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:

William W. Symes

Alan Levander

Scott Morton

Colin Zelt

Fenglin Niu

May, 2004

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 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 structure. The gradient of the objective function with respect to velocity is fomulated 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.

I can't end without thanking Dr. Fenglin Niu. He has been abundantly helpful and has given me excellent advice on preparing my thesis.

Contents

List of Figures

List of Tables

Chapter 1
Thesis introduction

1.1 Thesis structure

The thesis is written to consist of chapters in logical steps towards explanation of what I did on "wave equation migration velocity analysis by differential semblance optimization". This thesis work is composed of five major components in sequence: one-way wave equation prestack depth migration, construction of 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 velocity which is considered to be of smooth functions at least twice contiguously differentiable. The solution of the problem requires to understand 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 considered as 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 analysis 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 the singular behavior asymptotically. In order to do so, the 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 in before as velocities. This analysis thus reveals the relationship between velocity analysis and imaging. Mi-

gration, viewed as an operator defined on certain velocity function, will reconstruct the wavefront set of the coefficient asymptotically if the velocity is accurate. At 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

1.4.1 Classification of data domain method and image domain method

From a methodological point of view, velocity analysis is considered as applications of optimization theory combined with geophysical analysis. Depending on how the objective function is constructed, methods of velocity analysis divide into two categories: data domain method in which the objective function is formulated to measure the data residual obtained by subtracting predicted data from 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 of the first type, which is to minimize the data residual.

$$\min_c \|d^{obs} - d^{pre}\|,$$

likely, the latter is posed as an optimization problem of the second type

$$\min_c \|PI\|$$

where P provides required measurement on the image, and $\|\cdot\|$ indicate a proper norm for both data domain method and image domain method.

1.4.2 Data domain method

Early methods of velocity analysis try to find models that best explain the data and therefore fit in the category of data domain method. 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* [16]. Clearly travel time tomography is limited to fitting the first few phases of the observed data neglecting important information carried by later arrivals.

Waveform type of optimizations treat the observed data as it is and try to fit it

by solving wave equations. Least-squares waveform inversion require the starting velocity model be close enough to the true model in the sense that the predicted data are within half wavelength distanced from the observed data for the highest frequency used. The velocity updating is mainly driven by the gradient calculation which is formulated by 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 (assuming to be the Born data) reconstruct the image asymptotically, and thus is equivalent to reverse time prestack depth migration. The velocity analysis of data domain method usually takes the increasing frequency scheme where the data correspond to lower frequencies are first fitted, and then gradually increase to higher frequencies*Pratt, 1999* [17], *Pratt, at el., 1996* [18]. Time domain implementations using wave equation of such methods have been presented by *Tarantola 1987* [27], *Tarrantola & Vallette, 1982* [28].

The main disadvantages of data domain methods is that the predicted data using the current velocity model has to be within half wavelength to the observed data. This condition restricts the velocity inversion away from the highly nonlinear situa-

tion where the starting velocity model is far from the true velocity model. In another word, the objective function in data domain is easily trapped in local minima when the "half wavelength" condition is violated. Another disadvantage of the data domain methods is that the image obtained by migration using the optimized velocity may not be the optimum image because the methods do not directly pursue the optimality of the image. Finally, the data domain method depends on simulations of seismic wave propagation at correct physical model. Errors in assumptions of physical model often transforms into errors of the optimized velocity.

1.4.3 Image domain method

Recent researches of velocity analysis focus on image domain methods, in which objective functions are posed directly to improve the quality of the image. This brings out the fundamental question of how to quantify the quality of the image. The answer amounts to finding different ways to measure the singularities since image is understood as reconstructed wavefront sets of the medium. We shall mention two principal geophysical methods of measuring the singularities, angle domain measurements and subsurface offset domain measurements, each has a clear physical meaning through asymptotic analysis *Stolk & Symes, 2003* [24], *Shen P. at el. 2003* [22]. When ve-

locity is correct, the pair of rays from source and receiver 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 expanded in scattering angle should be flat. 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 in between rays are parameterized by the so called subsurface offset. The criterion for velocity analysis can thus be posed as that the common image gathers in subsurface offset is concentrated at zero offset.

1.4.4 Optimization criteria

We shall compare two image domain optimization criteria, one is differential semblance criteria, the other minimum image perturbation criterion.

1.4.4.1 Differential semblance criteria

The differential semblance criteria seek for differential semblance operators P that annihilate singularities of image gathers $I(x,h)$ or $I(x,\theta)$ at the correct velocity, where h and θ is subsurface offset and scattering angle, respectively. At incorrect velocities the application of operator P removes part of the singularities of the image. The result image gathers PI is called the image residual. The velocity updating is driven by minimizing the magnitude of the image residual defined by the objective function of the form

$$J = \frac{1}{2}\|PI\|^2 \tag{1.1}$$

with a proper norm $\|\cdot\|$. The operator P has several choices each of which measures the local properties of the image by taking differentials. In scattering angle domain, operator $\partial/\partial\theta$ or $\partial/\partial(tan(\theta))$ measure the non-flatness of the angle image gathers $I(x,\theta)$. In subsurface offset domain, the operator h measures the deviation from concentration of offset gathers $I(x,h)$ near $h=0$. The differential semblance operator used in this thesis is simply multiplication by subsurface offset parameter $P=h$. This differential semblance operator h requires no picking on offset (subsurface) domain image gathers, therefore is particular suitable for automatic velocity analysis. The corresponding objective function taken by L^2 norm is $J=\frac{1}{2}\|hI\|_2^2$. This differential semblance objective function varies smoothly with respect to the change of velocities

Stolk & Symes, 2003 [24], which can be optimized by gradient type of methods.

1.4.4.2 Minimum image perturbation criterion

The image perturbation $\delta I(x)$, introduced by *Sava & Biondi,2003*, [20], measures the perturbation of image $\delta I(x)$ due to perturbations of the velocity model δc. 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*, [20].

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

$$I(x; c_t) = I(x; c_b) + L\delta c \tag{1.2}$$

where

$$L = \frac{\partial I(x; c_t)}{\partial c}|_{c=c_b} \tag{1.3}$$

and

$$\delta c = c_t - c_b \tag{1.4}$$

The image perturbation is defined by

$$\delta I(x; c_t, c_b) = I(x; c_t) - I(x; c_b) \tag{1.5}$$

The optimization criterion expressed in the image perturbation is

$$\min_{c_b} ||\delta I(x; c_t, c_b)|| \tag{1.6}$$

which is achieved through solving a sequence of linear optimization problems

$$\min_{c} ||\delta I - L\delta c|| \tag{1.7}$$

The major difficulty for minimizing image perturbation is to obtain δI. As a matter of fact δI can never be obtained because the true velocity is unknown. Introduce the ratio of the improved velocity c and the current velocity c_b,

$$\rho = \frac{c}{c_b} \tag{1.8}$$

the image perturbation is instead approximated through a differential form

$$\delta I(x; c_t, c_b) \approx \frac{dI(x, c)}{d\rho}|_{\rho=1}\delta\rho \tag{1.9}$$

which can be further approximated as

$$\frac{dI(x, c)}{d\rho}|_{\rho=1}\delta\rho \approx (\frac{\partial I}{\partial k_z}\frac{\partial k_z}{\partial \rho})|_{\rho=1}\delta\rho \tag{1.10}$$

The image perturbation δI 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 $\rho = 1$

$$\frac{dI}{dk_z}|_{\rho=1} = -izI \tag{1.11}$$

Secondly, the derivative of the depth wavenumber with respect to velocity ratio is formulated through the Double-Square-Root equation as, *Sava, 2003*[21]

$$\frac{\partial k_z}{\partial \rho}|_{\rho=1} = \frac{\mu}{2\sqrt{\mu^2 - |k_s|}} + \frac{\mu^2}{2\sqrt{\mu^2 - |k_r|}} \tag{1.12}$$

where

$$\mu^2 = \frac{[4k_{z_0}^2 + (|k_r| - |k_s|)^2][4k_{z_0}^2 + (|k_r| + |k_s|)^2]}{16k_{z_0}^2} \tag{1.13}$$

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 (1.11) and (1.12) provide means to calculate $(dI/d\rho)|_{\rho=1}$. Finally, an optimum velocity ratio ρ_* can be obtained from the set of residually migrated images at various values of the parameter ρ, *Sava, 2000*[19]. The residual migration explained in *Sava, 2000*[19] is fast to compute and only works for constant velocity ratio $\rho(x) = const$. The optimality conditions for ρ_* is the flatness of the angle-domain image gathers, which can be measured through stack power, semblance or differential semblance.By setting

$$\delta\rho \approx \rho_* - 1 \tag{1.14}$$

we obtain an approximate of $\delta\rho$. Combining equations (1.11) (1.12) and (1.14) we

can calculate the image perturbation $\delta I(x)$. Once the image perturbation is calculated, the velocity perturbation δc can be obtained by solving the linear least-squares problem (1.7) using the conjugate gradient method. The velocity model is updated through $c_b \leftarrow c_b + \delta c$, and the process repeats until the image perturbation is sufficiently minimized.

1.4.4.3 Relationship between differential semblance criteria and the minimum image perturbation criterion

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

$$PI(c_t) = PI(c_b) + P\frac{\partial I}{\partial c}|_{c=c_b}\delta c \tag{1.15}$$

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

$$-PI(c_b) = P\frac{\partial I}{\partial c}|_{c=c_b}\delta c \tag{1.16}$$

which relates to equation (1.7) by replacing the image perturbation by the negative image residual PI and replacing L by PL. The image perturbation maintains the singularities of both the current image and the residually migrated image, whereas the

residual image tends to remove the singularities of the image. In this thesis I do not use equation (1.16) to optimize the differential semblance objective function, instead it is directly optimized by quasi-Newton algorithms which can provide quadratic update.

1.4.5 Kirchhoff migration based velocity analysis

Kirchhoff migration based velocity analysis usually works with image gathers in angle or surface offset domain. Residual move-outs in angle or surface offsets can be picked from semblance gathers (common image gathers) *Docherty at al., 2000*, [4]. 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 analysis in ray tracing regime.

1.5 Wave equation migration based velocity analysis

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. We shall, for time being, consider the image residual as some kind of imperfectness of the image, intuitively. Similar to Kirchhoff migration velocity analysis, the wave equation migration velocity analysis also needs to provide mechanism of projecting image residual back to model space alone the ray path. The information of ray path is not explicitly known in wave equation approach. In this thesis I define the differential migration as the Frechet derivative of migration as an operator with respect to the velocity. An important component of this thesis is to demonstrate that the adjoint of differential migration serves to find the ray paths (in high frequency asymptotic) that are implicitly used in wave equation migration. The gradient of image domain objective function (objective function of the second type) is formulated, in closed form, as the adjoint of differential migration as an operator applied to the image residual. It has a 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 useable for veloc-

ity updating because of the nonsmoothness 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 that of the high frequency.

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* [7]. the conservative scheme for wave equation migration preserves the total energy for downward extrapolated wavefields. The adjoint of differential migration using 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 conservative method is approximated four times of that of migration.

1.6 Thesis chapter summery

Chapter 2, Migration

The chapter gives introduction to basic concept and treatment of wave equation migration which is used as a platform for velocity analysis. The chapter answers why wave equation migration reconstruct the singularities of the subsurface structure and

how the implicit finite difference wave equation migration is implemented.

Chapter 3, Velocity analysis by differential semblance optimization

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

Chapter 4, Adjoint state calculation and inversion

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

Chapter 5, Data examples

The chapter demonstrates data examples that uses algorithm developed in this thesis. It shows that the this algorithm of automatic velocity analysis by differential semblance optimization works well for complex geology structures in two dimensions.

Chapter 2
Wave equation migration

Prestack depth migration is an important and effective tool in identifying the potential targets for oil and gas exploration. The Kirchhoff approach for prestack depth migration relies on the high-frequency asymptotic ray-tracing. Problems of ray tracing remain in caustics, multiple arrivals and shadow zones [5]. It is expected that the wave equation based methods can avoid these difficulties. Approaches by solving full wave equation is capable to handle very complex media. Such methods are usually very time consuming. One-way methods based on paraxial wave equation are much more efficient for reasons which will be explained in section (2.3.1). For completeness of the thesis writing, 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 sin-

gularities 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

$$\rho \frac{\partial^2}{\partial t^2} u_i(x,t) = f_i(x,t) + \sum_{jkl} (c_{ijkl} u_{k,l})_{,j}\,(x,t) \tag{2.1}$$

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

$$c_{ijkl} = \lambda \delta_{ij} \delta_{jl} + \mu(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) \tag{2.2}$$

with λ and μ known as Lame constants. The equation of motion reduces to

$$\rho \frac{\partial^2}{\partial t^2} u_i(x,t) = f_i(x,t) + \sum_{j} \sigma_{ij,j}(x,t) \tag{2.3}$$

where

$$\sigma_{ij} = \lambda \sum_{k} e_{kk}\delta_{ij} + 2\mu e_{ij} = \lambda\theta\delta_{ij} + 2\mu e_{ij} \tag{2.4}$$

$$e_{ij} = \frac{1}{2}(u_{i,j} + u_{j,i}) \tag{2.5}$$

and

$$\theta = \sum_k e_{kk} = \nabla \cdot u \tag{2.6}$$

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

$$\sigma_{ij} = (\lambda\theta + \frac{2}{3}\mu\theta)\delta_{ij} + 2\mu e_{ij} - \frac{2}{3}\mu\theta\delta_{ij} \tag{2.7}$$

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

$$\sigma_{ij} = -p\delta_{ij} \tag{2.8}$$

From equation(2.7), it is clear that

$$-dp = (\lambda + \frac{2}{3}\mu)d\theta \tag{2.9}$$

The bulk modulus κ is introduced as the ratio of the change of pressure to the change of fractional volume.

$$\kappa = \frac{-dp}{d\theta} = \lambda + \frac{2}{3}\mu \tag{2.10}$$

In the same time, the equation of motion for lithostatic pressure is

$$\rho\frac{\partial^2}{\partial t^2}u_i = f_i - p_{,i} \tag{2.11}$$

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

$$\frac{\partial}{\partial t}p = -\kappa\nabla\cdot v \tag{2.12}$$

$$\rho\frac{\partial}{\partial t}v = -\nabla p \tag{2.13}$$

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

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

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

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

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 wavefield observed at x subject to source located at x_s satisfies

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

One can decompose c' as summation of reference velocity c which is assumed to be smooth and perturbation velocity δc,

$$c'(x) = c(x) + \delta c(x) \tag{2.17}$$

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

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

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

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

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

$$(\frac{1}{c^2(x)}\frac{\partial^2}{\partial t^2} - \nabla_x^2)\delta\tilde{G}(x, x_s, t) = 2\frac{\delta c(x)}{c(x)}\frac{\omega^2}{c^2(x)}\frac{\partial^2}{\partial t^2}\tilde{G}(x, x_s, t) \tag{2.20}$$

Using the Green's function representation, we can write the solution evaluated at receiver location x_r on surface as

$$\delta\tilde{G}(x_r, x_s, t) = \int 2r(x)\frac{\omega^2}{c^2(x)}\tilde{G}(x, x_s, \omega)\tilde{G}(x, x_r, \omega)e^{iwt}dxd\omega \tag{2.21}$$

Here

$$r(x) = \delta c(x)/c(x) \tag{2.22}$$

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

$$(\frac{\omega^2}{c^2(x)} - \nabla_x^2)\tilde{G}(x, y, \omega) = \delta(x - y) \tag{2.23}$$

The Green's function satisfies the two-way Helmholtz equation is called the two-way Green's function. It propagates to all directions in the medium. If we consider the solution to equation(2.15) as the result of the map

$$\tilde{G} = F(c) \tag{2.24}$$

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

$$\delta\tilde{G} = \frac{\partial F}{\partial c} r \tag{2.25}$$

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 [11] 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 δp observed on, usually, 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 adjoint of Born modeling

Migration can be formulated by the adjoint of Born operator. In view of the forward Born mapping $\partial F/\partial c : r \rightarrow \delta p$, δp contends 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" through out the thesis. The concept of wavefront set is well explained in *Hörmander 1990*[10]. Taking the adjoint of operator in equation(2.21) and applying it to data, we obtain

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

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

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

Second, solve the adjoint state wavefield

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

where $\tilde{R}$ can be shown to satisfies a two-way Helmholtz equation. Third, propagate the source wavefield complex conjugated

$$\tilde{S}(x, x_s, \omega) = \overline{\tilde{G}(x, x_s, \omega)} \tag{2.29}$$

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

$$\tilde{I}(x) = \int dx_s \int d\omega \; e^{i\omega t}|_{t=0} \tilde{S}(x, x_s, \omega) \tilde{R}(x, x_s, \omega) \frac{\omega^2}{c^2} \tag{2.30}$$

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 may be called the reverse time migration. The "revers time" can be understood in view of equation (2.26) by its convolution representation in time.

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

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 frame work of shot-record migration. In high frequency asymptotic, the equation of migration equation(2.26) produces singularities of two types, one is the image, the other the refraction ray path. The statement is verified through the stationary phase analysis.

2.2.1 Method of stationary phase analysis

The Green's function $\tilde{G}(x, y, \omega)$ can be approximated by

$$\tilde{G}(x, y, \omega) = a(x, y, \omega)e^{-i\omega\phi(x,y)} \tag{2.32}$$

assuming that the wavefront is smooth and the rays from y do not intersect each other. Where $a(x, y, \omega) = \sum_j^{m+N} a_j(x, y)\frac{(i\omega)^m}{(i\omega)^j}$. When the differential operator P of Helmholtz equation acts on it this expression decays as ω^{-N} for any non-negative integer N if $x \neq y$. The sequence of amplitude terms $\{a_j\}$ satisfy the recursive system of transport equations, the first of which is called the first transport equation *Ikava, 2000* [12]

$$2\nabla_x\phi(x, y) \cdot \nabla_x a_0 + \nabla_x^2\phi(x, y) = 0. \tag{2.33}$$

See appendix B. The phase term can be shown to satisfy the eikonal equation

$$|\nabla_x \phi(x,y)|^2 = \frac{1}{c^2} \tag{2.34}$$

With proper choice of initial values of a_j the Helmholtz equation can be made to satisfy $P\tilde{G} = \delta(x-y)$ modulo a C^∞ function. In high frequency regime the expression of equation(2.32) is considered as a good a approximation for solutions of Helmholtz equation.

In the Green's function representation the data, in an first order approximation, can be expressed as summation of two terms, refracted arrival and reflection of single reflected arrival

$$\begin{aligned} d(x_r, x_s, \omega) &= a(x_r, x_s, \omega) e^{-i\omega\phi(x_r, x_s)} + \\ &\int 2r(x) a(x, x_s, \omega) a(x_r, x\omega) e^{-i\omega(\phi(x,x_s)+\phi(x_r,x))} dx \end{aligned} \tag{2.35}$$

Substitute $d(x_r, x_s, \omega)$ into equation of migration equation(2.26) and use the asymptotic expression for $\tilde{G}$, we derive

$$\begin{aligned} \tilde{I}(x) &= \int 2\frac{\omega^2}{c^2} A(x_r, x_s, x, \omega) e^{iw[\phi(x,x_s)+\phi(x_r,x)-\phi(x_r,x_s)]} dx_r dx_s d\omega + \\ &\int 2\frac{\omega^2}{c^2} B(x_r, x_s, y, \omega) e^{iw[\phi(x,x_s)+\phi(x_r,x)-\phi(y,x_s)-\phi(x_r,y)]} dy dx_r dx_s d\omega \end{aligned} \tag{2.36}$$

where A and B are products of amplitude terms with

$$\begin{aligned} A &= a(x_r, x_s, \omega) a(x, x_s, \omega) a(x_r, x, \omega) \\ B &= a(x, x_s, \omega) a(x_r, x, \omega) a(x', x_s, \omega) a(x_r, x', \omega). \end{aligned}$$

The first term of equation (2.36) becomes singular if

$$\phi(x, x_s) + \phi(x_r, x) - \phi(x_r, x_s) = 0 \tag{2.37}$$

Joshi [13]. The point x that satisfies equation(2.37) lies on the refracted ray path, thus produces the singularities of the first type. A stationary phase analysis presented in appendix C shows that the integral of second term picks up pairs of rays connecting source, point of reflection and receiver. The wavefront set of reflectivity distribution r is reconstructed when the migration is carried through the correct background velocity, and thus produces the singularities of the second type.

Refracted arrivals (including direct arrivals) can be removed from data using proper preprocessing. With the assumption that d corresponds only to data of single reflected reflections as described by the Born modeling equation(2.21), equation of migration equation(2.26) produces singularities of the second type, the image of velocity medium. The assumption may be violated when multiple reflections exist. Problems of de-multiple are not discussed in this thesis.

2.3 Wave equation migration

The migration of seismic data requires solving scalar wave equations in multidimen-

sions. Time domain solutions are natural which implicitly solves for all frequencies by stepping through the required time window. The stability condition for finite difference time stepping schemes is assured if time interval satisfies $\Delta t \leq \frac{e_1}{c_{max}} h$ for spatial interval h determined by $h = \frac{c_{min}}{e^2 f_{max}}$, where c_{min}, c_{max} are minimum and maximum velocities, e_1 is the coefficient related to finite difference approximations to the Laplacian operator. For 3D 4^{th} order scheme, e_1 is taken by 1/2, *Wu et al., 1996* [29]. The numerical dispersion relation requires e_2 to be at least 5.* Geological velocity models usually exhibit high velocity contrast with c_{max}/c_{min} greater than 3. With these estimations made above the number of time steps for a reasonable time window $t = 2sec$ needs

$$N_t = \frac{t f_{max} c_{max} e_2}{c_{min} e_1} (\approx 1200) \tag{2.38}$$

for $f_{max} \approx 20$Hz. The computational cost for each time stepping 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 computationally prohibitive. On the other hand, the number of frequencies needed is much less then the number of time steps. The Nyquist frequency f_N for time interval Δt is $f_N = \frac{1}{2\Delta t}$. The frequency interval is determined by $\Delta f = 2 f_N / N_t = 1/t$. The number

*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.

of frequencies N_f needed to reconstruct the time domain solution is

$$N_f = f_{max}/\Delta f = t f_{max} \tag{2.39}$$

which agrees with the number of frequencies used in practice at the order of 10^2. It follows from equation(2.39) and equation(2.38) that the relation

$$N_t/N_f = \frac{c_{max} e_2}{c_{min} e_1} (\approx 60) \tag{2.40}$$

indicates the number of time steps is at one order of magnitude greater than the number of frequencies for reasonable parameter estimations.

The Helmholtz equation is therefore a substitute for modeling of wave propagation. Write the receiver wavefield as

$$\tilde{R}(x, x_s, \omega) = \int \overline{\tilde{G}(x, x_r; \omega)} d(x_r, x_s; \omega) dx_r \tag{2.41}$$

which satisfies the full Helmholtz equation:

$$(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} + \frac{\omega^2}{c^2}) \tilde{R}(x_1, x_2, x_3; \omega) = \int \delta(x - x_r) d(x_r, x_s; \omega) dx_r \tag{2.42}$$

The full Helmholtz equation (2.42) solved using matrix inverse by Gaussian elimination 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 (2.42) can be derived by paraxial approximation, namely the one-way wave equation *Claerbout, 1985*

[1]. The implicity finite difference solution for one-way wave equation provides $O(N^3)$ floating point operations per frequency. We will give analysis of the computational cost of implicit finite difference method later in section (2.3.3).

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 one way traveling if the assumption

$$|k_1| > \epsilon\sqrt{k_1^2 + k_2^2}, \quad \epsilon \to 0 \tag{2.43}$$

is valid everywhere for waves propagated in the model, where $k \in \mathbb{R}^3$ is the wavenumber for full wavefield Green's functions and k_1 is the corresponding vertical wavenumber *Stolk & De Hoop, 2001* [25]. 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

their one way approximates and yet still explains 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

$$\frac{\partial}{\partial x_1}G^+(x,y,\omega) - \frac{i\omega}{c}\sqrt{1+\frac{c^2}{\omega^2}(\frac{\partial^2}{\partial x_2^2}+\frac{\partial^2}{\partial x_3^2})}G^+(x,y,\omega) = \delta(x-y) \tag{2.44}$$

or

$$\frac{\partial}{\partial x_1}G^-(x,y,\omega) + \frac{i\omega}{c}\sqrt{1+\frac{c^2}{\omega^2}(\frac{\partial^2}{\partial x_2^2}+\frac{\partial^2}{\partial x_3^2})}G^-(x,y,\omega) = \delta(x-y) \tag{2.45}$$

We use convention for the Fourier transform with kernel $e^{-ik\cdot x}$. This means the plane wave decomposition of G^+ is expressed in linear combinations of basis function $e^{ik\cdot x}$. Take vertical wave number be positive downward, then G^+ in equation(2.44) correspond to downgoing and G^- in equation(**??**) corresponds to upcoming waves. The one-way Helmholtz operator

$$\frac{\partial}{\partial x_1} \pm iB = \frac{\partial}{\partial x_1} \pm \frac{i\omega}{c}\sqrt{1+\frac{c^2}{\omega^2}(\frac{\partial^2}{\partial x_2^2}+\frac{\partial^2}{\partial x_3^2})} \tag{2.46}$$

is defined as a pseudo-differential operator of order one

$$\begin{aligned}(\frac{\partial}{\partial x_1} \pm iB)\tilde{p}(x) &= \frac{1}{(2\pi)^3}\int(-ik_1 \pm i\tilde{b}(x,\omega,k_2,k_3))\hat{\tilde{p}}(k)e^{ik\cdot x}dk + q(x) \\ &= \frac{1}{(2\pi)^3}\int(-ik_1 \pm i\tilde{b}(x,\omega,k_2,k_3))\tilde{p}(x')e^{ik\cdot(x-x')}dkdx' + q(x)\end{aligned} \tag{2.47}$$

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 $\frac{\omega}{c}\sqrt{1+\frac{c^2}{\omega^2}(\frac{\partial^2}{\partial x_2^2}+\frac{\partial^2}{\partial x_3^2})}$. The operator (2.46) 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 δG under the assumption of equation(2.43)

$$\delta G(x_r,x_s,t) = \int 2r(x)\frac{\omega^2}{c^2(x)}G^+(x,x_s,\omega)G^-(x_r,x,\omega)e^{iwt}dxd\omega \tag{2.48}$$

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

$$G^-(x,y,\omega) = G^+(y,x,\omega) \tag{2.49}$$

and is frequently used to write the one-way Born modeling equation(2.48) in the form

$$\delta G(x_r,x_s,t) = \int 2r\frac{\omega^2}{c^2(x)}G^+(x,x_s,\omega)G^+(x,x_r,\omega)e^{iwt}dxdw \tag{2.50}$$

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

$$\tilde{I}(x) = \int 2\frac{\omega^2}{c^2}\overline{G^+(x,x_s,\omega)G^+(x,x_r,\omega)}e^{-iwt}d(x_r,x_s,t)dx_rdx_sdtd\omega \tag{2.51}$$

A complete analysis of this integral requires the concept of the double exploding

model *Symes, 2002* [26] and survey sinking *Claerbout, 1985* [1] . 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 straight forward explanation of one-way migration formulated in equation(2.51) can be pursued in view of shot-record migration. Set

$$R(x, x_s, \omega) = \int \overline{G^+(x, x_r, \omega)} d(x_r, x_s, \omega) dx_r \tag{2.52}$$

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 equation(2.44) with a G^+ taking complex conjugate, we have

$$(\frac{\partial}{\partial x_1} - iB)\overline{G^+(x, y, \omega)} = \delta(x - y) \tag{2.53}$$

where $\overline{G^+}$ and G^+ satisfy the same one-way Helmholtz equation understood by looking at their Fourier integral representation. Notice the vertical coordinates of x_s and x_r are zero. Introduce notation $x' = (0, x_{2,3})$ for $x_{2,3} \in \mathbb{R}^2$. We apply operator $\frac{\partial}{\partial x_1} - iB$ to R and S, obtain

$$(\frac{\partial}{\partial x_1} - iB)R(x, x_s, \omega) \quad = \quad \int (\frac{\partial}{\partial x_1} - iB)\overline{G^+(x, x_r, \omega)} d(x_r, x_s, \omega) dx_r$$

$$
\begin{aligned}
&= \int \delta(x - x_r) d(x_r, x_s, \omega) dx_r \\
&= \delta(x_1) d(x', x_s, \omega) \qquad (2.54) \\
(\frac{\partial}{\partial x_1} - iB)S &= (\frac{\partial}{\partial x_1} - iB)\overline{G^+(x, x_s, \omega)} \\
&= \delta(x - x_s) \qquad (2.55)
\end{aligned}
$$

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)

$$(\frac{\partial}{\partial x_1} - iB)\tilde{p} = 0 \qquad (2.56)$$

is achieved through several steps of approximations. First we will ignore the smooth function $q(x)$ in equation (2.47).

$$(\frac{\partial}{\partial x_1} - iB)\tilde{p}(x) \approx \frac{1}{(2\pi)^3} \int (-ik_1 - i\tilde{b}(x, \omega, k_2, k_3))\tilde{p}(x') e^{ik \cdot (x - x')} dk dx' \qquad (2.57)$$

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

$$b(x, \omega, k_1, k_2) = \sigma_1(\tilde{b}) = \frac{w}{c(x)} \sqrt{1 - \frac{c^2(x)}{\omega^2}(k_2^2 + k_3^2)} \tag{2.58}$$

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

$$(\frac{\partial}{\partial x_1} - iB)\tilde{p}(x) \approx \frac{1}{(2\pi)^3} \int (-ik_1 - i\frac{w}{c(x)} \sqrt{1 - \frac{c^2(x)}{\omega^2}(k_2^2 + k_3^2)})\tilde{p}(x')e^{ik\cdot(x-x')} dk dx' \tag{2.59}$$

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

$$\frac{\partial}{\partial x_1} p(x) = \frac{1}{(2\pi)^3} \int i\frac{w}{c(x)} \sqrt{1 - \frac{c^2(x)}{\omega^2}(k_2^2 + k_3^2)} p(x')e^{ik\cdot(x-x')} dk dx' \tag{2.60}$$

The wavefield p agrees with $\tilde{p}$ in its leading order and preserves its wavefront set. The equation (2.60) 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 $\hat{p}(k) = \int p(x)e^{-ik\cdot x} dk$, 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 partial inverse Fourier transform with respect to k_2 and k_3, we obtain

$$\begin{aligned}\frac{\partial}{\partial x_1}p(x) &= \frac{1}{(2\pi)^2}\int i\frac{w}{c(x)}\sqrt{1-\frac{c^2(x)}{\omega^2}(k_2^2+k_3^2)}\hat{p}(x_1,k_2,k_3)e^{i(k_2x_2+k_3x_3)}dk_2dk_3 \\ &= \frac{1}{(2\pi)^2}\int ib(x,\omega,k_2,k_3)\hat{p}(x_1,k_2,k_3)e^{i(k_2x_2+k_3x_3)}dk_2dk_3 \qquad (2.61)\end{aligned}$$

Equation (2.61) has an solution in the integral form

$$p(x_1,x_2,x_3)=\int \hat{p}(x_0,k_2,k_3)e^{\int_{x_0}^{x_1} b(x_1',x_2,x_3,\omega,k_2,k_3)dx_1'}e^{i(k_2x_2+k_3x_3)}dk_2dk_3 \qquad (2.62)$$

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

$$p(x_0+\Delta x_1,x_2,x_3)=\frac{1}{(2\pi)^2}\int \hat{p}(x_0,k_2,k_3)e^{ib\Delta x_1}e^{i(k_2x_2+k_3x_3)}dk_2dk_3 \qquad (2.63)$$

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 $e^{i(k_2x_2+k_3x_3)}$ 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 *Claerbout, 1985* [1]. The polynomial coefficients can be further optimized for large propagation angles *Lee & Sang, 1981* [14]. Let

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

we have

$$ib = i\frac{w}{c(x)}\sqrt{1-\frac{c^2(x)}{\omega^2}(k_2^2+k_3^2)} \approx \frac{i\omega}{c}\{1+\sum_{i=1}^{m}\frac{\alpha_i S}{1+\beta_i S}\} \tag{2.64}$$

where α_i and β_i are coefficients derived by Lee and Suh [14]. 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 (2.64) can be further derived for $m = 1$

$$\begin{aligned} ib &\approx \frac{i\omega}{c}\{1+\frac{\alpha_1 S}{1+\beta_1 S}\} \\ &= \frac{i\omega}{c}\{1+\frac{\alpha_1 S_2}{1+\beta_1 S_2}+\frac{\alpha_1 S_3}{1+\beta_1 S_3}+\frac{-\alpha_1\beta_1 S_2 S_3-\alpha_1\beta_1^2(S_2S_2S_3+S_2S_3S_3)}{(1+\beta_1 S)(1+\beta_1 S_2)(1+\beta_1 S_3)}\} \end{aligned}$$

and dropping the cross-product terms *

$$ib \approx \frac{i\omega}{c}\{1+\frac{\alpha_1 S_2}{1+\beta_1 S_2}+\frac{\alpha_1 S_3}{1+\beta_1 S_3}\} \tag{2.65}$$

and substitute into equation(2.63), it yields

---

*

$$\begin{aligned} \frac{\alpha_1 S}{1+\beta_1 S} &= \frac{\alpha_1(S_2+S_3)(1+\beta_1 S_2)(1+\beta_1 S_3)}{(1+\beta_1 S)(1+\beta_1 S_2)(1+\beta_1 S_3)} \\ &= \frac{\alpha_1 S_2(1+\beta_1 S)(1+\beta_1 S_3)+\alpha_1 S_3(1+\beta_1 S)(1+\beta_1 S_2)-\alpha_1\beta_1^2(S_2^2S_3+S_2S_3^2)}{(1+\beta_1 S)(1+\beta_1 S_2)(1+\beta_1 S_3)} \\ &= \frac{\alpha_1 S_2}{1+\beta_1 S_2}+\frac{\alpha_1 S_3}{1+\beta_1 S_3}+\frac{-\alpha_1\beta_1 S_2 S_3-\alpha_1\beta_1^2(S_2S_2S_3+S_2S_3S_3)}{(1+\beta_1 S)(1+\beta_1 S_2)(1+\beta_1 S_3)} \\ &= \frac{\alpha_1 S_2}{1+\beta_1 S_2}+\frac{\alpha_1 S_3}{1+\beta_1 S_3}+O(S_2)O(S_3)+O(S_2^2)O(S_3)+O(S_3)O(S_3^2) \\ &\approx \frac{\alpha_1 S_2}{1+\beta_1 S_2}+\frac{\alpha_1 S_3}{1+\beta_1 S_3} \end{aligned}$$

$$
\begin{aligned}
p(x_0+\Delta x_1, x_2, x_3) &= \frac{1}{(2\pi)^2}\int \hat{p}(x_0,k_2,k_3) e^{\frac{i\omega}{c}(1+\frac{\alpha_1 S_2}{1+\beta_1 S_2}+\frac{\alpha_1 S_3}{1+\beta_1 S_3})\Delta x_1} e^{i(k_2x_2+k_3x_3)} dk_2 dk_3 \\
&= \frac{1}{(2\pi)^2}\int \hat{p}(x_0,k_2,k_3) e^{\frac{i\omega}{c}(\Delta x_{11}+f(k_2)\Delta x_{12}+f(k_3)\Delta x_{13})} e^{i(k_2x_2+k_3x_3)} dk_2 dk_3
\end{aligned}
$$

where $f(k_2) = -\frac{\alpha_1(c^2/\omega^2)k_2^2}{1-\beta_1(c^2/\omega^2)k_2^2}$, $f(k_3) = -\frac{\alpha_1(c^2/\omega^2)k_3^2}{1-\beta_1(c^2/\omega^2)k_3^2}$ and $\Delta x_{11} = \Delta x_{12} = \Delta x_{13} = \Delta x_1$. We solve $p(x_0+\Delta x_1, x_2, x_3)$ in three steps. Introduce

$$
p^1 = \frac{1}{(2\pi)^2}\int \hat{p}(x_0,k_2,k_3) e^{\frac{i\omega}{c}\Delta x_{11}} e^{i(k_2x_2+k_3x_3)} dk_2 dk_3 \tag{2.66}
$$

$$
p^2 = \frac{1}{(2\pi)^2}\int \hat{p}(x_0,k_2,k_3) e^{\frac{i\omega}{c}(\Delta x_{11}+f(k_2)x_{12})} e^{i(k_2x_2+k_3x_3)} dk_2 dk_3 \tag{2.67}
$$

$$
p^3 = \frac{1}{(2\pi)^2}\int \hat{p}(x_0,k_2,k_3) e^{\frac{i\omega}{c}(\Delta x_{11}+f(k_2)\Delta x_{12}+f(k_3)\Delta x_{13})} e^{i(k_2x_2+k_3x_3)} dk_2 dk_3 \tag{2.68}
$$

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

$$
\frac{\partial p^2}{\partial(\Delta x_{12})} = \frac{p^2-p^1}{\Delta x_1} \tag{2.69}
$$

$$
\frac{\partial p^3}{\partial(\Delta x_{13})} = \frac{p^3-p^2}{\Delta x_1} \tag{2.70}
$$

This can be verified by direct substitution. Take equation (2.69) as an example

$$
\begin{aligned}
\frac{p^2-p^1}{\Delta x_1} &= \frac{1}{(2\pi)^2}\int \hat{p}(x_0,k_2,k_3) e^{\frac{i\omega}{c}\Delta x_{11}} e^{i(k_2x_2+k_3x_3)} \frac{e^{\frac{i\omega}{c}f(k_2)\Delta x_1}-1}{\Delta x_1} dk_2 dk_3 \\
&= \frac{1}{(2\pi)^2}\int \hat{p}(x_0,k_2,k_3) \frac{i\omega}{c} f(k_2) e^{\frac{i\omega}{c}\Delta x_{11}} e^{i(k_2x_2+k_3x_3)} dk_2 dk_3 \\
&= \frac{\partial p^2}{\partial(\Delta x_{12})}
\end{aligned}
$$

Equation (2.70) can be similarly verified. The thin lens term p^1 can be evaluated straight forward from integral equation (2.66)

$$p^1 = p(x_0, x_2, x_3)e^{\frac{i\omega}{c}\Delta x_1} \tag{2.71}$$

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

$$\frac{\partial}{\partial(\Delta x_{12})}p^2 + \beta_1\frac{c^2}{\omega^2}\frac{\partial^2}{\partial x_2^2}(\frac{\partial}{\partial(\Delta x_{12})}p^2) = \alpha_1\frac{c^2}{\omega^2}\frac{i\omega}{c}\frac{\partial^2}{\partial(\Delta x_{12})^2}p^2 \tag{2.72}$$

$$\frac{\partial}{\partial(\Delta x_{13})}p^3 + \beta_1\frac{c^2}{\omega^2}\frac{\partial^2}{\partial x_3^2}(\frac{\partial}{\partial(\Delta x_{13})}p^3) = \alpha_1\frac{c^2}{\omega^2}\frac{i\omega}{c}\frac{\partial^2}{\partial(\Delta x_{13})^2}p^3 \tag{2.73}$$

Combining equations (2.71), (2.72), (2.73), (2.69), and (2.70) a 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 thing screen with thickness Δx_1. We wan 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(2.71) p^1 can be implemented as

$$p^1 = p^0 e^{\frac{i\omega}{c}\Delta x_1} \tag{2.74}$$

The Crank-Nicholson method is used to discretize equation(2.72) on the right hand side (likewise to equation(2.73))

$$\begin{aligned} &\frac{1}{\Delta x_1}\{p^2 - p^1 + \beta_1(c^2/\omega^2)[\frac{\partial^2}{\partial x_2^2}p^2 - \frac{\partial^2}{\partial x_2^2}p^1]\} \\ &= \frac{1}{2}[\alpha_1(c^2/\omega^2)\frac{i\omega}{c}\frac{\partial^2}{\partial x_2^2}p^2] + \frac{1}{2}[\alpha_1(c^2/\omega^2)\frac{i\omega}{c}\frac{\partial^2}{\partial x_2^2}p^1] \end{aligned} \tag{2.75}$$

The second order compact finite-difference operator can be introduced

$$\frac{\partial^2}{\partial x_2^2} = \frac{\delta_{x_2}^2}{1 + \lambda\Delta x_2^2\delta_{x_2}^2} \tag{2.76}$$

where

$$\delta_{x_2}^2 u = \frac{u(x_2 + \Delta x_2) - 2u(x_2) + u(x_2 - \Delta x_2)}{\Delta x_2^2} \tag{2.77}$$

is the central-difference operator, and $\lambda = 1/12$ is the compact coefficient to increase the finite-difference accuracy. Thus this finite-difference scheme becomes

$$\begin{aligned} &\frac{1}{\Delta x_1}\{p^2 - p^1 + (\lambda\Delta x_2^2 + \beta_1(c^2/\omega^2))[\delta_{x_2}^2 p^2 - \delta_{x_2}^2 p^1]\} \\ &\quad = \frac{1}{2}[\alpha_1(c^2/\omega^2)\frac{i\omega}{c}\delta_{x_2}^2 p^2] + \frac{1}{2}[\alpha_1(c^2/\omega^2)\frac{i\omega}{c}\delta_{x_2}^2 p^1] \end{aligned} \tag{2.78}$$

Denote $p^2_{i,j} = p^2(\cdot, i\Delta x_2, j\Delta x_3)$ discretised with interval Δx_2 and Δx_3 in x_2 and x_3 directions, (likewise for p^2 and p^3)

$$\begin{aligned} & p^{k+1}_{i,j}(\lambda + \beta_1 \frac{c^2}{\omega^2 \Delta x_2^2} - \frac{1}{2}\alpha_1 \frac{\Delta x_1}{\Delta x_2^2})[p^{k+1}_{i-1,j} - 2p^{k+1}_{i,j} + p^{k+1}_{i+1,j}] \\ & = p^k_{i,j}(\lambda + \beta_1 \frac{c^2}{\omega^2 \Delta x_2^2} - \frac{1}{2}\alpha_1 \frac{\Delta x_1}{\Delta x_2^2})[p^k_{i-1,j} - 2p^k_{i,j} + p^k_{i+1,j}] \end{aligned} \tag{2.79}$$

This can be rewritten as

$$a_i p^2_{i-1,j} + b_i p^2_{i,j} + c_i p^2_{i+1,j} = a^*_i p^1_{i-1,j} + b^*_i p^1_{i,j} + a^*_i p^1_{i+1,j} \tag{2.80}$$

where

$$\begin{aligned} a_i &= \lambda + \frac{\beta_1 c^2}{\Delta x_2^2 \omega^2} - \frac{1}{2}\alpha_1 \frac{\Delta x_1}{\Delta x_2^2}\frac{i\omega}{c} \\ a^*_i &= \text{complex conjugate of } a_i \\ b_i &= 1 - 2a_i \\ b^*_i &= \text{complex conjugate of } b_i \\ c_i &= a_i \end{aligned}$$

With the set of coefficients as above, the right hand side of equation(2.80) is viewed as a tridiagonal evaluation. A tridiagonal solver can be used to solve the vector of p^2 on the left hand side of equation(2.80). With the same set of coefficients as above except changing Δx_2^2 to Δx_3^2, the equation in x_3 direction can be solved similarly

$$a_i p^3_{i,j-1} + b_i p^3_{i,j} + c_i p^3_{i,j+1} = a^*_i p^2_{i-1,j} + b^*_i p^2_{i,j} + a^*_i p^2_{i+1,j} \tag{2.81}$$

For both 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

Singularities of seismic waves propagate along the bicharististic curves (rays) before they meet the reflector. The wavefront set of the reflectivity distribution, $WF(r)$, determines where the reflection happens as well as which pairs of incident and reflected rays are responsible for the arrival. Migration reconstructs the wavefront set of reflectivity by migrating singularities of reflection signals alone the rays path and bring constructive interference at correct position. The ray paths are sensitive to the change of velocity. At wrong velocities the constructive interference occurs at wrong position due to incorrect ray paths. Therefore, the reconstructed reflectivity by the wrong velocity deviate from the true reflectivity. How to measure this deviation is the main question we pursue to answer in this chapter. It will be not surprised for us to see that there are several valid measurements, and all the justified measurements should be equivalent in the sense that they all capture a image residual defined at the vicinity of reflection point x with singularities of the image removed. Project this image residual back to the velocity model is the main computational machinery for the 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 gradient is important for the purpose of velocity updating and will be discussed in section(4.2.2).

The reconstructed reflectivity (image) obtained by migration deviates from the true reflectivity. This deviation can by 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 offset domain, the other in angle domain. Differential semblance criteria based on local properties of CIGs is used for automatic velocity updating.

3.1 Generalized Born modeling

In order to study common image gathers at inexact velocities, we first generalize the Born modeling so that it invokes the inexact background velocity in the formulation. Introduce the generalized reflectivity distribution $r_g(x,y)$ such that it satisfies

$$\delta G(x_r, x_s, t) = \int 2r_g(x,y)\frac{\omega^2}{c(x)^2}G^+(x, x_s, \omega)G^+(y, x_r, \omega)e^{iwt}dxdyd\omega \tag{3.1}$$

where c is not necessarily the exact background velocity. The Born wavefield evaluated at an inexact 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(2.50) if $r_g(x,y) = r(x)\delta(x-y)$ for $r(x)$ the physical reflectivity distribution and the velocity c is exact.

$$\begin{aligned}\delta G(x_r, x_s, t) &= \int 2r(x)\delta(x-y)\frac{\omega^2}{c^2}G^+(x, x_s, \omega)G^+(y, x_r, \omega)e^{i\omega t}dxdyd\omega \\ &= \int 2r(x)\frac{\omega^2}{c^2}G^+(x, x_s, \omega)G^+(x, x_r, \omega)e^{i\omega t}dxdyd\omega\end{aligned}$$

It is clear that a physical reflection corresponds to the pair of the exact velocity and the physical reflectivity distribtion $r(x)$, 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* [26]. Using the relation $G^+(x,y,\omega) = G^-(y,x,\omega)$, equation (3.1) is expressed in terms of G^- as

$$\delta G(x_r, x_s, t) = \int 2r_g(x,y)\frac{\omega^2}{c^2}G^-(x_s, x, \omega)G^-(x_r, y, \omega)e^{iwt}dxdyd\omega \qquad (3.2)$$

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

The generalized born operator F_g can be introduced, $F_g : r_g \mapsto \delta p$. 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

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

In this section, we shall discuss the calculation of r_g^* through double-squre-root wave equation. It then follows naturally that the solution of double-square-root migration are equavalent 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

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

Let the correlation product of source and receiver wavefields be written as

$$\begin{aligned} u(x,y,x_s,\omega) &= \overline{S(x,x_s,\omega)}R(y,x_s,\omega) \\ &= G^+(x,x_s,\omega)\int G^+(y,x_r,\omega)d(x_r,x_s,\omega)dx_r, \end{aligned} \tag{3.4}$$

and introduce $U(x,y,\omega)$ to be the integral of u with respect to x_s, we arrive at

$$U(x,y,\omega) = \int u(x,y,x_s,\omega)dx_s \tag{3.5}$$

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 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 (2.54) and (2.55), we have

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

Using notations

$$\begin{aligned} x &= (x_1,x_{2,3}), \quad y = (y_1,y_{2,3}) \\ x' &= (0,x_{2,3}), \quad y' = (0,y_{2,3}), \quad x_{2,3},y_{2,3} \in \mathbb{R}^2, \end{aligned}$$

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

$$\begin{aligned} L^+u &= \delta(x_1)\delta(x' - x_s)R(y', x_s, \omega) + \overline{G^+(x', x_s, \omega)}\delta(y_1)d(y', x_s, \omega) \\ &= \delta(x_1)\delta(x' - x_s)d(y', x_s, \omega) + \delta(x' - x_s)\delta(y_1)d(y', x_s, \omega) \end{aligned} \tag{3.6}$$

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

$$\begin{aligned} R(y', x_s, \omega) &= d(y', x_s, \omega) \\ G^+(x', x_s, \omega) &= \delta(x' - x_s) \end{aligned}$$

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

$$\begin{aligned} L^+U &= \int L^+u dx_s \\ &= \int \delta(x_1)\delta(x' - x_s)d(y', x_s, \omega) + \delta(x' - x_s)\delta(y_1)d(y', x_s, \omega)dx_s \\ &= \delta(x_1)d(y', x', \omega) + \delta(y_1)d(y', x', \omega) \\ &= 2\delta(x_1)d(y', x', \omega) \end{aligned} \tag{3.7}$$

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 of survey sinking through the double-square-root continuation *Clarebout, 1985* [1]. Written with complete dependencies, the equation

$$(\frac{\partial}{\partial x_2} - iB(x, -i\partial_x, \omega) - iB(y, -i\partial_y, \omega))U(x, y, \omega) = 2\delta(y_1)d(y, x, \omega) \tag{3.8}$$

is known as the double-square-root wave equation. It is clear from the construction of $U(x, y, \omega) = \int \overline{S(x, x_s, \omega)}R(y, x_s, \omega)dx_s$ 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

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

$$r_g(x - h, x + h)^* = \int 2d(x_r, x_s, \omega)\frac{\omega^2}{c^2}\overline{G^+(x + h, x_s, \omega)G^+(x_r, x - h, \omega)}e^{i\omega t}dx_r dx_s dt d\omega \tag{3.9}$$

where h is understood as sub-surface offset. In view of survey sinking modeling, $|h|$ is half of the horizontal distance between the points at which the downward continued source and receiver wavefields are evaluated. The physical meaning of equation (3.9) can be explained as the reflection signals received at $x - h$ due to a source

located at $x+h$ through the Born propagation. Neglecting the weight factor $\frac{2\omega^2}{c^2}$ in equation(3.3), the singularities of r_g^* is preserved. We may simply write the image with offset parameter h as

$$\begin{aligned} I_h(x,h) &= \int \overline{S(x+h,x_s,\omega)} R(x-h,x_s,\omega) dx_s d\omega \\ &= \int d(x_r,x_s,\omega) \overline{G^+(x+h,x_s,\omega) G^+(x-h,x_r,\omega)} dx_r dx_s d\omega \end{aligned} \quad (3.10)$$

In view of shot-record migration h is not necessarily horizontal, however, for practical

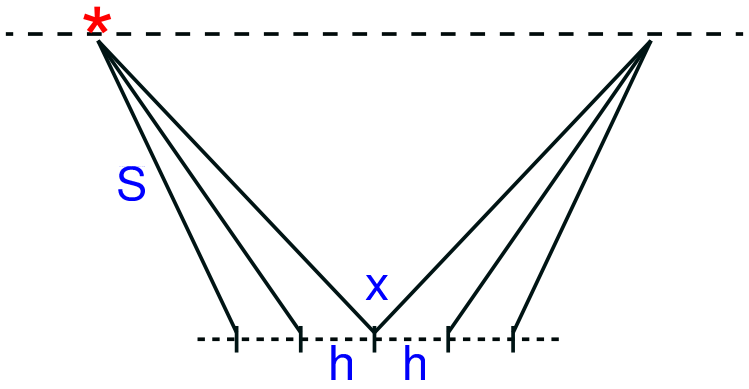


Figure 3.1 $I_h(x,h)$ is obtained by correlation of $S(x+h,x_s,\omega)$ and $R(x-h,x_s,\omega)$ at each point x in the model. The point of correlation x is not necessarily on the surface.

considerations, h will be taken only as horizontal offset. The image parameterized by offsets is called the common image gathers in domains of 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 true velocity, we see clearly that the constructive interference is obtained only at

$h = 0$. Reconstructed general reflectivity distribution with point support on h at origin can be written as

$$I_h(x,h) = I(x)\delta(h) \tag{3.11}$$

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

$$J = \frac{1}{2}||I_h(x,h)h||^2 \tag{3.12}$$

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 θ (see Fig(3.2)). The geometrical meaning of θ is half of the angle enclosed by the wave number k_s of downward continued source wavefield S and the negative of the wave number 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 for S and the local plane wave decomposed for R with their wave numbers bisected by the migration dip vector $\hat{n}$ Fig(3.2).

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

$$P_h : I_h \mapsto hI_h \tag{3.13}$$

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 surface but also an offset in the model. For practical reasons, 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, *Solk & Symes, 2003*[24], 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

$$I_\theta(x_1, x_2, \theta) = \int I_h(x_1 + qh, x_2, h)dh \tag{3.14}$$

where $q = -tan\theta$. For convenience, we have used two dimensional space variable $x = (x_1, x_2)$, and h is restricted to horizontal one-dimensional. The conclusion can be easily extended to three dimensions with careful treatment of the line of integration.

3.3.2 Relationship between I_h and I_θ

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

$$S(x, x_s, \omega) = \int \hat{S}(k_s, x_s, \omega) e^{ik_s \cdot x} dk_s \tag{3.15}$$

Similarly, for downward continued receive 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 ω, the wavenumbers 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}), likewise, k_r be decomposed as (k_{r1}, k_{r2}). Substituting equation(3.15) and equation(3.16) into

equation(3.10), we obtain

$$\begin{aligned} I_h(x,h) &= \int \overline{\hat{S}(k_s,x_s,\omega)}\hat{R}(k_r,x_s,\omega)e^{i(k_r-k_s)\cdot x}dx_s dk_s dk_r d\omega \\ &= \int \overline{\hat{S}(k_s,x_s,\omega)}\hat{R}(k_r,x_s,\omega)e^{-ik_{s1}x_1-k_{s2}(x_2+h)}e^{ik_{r1}x_1+k_{r2}(x_2-h)}dx_s d\omega dk_s dk_r \end{aligned} \tag{3.18}$$

Introduce local coordinate $(\hat{t},\hat{n})$ as shown in fig(3.2) with $\hat{n}$ the unit vector of migration dip and $\hat{t}$ the unit vector bisects the angle in 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

$$\begin{aligned} k_s &= \frac{\omega}{c}(\hat{t}\sin\theta + \hat{n}\cos\theta) \\ k_r &= \frac{\omega}{c}(\hat{t}\sin\theta - \hat{n}\cos\theta) \end{aligned} \tag{3.19}$$

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

$$\begin{aligned} [\mathscr{R}_q I_h(x,h)](x,q) &= \int I_h(x_1+qh,x_2,h)dh \\ &= \int \overline{\hat{S}(k_s,x_s,\omega)}\hat{R}(k_r,x_s,\omega)e^{-ik_{s1}(x_1+qh)-ik_{s2}(x_2+h)} \times \\ &\quad e^{ik_{r1}(x_1+qh)+ik_{r2}(x_2-h)}dh dx_s d\omega dk_s dk_r \\ &= \int \overline{\hat{S}(k_s,x_s,\omega)}\hat{R}(k_r,x_s,\omega)e^{i(k_r-k_s)\cdot x} \times \\ &\quad e^{i(-k_{s1}q-k_{s2}+k_{r1}q-k_{r2})h}dh dx_s dk_s dk_r d\omega \\ &= \int \overline{\hat{S}(k_s,x_s,\omega)}\hat{R}(k_r,x_s,\omega)e^{i(k_r-k_s)\cdot x} \times \end{aligned}$$

$$
\begin{aligned}
& 2\pi\delta[q(k_{r1}-k_{s1})-(k_{r2}+k_{s2})]dx_s dk_s dk_r d\omega \\
= \; & 2\pi\int_{L(k_s,k_r,q)=0}\overline{\hat{S}(k_s,x_s,\omega)}\hat{R}(k_r,x_s,\omega)e^{i(k_r-k_r)\cdot x}dk_s dk_r dx_s d\omega
\end{aligned}
\tag{3.20}
$$

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

$$
L(k_s,k_r,q)=\frac{k_{r2}+k_{s2}}{k_{r1}-k_{s1}}-q=0 \tag{3.21}
$$

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 $\overline{S}$ and 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 $I_\theta(x,\theta)$

$$
I_\theta(x,\theta)=\int_{L(k_s,k_r,q)=0}\overline{\hat{S}(k_s,x_s,\omega)}\hat{R}(k_r,x_s,\omega)e^{i(k_s+k_r)\cdot x}dk_s dk_r dx_s d\omega \tag{3.22}
$$

We have thus verified that the angle domain image I_θ relates to the offset domain image through the Radon transform

$$
I_\theta := I_q(x,q)|_{q=-\tan(\theta)}=[\mathscr{R}_q I_h(x,h)](x,q)|_{q=-\tan(\theta)} \tag{3.23}
$$

*Let α 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 $k_{s1}=\frac{\omega}{c}(-\sin\alpha\sin\theta+\cos\alpha\cos\theta)$, $k_{s2}=\frac{\omega}{c}(\cos\alpha\sin\theta-\sin\alpha\cos\theta)$, $k_{r1}=\frac{\omega}{c}(\cos\alpha\sin\theta+\sin\alpha\cos\theta)$, $k_{r2}=\frac{\omega}{c}(-\sin\alpha\sin\theta-\cos\alpha\cos\theta)$. The result follows by direct substitution $-\tan\theta=(k_{r2}+k_{s2})/(k_{r1}-k_{s1})$

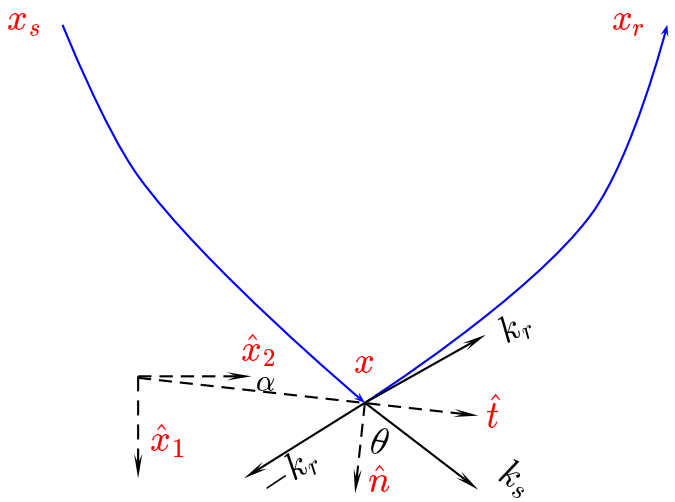


Figure 3.2 $\hat{n}$ migration dip unit vector, θ scattering angle. The wavenumber of $G^+(x, x_r, \omega)$ points downward. So the wavenumber of $\overline{G^+}(x, x_r, \omega)$ points upward

3.3.3 Angle domain DSO

For convenience of the writing 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

$$I_h(x,h) = I(x)\delta(h) \tag{3.24}$$

Thus, due to discussion from last section, we have

$$I_q(x,q) = \int I_h(x_1 + qh, x_2, h)dh = I(x) \tag{3.25}$$

which is independent of the choice of q. This result implies that the common image gathers in angle is 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

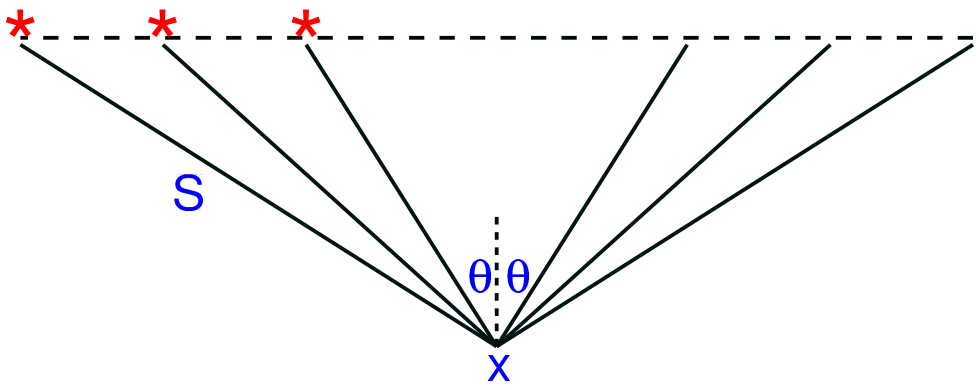


Figure 3.3

$$\frac{\partial I_\theta}{\partial \theta} = 0 \tag{3.26}$$

in the same time

$$\frac{\partial I_q}{\partial q} = 0 \tag{3.27}$$

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

$$\min \frac{1}{2}||P_q I_q||^2 \tag{3.28}$$

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* [24]. 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 $I_q(x,q) = \mathscr{R}_q\{I_h(x,h)\}(x,q)$. Written in terms of I_q, we have in angle

$$P_q I_q = P_q \mathscr{R}_q I_h \tag{3.29}$$

which can be inverse Radon transformed to an image gather in offset $\mathscr{R}_q^{-1} P_h \mathscr{R}_q I_q$. If differential semblance criteria on offset or angle are equivalent, then it suggests that the operators P_h and $\mathscr{R}_q^{-1} P_q \mathscr{R}_q$ should be equivalent. It also implies that $P_h P_h$ and $\mathscr{R}_q P_q P_q \mathscr{R}_q^{-1}$ are equivalent. the inverse Radon transform can be formuated in two dimensions as, *Deans, S. R., 1983*[3]*

*The formular givin in [3] is $\mathscr{R}^{-1} = \frac{-1}{2\pi}\frac{d}{dt} H_t \mathscr{R}^*$. The pseudo-differential operator $\Delta_t^{\frac{1}{2}}$ can be shown in one dimension as $\Delta_t^{\frac{1}{2}} = -i\frac{d}{dt}H_t$ where H_t is the Hilbert transform in one dimensional

$$\mathscr{R}_q^{-1} = i\Delta^{\frac{1}{2}}\mathscr{R}_q^* \tag{3.30}$$

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 a taking a derivative in a pseudo-differential sense.

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

$$\mathscr{R}_qf(x_1,x_2,h) := \int f(x_1+qh,x_2,h)dh = \check{f}(x_1,x_2,q) \tag{3.31}$$

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

$$\begin{aligned}(\mathscr{R}_qf)(x_1,q) &= \int f(t,h)\delta(t-qh-x_1)\ dt\ dh \\ &= \int \frac{1}{2\pi}f(t,h)e^{iw(t-qh-x_1)}dtdhd\omega \end{aligned} \tag{3.32}$$

variable t. Equation (3.30) is posed in the form to agree with the inverse Radon transform in n dimension which can be formuated in unified form using the fractional Laplacian $\mathscr{R}^{-1} = a_ni^{1-n}\Delta^{(n-1)/2}\mathscr{R}^*$ for a_n some real number dependent on n. [3].

*$\Delta^{\frac{1}{2}}$ applies to $f(x)$ for $x \in \mathbb{R}^n$ is understood as $\Delta^{\frac{1}{2}}f(x) = \frac{1}{(2\pi)^n}\int \hat{f}(k)i|k|e^{ik\cdot x}dk$ where $|k| = \sqrt{\sum_j k_j^2}$, and $\hat{f}(k) = \int f(x)e^{-ik\cdot x}dx$

We then see the kernel of the adjoint Radon transform is $\frac{1}{2\pi}e^{-iw(t-qh-x_1)}$. The adjoint Radon transform can be written as

$$\begin{aligned}(\mathscr{R}_q^*\check{f})(t,h) &= \int \frac{1}{2\pi}\check{f}(x_1,q)e^{-iw(t-qh-x_1)}dtdhd\omega \\ &= \int \check{f}(t-qh,q)\ dq\end{aligned}$$

Using the relation $P_q^* = -\frac{\partial}{\partial q}$, we derive for $\mathscr{R}_q^*P_q^*P_q\mathscr{R}_qI_h$.

$$\begin{aligned}\mathscr{R}_q^*P_q^*P_q\mathscr{R}_qI &= \mathscr{R}_q^*(-\frac{\partial^2}{\partial q^2})\mathscr{R}_qI_h \\ &= \mathscr{R}_q^*\int -\frac{\partial^2}{\partial q^2}I(t,h)e^{i\omega(t-qh-x_1)}\frac{1}{2\pi}dtdhd\omega \\ &= \int h'^2\omega'^2I(t',h')e^{i\omega'(t'-qh'-x_1)}\frac{1}{4\pi^2}e^{-i\omega(t-qh-x_1)}dx_1dqdt'dh'd\omega'd\omega \\ &= \int h'^2\omega^2I(t',h')e^{iw(t'-t)}e^{i\omega q(h'-h)}\frac{1}{2\pi}dqdh'dt'd\omega \\ &= \int h^2|\omega|I(t',h)e^{i\omega(t'-t)}dt'd\omega \\ &= \int_{-\infty}^{+\infty}dt'\ h^2I(t',h)\int_{-\infty}^{+\infty}|\omega|e^{i\omega(t'-t)}d\omega \\ &= -2h^2\int_{-\infty}^{+\infty}dt'\ \frac{I(t',h)}{(t-t')^2} \qquad (3.33)\end{aligned}$$

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

$$\int_{-\infty}^{\infty}|\omega|e^{iwz}d\omega = -\frac{2}{z^2} \qquad (3.34)$$

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

$$(\mathscr{R}_q^*P_q^*P_q\mathscr{R}_qI_h)(t,h) = P_h^*P_h\int -2\frac{I_h(t',h)}{(t-t')^2}dt' \qquad (3.35)$$

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

The transform $\frac{\partial}{\partial t}H_t$ can be analyzed using its Fourier representation, write

$$[(\frac{\partial}{\partial t}H_t)f(t',h)](t,h) = \int |\omega| f(t',h)e^{i\omega(t'-t)}dt'd\omega \tag{3.36}$$

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

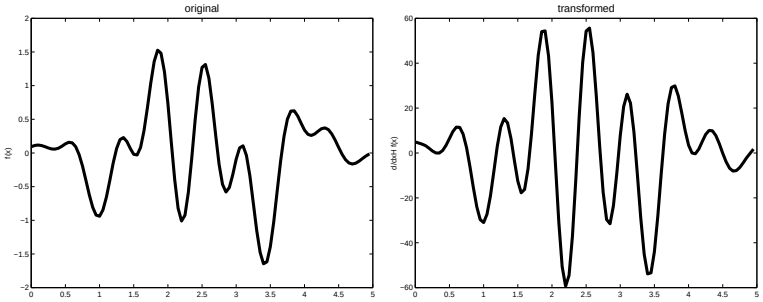


Figure 3.4 letf: original signal, right: transformed signal

*definition of Hilbert transform: $(Hf)(t) = \frac{1}{\pi}\int_{-\infty}^{+\infty} dt' \frac{f(t')}{t'-t}$.

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 frequencies. (see figure(**??**) Due to this reasoning, we consider differential semblance operator P_h is a good substitute for $P_q\mathscr{R}_q$.

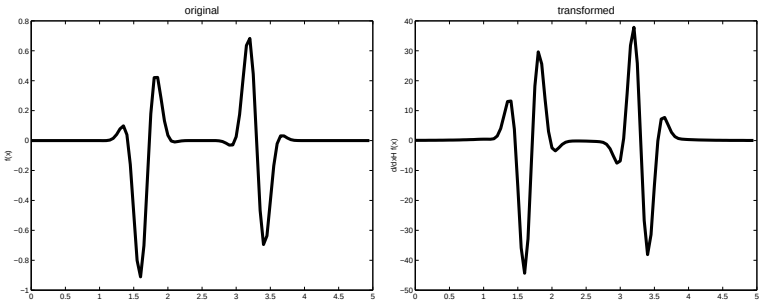


Figure 3.5 left: original signal with sharp peaks, right: transform signal

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

$$\frac{1}{2}||P_hI_h||^2 \;=\; \frac{1}{2} < P_hI_h, P_hI_h > \tag{3.37}$$

In discrete form, we have

$$\frac{1}{2}||P_hI_h||^2 = \sum_h \sum_x \sum_z h^2 \overline{I(x_1,x_2,h)} I(x_1,x_2,h) \tag{3.38}$$

Chapter 4
Adjoint state calculation

4.1 Gradient calculation and adjoint state analysis

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 integral representations and continuous as well as discrete wave equation representations. Extensive adjoint state analysis is used to derive for discrete gradient calculation within the frame work of one-way wave equation migration. The limited memory BFGS algorithm is implemented to solve the inverse problem where the smoothing properties of the gradient is analyzed and a B-spline smoothing scheme is developed.

4.1.1 Gradient formulation by the Green's function representation

vspace-0.5cm

The offset domain differential semblance objective function is written as

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

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

$$P_h : I_h(x,h) \rightarrow hI_h(x,h) \tag{4.2}$$

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

$$I_h = f[c] \tag{4.3}$$

and

$$\delta I_h = (DI_h)[c]\delta c = \frac{\partial I_h}{\partial c}\delta c \tag{4.4}$$

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 objective function written as an inner product $<,>$ can be derived

$$\begin{aligned}
\delta J &= \tfrac{1}{2} < \delta(P_h I_h), P_h I_h > + \tfrac{1}{2} < P_h I_h, \delta(P_h I_h) > \\
&= \tfrac{1}{2} < P_h \; DI_h \; \delta c, P_h I_h > + \tfrac{1}{2} < P_h I_h, P_h \; DI_h \; \delta c > \\
&= \tfrac{1}{2} < \delta c, DI_h^* P_h^* P_h I_h > + \tfrac{1}{2} < DI_h^* P_h^* P_h I_h, \delta c > \\
&= \tfrac{1}{2} < \delta c, DI_h^* P_h^* P_h I_h > + \tfrac{1}{2} < \delta c, \overline{DI_h^* P_h^* P_h I_h} > \\
&= < \delta c, Re(DI_h^* P_h^* P_h I_h) >
\end{aligned} \tag{4.5}$$

the gradient of J with respect to c can then be read*

$$\nabla_c J = Re\{(DI_h)^*(P_h^* P_h I_h)\} \tag{4.6}$$

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 conjugate of the integration kernel. Sub-surface offset semblance image is formulated by adjoint of the generalized Born modeling. Equation (3.10) is re-written in here as

$$I_h(x,h) = \int \overline{G^+(x-h,x_r,\omega)} d(x_r,x_s,t) \overline{G^+(x+h,x_s,\omega)} e^{-i\omega t} dt dx_r dx_s d\omega \tag{4.7}$$

where $d(x_r, x_s, \omega)$ is the data in angular frequency ω received at x_r due to a point source at x_s, $G(x, y, \omega)$ is the Green's function observed at point $x \in \mathbb{R}^3$ due to

*An angle domain differential semblance objective function can be written as $\tilde{J} = \frac{1}{2}||P_q \mathscr{R}_q I_h||_2^2$. The gradient of $\tilde{J}$ with respect to velocity c can be shown similarly as $\nabla_c \tilde{J} = Re(DI_h^* \mathscr{R}_q^* P_q^* P_q \mathscr{R}_q I_h)$. We see the gradient of the angle domain DSO and that of offset domain DSO differ in the operators $\mathscr{R}_q^* P_q^* P_q \mathscr{R}_q$ and $P_h^* P_h$.

a point source located at $y \in \mathbb{R}^3$. A 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 $G^+(x-h, x_r, \omega)$,

$$\begin{aligned}
\delta I_h(x,h) &= \int \overline{G^+(y,x_r,\omega)\frac{\delta c(y)\omega^2}{c(y)^2}G^+(x-h,y,\omega)}d(x_r,x_s,t)\overline{G^+(x+h,x_s,\omega)} \times \\
& e^{-i\omega t}dtdx_r dx_s dyd\omega \\
& + \int \overline{G^+(x-h,x_r,\omega)}d(x_r,x_s,t)\overline{G^+(x+h,y,\omega)\frac{\delta c(y)\omega^2}{c(y)^2}G^+(y,x_s,\omega)} \times \\
& e^{-i\omega t}dtdx_r dx_s dyd\omega \qquad (4.8)
\end{aligned}$$

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

$$\begin{aligned}
g(y) &= (DI_h)^* u \\
&= \int G^+(y,x_r,\omega)\frac{u(x,h)w^2}{c(y)^2}G^+(x-h,y,\omega) \times \\
& d(x_r,x_s,t)G^+(x+h,x_s,w)e^{i\omega t}dtdx_r dx_s dxdhd\omega \\
& + \int G^+(x-h,x_r,\omega)d(x_r,x_s,t)G^+(x+h,y,\omega) \times \qquad (4.9) \\
& \frac{u(x,h)w^2}{c(y)^2}G^+(y,x_s,\omega)e^{i\omega t}dtdx_r dx_s dxdhd\omega
\end{aligned}$$

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

$$g_1(y) = \int G^+(y,x_r\omega)\frac{u(x,h)w^2}{c(y)^2}G^+(x-h,y,\omega) \times$$

$$
\begin{aligned}
& d(x_r, x_s, t) G(x + h, x_s, \omega) e^{i\omega t} dx_r dt dx_s dx dh d\omega \\
= & \int \frac{\omega^2}{c(y)^2} dx_s d\omega \left(\int G^+(y, x_r, \omega) d(x_r, x_s, t) e^{i\omega t} dt dx_r \right) \times \\
& \int G^+(x + h, x_s, w) u(x, h) G^+(y, x - h, w) dx dh
\end{aligned} \tag{4.10}
$$

This integral carries in three steps: first, downward continue receiver wavefield complex conjugated,

$$
\begin{aligned}
\overline{R(y, x_s, \omega)} &= \int G^+(y, x_r \omega) d(x_r, x_s, t) e^{i\omega t} dx_r dt \\
&= \overline{\int \overline{G^+(x_r, y, w)} d(x_r, x_s, \omega) dx_r}
\end{aligned} \tag{4.11}
$$

which is just the downward continued receiver wavefield taken complex conjugate. Second, downward continued source wavefield, correlate in offset h with residual image $u(x, h)$ and then upward propagate to point y.

$$
g_s(y, x_s, \omega) = \int G^+(x + h, x_s, \omega) u(x, h) G^+(y, x - h, \omega) dx dh \tag{4.12}
$$

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

$$
g_1(y) = \int \frac{\omega^2}{c(y)^2} \overline{R(y, x_s, \omega)} g_s(y, x_s, \omega) dx_s d\omega \tag{4.13}
$$

Similarly, we can write the second term in equation (4.9) as

$$
g_2 = \int \frac{\omega^2}{c(y)^2} \overline{S(y, x_s, \omega)} g_r(y, x_s, \omega) dx_s d\omega \tag{4.14}
$$

where

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

is the downward continued source wavefield, and

$$g_r(y, x_s, \omega) = \int G^+(x - h, x_r, \omega) d(x_r, x_s, t) e^{i\omega t} G^+(x + h, y, \omega) dt dx dh dx_r \qquad (4.15)$$

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

$$g(y) = \int \frac{\omega^2}{c(y)^2} (\overline{R} g_s + S g_r)(y, x_s, \omega) dx_s d\omega \qquad (4.16)$$

The three conceptual steps for taking the integral shown in 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 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 is 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. Eqch sub-integral can be shown, depending on the Green's function used is two-way Green's function or one-way Green's function, to satisfy two-way or one-way wave equations, respectively. The gradient can be then formuated 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 formuate the gradient. Let's first repeat the migration formula using two-way Green's functions:

$$I_h(x,h) = \int \overline{\tilde{S}(x,x_s,\omega)}\tilde{R}(x,x_s,\omega)\ dx_s d\omega \tag{4.17}$$

where $\tilde{S}$ is the downward continued source wavefield, $\tilde{S}(x+h,x_s,\omega) = \tilde{G}(x+h,x_s,\omega)$, and $\tilde{R}$ is the downward continued receiver wavefield, $\tilde{R}(x,x_s,\omega) = \int d(x_r,x_s,\omega)\overline{\tilde{G}(x_r,x,\omega)}dx_r$. $\tilde{S}$ solves Cauchy problem:

$$\begin{aligned} (\nabla^2 + \frac{w^2}{c^2})\tilde{S}(x;x_s,\omega) &= \delta(x-x_s) \\ (\alpha\tilde{S} + \beta\nabla\tilde{S})|_\Sigma &= 0 \end{aligned} \tag{4.18}$$

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 α, β constant. Likely, $\tilde{R}$ solves

$$\begin{aligned}(\nabla^2 + \frac{w^2}{c^2})\tilde{R}(x;s,\omega) &= \int d(x_r;x_s,\omega)\delta(x-x_r)dx_r \\ (\alpha\tilde{R} + \beta\nabla\tilde{R})|_\Sigma &= 0\end{aligned} \tag{4.19}$$

Write perturbation of $I_h(x,h)$ as

$$\delta I_h = \int (\overline{\delta\tilde{S}}\tilde{R} + \overline{\tilde{S}}\delta\tilde{R})(x,x_s,h,\omega)dx_s d\omega \tag{4.20}$$

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

$$\begin{aligned}\langle\delta I_h, u\rangle &= \langle DI_h\delta c, u\rangle \\ &= \langle\delta c, (DI_h)^* u\rangle \\ &= \int (\overline{\delta I}(x,h)u(x,h)\ dxdh \\ &= \int \delta\tilde{S}(x+h,x_s,\omega)\overline{\tilde{R}(x-h,x_s,\omega)}u(x,h)dx_s d\omega dxdh + \\ &\quad \int \tilde{S}(x+h,x_s,\omega)\overline{\delta\tilde{R}}(x-h,x_s,\omega)\ u(x,h)dx_s d\omega dxdh \\ &= \int \delta\tilde{S}(x+h,x_s,\omega)\{\int \overline{\tilde{R}(x-2h,x_s,\omega)}u(x-h,h)dh\}dxdx_s d\omega + \\ &\quad \int \delta\overline{\tilde{R}(x-h,x_s,\omega)}\{\int \tilde{S}(x+2h,x_s,\omega)u(x+h,h)dh\}dxdx_s d\omega\end{aligned}$$

introduce $\tilde{g_r}$ and $\tilde{g_s}$ which satisfy

$$\begin{cases} (\nabla^2 + \frac{w^2}{c^2})\tilde{g_r}(x; x_s, \omega) &= \int \overline{\tilde{R}(x - 2h, x_s, \omega)} u(x - h, h) dh \\ \int_\Sigma (\nabla \tilde{S} \tilde{g_r} - \nabla \tilde{g_r} \tilde{S}) da &= 0 \end{cases} \tag{4.21}$$

$$\begin{cases} (\nabla^2 + \frac{w^2}{c^2})\tilde{g_s}(x, x_s, \omega) &= \int \tilde{S}(x + 2h, x_s, \omega) u(x + h, h) dh \\ \int_\Sigma (\nabla \tilde{R} \tilde{g_s} - \nabla \tilde{g_s} \tilde{R}) da &= 0 \end{cases} \tag{4.22}$$

Here da is the differential surface area on Σ. 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 conditionswill be used in the integration by parts later to yield desired results. Substitute g and q into eq(4.21), and then do integration by parts, we derive

$$\begin{aligned} \langle \delta I, u \rangle &= \int \{ \delta \tilde{S} (\nabla^2 + \frac{w^2}{c^2}) \tilde{g_r} + \delta \overline{\tilde{R}} (\nabla^2 + \frac{w^2}{c^2}) \tilde{g_s} \} dx_s d\omega dx \\ &= \int [(\nabla^2 + \frac{w^2}{c^2}) \delta \tilde{S}] \tilde{g_r} + [(\nabla^2 + \frac{w^2}{c^2}) \delta \overline{\tilde{R}}] \tilde{g_s} dx_s dx d\omega \\ &= \int \frac{2w^2}{c^3} \delta c(x) \tilde{S} \tilde{g_r} + \frac{2w^2}{c^3} \delta c(x) \overline{\tilde{R}} \tilde{g_s} dx_s d\omega dx \end{aligned} \tag{4.23}$$

here we have used the Born scattering relations

$$(\nabla^2 + \frac{w^2}{c^2}) \tilde{S} = \delta(x - x_s) \quad \Rightarrow \quad (\nabla^2 + \frac{w^2}{c^2}) \delta \tilde{S} = \frac{2w^2}{c^3} \delta c \tilde{S} \tag{4.24}$$

$$(\nabla^2 + \frac{w^2}{c^2}) \tilde{R} = d(x_r, x_s, \omega) \delta(x - x_s) \quad \Rightarrow \quad (\nabla^2 + \frac{w^2}{c^2}) \delta \tilde{R} = \frac{2w^2}{c^3} \delta c \tilde{R} \tag{4.25}$$

from eq(4.21), we can read off $(DI_h)^* u$ as

$$(DI_h)^*(u) = \int \frac{2w^2}{c^3} (\tilde{S} \tilde{g_r} + \overline{\tilde{R}} \tilde{g_s})(x, x_s, \omega) dx_s d\omega \tag{4.26}$$

Gradient by one way-wave equation

Next I shall derive $(DI_h)^*$ using one way wave equation. We start with a similar approach as we did using two-way wave equation. In last section, S, R, g and q are solved using two-way wave operator $\nabla^2 + \frac{w^2}{c^2}$. In this section we introduce one-way wave operators $\frac{\partial}{\partial x_1} + i\sqrt{\frac{\partial^2}{\partial x_{1,2}^2} + \frac{w^2}{c^2}}$ and $\frac{\partial}{\partial x_1} - i\sqrt{\frac{\partial^2}{\partial x_{1,2}^2} + \frac{w^2}{c^2}}$.

For convenience, let's first write down the corresponding one-way wave relations that S and R satisfy.

$$\begin{cases} (\frac{\partial}{\partial x_1} - i\sqrt{\frac{\partial^2}{\partial x_{1,2}^2} + \frac{w^2}{c^2}})S(z, x; s, \omega) = \delta(x - x_s) \\ \int_{\Sigma(z)}(\nabla_x Sg - \nabla_x gS)da = 0 \end{cases} \tag{4.27}$$

and

$$\begin{cases} (\frac{\partial}{\partial x_1} + i\sqrt{\frac{\partial^2}{\partial x_{1,2}^2} + \frac{w^2}{c^2}})R(z, x; s, w) = \int d(x_r, x_s, \omega)\delta(x - x_r)dx_r \\ \int_{\Sigma(z)}(\nabla_x Rg_s - \nabla_x g_s R)da = 0 \end{cases} \tag{4.28}$$

Introduce g_r and g_s, which satisfy the following equations.

$$\begin{cases} (\frac{\partial}{\partial x_1} + i\sqrt{\frac{\partial^2}{\partial x_{1,2}^2} + \frac{w^2}{c^2}})g_r(x, x_s, \omega) \;\; = \;\; \int \overline{R(x - 2h, x_s, \omega)}u(x - h, h)\; dh \\ g_r(x_1 = z_l, x_{1,2}, x_s, \omega) = 0 \\ \int_{\Sigma(x_1)}(\nabla_x Sg_r - \nabla_x g_r S)da = 0 \end{cases} \tag{4.29}$$

where z_l is the maximum depth that the computational propagation can reach.

$$\begin{cases} (\frac{\partial}{\partial x_1} + i\sqrt{\frac{\partial^2}{\partial x_{1,2}^2} + \frac{w^2}{c^2}})g_s(z,x;s,w) = \int S(z,x+2h;s,w)u(z,x+h,h)\ dh \\ g_r(z=z_l,x;s,w) = 0 \\ \int_{\Sigma(x_1)}(\nabla_x R g_s - \nabla_x g_s R)da = 0 \end{cases} \tag{4.30}$$

Here $\Sigma(x_1)$ represents the horizontal boundary for each x_1. substituting g_r and g_s into eq(4.21) and noticing,

$$\begin{aligned} \sqrt{\frac{\partial^2}{\partial x_{1,2}^2} + \frac{w^2}{c^2}} &\simeq \frac{w^2}{c^2} + \frac{1}{2}\frac{\partial^2}{\partial x_{1,2}^2} \\ \int \delta S(\frac{\partial}{\partial x_1} g_r)dx_1 &= \delta S g_r|_{x_1=z_l}^{x_1=0} - \int (\frac{\partial}{\partial x_1}\delta S)g_r dx_1 \\ &= -\int (\frac{\partial}{\partial x_1}\delta S)g_r dx_1 \end{aligned}$$

then performing integration by parts, we obtain

$$\begin{aligned} \langle \delta I, u\rangle &= \int \{\ \delta S(\frac{\partial}{\partial x_1} + i\sqrt{\frac{\partial^2}{\partial x_{1,2}^2} + \frac{w^2}{c^2}})g_r + \delta\overline{R}(\frac{\partial}{\partial x_1} + i\sqrt{\frac{\partial^2}{\partial x_{1,2}^2} + \frac{w^2}{c^2}})g_s\ \}dx_s d\omega dx \\ &= \int [(-\frac{\partial}{\partial x_1} + i\sqrt{\frac{\partial^2}{\partial x_{1,2}^2} + \frac{w^2}{c^2}})\delta S]g_r + [(-\frac{\partial}{\partial x_1} + i\sqrt{\frac{\partial^2}{\partial x_{1,2}^2} + \frac{w^2}{c^2}})\delta\overline{R}]g_s\ dx_s d\omega dx \\ &= \int (i(\frac{\partial^2}{\partial x_{1,2}^2} + \frac{w^2}{c^2})^{-\frac{1}{2}}\frac{2w^2}{c^3}S)(\delta c)g_r + (i(\frac{\partial^2}{\partial x_{1,2}^2} + \frac{w^2}{c^2})^{-\frac{1}{2}}\frac{2w^2}{c^3}\overline{R})(\delta c)g_s\ dx_s d\omega dx \end{aligned} \tag{4.31}$$

In the last $''='$, we have taken the perturbation of the complex conjugate of equations (14). If we let

$$A = \frac{\partial}{\partial x_1} - i\sqrt{\frac{\partial^2}{\partial x_{1,2}^2} + \frac{w^2}{c^2}} \tag{4.32}$$

and

$$\overline{A} = -\frac{\partial}{\partial x_1} + i\sqrt{\frac{\partial^2}{\partial x_{1,2}^2} + \frac{w^2}{c^2}} \tag{4.33}$$

then it can be verified that

$$[\frac{\partial \overline{A}}{\partial c} S^*]^* = (i(\frac{\partial^2}{\partial x_{1,2}^2} + \frac{w^2}{c^2})^{-\frac{1}{2}} \frac{2w^2}{c^3})(S) \tag{4.34}$$

and

$$[\frac{\partial \overline{A}}{\partial c} R]^* = (i(\frac{\partial^2}{\partial x_{1,2}^2} + \frac{w^2}{c^2})^{-\frac{1}{2}} \frac{2w^2}{c^3})(\overline{R}) \tag{4.35}$$

We then read $DI_h^* u$ from equation (12) as

$$\begin{aligned} DI_h^*(u) &= \int \{[i(\frac{\partial^2}{\partial x_{1,2}^2} + \frac{w^2}{c^2})^{-\frac{1}{2}} \frac{2w^2}{c^3}](S) g_r + [i(\frac{\partial^2}{\partial x_{1,2}^2} + \frac{w^2}{c^2})^{-\frac{1}{2}} \frac{2w^2}{c^3}](\overline{R}) g_s\} \; dx_s d\omega \\ &= \int \{[\frac{\partial \overline{A}}{\partial c} S^*]^* g_r + [\frac{\partial \overline{A}}{\partial c} R]^* g_s\} \; dx dx_s d\omega \end{aligned} \tag{4.36}$$

4.1.3 Gradient by discrete adjoint state calculation

Numerical optimization algorithms depend on accurate adjoint state calculation. There are in general two ways to calculate the adjoint: discretise the continuous form, or directly take the adjoint of discretised form. I shall in this section discuss adjoint state calculation of the second kind. 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

$$H : X \to Y \tag{4.37}$$

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

$$H^i : Z^{i-1} \to Z^i \qquad (i = 1, ..., N) \tag{4.38}$$

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

$$y = x \tag{4.39}$$

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. The algorithm H is considered as composition of differentiable functions H^is,

$$H = H^N \circ H^{N-1} \circ ... \circ H^1 := \prod_{i=1}^{N} \circ H^i \tag{4.40}$$

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

$$Z^i = H^i(Z^{i-1}) \qquad 1 \leq i \leq N \tag{4.41}$$

A variation δZ^i depends on a variation of the control variables δZ^{i-1}.

$$\delta Z^i = \frac{\partial H^i}{\partial Z^{i-1}}|_{Z^{i-1}=H(Z^{i-2})} \delta Z^{i-1} \tag{4.42}$$

Therefore, the perturbation at last step is

$$\delta Z^l = \frac{\partial H^N}{\partial Z^{N-1}}|_{Z^{N-1}=\prod_{i=1}^{N-1} \circ H^i} \cdot ... \cdot \frac{\partial H^2}{\partial Z^1}|_{Z^1=H(Z^0)} \cdot \frac{\partial H^1}{\partial Z^0} \delta Z^0 \tag{4.43}$$

The Jacobian is defined by

$$A(Z^0) = \frac{\partial H}{\partial Z^0} \tag{4.44}$$

which can be read off from eq(4.43) as

$$\frac{\partial H}{\partial Z^0} = \frac{\partial H^N}{\partial Z^{N-1}}|_{Z^{N-1}=\prod_{i=1}^{N-1} \circ H^i} \cdot ... \cdot \frac{\partial H^2}{\partial Z^1}|_{Z^1=H(Z^0)} \cdot \frac{\partial H^1}{\partial Z^0} \tag{4.45}$$

In matrix multiplication form, it can be expressed as

$$A(Z^0) = A^N(Z^{N-1}) \cdot A^{N-1}(Z^{N-2}) \cdot ... \cdot A^1(Z^0) \tag{4.46}$$

where

$$A^i(Z^{i-1}) = \frac{\partial H^i}{\partial Z^{i-1}}|_{Z^{i-1}=\prod_{j=1}^{i-1}\circ Hj} \tag{4.47}$$

4.1.3.2 Numerical adjoint code construction

The adjoint of Jacobian in matrix form is simply

$$\begin{aligned}(\frac{\partial H}{\partial Z^0})^* &= (A^1)^* \cdot ... \cdot (A^N)^* \\ &= (\frac{\partial H^1}{\partial Z^0})^* \cdot ... \cdot (\frac{\partial H^N}{\partial Z^{N-1}})^*|_{Z^{N-1}=\prod_{i=1}^{N-1}\circ H^i}\end{aligned} \tag{4.48}$$

Introduce adjoint state variables DZ^{N-1},...,DZ^0, which satisfy the inner product relation

$$< DZ^{i-1}, \delta Z^{i-1} > \;=\; < DZ^i, \delta Z^i > \tag{4.49}$$

This can be further derived

$$< DZ^i, \delta Z^i > \;=\; < DZ^i, A^i(Z^{i-1})\delta Z^{i-1} > \tag{4.50}$$

$$\;=\; < (A^i(Z^{i-1})^* DZ^i, \delta Z^{i-1} > \tag{4.51}$$

This holds for all δZ^{i-1}, so it follows

$$DZ^{i-1} = (A^i(Z^{i-1}))^* DZ^i \tag{4.52}$$

We use Equation 4.52 as the rule to perform one step adjoint calculation.

Numerical adjoint code construction obey 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 programming language to express a loop L:

$$\begin{array}{l} \text{for i} = 0 : \text{N} \\ L_i \\ \text{end i} \end{array} \tag{4.53}$$

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, parallel loop. The output of a parallel loop does not depend on the order by which the iterate is executed. The adjoint of a parallel loop is a loop of the same bounds without particular execution order for the iterate. It can be expressed as L^*:

$$\begin{array}{l} \text{for i} = 0 : \text{N} \\ L_i^* \\ \text{end i} \end{array} \tag{4.54}$$

where L_i^* represents the adjoint statement for the ith iterate. The adjoint of a sequential loop, however, by eq(4.48), has to be executed in the reverse order, L^*:

$$\begin{array}{l} \text{for i = N : 1 : -1} \\ L_i^* \\ \text{end i} \end{array} \tag{4.55}$$

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 at the interest of taking 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 show in the following table:

forward code	adjoint code
if (condition C_1)	if (condition C_1)
L_1	L_1^*
else if (condition C_2)	else if (condition C_2)
L_2	L_2^*
else (condition C_3)	else (condition C_2)
L_3	L_3^*

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

$$x^i = Ix^{i-1} \tag{4.56}$$

where I is the identity map, and x^{i-1}, x^i are the variable 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

$$x^i = f(x^{i-1}, \cdot) \tag{4.57}$$

for some function f. Adjoint of assignment on a linearized code results directly from eq(4.52).

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 w

$$I(z, x, h) = \sum_w R(z, x - h; w)S(z, x + h; w) \tag{4.58}$$

where z and x are vertical and horizontal coordinates, h is the horizontal offsets between two scalar field R and S. A linear perturbation of eq(4.58) yields

$$\delta I(z,x,h) = \sum_{w} \{\delta R(z,x-h;w)S(z,x+h;w) + R(z,x-h;w)\delta S(z,x+h;w)\} \tag{4.59}$$

A pseudocode for the summation is written as follows

```
for iw=min:dw:wmax

   for z=zmin:dz:zmax
      for x=xmin:dx:xmax
         for h=hmin:dh:hmax

            if (x+h and x-h do not lie outside of boundary)  then

            δI(z,x,h)+ = δR(z,x−h;w)S(z,x+h;w) + R(z,x−h;w)δS(z,x+h;w)
         end h

      end x

   end z

end w
```
(4.60)

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

$$\begin{pmatrix} \delta R^i(z,x-h;w) \\ \delta S^i(z,x+h;w) \\ \delta I^i(z,x,h) \end{pmatrix} = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ S(z,x+h;w) & R(z,x-h;w) & I \end{pmatrix} \cdot \begin{pmatrix} \delta R^{i-1}(z,x-h;w) \\ \delta S^{i-1}(z,x+h;w) \\ \delta I^{i-1}(z,x,h) \end{pmatrix} \tag{4.61}$$

The corresponding adjoint is

$$\begin{pmatrix} DR^{i-1}(z,x-h;w) \\ DS^{i-1}(z,x+h;w) \\ DI^{i-1}(z,x,h) \end{pmatrix} = \begin{pmatrix} I & 0 & \overline{S(z,x+h;w)} \\ 0 & I & \overline{R(z,x-h;w)} \\ 0 & 0 & I \end{pmatrix} \cdot \begin{pmatrix} DR^{i}(z,x-h;w) \\ DS^{i}(z,x+h;w) \\ DI^{i}(z,x,h) \end{pmatrix} \tag{4.62}$$

We also noticed that the loop over w,z,x and h are parallel, so the adjoint code is written as follows

$$\begin{aligned}
&for \quad iw = min : dw : wmax \\
&\quad for \quad z = zmin : dz : zmax \\
&\quad\quad for \quad x = xmin : dx : xmax \\
&\quad\quad\quad for \quad h = hmin : dh : hmax \\
&\quad\quad\quad\quad if \quad \text{(x+h and x-h do not lie outside of boundary)} \quad then \\
&\quad\quad\quad\quad DR^{i-1}(z,x-h;w) += \overline{S(z,x+h;w)} DI^{i}(z,x,h) \\
&\quad\quad\quad\quad DS^{i-1}(z,x+h;w) += \overline{R(z,x-h;w)} DI^{i}(z,x,h) \\
&\quad\quad\quad\quad end \quad if \\
&\quad\quad\quad end \quad h \\
&\quad\quad end \quad x \\
&\quad end \quad z \\
&end \quad w
\end{aligned} \tag{4.63}$$

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

$$
\begin{aligned}
&for\ i = 1 : n \\
&\quad for\ j = 1 : n_s \\
&\qquad S_j^i = H[c^{i-1}] S_j^{i-1} \\
&\qquad R_j^i = \overline{H[c^{i-1}]} R_j^{i-1} \\
&\qquad for\ h = (-H, -H) ... (H, H) \\
&\qquad\quad I_h^i = I_h^i + \overline{S_j^i(x+h)} R_j^i(x-h) \\
&\qquad end(h) \\
&\quad end(j) \\
&end(i)
\end{aligned}
\tag{4.64}
$$

The perturbation δI_h should satisfy the perturbation of imaging routine.

$$
\begin{aligned}
&for\ i = 1 : n \\
&\quad for\ j = 1 : n_s \\
&\qquad \delta S_j^i = H[c^{i-1}]\delta S_j^{i-1} + (\tfrac{\partial}{\partial c} H[c^{i-1}] S_j^{i-1})\delta c^{i-1} \\
&\qquad \delta R_j^i = \overline{H[c^{i-1}]}\delta R_j^{i-1} + (\tfrac{\partial}{\partial c}\overline{H[c^{i-1}]} R_j^{i-1})\delta c^{i-1} \\
&\qquad for\ h = (-H, -H)...(H, H) \\
&\qquad\quad S_j^i(x+h) = T(h) S_j^i \\
&\qquad\quad R_j^i(x-h) = T(-h) R_j^i \\
&\qquad\quad I_h^i = I_h^i + \overline{\delta S_j^i(x+h)} R_j^i(x-h) + \overline{S_j^i(x+h)}\delta R_j^i(x-h) \\
&\qquad end(h) \\
&\quad end(j) \\
&end(i)
\end{aligned}
\tag{4.65}
$$

Here, we distinguish the uses of $\overline{(\cdot)}$ and $(\cdot)^*$. For the former, it means an operator with coefficients being complex conjugate to the coefficients of the operator $(\cdot)$, whereas for the later, it means an adjoint operator with respect to the operator $(\cdot)$. In order to derive the adjoint of perturbation imaging routine, we first simplify equations (18) as a loop consists of 2 parts.

$$
\begin{array}{l}
for\ i = 1 : n \\
\quad for\ j = 1 : n_s \\
\quad\quad A \\
\quad\quad B \\
\quad end(j) \\
end(i)
\end{array}
\tag{4.66}
$$

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

$$
\begin{array}{l}
for\ i = n, n-1, ..., 1 \\
\quad for\ j = 1 : n_s \\
\quad\quad B^* \\
\quad\quad A^* \\
\quad end(j) \\
end(i)
\end{array}
\tag{4.67}
$$

Expressions B are expanded as,

$$
\begin{array}{l}
for\ h = (-H, -H), ..., (H, H) \\
\quad C \\
end(h)
\end{array}
\tag{4.68}
$$

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

$$
\begin{array}{l}
for\ h = (-H, -H), ..., (H, H) \\
\quad C^* \\
end(h)
\end{array}
\tag{4.69}
$$

We express A and C using matrix-vector product form. For C, we have

$$\begin{pmatrix} \delta I_h^i \\ \overline{\delta S_j^i(x+h)} \\ \delta R_j^i(x-h) \\ \overline{\delta S_j^i} \\ \delta R_j^i \end{pmatrix} = \begin{pmatrix} I & R_j^i(x-h) & \overline{S_j^i(x+h)} & 0 & 0 \\ 0 & 0 & 0 & T(h) & 0 \\ 0 & 0 & 0 & 0 & T(-h) \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & I \end{pmatrix} \begin{pmatrix} \delta I_h^i \\ \overline{\delta S_j^i(x+h)} \\ \delta R_j^i(x-h) \\ \overline{\delta S_j^i} \\ \delta R_j^i \end{pmatrix} \tag{4.70}$$

We then formulate the transposed expressions C^T,

$$\begin{pmatrix} DI_h^i \\ \overline{DS_j^i(x+h)} \\ DR_j^i(x-h) \\ \overline{DS_j^i} \\ DR_j^i \end{pmatrix} = \begin{pmatrix} I & 0 & 0 & 0 & 0 \\ \overline{R_j^i(x-h)} & 0 & 0 & 0 & 0 \\ S_j^i(x+h) & 0 & 0 & 0 & 0 \\ 0 & T(-h) & 0 & I & 0 \\ 0 & 0 & T(h) & 0 & I \end{pmatrix} \begin{pmatrix} DI_h^i \\ \overline{DS_j^i(x+h)} \\ DR_j^i(x-h) \\ \overline{DS_j^i} \\ DR_j^i \end{pmatrix} \tag{4.71}$$

or explicitly,

$$\begin{aligned} DR_j^i &= T(h)DR_j^i(x-h) + DR_j^i \\ D\overline{S_j^i} &= T(-h)D\overline{S_j^i(x+h)} + D\overline{S_j^i} \\ DR_j^i(x-h) &= S_j^i(x+h)DI_h^i \\ D\overline{S_j^i(x+h)} &= \overline{R_j^i(x-h)}DI_h^i \end{aligned} \tag{4.72}$$

Here we have used D to replace δ to indicate the adjoint variables. Substitute 27.3

into 27.1 and 27.4 into 27.2, we have,

$$B^* = \left\{ \begin{array}{lcl} DR_j^i & = & \sum_h T(h)\{S_j^i(x+h)DI_h^i\} + DR_j^i \\ D\overline{S_j^i} & = & \sum_h T(-h)\{\overline{R_j^i(x-h)}DI_h^i\} + D\overline{S_j^i} \end{array} \right. \tag{4.73}$$

For expressions A, the matrix representation is

$$\begin{pmatrix} \overline{\delta S_j^i} \\ \delta R_j^i \\ \overline{\delta S_j^{i-1}} \\ \delta R_j^{i-1} \\ \delta c^{i-1} \end{pmatrix} = \begin{pmatrix} 0 & 0 & \overline{H[c^{i-1}]} & 0 & ([\frac{\partial}{\partial c}\overline{H[c^{i-1}]}(\overline{S_j^{i-1}})] \\ 0 & 0 & 0 & \overline{H[c^{i-1}]} & (\frac{\partial}{\partial c}\overline{H[c^{i-1}]}R_j^{i-1}) \\ 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & I \end{pmatrix} \begin{pmatrix} \overline{\delta S_j^i} \\ \delta R_j^i \\ \overline{\delta S_j^{i-1}} \\ \delta R_j^{i-1} \\ \delta c^{i-1} \end{pmatrix} \tag{4.74}$$

The transposed expressions A^* can be written

$$\begin{pmatrix} D\overline{S_j^i} \\ DR_j^i \\ D\overline{S_j^{i-1}} \\ DR_j^{i-1} \\ Dc^{i-1} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \overline{H[c^{i-1}]}^* & 0 & I & 0 & 0 \\ 0 & \overline{H[c^{i-1}]}^* & 0 & I & 0 \\ ([\frac{\partial}{\partial c}\overline{H[c^{i-1}]}(\overline{S_j^{i-1}})]^* & (\frac{\partial}{\partial c}\overline{H[c^{i-1}]}R_j^{i-1})^* & 0 & 0 & I \end{pmatrix} \begin{pmatrix} D\overline{S_j^i} \\ DR_j^i \\ D\overline{S_j^{i-1}} \\ DR_j^{i-1} \\ Dc^{i-1} \end{pmatrix} \tag{4.75}$$

We then write explicitly,

$$A^* = \left\{ \begin{array}{lcl} Dc^{i-1} & = & ([\frac{\partial}{\partial c}\overline{H[c^{i-1}]}(\overline{S_j^{i-1}})]^* D\overline{S_j^i} + (\frac{\partial}{\partial c}\overline{H[c^{i-1}]}R_j^{i-1})^* DR_j^i + Dc^{i-1} \\ DR_j^{i-1} & = & \overline{H[c^{i-1}]}^* DR_j^i + DR_j^{i-1} \\ D\overline{S_j^i} & = & \overline{H[c^{i-1}]}^* D\overline{S_j^i} + D\overline{S_j^i} \end{array} \right. \tag{4.76}$$

Insert expressions for B^* and A^* in equations (17), we have

$$
\begin{aligned}
&for\ i = n, n-1, ..., 1 \\
&\quad for\ j = 1 : n_s \\
&\qquad DR_j^i = \textstyle\sum_h T(h)\{S_j^i(x+h)DI_h^i\} + DR_j^i \\
&\qquad DS_j^i = \textstyle\sum_h T(-h)\{\overline{R_j^i(x-h)}DI_h^i\} + DS_j^i \\
&\qquad Dc^{i-1} = ([\tfrac{\partial}{\partial c}\overline{H[c^{i-1}]}(\overline{S_j^{i-1}})]^* D\overline{S_j^i} + (\tfrac{\partial}{\partial c}\overline{H[c^{i-1}]}R_j^{i-1})^* DR_j^i + Dc^{i-1} \\
&\qquad DR_j^{i-1} = \overline{H[c^{i-1}]}^* DR_j^i + DR_j^{i-1} \\
&\qquad D\overline{S_j^{i-1}} = H[c^{i-1}]^* D\overline{S_j^i} + D\overline{S_j^{i-1}} \\
&\quad end(j) \\
&end(i)
\end{aligned}
\tag{4.77}
$$

We then unroll the sequential loop by isolating the first iteration $i = n$ and combine the equations (32.1) and (32.4) as well as equations (32.2) and (32.5) jointly into the rest of the loop. We arrive at

$i = n$
$for\ j = 1 : n_s$

$DR_j^i = \sum_h T(h)\{S_j^i(x+h)DI_h^i\}$

$DS_j^i = \sum_h T(-h)\{\overline{R_j^i(x-h)}DI_h^i\}$

$Dc^{i-1} = ([\frac{\partial}{\partial c}\overline{H[c^{i-1}]}(\overline{S_j^{i-1}})]^* D\overline{S_j^i} + (\frac{\partial}{\partial c}\overline{H[c^{i-1}]}R_j^{i-1})^* DR_j^i$

$end(j)$

$for\ i = n-1, ..., 1$ (4.78)

$for\ j = 1 : n_s$

$$DR_j^i = \sum_h T(h)\{S_j^i(x+h)DI_h^i\} + H^i[c^i]DR_j^{i+1}$$

$$DS_j^i = \sum_h T(-h)\{\overline{R_j^i(x-h)}DI_h^i\} + H^i[c^i]^* DS_j^{i+1}$$

$$Dc^{i-1} \mathrel{+}= ([\frac{\partial}{\partial c}\overline{H[c^{i-1}]}(\overline{S_j^{i-1}})]^* D\overline{S_j^i} + (\frac{\partial}{\partial c}\overline{H[c^{i-1}]}R_j^{i-1})^* DR_j^i$$

$end(j)$

$end(i)$

Comparing this routine and equations eq(4.34) eq(**??**) and eq(4.36), we are expecting to see that they are the continuous limit of this discretised formulation.

4.2 Inversion

We pose the problem of eismic velocity analysis is as a nonlinear optimization problem

$$\min_{c \in C^2} \frac{1}{2} ||P_h I_h||^2 \tag{4.79}$$

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 directly pursue the quality of semblance images , where in waveform type of inversion the optimized velocity does not necessarily produce the optimum image.

Nonlinear optimization by gradient type of methods requires that the objective function itself be smooth and has few local minima in order to avoid the solution trapped by local minima. The objective function constructed by differential semblance criterion varies smoothly due to change of velocities. Its local minima usually coincide with the global minimum. Therefore it is particular suitable for optimizations solving 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 the key role. In last section, we focus on the calculation of the gradient. In this section, we shall explain two things. One is the application of L-BFGS algorithm that uses the calculated gradient for optimization, and the other, the smoothness problems we face on using the gradients and how the gradient is smoothed: the B-spline smoothing scheme

4.2.1 Limited memory BFGS method

We use limited memory BFGS method to directly optimize the objective function eq(4.79). At each iteration, the BFGS search direction p minimizes a quadratic model of the objective function at the current iterate.

$$m_k(p) = J_k + \nabla_c J^t + \frac{1}{2} p^t H_k p \tag{4.80}$$

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

$$p_k = -H_k^{-1} \nabla_c J \tag{4.81}$$

and the new iterate is chosen by line search

$$c_{k+1} = c_k + \alpha_k p_k \tag{4.82}$$

to satisfy the sufficient decrease and curvature conditions. The particular formula to update H_k^{-1} *Nocedal& Wright, 2000*[15] 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. 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 the n value of between 3 and 20 produce satisfactory results. The limited memory BFGS algorithm can be stated formally as follows *Nocedal& Wright, 2000* [15].

```
Choose starting point c_0 and m > 0
k ← 0
Choose H_0^{-1}
repeat
    Compute p_k ← −H_k^{-1} ∇_c J
    Line search c_{k+1} = c_k + α_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 = ∇_c J_{k+1} − ∇_c J_k
    k ← k + 1
    Compute H_k^{-1}
until converge
```
(4.83)

4.2.2 Smoothing properties of the gradient

For a given velocity model paired with the observed data, it 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 gradient given in section (4.1.1). We rewrite the gradient in here

$$\begin{aligned}\nabla_c J(y) \;=\; & \int G^+(x_r, y, w)\frac{u(x,h)\omega^2}{c(y)^2}G^+(y, x-h, w)\times \\ & d(x_r, x_s, t)G^+(x+h, x_s, w)e^{i\omega t}dx_r dx_s dt\; dw\; dx\; dh \\ & + \int G^+(x_r, x-h, w)d(x_r, x_s, t)G^+(x+h, y, w)\times \end{aligned} \tag{4.84}$$

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

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

$$d(x_r,x_s,t) = \int G^+(x+h,x_s,\omega)r(x,h)G^+(x-h,x_r,\omega)e^{i\omega t}dxdhd\omega \tag{4.85}$$

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

$$d(x_r,x_s,t;x,h) = \int G^+(x+h,x_s,\omega)r(x,h)G^+(x-h,x_r,\omega)e^{i\omega t}d\omega \tag{4.86}$$

This data 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 gradient corresponds to the data of single reflection is reduced to

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

where

$$g_1 = \int G^+(x_r,y,w)\frac{u(x,h)\omega^2}{c(y)^2}G^+(y,x-h,w)d(x_r,x_s,t;x,h)G^+(x+h,x_s,w)e^{i\omega t}dtdw \tag{4.87}$$

$$g_2 = \int G^+(x_r, x-h, w) d(x_r, x_s, t) G^+(x+h, y, w) \frac{u(x,h) w^2}{c(y)^2} G^+(y, x_s, \omega) e^{i\omega t} dt d\omega \tag{4.88}$$

We shall see that g_1 is smooth 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. Parallel analysis apply to the second term due to symmetry.

Using asymptotic form of the Green's function

$$G^+(x, y, \omega) = a(x, y) e^{-i\omega\phi(x,y)} \tag{4.89}$$

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

$$d(x_r, x_s, t) = \int a(x_r, x-h, x+h, x_s) r(x,h) e^{i\omega(t-\phi(x+h,x_s)-\phi(x_r,x-h))} d\omega \tag{4.90}$$

With equations (4.89) and (4.90), g_1 can be written as

$$\begin{aligned} g_1 &= \int \frac{\omega^2}{c(y)^2} r(x,h) a(x_r, x-h, x-h, x_s) a(x_r, y) a(y, x-h) a(x+h, x_s) u(x,h) \times \\ &\quad e^{i\omega t} e^{i\omega'(t-\phi(x+h,x_s)-\phi(x_r,x-h))} e^{-i\omega\phi(x_r,y)} e^{-i\omega\phi(y,x-h)} e^{-i\omega\phi(x+h,x_s)} dt d\omega d\omega' \\ &= \int \frac{\omega^2}{c(y)^2} r(x,h) a(x_r, x-h, x-h, x_s) a(x_r, y) a(y, x-h) a(x+h, x_s) u(x,h) \times \\ &\quad 2\pi e^{i\omega(\phi(x_r,x-h)+\phi(x+h,x_s)-\phi(x_r,y)-\phi(y,x-h)-\phi(x+h,x_s))} d\omega \end{aligned}$$

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

We can write the integral in the form

$$g_1(y; x_r, x_s, x, h) = \int A(x_r, x_s, x, h, \omega) e^{i\Phi(x_r, y, x-h, \omega)]} d\omega \tag{4.91}$$

where

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

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 Φ is stationary with respect to ω,

$$\frac{\partial \Phi(x_r, y, x-h, \omega)}{\partial \omega}|_{y\in\gamma(x_r,x-h)} = [\phi(x_r, x-h) - \phi(x_r, y) - \phi(y, x-h)]_{y\in\gamma(x_r,x-h)} = 0 \tag{4.92}$$

Theories of oscilatory integral (Theorem 3.5 in *Joshi*[13]) 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

*$S^m_{\rho,\delta}(X \times \mathbb{R}^N)$ is the space of symbols of order m and of type (ρ, δ). It consists of all $a(x,\theta) \in C^\infty(X \times \mathbb{R}^N)$ such that for all compact $K \subset\subset X$ and all multi-index $\alpha \in N^n$, there is a constant $C = C_{K,\alpha,\beta}(a)$ such that $|\partial^\alpha_x \partial^\beta_\theta a(x,\theta) \leq C(1+|\theta|)^{m-\rho|\beta|+\delta|\alpha|}$ 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*, [9]. We see $A \in S^2_{1,0}(X \times \mathbb{R}^N)$ for $N = 1$.

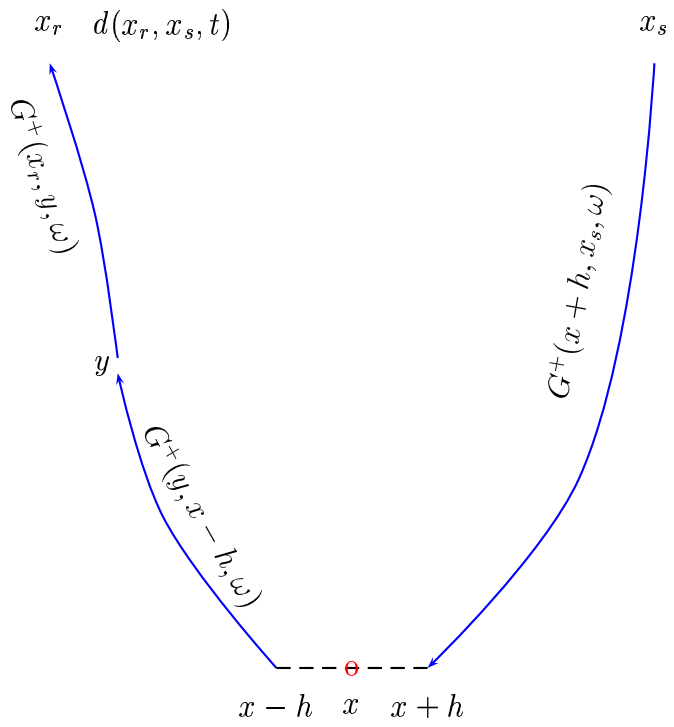


Figure 4.1 The ray path connecting from x_s to $x+h$, $x-h$ to y and y to x_r, where $G^+(x, y, \omega)$ is the Green's function observed at x emanated at y. The adjoint of differential migration picks up an image residual and smoothly distributed it along the corresponding ray paths.

path $\gamma(x_r, x-h)$. Taking derivative of the gradient along the bicharacteristic curve $\gamma(x_r, x-h)$ with respect to y means to differentiate equation (**??**) subject to equation (4.92). The differential operator ∂_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 along the bicharacteristic curves and non-smooth (singular) across the bicharacteris-

tic 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 singluarities 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 of 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 gradient 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. Certain smoothing scheme is necessary. 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 number of 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

$$c = Bm \tag{4.93}$$

where c is understood as the velocity in its image space representation with image space basis function defined as cartesian grid basis function, 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 derived

$$\frac{\partial c}{\partial m} = B^T \tag{4.94}$$

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

$$\frac{\partial J}{\partial m} = \frac{\partial c}{\partial m}\frac{\partial J}{\partial c} = B^T\frac{\partial J}{\partial c} \tag{4.95}$$

that they satisfy

$$\nabla_m J = B^T \nabla_c J \tag{4.96}$$

Where B^T is the adjoint of the B-spline projection. It is guaranteed from the properties of B-spline basis functions that

$$B\nabla_m J = BB^T \nabla_c J \tag{4.97}$$

is smooth, implying that the velocity update is smooth. Equation (4.96) suggests that instead of optimizing image space velocity subject to smoothing constraint

$$\min_{c \in C^2} J(c) \tag{4.98}$$

we can optimize its B-spline coefficients

$$\min_{m} J(Bm) \tag{4.99}$$

using third B-spline basis functions of degree 3. A complete optimization routine on B-spline parameters is shown as follows

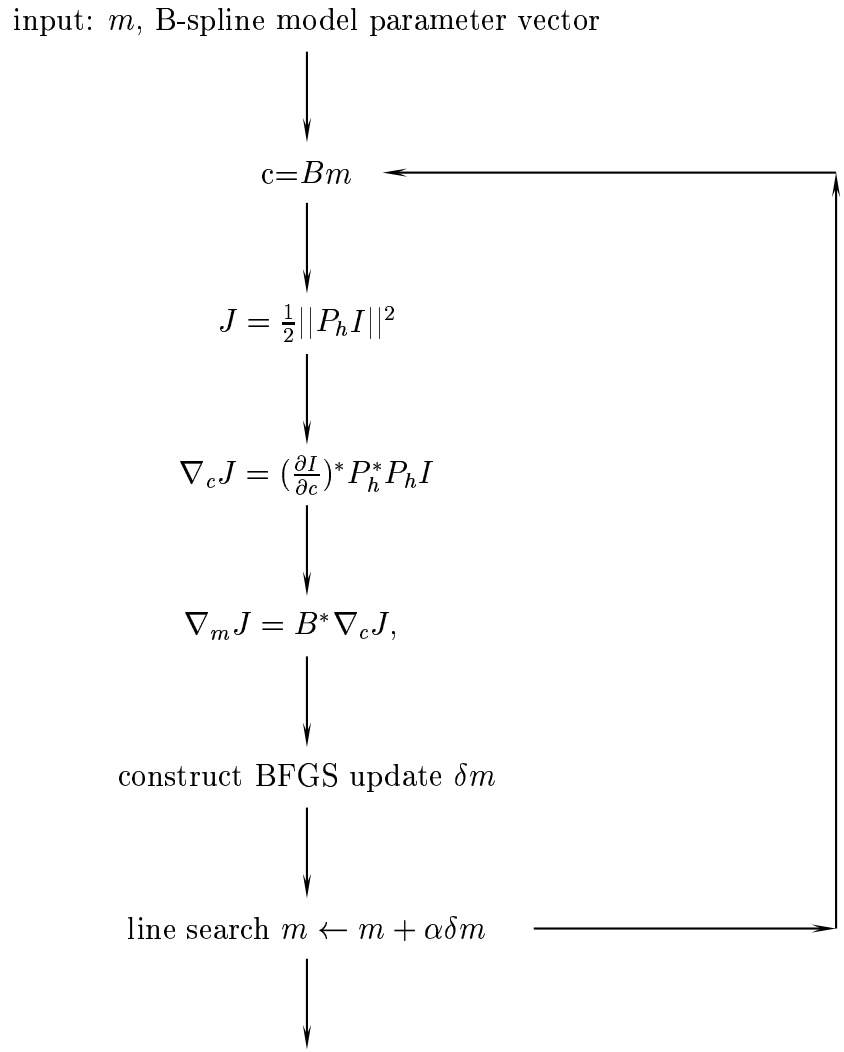


Figure 4.2 Inversion procedure. Forward project B-spline model parameters to obtain current velocity in image space through $c = Bm$, then evaluate objective function, calculate gradient with respect velocity in image space, project this gradient to model space through the adjoint of B-spline projection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

Chapter 5
Data examples

Chapter synopsis

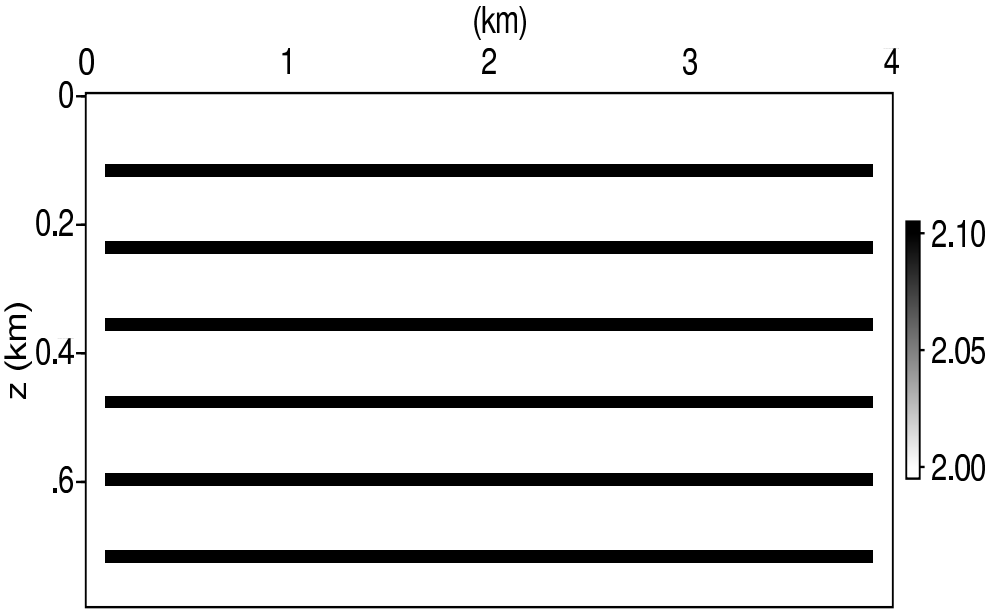
In this chapter four synthetic and real data examples are studied. Properties of image gathers in offset and the application problems associated with the differential semblance velocity analysis are analyzed. Questions of what should the gather look like at correct velocity and what is necessary in data 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 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. Reasonable results are obtained. Example IV is the application of differential semblance velocity analysis to Marmousi dataset. Shallow velocity structures can be shown well reconstructed by differential semblance optimization. Problems of inversions f complex velocity structure are analyzed in the discussion section (5.5).

5.1 Offset image gathers due to flat reflectors and constant velocity

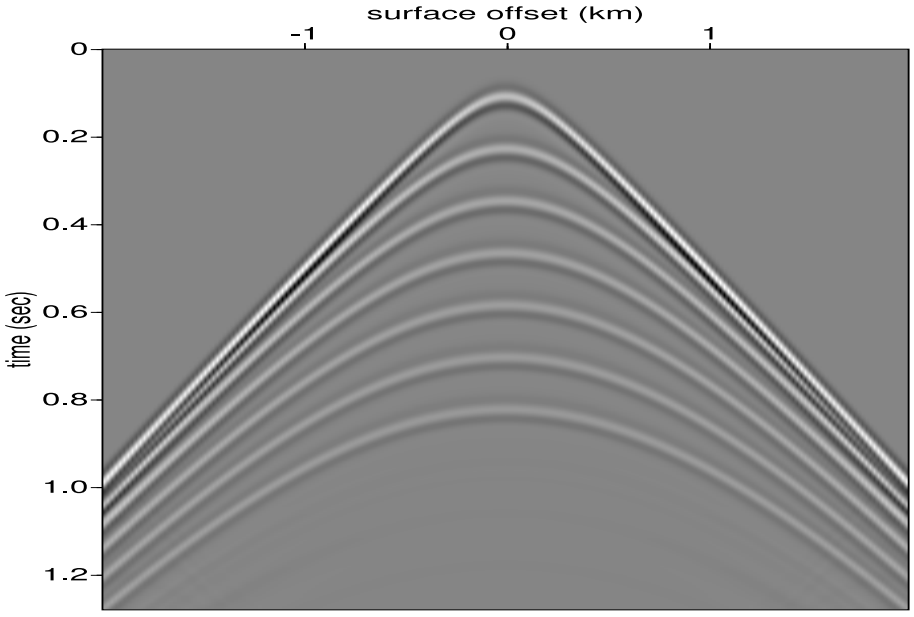
5.1.1 Synthetic data generation

We generate 2-D synthetic data using constant background velocity $c(z, x) = 2$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 [8] at about 10 grid points. Sources and receivers in a fixed receiver array are evenly distributed on the surface from $x = 0.1$ km to $x = 3.9$ km incremented using the same interval $\Delta x = 0.02$ km. Absorbing boundary condition described in appendix (A) is applied to remove free surface multiples and reflections from the computational boundaries. Two source gathers collected in the middle of the surface with direct arrival removed are shown are in Figure (5.1(b)).

*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.



(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 (a). Configuration of synthetic data generation using constant velocity and flat reflectors. (b). Reflection data for $x_s = 2.0$ km

5.1.2 Preprocessing

One-way wave equation usually can not handle waves that propagate in high angles with respect to the vertical direction. The complex 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 avoids evanescent energies in wavefield extrapolation. Ideally speaking, evanescent filtering should be applied to each depth extrapolated wavefields. However, a evanescent filter applied on the surface data once has shown to be adequate to remove evanescent energies for practical purposes.

We apply the time domain Fast Fourier transform to seismograms trace by trace.

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

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

$$D(x_r, x_s, \omega) = \mathscr{F}_{x_r}^{-1} \phi \mathscr{F}_{x_r}(d) \tag{5.2}$$

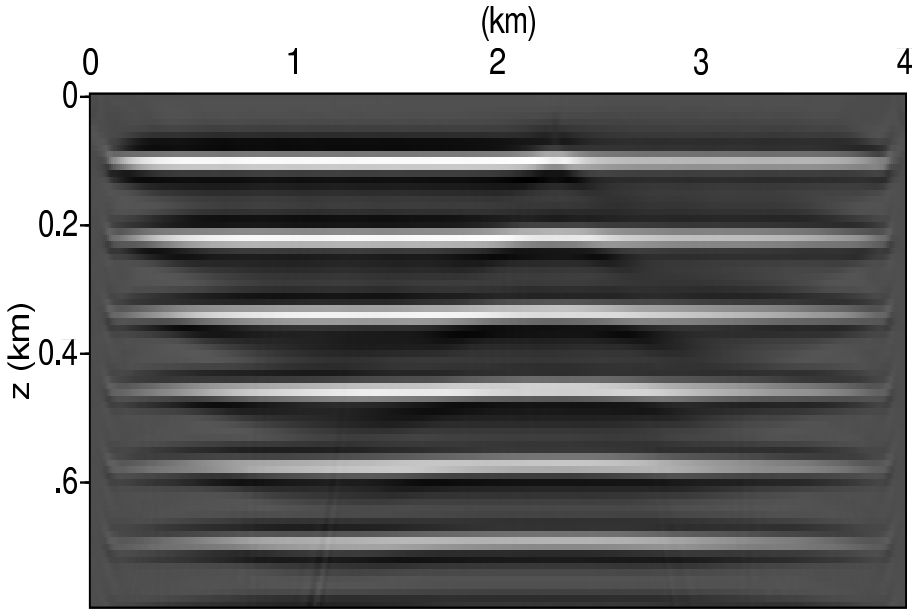
where

$$\mathscr{F}_{x_r}(d(x_r, x_s, \omega)) = \int d(x_r, x_s, \omega) e^{-ik \cdot x_r} dx_r$$

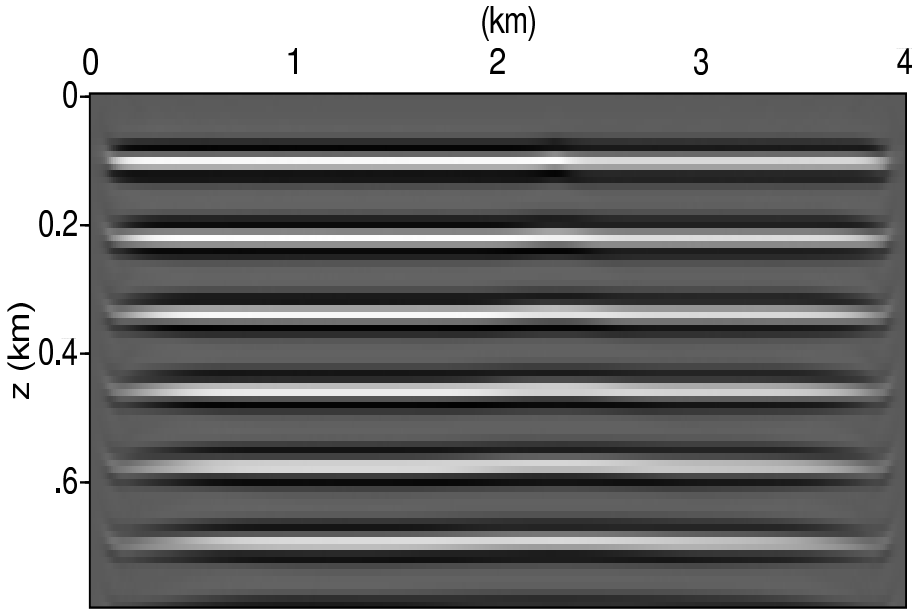
and

$$\phi(k) = \begin{cases} 0 & |k| > \frac{\omega}{c}\sin\theta \\ 1 & |k| < \rho\frac{\omega}{c}\sin\theta \\ \text{cosine tapered} & \text{elsewhere} \end{cases}$$

θ 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 to the surface plane, where c is used as the surface velocity and θ is chose to be 75°. Energies associated with propagating waves at angle higher than 75° are annihilated for all frequency 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 ρ 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.

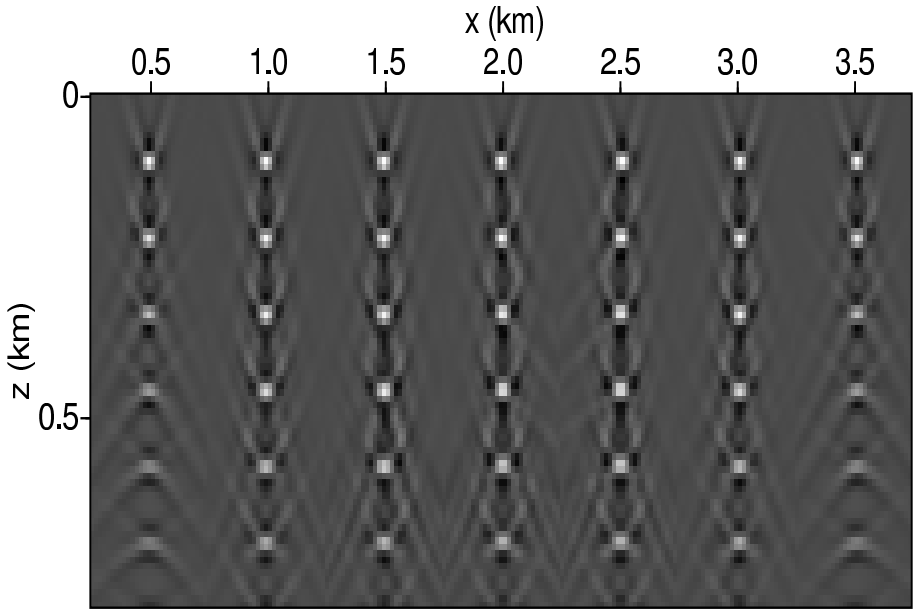


(a) $\rho = 0.9$

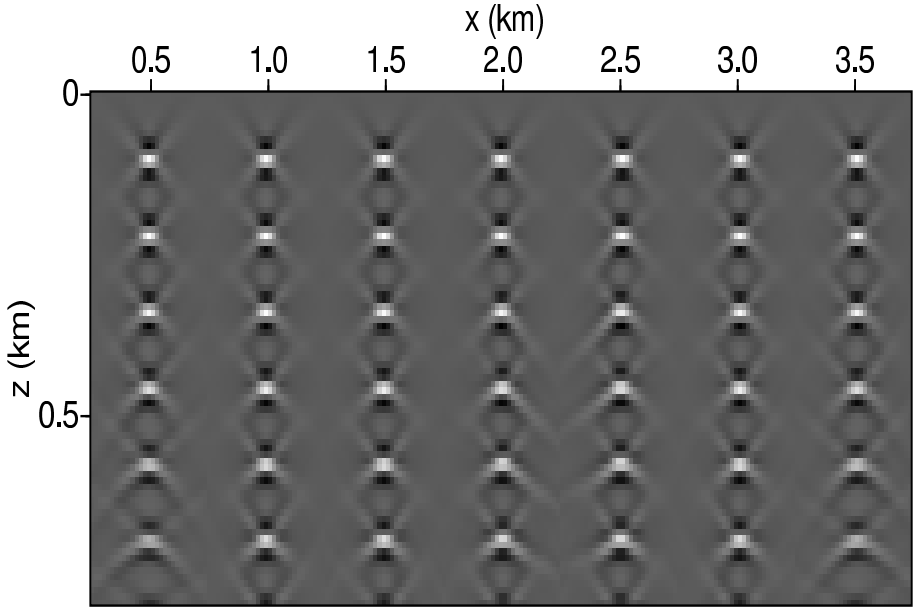


(b) $\rho = 0.5$

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 $\rho = 0.5$.



(a) $\rho = 0.9$

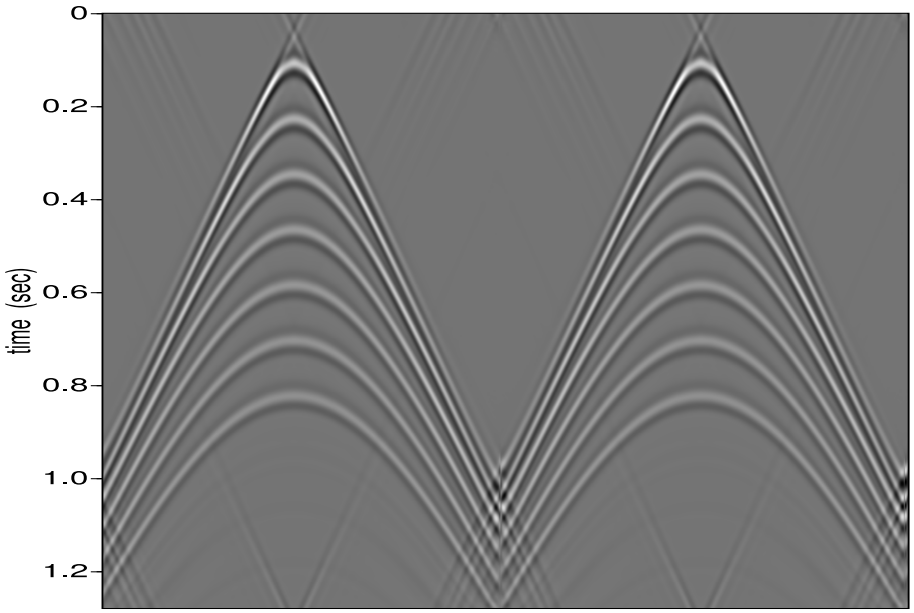


(b) $\rho = 0.5$

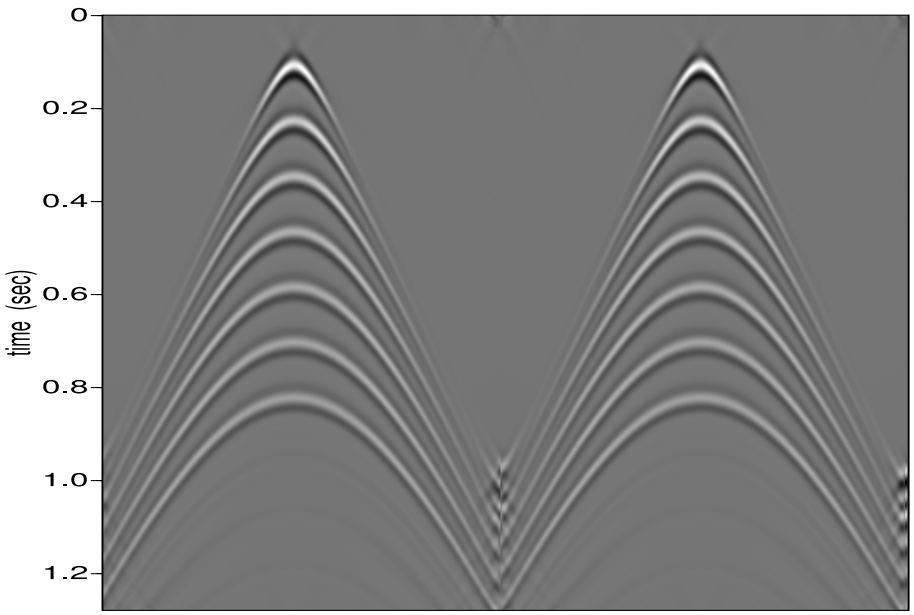
Figure 5.3 Effects of data preprocessing on offset gathers. (a). Offset gathers at the correct velocity correspond to preprocessed data using $\rho = 0.9$. (b). Offset gathers at the correct velocity correspond to preprocessed data using $\rho = 0.5$.

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 ρ value in the preprocessing. Figure (5.3) shows offset gathers migrated at the correct velocity using a preprocessed data corresponds to $\rho = 0.5$. It shows that the gathers become much clean 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 horizontally propagated waves are also muted by evanescent filter in Figure (5.4(b)).



(a) $\rho = 0.9$



(b) $\rho = 0.5$

Figure 5.4 Preprocessed data using $\rho = 0.9$ and $\rho = 0.5$ projected back into the (x, t) domain. Two source gathers are drawn for sources located in the middle of the model

5.1.3 Effects of acquisition geometry

Offset image gathers migrated through 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 correct velocity shown in Figure(5.5) demonstrates that the gather has a "X" pattern at around the zero offset.

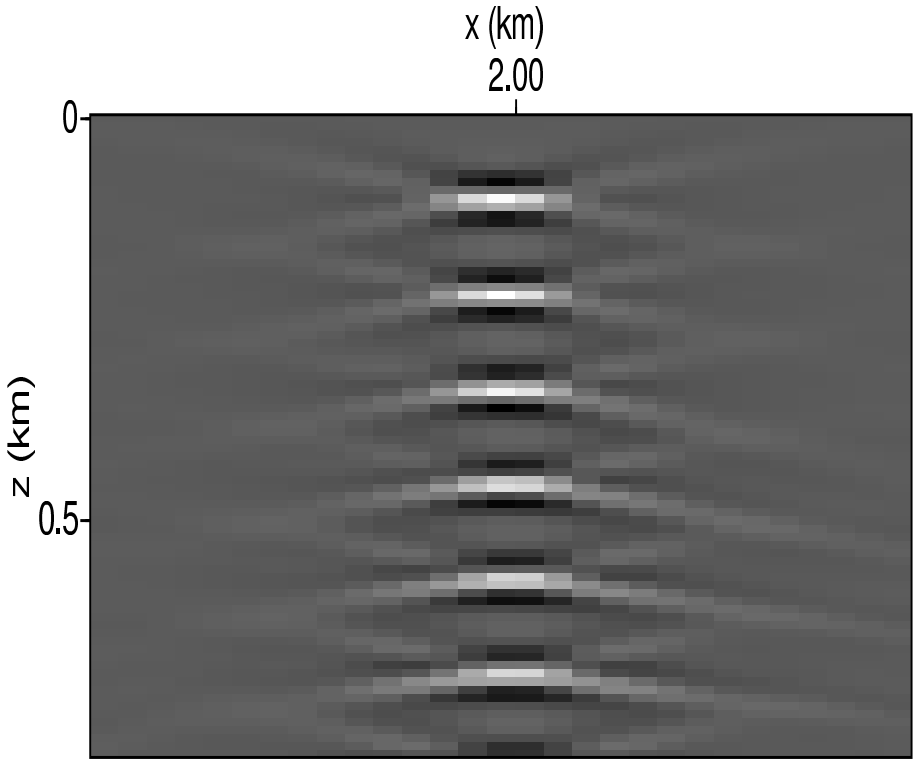


Figure 5.5 An "x" pattern in offset gathers obtained at correct velocity

*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.

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 understood as an integral over x_s as

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

where $u(z,x,x_s,h)$ is the product of downward continued source wavefield complex conjugated and the downward continued receiver wavefield integrated over ω.*

$$u(z,x,x_s,h,\omega) = \int \overline{S}(z,x+h,x_s,\omega)R(z,x-h,x_s,\omega)d\omega \tag{5.4}$$

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

$$\begin{aligned} u_q &= \int u(z+qh,x,x_s,h)dh \\ &= \int_{L(k_s,k_r,q)=0} \overline{\hat{S}(k_s,x_s)e^{ik_s\cdot x}}\hat{R}(k_r,x_s)e^{ik_r\cdot x}dk_r dk_s \end{aligned} \tag{5.5}$$

with

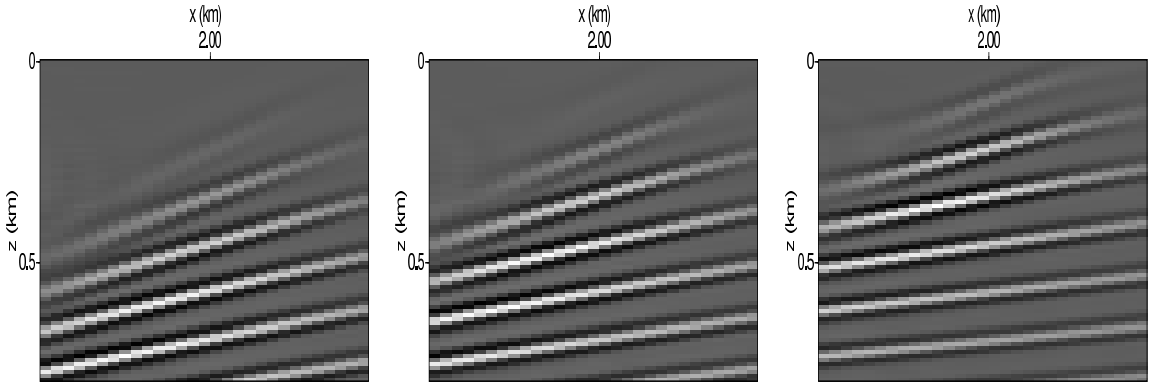
$$L(k_s,k_r,q) = \frac{k_{rx}+k_{sx}}{k_{rz}-k_{sz}} - q = 0$$

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 $\frac{k_{rx}+k_{sx}}{k_{rz}-k_{sz}} = q$. 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

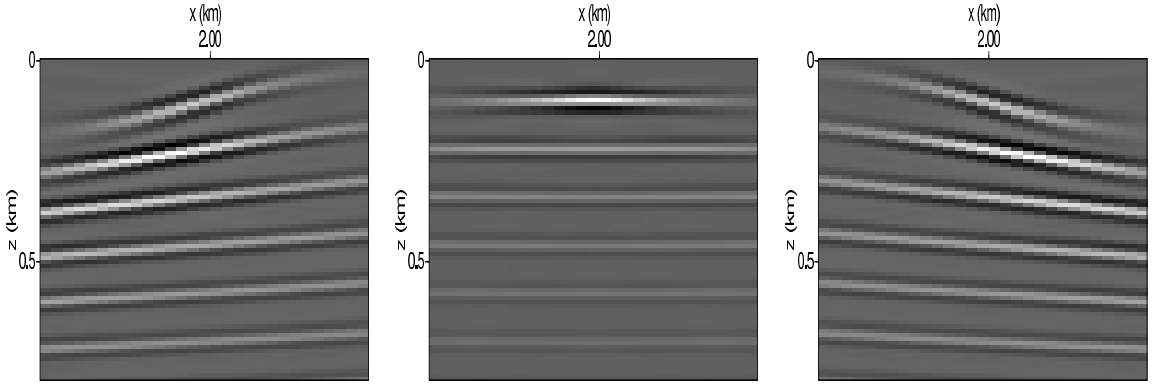
*The notation used in this chapter denotes z as the vertical coordinate and x as the horizontal coordinate.

rapidly decaying only when $q = -\tan(\theta)$ for θ is scattering angle as illustrated in Figure (3.2). We demonstrate this phenomena by carrying 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 $u(z, x, x_s)$.

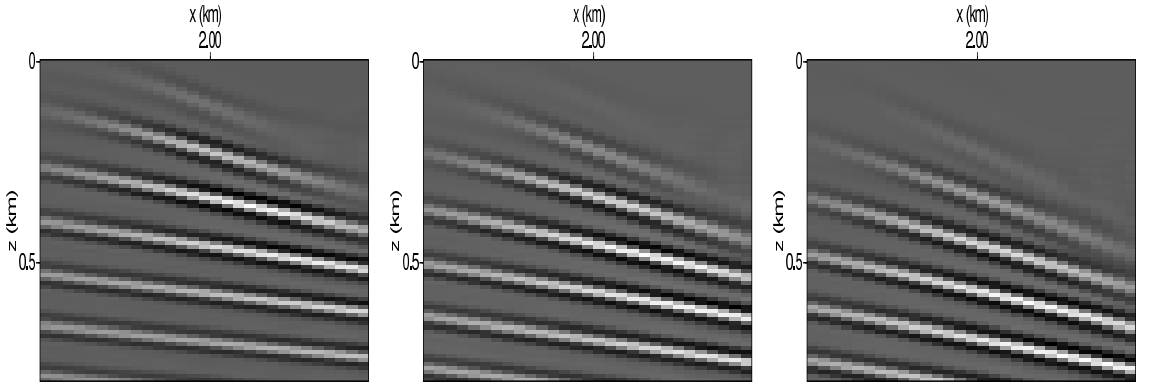
Offset gathers at the left most and right most panels present linear traces with highest dipping angles, which can be easily understood from the relation $q = -\tan(\theta)$. The offset gather $I_h(z, x, h)$ is expressed as an oscillatory integral of $u(z, x, x_s, h)$ over x_s. In view of Figure (5.6) the stacking of all offset gathers of single source 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. The "cut-off" in offset gathers of single source at high dipping angles is due to the limited acquisition geometry.



(a) $x_s = 0.1.6$ km (b) $x_s x = 1.7$ km (c) $x_s = 1.8$ km



(d) $x_s = 1.9$ km (e) $x_s = 2.0$ km (f) $x_s = 2.1$ km



(g) $x_s = 2.2$ km (h) $x_s = 2.3$ km (i) $x_s = 2.4$ km

Figure 5.6 Offset gathers due to migration of single source gathers at correct velocity. The reflectors in offset gather become a straight line. 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 due to multipathing *Stolk & Symes, 2002*[23]. Wave equation migration is free of multipathing induced kinematic artifacts and therefore provide ideal platform 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 offset domain differential semblance criterion is shown to be valid even for 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 evenascent filtering scheme correspond to $\rho = 0.5$.

5.2.2 Synthetic data generation and migration

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

$$c(x_1, x_2) = 2 - \alpha e^{\frac{(x_1-0.3)^2+(x_2-2)^2}{0.3}} \tag{5.6}$$

for $\alpha = 0.4, 0.6, 0.8, 1.0$ where α is an amplitude coefficient that controls the refracting strength. Data are collected using the same acquisition geometry as example I. Direct arrivals can be easily muted. As the refracting coefficient α increases from 0.4 to 1.0 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 (5.10).

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 $c = 2km/sec$ 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 images. 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. At locations where the image is distorted the offset image gathers show large amplitude at nonzero offset. Conversely, the off-

set image gathers are concentrated at zero offset when the flatness of the image is preserved as shown in Figure (5.13).

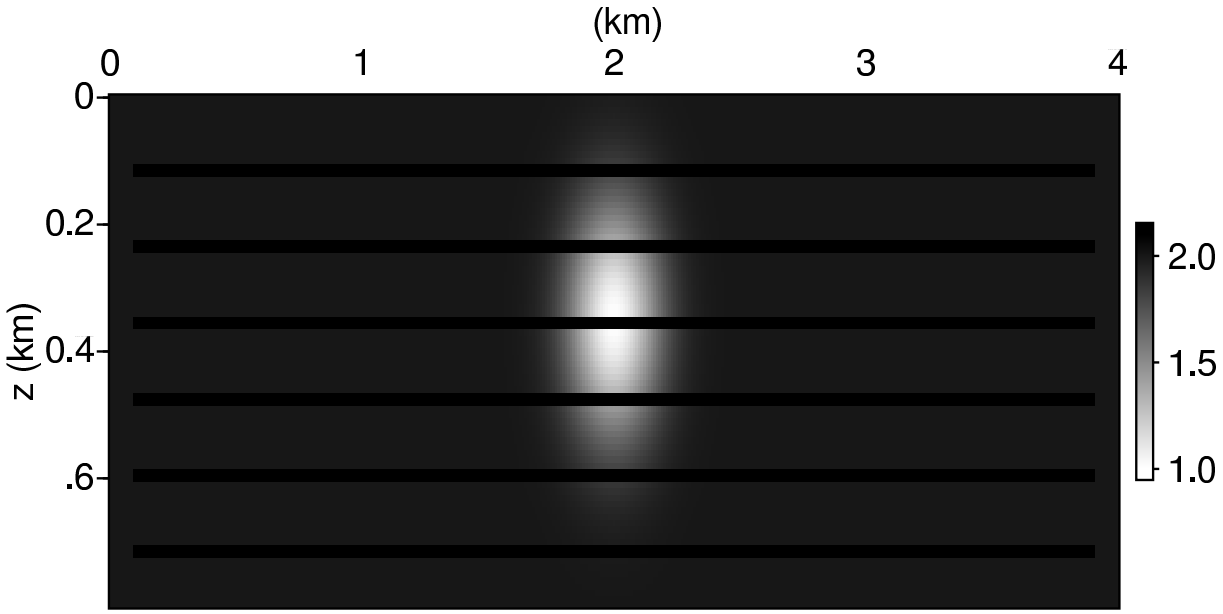
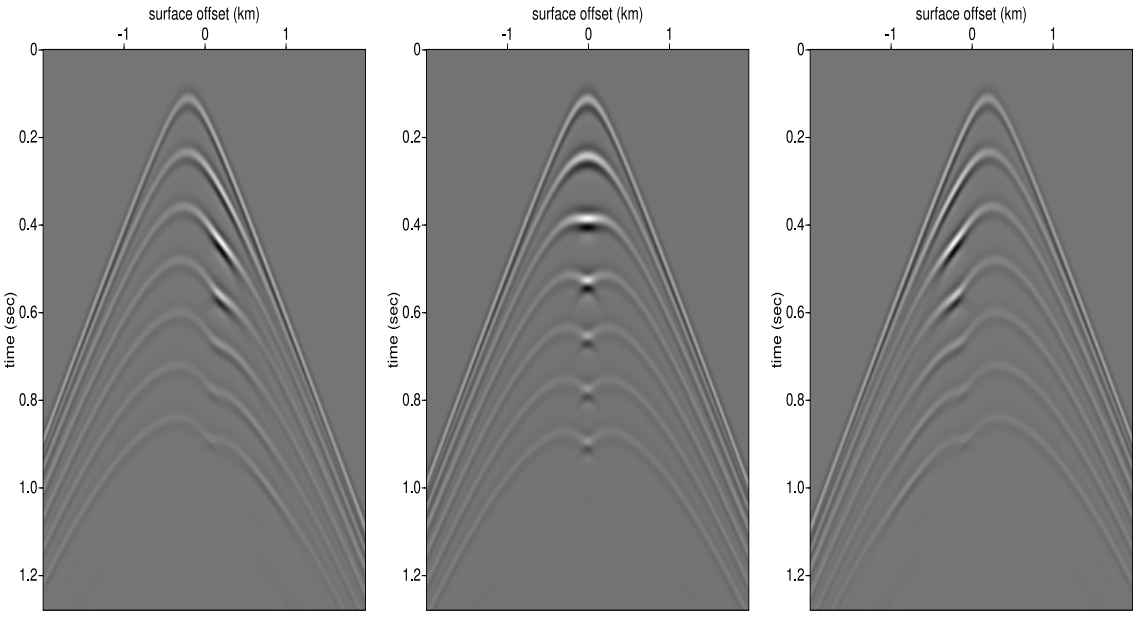


Figure 5.7 Velocity model of 0.8km by 4km with a low velocity lens imbedded in the middle. Sources and receivers are evenly distributed on the surface at the same spatial interval. Six horizontal reflectors are located at equal distance from $x_1 = 0.12$ to $x_1 = 0.78$.

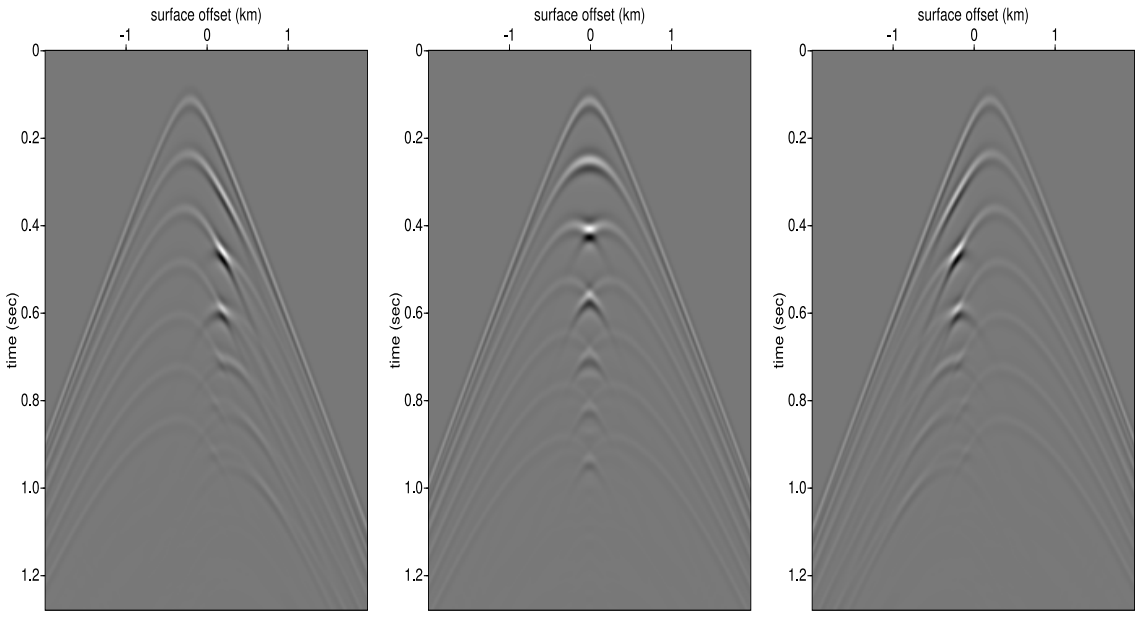


(a) source gather at x= 1.8 km

(b) source gather at x= 2.0 km

(c) source gather at x= 2.2 km

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

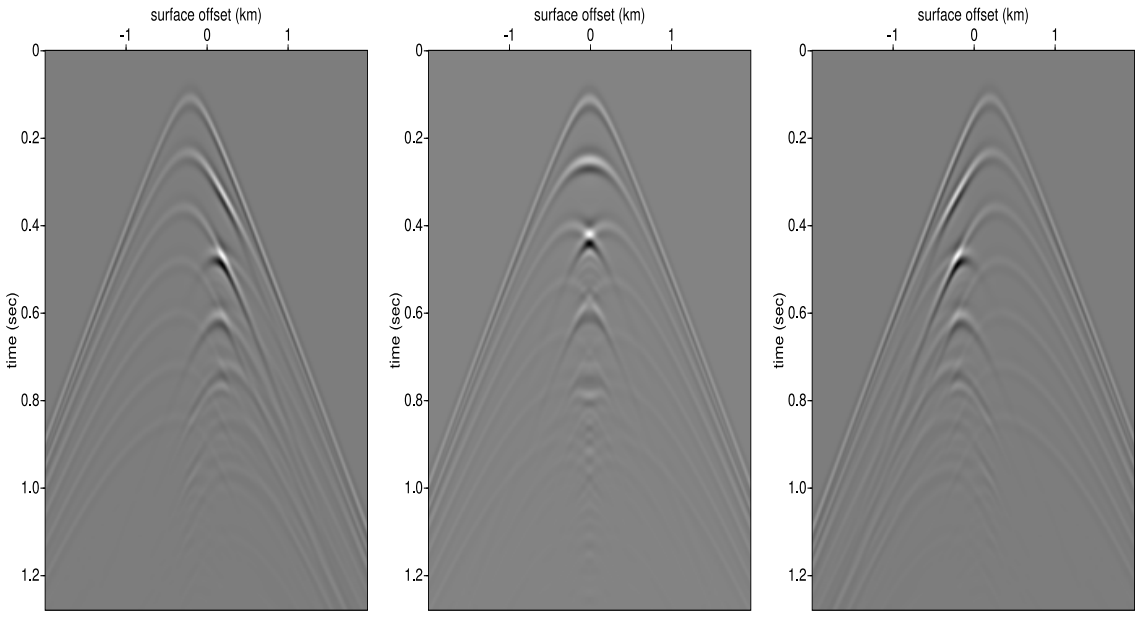


(a) source gather at x = 1.8 km

(b) source gather at x = 2.0 km

(c) source gather at x = 2.2 km

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

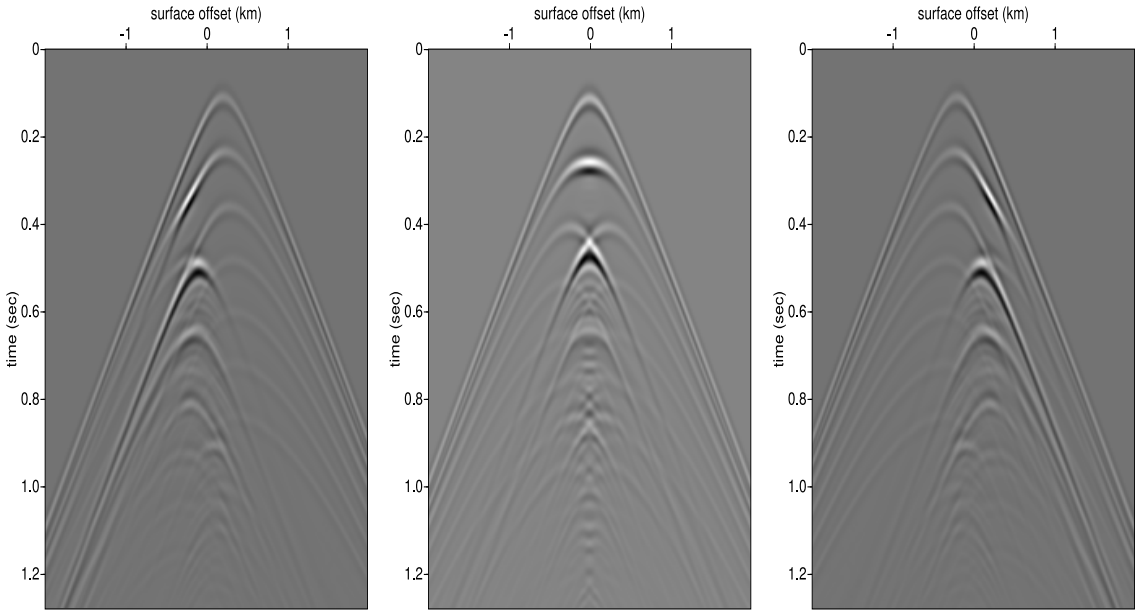


(a) source gather at x = 1.8 km

(b) source gather at x = 2.0 km

(c) source gather at x = 2.2 km

Figure 5.10 Source gathers obtained by acoustic simulation using velocity model eq(5.6) with refracting strength $\alpha = 0.8$. Multipathing arrivals are evident.

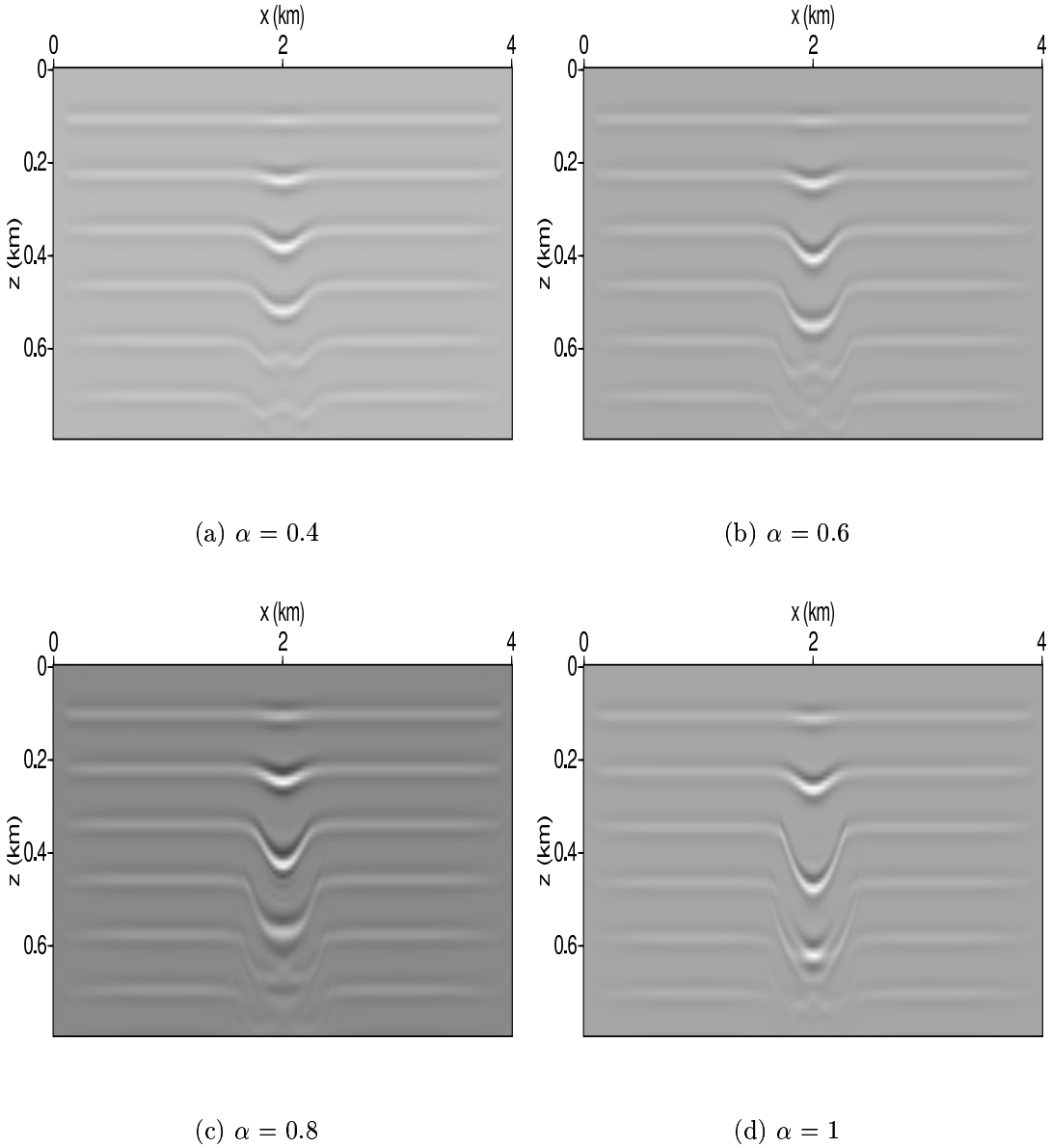


(a) source gather at x = 1.8 km

(b) source gather at x = 2.0 km

(c) source gather at x = 2.2 km

Figure 5.11 Synthetic source gathers obtained by acoustic simulation using velocity model eq(5.6) with refracting strength $\alpha = 1.0$ which generates strong multipathing arrivals. Maximum velocity perturbation is 50%.



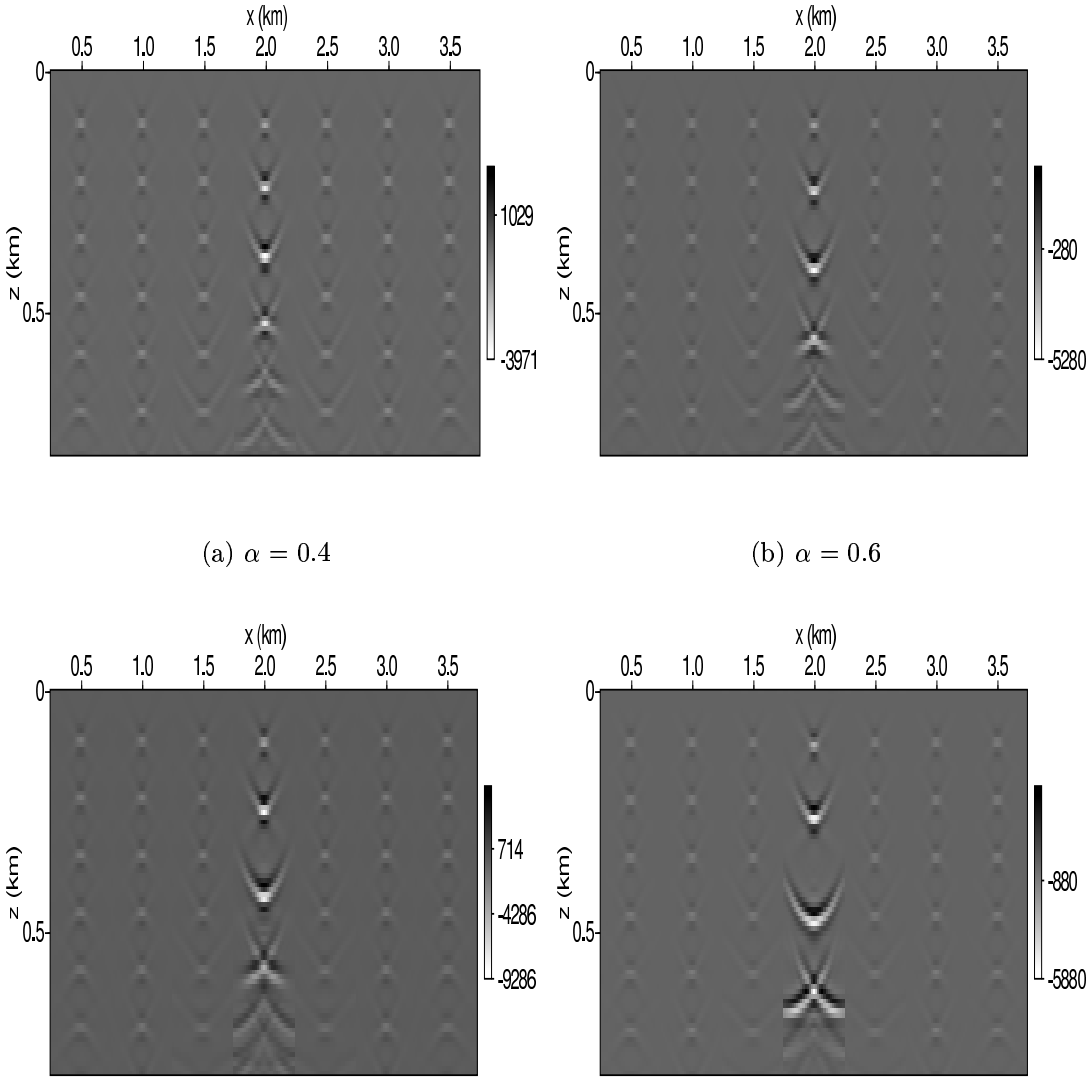
(a) $\alpha = 0.4$

(b) $\alpha = 0.6$

(c) $\alpha = 0.8$

(d) $\alpha = 1$

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



(a) $\alpha = 0.4$

(b) $\alpha = 0.6$

(c) $\alpha = 0.8$

(d) $\alpha = 1$

Figure 5.13 Initial image offsets obtained at constant velocity of 2 km/sec. Nonzero offset amplitudes are significant near the lens region.

5.2.2.1 Velocity inversion

Velocity analysis by differential semblance optimization tries to minimize the amplitude of nonzero offset images. Signals on nonzero offset are first amplified by multiplication of offset parameter h and then projected back to the velocity model by the adjoint differential migration. This delineates the procedure for computation of the gradient with respect 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 carried equivalently in 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 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.

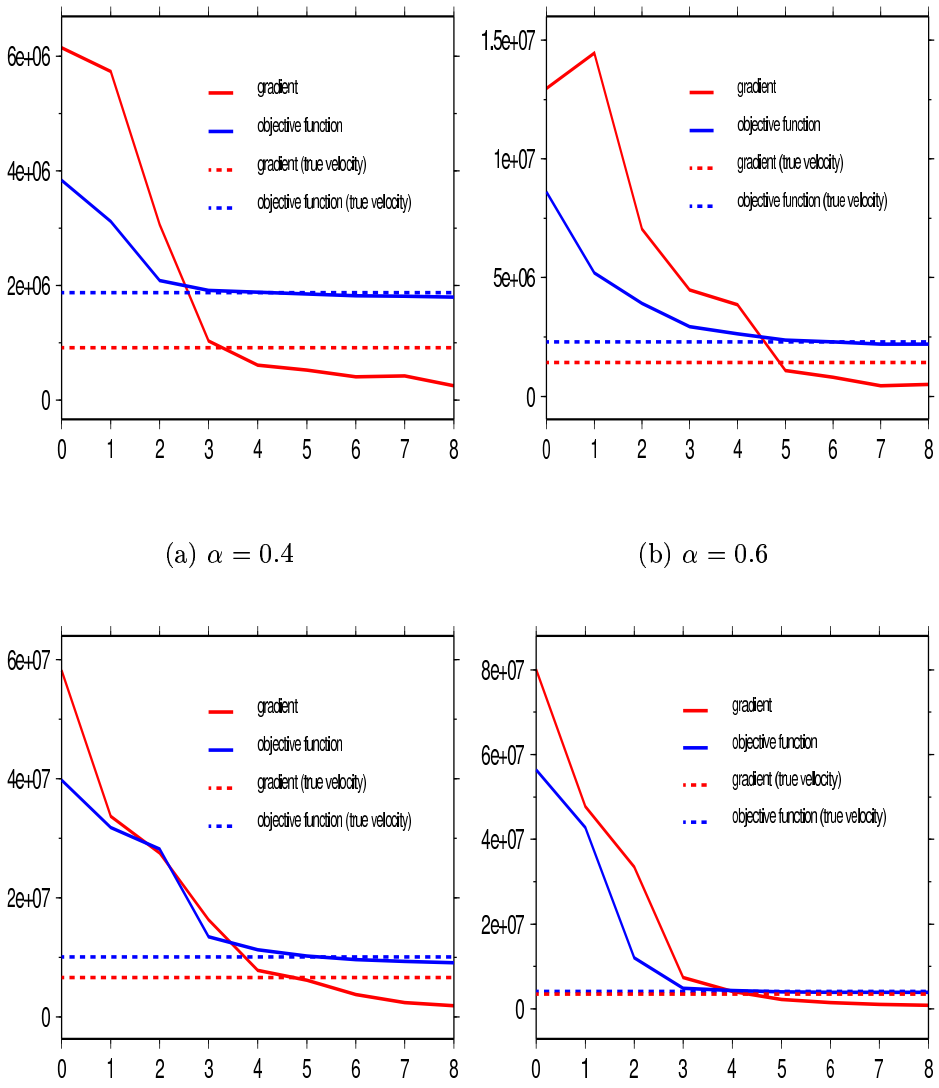
The differential semblance optimization does not need the linearization of the objec-

tive function which is essentially a Gauss-Newton's method used in many geophysical inversions [**?**],[8], 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 degrease as shown in Figure (5.14). The stopping criterion is provided depending on the desired accuracy with which the solution is to be found. The iteration terminates when

$$||g|| \leq \epsilon \max(1, ||m||) \tag{5.7}$$

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

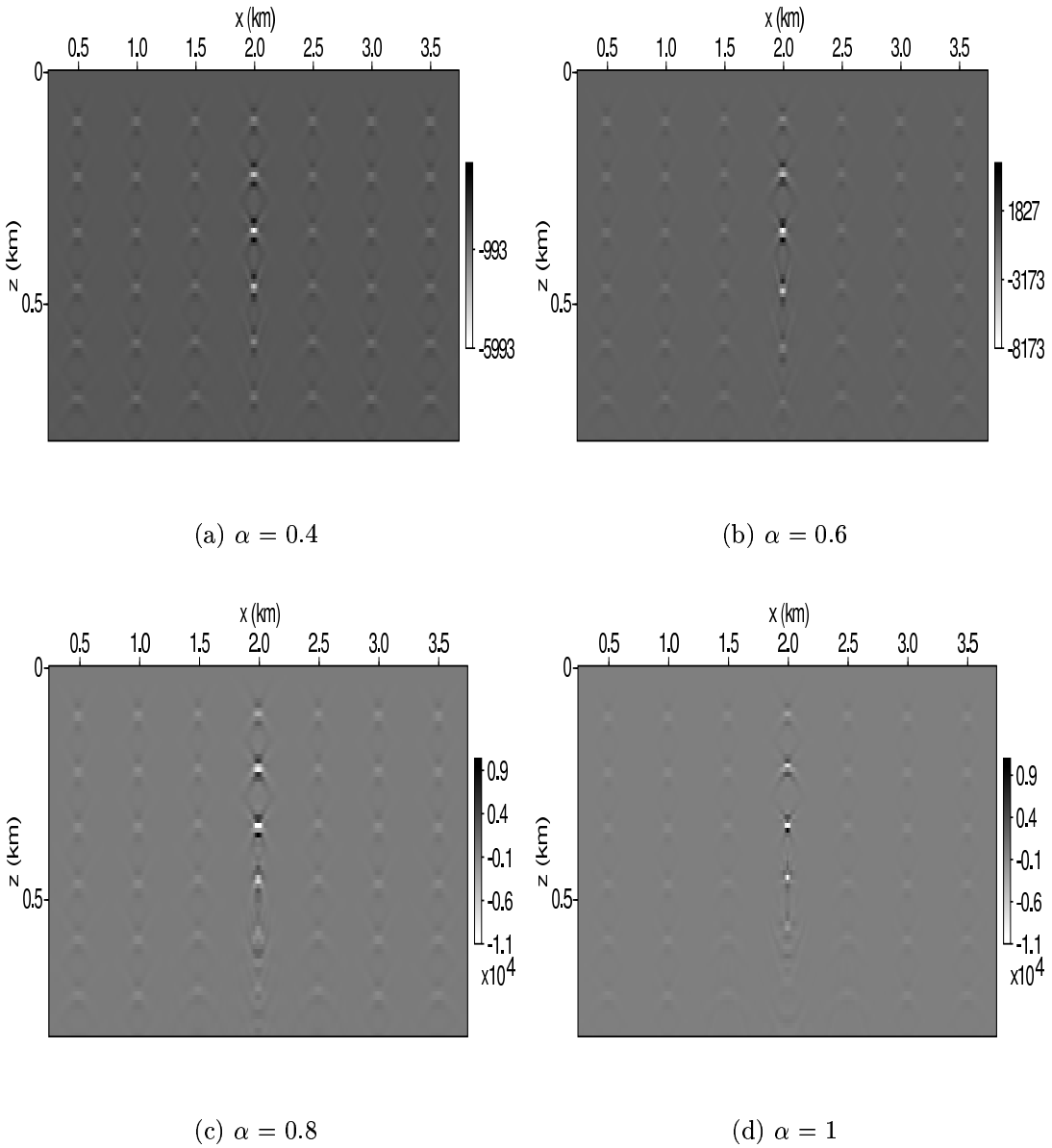
Output offset image gathers are shown to be well focused at zero offset in Figure (5.15) indicating the convergence has been reached. Comparing with the offset image gathers obtained at true velocity, we see the geometric difference is small as shown in Figure (5.18). Output images are also reconstructed to be flat (Figure (5.16) at the corresponding output velocities which are shown in good agreement to the true velocity. See Figure (5.19) and Figure (5.20).



(a) $\alpha = 0.4$ (b) $\alpha = 0.6$

(c) $\alpha = 0.8$ (d) $\alpha = 1$

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



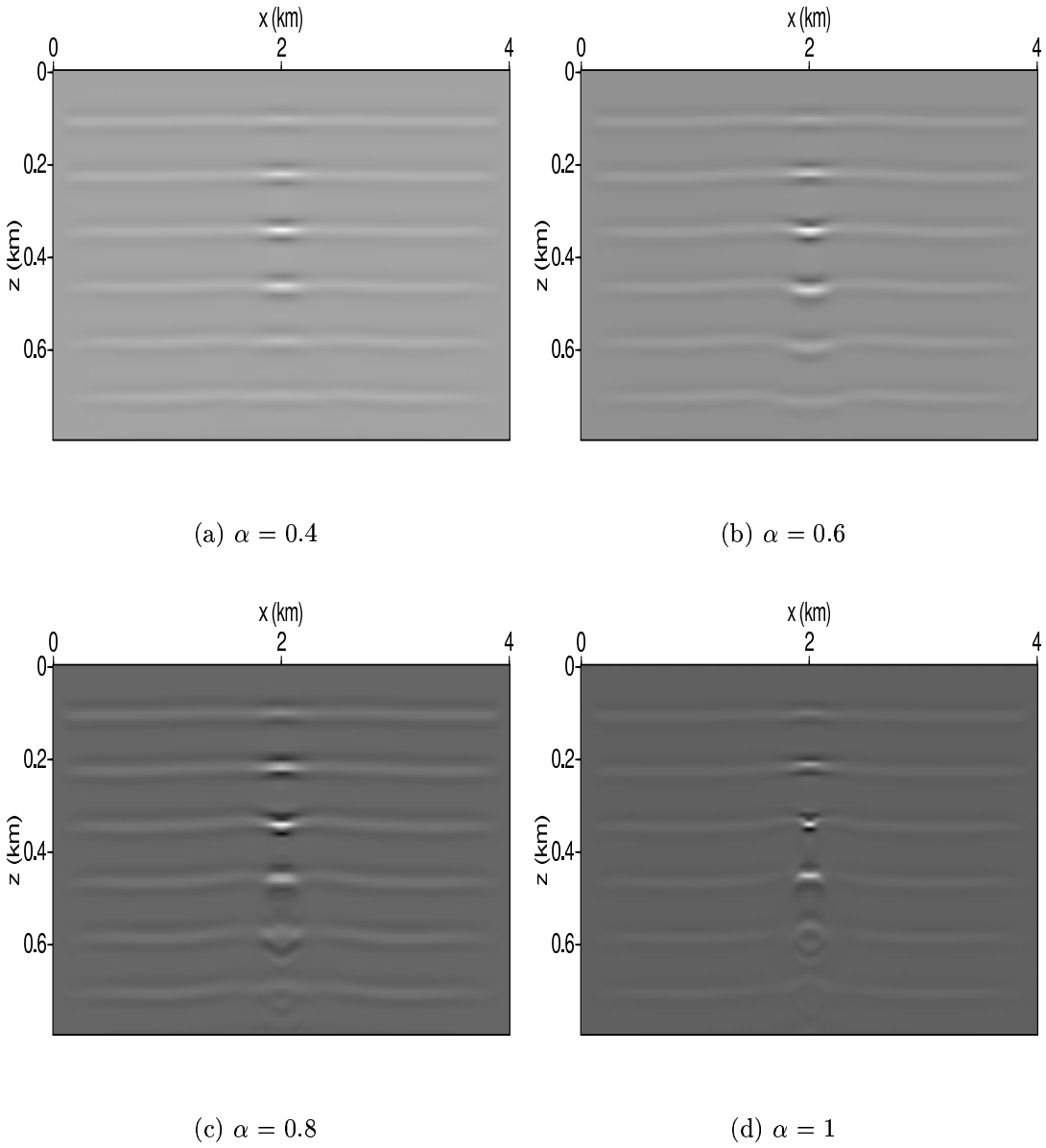
(a) $\alpha = 0.4$

(b) $\alpha = 0.6$

(c) $\alpha = 0.8$

(d) $\alpha = 1$

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



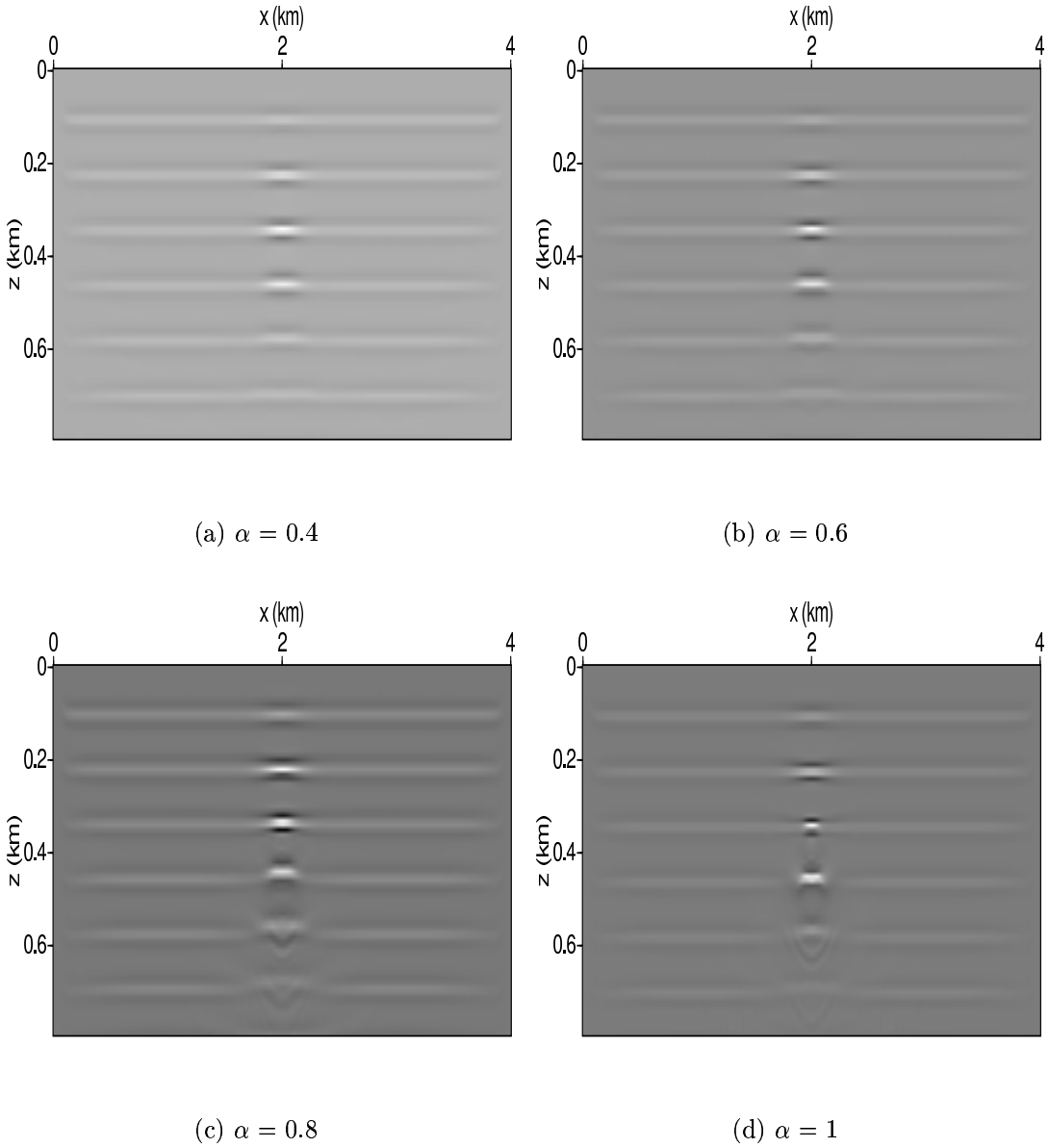
(a) $\alpha = 0.4$

(b) $\alpha = 0.6$

(c) $\alpha = 0.8$

(d) $\alpha = 1$

Figure 5.16 Images obtained at the fifth iteration. Reconstructed images are reasonably flattened.



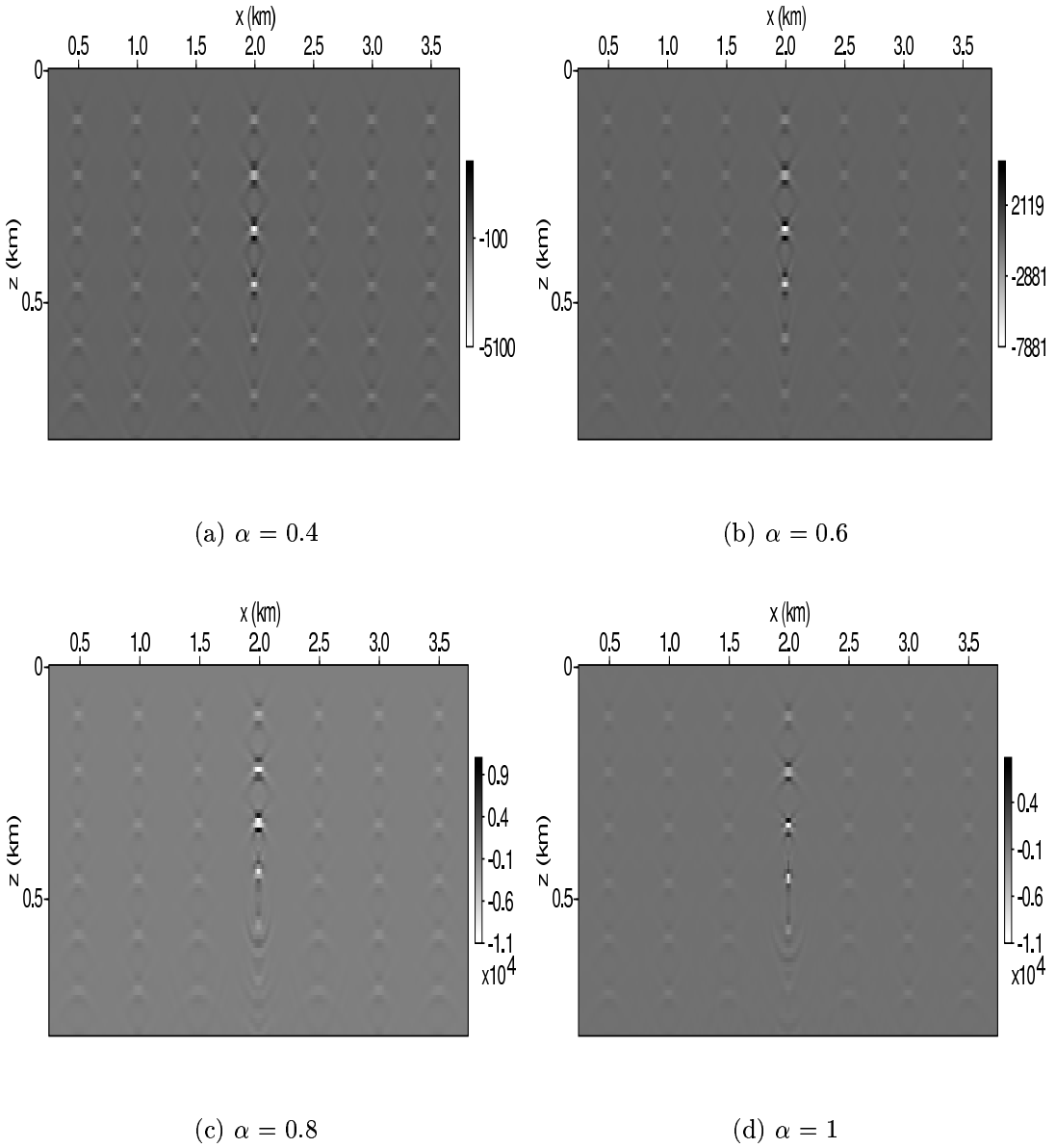
(a) $\alpha = 0.4$

(b) $\alpha = 0.6$

(c) $\alpha = 0.8$

(d) $\alpha = 1$

Figure 5.17 Images migrated using the correct velocities. The comparison with Figure (5.16) shows little differences.



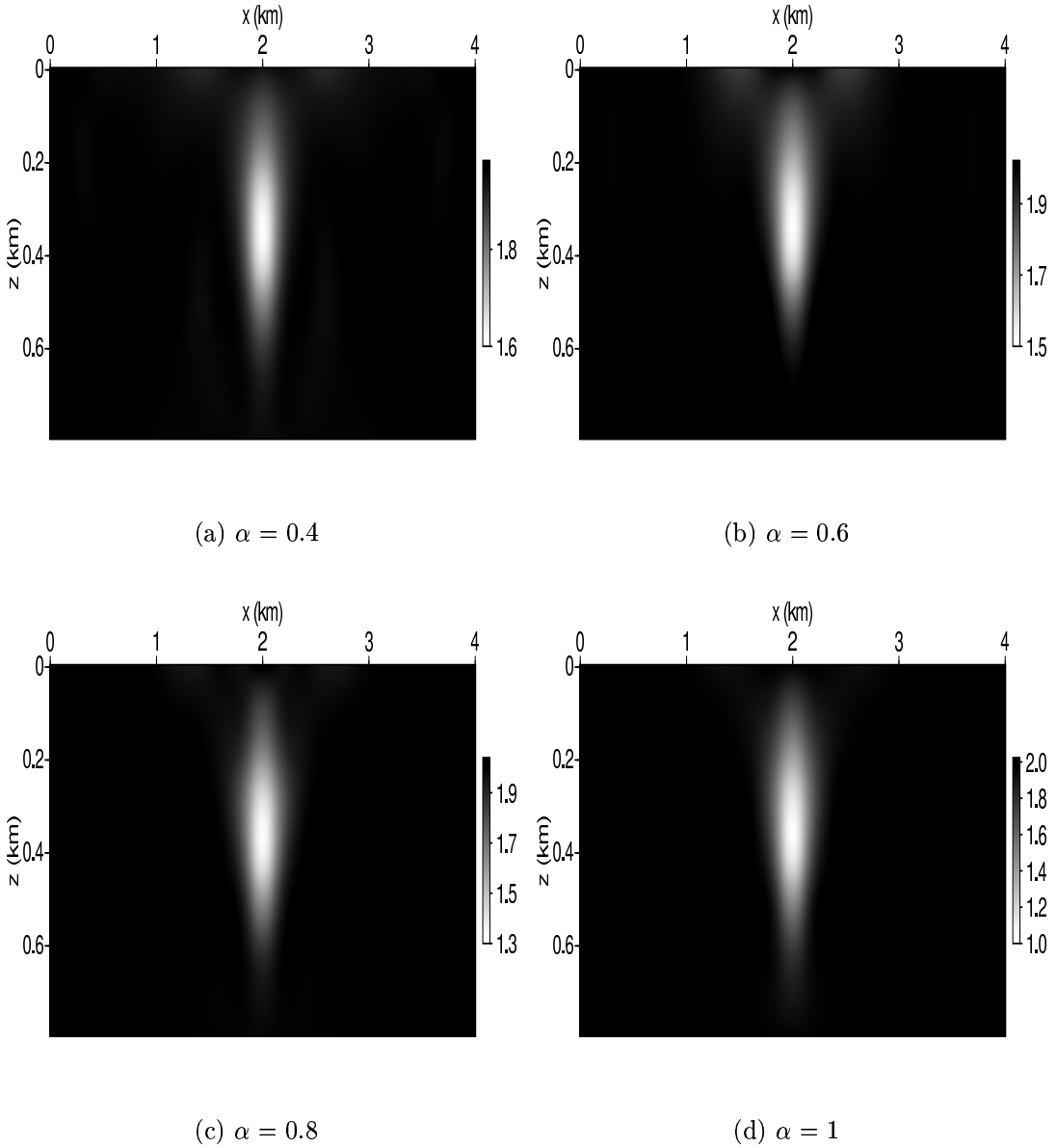
(a) $\alpha = 0.4$

(b) $\alpha = 0.6$

(c) $\alpha = 0.8$

(d) $\alpha = 1$

Figure 5.18 Offset image gathers migrated using the correct velocities. The comparison with Figure (5.15) show little geometrical differences.



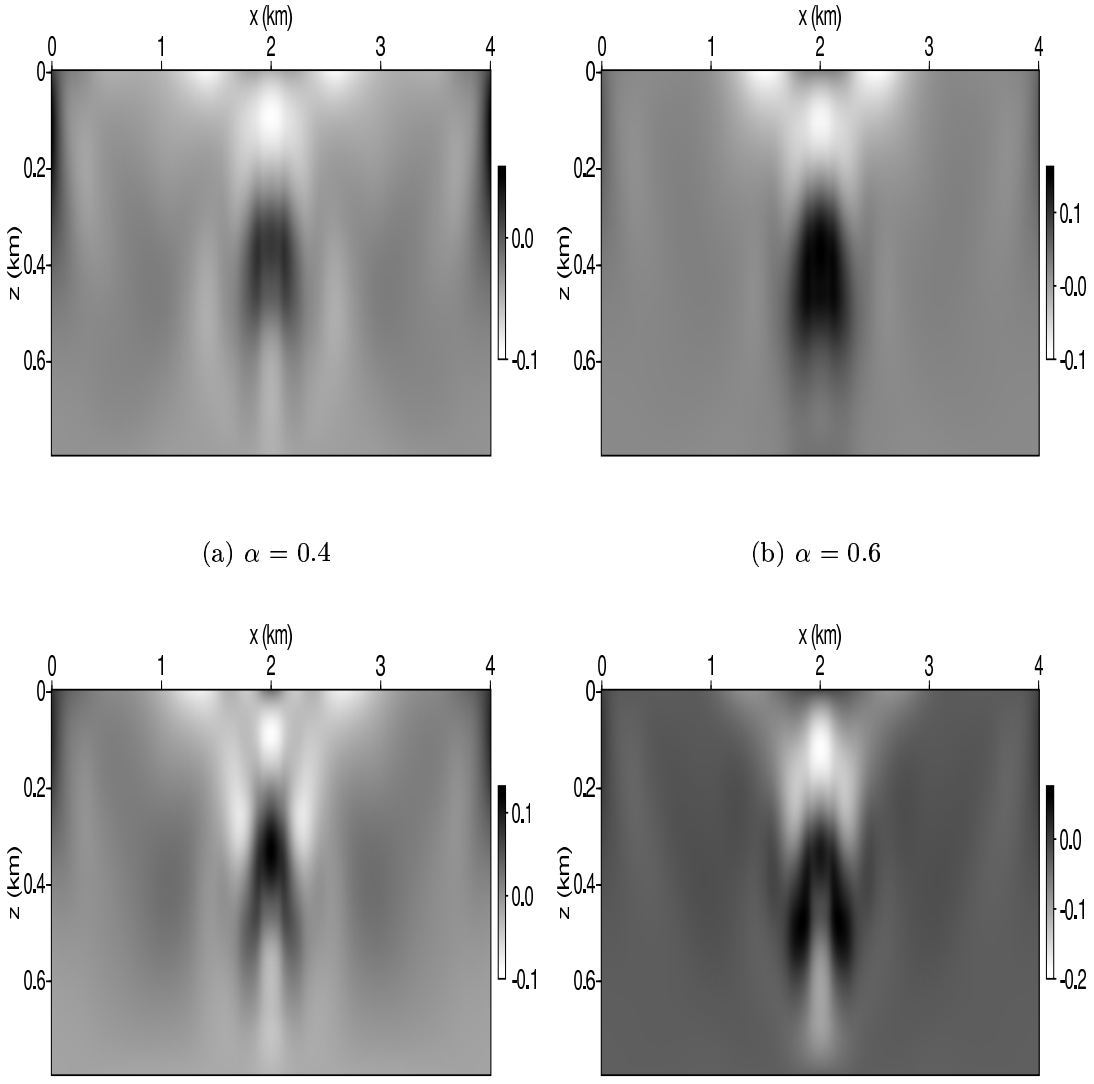
(a) $\alpha = 0.4$

(b) $\alpha = 0.6$

(c) $\alpha = 0.8$

(d) $\alpha = 1$

Figure 5.19 Inverted velocities at the fifth iteration for various lens models. The amplitude and the shape of the lens are well reconstructed.



(a) $\alpha = 0.4$

(b) $\alpha = 0.6$

(c) $\alpha = 0.8$

(d) $\alpha = 1$

Figure 5.20 The differences between inverted velocities and the correct velocities. The differences are in general with 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 (¡ 20 m) trichloro-ethene contaminated aquifer. The trichloro-eheme is a dense nonaqueous phase liquid (DNAPL). There are 5-20 meters of Quaternary sands, gravels and clays cover 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 aid further anti-remediation efforts.

The 3-D reflection seismic experiments made use of Texan portable seismographs with sources generated by 223 caliber single shot rifles fired in 6 cm drillholes. Acquisition geometry for the experiment is shown if Figure (5.21. For the VSP survey, surface shots fired every 0.7m to offsets of 21m were recorded by both receivers in two boreholes spaced at 0.5m increments and receivers on surface distanced at 0.35

increments. 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-borehole diffraction tomography.

5.3.2 Results from previous studies

The 223 rifle 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 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 can 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 is conducted using the VSP data *Gao, 2003* [6]. Both surface data and borehole data are fitted. The surface-borehole geometry provides reasonably good ray coverage which is in particular favorable to tomographic type of velocity analysis. See Figure (5.22(a)).

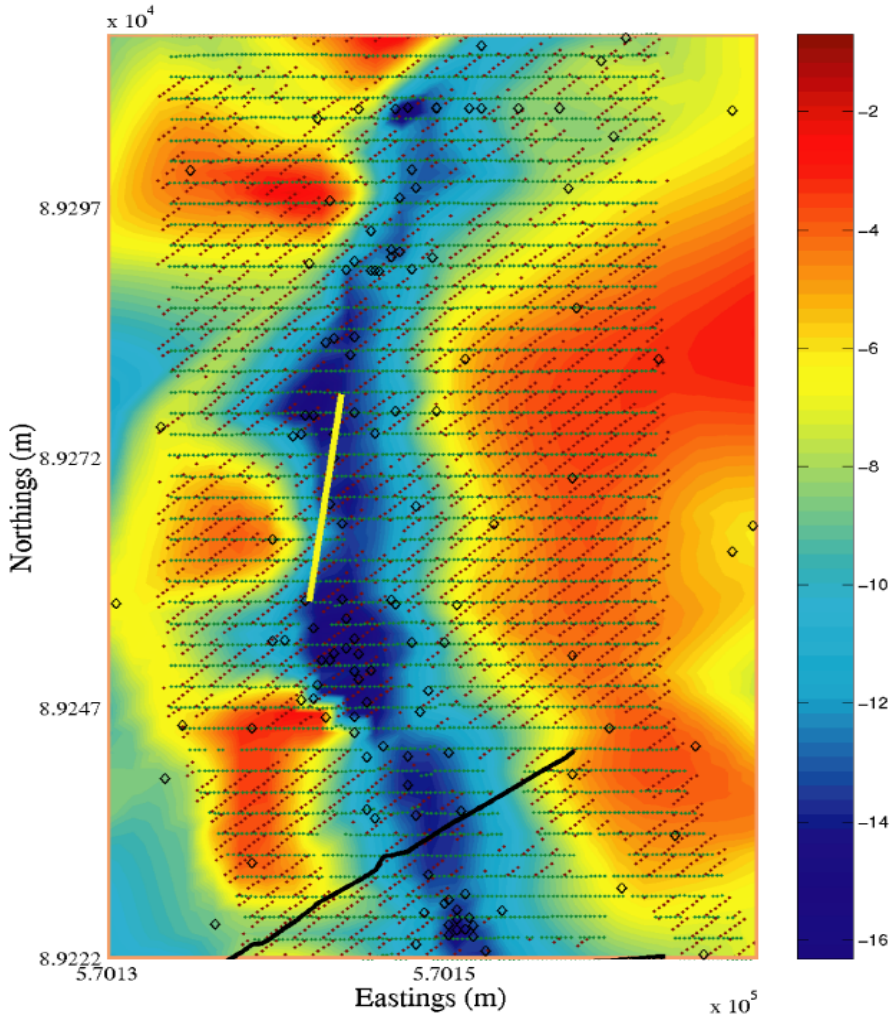
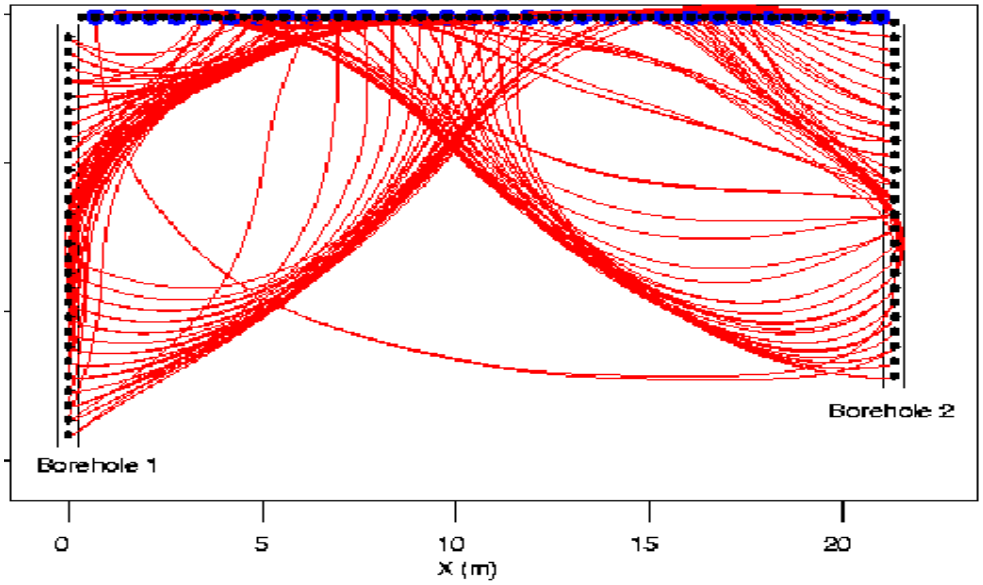
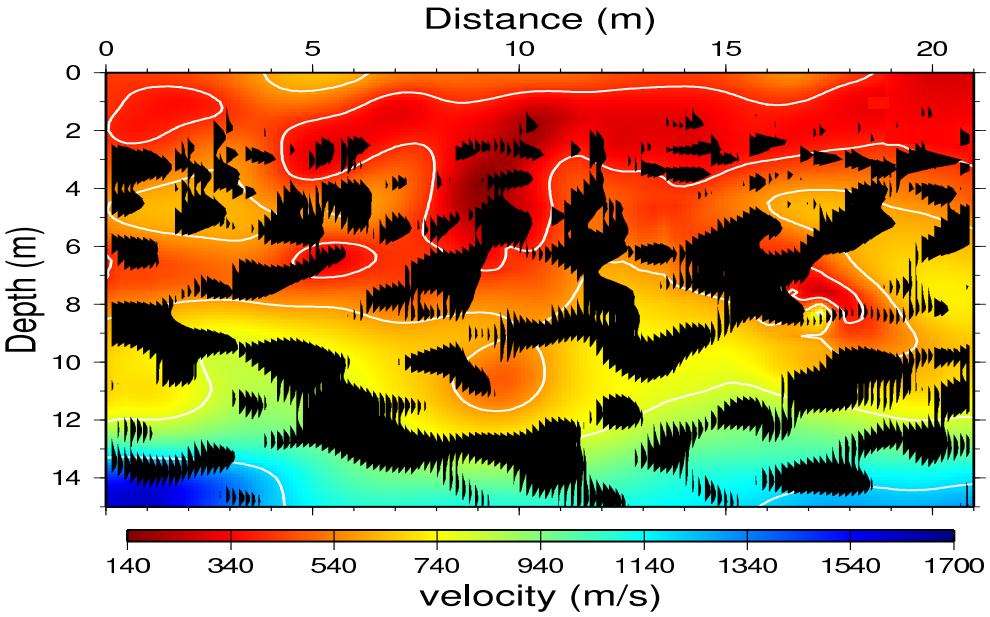


Figure 5.21 The 3-D reflection seismic experiments made use of 624 Texan portable seismographs, and 2 Geometrix multi-channel seismographs. For sources we used 223 caliber single shot rifles, fired in 6 cm drillholes. The Texan seismographs were deployed in 6 parallel lines with cross-line separations 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.



(a) Ray coverage using surface-borehole geometry



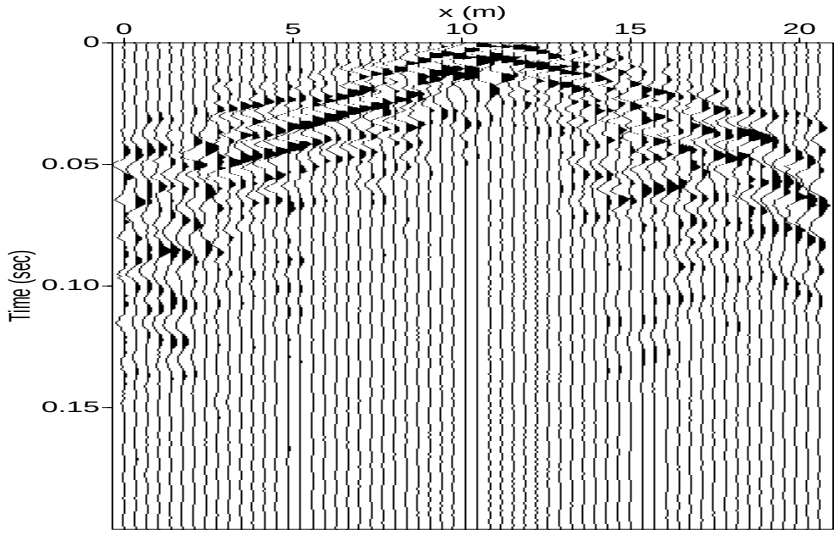
(b) Image (black) plotted on top of the velocity model optimized through diffraction tomography.

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

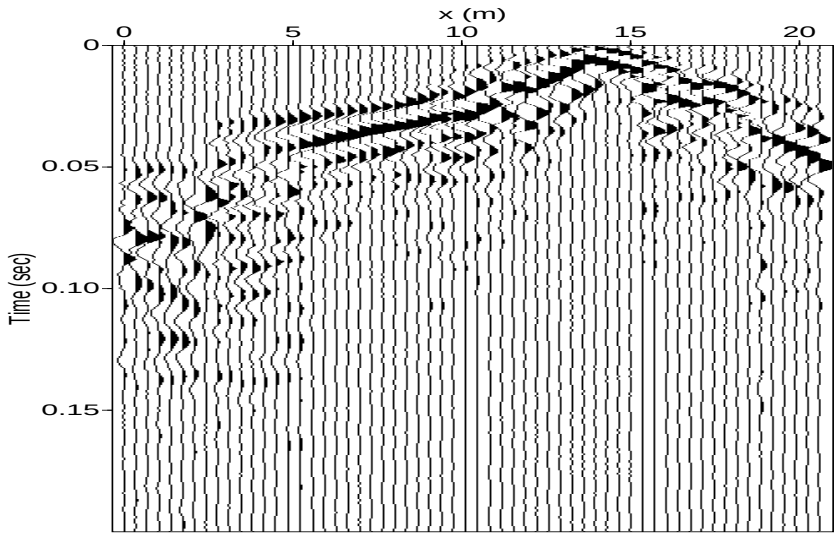
5.3.3 Results by differential semblance velocity analysis

The VSP data received on 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. The 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 5.18×10^9 compared to the initial objective function value 1.18×10^10. The offset gathers can not be further focused within $h \approx 2$m, the threshold estimated as the wave-length of the propagating waves. See Figure (5.26).

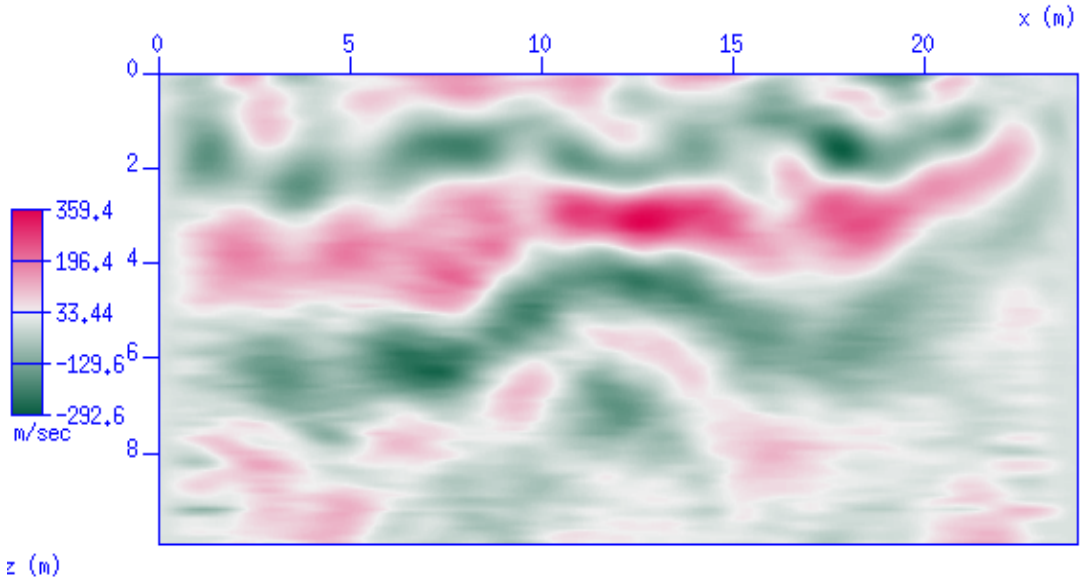


(a) Source gather of $x_s = 12.6$m preprocessed by trace-mixing.

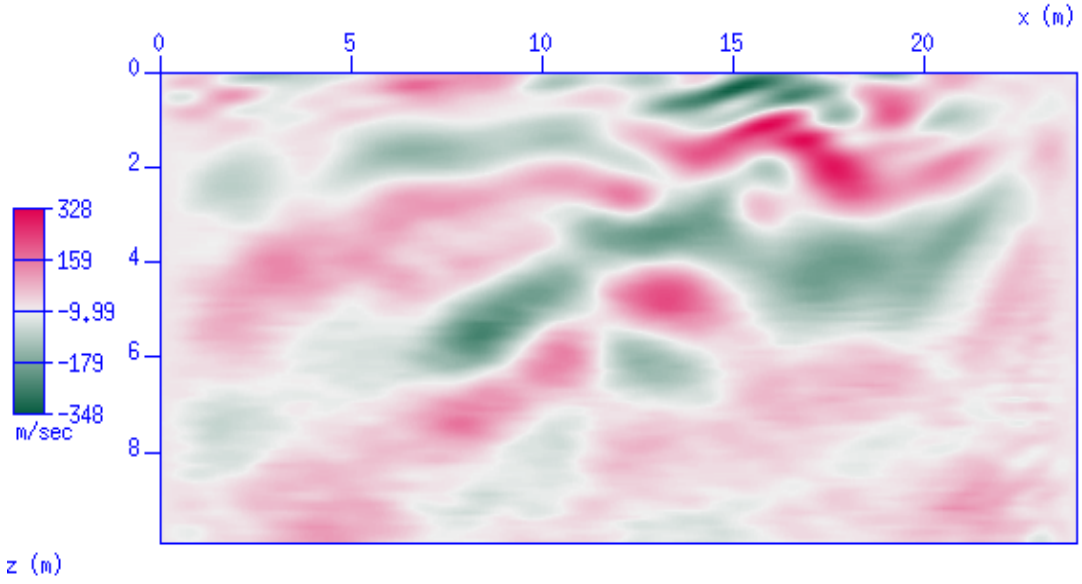


(b) Source gather of $x_s = 15.4$m preprocessed by trace-mixing.

Figure 5.23 Source gathers at $x_s = 12.6$ and $x_s = 15.4$. Data is surface wave dominated.

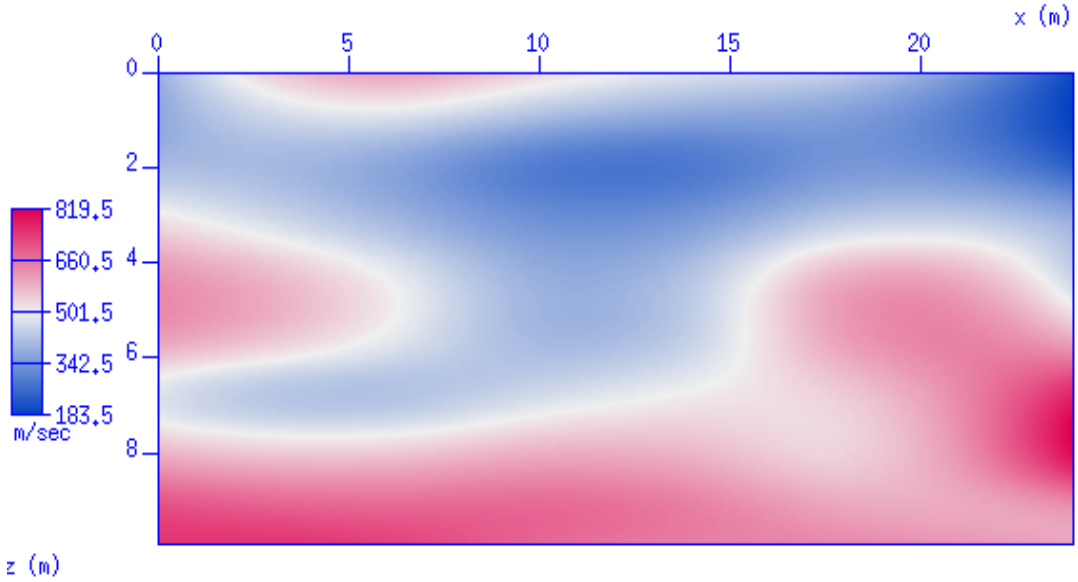


(a) Initial image.

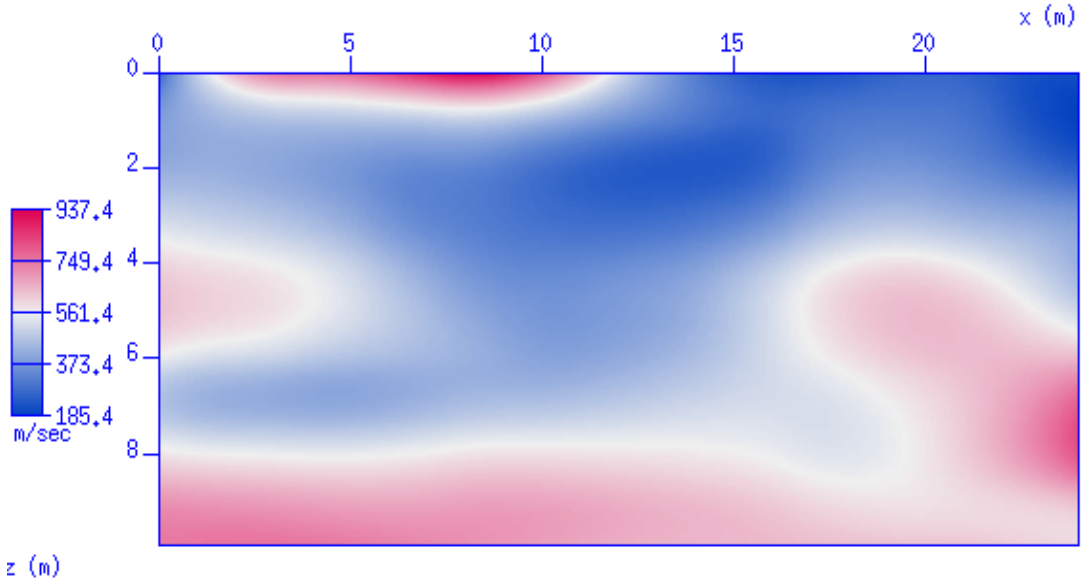


(b) Optimized image at the 4^{th} iteration.

Figure 5.24 Comparison of initial and final images. Deeper structures are better characterized in the final image.

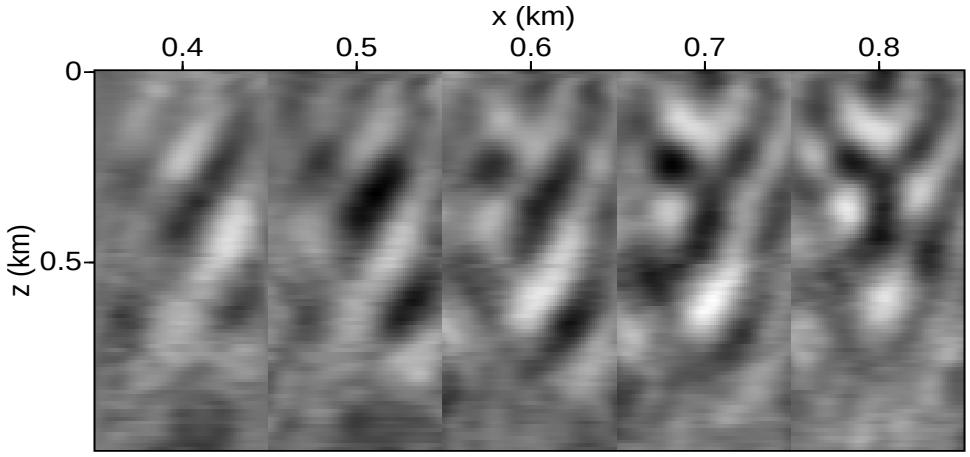


(a) Initial velocity model

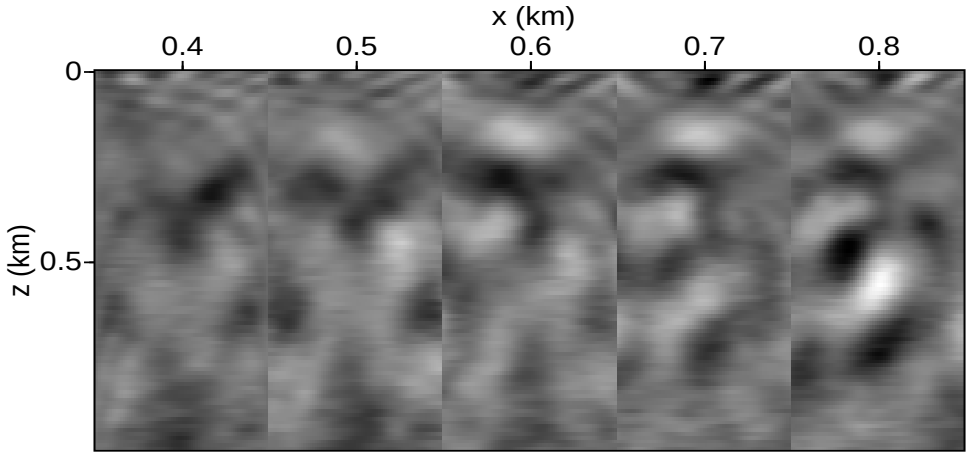


(b) Optimized velocity model at the 4^{th} iteration.

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-borehole data.



(a) Offset gathers at the initial velocity model



(b) Offset gathers at the optimized velocity model

Figure 5.26 Offset gathers migrated at initial and final (4^{th} 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.

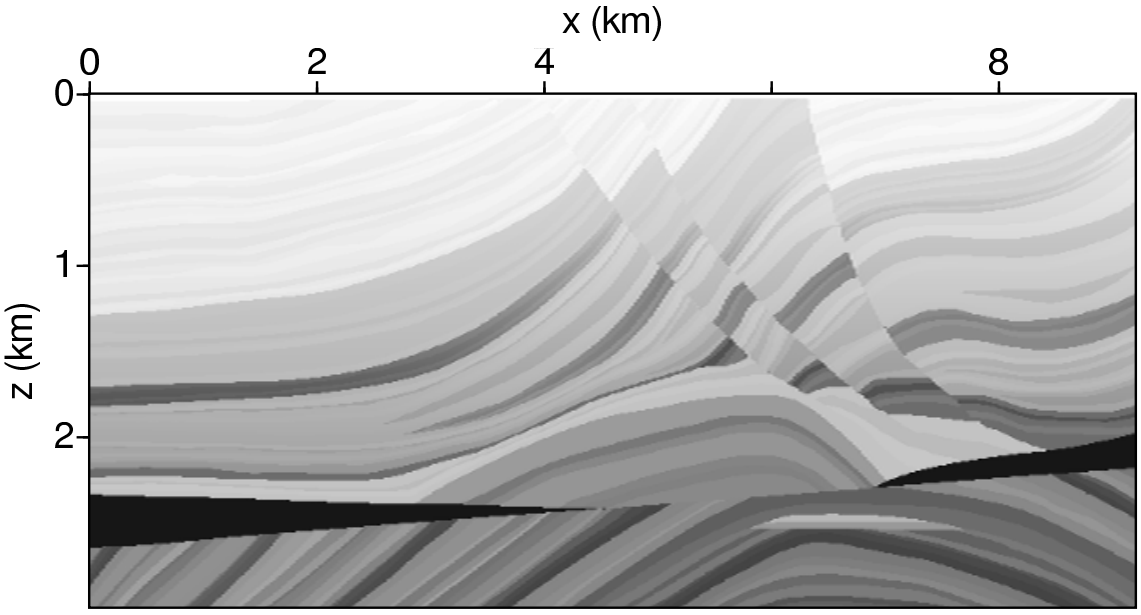
5.4 Marmousi data set

The original Marmousi dataset consists of 240 source gathers each of which is recorded by 96 geophones incremented at 25m at a trailing geometry. The data we use are the original Marmousi traces deconvolved with source signature and continue to zero surface offset. The continuation of to the 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.

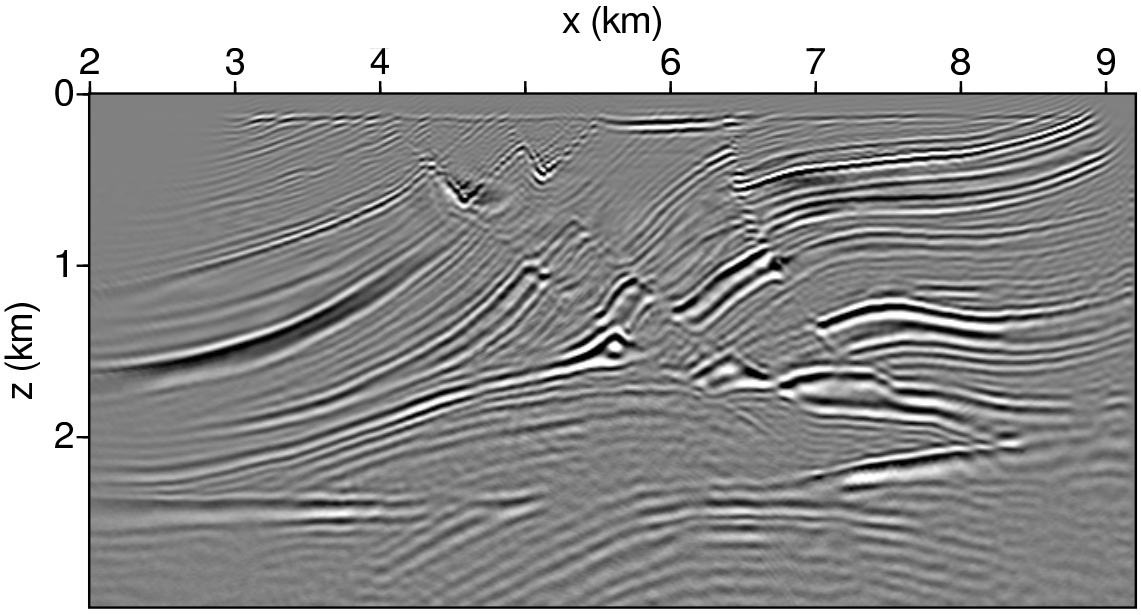
The image obtained using the true velocity model is presented in Figure (5.27(b)) showing the robustness of the migration algorithm in the presence of complex velocity structure. The differential semblance criterion in offset is still valid for this complex velocity structure. Figure (5.28(c)) shows the objective function evaluated at various velocity model 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 maximum 50% perturbation with respect to V. The objective function plotted against α draws a smooth curve centered at the true velocity in a wide range.

The main feature of the Marmousi velocity structure appears in the salt dome in the middle of the model covered by high angle dipping faults. The long fault planes intersected with the planes of sediments induce complicated ray path coverage 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 0.6km in depth. We have used 6×6 B-spline grid to represent velocities. The smoothed gradient at the initial constant (1.8km/sec) velocity model appears in Figure (**??**) with the B-spline smoothing implicitly used $BB^*\nabla_c J$. 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 values by roughly a factor of two, the length of the gradient by an order of magnitude in five iterations. The fifth velocity iterate appears as Figure (5.29(b)). Most of the shorter-scale lateral features in the velocity have now reconstructed. Comparison of image gathers at initial vs. iteration five velocities shows considerable improvement in focussing (Figure (5.30).

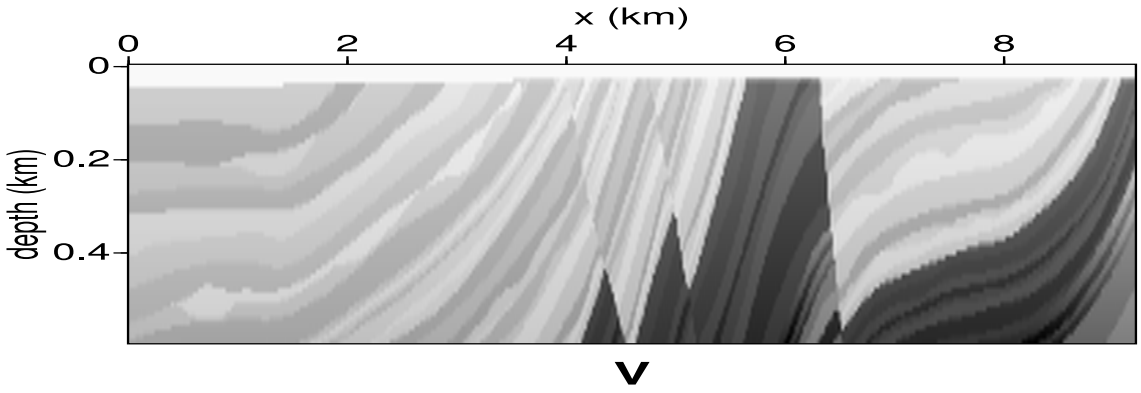


(a) The Marmousi velocity structure.

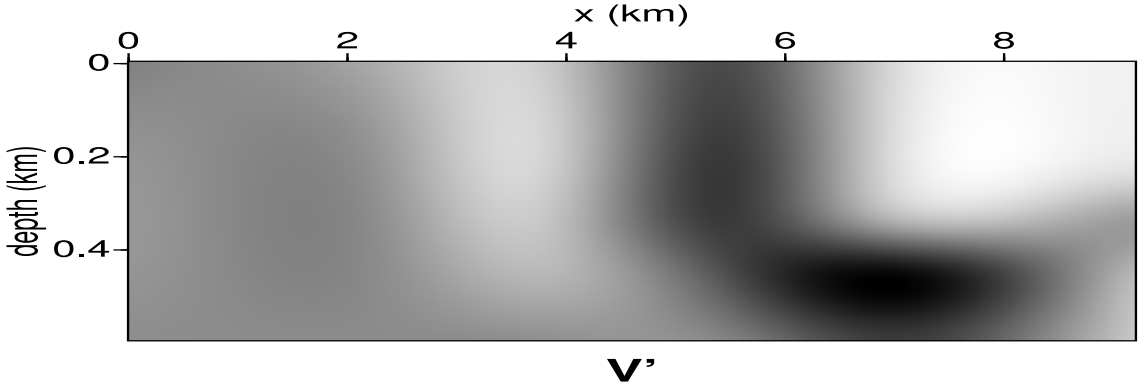


(b) Image obtained using the true velocity structure.

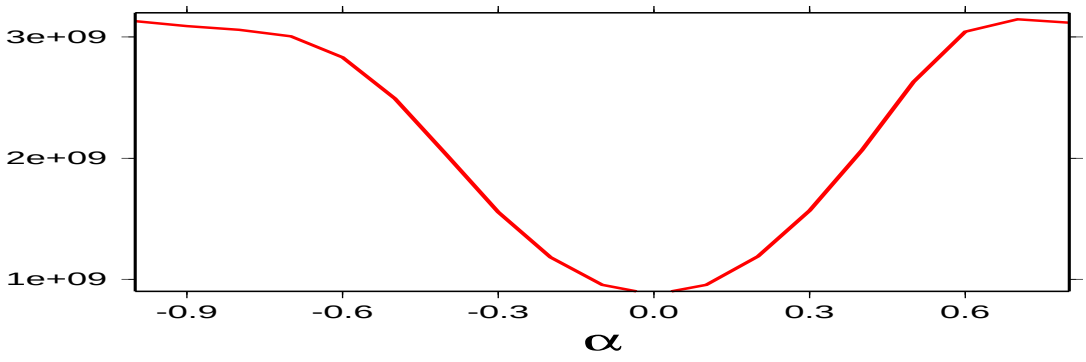
Figure 5.27 (a). The Marmousi velocity structure. (b). Image obtained at the correct velocity appeared in (a).



(a) The exact Marmousi velocity model V down to 0.6km.

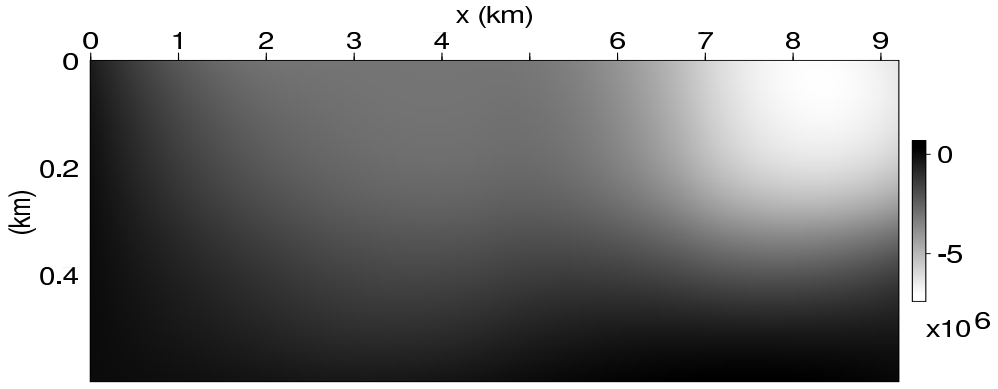


(b) Velocity perturbation V'.

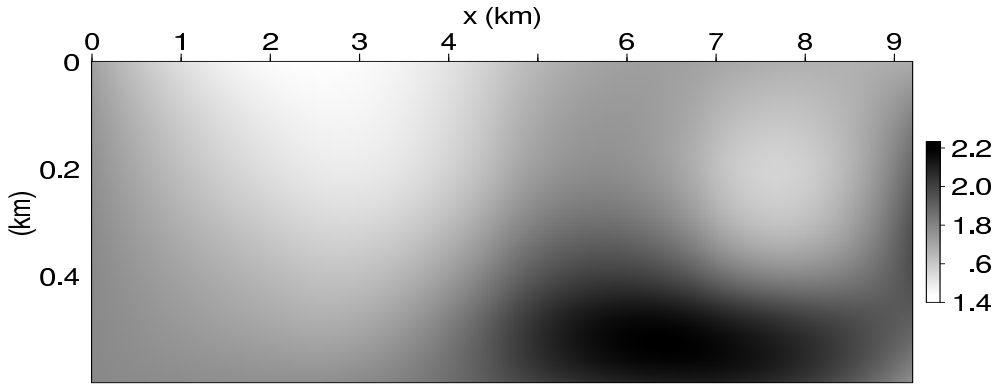


(c) Objective function evaluated at $J(V + \alpha V')$

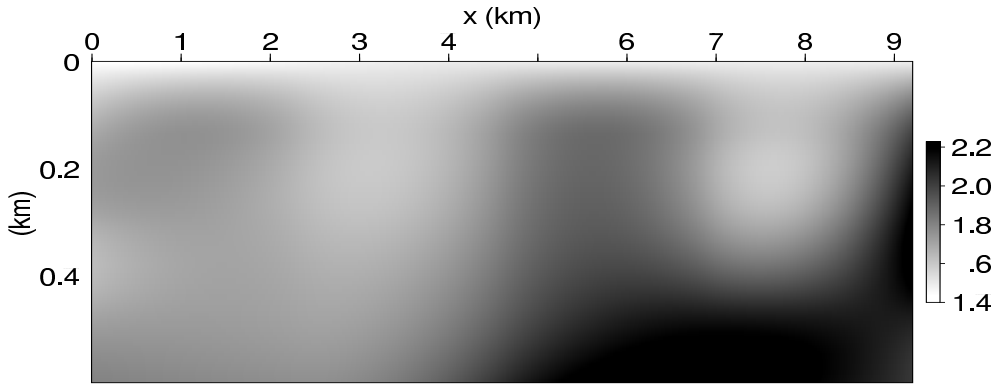
Figure 5.28 The differential semblance objective function varies in a wide and smooth curve with the minimum centered at the true velocity.



(a) Projected gradient at the constant (1.8km/sec) velocity.

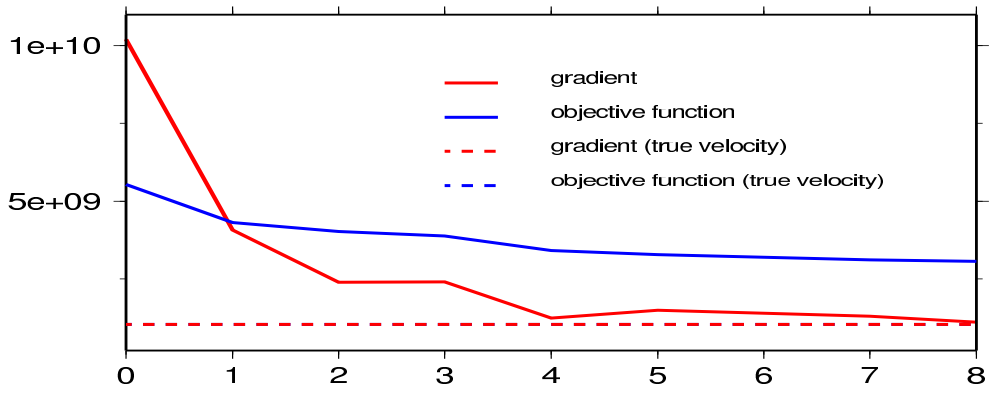


(b) The fifth velocity iterate.

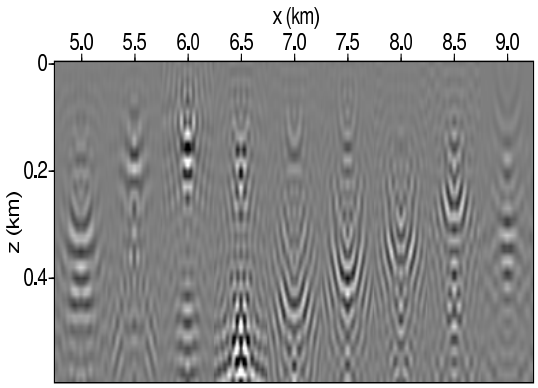


(c) B-spline projection of the true velocity.

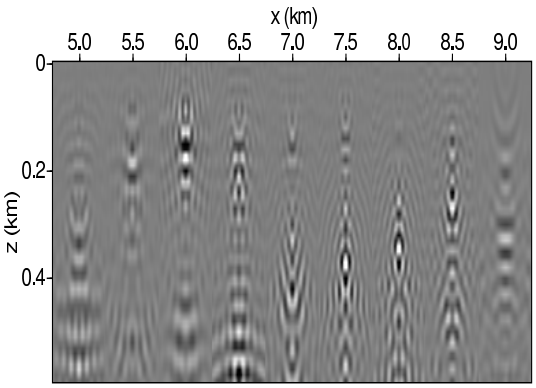
Figure 5.29 (a). B-spline model space gradient projected back to image space $B\nabla_m J = BB^*\nabla_c J$. (b). Velocity at the fifth iterate. (c). The best fit projected B-spline velocity $V_b = Bm_b$, where m_b solves $Bm = V$ in a least-squares sense for the true Marmousi velocity V defined on 921×60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



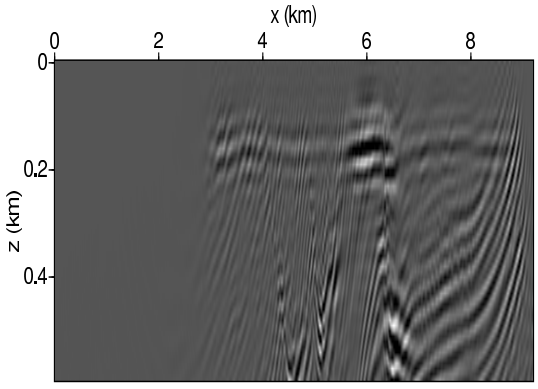
(a) Iterates of gradients and objective functions.



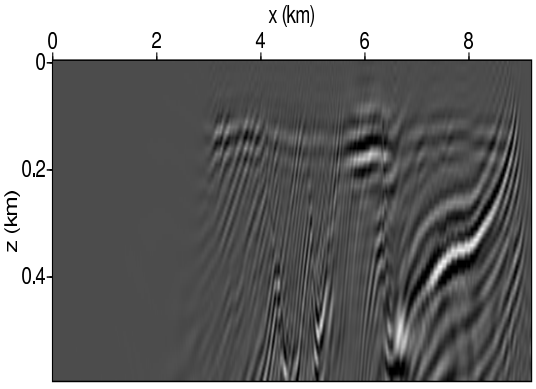
(b) Initial image offset gathers at constant (1.8km/sec) velocity.



(c) Initial image at constant (1.8km/sec) velocity.



(d) Image gathers at the fifth velocity iterate.



(e) Image at the fifty velocity iterate.

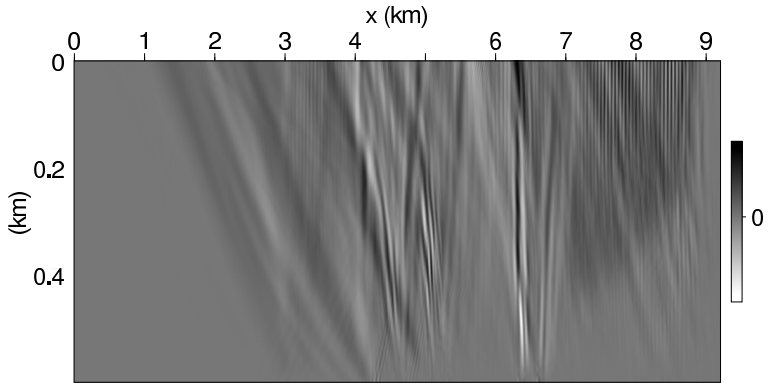
Figure 5.30 (a). Iterates of magnitude of gradient and objective function values. The magnitude of gradient at the true velocity coincide with the objective function value represented in the red dashed line. (b). Scattered image offset gathers at the initial constant velocity. (c). Initial image at constant 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

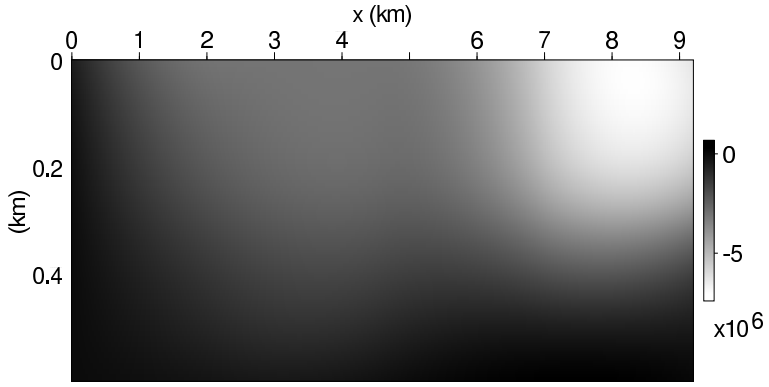
The analysis in section (4.2.2) indicates that the image space gradient, in high frequency asymptotic, recovers the ray path compatible with the velocity used in migration. Traces of ray paths can be appear in Figure (5.31) especially in the region $7km < x < 9km$, where horizontal reflectors creates ray path of vertical reflection. The fault plane in the middle of the model complicate 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 certain degrees of smoothness (which is controlled by the B-spline projection). Increasing the number of B-spline grid points 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 degreases at representations of the true velocity model with increased roughness. Using the same set of velocities, 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 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 (1.8km/sec) velocity.



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

Figure 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.

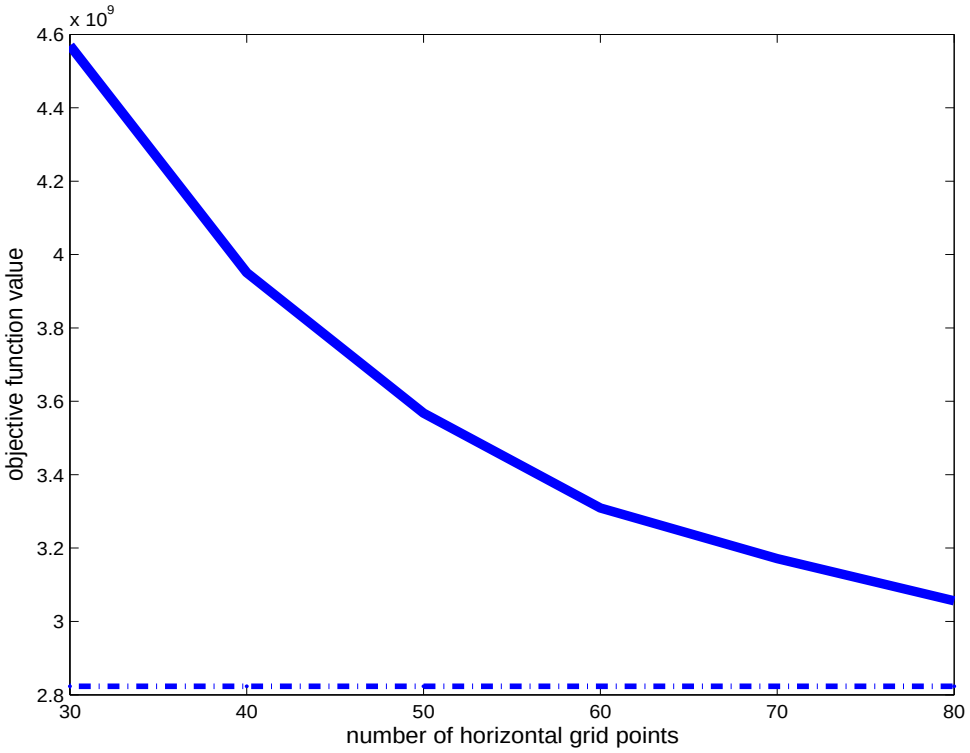
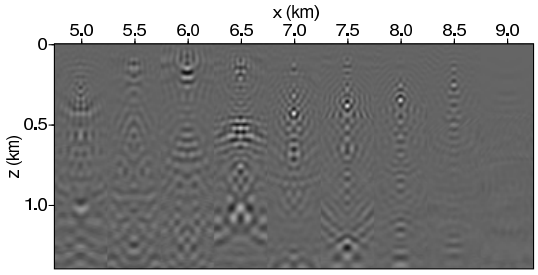
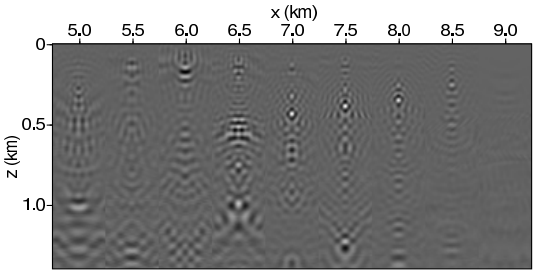


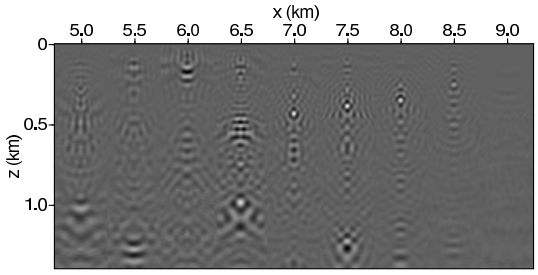
Figure 5.32 Objective function evaluated at representations of the true velocity from surface to 1.4km in depth with increased roughness. Number of grid points in horizontal direction of the B-spline model increases from 30 to 80 increments by 10.



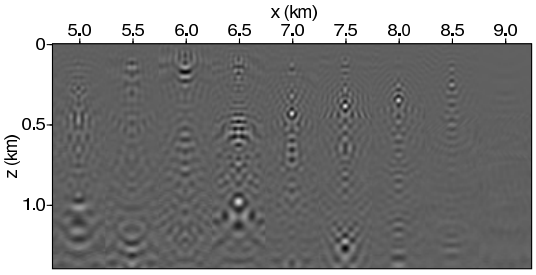
(a) Horizontal B-spline grid points bx = 30.



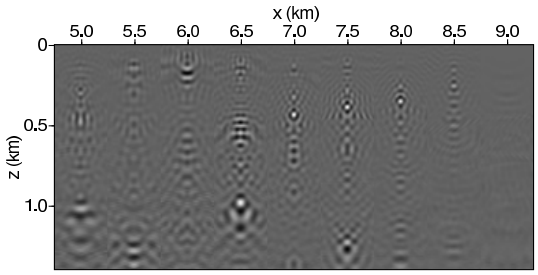
(b) Horizontal B-spline grid points bx = 30.



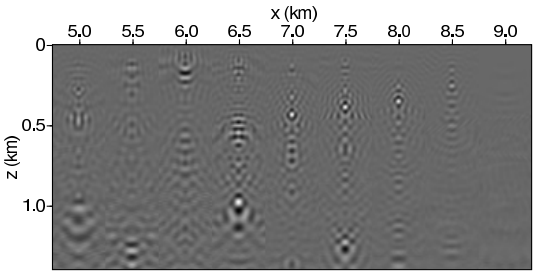
(c) Horizontal B-spline grid points bx = 30.



(d) Horizontal B-spline grid points bx = 30.

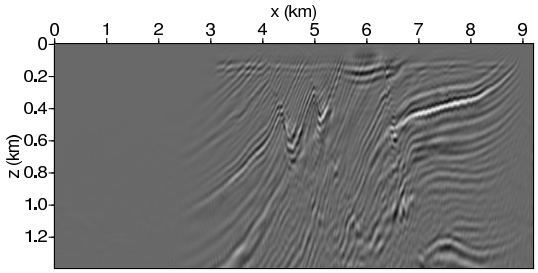


(e) Horizontal B-spline grid points bx = 30.

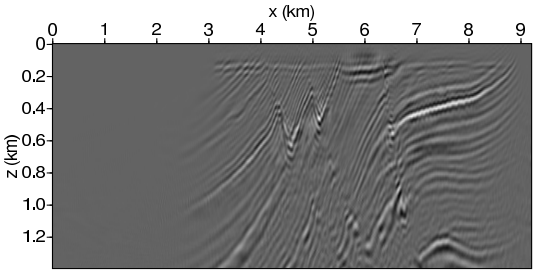


(f) Horizontal B-spline grid points bx = 30.

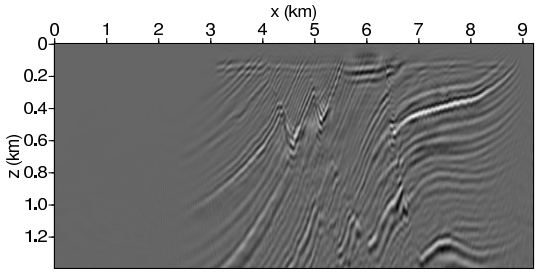
Figure 5.33 Offset gathers become more focused at velocities of higher degree of roughness.



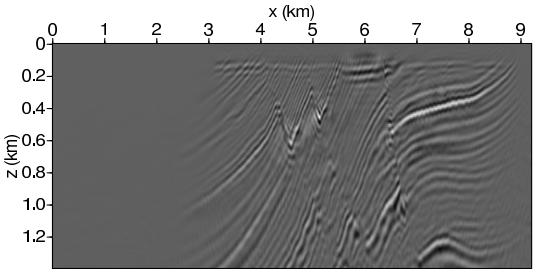
(a) Horizontal B-spline grid points bx = 30.



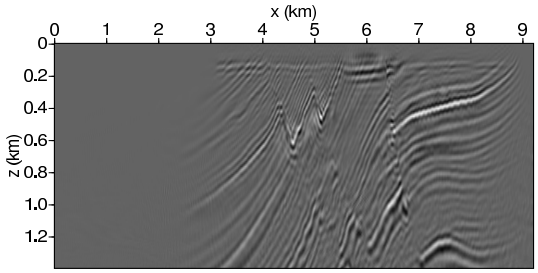
(b) Horizontal B-spline grid points bx = 40.



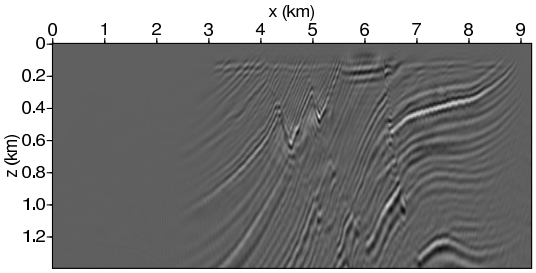
(c) Horizontal B-spline grid points bx = 50.



(d) Horizontal B-spline grid points bx = 60.



(e) Horizontal B-spline grid points bx = 70.



(f) Horizontal B-spline grid points bx = 80.

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 theists, which directly perseus the quality of the image through criterion of differential semblance optimization. Main contributions of this thesis appears in three aspects: 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

The absorbing boundary condition is often useful for synthetic seismic data simulation. This section gives an algorithm of the perfectly matched layer absorbing boundary condition for time domain acoustic wave propagation.

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

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

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

where v is the particle velocity field, p is the pressure, and κ 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 diffusive system by adding diffusive terms to eq(A.1) and eq(A.2). We have

$$\rho \frac{\partial}{\partial t} v = -\nabla p - \eta v \tag{A.3}$$

$$\frac{\partial}{\partial t} p = -\kappa \nabla \cdot v - \lambda p \tag{A.4}$$

where η is the linear viscous coefficient. The additional term, $-\eta v$, in eq(A.3) is responsible for the exponential energy decay. The constitutive relation, eq(A.2) (incompressibility condition) should be changed accordingly to match eq(A.3) through a thermodynamic argument. We continue to solve systems (A.3)+(A.4), assuming κ, λ, and η is slow varying in time and ρ is slow varying in space, we derive

$$\begin{aligned} \frac{\partial^2}{\partial t^2}p &= \kappa\nabla\cdot(\frac{\nabla p}{\rho}) + \kappa\nabla\cdot(\frac{\eta}{\rho}v) - \lambda\frac{\partial p}{\partial t} \\ &= \frac{\kappa}{\rho}\nabla^2 p - (\frac{\eta}{\rho}+\lambda)\frac{\partial p}{\partial t} - \frac{\eta\lambda}{\rho}p + \frac{\kappa}{\rho}\nabla\eta\cdot v \end{aligned} \tag{A.5}$$

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

$$v(\vec{x},t) = -\int_{t_0}^{t} e^{-\frac{\eta}{\rho}(t-s)}\frac{\nabla p}{\rho}(\vec{x},s)ds \tag{A.6}$$

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

$$\begin{array}{l} v_a = 0 \\ \text{for } t = 0 : dt : t_{max} \\ \quad v_a = v_a - e^{t\frac{\eta}{\rho}}\frac{\nabla p}{\rho}dt \\ \quad v = e^{-t\frac{\eta}{\rho}}v_a \\ \quad \text{update } p \text{ through equation(A.5)} \\ \text{end}(t) \end{array} \tag{A.7}$$

Q analysis:

The attenuation analysis follows the argument by *Cohen, 1999* [2]. The analysis is carried through a use of imaginary coordinates on which the energy of wavefield is attenuated most significantly parallel to the direction of propagation of wavefront. The main conclusion is that alone the direction of wave propagation, the wave attenuates according to

$$\frac{1}{Q(\omega)} = \eta \frac{\omega \lambda}{\pi c} \tag{A.8}$$

where as usual ω, λ, c denotes angular frequency, wavelength and velocity, respectively. At non-dispersive medium it is attenuated independent of freuency ω, $1/Q(\omega) = 2\eta$.

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

$$i\omega(\eta + \rho)\hat{\vec{v}} = -\nabla \hat{p} \tag{A.9}$$

$$i\omega(\lambda + 1)\hat{p} = -\kappa \nabla \cdot \hat{\vec{v}} \tag{A.10}$$

We further derive

$$\hat{p} = \frac{\kappa}{iw(\lambda+1)}\nabla \cdot \frac{\nabla \hat{p}}{iw(\eta+\rho)} \tag{A.11}$$

To ease the analysis, let's assume $\rho = 1$ and $\eta = \lambda$. Equation (A.11) can be written as

$$\omega^2 \hat{p} + \kappa \frac{i\omega}{i\omega(\eta+1)}\nabla \cdot \frac{i\omega \nabla \hat{p}}{i\omega(\eta+1)} \tag{A.12}$$

It reduces to

$$\omega^2 \hat{p} + \kappa \tilde{\nabla}^2 \hat{p} = 0 \tag{A.13}$$

by change of variables

$$\tilde{x_1} = x_1 + \frac{i}{\omega}\int_0^{x_1} \eta(s)ds \tag{A.14}$$

$$\tilde{x_2} = x_2 + \frac{i}{\omega}\int_0^{x_2} \eta(s)ds \tag{A.15}$$

A plane wave solution can be formulated

$$\hat{p} = e^{i(\vec{k}\cdot\tilde{x} - i\omega t)} \tag{A.16}$$

$$= e^{-\frac{k_1 \int_0^{x_1} \eta(s)ds + k_2 \int_0^{x_2} \eta(s)ds}{\omega}} e^{i(\vec{k}\cdot x - i\omega t)} \tag{A.17}$$

so

$$|p| = e^{-\frac{k_1 \int_0^{x_1} \eta(s)ds + k_2 \int_0^{x_2} \eta(s)ds}{\omega}} \tag{A.18}$$

and increment of amplitude in the direction of propagation is

$$\Delta|p| = -\eta|p|(k_1\Delta x_1 + k_2 \Delta x_2) \tag{A.19}$$

$$
\begin{aligned}
&= -\eta|p|(k_1\frac{k_1}{k}\Delta x + k_2\frac{k_2}{k}\Delta x) \\
&= -\eta k|p|\Delta x
\end{aligned}
$$

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

$$\frac{1}{Q(\omega)} = -\frac{\Delta|p|}{\pi|p|} \tag{A.20}$$

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

$$\frac{1}{Q(\omega)} = \eta\frac{\omega\Delta x}{\pi c} \tag{A.21}$$

For high frequencies, $\Delta x \to 0$ (Δx is the wave length), it follows

$$\frac{1}{Q(\omega)} = 2\eta \tag{A.22}$$

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 ω, $p(t, x; \omega)$ 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. Recall the wave equation is of the following form

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

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

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

for some integer N as $\omega \to \infty$. This means when ω is sufficiently large, since ω^{-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 satisfies eq(B.2) an asymptotic solution. So we first consider the following form

$$p(t, x; \omega) = e^{iw\Phi(t,x)}a(t, x; \omega) \tag{B.3}$$

with

$$a(t,x;\omega) = \sum_{j}^{m+N} a_j(x)\frac{(iw)^m}{(iw)^j} \tag{B.4}$$

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

$$\frac{\partial^2 p}{\partial t^2} = e^{iw\Phi}[(iw)^2(\frac{\partial\Phi}{\partial t})^2 a + iw(2\frac{\partial\Phi}{\partial t}\frac{\partial a}{\partial t} + \frac{\partial^2\Phi}{\partial t^2}a) + \frac{\partial^2 a}{\partial t^2}] \tag{B.5}$$

$$\frac{\partial^2 p}{\partial x_j^2} = e^{iw\Phi}[(iw)^2(\frac{\partial\Phi}{\partial x_j})^2 a + iw(2\frac{\partial\Phi}{\partial x_j}\frac{\partial a}{\partial x_j} + \frac{\partial^2\Phi}{\partial x_j^2}a) + \frac{\partial^2 a}{\partial x_j^2}] \tag{B.6}$$

Substituting them into the wave equation, we obtain

$$\begin{aligned}(\frac{\partial^2}{\partial t^2} - \nabla_x)p(t,x;\omega) = e^{iw\Phi}\Big\{&(iw)^2[\frac{1}{c^2}(\frac{\partial\Phi}{\partial t})^2 - \sum_j(\frac{\partial\Phi}{\partial x_j})^2]a \\ &+iw[\frac{1}{c^2}2\frac{\partial\Phi}{\partial t}\frac{\partial a}{\partial t} - 2\sum_j\frac{\partial\Phi}{\partial x_j}\frac{\partial a}{\partial x_j} + \frac{1}{c^2}\frac{\partial^2\Phi}{\partial t^2}a - \sum_j\frac{\partial^2\Phi}{\partial x_j^2}a] \\ &+[\frac{1}{c^2}\frac{\partial^2 a}{\partial t^2} - \sum_j\frac{\partial^2 a}{\partial x_j^2}]\Big\}\end{aligned} \tag{B.7}$$

Now introduce the notation for operator

$$\Lambda = \frac{1}{c^2}2\frac{\partial\Phi}{\partial t}\frac{\partial}{\partial t} - 2\sum_j\frac{\partial\Phi}{\partial x_j}\frac{\partial}{\partial x_j} + \frac{1}{c^2}\frac{\partial^2\Phi}{\partial t^2} - \sum_j\frac{\partial^2\Phi}{\partial x_j^2} \tag{B.8}$$

$$\nabla\Phi = (\frac{\partial\Phi}{\partial t}, \frac{\partial\Phi}{\partial x_1}, \frac{\partial\Phi}{\partial x_2}, \frac{\partial\Phi}{\partial x_3}) = (\Phi_t, \Phi_x) \tag{B.9}$$

and

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

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

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

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

$$\begin{aligned} e^{-iw\Phi}(\frac{\partial^2}{\partial t^2} - \nabla_x)p(t,x;\omega) &= (iw)^{m+2}H(x,\nabla\Phi)a_0 + \\ &+(iw)^{m+1}(H(x,\nabla\Phi)a_1 + \Lambda a_0) \\ &+(iw)^{m}(H(x,\nabla\Phi)a_2 + \Lambda a_1 + (\frac{\partial^2}{\partial t^2} - \nabla_x)a_0) \\ &+(iw)^{m-1}(H(x,\nabla\Phi)a_3 + \Lambda a_2 + (\frac{\partial^2}{\partial t^2} - \nabla_x)a_1) \\ &... \\ &+(iw)^{-N+2}(H(x,\nabla\Phi)a_{m+N} + \Lambda a_{m+N-1} + (\frac{\partial^2}{\partial t^2} - \nabla_x)a_{m+N-2}) \\ &+(iw)^{-N+1}(\Lambda a_{m+N} + (\frac{\partial^2}{\partial t^2} - \nabla_x)a_{m+N-1}) \end{aligned} \tag{B.12}$$

$$+(iw)^{-N}(\frac{\partial^2}{\partial t^2} - \nabla_x)a_{m+N}$$

To analyze the above equation, we think the case when $H(x, \nabla\Phi) \neq 0$. In order to make $(iw)^{m+2}$ term to vanish we need to have $a_0 = 0$. Next, with $a_0 = 0$, in order to make $(iw)^{m+1}$ term to vanish we need $v_1 = 0$. 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

$$H(x, \nabla\Phi) = 0 \tag{B.13}$$

In this case, if

$$\Lambda a_0 = 0 \tag{B.14}$$

then

$$\Lambda a_j = -(\frac{\partial^2}{\partial t^2} - \nabla_x)a_{j-1} \qquad (j = 1, 2, .., m + N) \tag{B.15}$$

Equation (B.13) is called the eikonal equation. For $\Phi(t, x)$ usually takes the form of $t - \phi(x)$. Substituting into eq(B.13), we see the usual eikonal equation used in geometric optics

$$|\nabla_x\phi(x)|^2 = \frac{1}{c^2} \tag{B.16}$$

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

$$2\nabla_x\phi \cdot \nabla_x a_0 + \nabla_x^2\phi a_0 = 0 \tag{B.17}$$

which is the form of the first transport equation given in most references.

Appendix C
Migration by an analysis of stationary phase

Now we take the phase function $\Phi(t,x)$ discussed in last section to be of the form

$$\Phi(t,x) = t - \phi(x) \tag{C.1}$$

and assume that the amplitude function a is independent of time t, then the Green's function solution of Hemholtz equation eq(**??**) can be approximated as

$$G(x,\omega) \;\; = \;\; a(x;\omega)e^{-iw\phi(x)} \tag{C.2}$$

To reflect the fact that the wave is propagating from source x to observation point y, we write the Green's function as

$$G(x,y;\omega) = a(x,y;\omega)e^{-iw\phi(x,y)} \tag{C.3}$$

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

$$d(x_r, x_s; \omega) \approx \delta p(x_r, x_s; \omega) = 2 \int a(\omega, x; x_s) a(\omega, x_r; x) \frac{\omega^2}{c^2} r(x) e^{-iw\phi(x,x_s)} e^{-iw\phi(x_r,x)} dx \tag{C.4}$$

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

$$|\nabla_x \phi(x, y)|^2 = 1/c(x)^2 \tag{C.5}$$

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 eq(B.4), the term $a(x, y; \omega)$ is a sum of slow varying amplitudes, each of which is at homogenous degree of, at most, m in ω. In particular the amplitude $a_0(x, y)$ in eq(B.4) satisfies the first transport equation

$$\nabla_x \cdot [a_0^2(x, y) \nabla_x \phi(x, y)] = 0 \tag{C.6}$$

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

$$d_{x'}(x_r, x_s; \omega) = 2 \int a(x, x_s; \omega) a(x_r, x; \omega) \frac{\omega^2}{c^2} \psi_{x'}(x) \hat{r}(\zeta) e^{i\zeta \cdot x} e^{iw[-\phi(x,x_s)-\phi(x_r,x)]} dx d\zeta \tag{C.7}$$

The adjoint of the integral operator in eq(C.7), using the same velocity function c, applied to data can be written

$$\tilde{r}_{x'}(y) = \int \overline{a(y, x_s; \omega) a(x_r, y; \omega)} e^{iw(\phi(y,x_s)+iw\phi(x_r,y))} d_{x'}(x_r, x_s; \omega) dx_r dx_s dw \tag{C.8}$$

Substituting eq(C.7) into eq(C.8), we obtain

$$\tilde{r}_{x'}(y) = \int A(y, x, x_s, x_r; \omega) \psi(x) \hat{r}(\zeta) e^{i\zeta \cdot x + iw[-\phi(x,x_s)-\phi(x_r,x)+\phi(y,x_s)+\phi(x_r,y)]} dx_r dx_s dx dw d\zeta \tag{C.9}$$

Introduce a cut-off function $\psi_{y'}(y) \in C_c^\infty$, identically 1 a and compactly supported in a neighborhood of y'. We study the asymptotic behavior of the form

$$\begin{aligned} J_{x',y'}(k) &= \int \tilde{r}_x(y) \psi(y) e^{-ik \cdot y} && \text{(C.10)} \\ &= \int A(\omega, x, x_s, x_r, y) \hat{r}(\zeta) \psi_{x'}(x) \psi_{y'}(y) \times && \text{(C.11)} \\ &\quad e^{i\zeta \cdot x - ik \cdot y + iw[-\phi(x,x_s)-\phi(x_r,x)+\phi(y,x_s)+\phi(x_r,y)]} dw d\zeta dx dy dx_s dx_r \end{aligned}$$

In agree to the form of stationary phase analysis, J can be written as

$$J_{x',y'}(k) = \int B(\omega, \zeta, \sigma) e^{i\Theta(k,\omega,\zeta,\sigma)} dw d\zeta d\sigma \tag{C.12}$$

where $d\sigma = dxdydx_s dx_r$. To proceed, we point out (1) 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 $a(\omega, x, y)e^{iw\phi(x,y)}$ in the neighborhood of x.

$$W_x(\xi) = \int a(\omega, x, y) e^{iw\phi(x,y)} e^{-i\xi \cdot x} \psi(x) dx \tag{C.13}$$

Asymptotically for large $|\xi|$, it picks up a non-trivial energy when

$$\omega \nabla_x \phi(x, y) = \xi \tag{C.14}$$

As we know, ξ 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 (2)

$$\nabla_x \phi(x, y)|_{x=y} = -\nabla_y \phi(x, y)|_{x=y} \tag{C.15}$$

which can be easily verified. We now turn back to the stationary phase analysis of the integral eq(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

$$\frac{\partial}{\partial\omega}\Theta = 0 \Rightarrow \quad \phi(x, x_s) + \phi(x_r, x) = \phi(y, x_s) + \phi(x_r, y) \tag{C.16}$$

$$\nabla_x\Theta = 0 \Rightarrow \quad \zeta + \omega\nabla_x\phi(x, x_s) + \omega\nabla_x\phi(x_r, x) = 0 \tag{C.17}$$

$$\nabla_y\Theta = 0 \Rightarrow \quad -k - \omega\nabla_y\phi(y, x_s) - \omega\nabla_y\phi(x_r, y) = 0 \tag{C.18}$$

$$\nabla_{x_s}\Theta = 0 \Rightarrow \quad \nabla_{x_s}\phi(x, x_s) = \nabla_{x_s}\phi(y, x_s) \tag{C.19}$$

$$\nabla_{x_r}\Theta = 0 \Rightarrow \quad \nabla_{x_r}\phi(x_r, x) = \nabla_{x_r}\phi(x_r, y) \tag{C.20}$$

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, eq(C.19) and eq(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 eq(C.18) and eq(C.17) that the wavefront set of discontinuity distribution r is reconstructed.

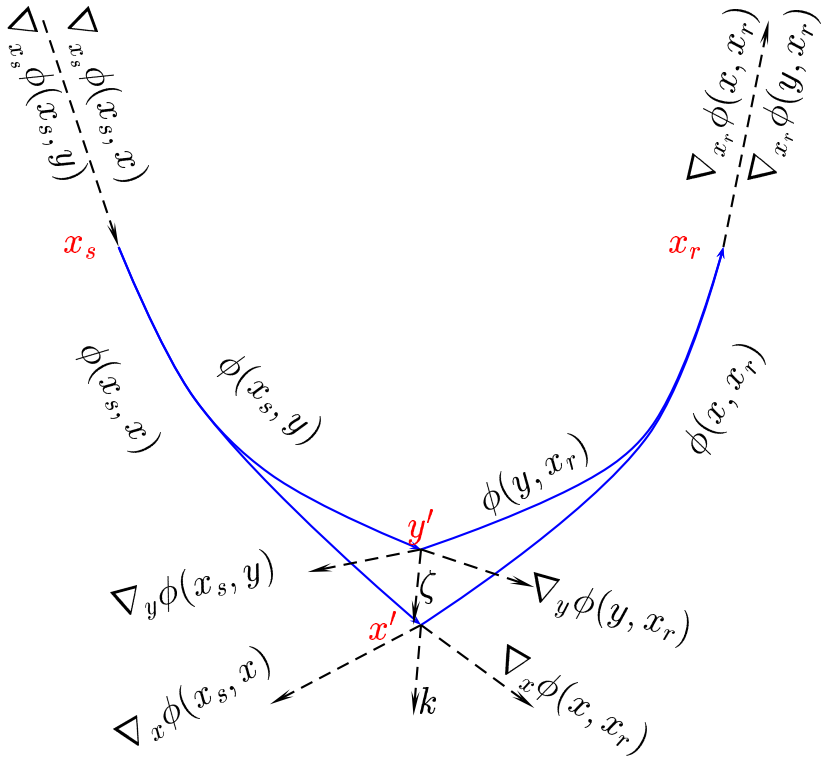


Figure C.1 The travel time is preserved. Snell's law applies to both image point and the reflection point. The ray parameters coincide at the surface.

References

1. J.F. Claerbout. *Imaging the Earth's Interior*. Blackwell Scientific Publishers, Oxford, 1985.

2. G. C. Cohen. *Higher-order numerical methods for transient wave equations*. Springer Verlag, 2001.

3. S. R. Deans. *The radon transform ad some of the applications*. John Wiley & Sons, Inc, 1983.

4. P. Docherty, C. Artley, M. Plumlee, M. Sullivan, and R. R Windels. Tomographic migration velocity analysis in 3-d. *Expanded Abstracts, Society of Exploration Geophysicists, 70nd Annual International Meeting*, pages 983–941, 2000.

5. T. Fei, M. Fehler, and S. T. Heldebrand. Depth migration artifacts associated with first-arrival travel-time. *Expanded Abstracts, Society of Exploration Geophysicists, 66nd Annual International Meeting*, pages 499–502, 1996.

6. Fuchun Gao. *Waveform tomography and its application to a ground water contamination site*. PhD thesis, Rice University, April 2003.

7. R. Giering and T. Kaminski. Recipes for adjoint code construction. *ACM trasactions on mathematical software*, pages 437–374, 1998.

8. M. Gockenbach, D. Reynolds, P. Shen, and W. Symes. Efficient and automatic implementation of the adjoint state method. *ACM Transactions on Mathematical Software*, 28:22–44, 2002.

9. A. Grigs and J. Sjöstrand. *Microlocal analysis for differential operators, An introduction*. Cambridge Universiy Press, 1994.

10. L Hörmander. *The analysis of linear partial differential operators II*. Springer-Verlag, 1990.

11. L Hörmander. *The analysis of linear partial differential operators IV*. Springer-Verlag, 1990.

12. M. Ikawa. *Hyperbolic partial differential equations and wave phenomena*. Americal Mathematical Society, Rhode Island, 2000.

13. M. S Joshi. *Introduction to pseudo-differential operators*. http://arxiv.org/abs/math.AP/9906155.

14. M. W. Lee and S. Y. Suh. Optimization of one-way wave equations. *Geophysics*, pages 1634–1637, 1981.

15. J. Nocedal and S. Wright. *Numerical Optimization*. Springer Verlag, New York, 2000.

16. G. Nolet. *Seismic Tomography*. D. Reidel & Co., Dordrecht, 1987.

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

18. R. G. Pratt, Song. Z. M., P. R. Williamson, and M. R. Warner. Two-dimensional velocity models from wide angle seismic data by waveform inversion. *Geophysical Journal International*, 61:549–560, 1996.

19. P. Sava. Prestack stolt residual migration for migration velocity analysis. *Expanded Abstracts, Society of Exploration Geophysicists,70nd Annual International Meeting*, pages 992–995, 2000.

20. P. Sava and B. Biodi. Wave-equation migration velocity analysis by inversion of differential image perturbations. *Expanded Abstracts, Society of Exploration Geophysicists,73nd Annual International Meeting*, pages 2124–2127, 2003.

21. P. Sava and B. Biodi. Wave-equation migration velocity anaysis by inversion of differential image perturbation. *Stanford Exploration Project Annual Report*, pages 299–313, 2003.

22. P. Shen, C. Stolk, and W. Symes. Automatic velocity analysis by differential semblance optimization. *Expanded Abstracts, Society of Exploration Geophysicists, 73nd Annual International Meeting*, pages 2132–2135, 2003.

23. C. Stolk and W. Symes. Artifacts in Kirchhoff common image gathers. *Expanded Abstracts, Society of Exploration Geophysicists, 72nd Annual International Meeting*, pages 1129–1132, 2002.

24. C. C. Stolk and W. W. Symes. Smooth objective functionals for seismic velocity inversion. *Inverse Problems*, 19:73–89, 2003.

25. Christiaan C. Stolk and Maarten V. De Hoop. Seismic inverse scattering in the 'wave-equation' approach. Preprint 2001-047, The Mathematical Sciences Research Institute, http://msri.org/publications/preprints/2001.html, December 2001.

26. W. W. Symes. C.i.m.e. shot course on seismic imaging. The Rice Inversion Project, http://www.trip.caam.rice.edu/txt/tripinfo/other_list.html, September 2002.

27. A. Tarantola. *Inverse Problem Theory*. Elsevier, 1987.

28. A. Tarantola and B. Vallette. Inverse problems: quest for information. *J. Geophysics*, 50:159–170, 1982.

29. W. Wu, L. R. Lines, and H. Lu. Analysis of higher-order finite-difference schemes in 3-d reverse-time migration. *Geophysics*, pages 854–856, 1996.