

5.29 (a). B-spline model space gradient projected back to image space \(B\nabla_m J = BB^*\nabla_c J\). (b). Velocity at the fifth iterate. (c). The best fit projected B-spline velocity \(V_b = Bm_b\), where \(m_b\) solves \(Bm = V\) in a least-squares sense for the true Marmousi velocity \(V\) defined on 921 × 60 image space grid and \(m\) defined on a 6 × 6 B-spline grid. . . . 159
5.30 (a). Iterates of magnitude of gradient and objective function values.

The magnitude of gradient at the true velocity coincide with the objective function value represented in the red dashed line. (b). Scattered image offset gathers at the initial constant velocity. (c). Initial image at constant velocity. (d). Image offset gathers at the fifth iteration shows improvements in focusing. (e) Image at the fifth iteration.

5.31 Image space gradient at constant (1.8km/sec) velocity. (a) The original gradient without smoothing. (b). B-spline smoothed gradient. Figure (5.29(a)) repeated.

5.32 Objective function evaluated at representations of the true velocity from surface to 1.4km in depth with increased roughness. Number of grid points in horizontal direction of the B-spline model increases from 30 to 80 increments by 10.

5.33 Offset gathers become more focused at velocities of higher degree of roughness.

5.34 Increased image quality at velocities of higher degree of roughness.

C.1 The travel time is preserved. Snell’s law applies to both image point and the reflection point. The ray parameters coincide at the surface.
List of Tables

4.1 Forward and adjoint conditional statements. ....................... 92
Chapter 1
Thesis introduction

1.1 Thesis structure

This thesis consists of chapters in logical steps towards explanation of what I did on “wave equation migration velocity analysis by differential semblance optimization”. It is composed of five major components in sequence: one-way wave equation prestack depth migration, construction of the differential semblance objective function, gradient calculation, B-spline smoothing scheme and inversion, and synthetic and real data examples.

The principal problem this thesis addresses is how to jointly optimize the image of reflectivity and the velocity. By stating the problem this way, we have already used the concept of separation of scales to differentiate the image, the singular part of the medium, from the smooth velocity which is at least twice contiguously differentiable.

The solution of the problem requires understanding how to measure the optimality of the image and how to project the image residual back to the velocity model. The answer to the first question brings out the construction of the differential semblance objective function, while the answer to the second is provided in the formulation of the gradient of the objective function. The wave equation prestack depth migration is running as a platform on which both the objective function and the gradient for
velocity updating are evaluated. The construction of the objective function is isolated as a chapter to emphasize its importance as a connecting step between wave equation migration and wave equation migration based velocity analysis. The gradient calculation formulated by one way wave equation migration is the major contribution of my thesis work. Analysis of smoothing properties of the gradient and the B-spline smoothing scheme are also important contributions to the complete solution of the inverse problem. Finally, the algorithm described in the thesis has been tested on synthetic as well as real data sets.

1.2 Separation of scales: a mathematical account of seismic inversion

The main objective of exploration geophysics is to characterize the physical properties of the medium through which seismic waves propagate. Mathematically, the physical properties of the medium are identified as coefficients of wave equation, which can be separated into two parts: part I contains all the singularities of the coefficients; part II is separated from part I by eliminating the singularities and remains as a smooth representation of the medium. The total coefficient is the summation of these two parts. The seismic inverse problem is posed as a combination of two steps correspondingly, reconstruction of the singular part of the coefficient and reconstruction of the
smooth part of the coefficient. The former is referred to as imaging, usually obtained by migration, whereas the latter is called the velocity model in the literature of exploration geophysics. The complete account of singularities is given by the concept of wavefront set which provides a refined characterization of singularities in phase-space and therefore draws the connection to the forward wave propagation by propagation of singularities on bicharacteristic curves. The reconstruction of the singular part of the coefficient is then formally posed as reconstruction of the wavefront set of the coefficient.

1.3 Why velocity analysis is important

Strictly speaking, the wavefront set of the coefficient can not be exactly recovered because of its singular nature. One can only reconstruct its singular behavior asymptotically. In order to do so, knowledge of the rays (bicharacteristic curves) are required so that the phase carried on each ray can meet at correct spatial locations to establish constructive interference. The rays, however, are sensitive to the smooth part of the coefficient of the wave equation which is mentioned before as the velocity model. Velocity analysis is the process of constructing a smooth velocity model which is consistent with the data. Migration, viewed as an operator defined velocity models, will reconstruct the wavefront set of the coefficient asymptotically if the velocity is
accurate. At an inaccurate velocity, the reconstructed wavefront set deviates from the true one. The quality of the imaging result relies on the accurateness of the velocity function.

1.4 Historical development of velocity analysis

Due to the non-linear relationship between the observed seismic data and the velocity model, seismic velocity analysis is usually posed as an optimization problem. Depending on how the objective function is constructed, I will divide methods of velocity analysis into two categories: data domain method in which the objective function is formulated to measure the deviation of the predicted data from the observed data; and image domain method in which the objective function measures the quality of the reconstructed image. The former is posed as an optimization problem minimizing the data residual.

$$\min_c \| P_d(d^{obs}, d^{pre}) \|,$$

likely, the latter is posed as an optimization problem

$$\min_c \| P_I(I) \|$$

where $\| \cdot \|$ indicate a proper norm for both the data domain and image domain.
methods, and $P_d$ and $P_I$ provide required measurement on the data and image, respectively.

### 1.4.1 Data domain methods

Early methods of velocity analysis try to find models that best explain the data and therefore fit in the category of data domain methods. When the observed data is simplified and restricted to arrival times, the associated velocity analysis is often referred to as travel time tomography. Algorithms of travel time tomography are well developed by *Nolet, 1987* [27]. Travel time tomography is usually limited to fitting the first few phases of the observed data neglecting important information carried by later arrivals.

The conventional data domain NMO-DMO stacking velocity analysis accounts for, in principle, all single reflected arrivals (*Claerbout, 1985* [6]). The conversion from the optimum stacking velocity to the interval velocity requires the simplification of the physical reflection model so that the lateral velocity variation is small enough to be neglected. This assumption is clearly is not valid for general earth models.

Gradient based waveform inversions try to fit the observed data by solving wave equations. The data residual is formulated by the linearized wave equation with respect to the velocity. This linearization requires that starting velocity model is close enough
to the true model in the sense that the predicted data are within half wavelength from the observed data for the highest frequency used (Gauthier at al., 1986[12]). The velocity updating is mainly driven by the gradient calculation which is formulated as the adjoint of the Born modeling. The physical meaning of the adjoint of Born modeling is understood as projection of the data residual back to the model space through downward continuation of source and receiver wave-fields. For low frequencies, such projection (gradient only) produces reasonable search directions for velocity updating. For high frequencies, the back projected data residual (assumed to be the Born data) reconstruct the image asymptotically, and thus is equivalent to reverse time prestack depth migration. Data domain velocity analysis usually uses the increasing frequency scheme where the data correspond to lower frequencies are first fitted and then gradually increase to higher frequencies (Pratt, 1999 [28], Pratt, et al., 1996[29]). Time domain implementations of such methods have been presented by Tarantola 1987 [43], Tarantola & Vallette, 1982 [44].

The velocity analysis by waveform inversion will succeed if the starting model is inside the valley of the global minimum. Velocity inversion starting from low frequency data will help this condition to be satisfied. However, the low frequency component is usually absent from the observed seismic data. It therefore makes the method of waveform inversion not practical in many real applications.

Methods of global optimizations are not constrained by the local linearization of the
wave equations and in general will avoid the local minima of the objective function. Fast algorithm of global search, such as genetic algorithms and the method of simulated annealing, have been tested for plane-wave seismograms (Stoffa & Sen, 1991[36], Sen & Stoffa, 1991[34]). However, the computational cost using global optimization methods is prohibitive for large data set with complex velocity models.

1.4.2 Image domain methods

The image relates linearly to the data through migrations. Image domain objective functions using proper semblance principle may suffer less on local minima for the velocity inversion compared with the waveform inversion in data domain. The semblance principle is based on kinematic descriptions of the image at the correct velocity. In principle, if the velocity is correct, then the image migrated from different data bins in the unstacked image volume should be the same, at least kinematically. The flatness semblance principles can be used to measure the kinematics of the image in common bin gathers, or common image gathers (CIGs). In general, the flatness principle in any binning scheme can serve as the criterion for the image domain velocity analysis. The common bin-wise velocity analysis has been reported by several authors: Al-Yahya, 1989[1] in common receiver gathers, Mulder & ten Kroode, 2002[24] and
Liu & Bleistein, 1995[22] in common offset gathers, Brandsbert-Dahl et al., 1999[5] in common scattering angle gathers and Symes & Carazzone, 1991[41] in common shot gathers. It has been reported by Nolan & Symes, 1996[26] that a single event may be imaged along several ray pairs, a phenomenon known as the kinematic artifact, in common shot migration. Kinematic migration artifacts in common bin gathers usually occur in the prestacked image volumes when the medium is strong refracting (Stolk & Symes, 2004[37]). As a result, the flatness principle may not be valid in the common bin migrated image gathers. To avoid the kinematic artifact for general velocity models, migration of full bin data is required to construct common image gathers.

We shall focus on two types of common image gathers which are obtained by migration of full bin data: angle domain common image gathers and subsurface offset domain common image gathers, each has a clear physical meaning through asymptotic analysis (Brandsbert-Dahl et al., 2003[4], Shen P. et al. 2003[35]). When the velocity model is correct, the pairs of rays from sources and receivers should bring constructive interference at the same spatial location regardless of the scattering angles enclosed between them. This gives the criterion of velocity analysis in angle domain: the common image gathers expressed in scattering angle should be flat(Fig (1.1). On the other hand, the criterion in subsurface offset domain takes a different thought, that is if the velocity is correct then the constructive interference should happen only when
the pair of rays from source and receiver intersect. The distance between rays are parameterized by the so called subsurface offset. The criterion for velocity analysis can thus be posed as the requirement that the common image gathers in subsurface offset be concentrated at zero offset (Fig (1.1). I shall emphasize that both image gathers in angle and in offset are obtained by migrating the full data, not by individual bin data.

1.4.3 Optimization criteria in common image gathers

Upon the construction of the common image gathers in angle or in offset we can use semblance or differential semblance criteria to formulate the image domain objective functionals. The conventional stacking power semblance criterion takes the integral of the image in CIGs. When the CIGs are highly non-flat, different events may cross-interface with each other. This will show up as local minima in the objective function,
and therefore unable to make suitable adjustment to the velocity (Claerbout, 1985[6]).

1.4.3.1 Differential semblance criterion

The differential semblance optimization (DSO) first proposed by Symes & Carazzone, 1991[41], avoids severe convergence difficulties associated with waveform least-squares inversion (Symes, 1999[42]). The DSO functional is expected to have much better global convexity properties than the least-squares functional and, therefore, to suffer significantly less from local minima.

In the thesis, I use the concept of the differential semblance to seek for operators $P_h$ or $P_\theta$ that annihilate singularities of image gathers $I_h(x, h)$ or $I_\theta(x, \theta)$ at the correct velocity, where $h$ and $\theta$ is subsurface offset and scattering angle, respectively. At incorrect velocities the application of operator $P$ removes part of the singularities of the image. I will call the resulting image gathers $PI_h$ or $PI_\theta$ the image residual. The velocity updating is driven by minimizing the magnitude of the image residual defined by the objective function of the form

$$J = \frac{1}{2} \|P_h I_h\|^2$$  \hspace{1cm} (1.1)

or
\[ J = \frac{1}{2} \| P_\theta I_\theta \|^2 \]  \hspace{1cm} (1.2)

with a proper norm \( \| \cdot \| \).

In the scattering angle domain, the operator \( P_\theta = \partial / \partial \theta \) or \( P_\theta = \partial / \partial (\tan(\theta)) \) measures, to the leading order, the kinematic non-flatness of the angle image gathers \( I_\theta(x, \theta) \). The image in angle can be directly obtained using an algorithm based on the inverse generalized Radon transform (Brandsbert-Dahl et al., 2003[4]). The objective functional in angle can then be evaluated as

\[ J = \frac{1}{2} \| \frac{\partial}{\partial \theta} I_\theta \|^2 \]  \hspace{1cm} (1.3)

Fast migration algorithms usually do not work directly in angle domain (Berkhout, 1981, 1982[2][3]). Direct computation of image in angle is expensive. The image in angle can be extracted from the image in subsurface offset through the Radon transform (Sava & Fomel, 2000[33]). The objective function using image in offset (subsurface offset) can then be formulated as

\[ J = \frac{1}{2} \| \frac{\partial}{\partial \theta} \mathcal{R}_{h-\theta} I_h \|^2 \]  \hspace{1cm} (1.4)

Here \( \mathcal{R}_{h-\theta} \) is the offset to angle Radon transform.

The operator \( P_h = \frac{\partial}{\partial \theta} \mathcal{R}_{h-\theta} \) is clearly one choice of the differential semblance operator in offset. Another choice of \( P_h \) can be simply taken as \( P_h = h \) In the subsurface offset domain, the operator \( P_h = h \) measures the deviation from the concentration of offset
gathers \( I_h(x, h) \) near \( h = 0 \). The differential semblance operator used in this thesis is simply multiplication by the subsurface offset parameter \( h \) which requires no picking in offset domain image gathers, therefore is particular suitable for automatic velocity analysis. The corresponding objective function using the \( L^2 \) norm is \( J = \frac{1}{2} \| hI \|_2^2 \). This differential semblance objective function varies smoothly with respect to the changes in the velocity model (Stolk \& Symes, 2003 [38]), which can be optimized by gradient type of methods.

1.4.3.2 Minimum image perturbation criterion

The image perturbation \( \delta I(x) \), introduced by Sava \& Biondi, 2003, [31], measures the perturbation of image \( \delta I(x) \) due to perturbations of the velocity model \( \delta c \). The image perturbation, evaluated only in spatial variables \( x \) not in its semblance form, usually contains the singularities of the image. I will give a brief outline of the algorithm developed by Sava \& Biondi, 2003, [31].

An image obtained at the true velocity \( c_t \) is expressed by Taylor expansion to the first order at the current velocity iterate \( c_b \) as

\[
I(x; c_t) = I(x; c_b) + \delta I(x; , c_b) \quad (1.5)
\]

\[
= I(x; c_b) + L\delta c \quad (1.6)
\]
where

$$L = \frac{\partial I}{\partial c}(x; c_b)$$  \hspace{1cm} (1.7)

An linearized inverse problem based on the first order perturbation of image can be formulated as

$$\min_{\delta c} ||\delta I - L\delta c||$$  \hspace{1cm} (1.8)

The major difficulty for minimizing image perturbation is to obtain $\delta I$. A natural candidate for $\delta I$ is taken as $\delta I(x; c_b) = I(x; c_t) - I(x; c_b)$, but this is not obtainable because the true velocity is unknown. Instead the image perturbation will be approximated to the leading order in a differential form as

$$\delta I(x; c_t, c_b) \approx \frac{dI(x, c_t)}{d\rho}|_{\rho=1}\delta \rho$$

$$\approx \left(\frac{\partial I}{\partial k_z} \frac{\partial k_z}{\partial \rho}\right)|_{\rho=1}\delta \rho.$$  \hspace{1cm} (1.9)

where $\delta \rho$ is the perturbation of the velocity ratio to be determined by semblance or differential semblance principle. The image perturbation $\delta I$ in equation (1.9) is evaluated in three steps. First, the image derivative in the Fourier domain, $\partial I/\partial k_z$, is straightforward to compute at $\rho = 1$,

$$\frac{dI}{dk_z}|_{\rho=1} = -izI.$$  \hspace{1cm} (1.10)

Secondly, the derivative of the depth wavenumber with respect to velocity ratio is formulated through the Double-Square-Root equation as, Sava, 2003[32]
\[
\frac{\partial k_z}{\partial \rho} \bigg|_{\rho=1} = \frac{\mu}{2 \sqrt{\mu^2 - |k_s|}} + \frac{\mu^2}{2 \sqrt{\mu^2 - |k_r|}}
\]  
(1.11)

where

\[
\mu^2 = \frac{[4k_{z_0}^2 + (|k_r| - |k_s|)^2][4k_{z_0}^2 + (|k_r| + |k_s|)^2]}{16k_{z_0}^2} 
\]  
(1.12)

\(k_r, k_s\) and \(k_{z_0}\) are spatial wavenumbers for the sources, receivers, and vertical component corresponds to the current background velocity, respectively. Equations (1.10) and (1.11) provide means to calculate \((dI/d\rho)\bigg|_{\rho=1}\). Finally, an optimum velocity ratio \(\rho_\ast\) can be picked, according to semblance or differential semblance criterion, through the repeated residual migration (Sava, 2000[30]) which is suitable only for constant velocity ratio but fast to compute. The perturbation of the velocity ratio is approximated by

\[
\delta \rho \approx \rho_\ast - 1. 
\]  
(1.13)

The image perturbation \(\delta I\) can be then evaluated combining equations (1.10) (1.11) and (1.13). Once the image perturbation is calculated, the velocity perturbation \(\delta c\) can be solved from the linear least-squares problem (equation (1.8)). The velocity model is updated through \(c_b \leftarrow c_b + \delta c\), and the process repeats until a stationary image with respect to the velocity is achieved.

The minimum image perturbation criterion is defined in each linearized step. However, the nonlinear objective function is not clearly defined, therefore it is hard to analyze its convergence properties.
1.4.3.3 Relationship between differential semblance criterion and the minimum image perturbation criterion

Minimization of image perturbation and differential semblance optimization are related. A Taylor expansion of residual image at the current velocity is

\[ PI(c_t) = PI(c_h) + P \frac{\partial I}{\partial c} \big|_{c=c_h} \delta c \]  

(1.14)

The differential semblance optimization criteria implies \( PI(c_t) \approx 0 \), therefore we have

\[ -PI(c_h) = P \frac{\partial I}{\partial c} \big|_{c=c_h} \delta c \]  

(1.15)

which relates to equation (1.8) by replacing the image perturbation by the negative image residual \( PI \) and replacing \( L \) by \( PL \). The image perturbation maintains the singularities of the current image, whereas the residual image tends to remove them. In this thesis I do not use the linearized version of the DSO objective function (equation (1.15)), instead the nonlinear objective function is directly optimized by quasi-Newton algorithms.
1.4.4 Wave equation migration and Kirchhoff migration based velocity analysis

The Kirchhoff migration is a widely used as a platform for velocity analysis. Residual move-outs in common bin gathers always have a ray-geometric interpretation (Meng at al., 1999[23], Liu & Bleistein, 1995[22]). Angle or surface offsets can be picked from common image gathers (Docherty at al., 2000[9]). Search directions of the velocity updating are obtained by projecting residual move-outs alone the ray path through ray tracing. Problems of Kirchhoff migration based velocity analysis are deeply rooted in ray tracing. Kinematic artifacts and lack of ray coverage in low velocity zones are unavoidable problems for both migration and associated velocity analyzed in Kirchhoff methods.

Wave equation migrations provide certain wave-field coverage on low velocity zones. Wave equation migrations using full bin data are free of kinematic artifacts and are therefore ideal platforms for velocity updating. Similar to Kirchhoff migration velocity analysis, wave equation migration velocity analysis also needs to provide a mechanism for projecting image residual back to model space alone the ray path. The ray path is not explicitly known in wave equation approach. In this thesis I define differential migration as the Frechet derivative of the migration operator with respect to the velocity. An important component of this thesis is to demonstrate that the adjoint of the differential migration serves to find the ray paths (in high frequency asympt-
totic) that are implicitly used in wave equation migration. The gradient of image domain objective function is formulated, in closed form, as the adjoint of differential migration as an operator applied to the image residual. It has the clear physical meaning of projecting image residuals smoothly alone the ray paths (section 4.2.2) instead of purely a mathematical treatment for solving inverse problems. When high frequency data are used, this gradient effectively reconstructs ray paths compatible with data and the velocity used in migration, and thus not directly usable for velocity updating because of the non-smoothness introduced by singular behaviors of the ray paths. It should not be surprising to see that the gradient with low frequency data are smoother than with high frequency data.

The central machinery for wave equation migration based velocity analysis is the computation of adjoint of differential migration implemented by extensive use of adjoint state analysis Giering & Kaminski, 1998 [13]. The conservative scheme for wave equation migration preserves the total energy for downward extrapolated wavefields. The adjoint of differential migration using the conservative scheme consists of three parts, the depth reverse wave-field recalculation, adjoint of wave-field perturbations, and the adjoint of velocity perturbations. The computational cost for each gradient using the conservative method is approximately four times of that of migration.
1.5 Thesis chapter summary

Chapter 2. Wave equation migration

This chapter gives an introduction to the basic concept and treatment of wave equation migration which is used here as a platform for velocity analysis. It answers why wave equation migration reconstructs the singularities of the subsurface structure and how an implicit finite difference wave equation migration is implemented.

Chapter 3. Velocity analysis by differential semblance optimization

This chapter mainly discusses the differential semblance minimization criterion. Angle domain and offset domain differential semblance criteria are compared and the relationship between them is explained. I give a construction of the objective function and set up the optimization problem in this chapter.

Chapter 4. Adjoint state calculation and inversion

This chapter answers how the velocity analysis as an optimization problem is solved by gradient type of method. The focus is on the gradient calculation, its formulation, computation, and physical meaning. The inverse problem is solved by L-BFGS method with the aid of B-spline smoothing scheme.

Chapter 5. Data examples

This chapter demonstrates data examples that use the algorithm developed in this thesis. It shows that this algorithm of automatic velocity analysis by differential
semblance optimization works well for complex geology structures in two dimensions.
Chapter 2
Wave equation migration

Chapter synopsis

Prestack depth migration is an important and effective tool in identifying potential targets for oil and gas exploration. The Kirchhoff approach to prestack depth migration relies on the high-frequency asymptotic ray-tracing. Problems of ray tracing occur in caustics, multiple arrivals and shadow zones [10]. It is expected that the wave equation based methods can avoid these difficulties. Although full wave equation methods are able to handle very complex media, such methods are usually very time consuming. One-way methods based on the paraxial wave equation are much more efficient for reasons which will be explained in section (2.3.1). For completeness of this thesis, I shall give an outline on theories and practice of depth migration.
2.1 Forward propagation

Migration reconstructs the singular behavior of the reflectivity distribution from the reflection seismic data which, in turn, can be considered as the response to the singularities of the reflectivity distribution. Therefore it is important to understand the phenomena of forward wave propagation.

2.1.1 Seismic wave propagation

Mechanical waves propagating in elastic medium can be described by

$$\rho \frac{\partial^2}{\partial t^2} u_i(x, t) = f_i(x, t) + \sum_{jkl} (c_{ijkl} u_{k,l})_j (x, t)$$

(2.1)

where $u_i(x, t)$ is the displacement of a particle moving in the direction $i$ at point $x$ and time $t$, $f$ is the force per unit volume, and $c_{ijkl}$ is the elastic moduli. For isotropic material

$$c_{ijkl} = \lambda \delta_{ij} \delta_{jl} + \mu (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk})$$

(2.2)

with $\lambda$ and $\mu$ known as Lame constants. The equation of motion reduces to

$$\rho \frac{\partial^2}{\partial t^2} u_i(x, t) = f_i(x, t) + \sum_j \sigma_{ij,j}(x, t)$$

(2.3)

where
\[
\sigma_{ij} = \lambda \sum_k e_{kk} \delta_{ij} + 2\mu e_{ij} = \lambda \theta \delta_{ij} + 2\mu e_{ij} \quad (2.4)
\]

\[
e_{ij} = \frac{1}{2} (u_{i,j} + u_{j,i}) \quad (2.5)
\]

and

\[
\theta = \sum_k e_{kk} = \nabla \cdot u \quad (2.6)
\]

known as the dilatation, which has physical significance because it gives the change in volume per unit volume associated with the deformation. Equation (2.4) can also be written as

\[
\sigma_{ij} = (\lambda \theta + \frac{2}{3} \mu \theta) \delta_{ij} + 2\mu e_{ij} - \frac{2}{3} \mu \theta \delta_{ij} \quad (2.7)
\]

to separate the diagonal and the off-diagonal elements. We are interested in the case where the stress is lithostatic. The stress can be simply related to the pressure \( p \) through

\[
\sigma_{ij} = -p \delta_{ij} \quad (2.8)
\]

From equation (2.7), it is clear that

\[
-dp = (\lambda + \frac{2}{3} \mu) d\theta \quad (2.9)
\]

The bulk modulus \( \kappa \) is introduced as the ratio of the change of pressure to the change of fractional volume.

\[
\kappa = \frac{-dp}{d\theta} = \lambda + \frac{2}{3} \mu \quad (2.10)
\]
In the same time, the equation of motion for lithostatic pressure is

$$\rho \frac{\partial^2}{\partial t^2} u_i = f_i - p_{ri} \quad (2.11)$$

The pair of equations equation(2.10) and equation(2.11) are usually re-written as

$$\frac{\partial}{\partial t} p = -\kappa \nabla \cdot v \quad (2.12)$$

$$\rho \frac{\partial}{\partial t} v = -\nabla p \quad (2.13)$$

here, $v$ is the particle velocity defined as $v = \frac{\partial}{\partial t} u$. Taking the divergent of equation(2.11) and assume $\rho$ is constant, it follows from equation(2.10) and equation(2.11) that

$$\frac{1}{\kappa/\rho} \frac{\partial^2}{\partial t^2} p - \nabla^2 p = -\nabla \cdot f \quad (2.14)$$

This equation is usually referred as acoustic wave equation of constant density with sound velocity $c = \sqrt{\kappa/\rho}$.

### 2.1.2 Forward Born modeling

The acoustic wave equation with constant density function can be written as

$$\left( \frac{1}{c'(x)} \frac{\partial^2}{\partial t^2} - \nabla^2 \right) p(x, x_s, t) = -\nabla_x \cdot f(x, x_s, t) \quad (2.15)$$

here $f$ is the body force per unit volume, $p$ is the pressure wavefield, and $c'$ is the
sound velocity of the medium. For convenience, let’s assume $-\nabla \cdot f = \delta(x - x_s)\delta(t)$.

The fundamental solution $\tilde{G}'(x, x_s, t)$ for the wavefield observed at $x$ subject to source located at $x_s$ satisfies

$$\left(\frac{1}{c^2(x)} \frac{\partial^2}{\partial t^2} - \nabla_x^2\right) \tilde{G}'(x, x_s, t) = \delta(x - x_s)\delta(t) \quad (2.16)$$

One can write $c'$ as the sum of a reference velocity $c$ which is assumed to be smooth and perturbation velocity $\delta c$,

$$c'(x) = c(x) + \delta c(x) \quad (2.17)$$

where we can make the singular support of perturbation velocity $\delta c$ be equal to that of the total velocity $\text{sing supp}(\delta c) = \text{sing supp}(c')$. Substitute into equation (2.15) and replace $p$ by $\tilde{G} + \delta \tilde{G}$, it yields

$$\left(\frac{1}{(c + \delta c)^2(x)} \frac{\partial^2}{\partial t^2} - \nabla_x^2\right)(\tilde{G} + \delta \tilde{G})(x, x_s, t) = \delta(x - x_s)\delta(t) \quad (2.18)$$

where $\tilde{G}$ is assumed to satisfy

$$\left(\frac{1}{c^2(x)} \frac{\partial^2}{\partial t^2} - \nabla_x^2\right)\tilde{G}(x, x_s, t) = \delta(x - x_s)\delta(t) \quad (2.19)$$

Subtracting equation (2.19) from equation (2.18), and keeping only the first order terms by the Born approximation, we get

$$\left(\frac{1}{c^2(x)} \frac{\partial^2}{\partial t^2} - \nabla_x^2\right)\delta \tilde{G}(x, x_s, t) = \frac{\delta c(x)}{c(x)} \frac{1}{c^2(x)} \frac{\partial^2}{\partial t^2} \tilde{G}(x, x_s, t) \quad (2.20)$$
Using the Green's function representation, we can write the solution evaluated at receiver location \( x_r \) on surface as

\[
\delta \tilde{G}(x_r, x_s, t) = \int 2r(x) \frac{\omega^2}{c^2(x)} \tilde{G}(x, x_s, \omega) \tilde{G}(x, x_r, \omega) e^{i\omega t} \, dx \, d\omega
\]  

(2.21)

Here

\[
r(x) = \delta c(x)/c(x)
\]  

(2.22)

is the reflectivity function. \( \tilde{G}(x, y, \omega) \) solves the two-way Helmholtz equation

\[
(\frac{\omega^2}{c^2(x)} - \nabla_x^2) \tilde{G}(x, y, \omega) = \delta(x - y)
\]  

(2.23)

It propagates to all directions in the medium. If we consider the solution to equation (2.15) as the result of the map

\[
\tilde{G} = F(c)
\]  

(2.24)

then the solution to equation (2.21) is the result of the differential map

\[
\delta \tilde{G} = \frac{\partial F}{\partial c} r
\]  

(2.25)

Let's call \( \frac{\partial F}{\partial c} \) the Born operator which is a Fourier integral operator. Discussion of properties of the Fourier integral operator is beyond the scope of this thesis. Readers should refer to Hörmander [17] for more details. Later in this chapter we shall introduce the adjoint of \( \frac{\partial F}{\partial c} \) defined as the imaging operator. In seismic literature, \( \frac{\partial F^*}{\partial c} \) is usually referred to as migration operator. It reconstructs part of the wavefront set of
the reflectivity distribution $r = \delta c / c$ from the wavefront set of $\delta p$ usually observed on the surface of the earth. To understand how the wavefront set is reconstructed, we need to have certain knowledge of propagation of singularities by differential and pseudo-differential operators.
2.2 Migration as the adjoint of Born modeling

Migration can be formulated as the adjoint of the Born operator. In view of the forward Born mapping $\partial F/\partial c : r \rightarrow \delta p$, $\delta p$ contains the information of the wavefront set of reflectivity distribution. The adjoint operator $(\partial F/\partial c)^* : \delta p \rightarrow \tilde{r}$ can be used to reconstruct the wavefront set of $r$. The wavefront set of $r$ is used as a definition of “image” throughout this thesis. The concept of wavefront set is well explained in Hörmander 1990[16]. Taking the adjoint of the operator in equation (2.21) and applying it to data, we obtain

$$
\tilde{I}(x) = \int 2d(x_r, x_s, t) \frac{\omega^2}{c^2} \overline{G(x, x_s, \omega)} \overline{G(x, x_r, \omega)} e^{-i\omega t} dx_r dx_s dt d\omega
$$

$$
= \int dx_s d\omega \frac{\omega^2}{c^2} \overline{G(x, x_s, \omega)} \int dx_r \overline{G(x, x_r, \omega)} \int dt e^{-i\omega t} d(x_r, x_s, t) (2.26)
$$

The migration procedure described by equation (2.26) can be understood as a 4-step process. First taking the Fourier transform of the data with respect to time $t$, we obtain the data in the frequency domain

$$
d(x_r, x_s, \omega) = \int dt e^{-i\omega t} d(x_r, x_s, t) \quad (2.27)
$$

Second, calculate the adjoint state wavefield

$$
\tilde{R}(x, x_s, \omega) = \int dx_r \overline{G(x, x_r, \omega)} d(x_r, x_s, \omega) \quad (2.28)
$$

where $\tilde{R}$ can be shown to satisfy the two-way Helmholtz equation with velocity $c(x)$. 

Third, propagate the source wavefield complex conjugated

\[ \tilde{S}(x, x_s, \omega) = \bar{G}(x, x_s, \omega) \]  \hspace{1cm} (2.29)

Finally, inverse Fourier transform the product of \( \tilde{S} \) and \( \tilde{R} \) weighted by \( \frac{\omega^2}{c^2} \) at each \( x \) evaluated at time zero and stack for all sources

\[ \tilde{I}(x) = \int dx_s \int d\omega \ e^{i\omega t} \big|_{t=0} \tilde{S}(x, x_s, \omega) \tilde{R}(x, x_s, \omega) \frac{\omega^2}{c^2} \]  \hspace{1cm} (2.30)

Because their associated Green’s functions are two-way Green’s functions, both \( \tilde{R} \) and \( \tilde{S} \) satisfy two-way Helmholtz equations. The adjoint of Born modeling formulated through two-way Green’s functions is called reverse time migration when performed in the time domain. The “reverse time” can be understood in view of equation (2.26) by its convolution representation in time

\[ \tilde{I}(x) = \frac{\partial^2}{\partial t^2} \int \frac{1}{c^2(x)} d(x_r, x_s, t) \tilde{G}(x, x_s, t') \tilde{G}(x, x_r, t - t') dt' dt dx_r dx_s \]  \hspace{1cm} (2.31)

A migration process that calculates step 2 and 3 separately is called shot-record migration. The source and receiver wavefields are propagated in the model through the smooth background velocity \( c(x) \). The objective for migration velocity analysis is to find the optimum \( c(x) \) such that migration gives the best image.

The method of wave equation migration velocity analysis presented in this thesis is under the framework of shot-record migration in which the source and receiver wavefield are calculated separately. Depending on which equation we want to use
to construct the image, equation (2.30) or equation (2.31), the source or receiver wavefield can be solved in time or in frequency.
2.3 Wave equation migration

The migration of seismic data requires solving scalar wave equations in multi-dimensions. Time domain solutions are natural which implicitly solves for all frequencies by stepping through the required time window. The stability condition for explicit finite difference time stepping schemes is assured if the time step $\Delta t$ satisfies $\Delta t \leq \frac{c_1}{c_{\text{max}}} h$ for spatial interval $h$ determined by $h = \frac{c_{\text{min}}}{e_2 f_{\text{max}}}$, where $c_{\text{min}}$, $c_{\text{max}}$ are minimum and maximum velocities, $e_1$ is the coefficient related to finite difference approximations to the Laplacian operator. For 3D 4th order scheme, $e_1$ is taken by 1/2, Wu et al., 1996 [45]. The numerical dispersion relation requires $e_2$ to be at least 5.\textsuperscript{*} Geological velocity models usually exhibit high velocity contrast with $c_{\text{max}}/c_{\text{min}}$ greater than 3. With these estimations made above the number of time steps for a reasonable time window $t = 12\text{sec}$ needs

$$N_t = \frac{tf_{\text{max}}c_{\text{max}}e_2}{c_{\text{min}}e_1} \approx 7200$$

(2.32)

for $f_{\text{max}} \approx 20\text{Hz}$. The computational cost for each time step is $O(N^3)$ for models of size $N^3$. Obviously the total computational cost for all time steps is $N_t O(N^3)$ which is at the order of $10^3 O(N^3)$. The number of time steps at the order of $10^3$ for a reasonable simulation makes the time domain finite difference solution computation-

\textsuperscript{*}The number of grid points per wavelength varies according to different finite difference approximations to the Laplacian operator. A reasonable high order approximation usually requires no less than five grid points per wavelength. The smallest possible number of grid points per wavelength is two for the Nyquist frequency.
ally prohibitive. On the other hand, the number of frequencies needed is much less than the number of time steps. The Nyquist frequency \( f_N \) for time interval \( \Delta t \) is 
\[ f_N = \frac{1}{2\Delta t} \]
. The frequency interval is determined by \( \Delta f = 2f_N/N_t = 1/t \). The number of frequencies \( N_f \) needed to reconstruct the time domain solution is
\[ N_f = f_{\text{max}}/\Delta f = tf_{\text{max}} \]  
(2.33)

which agrees with the number of frequencies used in practice at the order of \( 10^2 \). It follows from equation (2.33) and equation (2.32) that the relation
\[ N_t/N_f = \frac{c_{\text{max}}e_2}{c_{\text{min}}e_1} (\approx 30) \]  
(2.34)
indicates the number of time steps is at one order of magnitude greater than the number of frequencies for reasonable parameter estimations.

The computational cost for solving time domain wave equation is prohibitive in practice. The Helmholtz equation is therefore a substitute for modeling of wave propagation. Write the receiver wavefield as
\[ \tilde{R}(x, x_s, \omega) = \int \tilde{G}(x, x_r; \omega)d(x_r, x_s; \omega)dx_r \]  
(2.35)

which satisfies the full Helmholtz equation:
\[ \left( \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} + \frac{\omega^2}{c^2} \right) \tilde{R}(x_1, x_2, x_3; \omega) = \int \delta(x - x_r)d(x_r, x_s; \omega)dx_r \]  
(2.36)
The full Helmholtz equation (2.36) solved using matrix inverse by Gaussian elimina-
tion needs $O((N^3)^2)$ floating point operations for each frequency, where the model is assumed to be of size $O(N^3)$. Fast algorithms for solving equation (2.36) can be derived using the paraxial approximation, namely the one-way wave equation Claerbout, 1985 [6]. The implicit finite difference solution for a one-way wave equation provides $O(N^3)$ floating point operations per frequency (the computational cost of implicit finite difference method in section (2.3.3)).

Let me give an absolute computational cost analysis comparing the explicit time domain migration and the frequency domain one-way migration for the size of real seismic data set. The size of the velocity model for real data is estimated as $N_x \approx N_y \approx N_z \approx 10^3$. The number of shots $N_{\text{shot}}$ is roughly at the order of $10^5$. For seismograms recorded at 15sec with peak frequency at 25Hz the number of frequency $N_\omega$ used in frequency domain migration is about 250, whereas the number of time steps $N_t$ used in explicit time domain migration is about $10^4$. The industry computational power is assumed to be $10^{12}$flops/sec equivalent to $8.64 \times 10^{16}$flops/day.

**Explicit time domain migration**: In a explicit time-stepping 2-4 finite difference scheme, each time step require roughly $\alpha_t = 10$ visits to each grid points. The total number of floating point operations is estimated as $N_{\text{full}}^{\text{time}} = N_x N_y N_z \alpha_t N_{\text{shot}} N_t \approx 10^{19}$. The computation time is approximately $T \approx 10^{19}/8.64 \times 10^{16} \approx 100$ (days).

**Frequency domain one-way migration**: The implicit finite difference one-way wave equation is solved in complex numbers. In each depth extrapolation step we
can estimate the number of floating point operations to be \( N_{\text{step}} \approx 100N_xN_y \). The total amount of computation is \( N_{\text{freq}}^{\text{one way}} = N_{\text{step}}N_xN_{\text{shot}}N_\omega \approx 2.5 \times 10^{18} \) and the total computation time is estimated as \( T \approx 30 \) days.

The two-way Helmholtz equation can be solved through explicit matrix inverses. The LU decomposition for a square matrix inverse requires in principle \( N^2 \) number of memory units, where \( N \) is the number of unknowns. For a real seismic data set \( N^2 = (N_xN_yN_z)^2 \approx 10^{19} \), which makes the exact solution rarely obtainable in 3D due to the prohibitive computer memory request. This set of comparisons clearly suggest that both the frequency domain two-way migration algorithms and the explicit time domain migration are not suitable for the real seismic data processing. We shall focus in the algorithm for one-way wave equations.

### 2.3.1 One-way wave equation

The key in connection from two-way Helmholtz equation to one-way Helmholtz equation is to realize that the full wavefield Green's function in the Born modeling equation (2.21) is essentially traveling upward or downward if the assumption

\[
|k_1| > \epsilon \sqrt{k_2^2 + k_3^2}, \quad \epsilon \to 0
\]

(2.37)

is valid everywhere for waves propagated in the model, where \( k \in \mathbb{R}^3 \) is the wavenum-
ber for full wavefield Green's functions and $k_1$ is the corresponding vertical wavenumber. Stolk & De Hoop, 2001 [39]. The case when $k_1 = 0$ corresponds to turning waves which propagate horizontally in space. Much of the turning waves are refracted waves. The assumption is usually satisfied for most part of the model by removing refracted arrivals from the data. We shall then limit ourselves to consider the data that is contributed from waves either going up or down but never turns to horizontal. The Green's function in the Born modeling equation (2.21) can therefore be replaced by its one way approximate and yet still explain roughly the same physics under the assumption of $|k_1| > 0$. The one-way Green's function satisfies the one-way Helmholtz equation written as

$$ \frac{\partial}{\partial x_1} G^+(x, y, \omega) - \frac{i \omega}{c} \sqrt{1 + \frac{c^2}{\omega^2} \left( \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} \right)} G^+(x, y, \omega) = \delta(x - y) \quad (2.38) $$

or

$$ \frac{\partial}{\partial x_1} G^-(x, y, \omega) + \frac{i \omega}{c} \sqrt{1 + \frac{c^2}{\omega^2} \left( \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} \right)} G^-(x, y, \omega) = \delta(x - y) \quad (2.39) $$

We use the convention for the Fourier transform kernel $e^{-ikx}$. This means the plane wave decomposition of $G^+$ is expressed in linear combinations of basis function $e^{ikx}$.

Take the vertical wave number to be positive downward, then $G^+$ in equation (2.38) corresponds to downgoing propagation and $G^-$ in equation (2.39) corresponds to up-coming waves. The one-way Helmholtz operator

$$ \frac{\partial}{\partial x_1} \pm iB = \frac{\partial}{\partial x_1} \pm \frac{i \omega}{c} \sqrt{1 + \frac{c^2}{\omega^2} \left( \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} \right)} \quad (2.40) $$
is defined as a pseudo-differential operator of order one

\[
\left( \frac{\partial}{\partial x_1} \pm iB \right) \tilde{p}(x) = \frac{1}{(2\pi)^3} \int \left( -ik_1 \pm i\tilde{b}(x, \omega, k_2, k_3) \right) \tilde{p}(k)e^{ik \cdot x} dk + q(x) \\
= \frac{1}{(2\pi)^3} \int \left( -ik_1 \pm i\tilde{b}(x, \omega, k_2, k_3) \right) \tilde{p}(x')e^{ik \cdot (x-x')} dk dx' + q(x)
\]

(2.41)

for \( q \in C^\infty \) and \( \tilde{p} \in \{ f(x) : \hat{f}(k) = O(|k|^{-N}) \} \) for \( k_1 \leq \epsilon \sqrt{k_2^2 + k_3^2} \) at \( \epsilon \to 0 \) and \( N \) any integer \}. Here \( \tilde{b}(x, \omega, k_2, k_3) \) is the full symbol of the square-root operator

\[
\frac{\omega}{c} \sqrt{1 + \frac{c^2}{\omega^2} \left( \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} \right)}.
\]

The operator (2.40) preserves part of the bicharacteristics of the full Helmholtz equation on which \( |k_1| > \epsilon \sqrt{k_2^2 + k_3^2} \) at \( \epsilon \to 0 \).

The wavefront set of the Born modeling for \( \delta \tilde{G} \) is preserved in its one-way version \( \delta G \) under the assumption of equation(2.37)

\[
\delta G(x_r, x_s, t) = \int 2r(x) \frac{\omega^2}{c^2(x)} G^+(x, x_s, \omega)G^-(x_r, x, \omega)e^{iwt} dx d\omega. \tag{2.42}
\]

A useful relation of one-way Green's functions concerning switching source point \( y \) and observation point \( x \) can be easily shown as

\[
G^-(x, y, \omega) = G^+(y, x, \omega) \tag{2.43}
\]

and is frequently used to write the one-way Born modeling equation(2.42) in the form

\[
\delta G(x_r, x_s, t) = \int 2r \frac{\omega^2}{c^2(x)} G^+(x, x_s, \omega)G^+(x_r, x, \omega)e^{iwt} dx dw. \tag{2.44}
\]
We have thus derived the one-way Born operator. The adjoint of the one-way Born operator applies to data is considered as one-way migration

\[
\tilde{I}(x) = \int \frac{\omega^2}{c^2} G^+(x, x_s, \omega) G^+(x, x_r, \omega) e^{-i\omega t} d(x_r, x_s, t) dx_r dx_s dt d\omega.
\] (2.45)

A complete analysis of this integral requires the concept of the double exploding model \textit{Symes, 2002} [40] and survey sinking \textit{Claerbout, 1985} [6]. We shall delay this discussion to the next chapter where a generalized Born modeling is introduced which is closely related to the double exploding model. The straightforward explanation of one-way migration formulated in equation(2.45) can be pursued in view of shot-record migration. Set

\[
R(x, x_s, \omega) = \int G^+(x, x_r, \omega) d(x_r, x_s, \omega) dx_r
\] (2.46)
as the downward continued receiver wavefield and \( S(x, x_s, \omega) = \overline{G^+(x, x_s, \omega)} \) as the downward continued source wavefield complex conjugated. The wavefields \( R \) and \( S \) differ from their two-way versions \( \tilde{R} \) and \( \tilde{S} \) not only in the directions they travel but also in amplitudes. The one-way wavefields preserve the downgoing kinematics of the two-way wavefields. Using one-way wave equation(2.38) taking complex conjugate, we have

\[
\left( \frac{\partial}{\partial x_1} + iB \right) G^+(x, y, \omega) = \delta(x - y)
\] (2.47)

Notice the vertical coordinates of \( x_s \) and \( x_r \) are zero. We apply operator \( \frac{\partial}{\partial x_1} + iB \) to \( R \) and \( S \), obtain
\[
\left( \frac{\partial}{\partial x_1} + iB \right) R(x, x_s, \omega) = \int \left( \frac{\partial}{\partial x_1} + iB \right) G^+(x, x_r, \omega) d(x_r, x_s, \omega) dx_r \\
= \int \delta(x - x_r) d(x_r, x_s, \omega) dx_r \\
= \delta(x_1) d(x, x_s, \omega) 
\]

(2.48)

\[
\left( \frac{\partial}{\partial x_1} + iB \right) S = \left( \frac{\partial}{\partial x_1} - iB \right) G^+(x, x_s, \omega) \\
= \delta(x - x_s) 
\]

(2.49)

The downward continued receiver and complex conjugated source wavefields are modeled by the same one-way wave equation of single square root but using different source terms. We are now ready to show algorithms for solving such single square root equations deployed in this thesis.

2.3.2 Computation of single square root equation

The solution to the one-way Helmholtz equation (taking minus sign only)

\[
\left( \frac{\partial}{\partial x_1} - iB \right) \tilde{p} = 0 
\]

(2.50)

is achieved through several steps of approximations. First we will ignore the smooth function \( q(x) \) in equation (2.41).

\[
\left( \frac{\partial}{\partial x_1} - iB \right) \tilde{p}(x) \approx \frac{1}{(2\pi)^3} \int (-ik_1 - i\tilde{b}(x, \omega, k_2, k_3)) \tilde{p}(x') e^{ik(x-x')} dk dx' 
\]

(2.51)

By ignoring the smooth function \( q \), we keep the wavefront set of \( \tilde{p} \) unchanged. Second,
the full symbol $\tilde{b}(x, \omega, k_2, k_3)$ of the square root operator $B$ is often approximated by its principal symbol

$$b(x, \omega, k_1, k_2) = \sigma_1(\tilde{b}) = \frac{w}{c(x)} \sqrt{1 - \frac{c^2(x)}{\omega^2} (k_2^2 + k_3^2)}.$$ (2.52)

The symbol of the difference $\tilde{b} - b$ corresponds to a pseudo-differential operator of order zero and is often ignored due to asymptotic arguments. The one-way Helmholtz equation is further approximated as

$$(\frac{\partial}{\partial x_1} - iB)\tilde{p}(x) \approx \frac{1}{(2\pi)^3} \int (-i)k_1 - i \frac{w}{c(x)} \sqrt{1 - \frac{c^2(x)}{\omega^2} (k_2^2 + k_3^2)} \tilde{p}(x') e^{ik(x-x')} dk dx'.$$ (2.53)

Introduce $p(x)$ that satisfies the following equation

$$\frac{\partial}{\partial x_1} p(x) = \frac{1}{(2\pi)^3} \int i \frac{w}{c(x)} \sqrt{1 - \frac{c^2(x)}{\omega^2} (k_2^2 + k_3^2)} p(x') e^{ik(x-x')} dk dx'.$$ (2.54)

The wavefield $p$ agrees with $\tilde{p}$ in its leading order and preserves its wavefront set. The equation (2.54) is what we solve as a substitute for the one-way Helmholtz equation.

For convenience, we will denote $\hat{p}(k)$ as the multi-dimensional Fourier transform

$\hat{p}(k) = \int p(x) e^{-ikx} dk$, and $\hat{p}(x_1, k_2, k_3)$ as a partial inverse Fourier transform $\hat{p}(x_1, k_2, k_3) = \frac{1}{2\pi} \int \hat{p}(k) e^{ik_1 x_1} dk_1$. Taking the partial inverse Fourier transform with respect to $k_2$ and $k_3$, we obtain

$$\frac{\partial}{\partial x_1} p(x) = \frac{1}{(2\pi)^2} \int i \frac{w}{c(x)} \sqrt{1 - \frac{c^2(x)}{\omega^2} (k_2^2 + k_3^2)} \hat{p}(x_1, k_2, k_3) e^{i(k_2 x_2 + k_3 x_3)} dk_2 dk_3.$$
\begin{equation}
\frac{1}{(2\pi)^2} \int i b(x, \omega, k_2, k_3) \hat{p}(x_1, k_2, k_3)e^{i(k_2 x_2 + k_3 x_3)} \, dk_2 dk_3
\end{equation}

Equation (2.55) has an solution in the integral form

\begin{equation}
p(x_1, x_2, x_3) = \int \hat{p}(x_0, k_2, k_3)e^{i x_0 b(x_1', x_2, x_3, \omega, k_2, k_3)}dx_1'e^{i(k_2 x_2 + k_3 x_3)} \, dk_2 dk_3
\end{equation}

which can be approximated in the finite difference limit \( x_1 - x_0 = \Delta x_1 \to 0 \),

\begin{equation}
p(x_0 + \Delta x_1, x_2, x_3) = \frac{1}{(2\pi)^2} \int \hat{p}(x_0, k_2, k_3)e^{i b \Delta x_1}e^{i(k_2 x_2 + k_3 x_3)} \, dk_2 dk_3
\end{equation}

Although \( \hat{p}(x_0, k_2, k_3) \) can be easily obtained by the Fast Fourier Transform, it is difficult to compute at each \((x_2, x_3)\) the wavenumbers \((k_2, k_3)\) and therefore \(b = b(x, \omega, k_2, k_3)\) are hard to evaluate. Alternative methods are developed to transform \(e^{i(k_2 x_2 + k_3 x_3)}\) into derivatives in \(x_2\) and \(x_3\). To do so, we need rational expansion of the symbol \(b\). The rational expansion of the symbol \(b\) through an series of polynomial ratios originates from a continued-fraction expansion \textit{Claerbout, 1985} [6]. The polynomial coefficients can be further optimized for large propagation angles \textit{Lee & Suh, 1981} [20]. Let

\begin{align*}
S_2 &= -\frac{c^2(x)}{\omega^2} k_2^2 \\
S_3 &= -\frac{c^2(x)}{\omega^2} k_3^2 \\
S &= S_2 + S_3 = -\frac{c^2(x)}{\omega^2}(k_2^2 + k_3^2)
\end{align*}

we have
\[ \frac{i}{c} \left\{ 1 + \alpha_1 \frac{S}{1 + \beta_1 S} \right\} \approx \frac{i\omega}{c} \left\{ 1 + \sum_{i=1}^{m} \frac{\alpha_i S}{1 + \beta_i S} \right\} \] (2.58)

where \( \alpha_i \) and \( \beta_i \) are coefficients derived by Lee and Suh [20]. Commonly, one term of the series expansion \( (m = 1) \) is accurate for propagation angles up to 75°. We shall use \( m = 1 \) for the rest of the discussion. Approximation (2.58) can be further derived for \( m = 1 \)

\[
ib \approx \frac{i\omega}{c} \left\{ 1 + \alpha_1 \frac{S}{1 + \beta_1 S} \right\}
= \frac{i\omega}{c} \left\{ 1 + \frac{\alpha_1 S_2}{1 + \beta_1 S_2} + \frac{\alpha_1 S_3}{1 + \beta_1 S_3} - \frac{\alpha_1 \beta_1 S_2 S_3 - \alpha_1 \beta_1^2 (s_2 s_2 s_3 + s_2 s_3 s_3)}{(1 + \beta_1 S_2)(1 + \beta_1 S_3)} \right\}
\]

and dropping the cross-product terms *

\[
ib \approx \frac{i\omega}{c} \left\{ 1 + \frac{\alpha_1 S_2}{1 + \beta_1 S_2} + \frac{\alpha_1 S_3}{1 + \beta_1 S_3} \right\} \]
(2.59)

and substitute into equation (2.57), it yields

\[
p(x_0 + \Delta x_1, x_2, x_3) = \frac{1}{(2\pi)^2} \int \hat{p}(x_0, k_2, k_3) e^{i\omega(1 + \frac{\alpha_1 S_2}{1 + \beta_1 S_2} + \frac{\alpha_1 S_3}{1 + \beta_1 S_3})} d k_2 d k_3
\]

* 

\[
\frac{\alpha_1 S}{1 + \beta_1 S} = \frac{\alpha_1 (s_2 + s_3)(1 + \beta_1 S_2)(1 + \beta_1 S_3)}{(1 + \beta_1 S)(1 + \beta_1 S_2)(1 + \beta_1 S_3)}
= \frac{\alpha_1 S_2(1 + \beta_1 S)(1 + \beta_1 S_3) + \alpha_1 S_3(1 + \beta_1 S)(1 + \beta_1 S_2) - \alpha_1 \beta_1^2 (s_2 s_3 + s_2 s_3)}{(1 + \beta_1 S)(1 + \beta_1 S_2)(1 + \beta_1 S_3)}
= \frac{\alpha_1 S_2(1 + \beta_1 S_2) + \alpha_1 S_3(1 + \beta_1 S_3) - \alpha_1 \beta_1 S_2 S_3 - \alpha_1 \beta_1^2 (s_2 s_2 s_3 + s_2 s_3 s_3)}{(1 + \beta_1 S)(1 + \beta_1 S_2)(1 + \beta_1 S_3)}
= \frac{\alpha_1 S_2}{1 + \beta_1 S_2} + \frac{\alpha_1 S_3}{1 + \beta_1 S_3} + O(s_2)O(s_3) + O(s_2^2)O(s_3) + O(s_3)O(s_2^2)
\approx \frac{\alpha_1 S_2}{1 + \beta_1 S_2} + \frac{\alpha_1 S_3}{1 + \beta_1 S_3}
\]
\[
\begin{align*}
&= \frac{1}{(2\pi)^2} \int \tilde{p}(x_0, k_2, k_3) e^{i\omega (\Delta x_{11} + f(k_2) \Delta x_{12} + f(k_3) \Delta x_{13})} e^{i(k_2 x_2 + k_3 x_3)} dk_2 dk_3 \\
\text{where } f(k_2) &= -\frac{\alpha_1 (c^2 / \omega^2) k_2^3}{1 - \beta_1 (c^2 / \omega^2) k_2^3}, \quad f(k_3) = -\frac{\alpha_1 (c^2 / \omega^2) k_3^3}{1 - \beta_1 (c^2 / \omega^2) k_3^3} \text{ and } \Delta x_{11} = \Delta x_{12} = \Delta x_{13} = \Delta x_1. \text{ We solve } p(x_0 + \Delta x_1, x_2, x_3) \text{ in three steps. Introduce} \\
p^1 &= \frac{1}{(2\pi)^2} \int \tilde{p}(x_0, k_2, k_3) e^{i\omega \Delta x_{11}} e^{i(k_2 x_2 + k_3 x_3)} dk_2 dk_3 \\
p^2 &= \frac{1}{(2\pi)^2} \int \tilde{p}(x_0, k_2, k_3) e^{i\omega (\Delta x_{11} + f(k_2) \Delta x_{12})} e^{i(k_2 x_2 + k_3 x_3)} dk_2 dk_3 \\
p^3 &= \frac{1}{(2\pi)^2} \int \tilde{p}(x_0, k_2, k_3) e^{i\omega (\Delta x_{11} + f(k_2) \Delta x_{12} + f(k_3) \Delta x_{13})} e^{i(k_2 x_2 + k_3 x_3)} dk_2 dk_3
\end{align*}
\]

(2.60) \quad (2.61) \quad (2.62)

Obviously, \( p^3 \) is what we want to solve,

\[p^3 = p(x_1 + \Delta x_1, x_2, x_3)\]

The intermediate wavefields \( p^1, p^2, \) and \( p^3 \) have finite difference relations

\[
\frac{\partial p^2}{\partial (\Delta x_{12})} = \lim_{\Delta x_1 \to 0} \frac{p^2 - p^1}{\Delta x_1} \quad (2.63)
\]

\[
\frac{\partial p^3}{\partial (\Delta x_{13})} = \lim_{\Delta x_1 \to 0} \frac{p^3 - p^2}{\Delta x_1} \quad (2.64)
\]

This can be verified by direct substitution. Take equation (2.63) as an example

\[
\begin{align*}
\lim_{\Delta x_1 \to 0} \frac{p^2 - p^1}{\Delta x_1} &= \lim_{\Delta x_1 \to 0} \frac{1}{(2\pi)^2} \int \tilde{p}(x_0, k_2, k_3) e^{i\omega \Delta x_{11}} e^{i(k_2 x_2 + k_3 x_3)} e^{i\omega f(k_2) \Delta x_1} - 1 \frac{\partial p^2}{\partial (\Delta x_{12})} dk_2 dk_3 \\
&= \lim_{\Delta x_1 \to 0} \frac{1}{(2\pi)^2} \int \tilde{p}(x_0, k_2, k_3) \frac{i\omega}{c} f(k_2) e^{i\omega \Delta x_{11}} e^{i(k_2 x_2 + k_3 x_3)} dk_2 dk_3 \\
&= \frac{\partial p^2}{\partial (\Delta x_{12})}
\end{align*}
\]
Equation (2.64) can be similarly verified. The thin lens term $p^1$ can be evaluated straightforward from integral equation (2.60)

$$p^1 = p(x_0, x_2, x_3)e^{i\omega \Delta x_1}$$  \hspace{1cm} (2.65)

The diffraction terms, $p^2$ and $p^3$, can be shown to satisfy the differential equations

$$\frac{\partial}{\partial(\Delta x_{12})}p^2 + \beta_1 \frac{c^2}{\omega^2} \frac{\partial^2}{\partial x_2^2}(\frac{\partial}{\partial(\Delta x_{12})})p^2 = \alpha_1 \frac{c^2}{\omega^2} \frac{i\omega}{c} \frac{\partial}{\partial(\Delta x_{12})^2}p^2$$  \hspace{1cm} (2.66)

$$\frac{\partial}{\partial(\Delta x_{13})}p^3 + \beta_1 \frac{c^2}{\omega^2} \frac{\partial^2}{\partial x_3^2}(\frac{\partial}{\partial(\Delta x_{13})})p^3 = \alpha_1 \frac{c^2}{\omega^2} \frac{i\omega}{c} \frac{\partial}{\partial(\Delta x_{13})^2}p^3$$  \hspace{1cm} (2.67)

Combining equations (2.65), (2.66), (2.67), (2.63), and (2.64) an implicit finite difference scheme can be used to solve for $p(x_0 + \Delta x_1, x_2, x_3)$.

2.3.3 Solution by implicit finite difference scheme

We consider depth extrapolation in a thin screen with thickness $\Delta x_1$. We want to calculate $p^3 = p(x_0 + \Delta x_1, x_2, x_3)$ from $p^0 = p(x_0, x_2, x_3)$. According to equation (2.65) $p^1$ can be implemented as

$$p^1 = p^0 e^{i\omega \Delta x_1}$$  \hspace{1cm} (2.68)

The Crank-Nicholson method is used to discretize equation (2.66) on the right hand side

$$\frac{1}{\Delta x_1}\{p^2 - p^1 + \beta_1 (c^2/\omega^2)[\frac{\partial^2}{\partial x_2^2}p^2 - \frac{\partial^2}{\partial x_2^2}p^1]\}$$
\[
\frac{\partial^2}{\partial x_2^2} = \frac{\delta_{x_2}^2}{1 + \lambda \Delta x_2^2 \delta_{x_2}^2}
\]  
(2.70)

where
\[
\delta_{x_2}^2 u = \frac{u(x_2 + \Delta x_2) - 2u(x_2) + u(x_2 - \Delta x_2)}{\Delta x_2^2}
\]  
(2.71)

is the central-difference operator, and \(\lambda = 1/12\) is the compact coefficient which increases the finite-difference accuracy. Thus this finite-difference scheme becomes
\[
\frac{1}{\Delta x_1} \left\{ p^2 - p^1 + (\lambda \Delta x_2^2 + \beta_1 (c^2 / \omega^2)) [\delta_{x_2}^2 p^2 - \delta_{x_2}^2 p^1] \right\}
\]
\[
= \frac{1}{2} [\alpha_1 (c^2 / \omega^2) \frac{i \omega}{c} \delta_{x_2}^2 p^2] + \frac{1}{2} [\alpha_1 (c^2 / \omega^2) \frac{i \omega}{c} \delta_{x_2}^2 p^1]
\]  
(2.72)

Denote \(p_{i,j}^2 = p^2(i \Delta x_2, j \Delta x_3)\) discretized with interval \(\Delta x_2\) and \(\Delta x_3\) in \(x_2\) and \(x_3\) directions, we have for \(k = 1, 2\)
\[
p_{i,j}^{k+1} + (\lambda + \beta_1 \frac{c^2}{\omega^2 \Delta x_2^2} - \frac{1}{2} \alpha_1 \frac{c}{\omega} \frac{\Delta x_1}{\Delta x_2^2}) [p_{i-1,j}^{k+1} - 2p_{i,j}^{k+1} + p_{i+1,j}^{k+1}]
\]
\[
= p_{i,j}^{k} + (\lambda + \beta_1 \frac{c^2}{\omega^2 \Delta x_2^2} - \frac{1}{2} \alpha_1 \frac{c}{\omega} \frac{\Delta x_1}{\Delta x_2^2}) [p_{i-1,j}^{k} - 2p_{i,j}^{k} + p_{i+1,j}^{k}]
\]  
(2.73)

This can be rewritten for \(k = 1\) as
\[
a_i p_{i-1,j}^2 + b_i p_{i,j}^2 + c_i p_{i+1,j}^2 = a_i^* p_{i-1,j}^1 + b_i^* p_{i,j}^1 + c_i^* p_{i+1,j}^1
\]  
(2.74)
where

\[ a_i = \lambda + \frac{\beta_1 c^2}{\Delta x_2^2 \omega^2} - \frac{1}{2} \frac{\alpha_1}{\Delta x_2^2} \frac{c}{\omega} \]

\[ a_i^* = \text{complex conjugate of } a_i \]

\[ b_i = 1 - 2a_i \]

\[ b_i^* = \text{complex conjugate of } b_i \]

\[ c_i = a_i \]

With the set of coefficients as above, the right hand side of equation (2.74) is viewed as a tridiagonal evaluation. A tridiagonal solver can be used to solve for the vector of \( p^2 \) on the left hand side of equation (2.74). With the same set of coefficients as above except changing \( \Delta x_2^2 \) to \( \Delta x_3^2 \), the equation in \( x_3 \) direction can be solved similarly

\[ a_i p_{i,j-1}^3 + b_i p_{i,j}^3 + c_i p_{i,j+1}^3 = a_i^* p_{i,j-1,j}^2 + b_i^* p_{i,j}^2 + a_i^* p_{i,j+1,j}^2 \quad (2.75) \]

For both the tridiagonal evaluation and tridiagonal solver the computational cost is approximately \( 3N \) on \( x_2 \) and \( x_3 \) for the model of size \( N^3 \). It justified the statement in the beginning of the section (2.3) that the computational cost of solving one-way Helmholtz equation is \( O(N^3) \).
Chapter 3
Velocity analysis by differential semblance optimization

Chapter synopsis

Seismic waves propagate along the bicharacteristic curves (rays) before they meet the reflector. Discontinuities in the reflectivity determines where the reflection happens as well as which pairs of incident and reflected rays are responsible for the arrival. Migration reconstructs discontinuities in the reflectivity by migrating singularities of reflection signals along the rays path and bring constructive interference at correct position. The ray paths are sensitive strong change to the velocity. At wrong velocities the constructive interference of migration occurs at the wrong position due to incorrect ray paths. Therefore, the reconstructed reflectivity by the wrong velocity deviates from the true reflectivity. How to measure this deviation is the main question we pursue to answer in this chapter.

An image residual will be defined in section (3.2.2). Project this image residual back into the velocity model is the main computational machinery needed for velocity analysis. The back-projection of the image residual is formulated as the gradient of the objective function which we shall introduce in this chapter. Analyzing the smoothing properties of the gradient is important for the purpose of velocity updat-
ing and will be discussed in section 4.2.2.

The reconstructed reflectivity (image) obtained by migration deviates from the true reflectivity. This deviation can be measured in common image gathers (CIGs) where the image is expanded using one additional variable. We will discuss two major types of CIGs, one in the offset domain, the other in the angle domain. Differential semblance criteria based on local properties of CIGs are used for automatic velocity updating.

3.1 Generalized Born modeling

In order to study common image gathers at incorrect velocities, we first generalize the Born modeling so that it invokes the incorrect background velocity in the formulation. Introduce the generalized reflectivity distribution, \( r_g(x, y) \) \( x, y \in \mathbb{R}^3 \), such that it satisfies

\[
\delta G(x_r, x_s, t) = \int 2r_g(x, y) \frac{\omega^2}{c(x)^2} G^+(x, x_s, \omega) G^+(y, x_r, \omega) e^{i\omega t} dx dy d\omega
\]

where \( c \) is not necessarily the correct background velocity. The Born wavefield \( \delta G \) evaluated at an incorrect velocity indicates that there can be a pair of a wrong velocity \( c \) and a generalized reflectivity distribution \( r_g(x, y) \) such that the observed data can be explained by a general reflection due to \( r_g(x, y) \) just as well as the physical reflection
due to $r_g = r(x)\delta(x - y)$. Notice equation (3.1) is reduced to equation (2.44) if $r_g(x, y) = r(x)\delta(x - y)$ for $r(x)$ the physical reflectivity distribution and the velocity $c$ is correct.

$$\delta G(x_r, x_s, t) = \int \frac{2r(x)\delta(x - y)}{c^2} \omega^2 G^+(x, x_s, \omega) G^+(y, x_r, \omega) e^{i\omega t} dxdy d\omega$$

$$= \int \frac{2r(x)\omega^2}{c^2} G^+(x, x_s, \omega) G^+(x, x_r, \omega) e^{i\omega t} dxdy d\omega.$$  

It is clear that a physical reflection corresponds to the pair of the correct velocity and the physical reflectivity distribution $r(x)$, which is a special case of equation (3.1).

The generalized Born modeling can be understood from a point of view of the double exploding model, Symes, 2002 [40]. Using the relation $G^+(x, y, \omega) = G^-(y, x, \omega)$, equation (3.1) is expressed in terms of $G^-$ as

$$\delta G(x_r, x_s, t) = \int 2r_g(x, y)\frac{\omega^2}{c^2} G^-(x_s, x, \omega) G^-(x_r, y, \omega) e^{i\omega t} dxdy d\omega$$  

(3.2)

which characterize the same data due to wave propagation as if there are two explosions of sources $r_g(x, y)\delta(x)$ and $r_g(x, y)\delta(y)$, each of which probates one-way waves up to the surface.

The above can be viewed as the generalized born operator, $F_g : r_g \mapsto \delta p$. The adjoint operator of $F_g$ will map the data to singularities of the generalized reflectivity, $F^*_g : \delta p \mapsto r^*_g$.  


3.2 Adjoint of generalized Born modeling

Application of operator $F_g^*$ to data can be read from equation (3.1) as

$$r_g^*(x, y) = \int d(x_r, x_s, t) \frac{2\omega^2}{c(x)^2} \overline{G^+(x, x_s, \omega)G^+(y, x_r, \omega)e^{-i\omega t}} dx_s dx_r dt d\omega \quad (3.3)$$

In this section, we shall discuss the calculation of $r_g^*$ through double-square-root wave equation. It then follows naturally that the solution of double-square-root migration are equivalent to that of shot-record migration. Applying a change of variables on $r_g^*$, we introduce subsurface offset image gathers on which the offset domain differential semblance optimization is posed.

3.2.1 Double-square-root wave equation

Recall the downward continued source and receiver wavefields are defined, respectively, as

$$S(x, x_s, \omega) = \overline{G^+(x, x_s, \omega)}$$

$$R(x, x_s, \omega) = \int \overline{G^+(x, x_r, \omega)}d(x_r, x_s, t)e^{i\omega t} dt dx_r$$

$$= \int \overline{G^+(x, x_r, \omega)}d(x_r, x_s, \omega)dx_r$$

The correlation product of source and receiver wavefields can be written as
\[ u(x, y, x_s, \omega) = S(x, x_s, \omega)R(y, x_s, \omega) \]
\[ = \overline{G^+(x, x_s, \omega)} \int \overline{G^+(y, x_r, \omega)}d(x_r, x_s, \omega)dx_r, \quad (3.4) \]

and we also introduce \( U(x, y, \omega) \)

\[ U(x, y, \omega) = \int u(x, y, x_s, \omega)dx_s. \quad (3.5) \]

When \( x \) and \( y \) are restricted to have the same vertical coordinates \( x_1 = y_1 \), \( U(x, y, \omega) \)
can be shown to satisfy the double-square-root wave equation. Introduce a notation of the double-square-root operator

\[ L^+(x, y, -i\partial_x, -i\partial_y, \omega) = \frac{\partial}{\partial x_1} - iB(x, -i\partial_x, \omega) - iB(y, -i\partial_y, \omega). \]

Applying \( L^+ \) to \( u \) and using equations (2.48) and (2.49), we have

\[ L^+u = \left[(\frac{\partial}{\partial x_1} - iB(x, -i\partial_x, \omega))\overline{G^+}\right]R + \overline{G^+}\left[(\frac{\partial}{\partial x_1} - iB(y, -i\partial_y, \omega))R\right] \]
\[ = \delta(x - x_s)R(y, x_s, \omega) + \overline{G^+(x, x_s, \omega)} \int \delta(y - x_r)d(x_r, x_s, \omega)dx_r \]

Using notations

\[ x = (x_1, x_{2,3}), \quad y = (y_1, y_{2,3}) \]
\[ x' = (0, x_{2,3}), \quad y' = (0, y_{2,3}), \quad x_{2,3}, y_{2,3} \in \mathbb{R}^2, \]

and noticing that \( x_s \) and \( x_r \) are essentially surface coordinates

\[ x_s = (0, x_{s2,3}), x_r = (0, x_{r2,3}), \quad x_{s2,3}, x_{r2,3} \in \mathbb{R}^2 \]
we can further derive equation (3.6) as

\[
L^+ u = \delta(x_1)\delta(x' - x_s)R(y', x_s, \omega) + \overline{G^+(x', x_s, \omega)}\delta(y_1)d(y', x_s, \omega)
\]

\[
= \delta(x_1)\delta(x' - x_s)d(y', x_s, \omega) + \delta(x' - x_s)\delta(y_1)d(y', x_s, \omega)
\]  \hspace{1cm} (3.6)

In the second "=" we have used the relations of source and receiver wavefields at the surface

\[
R(y', x_s, \omega) = d(y', x_s, \omega)
\]

\[
G^+(x', x_s, \omega) = \delta(x' - x_s)
\]

The double-square-root equation can be then written as

\[
L^+ U = \int L^+ u dx_s
\]

\[
= \int [\delta(x_1)\delta(x' - x_s)d(y', x_s, \omega) + \delta(x' - x_s)\delta(y_1)d(y', x_s, \omega)]dx_s
\]

\[
= \delta(x_1)d(y', x', \omega) + \delta(y_1)d(y', x', \omega)
\]

\[
= 2\delta(x_1)d(y', x', \omega)
\]  \hspace{1cm} (3.7)

where the last "=" is valid because \( x_1 = y_1 \). The surface coordinates \( y' \) and \( x' \) in \( d(y', x', \omega) \) corresponds to the source and receiver locations spanning the entire surface. Equation (3.7) is understood as modeling survey sinking through the double-square-root continuation \textit{Claerbout, 1985} [6]. Written with complete dependencies, the equation

\[
\left( \frac{\partial}{\partial x_1} - iB(x, -i\partial_x, \omega) - iB(y, -i\partial_y, \omega) \right)U(x, y, \omega) = 2\delta(y_1)d(y, x, \omega)
\]  \hspace{1cm} (3.8)
is known as the double-square-root wave equation. It is clear from the construction of $U(x, y, \omega) = \int S(x, x_s, \omega)R(y, x_s, \omega)dx_s$ that it can be solved also by shot-record migration approach, where source and receiver wavefields are downward continued separately to $x$ and $y$ with the restriction $x_1 = y_1$ and then correlate. The adjoint of generalized Born modeling in this thesis is implemented through the shot-record migration approach.

### 3.2.2 Sub-surface offset image gathers

Applying a change of variable $h = (y - x)/2$ to equation (3.3), we can write the generalized reflectivity distribution to the symmetric form

$$
\mathcal{r}_g^* (x - h, x + h) = \int 2d(x_r, x_s, \omega) \frac{\omega^2}{c^2} G^+ (x + h, x_s, \omega)G^+ (x_r, x - h, \omega) e^{i\omega t} dx_r dx_s d\omega d\omega
$$

(3.9)

where $h$ is understood as the sub-surface offset. In view of shot-record migration, $|h|$ is half of the horizontal distance between the points at which the downward continued source and receiver wavefields are evaluated. The equation (3.9) can be explained as the reflection signals received at $x - h$ due to a source located at $x + h$ through the non-physical Born propagation at some wrong velocity. Obviously, no signals can be received at non-zero offset due to physical propagations. Neglecting the weight factor $\frac{\omega^2}{c^2}$ in equation(3.3), the singularities of $\mathcal{r}_g^*$ are preserved. We may simply write the
image with offset parameter $h$ as

$$I_h(x, h) = \int S(x + h, x_s, \omega) R(x - h, x_s, \omega) dx_s d\omega$$

$$= \int d(x_r, x_s, \omega) G^+(x + h, x_s, \omega) G^+(x - h, x_r, \omega) dx_r dx_s d\omega \quad (3.10)$$

The offset $h$ is not necessarily horizontal, however, for practical considerations, $h$ will be taken only as horizontal. The image parameterized by offsets is called common image gathers in sub-surface offset, or in short, offset domain. A geometric interpretation is shown in Figure(3.1), where a sequence of horizontal offsets are used to correlate between $\overline{S}$ and $R$. When migrated at the true velocity, we see clearly that the constructive interference is obtained only at $h = 0$. The reconstructed general reflectivity distribution with point support on $h$ at origin can be written as

$$I_h(x, h) = I(x) \delta(h) \quad (3.11)$$
where \( I(x) \) is the image obtained by correlation of \( \bar{S} \) and \( R \) at zero offset, \( I(x) = \int \bar{S}(x, x_s, \omega) R(x, x_s, \omega) d(x_r, x_s, \omega) dx_s d\omega \). The equation (3.11) suggest the Euclidean norm

\[
J = \frac{1}{2} \| I_h(x, h) h \|^2
\]  

be minimized at the true velocity. We thus obtain an objective function that gives its minimum at the true velocity.
3.3 Differential semblance criteria

$I_h(x, h)$ is considered as a common image gather for $I(x)$ expanded by an offset parameter $h$, where $|h|$, according to equation (3.3), is the half of the correlation distance between downward continued source wavefield and the receiver wavefield inside the model. There are several other ways to expand the image $I(x)$ by introducing one an additional parameter, one of which is to use scattering angle $\theta$ (see Fig (3.2)). The geometrical meaning of $\theta$ is half of the angle enclosed by the wave number vector $k_s$ of downward continued source wavefield $S$ and the negative of the wave number vector $k_r$ of downward continued receiver wavefield $R$ at image point. A common image gather expanded in scattering angle is called an angle domain CIG, $I_\theta(x, \theta)$. The common image gather $I_\theta(x, \theta)$ is obtained by correlating the local plane wave decomposed from $S$ and the local plane wave decomposed for $R$ with their wave numbers bisected by the migration dip vector $\hat{n}$ Fig (3.2).

The differential semblance operator (DSO) provides measurements of common image gathers in some differential sense local to the imaging point $x$. We will compare and explain the relationship between two differential semblance operators, one in offset domain, the other in angle domain.
3.3.1 Offset domain DSO

We introduce operator $P_h$ not dependent on velocity

$$P_h : I_h \mapsto hI_h$$

(3.13)

as the offset domain differential semblance operator. The operator $P_h$ annihilates the singularities of $I_h(x, h)$ supported at $h = 0$ and magnifies signals of $I_h(x, h)$ away from $h = 0$. It thus provides a way to measure the deviation of offset domain CIG away from concentration in offset. The operator $P_h$ justifies itself as a differential semblance operator on image gather $I_h(x, h)$ for reasons which will be given later in this chapter.

The offset $h$ is not only an offset on the surface but also an offset in the model. For practical reasons, the offset parameter $h$ will only be taken as horizontal vectors. This choice of $h$ is proved to have maximum resolution power when subsurface structures are horizontal, Stolk & Symes, 2003[38], which is, in general, the case of sedimentary structure in the earth.

The offset domain CIGs are closely related to angle domain CIGs. Sava (2001) gives an geometrical analysis between them and showed that $I_\theta(x, \theta)$ relates to $I_h(x, h)$ through the Radon transform

$$I_\theta(x_1, x_2, \theta) = \int I_h(x_1 + qh, x_2, h)dh$$

(3.14)

where $q = -\tan\theta$. For convenience, we have used two dimensional space variable
x = (x₁, x₂), and h is restricted to horizontal one-dimensional. The conclusion can be easily extended to three dimensions with careful treatment of the plane of integration.

3.3.2 Relationship between \( I_h \) and \( I_0 \)

I shall give in here an independent analysis on the relationship between \( I_h \) and \( I_0 \), which is, hopefully, able to lead us to more general conclusions. Let the downward continued source wavefield \( S(x) \) have a plane wave decomposition locally at the image point \( x \),

\[
G^+(x, x_s, \omega) = \int \hat{G}^+(k_s, x_s, \omega) e^{ik_s \cdot x} dk_s. \tag{3.15}
\]

Similarly, for the downward continued receiver wavefield

\[
R(x, x_s, \omega) = \int \hat{R}(k_r, x_s, \omega) e^{ik_r \cdot x} dk_r. \tag{3.16}
\]

For a given angular frequency \( \omega \), the wavenumbers vector \( k_s \) and \( k_r \) are restricted for propagating plane waves as

\[
|k_s| = |k_r| = \omega/c(x). \tag{3.17}
\]

Let \( k_s \) be decomposed into vertical and horizontal components \((k_{s1}, k_{s2})\), and likewise, let \( k_r \) be decomposed into \((k_{r1}, k_{r2})\). Substituting equation(3.15) and equation(3.16) into equation(3.10), we obtain
\[ I_h(x, h) = \int \hat{G}^+(k_s, x_s, \omega) \hat{R}(k_r, x_s, \omega) e^{i(k_r - k_s) \cdot x} dx_s dk_s dk_r d\omega \]  
\[ = \int \hat{G}^+(k_s, x_s, \omega) \hat{R}(k_r, x_s, \omega) e^{-ik_s x_1 - k_s (x_2 + h)} e^{ik_r x_1 + k_r (x_2 - h)} dx_s d\omega dk_s dk_r \]  
(3.18)

Introduce the local coordinates \( (\hat{t}, \hat{n}) \) as shown in Figure(3.2) with \( \hat{n} \) the unit vector of the migration dip and \( \hat{t} \) the unit vector bisecting the angle between \( k_r \) and \( k_s \). It can be shown through a stationary phase analysis that the constructive interference in \( I_h \) is due to plane waves associated with wavenumbers \( k_s \) and \( k_r \) such that the following is satisfied

\[ k_s = \frac{\omega}{c} (\hat{t} \sin \theta + \hat{n} \cos \theta) \]

\[ k_r = \frac{\omega}{c} (\hat{t} \sin \theta - \hat{n} \cos \theta) \]  
(3.19)

The Radon transform of \( I_h(x, h) \) with respect to \( h \) in \( q = \tan(\theta) \) can be shown

\[ [\mathcal{R}_q I_h(x, h)](x, q) = \int I_h(x_1 + qh, x_2, h) dh \]

\[ = \int \hat{G}^+(k_s, x_s, \omega) \hat{R}(k_r, x_s, \omega) e^{-ik_s x_1 - ih q} e^{ik_s (x_2 + h)} \times \]

\[ e^{ik_r x_1 + k_r (x_2 - h)} dh dx_s d\omega dk_s dk_r \]

\[ = \int \hat{G}^+(k_s, x_s, \omega) \hat{R}(k_r, x_s, \omega) e^{i(k_r - k_s) \cdot x} \times \]

\[ e^{(-k_s x_1 - k_s x_2 + k_r q)h} dh dx_s dk_s dk_r d\omega \]

\[ = \int \hat{G}^+(k_s, x_s, \omega) \hat{R}(k_r, x_s, \omega) e^{i(k_r - k_s) \cdot x} \times \]

\[ 2\pi \delta[q(k_{r_1} - k_{s_1}) - (k_{r_2} + k_{s_2})] dx_s dk_s dk_r d\omega \]
\[
= 2\pi \int_{L(k_s, k_r, q)=0} \hat{G}^+(k_s, x_s, \omega) \hat{R}(k_r, x_s, \omega) e^{i(k_r-k_r')x} dk_s dk_r dx_s d\omega
\]

(3.20)

where the integration with respect to \(k_r\) and \(k_s\) is taken along the curve in wavenumber domain

\[
L(k_s, k_r, q) = \frac{k_{r2} + k_{s2}}{k_{r1} - k_{s1}} - q = 0.
\]

(3.21)

Combining equation (3.21) and equation (3.19), it is not difficult to see that \(q = -\tan(\theta)^*\). This integral of the Fourier components of \(\hat{G}^+\) and \(\hat{R}\) restricted to condition \(L(k_s, k_r, q) = 0\) with \(q = -\tan(\theta)\) can be viewed as a definition of angle domain image gathers \(I_\theta(x, \theta)\)

\[
I_\theta(x, \theta) = \int_{L(k_s, k_r, q)=0} \hat{G}^+(k_s, x_s, \omega) \hat{R}(k_r, x_s, \omega) e^{i(k_s+k_r)x} dk_s dk_r dx_s d\omega.
\]

(3.22)

We have thus verified that the angle domain image \(I_\theta\) relates to the offset domain image through the Radon transform

\[
I_\theta := I_q(x, q)|_{q=-\tan(\theta)} = [\mathcal{R}_q I_h(x, h)](x, q)|_{q=-\tan(\theta)}.
\]

(3.23)

*Let \(\alpha\) be the angle enclosed by \(\hat{x}_2\) and \(\hat{t}\), substituting \(\hat{t} = \hat{x}_2 \cos \alpha - \hat{x}_1 \sin \alpha\), \(\hat{n} = -\hat{x}_2 \sin \alpha + \hat{x}_2 \cos \alpha\) into equation (3.19), regrouping terms in \(\hat{x}_1\) and \(\hat{x}_2\), we have \(k_{s1} = \frac{\hat{x}_2}{\hat{c}}(-\sin \alpha \sin \theta + \cos \alpha \cos \theta), \ k_{s2} = \frac{\hat{x}_2}{\hat{c}}(\cos \alpha \sin \theta - \sin \alpha \cos \theta), \ k_{r1} = \frac{\hat{x}_2}{\hat{c}}(\cos \alpha \sin \theta + \sin \alpha \cos \theta), \ k_{r2} = \frac{\hat{x}_2}{\hat{c}}(-\sin \alpha \sin \theta - \cos \alpha \cos \theta). \) The result follows by direct substitution \(-\tan \theta = (k_{r2} + k_{s2})/(k_{r1} - k_{s1})\)
Figure 3.2  \( \hat{n} \) is migration dip unit vector, and \( \theta \) is the scattering angle. The wavenumber of \( G^+(x, x_r, \omega) \) points downward. So the wavenumber of \( \overline{G}^+(x, x_r, \omega) \) points upward.
3.3.3 Angle domain DSO

For convenience we shall always set

\[ q = -\tan(\theta) \]

and write when there is no confusion

\[ I_\theta = I_q. \]

It is clear that at the true velocity, \( I_h(x, h) \) has the form

\[ I_h(x, h) = I(x)\delta(h) \quad (3.24) \]

Thus, due to discussion from last section equation (3.14), we have

\[ I_q(x, q) = \int I_h(x_1 + qh, x_2, h)dh \quad (3.25) \]

which is independent of the choice of \( q \). This result implies that the common image gathers in angle are flat with respect to \( \tan^{-1} q \). A geometric interpretation is shown in figure(3.3). The statement suggests another measurement of the reconstructed reflectivity as apposed to common image gathers in offset domain. Introduce \( I_\theta(x, \theta) = I_q(x, q)|_{q=-\tan \theta} \). At the true velocity we have

\[ \frac{\partial I_\theta}{\partial \theta} = 0 \quad (3.26) \]
Figure 3.3  Common image gathers in angle are flat with respect to $\theta$.

or

$$\frac{\partial I_q}{\partial q} = 0. \quad (3.27)$$

Equations (3.26) and (3.27) provide measurements of deviation from flatness for angle domain image gathers. The vanishing derivative indicates incident and reflected rays meet at the same imaging point for all scattering angles $\theta$. Introducing $P_q = \theta$ as the angle domain differential semblance operator, the angle domain differential semblance criterion is posed as

$$\min \frac{1}{2} ||P_q I_q||^2. \quad (3.28)$$
3.3.4 Relationship between $P_h$ and $P_q$

A differential semblance operator $P$, according to different types of common image gathers it applies to, can be formulated in two forms as shown in this chapter. Other forms of differential semblance operators are proposed by Stolk and Symes, 2003 [38]. In this section, we discuss the relations between two formulations of $P$, one is simply $P_h = h$, where $h$ is the offset parameter, defined on offset domain CIG; and the other $P_q = \frac{\partial}{\partial q}$ defined on angle domain CIG.

The angle domain image gathers relates to offset domain image gathers through the Radon transform $I_q(x, q) = \mathcal{R}_q\{I_h(x, h)\}(x, q)$. Written in terms of $I_q$, we have in angle

\[ P_qI_q = P_q\mathcal{R}_qI_h \]  \hspace{1cm} (3.29)

which can be inverse Radon transformed to an image gather in offset $\mathcal{R}_q^{-1}P_h\mathcal{R}_qI_q$. If differential semblance criteria on offset or angle are equivalent, then it suggests that the operators $P_h$ and $\mathcal{R}_q^{-1}P_q\mathcal{R}_q$ should be equivalent. It also implies that $P_hP_h$ and $\mathcal{R}_qP_q\mathcal{R}_q^{-1}$ are equivalent. The inverse Radon transform can be formulated in two dimensions as, Deans, S. R., 1983[8]*

*The formula given in [8] is $\mathcal{R}^{-1} = -\frac{1}{2\pi t}H_2\mathcal{R}^*$. The pseudo-differential operator $\Delta^\frac{1}{2}_t$ can be shown in one dimension as $\Delta^\frac{1}{2}_t = -i\frac{2}{\pi t}H_t$ where $H_t$ is the Hilbert transform in one dimensional variable $t$. Equation (3.30) is posed in the form to agree with the inverse Radon transform in $n$ dimensions which can be formulated in unified form using the fractional
\[ R_q^{-1} = i \Delta^{\frac{1}{2}} R_q^*. \]

The operator \( i \Delta^{\frac{1}{2}} \), properly defined as a pseudo-differential operator, * does not change the phase of the signal transformed by adjoint Radon transform. It acts like taking a derivative in a pseudo-differential sense.

Instead of comparing \( P_h P_h \) (which is equal to \( P_h^* P_h \)) and \( R_q P_q P_q R_q^{-1} \) (which is equal to \( -R_q P_q P_q R_q^{-1} \)), we will compare \( P_h^* P_h \) and \( R_q^* P_q^* P_q R_q \). We shall see in the next chapter the gradient with respect velocity will pick up the image residual which has the form \( P_h^* P_h I_h \) in offset or \( R_q^* P_q^* P_q R_q I_h \) in angle. So the difference between those two forms are of the interest. To begin with, we first analyze the Radon transform and its adjoint. The forward Radon transform is defined as

\[ R_q f(x_1, x_2, h) := \int f(x_1 + qh, x_2, h) dh = \hat{f}(x_1, x_2, q) \quad (3.31) \]

We will suppress the dependency on \( x_2 \) for convenience. To see the adjoint of the Radon transform, we write equation (3.31) in a slightly different form.

\[
(R_q f)(x_1, q) = \int f(t, h) \delta(t - qh - x_1) \, dt \, dh = \int \frac{1}{2\pi} f(t, h) e^{iu(t-qh-x_1)} \, dt \, dh \quad (3.32)
\]

---

\*Laplacian \( \Delta^{-1} = a_n i^{1-n} \Delta^{(n-1)/2} R^* \) for \( a_n \) some real number dependent on \( n \). [8].

\*\( \Delta^{\frac{1}{2}} \) applies to \( f(x) \) for \( x \in \mathbb{R}^n \) is understood as \( \Delta^{\frac{1}{2}} f(x) = \frac{1}{(2\pi)^n} \int \hat{f}(k)|k|e^{i k \cdot x} \, dk \) where \( |k| = \sqrt{\sum_j k_j^2} \), and \( \hat{f}(k) = \int f(x)e^{-i k \cdot x} \, dx \).
We then see the kernel of the adjoint Radon transform is \( \frac{1}{2\pi} e^{-iw(t-q_h-x_1)} \). The adjoint Radon transform can be written as

\[
(\mathcal{R}_q^* f)(t, h) = \int \frac{1}{2\pi} f(x_1, q) e^{-iw(t-q_h-x_1)} dt dh d\omega
\]

\[
= \int f(t - q_h, q) dq
\]

Using the relation \( P_q^* = -\frac{\partial}{\partial q} \), we derive

\[
\mathcal{R}_q^* P_q^* P_q \mathcal{R}_q I = \mathcal{R}_q^* \left(-\frac{\partial^2}{\partial q^2}\right) \mathcal{R}_q I_h
\]

\[
= \mathcal{R}_q^* \int \frac{\partial^2}{\partial q^2} I(t, h) e^{i\omega(t-q_h-x_1)} \frac{1}{2\pi} dt dh d\omega
\]

\[
= \int h'^2 \omega^2 I(t', h') e^{i\omega(t'-q_h'-x_1)} \frac{1}{4\pi^2} e^{-i\omega(t-q_h-x_1)} dx_1 dq dt' dh' d\omega' d\omega
\]

\[
= \int h'^2 \omega^2 I(t', h') e^{i\omega(t'-t)} e^{i\omega(q(h'-h))} \frac{1}{2\pi} dq dh' dt' d\omega
\]

\[
= \int h'^2 |\omega| I(t', h) e^{i\omega(t'-t)} dt' d\omega
\]

\[
= \int_{-\infty}^{+\infty} dt' h'^2 I(t', h) \int_{-\infty}^{+\infty} |\omega| e^{i\omega(t'-t)} d\omega
\]

\[
= -2h^2 \int_{-\infty}^{+\infty} dt' \frac{I(t', h)}{(t-t')^2}
\]

(3.33)

In the last step, the principal value integration is understood using the relation

\[
\int_{-\infty}^{\infty} |\omega| e^{i\omega z} d\omega = -\frac{2}{z^2}.
\]

(3.34)

Equation (3.33) justifies that \( P_2 = h \) is a differential semblance operator, since \( P_h^* P_h I_h \)
relates to \( \mathcal{R}_q^* P_q^* P_q \mathcal{R}_q I_h \) through the integral transform

\[
(\mathcal{R}_q^* P_q^* P_q \mathcal{R}_q I_h)(t, h) = P_h^* P_h \int_{-\infty}^{+\infty} \frac{I_h(t', h)}{(t-t')^2} dt'.
\]

(3.35)
The question is how different is it between \( P_h^s P_h I_h \) and \( \mathcal{R} P^s P_q \mathcal{R} I_h \)? The answer, according to equation (3.35) depends on the relation between \( I_h(t, h) \) and \( \int -2I_h(t, h) \, dt' \), which is an integral transform recognized by Hilbert transform followed by differential. Write \( DH_t = -2\pi \frac{\partial}{\partial t} H_t \). Here \( H_t \) is an Hilbert transform in \( t \).*

The transform \( DH_t \) can be analyzed using its Fourier representation, write

\[
[(\frac{\partial}{\partial t} H_t)f(t', h)](t, h) = \int |\omega| f(t', h)e^{i\omega(t'-t)} \, dt' \, d\omega
\]

(3.36)

This transform doesn’t change the phase of the original signal, but the amplitude is weighted according to the magnitude of its Fourier frequencies. Its spectrum of high frequencies components have higher weights while low frequencies have relatively small weights. Figure (3.4) demonstrates this phenomenon through the comparison between an untransformed signal \( f(x) = e^{-0.5(x-2.5)^2}[\cos(3x) + \cos(10x) + \sin(5x)] \) and its \( \frac{\partial}{\partial t} H_t \) transformation. In conclusion, the \( \frac{\partial}{\partial t} H_t \) transformed wavelet will always be in phase with the original wavelet. At places where there are high frequency signals, the \( \frac{\partial}{\partial t} H_t \) transformation tend to scale up their amplitudes. When low frequency components carry energy close to zero, the \( \frac{\partial}{\partial t} H_t \) transformed wavelet shows significant similarities comparing with the original one. For general seismic common-image offset gathers regarded as function in \( x_1 \) and \( h \), wavelets in \( x_1 \) are usually zero energy for low

*definition of Hilbert transform: \( (Hf)(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} dt' \frac{f(t')}{t'-t} \).
frequencies. (see Figure 3.5) For this reasoning, we consider differential semblance operator $P_h$ a good substitute for $P_q \mathcal{R}_q$.

For the rest of the thesis, the differential semblance objective function will be the form

$$\frac{1}{2} ||P_h I_h||^2 = \frac{1}{2} < P_h I_h, P_h I_h >,$$  \hspace{1cm} (3.37)

which in discrete form, we have
\[
\frac{1}{2} \| P_h I_h \|^2 = \sum_h \sum_x \sum_z h^2 I(x_1, x_2, h) I(x_1, x_2, h). \tag{3.38}
\]
3.4 Kinematics of the image in offset

In this section I will give a simple analysis to the kinematics of the image in offset in two-dimensions, $I_h(x, z, h)$, in the case of constant velocity and flat reflectors, where $x$, $z$ and $h$ are horizontal, vertical and offset coordinates, respectively. The singular support of $I_h$ describes the kinematic behavior of the image which can be understood through a stationary phase argument.

3.4.1 Kinematics of image in offset parameterized by surface offset $I_h(x, z, h; H)$

We construct first a model coordinates in two-dimensions $(x, z)$ using $x$ for the horizontal coordinate and $z$ for the vertical coordinate. Let a horizontal reflector be located in depth at $Z$. A reflection event at the point $(x, Z)$ can be associated to the pair of source and receiver located on the surface in coordinates $(x - H, 0)$ and $(x + H, 0)$, respectively, separated by a surface offset $2H$ with respect to the mid-point $(x, 0)$. See Figure (3.6(a)).

We can parameterize by $H$ the image in offset due to such single reflection as $I_h(x, z, h; H)$. The kinematics of $I_h(x, z, h; H)$ for the pair of receivers and the reflection point $(x, Z)$ is the set of points in a sub-surface coordinates $[x, z, h]$ through which the travel time in migration equals the physical travel time. Since we will limit ourselves to the re-
flection due to a fixed mid-point \(x\) only, the sub-surface coordinates can be simplified as \([z, h]\) and the \(H\) parameterized image in offset is simply written as \(I_h(z, h; H)\).

We write the physical travel time from the source \((x + H, 0)\) to the receiver \((x - H, 0)\) reflected by the horizontal reflector \((x, Z)\) as

\[
T_{H,Z} = \frac{2\sqrt{H^2 + Z^2}}{c}
\]  

(3.39)

for \(c\) the constant true velocity. The migration travel time in offset denoted by the sub-surface coordinate \([x, h]\) is expressed as the sum of the migration travel time from the source \((x + H, 0)\) to a sub-surface point \((x + h, z)\) and the travel time from the receiver \((x - H, 0)\) to another point \((x - h, z)\),

\[
t_{H,z} = \frac{2\sqrt{(H - h)^2 + z^2}}{u}
\]  

(3.40)

for \(u\) the constant migration velocity. The constructive interference (or the singular support) of \(I_h(z, h; H)\) happens at the sub-surface coordinates \([h, z]\) such that \(T_{H,Z} = t_{H,z}\). We derive from equation (3.39) and (3.40)

\[
(h - H)^2 + z^2 = \rho^2(H^2 + Z^2)
\]  

(3.41)

where

\[
\rho = \frac{u}{c}
\]

is the ratio of the migration velocity and the true velocity. Equation (3.41) repre-
sents a circle centered at the surface \([0, H]\) with radius \(\rho \sqrt{H^2 + Z^2}\) in the sub-surface coordinate \([h, z]\) (the red circle in Figure (3.6(a))). Locally near \(h = 0\) this circle has a tangent line at an angle of \(\arctan(H/Z)\) to the vertical line if \(u = v\). A local slant stack in angle \(\theta\) (Radon transform \(\mathcal{R}_q|_{q=\tan\theta}\)) of \(I_h\) will pick up the energy on a line tilted in \(\theta\) with respect to the vertical line only when \(\theta = \arctan(H/Z)\).

### 3.4.2 The envelope of one-parameter family of curves

Equation (3.41) characterize a one-parameter family of curves in the sub-surface coordinates. Write these set of curves parameterized by \(H\) as

\[
F(z, h; H) = (h - H)^2 + z^2 - \rho^2(H^2 + Z^2)
\]  

(3.42)

The kinematics of the total image in offset \(I_h(z, h)\) is represented by the envelope of the curves \(F(z, h; H)\). The envelope is determined jointly by the contact condition

\[
\frac{\partial F(z, h; H)}{\partial H} = 0
\]  

(3.43)

and equation (3.42). Eliminating \(H\) by solving equation (3.42) and (3.43) we derive

\[
\frac{h^2}{(\rho^2 - 1)Z^2} + \frac{z^2}{\rho^2 Z^2} = 1.
\]  

(3.44)

Equation (3.44) characterizes the kinematics of the common image in offset for some fixed mid-point \(x\). When the migration velocity \(u\) is greater than the true velocity
\( c, \rho > 1 \) and equation (3.44) is an eclipse as shown in Figure (3.6(a)). When the migration velocity \( u \) is less than the true velocity \( c \), we have \( \rho < 1 \), and equation (3.44) becomes a hyperbola as shown in Figure (3.6(b)). At \( \rho = 1 \), the migration velocity equals the true velocity, there is no envelope to the set of curves described by equation (3.42). The kinematics of total \( I_h(z, h) \) fall into a point at \([Z, 0]\). See Figure (3.6(c)).

3.4.3 Offset range in relation to surface acquisition limit

The surface acquisition limit can be characterized by the maximum surface offset \( H \) for each mid-point \( x \). The limited range in \( H \) results in the limited range in offset \( h \) on the quadratic curve described by equation (3.42). Although it is not used in the thesis, the determination of the maximum \( h \) due to limited \( H \) is a very useful information for the construction of the objective function in offset. I will show the relation of maximum offset \( h_{\text{max}} \) induced by the maximum surface offset \( H_{\text{max}} \).

The \( h_{\text{max}} \) is determined by the contact point between the circle parameterized by \( H \)

\[
(h - H)^2 + z^2 = \rho^2(H^2 + Z^2)
\]

(equation (3.41 repeat) evaluated at \( H_{\text{max}} \) and the envelope

\[
\frac{\rho^2}{(\rho^2 - 1)Z^2} + \frac{z^2}{\rho^2 Z^2} = 1.
\]
(equation (3.44 repeat). At the contact point equation (3.41) and (3.44) share the same tangent line,

\[
\text{equation (3.41)} \Rightarrow \frac{dz}{dh} = \frac{H_{\text{max}} - h}{z}
\]

\[
\text{equation (3.44)} \Rightarrow \frac{dz}{dh} = \frac{\rho^2 h_{\text{max}}}{1 - \rho^2 z}
\]

which is solved as

\[
h_{\text{max}} = H_{\text{max}}(1 - \rho^2)
\]  (3.45)
(a) $\rho < 1$

(b) $\rho > 1$

(c) $\rho > 1$
Chapter 4

Adjoint state calculation and inversion

Chapter Synopsis

Wave equation migration velocity analysis is posed as a nonlinear optimization problem with the objective function defined on the image domain. The gradient of the objective function with respect to velocity provides search directions for iterative velocity updating. Fast algorithm of quasi-Newton’s method can be implemented to solve the nonlinear inverse problem with the aid of the gradient calculation.

I shall give formulations of the gradient of the objective function in both continuous as well as discrete wave equation representations. Extensive adjoint state analysis is used to derive gradient calculation within the frame work of one-way wave equation migration. The limited memory BFGS algorithm, Nocedal & Wright, 2000[25], is implemented to solve the inverse problem where the smoothing properties of the gradient are analyzed and a B-spline smoothing scheme is developed.
4.1 Gradient calculation and adjoint state analysis

In this section I will give three formulations of the gradient calculation. First, I will give an integral representation of the gradient using the Green’s functions. This formulation is employed to analyze the gradient smoothness later in section (4.2.2). Second, I will show a formulation of the gradient by solving differential equations which serves as a guideline of the calculation in theory. Third, I shall give an explicit discrete formulation of the gradient calculation, which describes how the gradient is implemented in this thesis work.

4.1.1 Gradient formulation by the Green’s function representation

The offset domain differential semblance objective function is written as

\[ J[c] = \frac{1}{2} \| P_h I_h \|_2^2 \] (4.1)

where \( \| \cdot \|_2 \) denotes the \( L_2 \) norm, \( I_h(x, h) \) is the image gather parameterized by subsurface offset parameter \( h \), and \( P_h \) is a differential semblance operator defined in offset

\[ P_h : I_h(x, h) \rightarrow h I_h(x, h). \] (4.2)

The image gather in offset \( I_h(x, h) \) is nonlinearly dependent on the velocity \( c \), write

\[ I_h = f[c] \] (4.3)
and

\[ \delta I_h = (DI_h)[c] \delta c = \frac{\partial I_h}{\partial c} \delta c \]  

(4.4)

where \( DI_h \) is the differential migration with respect to velocity obtained by taking first order derivative of \( f \) with respect to \( c \). The perturbation of the objective function written as an inner product \( <,> \) is

\[ \delta J = \frac{1}{2} < \delta(P_h I_h), P_h I_h > + \frac{1}{2} < P_h I_h, \delta(P_h I_h) > \]

\[ = \frac{1}{2} < P_h DI_h \delta c, P_h I_h > + \frac{1}{2} < P_h I_h, P_h DI_h \delta c > \]

\[ = \frac{1}{2} < \delta c, DI_h^* P_h^* P_h I_h > + \frac{1}{2} < DI_h^* P_h^* P_h I_h, \delta c > \]  

(4.5)

\[ = \frac{1}{2} < \delta c, DI_h^* P_h^* P_h I_h > + \frac{1}{2} < \delta c, DI_h^* P_h^* P_h I_h > \]

\[ = < \delta c, Re(DI_h^* P_h^* P_h I_h) > \]

the gradient of \( J \) with respect to \( c \) can then be read

\[ \nabla_c J = Re\{ (DI_h)^* (P_h^* P_h I_h) \} \]  

(4.6)

which is understood as the real part of the composition of adjoint differential migration \( DI_h^* \) and \( P_h^* \) applied to image residual \( P_h I_h \). The adjoint differential migration is the central concept to be investigated in this section. To formulate the adjoint differential migration we first derive a Green’s function representation of the differential migration, the adjoint of which will then follow easily by taking the complex conjun-

---

*An angle domain differential semblance objective function can be written as \( \tilde{J} = \frac{1}{2} ||P_q^* \tilde{R}_q I_h||^2 \). The gradient of \( \tilde{J} \) with respect to velocity \( c \) can be shown similarly as \( \nabla_c \tilde{J} = Re(DI_h^* \tilde{R}_q^* P_q^* P_q \tilde{R}_q I_h) \). We see the gradient of the angle domain DSO and that of offset domain DSO differ only in the operators \( \tilde{R}_q^* P_q^* P_q \tilde{R}_q \) and \( P_h^* P_h \).
gate of the integration kernel. The sub-surface offset semblance image is formulated by adjoint of the generalized Born modeling. Equation (3.10) is re-written in here as

\[
I_h(x, h) = \int G^+(x - h, x_r, \omega)d(x_r, x_s, \omega)G^+(x + h, x_s, \omega)dx_r dx_s d\omega
\]  

(4.7)

where \(d(x_r, x_s, \omega)\) is the data in angular frequency \(\omega\) received at \(x_r\) due to a point source at \(x_s\) and \(G(x, y, \omega)\) is the Green’s function observed at point \(x \in \mathbb{R}^3\) due to a point source located at \(y \in \mathbb{R}^3\). An image perturbation \(\delta I_h = (DI_h)\delta c\) can be formulated through Born perturbations of Green’s functions \(G^+(x + h, x_s, \omega)\) and \(G^+(x - h, x_r, \omega)\),

\[
\delta I_h(x, h) = \int \frac{G^+(y, x_r, \omega)\delta c(y)\omega^2}{c(y)^2} G^+(x - h, y, \omega)d(x_r, x_s, \omega)G^+(x + h, x_s, \omega) \times dx_r dx_s dy d\omega
\]

\[
+ \int \frac{G^+(x - h, x_r, \omega)d(x_r, x_s, t)G^+(x + h, y, \omega)\delta c(y)\omega^2}{c(y)^2} G^+(y, x_s, \omega) \times dx_r dx_s dy d\omega
\]  

(4.8)

The adjoint of \(DI_h\), \((DI_h)^* : u(x, h) \rightarrow g(y)\) can be read from equation (4.8)

\[
g(y) = (DI_h)^* u
\]

\[
= \int \frac{G^+(y, x_r, \omega)u(x, h)\omega^2}{c(y)^2} G^+(x - h, y, \omega) \times dx_r dx_s dx dhd\omega
\]

\[
+ \int \frac{G^+(x - h, x_r, \omega)d(x_r, x_s, \omega)G^+(x + h, y, \omega)\times u(x, h)\omega^2}{c(y)^2} G^+(y, x_s, \omega) dx_r dx_s dx dhd\omega.
\]  

(4.9)
When \( u(x, h) = P_h^* P_h I_h \), equation (4.9) gives the gradient of the offset domain differential semblance objective function. The two terms in equation (4.9) are symmetric to each other. We shall analyze the first term, and then carry the similar analysis to the second term. We have

\[
g_1(y) = \int \frac{\omega^2}{c(y)^2} dx_d d\omega \left( \int G^+(y, x_r, \omega) \overline{d(x_r, x_s, \omega)} dx_r \right) \times \\
\int dx_d h G^+(x + h, x_s, \omega) u(x, h) G^+(y, x - h, \omega)
\]

This integral consists of three steps: first, downward continue the receiver wavefield complex conjugated,

\[
\overline{R}(y, x_s, \omega) = \int G^+(y, x_r \omega) \overline{d(x_r, x_s, \omega)} dx_r \\
= \int G^+(x_r, y, \omega) d(x_r, x_s, \omega) dx_r. \quad (4.10)
\]

Second, downward continue source wavefield, correlate in offset \( h \) with the residual image \( u(x, h) \) and then upward propagate to point \( y \),

\[
g_s(y, x_s, \omega) = \int G^+(x + h, x_s, \omega) u(x, h) G^+(y, x - h, \omega) dx_d h. \quad (4.11)
\]

Finally, multiply \( \overline{R} \) by \( g_s \) then integrate

\[
g_1(y) = \int \overline{R}(y, x_s, \omega) g_s(y, x_s, \omega) dx_s d\omega \\
= \int \overline{R}(y, x_s, \omega) g_s(y, x_s, \omega) dx_s d\omega \quad (4.12)
\]

Similarly, we can write the second term in equation (4.9) as
\[ g_2 = \int S(y, x_s, \omega)g_r(y, x_s, \omega)dx_sd\omega \]  

(4.13)

where

\[ S(y, x_s, \omega) = G^+(y, x_s, \omega) \]

is the downward continued source wavefield complex conjugated, and

\[ g_r(y, x_s, \omega) = \int G^+(x - h, x_r, \omega)d(x_r, x_s, \omega)G^+(x + h, y, \omega)dxdhdx_r. \]  

(4.14)

We shall call \( g_s \) the adjoint state perturbation source wavefield and \( g_r \) the adjoint state perturbation receiver wavefield. The gradient formulated using the Green's function representation, equation (4.9), can be rewritten in terms of \( S, \overline{R}, g_r \) and \( g_s \) as

\[ g(y) = \int \frac{\omega^2}{c(y)^2}(\overline{R}g_s + \overline{S}g_r)(y, x_s, \omega)dx_sd\omega. \]  

(4.15)

The three conceptual steps for taking the integral shown here is well correlated to the computational procedure for the gradient derived later in section (4.1.2). Instead of expressing the gradient as an integral, we shall separate the integral into parts each of which satisfies a wave equation. The advantage of expressing the gradient as an integral is fully utilized in section (4.2.3), where the smoothing properties of the gradient are analyzed through the method of stationary phase.
4.1.2 Gradient formulation by wave equation

The gradient integral in the Green's function representation derived in the last section can be partitioned into several sub-integrals. Each sub-integral can be shown, depending on whether the Green's function used is a two-way Green's function or a one-way Green's function, to satisfy two-way or one-way wave equations, respectively. The gradient can then be formulated using two-way or one-way wave equations accordingly.

**Gradient by two-way wave equation**

In this section I will use the two-way Green's function to formulate the gradient. Let's first repeat the migration formula using two-way Green's functions:

\[
I_h(x, h) = \int \tilde{S}(x, x_s, \omega) \tilde{R}(x, x_s, \omega) \, dx_s d\omega
\] (4.16)

where \( \tilde{S} \) is the downward continued source wavefield by the two-way Green's function, \( \tilde{S}(x + h, x_s, \omega) = \tilde{G}(x + h, x_s, \omega) \), and \( \tilde{R} \) is the downward continued receiver wavefield, \( \tilde{R}(x, x_s, \omega) = \int d(x_r, x_s, \omega) \tilde{G}(x_r, x, \omega) dx_r \). \( \tilde{S} \) solves Cauchy problem:

\[
(\nabla^2 + \frac{\omega^2}{c^2})\tilde{S}(x, x_s, \omega) = \delta(x - x_s)
\] (4.17)

\[
(\alpha\tilde{S} + \beta \nabla \tilde{S})|_{\Sigma} = 0
\]
where we have extended the Dirichlet boundary condition on the boundary of the computational domain to the form \((\alpha \tilde{S} + \beta \nabla \tilde{S})|\Sigma = 0\) with \(\alpha, \beta\) constant. Likewise, \(\tilde{R}\) solves

\[
(\nabla^2 + \frac{w^2}{c^2})\tilde{R}(x, x_s, \omega) = \int d(x_r, x_s, \omega)\delta(x - x_r)dx_r \tag{4.18}
\]

\[
(\alpha \tilde{R} + \beta \nabla \tilde{R})|\Sigma = 0.
\]

Write the perturbation of \(I_h(x, h)\) as

\[
\delta I_h = \int (\delta \tilde{S}\tilde{R} + \tilde{S}\delta \tilde{R})(x, x_s, h, \omega)dx_sd\omega. \tag{4.19}
\]

To determine \((DI_h)^*\), we construct a scalar product of \(I_h\) with an arbitrary image gather in offset \(u(x, h)\)

\[
\langle \delta I_h, u \rangle = \langle DI_h \delta c, u \rangle
\]

\[
= \int (\delta I(x, h)u(x, h) \ dx dh
\]

\[
= \int \delta \tilde{S}(x + h, x_s, \omega)\tilde{R}(x - h, x_s, \omega)u(x, h)dx_sd\omega dx dh + \int \tilde{S}(x + h, x_s, \omega)\delta \tilde{R}(x - h, x_s, \omega)u(x, h)dx_sd\omega dx dh
\]

\[
= \int \delta \tilde{S}(x + h, x_s, \omega)\int \tilde{R}(x - 2h, x_s, \omega)u(x - h, h)dh dx_s d\omega + \int \delta \tilde{R}(x - h, x_s, \omega)\int \tilde{S}(x + 2h, x_s, \omega)u(x + h, h)dh dx_s d\omega
\]

\[
= \langle \delta c, (DI_h)^*u \rangle
\]

We now introduce \(\tilde{g}_r\) and \(\tilde{g}_s\) which satisfy
\[(\nabla^2 + \frac{w^2}{c^2})\tilde{g}_r(x, x_s, \omega) = \int \tilde{R}(x - 2h, x_s, \omega)u(x - h, h)dh\]  
\[\int_\Sigma (\nabla \tilde{S}\tilde{g}_r - \nabla \tilde{g}_r\tilde{S})da = 0\]  
\[(\nabla^2 + \frac{w^2}{c^2})\tilde{g}_s(x, x_s, \omega) = \int \tilde{S}(x + 2h, x_s, \omega)u(x + h, h)dh\]  
\[\int_\Sigma (\nabla \tilde{R}\tilde{g}_s - \nabla \tilde{g}_s\tilde{R})da = 0\]  

Here \(da\) is the differential surface area on \(\Sigma\). We have extended the Dirichlet boundary condition to the form \(\int_\Sigma (\nabla \tilde{S}\tilde{g}_r - \nabla \tilde{g}_r\tilde{S})da = 0\) and \(\int_\Sigma (\nabla \tilde{R}\tilde{g}_s - \nabla \tilde{g}_s\tilde{R})da = 0\) for \(\tilde{g}_r\) and \(\tilde{g}_s\), respectively. The boundary conditions will be used in the integration by parts later to yield the desired results. Substitute \(\tilde{g}_s\) and \(\tilde{g}_r\) into equation (4.20), and then do integration by parts, we derive

\[
\langle \delta I, u \rangle = \int \{\delta \tilde{S}(\nabla^2 + \frac{w^2}{c^2})\tilde{g}_r + \delta \tilde{R}(\nabla^2 + \frac{w^2}{c^2})\tilde{g}_s\}dx_s d\omega dx
\]
\[
= \int [(\nabla^2 + \frac{w^2}{c^2})\delta \tilde{S}]\tilde{g}_r + [(\nabla^2 + \frac{w^2}{c^2})\delta \tilde{R}]\tilde{g}_s dx_s d\omega dx
\]
\[
= \int \frac{2w^2}{c^2} \delta c(x)\tilde{S}\tilde{g}_r + \frac{2w^2}{c^2} \delta c(x)\tilde{R}\tilde{g}_s dx_s d\omega dx
\]

where here we have used the Born scattering relations

\[(\nabla^2 + \frac{w^2}{c^2})\tilde{S} = \delta(x - x_s) \Rightarrow (\nabla^2 + \frac{w^2}{c^2})\delta \tilde{S} = \frac{2w^2}{c^3} \delta c \tilde{S}\]
\[(\nabla^2 + \frac{w^2}{c^2})\tilde{R} = d(x_r, x_s, \omega)\delta(x - x_s) \Rightarrow (\nabla^2 + \frac{w^2}{c^2})\delta \tilde{R} = \frac{2w^2}{c^3} \delta c \tilde{R}\]

From equation (4.20), we can read off \((DI_h)^*u\) as

\[(DI_h)^*u = \int \frac{2w^2}{c^3} (\tilde{S}\tilde{g}_r + \tilde{R}\tilde{g}_s)(x, x_s, \omega)dx_s d\omega,\]
which is two-way version gradient formulation comparing with equation (4.15).
Gradient by one way-wave equation

Next I shall derive \((D\Omega_h)^*\) using the one-way wave equation. We start with a similar approach as we did using two-way wave equation. In last section, \(S\), \(R\), \(g_s\) and \(g_r\) are solved using two-way wave operator \(\nabla^2 + \frac{u^2}{c^2}\). In this section I show how they may be solved using one-way wave operators \(\frac{\partial}{\partial x_1} + i\sqrt{\frac{\partial^2}{\partial x_{2,3}^2} + \frac{u^2}{c^2}}\) and \(\frac{\partial}{\partial x_1} - i\sqrt{\frac{\partial^2}{\partial x_{2,3}^2} + \frac{u^2}{c^2}}\).

For convenience, let’s first write down the corresponding one-way wave relations that \(S\) and \(R\) satisfy,

\[
\left(\frac{\partial}{\partial x_1} + i\sqrt{\frac{\partial^2}{\partial x_{2,3}^2} + \frac{u^2}{c^2}}\right)S(x, x_s, \omega) = \delta(x - x_s) \tag{4.26}
\]

\[
\int_{\Sigma(z)} (\nabla_x S g - \nabla_x g S)da = 0
\]

and

\[
\left(\frac{\partial}{\partial x_1} + i\sqrt{\frac{\partial^2}{\partial x_{2,3}^2} + \frac{u^2}{c^2}}\right)R(x, x_s, \omega) = \int d(x_r, x_s, \omega)\delta(x - x_r)dx_r \tag{4.27}
\]

\[
\int_{\Sigma(x_1)} (\nabla_x R g_s - \nabla_x g_s R)da = 0.
\]

The one-way adjoint state wavefields, \(g_r\) and \(g_s\), introduced in section(4.1.1) can be shown to satisfy the following equations.

\[
\left(\frac{\partial}{\partial x_1} + i\sqrt{\frac{\partial^2}{\partial x_{2,3}^2} + \frac{u^2}{c^2}}\right)g_r(x, x_s, \omega) = \int R(x - 2h, x_s, \omega)u(x - h, h)\, dh
\]

\[g_r(x_1 = z_l, x_{2,3}, x_s, \omega) = 0\tag{4.28}\]

\[
\int_{\Sigma(x_1)} (\nabla_x S g_r - \nabla_x g_r S)da = 0
\]

where \(z_l\) is the maximum depth that the computational propagation reaches.
\[
(\frac{\partial}{\partial x_1} + i \sqrt{\frac{\partial^2}{\partial x_{2,3}^2} + \frac{w^2}{c^2}}) g_s (z, x; s, w) = \int \tilde{S}(z, x + 2h; s, w) u(z, x + h, h) \, dh
\]
\[
g_r (z = z_1, x; s, w) = 0
\]  
(4.29)
\[
\int_{\Sigma(x_1)} (\nabla_x R g_s - \nabla_x g_s R) \, da = 0
\]

Here \(\Sigma(x_1)\) represents the horizontal boundary for each \(x_1\). Substituting \(g_r\) and \(g_s\) into equation (4.20) and noticing,

\[
\{i \sqrt{\frac{\partial^2}{\partial x_{2,3}^2} + \frac{w^2}{c^2}}\}^* = -i \sqrt{\frac{\partial^2}{\partial x_{2,3}^2} + \frac{w^2}{c^2}}
\]
\[
(\frac{\partial}{\partial x_1})^* = -\frac{\partial}{\partial x_1}
\]

then performing integration by parts, we obtain

\[
\langle \delta I_h, u \rangle = \int \overline{\delta S(x + h, x_s, \omega)} R(x - h, x_s, \omega) u(x, h) +
\]
\[
\overline{S(x + h, x_s, \omega)} \delta R(x - h, x_s, \omega) u(x, h) \, dx \, dh \, dx_s \, d\omega
\]
\[
= \int \overline{\delta S(x, x_s, \omega)} R(x - 2h, x_s, \omega) u(x, h) +
\]
\[
\overline{\delta R(x, x_s, \omega)} S(x + 2h, x_s, \omega) u(x, h) \, dx \, dh \, dx_s \, d\omega
\]
\[
= \int [\overline{\delta \left( \frac{\partial}{\partial x_1} + i \sqrt{\frac{\partial^2}{\partial x_{2,3}^2} + \frac{w^2}{c^2}} \right) g_r + \overline{\delta R \left( \frac{\partial}{\partial x_1} + i \sqrt{\frac{\partial^2}{\partial x_{2,3}^2} + \frac{w^2}{c^2}} \right) g_s}}] \, dx \, dx_s \, d\omega
\]
\[
= -\int \left[ (\frac{\partial}{\partial x_1} + i \sqrt{\frac{\partial^2}{\partial x_{2,3}^2} + \frac{w^2}{c^2}}) \delta S g_r + (\frac{\partial}{\partial x_1} + i \sqrt{\frac{\partial^2}{\partial x_{2,3}^2} + \frac{w^2}{c^2}}) \delta R g_s \right] \, dx \, dx_s \, d\omega
\]
\[
= \int \left[ (\overline{\delta L^+} - \frac{\partial}{\partial c} S) \delta c g_r + (\overline{\delta L^+} - \frac{\partial}{\partial c} R) \delta c g_s \right] \, dx \, dx_s \, d\omega
\]
\[
= \langle \delta c, (DI_h)^* u \rangle
\]  
(4.30)

where we recall
\[ L^+ = \frac{\partial}{\partial x_1} + i \sqrt{\frac{\partial^2}{\partial x_{2,3}^2} + \frac{w^2}{c^2}} \]

and

\[ L^+ S = \delta(x - x_s) \Rightarrow L^+ \delta S = -\left( \frac{\partial L^+}{\partial c} S \right) \delta c \]

We then read \((DI_h)^*u\) from equation (4.30) as

\[
(DI_h)^*u = \int \left\{ \left( \frac{\partial L^+}{\partial c} S \right) g_r + \left( \frac{\partial L^+}{\partial c} R \right) g_s \right\} dxdx_sd\omega
\]
4.1.3 Gradient by discrete adjoint state calculation

Numerical optimization algorithms depend on accurate adjoint state calculation Nocedal & Wright, 2000[25]. There are in general two ways to calculate the adjoint: discretise the continuous form, or directly take the adjoint of the discretized form. I shall in this section discuss adjoint state calculations of the second kind. The discussion follows mainly that of Ralf & Kaminski, 1998[13]. To start with, let’s look at the chain rule in discrete formulation.

4.1.3.1 The chain rule

Let a numerical algorithm be defined as a function

\[ H : X \rightarrow Y \]  \hspace{1cm} (4.31)

This can be decomposed into \( N \) steps, each having an explicit representation

\[ H^i : Z^{i-1} \rightarrow Z^i \] \hspace{1cm} (i = 1, ..., N)  \hspace{1cm} (4.32)

The variables \( Z^i \) holds all intermediate results that remains after \( i \)th step of the algorithm. For example, a numerical algorithm at the first step

\[ y = x \]  \hspace{1cm} (4.33)
has the input variable $Z^0 = (x^0, y^0)^T$ and output variable $Z^1 = (x^1, y^1)^T$, where $x^0$ and $y^0$ represent the variables before the assignment, and $x^1$ and $y^1$ are the variables after the assignment. We then have

$$
\begin{pmatrix}
  y^1 \\
  x^1
\end{pmatrix} =
\begin{pmatrix}
  0 & 1 \\
  0 & 1
\end{pmatrix}
\begin{pmatrix}
  y^0 \\
  x^0
\end{pmatrix}.
$$

The matrix representation of $H^1$ is identified as

$$
H =
\begin{pmatrix}
  0 & 1 \\
  0 & 1
\end{pmatrix},
$$

and

$$
Z^0 =
\begin{pmatrix}
  y^0 \\
  x^0
\end{pmatrix},
$$

$$
Z^1 =
\begin{pmatrix}
  y^1 \\
  x^1
\end{pmatrix}.
$$

The algorithm $H$ is considered as composition of differentiable functions $H_i$'s,

$$
H = H^N \circ H^{N-1} \circ ... \circ H^1 := \prod_{i=1}^{N} o H^i
$$

(4.34)

which can be differentiated according to the chain rule. To see the chain rule, let’s look at an intermediate step,

$$
Z^i = H^i(Z^{i-1}) \quad 1 \leq i \leq N
$$

(4.35)
A variation $\delta Z^i$ depends on a variation of the control variables $\delta Z^{i-1}$,

$$
\delta Z^i = \frac{\partial H^i}{\partial Z^{i-1}}|_{Z^{i-1} = H(Z^{i-2})} \delta Z^{i-1}.
$$

(4.36)

Therefore, the perturbation at last step is

$$
\delta Z^i = \frac{\partial H^N}{\partial Z^{N-1}}|_{Z^{N-1} = F_{n-1} \circ H^i \circ \ldots \circ H^1|_{Z^1 = H(Z^0)} \frac{\partial H^1}{\partial Z^0} \delta Z^0.
$$

(4.37)

The Jacobian is defined by

$$
A(Z^0) = \frac{\partial H}{\partial Z^0},
$$

(4.38)

which can be read off from equation (4.37) as

$$
\frac{\partial H}{\partial Z^0} = \frac{\partial H^N}{\partial Z^{N-1}}|_{Z^{N-1} = F_{n-1} \circ H^i \circ \ldots \circ H^1|_{Z^1 = H(Z^0)} \frac{\partial H^1}{\partial Z^0}.
$$

(4.39)

In matrix multiplication form, it can be expressed as

$$
A(Z^0) = A^N(Z^{N-1}) \cdot A^{N-1}(Z^{N-2}) \cdot \ldots \cdot A^1(Z^0)
$$

(4.40)

where

$$
A^i(Z^{i-1}) = \frac{\partial H^i}{\partial Z^{i-1}}|_{Z^{i-1} = F_{n-1} \circ H^i}.
$$

(4.41)
4.1.3.2 Numerical adjoint code construction

The adjoint of the Jacobian in matrix form is simply

\[
\left( \frac{\partial H}{\partial Z^0} \right)^* = (A^1)^* \cdots (A^N)^* \]

\[
= \left( \frac{\partial H^1}{\partial Z^0} \right)^* \cdots \left( \frac{\partial H^N}{\partial Z^{N-1}} \right)^* |_{Z^{N-1} = \Pi_{i=1}^{N-1} \circ H^i} .
\]  

(4.42)

The adjoint state calculation can be formulated using the adjoint state variables. Introduce the adjoint state variables \( DZ^{N-1}, \ldots, DZ^0 \), which satisfy the inner product relation

\[
< DZ^{i-1}, \delta Z^{i-1} > = < DZ^i, \delta Z^i > .
\]  

(4.43)

This can be further derived

\[
< DZ^i, \delta Z^i > = < DZ^i, A^i(Z^{i-1}) \delta Z^{i-1} >
\]

\[
= < (A^i(Z^{i-1})^* DZ^i, \delta Z^{i-1} > .
\]  

(4.44)

(4.45)

This holds for all \( \delta Z^{i-1} \), so it follows that

\[
DZ^{i-1} = (A^i(Z^{i-1}))^* DZ^i .
\]  

(4.46)

We use Equation 4.46 as the rule to perform one step of an adjoint calculation.

Numerical adjoint code construction obeys certain rules. These rules apply in order to convert computer programs, without any aliasing, to the mathematical expressions we have just derived. In this section, we shall discuss three commonly used
programming semantics, loops, conditionals and assignments and their corresponding adjoints.

**Loops**

We differentiate loops into sequential loops and parallel loops. For convenience, we use C style pseudo-code to express a loop $L$:

\[
\text{for } i = 0 : N \\
L_i \\
\text{end } i
\]  

(4.47)

where $L_i$ denotes the statements inside the loop indexed by $i$. If the iterate $L_i$ depends on the output of $L_{i-1}$, we called it a sequential loop, otherwise, a parallel loop. The output of a parallel loop does not depend on the order in which the iterations are executed. The adjoint of a parallel loop is a loop of the same bounds without particular execution order for the iteration. It can be expressed as $L^*$:

\[
\text{for } i = 0 : N \\
L_i^* \\
\text{end } i
\]  

(4.48)

where $L_i^*$ represents the adjoint statement for the $i$th iterate. The adjoint of a sequential loop, however, by equation (4.42), has to be executed in the reverse order, $L^*$:

\[
\text{for } i = N : 1 : -1 \\
L_i^* \\
\text{end } i
\]  

(4.49)
Conditionals

Conditional statements measure the boolean values of certain conditions. The conditional statement itself does not have an adjoint. The statements branched by conditionals are to be taken adjoint. Consequently, the boolean values of the conditions must be known in the adjoint code in order to decide which branch of the conditional has to be taken adjoint. This procedure is shown in the following table:

<table>
<thead>
<tr>
<th>forward code</th>
<th>adjoint code</th>
</tr>
</thead>
<tbody>
<tr>
<td>if (condition $C_1$)</td>
<td>if (condition $C_1$)</td>
</tr>
<tr>
<td>$L_1$</td>
<td>$L_1^*$</td>
</tr>
<tr>
<td>else if (condition $C_2$)</td>
<td>else if (condition $C_2$)</td>
</tr>
<tr>
<td>$L_2$</td>
<td>$L_2^*$</td>
</tr>
<tr>
<td>else (condition $C_3$)</td>
<td>else (condition $C_3$)</td>
</tr>
<tr>
<td>$L_3$</td>
<td>$L_3^*$</td>
</tr>
</tbody>
</table>

Table 4.1  Forward and adjoint conditional statements.

Assignments

Assignment has two types, passive and active. Passive assignment in step $i$ can be formulated as

$$x^i = I x^{i-1}$$  \hspace{1cm} (4.50)

where $I$ is the identity map, and $x^{i-1}$, $x^i$ are the variables before and after the assignment, respectively. Passive assignment happens when a variable does not change its
value during the execution of the step. On the other hand, active assignment modifies the value of a variable, which can be formulated as

\[ x^i = f(x^{i-1}, \cdot) \]  

(4.51)

for some function \( f \). In the example of equation (4.33), \( x^1 = x^0 \) is passive whereas \( y^1 = x^0 \) is active. Adjoint of assignment on a linearized code results directly from equation (4.46) by taking the complex transpose of \( A^i(Z^{i-1}) \).

**Example**

To show a non-trivial example of adjoint code construction, let’s consider a code which correlates two scalar fields separated by a sequence of offsets, and then accumulate over frequency \( \omega \),

\[ I(z, x, h) = \sum_\omega R(z, x - h, \omega)S(z, x + h, \omega) \]  

(4.52)

where \( z \) and \( x \) are vertical and horizontal coordinates, respectively, and \( h \) is the horizontal offset between two scalar fields \( R \) and \( S \). A linear perturbation of equation (4.52) yields

\[ \delta I(z, x, h) = \sum_\omega \{ \delta R(z, x - h, \omega)S(z, x + h, \omega) + R(z, x - h, \omega)\delta S(z, x + h, \omega) \} \]  

(4.53)
A pseudo-code for the summation is written as follows

\[
\text{for } w = \text{min} : dw : w_{\text{max}} \\
\text{for } z = z_{\text{min}} : dz : z_{\text{max}} \\
\quad \text{for } x = x_{\text{min}} : dx : x_{\text{max}} \\
\qquad \text{for } h = h_{\text{min}} : dh : h_{\text{max}} \\
\quad \quad \text{if } (x+h \text{ and } x-h \text{ do not lie outside of boundary)} \quad \text{then} \\
\quad \quad \quad \delta I(z, x, h)^+ = \delta R(z, x - h, \omega)S(z, x + h, \omega) + R(z, x - h, \omega)\delta S(z, x + h, \omega) \\
\quad \quad \text{end } h \\
\quad \quad \text{end } x \\
\quad \text{end } z \\
\text{end } \omega
\]

(4.54)

Each inner most iterate under the conditional written in matrix form is

\[
\begin{pmatrix}
\delta R^i(z, x - h, \omega) \\
\delta S^i(z, x + h, \omega) \\
\delta I^i(z, x, h)
\end{pmatrix}
= \begin{pmatrix}
I & 0 & 0 \\
0 & I & 0 \\
S(z, x + h, \omega) & R(z, x - h, \omega) & I
\end{pmatrix}
\begin{pmatrix}
\delta R^{i-1}(z, x - h, \omega) \\
\delta S^{i-1}(z, x + h, \omega) \\
\delta I^{i-1}(z, x, h)
\end{pmatrix}
\]

(4.55)

The corresponding adjoint is

\[
\begin{pmatrix}
DR^{i-1}(z, x - h, \omega) \\
DS^{i-1}(z, x + h, \omega) \\
DI^{i-1}(z, x, h)
\end{pmatrix}
= \begin{pmatrix}
I & 0 & S(z, x + h, \omega) \\
0 & I & R(z, x - h, \omega) \\
0 & 0 & I
\end{pmatrix}
\begin{pmatrix}
DR^i(z, x - h, \omega) \\
DS^i(z, x + h, \omega) \\
DI^i(z, x, h)
\end{pmatrix}
\]

(4.56)

We also noticed that the loop over \( \omega, z, x \) and \( h \) are parallel, so the adjoint code is written as follows
for \( \omega = \text{min} : d\omega : \omega_{\text{max}} \)
for \( z = z_{\text{min}} : dz : z_{\text{max}} \)
for \( x = x_{\text{min}} : dx : x_{\text{max}} \)
for \( h = h_{\text{min}} : dh : h_{\text{max}} \)
if (x+h and x-h do not lie outside of boundary) then
\[
DR^{i-1}(z, x - h, \omega) + = S(z, x + h, \omega)DI^i(z, x, h) \\
DS^{i-1}(z, x + h, \omega) + = R(z, x - h, \omega)DI^i(z, x, h)
\] (4.57)
end if
end h
end x
end z
end \omega
4.1.3.3 A discrete formulation for differential migration and its adjoint

Let $i$ be the index for depth, $j$ be the source index, and $h$ be offset index. The imaging routine can be written as

\[
\text{for } i = 1 : n \\
\quad \text{for } j = 1 : n_s \\
\quad \quad S_j^i = H[c^{i-1}]S_j^{i-1} \\
\quad \quad R_j^i = H[c^{i-1}]R_j^{i-1} \\
\quad \quad \text{for } h = (-H, -H) \ldots (H, H) \\
\quad \quad \quad I_h^i = I_h^i + S_j^i(x + h)R_j^i(x - h) \\
\quad \quad \text{end } h \\
\quad \text{end } j \\
\text{end } i
\]

(4.58)

Here $H[c^{i-1}]$ is the downward propagator dependent on the velocity of $i - 1$th depth level for both receiver wavefield and the complex conjugated source wavefield. The perturbation $\delta I_h$ should satisfy the perturbation of imaging routine,

\[
\text{for } i = 1 : n \\
\quad \text{for } j = 1 : n_s \\
\quad \quad \delta S_j^i = H[c^{i-1}]\delta S_j^{i-1} + (\frac{\partial}{\partial c} H[c^{i-1}]S_j^{i-1})\delta c^{i-1} \\
\quad \quad \delta R_j^i = H[c^{i-1}]\delta R_j^{i-1} + (\frac{\partial}{\partial c} H[c^{i-1}]S_j^{i-1})\delta c^{i-1} \\
\quad \quad \text{for } h = (-H, -H) \ldots (H, H) \\
\quad \quad \quad S_j^i(x + h) = T(h)S_j^i \\
\quad \quad \quad R_j^i(x - h) = T(-h)R_j^i \\
\quad \quad \quad I_h^i = I_h^i + \delta S_j^i(x + h)R_j^i(x - h) + S_j^i(x + h)\delta R_j^i(x - h) \\
\quad \text{end } h \\
\quad \text{end } j \\
\text{end } i
\]

(4.59)

Here $T(h)$ is a translate operator, $T(h)S(x) = S(x + h)$. We distinguish the uses of
\( (\cdotp) \) and \( (\cdotp)^* \). For the former, it means an operator with coefficients being complex conjugate to the coefficients of the operator \( (\cdotp) \), whereas for the later, it means an adjoint operator with respect to the operator \( (\cdotp) \). In order to derive the adjoint of the perturbation imaging routine, we first simplify equations (4.59) as a loop consists of 2 parts.

\[
\begin{align*}
\text{for } i = 1 : n \\
\text{for } j = 1 : n_s \\
A \\
B \\
\text{end } j \\
\text{end } i 
\end{align*}
\]

and note that the loop over source index \( j \) is parallel embedded in a sequential loop indexed by \( i \). The adjoint routine should be, at simplified level, written as

\[
\begin{align*}
\text{for } i = n, n - 1, \ldots, 1 \\
\text{for } j = 1 : n_s \\
B^* \\
A^* \\
\text{end } j \\
\text{end } i 
\end{align*}
\]

Expressions \( B \) are expanded as,

\[
\begin{align*}
\text{for } h = (-H, -H), \ldots, (H, H) \\
C \\
\text{end } h 
\end{align*}
\]

Notice that the loop over \( h \) is parallel, we write adjoint expressions \( B^* \) as,

\[
\begin{align*}
\text{for } h = (-H, -H), \ldots, (H, H) \\
C^* \\
\text{end } h 
\end{align*}
\]

We express \( A \) and \( C \) using matrix-vector product form. For \( C \), we have
\[
\begin{pmatrix}
\delta I_h^i \\
\delta S_j^i(x + h) \\
\delta R_j^i(x - h)
\end{pmatrix} =
\begin{pmatrix}
I & R_j^i(x - h) & S_j^i(x + h) & 0 & 0 \\
0 & 0 & 0 & T(h) & 0 \\
0 & 0 & 0 & 0 & I \\
0 & 0 & 0 & 0 & I
\end{pmatrix}
\begin{pmatrix}
\delta I_h^i \\
\delta S_j^i(x + h) \\
\delta R_j^i(x - h)
\end{pmatrix}
\]

(4.64)

We then formulate the adjoint expressions \( C^* \),

\[
\begin{pmatrix}
DI_h^i \\
DS_j^i(x + h) \\
DR_j^i(x - h) \\
\end{pmatrix} =
\begin{pmatrix}
I & 0 & 0 & 0 \\
\frac{R_j^i(x - h)}{S_j^i(x + h)} & 0 & 0 & 0 \\
0 & T(-h) & 0 & I \\
0 & 0 & T(h) & 0 & I
\end{pmatrix}
\begin{pmatrix}
DI_h^i \\
DS_j^i(x + h) \\
DR_j^i(x - h)
\end{pmatrix}
\]

(4.65)

or explicitly,

\[
\begin{align*}
DR_j^i &= T(h)DR_j^i(x - h) + DR_j^i \\
DS_j^i &= T(-h)DS_j^i(x + h) + DS_j^i \\
DR_j^i(x - h) &= \overline{S_j^i(x + h)DI_h^i} \\
DS_j^i(x + h) &= \overline{R_j^i(x - h)DI_h^i b}
\end{align*}
\]

(4.66)

Here we have used \( D \) to replace \( \delta \) to indicate the adjoint variables. Substitute equation (4.65) into equations (4.63), we have,

\[
B^* : \begin{cases}
DR_j^i &= \sum_h T(h)\{S_j^i(x + h)DI_h^i\} + DR_j^i \\
DS_j^i &= \sum_h T(-h)\{R_j^i(x - h)DI_h^i\} + DS_j^i
\end{cases}
\]

(4.67)

For expression \( A \), the matrix representation is

\[
\begin{pmatrix}
\delta S_j^i \\
\delta R_j^{i-1} \\
\delta S_j^{i-1} \\
\delta R_j^{i-1} \\
\delta c^{i-1}
\end{pmatrix} =
\begin{pmatrix}
0 & 0 & H[c^{i-1}] & 0 & \left( \frac{\partial}{\partial c} H[c^{i-1}] S_j^{i-1} \right) \\
0 & 0 & 0 & H[c^{i-1}] & \left( \frac{\partial}{\partial c} H[c^{i-1}] R_j^{i-1} \right) \\
0 & 0 & I & 0 & 0 \\
0 & 0 & 0 & I & 0 \\
0 & 0 & 0 & 0 & I
\end{pmatrix}
\begin{pmatrix}
\delta S_j^i \\
\delta R_j^{i-1} \\
\delta S_j^{i-1} \\
\delta R_j^{i-1} \\
\delta c^{i-1}
\end{pmatrix}
\]

(4.68)
The transposed expressions $A^*$ can be written

$$
\begin{pmatrix}
DS_i^j \\
DR_i^j \\
DS_i^{i-1} \\
DR_i^{i-1} \\
Dc^{i-1}
\end{pmatrix}
= 
\begin{pmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
(H[c^{i-1}])^* & 0 & I & 0 & 0 \\
0 & (H[c^{i-1}])^* & 0 & I & 0 \\
\left(\frac{\partial}{\partial c} H[c^{i-1}]S_j^{i-1}\right)^* & \left(\frac{\partial}{\partial c} H[c^{i-1}]R_j^{i-1}\right)^* & 0 & 0 & I
\end{pmatrix}
\begin{pmatrix}
DS_i^j \\
DR_i^j \\
DS_i^{i-1} \\
DR_i^{i-1} \\
Dc^{i-1}
\end{pmatrix}
(4.69)
$$

We then write explicitly,

$$
A^* : \begin{cases}
Dc^{i-1} = (\frac{\partial}{\partial c} H[c^{i-1}]S_j^{i-1})^* DS_j^i + (\frac{\partial}{\partial c} H[c^{i-1}]R_j^{i-1})^* DR_j^i + Dc^{i-1} \\
DR_j^{i-1} = (H[c^{i-1}])^* DR_j^i + DR_j^{i-1} \\
DS_j^i = (H[c^{i-1}])^* DS_j^i + DS_j^i
\end{cases}
(4.70)
$$

Insert expressions for $B^*$ and $A^*$ in equations (4.61), we have

$$
\text{for } i = n, n-1, \ldots, 1 \\
\text{for } j = 1 : n_x,

\begin{align*}
DR_j^i &= \sum_h T(h)\{S_j^i(x+h)DI_h^i\} + DR_j^i \\
DS_j^i &= \sum_h T(-h)\{R_j^i(x-h)DI_h^i\} + DS_j^i \\
Dc^{i-1} &= (\frac{\partial}{\partial c} H[c^{i-1}]S_j^{i-1})^* DS_j^i + (\frac{\partial}{\partial c} H[c^{i-1}]R_j^{i-1})^* DR_j^i + Dc^{i-1} \\
DR_j^{i-1} &= (H[c^{i-1}])^* DR_j^i + DR_j^{i-1} \\
DS_j^{i-1} &= (H[c^{i-1}])^* DS_j^i + DS_j^{i-1}
\end{align*}
\text{end } j
\text{end } i
(4.71)
$$

We then unroll the sequential loop by isolating the first iteration $i = n$ and arrive at
\begin{align*}
    i = n \\
    \text{for } j = 1 : n_s \\
    DR_j^i &= \sum_h T(h)\{S_j^i(x + h)DI_h^i\} \\
    DS_j^i &= \sum_h T(-h)\{R_j^i(x - h)DI_h^i\} \\
    Dc^{i-1} &= (\frac{\partial}{\partial c} H[c^{i-1}]S_j^{i-1})^* DS_j^i + (\frac{\partial}{\partial c} H[c^{i-1}]R_j^{i-1})^* DR_j^i \\
    \text{end } j \\
    \text{for } i = n - 1, \ldots, 1 \\
    \text{for } j = 1 : n_s \\
    DR_j^i &= \sum_h T(h)\{S_j^i(x + h)DI_h^i\} + (H[c^{i-1}])^* DR_j^{i+1} \\
    DS_j^i &= \sum_h T(-h)\{R_j^i(x - h)DI_h^i\} + (H[c^{i-1}])^* DS_j^{i+1} \\
    Dc^{i-1} &= (\frac{\partial}{\partial c} H[c^{i-1}]S_j^{i-1})^* DS_j^i + (\frac{\partial}{\partial c} H[c^{i-1}]R_j^{i-1})^* DR_j^i \\
    \text{end } j \\
    \text{end } i
\end{align*}

We can recognize from equation (4.72) that \( R_j^i \) and \( S_j^i \) are the discrete adjoint state receiver and source wavefields. Their corresponding full wavefield continuous formulations can be found in equation (4.25) and one-way formulation can be found in equation (4.31). one-way wavefields. The gradient is identified as \( Dc \).
4.2 Inversion

We have posed the problem of seismic velocity analysis as a nonlinear optimization problem

$$\min_{c \in \mathbb{C}^2} \frac{1}{2} \| P_h I_h \|^2$$  \hspace{1cm} (4.73)

where the semblance image gather $I_h(c)$ is nonlinearly dependent on velocity. At the global optimum velocity the semblance image becomes concentrated in offset domain and flat in angle domain, indicating the best image is achieved. This is the main difference between waveform type of inversion and inversions which directly pursue the quality of images; in waveform type of inversion the optimized velocity does not necessarily produce the optimum image.

Nonlinear optimization by gradient type of methods require that the objective function itself be smooth and have few local minima in order to avoid the solution getting trapped by a local minimum. The objective function constructed by differential semblance criterion varies smoothly with respect to changes in velocities. Its local minima usually coincide with the global minimum *Stolk et Symes, 2003*[38]. Therefore it is particular suitable for optimizations by gradient methods.

Methods for solving nonlinear optimization problems categorize into two strategies,
line search methods and trust region methods. In both strategies, information of the gradient of the objective function play a key role. In last section, we focused on the calculation of the gradient. In this section, we shall explain two things. One is the application of the L-BFGS algorithm that uses the calculated gradient for optimization, and the other, the smoothness problems we face on using the gradient and how the gradient is smoothed: the B-spline smoothing scheme

4.2.1 Limited memory BFGS method

We use limited memory BFGS method, Nocedal & Wright, 2000[25], to directly optimize the objective function equation (4.73). At each iteration, the BFGS search direction \( p \) minimizes a quadratic model of the objective function at the current iterate.

\[
m_k(p) = J_k + \nabla_c J^t p + \frac{1}{2} p^t H_k p
\]  

(4.74)

Here \( H_k \) is a positive definite matrix that will be updated at every iteration. We can write the minimizer \( p \) of this quadratic model explicitly as

\[
p_k = -H_k^{-1} \nabla_c J
\]  

(4.75)

and the new iterate is chosen by line search

\[
c_{k+1} = c_k + \alpha_k p_k
\]  

(4.76)

to satisfy the sufficient decrease and curvature conditions. The particular formula to
update $H_k^{-1}$ (Nocedal & Wright, 2000[25]) is the defining formula for BFGS method. We emphasize the fact that the BFGS update is not obtained by solving a linearized equation at the current iterate, which is a method frequently used in geophysical inversions. Optimizing by quadratic match at each iteration instead of solving a linearized equation greatly reduces the number of line searches, and therefore significantly reduces the computational cost.

The problem with BFGS method is that the inverse Hessian matrix $H_k^{-1}$ may be dense and require fairly large storage and computational cost. The limited memory BFGS method modify the BFGS method so that the inverse Hessian can be stored compactly in just a few vectors of length $n$, where $n$ is the number of unknown model parameters. Once a new iterate is computed, the old vector is deleted and replaced by the new ones obtained from the current step. Practical experience suggests that $n$ between 3 and 20 produces satisfactory results. The limited memory BFGS algorithm can be stated formally as follows Nocedal & Wright, 2000 [25].

Choose starting point $c_0$ and $m > 0$
$k ← 0$
Choose $H_0^{-1}$
repeat
  Compute $p_k ← -H_k^{-1}\nabla_c J$
  Line search $c_{k+1} = c_k + \alpha_k p_k$
  if $k > m$
    Discard the vector pair $(s_{k-m}, y_{k-m})$
  Compute and save $s_k ← x_{k+1} - x_k, y_k = \nabla_c J_{k+1} - \nabla_c J_k$
  $k ← k + 1$
  Compute $H_k^{-1}$
u nit until convergence criteria are satisfied
4.2.2 Smoothing properties of the gradient

A given velocity model paired with the observed data determines a certain coverage of bicharacteristic curves (ray coverage). The basic claim of this section is that the gradient with respect to the velocity of the differential semblance objective function is smooth along the bicharacteristic curves (rays) that correspond to the given velocity and observed data, and the gradient is, in general, not smooth across such rays. The analysis follows from the closed form formula of the gradient given in section (4.1.1).

We rewrite the gradient as

\[
\nabla_c J(y) = \int G^+(x_r, y, w) \frac{u(x, h)\omega^2}{c(y)^2} G^+(y, x - h, w) \times
\]

\[d(x_r, x_s, t)G^+(x + h, x_s, w)e^{i\omega t} dx_r dx_s dt dw dx dh\]

\[+ \int G^+(x_r, x - h, w)d(x_r, x_s, t)G^+(x + h, y, w) \times\]

\[\frac{u(x, h)\omega^2}{c(y)^2} G^+(y, x_s, \omega)e^{i\omega t} dx_r dx_s dt dw dx dh\]

(4.78)

where \(u(x, h) = h^2 I_h(x, h)\). Recall that the data can be explained by the generalized Born modeling

\[d(x_r, x_s, t) = \int G^+(x + h, x_s, \omega)r(x, h)G^+(x - h, x_r, \omega)e^{i\omega t} dxdhd\omega\]

(4.79)
Let’s label the data due to a single reflection at \((x, h)\) from a point source located at \(x_s\) and received at \(x_r\) by \(d(x_r, x_s; x, h)\). We have

\[
d(x_r, x_s, t; x, h) = \int G^+(x + h, x_s, \omega)r(x, h)G^+(x - h, x_r, \omega)e^{i\omega t} d\omega
\]

(4.80)

This datum of single reflection corresponds to the union of two ray paths

\[
\Gamma = \gamma(x_r, x - h) \cup \gamma(x_s, x + h)
\]

where \(\gamma(x_r, x - h)\) and \(\gamma(x_s, x + h)\) are ray paths of \(G^+(x-h, x_r, \omega)\) and \(G^+(x+h, x_s, \omega)\), respectively. The objective function gradient corresponding to this datum of single reflection

\[
\nabla_c J(y; x_r, x_s, x, h) = g_1(y; x_r, x_s, x, h) + g_2(y; x_r, x_s, x, h)
\]

where

\[
g_1 = \int G^+(x_r, y, w)\frac{u(x, h)\omega^2}{c(y)^2}G^+(y, x - h, w)d(x_r, x_s, t; x, h)G^+(x + h, x_s, w)e^{i\omega t} dtd\omega
\]

(4.81)

\[
g_2 = \int G^+(x_r, x - h, w)d(x_r, x_s, t)G^+(x + h, y, w)\frac{u(x, h)\omega^2}{c(y)^2}G^+(y, x_s, \omega)e^{i\omega t} dtd\omega
\]

(4.82)

We shall see that \(g_1\) is smoothly varying if \(y\) moves along the ray path \(\gamma(x_r, x - h)\).

Similarly, \(g_2\) varies smoothly if \(y\) moves along the ray path \(\gamma(x_s, x + h)\). I will give
an analysis of the first term in detail. A parallel analysis applies to the second term by symmetry.

Using the asymptotic form of the Green's function,

\[ G^+(x, y, \omega) = a(x, y)e^{-i\omega \Phi(x, y)}, \quad (4.83) \]

we can write the data of single reflection in its asymptotic form as

\[ d(x_r, x_s, t) = \int a(x_r, x - h, x + h, x_s) r(x, h) e^{i\omega(t - \Phi(x + h, x_s) - \Phi(x_r, x_h))} d\omega \quad (4.84) \]

With equations (4.83) and (4.84), \( g_1 \) can be written as

\[
g_1 = \int \frac{\omega^2}{c(y)^2} r(x, h) a(x_r, x - h, x - h, x_s) a(x_r, y) a(y, x - h) a(x + h, x_s) u(x, h) \times e^{i\omega t} e^{i\omega'(t - \Phi(x + h, x_s) - \Phi(x_r, x_h))} e^{-i\omega \Phi(x_r, x_s)} e^{-i\omega \Phi(y, x_h)} e^{-i\omega \Phi(x_r, x_h)} dt d\omega d\omega' \]

\[
= \int \frac{\omega^2}{c(y)^2} r(x, h) a(x_r, x - h, x - h, x_s) a(x_r, y) a(y, x - h) a(x + h, x_s) u(x, h) \times 2\pi e^{i\omega(\Phi(x_r, x_h) + \Phi(x + h, x_s) - \Phi(x_r, y) - \Phi(y, x_h) - \Phi(x + h, x_s))} d\omega \]

\[
= \int \frac{\omega^2}{c(y)^2} r(x, h) a(x_r, x - h, x - h, x_s) a(x_r, y) a(y, x - h) a(x + h, x_s) u(x, h) \times 2\pi e^{i\omega(\Phi(x_r, x_h) - \Phi(x_r, y) - \Phi(y, x_h))} d\omega \]

We can write the integral in the form

\[ g_1(y; x_r, x_s, x, h) = \int A(x_r, x_s, x, h, \omega) e^{i\Phi(x_r, y, x_h, \omega)} d\omega \quad (4.85) \]
where

\[ A(x_r, x_s, x, h, \omega) = \frac{\omega^2}{c^2} T(x, h)a(x_r, x - h, x + h, x_s)a(x_r, y)a(y, x - h)a(x + h, x_s)u(x, h) \]

is a symbol* of type \( S^2_{1,0} \) and

\[ \Phi(x_s, y, x - h, \omega) = \omega(\phi(x_r, x - h) - \phi(x_r, y) - \phi(y, x - h)) \]

When \( y \) varies along the ray curve \( \gamma(x_r, x - h) \), the total phase \( \Phi \) is stationary with respect to \( \omega \),

\[ \frac{\partial \Phi(x_r, y, x - h, \omega)}{\partial \omega}|_{y \in \gamma(x_r, x-h)} = [\phi(x_r, x - h) - \phi(x_r, y) - \phi(y, x - h)]_{y \in \gamma(x_r, x-h)} = 0 \] (4.86)

Theories of oscillatory integral (Theorem 3.5 in Joshi[19]) imply that \( g_1 \) is singular when \( y \in \gamma(x_r, x-h) \). We see singularities of \( g_1 \) lie on the section of ray path \( \gamma(x_r, x-h) \). Taking the derivative of the gradient along the bicharacteristic curve \( \gamma(x_r, x-h) \) with respect to \( y \) means to differentiate equation (4.85) subject to equation (4.86).

The differential operator \( \partial_y \) applies to amplitude terms only. Since the amplitude terms satisfy the first transport equation, the derivatives are smooth. We have thus established the claim in the beginning of this section that the gradient is smooth.

---

* \( S^{m}_{\rho,\delta}(X \times \mathbb{R}^N) \) is the space of symbols of order \( m \) and of type \((\rho,\delta)\). It consists of all \( a(x, \theta) \in C^m(X \times \mathbb{R}^N) \) such that for all compact \( K \subset X \) and all multi-index \( \alpha \in \mathbb{N}^n \), there is a constant \( C = C_{K,\alpha,\delta}(a) \) such that \( |\partial_x^\alpha \partial_\theta^\beta a(x, \theta)| \leq C(1 + |\theta|)^{m-\rho|\beta|+\delta|\alpha|} \) for \( (x, \theta) \in K \times \mathbb{R}^N \). We adopt this definition of symbol space from standard references of pseudo-differential operators, i.e. Grigs & Sjöstrand, 1994, [15]. We see \( A \in S^2_{1,0}(X \times \mathbb{R}^N) \) for \( N = 1 \).
Figure 4.1  The ray path connecting from $x_s$ to $x + h$, $x - h$ to $y$ and $y$ to $x_r$, where $G^+(x, y, \omega)$ is the Green's function observed at $x$ emanated at $y$. The adjoint of differential migration picks up an image residual and smoothly distributed it along the corresponding ray paths.

along the bicharacteristic curves and non-smooth (singular) across the bicharacteristic curves.

Parallel analysis can be applied to the second term in gradient formulation. The second term of gradient $g_2$ is singular for $y \in \gamma(x_s, x + h)$. Similar conclusions on smoothness follow accordingly for $y \in \gamma(x_s, x + h)$. The total gradient will have its singularities distributed along the curve $\Gamma = \gamma(x_r, x - h) \cup \gamma(x_s, x + h)$. In view of
gradient formulated as adjoint of differential migration applied to the residual image, the operator of adjoint differential migration picks up an image residual defined in \((x, h)\) domain and smoothly distributes it along the corresponding ray paths. The analysis laid in this section suggests that the smooth property of the gradient is determined by the ray coverage. In regions where there is sparse ray coverage, the gradient may show non-smoothness behavior.

Velocity updates using search directions constructed by non-smooth gradients are therefore non-smooth. It thus violates the assumption we have on the smoothness of the velocity function and yields unstable inversion results.
4.2.3 Problem

How to use a non-smooth gradient to produce smooth velocity updates is another key component in this thesis work. A smoothing scheme is necessary. The choices of which smoothing schemes to use is motivated by the following two considerations: first, to represent a smooth velocity function, the basis functions on cartesian grids are redundant. The velocity function should be parameterized by fewer parameters than the number of grid points on which the velocity is used for imaging. Second, variations of velocities are local in space. Therefore we look for smooth basis functions that have compact support in space to decompose the velocity function. B-spline basis functions of order greater than one are smooth and compactly supported. Velocity functions decomposed by proper B-spline basis functions are necessarily smooth, and therefore provides a good smoothing scheme. Write the decomposition of velocity \( c \) as

\[
c = Bm
\]  

(4.87)

where \( c \) is understood as the velocity in its image space representation with image space basis functions defined as cartesian grid basis functions, \( B \) is the matrix of B-spline basis functions represented in image space basis functions, and \( m \) is the vector of B-spline coefficients. The gradient of the image space velocity with respect to B-spline parameters is
\[
\frac{\partial c}{\partial m} = B^T
\]  \hspace{1cm} (4.88)

and the relation between the gradient of the objective function with respect to the
image space velocity and gradient with respect to B-spline parameters is seen from

\[
\frac{\partial J}{\partial m} = \frac{\partial c}{\partial m} \frac{\partial J}{\partial c} = B^T \frac{\partial J}{\partial c}
\]  \hspace{1cm} (4.89)

that they satisfy

\[
\nabla_m J = B^T \nabla_c J
\]  \hspace{1cm} (4.90)

Where \(B^T\) is the adjoint of the B-spline projection. It is guaranteed from the prop-
erties of B-spline basis functions that

\[
B \nabla_m J = B B^T \nabla_c J
\]  \hspace{1cm} (4.91)

is smooth, implying that the velocity update is smooth. Equation (4.90) suggest that
instead of optimizing the image space velocity subject to smoothing constraint

\[
\min_{c \in C^2} J(c)
\]  \hspace{1cm} (4.92)

we can optimize its B-spline coefficients

\[
\min_m J(Bm)
\]  \hspace{1cm} (4.93)

using third B-spline basis functions of degree 3. A complete optimization routine on
B-spline parameters is shown as follows
input: $m$, B-spline model parameter vector

$\mathbf{c} = B\mathbf{m}$

$J = \frac{1}{2} || P_h I ||^2$

$\nabla_c J = \left( \frac{\partial L}{\partial c} \right)^* P_h^* P_h I$

$\nabla_m J = B^* \nabla_c J,$

construct BFGS update $\delta m$

line search $m \leftarrow m + \alpha \delta m$

output: optimized velocity, optimized image

**Figure 4.2** Inversion procedure. Forward project the B-spline model parameters to obtain the current velocity in the image space through $\mathbf{c} = B\mathbf{m}$, then evaluate the objective function, calculate the gradient with respect to the velocity in the image space, project this gradient to the model space through the adjoint of the B-spline projection to obtain the gradient with respect to the B-spline model parameters. The search direction is constructed by the L-BFGS method in the model space.
Chapter 5
Data examples

Chapter synopsis

In this chapter four synthetic and real data examples are studied. Properties of image
gathers in offset and the application problems associated with the differential semblance velocity analysis are analyzed. Questions of what the gather should look like
at the correct velocity and what is necessary in data preprocessing in order to obtain
clean image gathers are addressed in example I. Example II studies the robustness
of the method presented in this thesis in response to various degree of the nonlinear
effect due to low velocity perturbations; a sequence of low velocity lens with increasing refracting strength are tested for velocity inversions. The differential semblance velocity analysis is applied to real seismic data collected at Hill Air Force Base in
example III, where reasonable results are obtained. Example IV is the application of
differential semblance velocity analysis to Marmousi dataset. Shallow velocity struc-
tures are shown well reconstructed by differential semblance optimization. Problems
of inversions of complex velocity structure are analyzed in the discussion of section
(5.5).

Examples used in this chapter are 2 dimensional. We will use z to denote for vertical
coordinate and x for horizontal coordinate through out this chapter.
5.1 Offset image gathers due to flat reflectors and constant velocity

In this section we study a very simple synthetic example, seismic reflections due to a constant velocity with flat reflectors. We will see that the offset gathers migrated using the correct velocity do not focus perfectly at the zero offset. Reasons for this non-focusing phenomenon will be analyzed. Proper pre-processing treatments are employed to reduce the non-focusing phenomenon.

5.1.1 Synthetic data generation

We generate 2-D synthetic data using a constant background velocity \( c(z, x) = 2 \text{km/sec} \) with six horizontal reflectors marked by a 5% velocity perturbation as shown in Figure (5.1(a)). Signal responses to the Ricker wavelet with peak frequency at 18 Hz are simulated through finite difference time domain simulation [14] at about 10 grid points per wavelength for the highest frequency. Sources and receivers are in a fixed receiver array are evenly distributed on the surface from \( x = 0.1 \text{ km} \) to \( x = 3.9 \text{ km} \) incremented using the same interval \( \Delta x = 0.02 \text{ km} \). The absorbing boundary condition described in appendix (A) is applied to remove free surface multiples and reflections from the computational boundaries. One source gather from the middle of the surface with the direct arrival removed is shown in Figure (5.1(b)).
(a) Flat reflectors at constant velocity. Sources and receivers are evenly distributed on the surface.

(b) Source gather at x = 2.0 km.

**Figure 5.1** Synthetic data generated at flat reflectors and constant velocity.
5.1.2 Preprocessing

The one-way wave equation usually can not handle waves that propagate of high angles with respect to the vertical direction. The imaginary vertical wavenumber induced by the one-way wave equation at high angles produces evanescent waves with exponentially decaying or exploding amplitudes. The main purpose of preprocessing is to remove horizontally propagated waves observed on the surface in order to avoid evanescent energy during wavefield extrapolation. Ideally speaking, evanescent filtering should be applied to the extrapolated wavefields at each depth. However, an evanescent filter applied on the surface data once has shown to be adequate to remove evanescent energies for practical purposes.

The evanescent filter on the surface will remove, for each $\omega$, the energy in the data corresponds to the horizontal wavenumber $\frac{\omega}{c_r} \sin(\theta)$, where $\theta$ is the maximum propagation angle and $c_r$ is the velocity at the surface of data acquisition. To explain this evanescent filtering scheme, we first apply the time domain Fast Fourier transform to seismograms trace by trace.

$$d(x_r, x_s, \omega) = \int d(x_r, x_s, \omega) e^{-i\omega t} dt \quad (5.1)$$

Let $\phi$ be a smooth cut-off function in wave number domain, the evanescent filtering scheme is represented as a low pass filter

$$D(x_r, x_s, \omega) = \mathcal{F}^{-1}_{x_r} \phi \mathcal{F}_{x_r}(d) \quad (5.2)$$
where
\[
\mathcal{F}_{x_r}(d(x_r, x_s, \omega)) = \int d(x_r, x_s, \omega) e^{-ik \cdot x_r} dx_r
\]
and
\[
\phi(k) = \begin{cases} 
0 & |k| > k_1 |k| > \frac{\omega}{c_r} \sin \theta \\
1 & |k| < k_2 |k| < \rho \frac{\omega}{c_r} \sin \theta, \quad 0 < \rho < 1 \\
\cos^2\left(\frac{\pi/2(k-k_2)}{k_1-k_2}\right) & \text{elsewhere}
\end{cases}
\]

\(\theta\) is the angle between direction of wave propagation and the vertical direction. \(\theta = 90^\circ\) corresponds to horizontally propagated waves. The upper threshold \(\sin(\theta)\omega/c\) is the projection of the total wavenumber onto the surface plane, where \(c_r\) is used as the surface velocity and \(\theta\) is chosen to be 75°. Energies associated with propagating waves at angle higher than 75° are annihilated for all frequencies on the surface. To minimize frequency aliasing a smooth cosine taper is applied in the range \(\rho \sin(\theta)\omega/c < k < \sin(\theta)\omega/c\) with \(0 < \rho < 1\). The value of \(\rho\) hardly has any influence to the kinematics of the image as shown in Figure (5.2). However, it influences the quality of offset image gathers significantly according to Figure (5.3), where preprocessed data corresponds to \(\rho = 0.9\) is migrated at the correct velocity showing non-negligible noise at nonzero offset. It is important to remove such noises because noise in the image residuals will be projected onto the gradient which leads to erroneous velocity updating.
Figure 5.2  Effects of data preprocessing on images. (a). Images at the correct velocity correspond to preprocessed data using $\rho = 0.9$. (b). Images at the correct velocity correspond to preprocessed data using $\rho = 0.5$. 
Figure 5.3  Effects of data preprocessing on offset gathers. (a). Offset gathers at the correct velocity correspond to preprocessed data using $\rho = 0.9$. (b). Offset gathers at the correct velocity correspond to preprocessed data using $\rho = 0.5$. 
The noise at nonzero offset can be explained due to spatial high frequency aliasing induced by the evanescent filter. Such noise can be reduced by lowering the $\rho$ value in the preprocessing. Figure (5.3) shows offset gathers migrated at the correct velocity using a preprocessed data corresponds to $\rho = 0.5$. Although trade-offs of less spatial bandwidth resolutions exhibit comparing (a) and (b) in Figure (5.3), it shows that the gathers of $\rho = 0.5$ become much cleaner on nonzero offsets.

When the preprocessed data projected back to space-time domain, it clearly shows that the scheme of $\rho = 0.9$ introduces significant spatial high frequency aliasing as compared to the case of $\rho = 0.5$. See Figure (5.4). Large amount of near horizontally propagated waves are also muted by evanescent filter in Figure (5.4(b)).
Figure 5.4  Preprocessed data using $\rho = 0.9$ and $\rho = 0.5$ projected back into the $(x,t)$ domain. Two source gathers are shown for sources located in the middle of the model.
5.1.3 Effects of acquisition geometry

Offset image gathers migrated through the correct velocity are not perfectly concentrated at zero offset even with the noise (as discussed in last section) removed. An offset gather \( I_h(z, x, h)|_{x=2km} \) migrated at the correct velocity shown in Figure 5.5 demonstrates that the gather has a “X” pattern around the zero offset.

Figure 5.5 An “x” pattern in offset gathers obtained at the correct velocity.

This phenomenon can be explained by the relationship between offset domain images and the angle domain images as described in Chapter 3. The image gather shown in
Figure (5.5) is an integral over \( x_s \)

\[
I_h(z, x, h) = \int u(z, x, x_s, h)dx_s \tag{5.3}
\]

where \( u(z, x, x_s, h) \) is the product of the downward continued source wavefield complex conjugated and the downward continued receiver wavefield integrated over \( \omega \),

\[
u(z, x, x_s, h) = \int G^+(z, x + h, x_s, \omega) R(z, x - h, x_s, \omega) d\omega. \tag{5.4}\]

By a similar argument using plane wave decomposition as described in section (3.3.2), a slant stack of \( u(x, x_s, h) \) can be derived in 2-dimensions

\[
u_q = \int u(z + qh, x, x_s, h)dh \]

\[
= \int_{L(k_s, k_r, q) = 0} \tilde{G}^+(k_s, x_s) e^{ik_s x} \tilde{R}(k_r, x_s) e^{ik_r x} dk_r dk_s \tag{5.5}
\]

with

\[
L(k_s, k_r, q) = \frac{k_{rx} + k_{sz}}{k_{rx} - k_{sz}} - q = 0
\]

The meaning of equation (5.5) is understood as the integration of plane waves of \( S \) and \( R \) along the curve in wavenumber domain at \( \frac{k_{rx} + k_{sz}}{k_{rx} - k_{sz}} = q \). The stationary phase analysis indicates that this integral contributes one pair of rays that connect source, reflector and receiver by the Snell’s law. For a fixed source it follows that \( u_q \) is not rapidly decaying only when \( q = -\tan(\theta) \) where \( \theta \) is the scattering angle as illustrated in Figure (3.2). We demonstrate this phenomena by constructing a sequence of image offset gathers where each offset gather is constructed using one source only. Figure
(5.6) shows the sequence of offset gathers in the middle of the model each of which is obtained by migration of a single source gather. The slant trace in each offset gather provides direct evidence of the above analysis about properties of $u(z, x, x_s)$.

Offset gathers at the left most and right most panels present linear traces with highest dipping angles, which can be easily understood from the relation $q = -\tan(\theta)$. The offset gather $I_h(z, x, h)$ is expressed as an oscillatory integral of $u(z, x, x_s, h)$ over $x_s$. The stacking of all the offset gathers, in Figure (5.6) will cancel each other due to the oscillatory nature of the integrand except at the boundary of high dipping angles outside of which there is no images around $h = 0$. The “cut-off” in offset gathers of single source at high dipping angles is due to the limited acquisition geometry.
Figure 5.6  Offset gathers due to migration of single source gathers at correct velocity. The reflector become a straight line in each offset gather. The phenomenon is closely related to relationship between offset domain image and angle domain image.
5.2 Optimization of low velocity lens model

5.2.1 Introduction

Strongly refracting velocity model produces kinematic image artifacts in common bin migration due to multipathing Stolk \\& Symes, 2003 [38]. Wave equation migrations that use multi-bin data, such as shot record migration, are free of multipathing induced kinematic artifacts and therefore are ideal platforms for velocity analysis. In this section we study a series of simple examples where the medium contains a sequence of horizontal reflectors and a low velocity lens that leads to multipathing. Offset image gathers are constructed on which the offset domain differential semblance criterion is shown to be valid even for a very strong refracting lens model. Faithful velocity reconstruction is obtained by the wave equation migration based velocity analysis discussed in this thesis. All data used in this section are pre-processed using the evanescent filtering scheme as described in section (??).

5.2.2 Synthetic data generation and migration

Four 2-D synthetic data sets are generated using velocity models given by

\[ c(z, x) = 2 - \alpha e^{\frac{(z-0.3)^2 + (x-2)^2}{0.3}} \]  

(5.6)
for $\alpha = 0.4, 0.6, 0.8, 1.0$ where $\alpha$ is an amplitude coefficient that controls the refracting strength. Data are collected using the same full acquisition geometry as example I, with direct arrivals muted. Figure (5.7).

As the refracting coefficient $\alpha$ increases from 0.4 to 1.0, Figure (5.8), (5.9), (5.10) and (5.11), the data shows increased complexity due to refracting ray paths. For $\alpha = 0.4$ and $\alpha = 0.6$ the data demonstrate triplication arrivals but still identifiable as reflections due to horizontal reflectors. At $\alpha = 1.0$ the refracting ray paths distort the data significantly so that it is hard to be identified as reflection of flat reflectors as shown in Figure (5.11).

The effect of velocity to the imaging results is significant when the velocity is far away from the true one. The image obtained at constant velocity $c = 2\text{km/sec}$ show large errors, particularly for $\alpha = 1.0$. Images are shifted to deeper depths in the center of the model due to the low velocity lens at which the data is generated. Far away from the center of the model the migration velocity agrees well with the true velocity and therefore produces flat reflectors, in agreement with the true model. See Figure (5.12).

Strong signals at nonzero offset can be found in the center of the offset image gathers when migrated using the constant velocity. See Figure (5.12). At locations where the image is distorted the offset image gathers show large amplitude at nonzero offset.
Conversely, the offset image gathers are concentrated at zero offset when the flatness of the image is preserved as shown in Figure (5.13).

**Figure 5.7** Velocity model of 0.8km by 4km with a low velocity lens embedded in the middle. Sources and receivers are evenly distributed on the surface at the same spatial interval. Six horizontal reflectors are located at equal distance from $z = 0.12$ to $z = 0.78$. 
(a) $x_s = 1.8\text{km}$ source gather  
(b) $x_s = 2.0\text{km}$ source gather  
(c) $x_s = 2.2\text{km}$ source gather

**Figure 5.8** Source gathers obtained by acoustic simulation using velocity model equation (5.6) with refracting strength $\alpha = 0.4$. 
Figure 5.9 Source gather obtained by acoustic simulation using velocity model equation (5.6) with refracting strength $\alpha = 0.6$. Multipathing arrivals start to establish.
Figure 5.10  Source gathers obtained by acoustic simulation using velocity model equation (5.6) with refracting strength $\alpha = 0.8$. Multipathing arrivals are evident.
Figure 5.11  Synthetic source gathers obtained by acoustic simulation using velocity model equation (5.6) with refracting strength $\alpha = 1.0$ which generates strong multipathing arrivals. Maximum velocity perturbation is 50%.
Figure 5.12 Images obtained at constant velocity of 2 km/sec. Large imaging errors are evident, in particular for $\alpha = 1$. 
Figure 5.13  Offset gathers obtained at constant velocity of 2 km/sec. Nonzero offset amplitudes are significant near the lens region.
5.2.3 Velocity inversion

Velocity analysis by differential semblance optimization tries to minimize the image amplitude at nonzero offsets. Signals at nonzero offset are first amplified by multiplication with the offset parameter $h$ and then projected back to the velocity model by adjoint differential migration. This delineates the procedure for computation of the gradient with respect to velocity in the image space. A B-spline smoothing scheme introduced in section (4.2.3) can be applied in two steps. First project the image space gradient to the B-spline model space gradient by the adjoint B-spline projection $\nabla_{m} J = B^* \nabla_{x} J$. Second, interpolate the updated B-spline model parameters to obtain the image space velocities through forward B-spline projection $c = Bm$. The model updating by limited memory BFGS (L-BFGS) algorithm is carried on the B-spline model space. The same algorithm can be used with the image space using a self-adjoint low pass filter $BB^*$ to smooth the image space gradient $\nabla_c J$. The disadvantage of working with image space sampled models is that the it has to provide large memories required by the L-BFGS algorithm to store many vectors at a size of the image space gradient whereas working with the B-spline model space requires virtually very little storage.

Differential semblance optimization does not solve the linearization of the objective function which is essentially a Gauss-Newton’s method as used in many geophys-
ical inversions (Sava & Biondi, 2003[31], Gockenbach et al., 2001[14]), instead the objective function can be directly optimized by efficient quadratic matching at each iteration using the L-BFGS algorithm. The computational cost is greatly reduced because the number of line searches are reduced due to fast convergence of L-BFGS algorithm. Usually 2 to 3 iterations (each needs one line search) yield sufficient objective function decrease as shown in Figure (5.14). The stopping criterion depends on the desired accuracy with which the solution is to be found. The BFGS iteration terminates when

\[ \|g\| \leq \epsilon \max(1, \|m\|) \]

(5.7)

where \(\| \cdot \|\) is the Euclidean norm, \(\|g\|\) is the length of the gradient and \(\|m\|\) is the length of the model parameters.

Figure (5.14) shows the objective function values and the magnitude of the gradients through iterations of the optimization of the four data sets. The objective function values decay rapidly within the first few iterations. Output offset image gathers are shown to be well focused at zero offset in Figure (5.15) indicating the convergence has been reached. Comparing with the offset image gathers obtained at true velocity, we see the geometric difference is small as shown in Figure (5.16). Output images are also reconstructed to be flat (Figure (5.17) at the corresponding output velocities which are shown in good agreement to that of the true velocity. See Figure (5.19) and Figure (5.20). The inverted velocity models for each data set, Figure (5.19), agree well
with the corresponding true velocity models. The magnitude of the velocity difference between the inverted model and the true model is less then 5% of the background velocity for data sets of refracting strength $\alpha = 0.4$, 0.6 and 0.8.
Figure 5.14  Decay of objective function value and gradient magnitude. Values of objective function and magnitude of gradient at the correct velocity are shown by the dashed lines. Red and blue correspond to the objective function and the magnitude of gradient, respectively.
Figure 5.15  Offset image gather at the fifth iteration. Inverted offset gather are concentrated at the zero offset.
Figure 5.16  Offset image gathers migrated using the correct velocities. The comparison with Figure (5.15) show little geometrical differences.
Figure 5.17  Images obtained at the fifth iteration. Reconstructed images are reasonably flattened.
Figure 5.18  Images migrated using the correct velocities. The comparison with Figure (5.17) shows little differences.
Figure 5.19  Inverted velocities at the fifth iteration for various lens models. The amplitude and the shape of the lens are well reconstructed.
Figure 5.20  The differences between inverted velocities and the correct velocities. The magnitude of the differences is in general within 5% of the background velocity.
5.3 Optimization of HAFB high resolution data

5.3.1 Introduction

In July and August 2000 a field crew lead by Rice personnel conducted 3-D reflection experiments including two vertical seismic profiles and six check shot surveys at Hill Air Force Base Operable Unit 2 (OU-2). The surveys were designed for environmental characterization of a shallow (less than 20m) trichloro-ethene contaminated aquifer. The trichloro-ethene is a dense nonaqueous phase liquid (DNAPL). There are 5-20 meters of Quaternary sands, gravels and clays covering the Quaternary clay which is incised by ponds of DNAPL. The principal goal of the experiments is to characterize the base of the paleochannel through seismic exploration methods with resolution of about 40 centimeters to further aid anti-remediation efforts.

The 3-D seismic reflection experiments made use of Texan portable seismographs with sources generated by 223 caliber single shot rifles fired in 6 cm drill-holes. Acquisition geometry for the experiment is shown in Figure (5.21). For the VSP survey, surface shots fired every 0.7m in offsets of 21m were recorded by both receivers in two boreholes spaced at 0.5m and receivers on surface spaced at 0.35 apart. See Figure (5.21).
Differential semblance analysis is applied to the surface VSP data. The 2-D optimized images show improvement in correlation with velocity patterns. The velocity inversion of differential semblance optimization shows good agreement with results obtained by surface-bore-hole diffraction tomography.

5.3.2 Results from previous studies

The 223 rifle shot produces a broad bandwidth (50-350Hz) signal of large amplitude. However, the VSP data contends no significant energy for frequencies greater than 200Hz. Previous studies have shown the velocity increases rapidly from 200m/sec on the surface to 1000m/sec at depth of 15m deep in the model. The average velocity from surface to 15m in depth is estimated at 500m/sec. For VSP imaging targets at scales of 5m in a model with average velocity 500m/sec, the survey is considered as a low resolution survey. Previous imaging studies did not provide identification of continuous reflectors in shallow regions. Images from a 3-D Kirchhoff migration interpolated in the cross-section of the 2-D VSP model hardly identify any consistent pattern with the background velocity. See Figure (5.22(b)).

A 2-D waveform tomographic velocity inversion has been conducted using the VSP data Gao, 2003 [11]. Both surface data and bore-hole data are fitted. The surface-bore-hole geometry provides reasonably good ray coverage which is particularly fa-
voroble to tomographic type of velocity analysis. See Figure (5.22(a)).
Figure 5.21  The 3-D reflection seismic experiments made use of 624 Texan portable seismographs, and 2 Geometrix multi-channel seismographs. For sources we used 223 caliber single shot rifles, fired in 6 cm drill-holes. The Texan seismographs were deployed in 6 parallel lines with cross-line separations of 2.1m and with inline geophone spacing of 35 cm. Shots were fired in a rotated staggered brick pattern, with 120 shots/line. Forty-six seismic lines were occupied, producing a survey area of 94.5 by 36.05 m. The yellow line in the channel marks the VSP cross-section.
Figure 5.22  Results of previous studies. (a). Ray coverage of surface-bore-hole geometry. (b). Images of 3-D Kirchhoff migration interpolated in the VSP model are plotted on top of the velocity model obtained by surface-bore-hole diffraction tomography. The images are plotted in black.
5.3.3 Results by differential semblance velocity analysis

The VSP data received on the surface is dominated by surfaces waves. Part of the surface waves are removable by a standard trace-mixing technique (Figure (5.23). Large amount of surface waves are further removed by evanescent filtering. Our differential semblance velocity analysis uses only surface VSP data.

The starting velocity model is derived from the output velocity of diffraction tomography filtered by the B-spline low pass filter $BB^*$. The initial and optimized image at iteration 4, shown in Figure (5.24), suggest the deeper structures are identified more clearly in the optimized image. Simultaneously, the optimized velocity model agrees in pattern with the optimized image and exhibits more velocity variation than the initial velocity model (Figure (5.25)). The optimized offset gathers are better focused than the initial offset gathers indicating the differential semblance objective function is being minimized. The objective function value is reduced by half at the fourth iteration to $5.18 \times 10^9$ compared to the initial objective function value $1.18 \times 10^{10}$. The offset gathers can not be further focused within $h \approx 2m$, the threshold estimated as the wave-length of the propagating waves. See Figure (5.26).
Figure 5.23  Source gathers at $x_s = 12.6$ and $x_s = 15.4$. The data is surface wave dominated.
Figure 5.24  Comparison of initial and final images. The deeper structures are better characterized in the final image.
Figure 5.25  Comparison of initial and final velocity models. The final velocity model is obtained using surface data only. It agrees well with the inverted velocity through diffraction tomography which uses both the surface and the surface-bore-hole data.
Figure 5.26 Offset gathers migrated at initial and final (4th iteration) velocity models. The offset range for each panel is \((h_{\text{min}}, h_{\text{max}}) = (-4m, 4m)\). Better focused offset gathers are obtained at the optimized velocity model.
5.4 Marmousi data set

The original Marmousi dataset consists of 240 source gathers each of which is recorded by 96 geophones spaced 25m apart in a trailing geometry. The data we use in this analysis are the original Marmousi traces deconvolved with the source signature and continued to zero surface offset. The continuation to zero surface offset creates a horizontal image artifact located at 120m in depth (Figure 5.27(b)). The amplitude spectrum of the data is peaked at about 27Hz. The migration and the velocity inversion take up to 25Hz of the frequency spectrum. The computational grid spacing of 10m which provides a wavefield sampling of approximately 10 grid points per wavelength for an average velocity estimated at 3km/sec.

The image obtained using the true velocity model is presented in Figure 5.27(b)) showing the robustness of the migration algorithm in the presence of complex velocity structure. The differential semblance criterion in offset is still valid for this complex velocity structure. Figure 5.28(c)) shows the objective function evaluated at various velocity models through a line search fashion $V + \alpha V'$, where $V$ is the true velocity for shallow 0.6km and $V'$ a perturbation velocity corresponds to a maximum of 50% perturbation with respect to $V$. The objective function plotted against $\alpha$ draws a smooth curve centered at the true velocity in a wide range.

The target feature of the Marmousi velocity structure is in the salt dome in the mid-
dle of the model which is covered by high angle dipping faults. The long fault planes intersecting with the planes of sediments induce complicated ray paths and make the velocity inversion particularly difficult. The velocity analysis by differential semblance optimization has given satisfactory results for shallow Marmousi from the surface to 0.6 km in depth. We have used a 6 × 6 B-spline grid to represent the velocity model. The smoothed gradient for an initial constant (1.8 km/sec) velocity model appears in Figure (5.29(b)) with the B-spline smoothing implicitly used $BB^* \nabla_J$. While it lacks the shorter-scale lateral features of the B-spline projection of the true velocity (Figure (5.29(c))), it is certainly a constructive update direction.

The L-BFGS method requires one or sometimes two migrations for each BFGS iteration. The L-BFGS method reduced the objective function by roughly a factor of two, and the length of the gradient by an order of magnitude in five iterations. The fifth velocity iterate appears in Figure (5.29(b)). Most of the shorter-scale lateral features in the velocity have now reconstructed. Comparison of the image gathers at initial vs. fifth iteration velocities shows considerable improvement in focusing (Figure (5.30)).
(a) The Marmousi velocity structure.

(b) Image obtained using the true velocity structure.

Figure 5.27  (a). The Marmousi velocity structure. (b). Image obtained at the correct velocity.
(a) The exact Marmousi velocity model $V$ down to 0.6km.

(b) Velocity perturbation $V'$.

(c) Objective function evaluated at $J(V + \alpha V')$

**Figure 5.28** The differential semblance objective function varies in a wide and smooth curve with the minimum centered at the true velocity.
(a) Projected gradient at the constant (1.8 km/sec) velocity.

(b) The fifth velocity iterate.

(c) B-spline projection of the true velocity.

**Figure 5.29**  (a). B-spline model space gradient projected back to image space $B\nabla_m J = BB^*\nabla_c J$. (b). Velocity at the fifth iterate. (c). The best fit projected B-spline velocity $V_b = Bm_b$, where $m_b$ solves $Bm = V$ in a least-squares sense for the true Marmousi velocity $V$ defined on $921 \times 60$ image space grid and $m$ defined on a $6 \times 6$ B-spline grid.
Figure 5.30  (a). Iterates of magnitude of gradient and objective function values. The magnitude of gradient at the true velocity coincide with the objective function value represented in the red dashed line. (b). Scattered image offset gathers at the initial constant velocity. (c). Initial image at constant velocity. (d). Image offset gathers at the fifth iteration shows improvements in focusing. (e) Image at the fifth iteration.
5.5 Discussion: problems associated with the model roughness

The analysis in section (4.2.2) indicates that the image space gradient, in high frequency asymptotic, recovers the ray path compatible with the velocity used in migration. Traces of ray paths can be appear in Figure (5.31) especially in the region $7km < x < 9km$, where horizontal reflectors creates ray path of vertical reflection. The fault plane in the middle of the model complicates the ray geometry. Applying operator $BB^*$ to the image space gradient we obtain the smoothed gradient Figure (5.29(a)), which points to the correct direction to update the velocity.

The smoothing of the image space gradient is one of the crucial steps for velocity inversion. The velocity inversion requires a certain degree of smoothness (which is controlled by the B-spline projection). Decreasing the B-spline grid spacing tends to increase the degree of roughness and reduce the degree of the smoothness. The minimum of the differential semblance objective function is found at the exact velocity model, which is a rough model. Figure (5.32) demonstrates that the objective function value decreases at representations of the true velocity model with increased roughness. Using the same set of velocities derived from the true velocity model, the offset image gathers are more focused at zero offset at velocities of higher degree of roughness (Figure (5.33)). The offset gathers may not be well concentrated at zero
offset even at the correct velocity with too much smoothness. The conflicting requirement of roughness in order to bring offset gathers into focus and the requirement of smoothness for stable velocity inversions is the main difficulty for further studies of the subject.
(a) Image space gradient at constant (1.8 km/sec) velocity.

(b) B-spline smoothed gradient at constant (1.8 km/sec) velocity. Repeated Figure (5.29(a)).

Figure 5.31 Image space gradient at constant (1.8 km/sec) velocity. (a) The original gradient without smoothing. (b) B-spline smoothed gradient. Figure (5.29(a)) repeated.
Figure 5.32  Objective function evaluated at representations of the true velocity from surface to 1.4km in depth with increased roughness. Number of grid points in horizontal direction of the B-spline model increases from 30 to 80 increments by 10.
Figure 5.33  Offset gathers become more focused at velocities of higher degree of roughness.
(a) Horizontal B-spline grid points $\Delta x = 300m$.

(b) Horizontal B-spline grid points $\Delta x = 230m$.

(c) Horizontal B-spline grid points $\Delta x = 180m$.

(d) Horizontal B-spline grid points $\Delta x = 150m$.

(e) Horizontal B-spline grid points $\Delta x = 130m$.

(f) Horizontal B-spline grid points $\Delta x = 115m$.

**Figure 5.34** Increased image quality at velocities of higher degree of roughness.
Chapter 6
Conclusions

Shot-record wave equation migration is free of kinematic imaging artifacts and provides an ideal platform for migration velocity analysis. A version of wave equation migration is presented in this thesis, which directly pursues the quality of the image through criterion of differential semblance optimization. There are three main contributions of this thesis: first, gradient of the differential semblance objective function with respect to velocity is formulated through extensive use of adjoint state analysis. The construction of the gradient calculation can be easily extended to other migration methods. Computation of the gradient in two dimensions is implemented and can be accordingly extended to 3 dimensions; Second, a physical meaning of the gradient is analyzed using the Green's function representation.; Third, B-spline forward interpolation and adjoint projection is found to be suitable to smooth the image space gradient. The use of B-spline smoothing scheme is crucial for obtaining constructive search directions for velocity updating.

Success of the algorithm developed in this thesis has been shown in both synthetic and real data examples. The discussion held in last chapter indicates the difficulty
for further studies in the direction of wave equation migration velocity analysis.
Appendix A
PML absorbing boundary condition and its attenuation analysis

Absorbing boundary conditions are often useful for synthetic seismic data simulation. This section gives an algorithm for the perfectly matched layer absorbing boundary condition for time domain acoustic wave propagation Cohen, 2001[7].

The scalar wave equation introduced in section (2.1) can be written as a system of first order partial differential equations.

\[
\rho \frac{\partial v}{\partial t} = -\nabla p \tag{A.1}
\]

\[
\frac{\partial p}{\partial t} = -\kappa \nabla \cdot v \tag{A.2}
\]

where \(v\) is the particle velocity field, \(p\) is the pressure, and \(\kappa\) is the incompressibility coefficient. Equation (A.1) is the Newton’s law for a kinematic energy conservative system free of external forces. Equation (A.2) is the elastic lithostatic constitutive relation between pressure change and volume change. Introduce energy dissipative system by adding diffusive terms to equation(A.1) and equation(A.2). We have

\[
\rho \frac{\partial v}{\partial t} = -\nabla p - \alpha v \tag{A.3}
\]

\[
\frac{\partial p}{\partial t} = -\kappa \nabla \cdot v - \beta p \tag{A.4}
\]
where \( \alpha \) is the linear viscous coefficient. The additional term, \(-\alpha v\), in equation (A.3) is responsible for the exponential energy decay. The constitutive relation, equation (A.2) (incompressibility condition) should be changed accordingly to match equation (A.3) through a thermodynamic argument. We continue to solve systems (A.3)+(A.4), assuming \( \kappa, \beta, \) and \( \alpha \) is slow varying in time and \( \rho \) is slow varying in space, we derive

\[
\frac{\partial^2}{\partial t^2} p = \frac{\kappa}{\rho} \nabla \cdot \left( \frac{\nabla p}{\rho} \right) + \kappa \nabla \cdot \left( \frac{\alpha}{\rho} v \right) - \beta \frac{\partial p}{\partial t} \tag{A.5}
\]

\[
= \frac{\kappa}{\rho} \nabla^2 p - \left( \frac{\alpha}{\rho} + \beta \right) \frac{\partial p}{\partial t} - \frac{\alpha \beta}{\rho} p + \frac{\kappa}{\rho} \nabla \alpha \cdot v
\]

From equation (A.5) we have that when \( \alpha \) and \( \beta \) are relatively small, it is energy conservative; when \( \alpha \) is relatively large, it describes an energy dissipative system. The only problem of equation (A.5) is that the particle velocity is unknown. From equation (A.3) an analytical solution of \( v \) is obtained assuming \( p \) is known and \( v|_{t=0} = 0 \):

\[
v(\bar{x}, t) = -\int_{t_0}^t e^{-\frac{\alpha}{\rho}(t-s)} \frac{\nabla p}{\rho}(\bar{x}, s) ds \tag{A.6}
\]

Numerically the time domain scalar wave field with PML absorbing boundary condition can be implemented as:

\[
\begin{align*}
v_a &= 0 \\
\text{for } t &= 0 : dt : t_{\text{max}} \\
v_a &= v_a - e^{\frac{\alpha}{\rho}} \frac{\nabla p}{\rho} dt \\
v &= e^{-\frac{\alpha}{\rho}} v_a \\
\text{update } p \text{ through equation (A.5)}
\end{align*} \tag{A.7}
\]

end(t)
Q analysis:

The attenuation analysis follows the argument by Cohen, 1999 [7]. The analysis is carried through a use of imaginary coordinates on which the energy of the wavefield is attenuated most significantly parallel to the direction of propagation of the wavefront. The main conclusion is that along the direction of wave propagation, the wave attenuates according to

$$\frac{1}{Q(\omega)} = \alpha \frac{\omega \lambda}{\pi c}$$  \hspace{1cm} (A.8)

where as usual $\omega$, $\lambda$, $c$ denotes angular frequency, wavelength and velocity, respectively. In non-dispersive medium the phase velocity $c$ is independent of frequency $\omega$, we have $1/Q(\omega) = 2\alpha$.

The Fourier transform of equations (A.3) and (A.4) are,

$$i\omega(\alpha + \rho)\hat{v} = -\nabla\hat{p}, \hspace{1cm} (A.9)$$

$$i\omega(\beta + 1)\hat{p} = -\kappa \nabla \cdot \hat{v}. \hspace{1cm} (A.10)$$

Eliminating $\hat{v}$ we find
\[ \dot{p} = \frac{\kappa}{iw(\beta + 1)} \nabla \cdot \frac{\nabla \dot{p}}{iw(\alpha + \rho)} \]  

(A.11)

To ease the analysis, let’s assume \( \rho = 1 \) and \( \alpha = \beta \). Equation (A.11) can be written as

\[ \omega^2 \ddot{p} + \kappa \frac{i\omega}{i\omega(\alpha + 1)} \nabla \cdot \frac{i\omega \nabla \dot{p}}{i\omega(\alpha + 1)} \]  

(A.12)

It reduces to

\[ \omega^2 \ddot{p} + \kappa \nabla^2 \dot{p} = 0 \]  

(A.13)

by the change of variables

\[ \ddot{x}_1 = x_1 + \frac{i}{\omega} \int_0^{x_1} \alpha(s)ds \]  

(A.14)

\[ \ddot{x}_2 = x_2 + \frac{i}{\omega} \int_0^{x_2} \alpha(s)ds \]  

(A.15)

A plane wave solution can be formulated

\[ \dot{p} = e^{i(\vec{k} \cdot \vec{x} - \omega t)} \]  

(A.16)

\[ = e^{-\frac{k_1 \int_0^{x_1} \alpha(s)ds + k_2 \int_0^{x_2} \alpha(s)ds}{\omega}} e^{i(\vec{k} \cdot \vec{x} - \omega t)} \]  

(A.17)

so

\[ |p| = e^{-\frac{k_1 \int_0^{x_1} \alpha(s)ds + k_2 \int_0^{x_2} \alpha(s)ds}{\omega}} \]  

(A.18)

and increment of amplitude in the direction of propagation is

\[ \Delta |p| = -\alpha |p| (k_1 \Delta x_1 + k_2 \Delta x_2) \]  

(A.19)
\[ = -\alpha|p|(k_1 \frac{k_1}{k} \Delta x + k_2 \frac{k_2}{k} \Delta x) \]
\[ = -\alpha k|p| \Delta x \]

where \( k = |\vec{k}| \) and \( \Delta x = |\vec{\Delta x}| \). Use the definition of \( Q \)

\[ \frac{1}{Q(\omega)} = -\frac{\Delta|p|}{\pi|p|} \]  \( \text{(A.20)} \)

where \( \Delta|p| \) is the amplitude increase in each cycle. We finally arrive at the expression of \( Q \) in the direction of wave propagation:

\[ \frac{1}{Q(\omega)} = \alpha \frac{\omega \Delta x}{\pi c} \]  \( \text{(A.21)} \)

For high frequencies, \( \Delta x \rightarrow 0 \) (take \( \Delta x \) as the wave length), it follows

\[ \frac{1}{Q(\omega)} = 2\alpha \]  \( \text{(A.22)} \)
Appendix B
Construction of asymptotic solutions

For hyperbolic equations, we can construct a function $p(t, x; \omega)$ with the purpose that

$$(\frac{\partial^2}{\partial t^2} - \nabla_x)p = O(\omega^{-N})$$

for certain positive integer $N$. This function proves to be very useful in the study of the phenomena governed by the equation $$(\frac{\partial^2}{\partial t^2} - \nabla_x)p = 0.$$ For large $\omega$, $p(t, x; \omega)$ corresponds to waves with high frequency. Using this high frequency asymptotic solution, we can study the propagation of the waves and its reflection and refraction. Derivations in this appendix follows mainly that of Ikawa, 2000[18]. Recall the wave equation is of the following form

$$(\frac{\partial^2}{\partial t^2} - \nabla_x)p(t, x) = 0 \quad (B.1)$$

We want to find a function $p(t, x; \omega)$, with parameter $\omega \in \mathbb{R}$, such that

$$(\frac{\partial^2}{\partial t^2} - \nabla_x)p(t, x; \omega) = O(\omega^{-N}) \quad (B.2)$$

for some integer $N$ as $\omega \to \infty$. This means when $\omega$ is sufficiently large, since $\omega^{-N}$ is exceedingly small we have a reasonable justification to regard $(\frac{\partial^2}{\partial t^2} - \nabla_x)p(t, x; \omega)$ as 0. We call the $p$ that satisfies equation(B.2) an asymptotic solution. So we first consider the following form
\[ p(t, x; \omega) = e^{i\omega \Phi(t, x)} a(t, x; \omega) \]  

(B.3)

with

\[ a(t, x; \omega) = \sum_j a_j(x) \frac{(i\omega)^m}{(i\omega)^j} \]  

(B.4)

for some integer \( m > 0 \). Derive second order derivatives with respect to \( t \) and \( x \)

\[
\begin{align*}
\frac{\partial^2 p}{\partial t^2} &= e^{i\omega \Phi} \left( (i\omega)^2 \left( \frac{\partial \Phi}{\partial t} \right)^2 a + i\omega \left( 2 \frac{\partial \Phi}{\partial t} \frac{\partial a}{\partial t} + \frac{\partial^2 \Phi}{\partial t^2} a + \frac{\partial^2 a}{\partial t^2} \right) \right) \\
\frac{\partial^2 p}{\partial x_j^2} &= e^{i\omega \Phi} \left( (i\omega)^2 \left( \frac{\partial \Phi}{\partial x_j} \right)^2 a + i\omega \left( 2 \frac{\partial \Phi}{\partial x_j} \frac{\partial a}{\partial x_j} + \frac{\partial^2 \Phi}{\partial x_j^2} a + \frac{\partial^2 a}{\partial x_j^2} \right) \right)
\end{align*}
\]  

(B.5) \hspace{1cm} (B.6)

Substituting them into the wave equation, we obtain

\[
\begin{align*}
\left( \frac{\partial^2}{\partial t^2} - \nabla_x \right) p(t, x; \omega) &= e^{i\omega \Phi} \left\{ (i\omega)^2 \left[ \frac{1}{c^2} \left( \frac{\partial \Phi}{\partial t} \right)^2 \right] - \sum_j \left( \frac{\partial \Phi}{\partial x_j} \right)^2 a \right. \\
&+ i\omega \left[ \frac{1}{c^2} \frac{\partial \Phi}{\partial t} \frac{\partial a}{\partial t} - 2 \sum_j \frac{\partial \Phi}{\partial x_j} \frac{\partial a}{\partial x_j} + \frac{1}{c^2} \frac{\partial^2 \Phi}{\partial t^2} a - \sum_j \frac{\partial^2 \Phi}{\partial x_j^2} a \right] \\
&\left. + \left[ \frac{1}{c^2} \frac{\partial^2 a}{\partial t^2} - \sum_j \frac{\partial^2 a}{\partial x_j^2} \right] \right\}
\end{align*}
\]  

(B.7)

Now introduce the notation for operator

\[
\Lambda = \frac{1}{c^2} \frac{\partial \Phi}{\partial t} \frac{\partial}{\partial t} - 2 \sum_j \frac{\partial \Phi}{\partial x_j} \frac{\partial}{\partial x_j} + \frac{1}{c^2} \frac{\partial^2 \Phi}{\partial t^2} - \sum_j \frac{\partial^2 \Phi}{\partial x_j^2}
\]  

(B.8)
\[ \nabla \Phi = \left( \frac{\partial \Phi}{\partial t}, \frac{\partial \Phi}{\partial x_1}, \frac{\partial \Phi}{\partial x_2}, \frac{\partial \Phi}{\partial x_3} \right) = (\Phi_t, \Phi_x) \] (B.9)

and

\[ H(t, x, \omega, \xi) = \frac{\omega^2}{c^2} - \sum_j \xi_j^2 \] (B.10)

Although \( H \) has no dependencies on \( t \), we still put \( t \) in just in order to keep the conjugacy between \( (t, x) \) and \( (\omega, \xi) \). We will use simplified notation \( H(x, \nabla \Phi) \) for \( H(t, x, \Phi_t, \Phi_x) \)

\[ H(x, \nabla \Phi) = \frac{\Phi_t^2}{c^2} - \sum_j \Phi_{x_j}^2 \] (B.11)

We substitute equation (B.4) into equation (B.7) and then rearrange the expression in terms of powers of \( iw \) in descending order

\[ e^{-i\omega t} \left( \frac{\partial^2}{\partial t^2} - \nabla_x \right)p(t, x; \omega) = (i\omega)^{m+2} H(x, \nabla \Phi)a_0 + \] (B.12)

\[ + (i\omega)^{m+1} (H(x, \nabla \Phi)a_1 + \Lambda a_0) \]

\[ + (i\omega)^m (H(x, \nabla \Phi)a_2 + \Lambda a_1 + \left( \frac{\partial^2}{\partial t^2} - \nabla_x \right)a_0) \]

\[ + (i\omega)^{m-1} (H(x, \nabla \Phi)a_3 + \Lambda a_2 + \left( \frac{\partial^2}{\partial t^2} - \nabla_x \right)a_1) \]
\[ \ldots + (i\omega)^{-N+2}(H(x, \nabla \Phi)a_{m+N} + \Lambda a_{m+N-1} + \left(\frac{\partial^2}{\partial t^2} - \nabla_x\right)a_{m+N-2}) + (i\omega)^{-N+1}(\Lambda a_{m+N} + \left(\frac{\partial^2}{\partial t^2} - \nabla_x\right)a_{m+N-1}) + (i\omega)^{-N}(\frac{\partial^2}{\partial t^2} - \nabla_x)a_{m+N} \]

To analyze the above equation, we think the case when \( H(x, \nabla \Phi) \neq 0 \). In order to make \((i\omega)^{m+2}\) term to vanish we need to have \( a_0 = 0 \). Next, with \( a_0 = 0 \), in order to make \((i\omega)^{m+1}\) term to vanish we need \( \nu_1 = 0 \). Continue in this process, we end up to a trivial solution. Therefore, for non-trivial solution for wave propagation it is necessary that the following is satisfied

\[ H(x, \nabla \Phi) = 0 \quad (B.13) \]

In this case, if

\[ \Lambda a_0 = 0 \quad (B.14) \]

then

\[ \Lambda a_j = -\left(\frac{\partial^2}{\partial t^2} - \nabla_x\right)a_{j-1} \quad (j = 1, 2, \ldots, m + N) \quad (B.15) \]

Equation (B.13) is called the eikonal equation. For \( \Phi(t, x) \) usually takes the form of \( t - \phi(x) \). Substituting into equation (B.13), we see the usual eikonal equation used in geometric optics
\[ |\nabla_x \phi(x)|^2 = \frac{1}{c^2} \]  \hspace{1cm} (B.16)

Equation (B.14) is called the first transport equation. If we assume \( a \) is independent of time \( t \) and the choice of \( \Phi(t, x) = t - \phi(x) \) implies that \( \frac{\partial^2}{\partial t^2} \Phi = 0 \). This gives

\[ 2\nabla_x \phi \cdot \nabla_x a_0 + \nabla_x^2 \phi a_0 = 0 \]  \hspace{1cm} (B.17)

which is the form of the first transport equation given in most references.
Appendix C

Migration by an analysis of stationary phase

The seismic migration reconstructs partially the wavefront set of the medium. We will directly verify this by an asymptotic analysis. Now we take the phase function $\Phi(t, x)$ discussed in appendix (C) to be of the form

$$\Phi(t, x) = t - \phi(x)$$  \hspace{1cm} (C.1)

and assume that the amplitude function $a$ is independent of time $t$, then the Green's function solution of Helmholtz equation (2.23) can be approximated as

$$G(x, \omega) = a(x, \omega) e^{-i\omega\phi(x)}$$  \hspace{1cm} (C.2)

To reflect the fact that the wave is propagating from source $x$ to observation point $y$, we write the Green's function as

$$G(x, y; \omega) = a(x, y; \omega) e^{-i\omega\phi(x,y)}$$  \hspace{1cm} (C.3)

so the most singular part of the seismogram by the Born approximation observed at $x_r$ due to a point source $x_s$ can be written as

$$d(x_r, x_s; \omega) \approx \delta p(x_r, x_s; \omega) = 2 \int a(\omega, x; x_s) a(\omega, x_r; x) \frac{\omega^2}{c^2} r(x) e^{-i\omega\phi(x,x_s)} e^{-i\omega\phi(x_r,x)} dx$$  \hspace{1cm} (C.4)

where $r = \delta c/c$ is the reflectivity coefficient and $\phi(x, y)$ are phase functions implicitly
dependent on velocity $c$, which satisfies the eikonal equation

$$|\nabla_x \phi(x, y)|^2 = 1/c(x)^2 \quad (C.5)$$

The phase functions $\phi(x, y)$ are continuously differentiable with respect to $x$ or $y$ except at points of $(x, y)$ such that $\phi(x, y) = 0$. Referring to equation (B.4), the term $a(x, y; \omega)$ is a sum of slow varying amplitudes, each of which is of homogenous degree of, at most, $m$ in $\omega$. In particular the amplitude $a_0(x, y)$ in equation (B.4) satisfies the first transport equation

$$\nabla_x \cdot [a_0^2(x, y) \nabla_x \phi(x, y)] = 0 \quad (C.6)$$

We want to isolate the reflection to the neighboring points of discontinuities. Introduce cut-off function $\psi_{x'}(x) \in C_c^\infty$ which is identically 1 and compactly supported in a neighborhood close enough to the point of discontinuity at $x'$. Supposing $r(x)$ has an inverse Fourier transform, the data due to the reflection from discontinuities near $x$ can be written as

$$d_{x'}(x_r, x_s; \omega) = 2 \int a(x, x_s; \omega) a(x_r, x; \omega) \frac{\omega^2}{c^2} \psi_{x'}(x) \hat{r}(\zeta) e^{i \zeta \cdot x} e^{i \omega [-\phi(x, x_r) - \phi(x_r, x_s)]} dx dx' d\zeta \quad (C.7)$$

The adjoint of the integral operator in equation (C.7), using the same velocity function $c$, applied to data can be written

$$\hat{r}_{x'}(y) = \int a(y, x_s; \omega) a(x_r, y_r; \omega) e^{i \omega \phi(y, x_s) + i \omega \phi(y, x)} d_{x'}(x_r, x_s; \omega) dx_r dx_s dw \quad (C.8)$$

Substituting equation (C.7) into equation (C.8), we obtain
\[ \mathcal{F}_{x'}(y) = \int A(y, x, x_s, x_r; \omega) \psi(x) \mathcal{F}(\zeta) e^{i\zeta x + iw[-\phi(x, x_s) - \phi(x_r, x) + \phi(y, x_s) + \phi(x_r, y)]} dx_r dx_s dxdwd\zeta \]  
\hspace{1cm} (C.9)

Introduce a cut-off function \( \psi_{\gamma'}(y) \in \mathcal{C}_c^{\infty} \), identically 1 and compactly supported in a neighborhood of \( y' \). We study the asymptotic behavior of the form

\[
J_{x', y'}(k) = \int \mathcal{F}_{x}(y) \psi(y) e^{-ik \cdot y} 
= \int A(\omega, x, x_s, x_r, y) \mathcal{F}(\zeta) \psi_{x'}(x) \psi_{y'}(y) \times
\hspace{1cm} e^{i\zeta x - ik \cdot y + iw[-\phi(x, x_s) - \phi(x_r, x) + \phi(y, x_s) + \phi(x_r, y)]} dwd\zeta dxdydxd_r dx_r 
\hspace{1cm} (C.10)
\]

In agree to the form of stationary phase analysis, \( J \) can be written as

\[
J_{x', y'}(k) = \int B(\omega, \zeta, \sigma) e^{i\Theta(k, \omega, \zeta, \sigma)} dwd\zeta d\sigma 
\hspace{1cm} (C.12)
\]

where \( d\sigma = dxdydxd_r dx_r \). To proceed, we point out first that \( \nabla_x \phi(x, y) \) points to the propagation direction of the wavefront. In order to verify this, we look at the local plane wave decomposition of \( a(\omega, x, y) e^{i\omega \phi(x, y)} \) in the neighborhood of \( x \).

\[
W_x(\xi) = \int a(\omega, x, y) e^{i\omega \phi(x, y)} e^{-i\xi \cdot x} \psi(x) dx 
\hspace{1cm} (C.13)
\]

Asymptotically for large \( |\xi| \), it picks up a non-trivial energy when

\[
\omega \nabla_x \phi(x, y) = \xi 
\hspace{1cm} (C.14)
\]

As we know, \( \xi \) is the wave number of the local plane wave. The above equation means: the wave represented by \( a(\omega, x, y) e^{i\phi(x, y)} \) propagates most significantly to the
particular direction of $\nabla_x \phi(x, y)$, which is parallel to the wave number of a plane wave decomposed at the vicinity of $x$. And secondly we notice
\[
\nabla_x \phi(x, y)|_{x=y} = -\nabla_y \phi(x, y)|_{x=y}
\] (C.15)

which can be easily verified. We now turn back to the stationary phase analysis of the integral equation (C.12). The stationary phase analysis concerns the points such that $\nabla_{\omega, \zeta, \sigma} \Theta = 0$. Explicitly, the result of the integral comes from the points such that they satisfy jointly
\[
\frac{\partial}{\partial \omega} \Theta = 0 \Rightarrow \phi(x, x_s) + \phi(x_r, x) = \phi(y, x_s) + \phi(x_r, y)
\] (C.16)
\[
\nabla_x \Theta = 0 \Rightarrow \zeta + \omega \nabla_x \phi(x, x_s) + \omega \nabla_x \phi(x_r, x) = 0
\] (C.17)
\[
\nabla_y \Theta = 0 \Rightarrow -k - \omega \nabla y \phi(y, x_s) - \omega \nabla y \phi(x_r, y) = 0
\] (C.18)
\[
\nabla_{x_s} \Theta = 0 \Rightarrow \nabla_{x_s} \phi(x, x_s) = \nabla_{x_s} \phi(y, x_s)
\] (C.19)
\[
\nabla_{x_r} \Theta = 0 \Rightarrow \nabla_{x_r} \phi(x_r, x) = \nabla_{x_r} \phi(x_r, y)
\] (C.20)

Equation (C.16) requires that the travel time connected from $x_s$ to $x'$ to $x_r$ should be equal to the travel time connected from $x_s$ to $y'$ to $x_r$. Moreover, equation (C.19) and equation (C.20) indicate that both ray paths associated with the two travel times have the same ray parameter on the surface, that is the ray parameters are the same at each source point $x_s$ and receiver points $x_r$. If we force $x' = y'$, meaning we want to look at the point in the image from which the reflection data is generated. It is obvious from equation (C.18) and equation (C.17) that the wavefront set of
Figure C.1  The travel time is preserved. Snell’s law applies to both image point and the reflection point. The ray parameters coincide at the surface.

discontinuity distribution $r$ is reconstructed.
References


