Full waveform inversion with image-guided gradient

Yong Ma

— Master of Science Thesis —
Geophysics

Defended May 12, 2010

Committee Chair: Dr. André Revil
Advisor: Dr. Dave Hale

Committee members:
Dr. Paul Sava

Abstract

The objective of seismic full waveform inversion (FWI) is to estimate a model of the subsurface that minimizes the difference between recorded seismic data and synthetic data simulated for that model. Although FWI can yield accurate and high-resolution models, multiple problems have prevented widespread application of this technique in practice. First, FWI is computationally intensive, in part because it typically requires many iterations of costly gradient-descent calculations to converge to a solution model. Second, FWI often converges to spurious local minima in the data misfit function of the difference between recorded and synthetic data. Third, FWI is an underdetermined inverse problem with many solutions, most of which may make no geological sense. These problems are related to a typically large number of model parameters and to the absence of low frequencies in recorded data.

FWI with an image-guided gradient mitigates these problems by reducing the number of parameters in the subsurface model. We represent the subsurface model with a sparse set of values, and from these values, we use image-guided interpolation (IGI) to compute finely- and uniformly-sampled gradients of the data misfit function in FWI. Because the interpolation is guided by seismic images, gradients computed in this way conform to geologic structures and subsequently yield models that also agree with subsurface structures. Because models are parameterized sparsely, IGI makes the models more blocky than finely-sampled models, and this blockiness from the model space mitigates the absence of low frequencies in recorded data. A smaller number of parameters to invert also reduces the number of iterations required to converge to a solution model. Tests with a synthetic model and data demonstrate these improvements.

Table of Contents

Acknowledgments

Chapter 1

Introduction

The ultimate goal of exploration seismology is to estimate quantitatively accurate models **m** of the subsurface from measured seismic data **d**. In general, this requires solving an inverse problem governed by a forward operator **F** (see Figure 1.1). Seismic migration is used to extract and locate reflectivities of the subsurface by producing a structural image (Claerbout, 1985). However, a structural image alone cannot supply sufficient information to fully interpret properties of the model of the subsurface. Among many reasons for this insufficiency, the most obvious and intrinsic one is that seismic migration behaves as an adjoint operator rather than the inverse operator. In other words, a migrated image only approximates the solution of the inverse problem.

Seismic migration has evolved from producing a simple structural image of the subsurface to rendering an image with correct amplitudes and other attributes that can describe the interior of the earth. For this to happen, seismic depth migration depends more on velocity information than do other seismic processing steps (Etgen *et al.*, 2009), because the accuracy of seismic depth imaging depends greatly on the accuracy of the velocity model, especially in complex media. For example, the estimation of an accurate velocity model with high resolution in and below a salt body is essential for subsalt imaging(Oezsen, 2004; Wang *et al.*, 2008; Jiao *et al.*, 2008). High resolution velocity models also benefit seismic pore pressure prediction (Batzle & Wang, 1992), which helps geoscientists improve understanding of hydrocarbon reservoirs, identify potential drilling hazards, and improve well positioning.

1.1   Methods for Estimating Velocity Models

In the exploration and production industry, several methodologies have been developed to build velocity models of the earth. These methodologies can be mainly divided into three categories: traveltime tomography (Stork, 1992; Woodward, 1992; Vasco & Majer, 1993; Zelt & Barton, 1998), migration velocity analysis (MVA) (Yilmaz & Chambers,
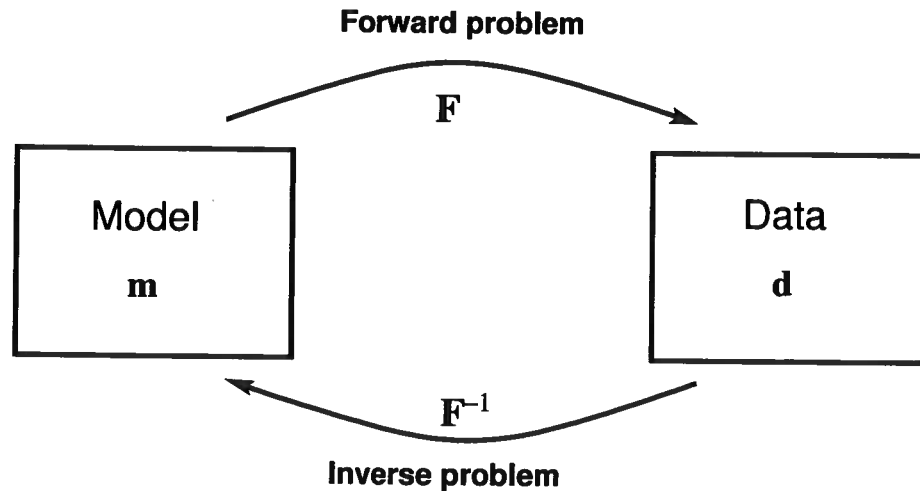
Forward problem

Figure 1.1: Schematic of the relationship between the model space $\mathbf{m}$ and the data space $\mathbf{d}$. Simulating data in a model with a forward operator $\mathbf{F}$ illustrates a forward problem $\mathbf{d} = \mathbf{F}(\mathbf{m})$. Retrieving model parameters or properties from recorded data with an inverse operator $\mathbf{F}^{-1}$ is a corresponding inverse problem, $\mathbf{m} = \mathbf{F}^{-1}(\mathbf{d})$. For generality, the operator $\mathbf{F}$ depends nonlinearly on model $\mathbf{m}$.

1984; Sava & Biondi, 2004a,b), and full waveform inversion (FWI) (Tarantola, 1984; Pratt *et al.*, 1998; Pratt, 1999; Symes, 2008).

1.1.1   Traveltime Tomography

Tomography that uses data collected on a surface to estimate models was developed for medical imaging (such as X-ray computerized tomography) before being performed using seismic data. A tomographic method used with success in exploration seismology begins with extraction of refraction or reflection traveltime from large sets of recorded seismic data. An inverse problem is then posed for which the measured data are these traveltimes. In seismic traveltime tomography, the objective is to match traveltimes obtained from recorded seismic data and corresponding synthetic seismic data simulated in a trial model of the subsurface.

1.1.2   Migration Velocity Analysis

Migration velocity analysis (MVA) is the process of estimating velocity model in the image domain by iteratively performing the following three steps:

(i) migrate data with the current estimate of velocity model;

(ii) measure kinematic errors in the prestack images;

(iii) invert measured kinematic errors to obtain velocity updates by a tomographic process.

To measure kinematic errors caused by velocity errors, MVA exploits data redundancy in common image gathers (CIGs) by measuring coherency across images obtained for different offsets, or for different reflection angles. The objective in MVA is to flatten these CIGs. Wave-equation migration velocity analysis (WEMVA) (Sava & Biondi, 2004a,b) goes beyond flattening CIGs; i.e., WEMVA measures image focusing to estimate velocity errors. To do so, migrated images produced by extended imaging conditions are needed. WEMVA works better than MVA in complex media because WEMVA combines both kinematics (through wave propagation) and data attributes (through image focusing). In contrast, MVA employs only the kinematic information. Symes (2008) further states that MVA is an approximate special case of waveform inversion.

Both traveltime tomography and MVA are able to estimate velocity models of the subsurface in large-scale (low-resolution), but neither of them can estimate velocity variation in small-scale (high-resolution).

1.1.3   Full Waveform Inversion

With increasing computing power, seismic full waveform inversion (FWI) (Tarantola, 1984; Pratt *et al.*, 1998; Pratt, 1999; Symes, 2008) has become an increasingly practical tool for estimating subsurface parameters, which is the ultimate goal in exploration seismology. FWI iteratively updates an estimated subsurface model and computes corresponding synthetic data to reduce the difference (the data misfit) between the synthetic and recorded data. The objective of FWI is to match the synthetic data $\mathbf{F}(\mathbf{m})$ and recorded data $\mathbf{d}$ in a comprehensive way, such that all information in waveforms (e.g., traveltimes, amplitudes, converted waves, multiples, etc.) is accounted for in the data misfit $E(\mathbf{m})$. An often used form of $E(\mathbf{m})$ is a least-squares function $E(\mathbf{m}) = \frac{1}{2}\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|^2$.

The FWI technique is attractive in its capability to estimate a subsurface model with generally higher resolution (Operto *et al.*, 2004) than does traveltime tomography (Stork, 1992; Woodward, 1992; Vasco & Majer, 1993; Zelt & Barton, 1998) and migration velocity analysis (MVA) (Yilmaz & Chambers, 1984; Sava & Biondi, 2004a,b). Another advantage of FWI over traveltime tomography or MVA is that FWI can estimate multiple parameters (e.g., velocity, density, attenuation, etc.). In practice, a macromodel generated by traveltime tomography or MVA may serve as an initial model for FWI.

1.2   FWI challenges

Although FWI has a long history and definite benefits, two major obstacles have prevented its widespread application in exploration seismology.

1.2.1   High Computational Cost

One obstacle is computational cost. FWI requires huge amounts of simulations and reconstructions of seismic wavefields, and computational cost is proportional to the number of sources or the number of shots. For large 3D models and seismic data sets, this cost may be prohibitive. Therefore, various efforts from different perspectives have been expended to reduce computational cost. One such method is to apply phase-encoding techniques (Krebs *et al.*, 2009) that combine all shots together to form a simultaneous source. The computational cost of FWI using encoding techniques is thereby reduced by a factor roughly equal to the number of encoded shots divided by the number of recorded shots.

FWI also requires multiple iterations of gradient descent to minimize the data misfit (see Figure 1.2), and computational cost is proportional to the number of required iterations. To reduce this number, one may reduce the number of model parameters used to represent the subsurface model. To reduce the number of parameters, one can represent a finely-sampled model as some basis functions of a sparse set of parameters. Many different compression methods employed for this purpose, such as Fourier transform, wavelet transform, curvelet transform, etc., share the same principle of projecting a model into another sparse domain. Through the sparse representation, one discards unwanted or unresolvable details that could be present in a more finely sampled model. The wavelet transform is a representative technique used in inverse problems (Meng & Scales, 1996). However, such methods do not account for geological structures in the subsurface that may be apparent in seismic images, and so may yield models that are geologically unreasonable.

1.2.2   No Unique Solution

A second obstacle is that the inverse problem posed by FWI has no unique solution. This is first because FWI is a typical underdetermined problem. Many different models may yield synthetic data that match recorded data within a reasonable tolerance that accounts for uncertainties and inadequacies in both recorded data and the theory underlying computed synthetic data.

The presence of local minima in the data misfit $E(\mathbf{m})$ is another source of the nonuniqueness problem. The forward operator $\mathbf{F}$ is generally a nonlinear function of the
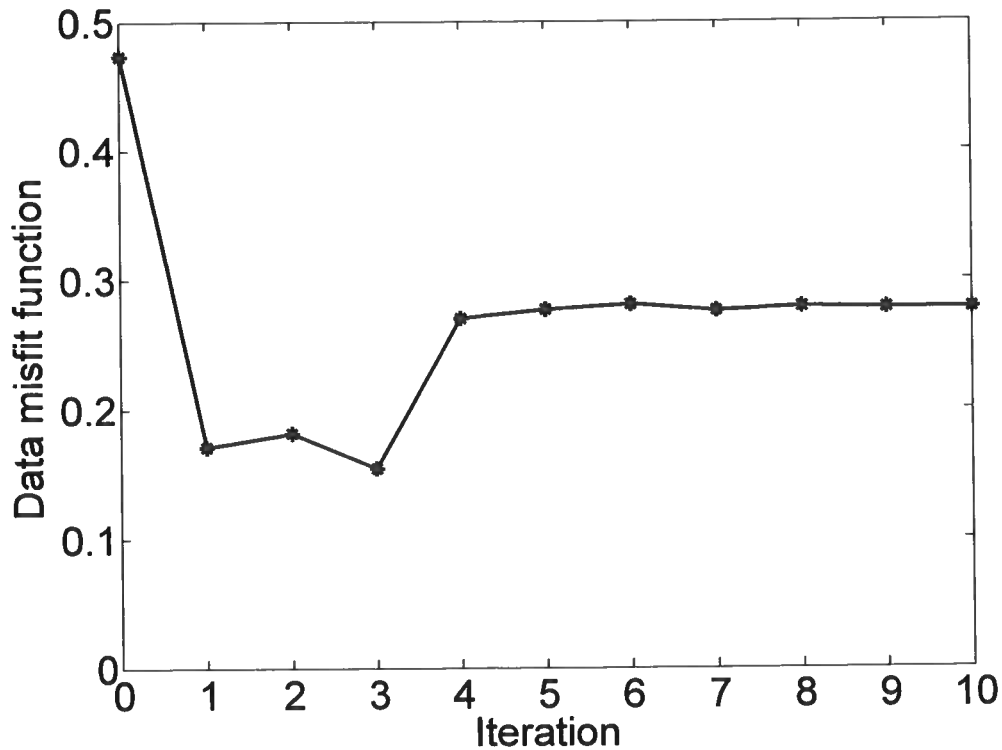
Figure 1.2: Illustration of many iterations required to convergence to a solution model.

model $\mathbf{m}$, thereby generating local minima in the data misfit function $\frac{1}{2}\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|^2$, as illustrated in Figure 1.3. Iterative inversion methods may converge, in the presence of local minima, to the global minimum if the initial model is close to the true model. If the initial model is too far away from the true model, iterative methods may converge to a local minimum. To eliminate local minima, one can linearize the inverse problem by linearizing the forward operator $\mathbf{F}$ (Tarantola, 1984). As a consequence, the data misfit function becomes $\frac{1}{2}\|\mathbf{Fm} - \mathbf{d}\|^2$, which is a purely quadratic function of $\mathbf{m}$ and contains only a global minimum (Snieder, 1998). However, this linearization approach is valid only if the initial model is in the vicinity of the true model.

Cycle-skipping also causes nonunique solutions to FWI. As shown in Figure 1.4, cycle-skipping occurs if the phase difference (time delay) between synthetic and recorded data is larger than half a period of the dominant wavelet. In practice, the cycle-skipping problem appears because it can be difficult to obtain an adequate initial model that is consistent with unrecorded low frequencies. In particular, low-wavenumber components of models are often poorly recovered by FWI because corresponding low-frequency content in data is
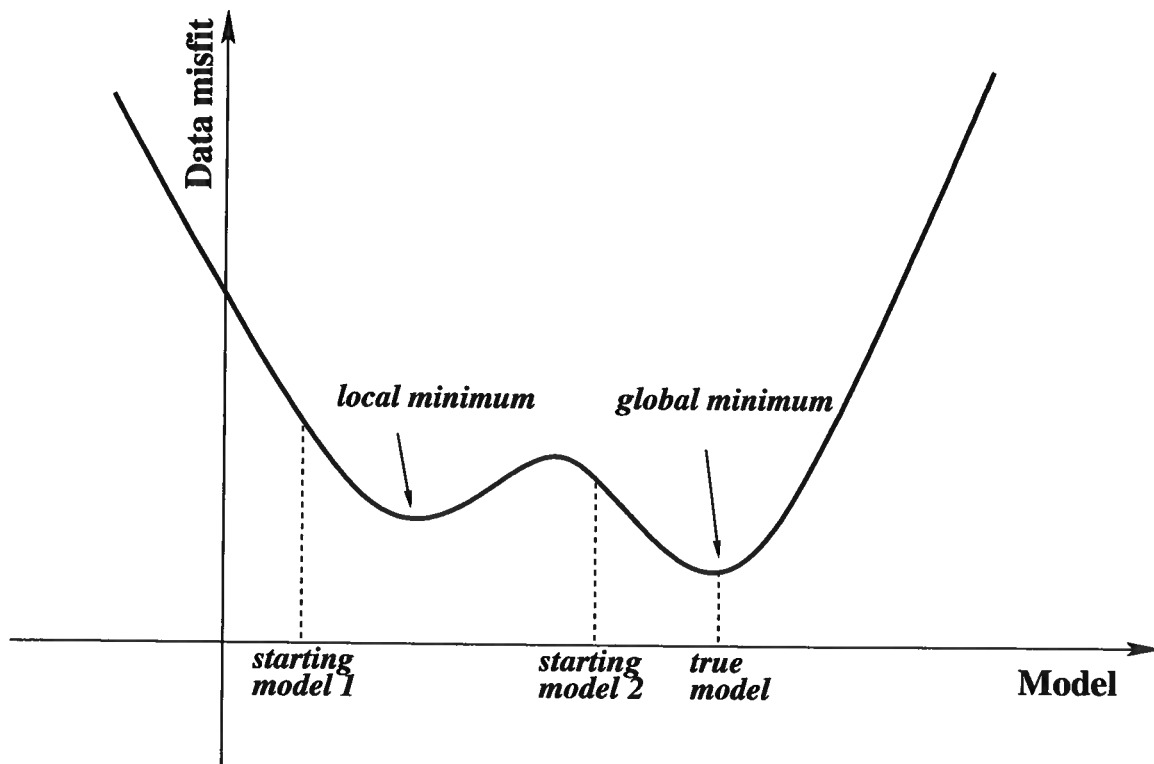
Figure 1.3: Schematic of the local-minima problem in FWI. The data misfit has spurious local minima because of the nonlinearity in forward modeling $\mathbf{F}(\mathbf{m}) = \mathbf{d}$. If FWI starts from a initial model (starting model 1) that is too far way from the true model, FWI may converge to a local minimum. If FWI starts from a initial model (starting model 2) that is close to the true model, FWI may converge to the global minimum.

rarely recorded.

Both local-minima and cycle-skipping problems lead to models that poorly approximate the subsurface. To mitigate such problems, multiscale approaches (Bunks, 1995; Sirgue & Pratt, 2004; Boonyasiriwat *et al.*, 2009) have been proposed. These methods recursively add higher-frequency details to models first computed from lower-frequency data. The fidelity of multiscale techniques depends fundamentally on the fidelity of low-frequency content in recorded data. In practice, the low frequencies required to bootstrap a multiscale FWI technique may be unavailable. Other methods for addressing these problems have been proposed as well. These include minimizing data misfit functions in logarithmic and Laplace domains (Shin & Min, 2006; Shin & Ha, 2008).

To obtain better subsurface models, *a priori* information is useful to make an inverse problem well determined. The *a priori* knowledge can take different forms. For example, both geological and geophysical data, such as those obtained from boreholes, may provide useful *a priori* constraints. Other useful constraints may be specified shapes and orientations of geologic structures in the subsurface. Meng (2009) uses estimated dips of reflectors as a constraint. Such *a priori* information constrains gradients and updates to models computed iteratively in FWI.

1.3   Thesis Overview

In this thesis, I develop image-guided FWI, a new technique for decreasing computational cost by reducing the number of iterations required for convergence. This image-guided FWI calculates and guides gradients using structural information derived from seismic images as the *a priori* constraints. Subsurface models computed from these image-guided gradients conform to geologic structures apparent in those images.

In **Chapter 2**, I first revisit basic concepts of FWI, analyze computational cost of FWI, and illustrate some of its practical problems with a synthetic example.

In **Chapter 3**, I first exploit the relationship between convergence and model parameters, then propose the inverse problem in the sparse domain with fewer model parameters, finally show how image-guided interpolation (Hale, 2009a) and its adjoint may be used to constrain the calculations of gradients in image-guided FWI. The same synthetic example used in Chapter 2 is used to illustrate the effectiveness of image-guided FWI.

Chapter 4 discusses the advantages of image-guided FWI.

Chapter 5 summarizes the thesis and discusses ideas for future investigation.
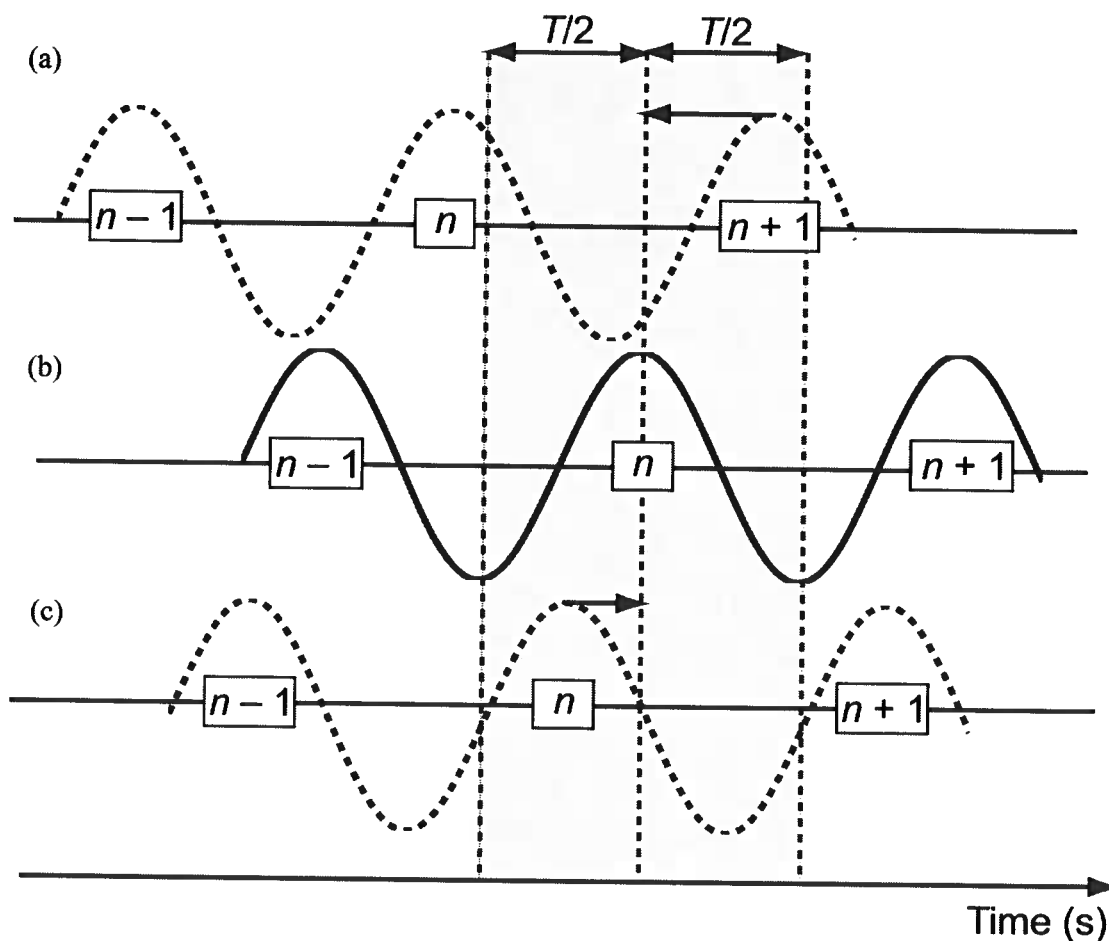
Figure 1.4: Two synthetic monochromatic seismograms with a period T in (a) and (c) are compared with the recorded seismogram in (b) to illustrate the cycle-skipping problem in FWI. The synthetic monochromatic seismogram in (a) has a time delay larger than half a period, and in the case FWI will update the model such that the $(n + 1)$th cycle in (a) matches the nth cycle in (b). Therefore, FWI produces an inaccurate model. The seismogram in (c) has a time delay smaller than half a period, and in the case FWI will update the model to make the nth cycle in (c) match the nth cycle in (b), leading to a correct model. Image courtesy of Virieux & Operto (2009).

Chapter 2

Full waveform inversion

2.1 Summary

This chapter first provides some background knowledge about full waveform inversion (FWI). I then analyze the composition of the computational cost of FWI, based on a typical implementation of FWI. Finally, I provide a synthetic FWI example and discuss some practical problems that could be encountered in FWI.

2.2 FWI as An Optimization Problem

Full waveform inversion (Tarantola, 2005) uses recorded seismic data $\mathbf{d}$ to estimate parameters of a subsurface model $\mathbf{m}$, given a forward operator $\mathbf{F}$ that synthesizes data. In FWI, we seek a model $\mathbf{m}$ that minimizes the difference $\mathbf{d} - \mathbf{F}(\mathbf{m})$. In seismic inversion, as for most geophysical inversion problems, the forward data-synthesizing operator $\mathbf{F}$ is a non-linear function of model parameters, such as seismic wave velocities.

Unfortunately, for almost any geophysical inverse problem, the forward operator $\mathbf{F}$ has no inverse $\mathbf{F}^{-1}$, so we cannot simply invert for the model from the data using $\mathbf{m} = \mathbf{F}^{-1}(\mathbf{d})$. Therefore, FWI is usually formulated as a least-squares optimization problem, in which we compute a model $\mathbf{m}$ that minimizes the data misfit function

$$E(\mathbf{m}) = \frac{1}{2}\|\mathbf{d} - \mathbf{F}(\mathbf{m})\|^2 , \qquad (2.1)$$

where $\|.\|$ denotes an L2 norm. All information in recorded seismic waveforms should, in principle, be taken into account in the data misfit function. Therefore, FWI comprehensively minimizes the difference in traveltimes, amplitudes, converted waves, multiples, etc. between recorded and synthetic data. This all-or-nothing approach distinguishes FWI from other methods, such as traveltime tomography, which focuses only on traveltime differences. Monte Carlo (random) methods (Nocedal & Wright, 2000; Tarantola, 2005) test randomly generated models to find one that minimizes the data misfit function $E(\mathbf{m})$. However, the

typically large number of model parameters makes such Monte Carlo methods impractical.

2.2.1   Newton-like Method

Gradient descent is a more practical alternative to a random search. We begin with an initial model $\mathbf{m}_0$, which can be found using other inversion methods (e.g., traveltime tomography or migration velocity analysis); then we use the gradient of the data misfit function $\mathbf{g} \equiv \boldsymbol{\nabla}_\mathbf{m} E = \dfrac{\partial E}{\partial \mathbf{m}}$ evaluated at $\mathbf{m}_0$ to search locally for a model $\mathbf{m} = \mathbf{m}_0 + \delta\mathbf{m}$ that reduces the data misfit $E\left(\mathbf{m}\right)$.

The Taylor series expansion of equation 2.1 about the initial model is

$$E\left(\mathbf{m}_0 + \delta\mathbf{m}\right) = E\left(\mathbf{m}_0\right) + \delta\mathbf{m}^T \mathbf{g}_0$$
$$+ \frac{1}{2}\delta\mathbf{m}^T \mathbf{H}_0 \delta\mathbf{m} + \dots , \tag{2.2}$$

where $E\left(\mathbf{m}_0\right)$ denotes the data misfit evaluated at $\mathbf{m}_0$, $\mathbf{g}_0 = \mathbf{g}\left(\mathbf{m}_0\right)$ and $\mathbf{H}_0 = \mathbf{H}\left(\mathbf{m}_0\right)$ denotes the Hessian matrix comprised of the 2nd partial derivatives of $E\left(\mathbf{m}\right)$, again evaluated at $\mathbf{m}_0$. If we ignore any term higher than the 2nd order in equation 2.2, this Taylor approximation is quadratic in the model perturbation $\delta\mathbf{m}$, and we can minimize the data misfit $E\left(\mathbf{m}\right)$ by solving a set of linear equations:

$$\mathbf{H}_0 \delta\mathbf{m} = -\mathbf{g}_0 \tag{2.3}$$

with a solution

$$\delta\mathbf{m} = -\mathbf{H}_0^{-1}\mathbf{g}_0 . \tag{2.4}$$

In Newton's method for minimization of the data misfit $E\left(\mathbf{m}\right)$, we begin with the initial model $\mathbf{m}_0$ and solve iteratively for

$$\delta\mathbf{m}_i = -\mathbf{H}_i^{-1}\mathbf{g}_i , \tag{2.5}$$

and

$$\mathbf{m}_{i+1} = \mathbf{m}_i - \mathbf{H}_i^{-1}\mathbf{g}_i , \tag{2.6}$$

where $\mathbf{g}_i \equiv \mathbf{g}\left(\mathbf{m}_i\right)$ and $\mathbf{H}_i$ is the Hessian matrix for the model $\mathbf{m}_i$. If we neglect nonlinearity (e.g., multiple scattering) in the forward operator $\mathbf{F}$, we obtain a Gauss-Newton method (Pratt *et al.*, 1998). However, in practice, the large size of the Hessian matrix $\mathbf{H}_i$, which depends on the number of parameters in the model, prevents the application of Newton-like

methods.

2.2.2 Iterative Method

Alternatively, the model update in equation 2.6 can be iteratively approximated by replacing the inverse of the Hessian matrix with a scalar step length α_i:

$$\mathbf{m}_{i+1} = \mathbf{m}_i - \alpha_i \mathbf{h}_i \ , \tag{2.7}$$

where the search direction $\mathbf{h}_i$ is determined by conjugate gradients (Vigh & Starr, 2008):

$$\mathbf{h}_0 = \mathbf{g}_0 \ ,$$
$$\beta_i = \frac{\mathbf{g}_i^T \left(\mathbf{g}_i - \mathbf{g}_{i-1} \right)}{\mathbf{g}_{i-1}^T \mathbf{g}_{i-1}} \ ,$$
$$\mathbf{h}_i = \mathbf{g}_i + \beta_i \mathbf{h}_{i-1} \ . \tag{2.8}$$

In each iteration, we compute the step length α_i using a quadratic line search algorithm (Nocedal & Wright, 2000).

2.3 Implementation of FWI

2.3.1 A Four-step Procedure

A gradient-descent implementation of FWI consists of four steps performed iteratively, beginning with an initial model $\mathbf{m}_0$:

(i) Compute $\mathbf{d} - \mathbf{F}(\mathbf{m}_i)$, the difference between recorded data $\mathbf{d}$ and synthetic data $\mathbf{F}(\mathbf{m}_i)$ computed for the current model $\mathbf{m}_i$;

(ii) Compute the gradient $\mathbf{g}_i = \nabla_{\mathbf{m}} E_i$;

(iii) Search for a step length α_i in the conjugate direction $\mathbf{h}_i$;

(iv) Compute the updated model $\mathbf{m}_{i+1}$ using equation 2.7.

2.3.2 Time Domain versus Frequency Domain

This gradient-descent version of FWI can be implemented both in the time domain (Tarantola, 1984, 1986; Mora, 1989) and in the frequency domain (Pratt, 1999). Perhaps

the greatest benefit of using frequency domain FWI is that we can select only a few frequencies for inversion (Sirgue & Pratt, 2004). Unfortunately, this advantage does not extend to inversion for deep subsurface models that require more frequencies. The advantages of implementing FWI in the time domain, as noted by Vigh & Starr (2008), include increased parallelism and reduced memory requirements, thereby making FWI more applicable to large 3D models and data sets. In the examples shown in this report, I use reverse time migration (RTM) and implement FWI in the time domain.

2.4 Analysis of Computational Cost

Most of the computational cost of FWI in this gradient-descent based implementation lies in steps (ii) and (iii), and meanwhile the computational cost in steps (i) or (iv) is neglectable.

2.4.1 Calculation of the Gradient

Although the gradient of the data misfit function $\mathbf{g} \equiv \nabla_{\mathbf{m}} E$ can be intuitively obtained by taking the first derivative of the data misfit function $E(\mathbf{m})$ with respect to each element in $\mathbf{m}$, the gradient calculated in this way is extremely expensive, especially when the model $\mathbf{m}$ contains a large number of parameters. Alternatively, we can compute the gradient in the following way by first rewriting the data misfit function as:

$$E(\mathbf{m}) = \frac{1}{2} (\mathbf{d} - \mathbf{F}(\mathbf{m}))^T (\mathbf{d} - \mathbf{F}(\mathbf{m})) \ . \tag{2.9}$$

Subsequently, the gradient should be

$$\mathbf{g} = \frac{\partial E(\mathbf{m})}{\partial \mathbf{m}} = \frac{\partial \mathbf{F}(\mathbf{m})}{\partial \mathbf{m}} (\mathbf{F}(\mathbf{m}) - \mathbf{d}) = \left(\frac{\partial \mathbf{F}(\mathbf{m})}{\partial \mathbf{m}} \right)^T \delta \mathbf{d} \ , \tag{2.10}$$

where $\delta \mathbf{d} = \mathbf{F}(\mathbf{m}) - \mathbf{d}$ is the data residual.

Applying the Born approximation to the forward modeling operator $\mathbf{F}$, we have

$$\mathbf{F}(\mathbf{m} + \delta \mathbf{m}) \approx \mathbf{F}(\mathbf{m}) + \mathbf{G}(\mathbf{m}) \delta \mathbf{m} \ , \tag{2.11}$$

where $\mathbf{G}(\mathbf{m})$ describes the single-scattering of model perturbation $\delta \mathbf{m}$. The single-scattering operator $\mathbf{G}$ depends nonlinearly on the background velocity $\mathbf{m}$ and in this way $\mathbf{m}$ accounts for the kinematic of data. Therefore, we can substitute $\mathbf{G}$ for $\dfrac{\partial \mathbf{F}(\mathbf{m})}{\partial \mathbf{m}}$ in equation 2.10 and
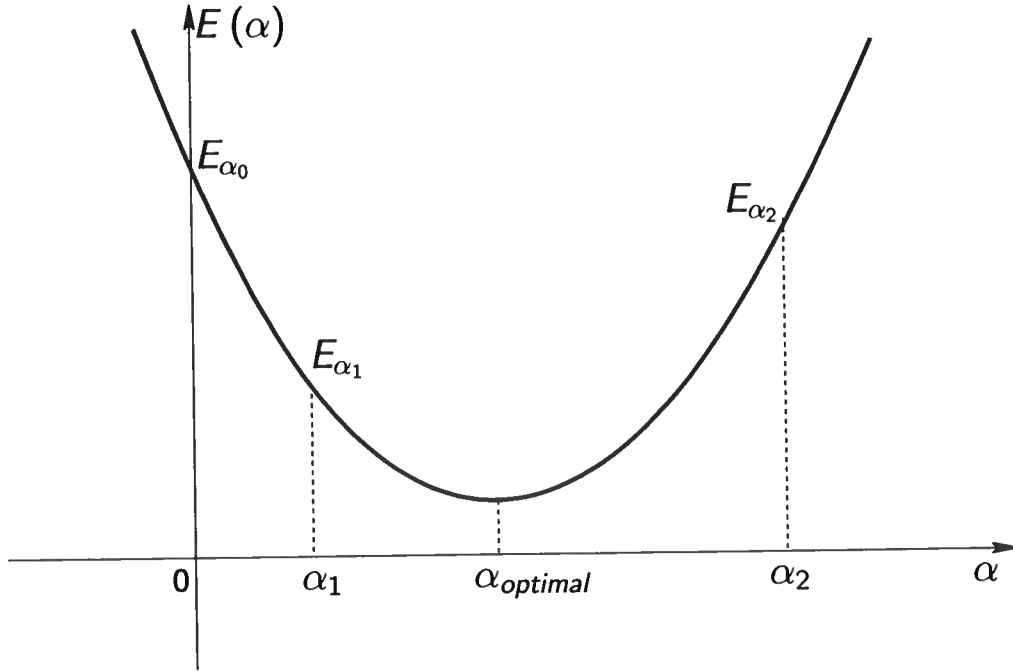
Figure 2.1: Schematic of a quadratic line-search method. Except for $(0, E_{\alpha_0})$, this quadratic search requires at least two trials of (step-length, misfit) pairs (e.g., (α_1, E_{α_1}) and (α_2, E_{α_2})) to determine an optimal step-length $\alpha_{optimal}$.

obtain the gradient as

$$\mathbf{g} = \mathbf{G}^T \delta \mathbf{d} \ . \tag{2.12}$$

The gradient $\mathbf{g}$ in equation 2.12 is a migrated image of the data residual $\delta \mathbf{d}$, because the adjoint $\mathbf{G}^T$ of the operator of $\mathbf{G}$ is migration. To compute such an image, we first backpropagate the data residual $\delta \mathbf{d}$ as a single simultaneous source in the background velocity model $\mathbf{m}$ and then correlate this backpropagated wavefield with a forward wavefield simulated in a model $\mathbf{m}$. The gradient computed in this way is equivalent to a result of prestack reverse time migration (RTM). This idea is introduced in Tarantola (1984) and later is also referred to as the adjoint-state method (Tromp $et\ al.$, 2005).

The major computational cost in RTM is spent in simulating two wavefields in a full model space for all the shots. We approximately write this part of cost in FWI as $2 N_s C_{\mathbf{F(m)}}$, where N_s is the number of shots and $C_{\mathbf{F(m)}}$ is the unit cost spent in simulating a wavefield for a single shot.

Components	Description	Reduction method
N_i	the number of iterations	speed up convergence
N_s	the number of shots	phase-encoding technique
$C_{\mathbf{F(m)}}$	cost per shot in simulating wavefields	faster forward modeling kernel

Table 2.1: Components that control the computational cost in FWI.

2.4.2   Cost of Line-search

Another part of the computational cost lies in searching for a step-length α to update the model. Figure 2.1 illustrates an example of quadratic line-search. At least two trials of step length are necessary in this quadratic search. For each trial, it requires simulating a wavefield in the whole model space for all the shots. Therefore, this part of cost is larger than or equal to $2N_s C_{\mathbf{F(m)}}$.

2.4.3   Total Cost in FWI

Because those four steps are performed iteratively, the total computational cost in FWI C_{FWI} is

$$C_{FWI} \propto N_i N_s C_{\mathbf{F(m)}} \, , \tag{2.13}$$

where N_i is the number of iterations required to converge to a solution model.

Table 2.1 summarizes these three components that directly control the computational cost of FWI. From the point of decreasing computational cost view, a variety of efforts can be expended to reduce N_i, N_s, and $C_{\mathbf{F(m)}}$, respectively. Among them, reducing the unit cost $C_{\mathbf{F(m)}}$ is a key ingredient, which involves developing a more efficient forward modeling algorithm. To reduce N_s, phase-encoding techniques have been recently employed in FWI. Furthermore, speeding up the convergence of FWI is helpful to reduce the number of iterations N_i, thereby reducing the total cost.

2.5   Synthetic Example

Figure 2.2(a) depicts a subsurface velocity model with two anomalies. One is a low-velocity zone and another is a high-velocity bar, as shown separately in Figure 2.2(c). We

refer to the model in Figure 2.2(a) as the true model $\mathbf{m}$. Figure 2.2(b) displays the initial model $\mathbf{m}_0$ that we used in FWI; it is simply the true model $\mathbf{m}$ without the two anomalies.

To test FWI, we first create data $\mathbf{d} = \mathbf{F}(\mathbf{m})$ using the true model $\mathbf{m}$. Henceforth, for consistency with the discussion above, we refer to these data as the "recorded" data, even though we compute these noise-free data using the forward operator $\mathbf{F}$, a finite-difference constant-density solution to the 2D acoustic wave equation. A total of 25 shots are evenly distributed on the top surface with an interval of 120 m; the receiver interval is 10 m. The source is a Ricker wavelet with a peak frequency of 15 Hz. For example, Figure 2.3(a) shows a common-shot gather for shot number 13 of the recorded data $\mathbf{d}$. Figure 2.3(b) shows the corresponding synthetic data $\mathbf{F}(\mathbf{m}_0)$ computed for the initial model $\mathbf{m}_0$ displayed in Figure 2.2(b). Figure 2.3(c) displays the difference $\mathbf{d} - \mathbf{F}(\mathbf{m}_0)$, which is also known as the data residual, that part of the recorded data that cannot be explained by the current model. In the four steps of FWI, computation of this data residual is step (i).

In step (ii), we compute the gradient of the data misfit. As discussed by (Tarantola & Valette, 1982; Pratt, 1999), this gradient is equal to the output of RTM applied to the data residual shown in Figure 2.3(c), using the current model $\mathbf{m}_0$ shown in Figure 2.2(b). This method for the calculation of gradient is also referred to as the adjoint-state method (Tromp et al., 2005). Figure 2.4(a) shows the gradient $\mathbf{g}_0$ computed in this way for the first iteration of FWI.

In step (iii), we then compute a step length α_0 that determines how much to change our velocity model in this first iteration. We compute the step length using a quadratic line search algorithm and search in a direction defined by conjugate gradients (Gong et al., 2008). This line search requires computation of at least two synthetic data sets.

Finally, in step (iv), we update the current velocity model according to equation 2.7. Figure 2.5(a) is the change $\delta\mathbf{m}_0$ in velocity computed in the 1st iteration; in this 1st iteration, this change is simply a scaled version of the gradient computed in step (ii).

In subsequent iterations, the iterative four-step FWI process introduces additional details, as indicated by the gradients displayed in Figure 2.4(b) and (c), which correspond to the 2nd and 5th iterations, respectively. Figure 2.5(b) and (c) show the corresponding accumulated velocity updates, the difference between the current and initial velocity models.
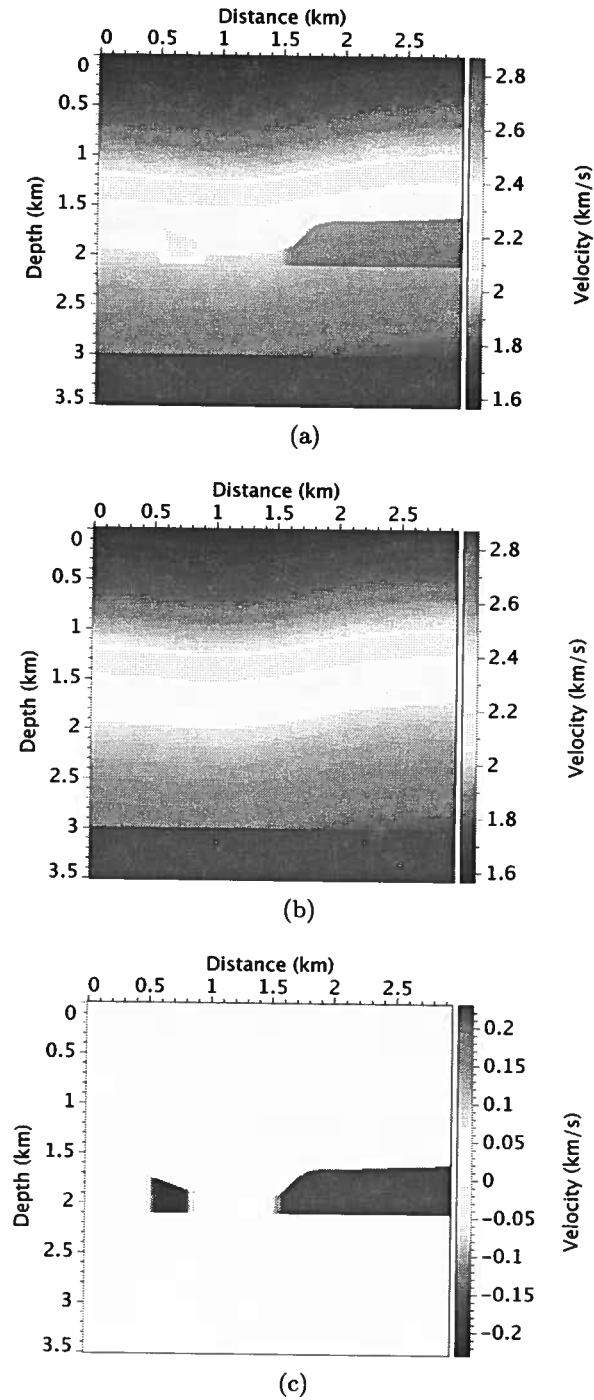
(a)

(b)

(c)

Figure 2.2: (a) The LVZ model courtesy of ConocoPhillips. (b) The initial velocity model in FWI. (c) Velocity anomalies (one low velocity zone and one high velocity bar) created by subtracting the initial model in (b) from the true model in (a).
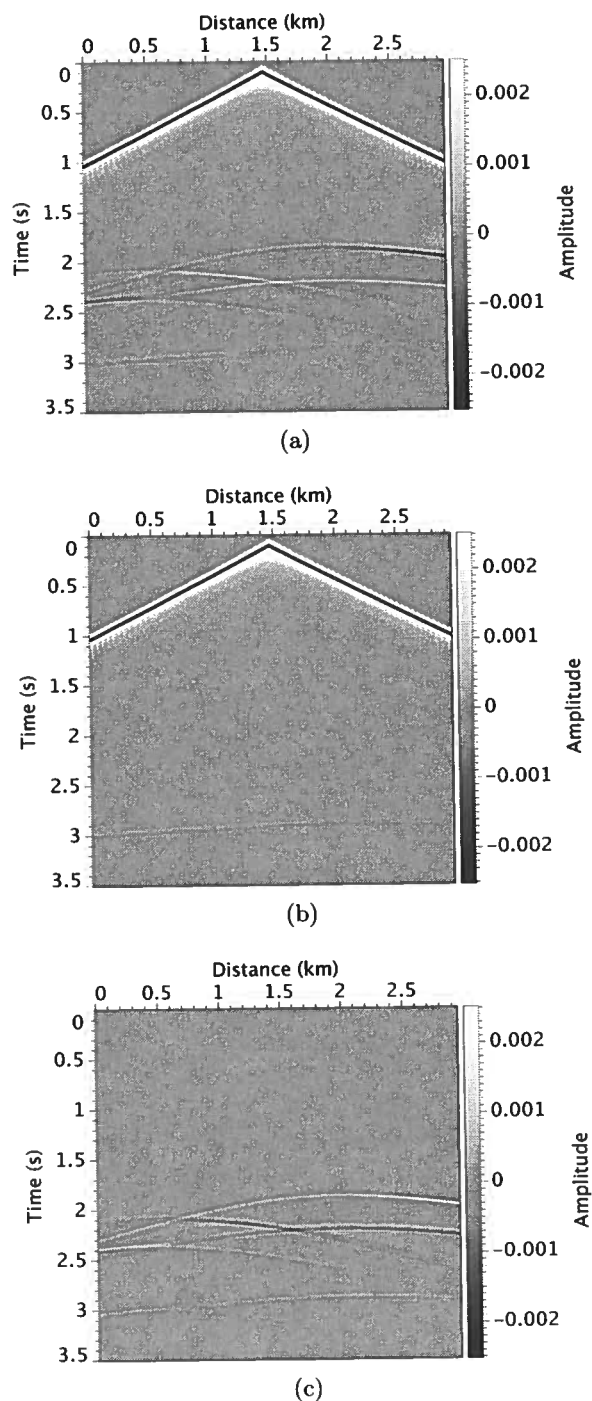
(a)

(b)

(c)

Figure 2.3: (a) The common-shot gather of shot number 13 in the recorded data set. (b) The corresponding synthetic common-shot gather simulated in the initial velocity model (Figure 2.2(b)). (c) The data residual for this shot. Intuitively, FWI needs to recover the velocity anomalies in Figure 2.2(c), starting from an initial model in Figure 2.2(b) and the data residual in (c).
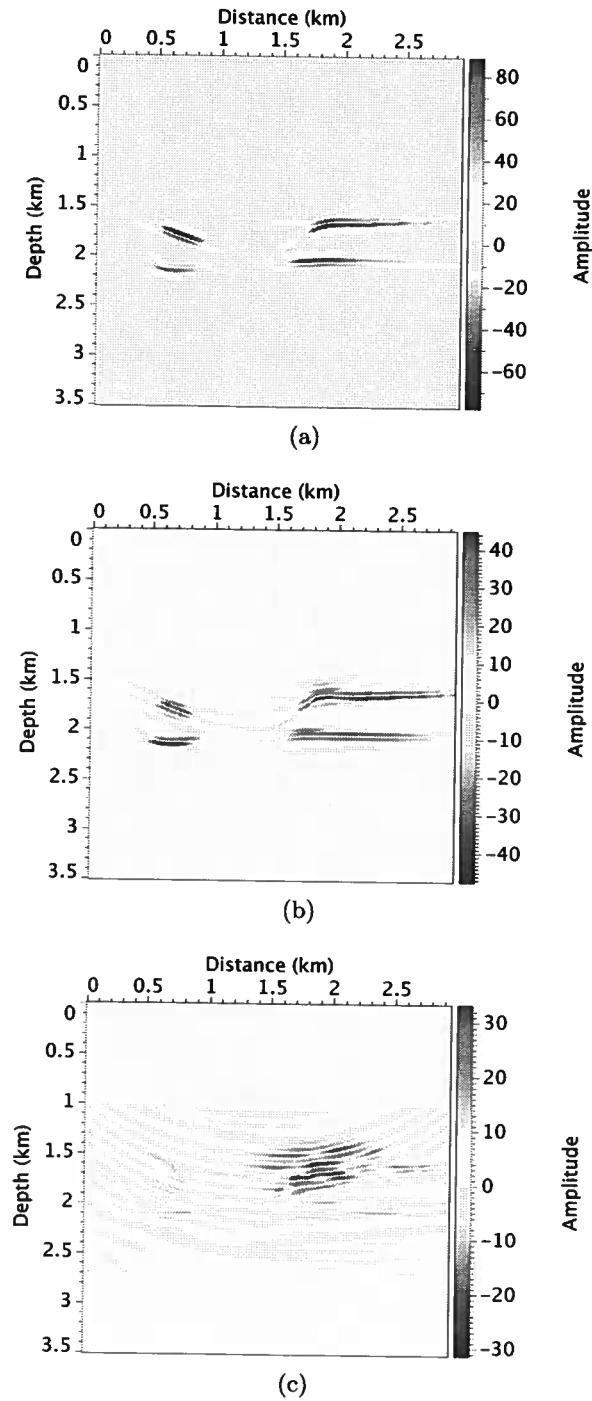
Figure 2.4: Gradient of the data misfit function in (a) the first iteration, (b) the second iteration and (c) the fifth iteration. The gradient in each iteration is the result of RTM applied to the current data residual, using the current velocity model.
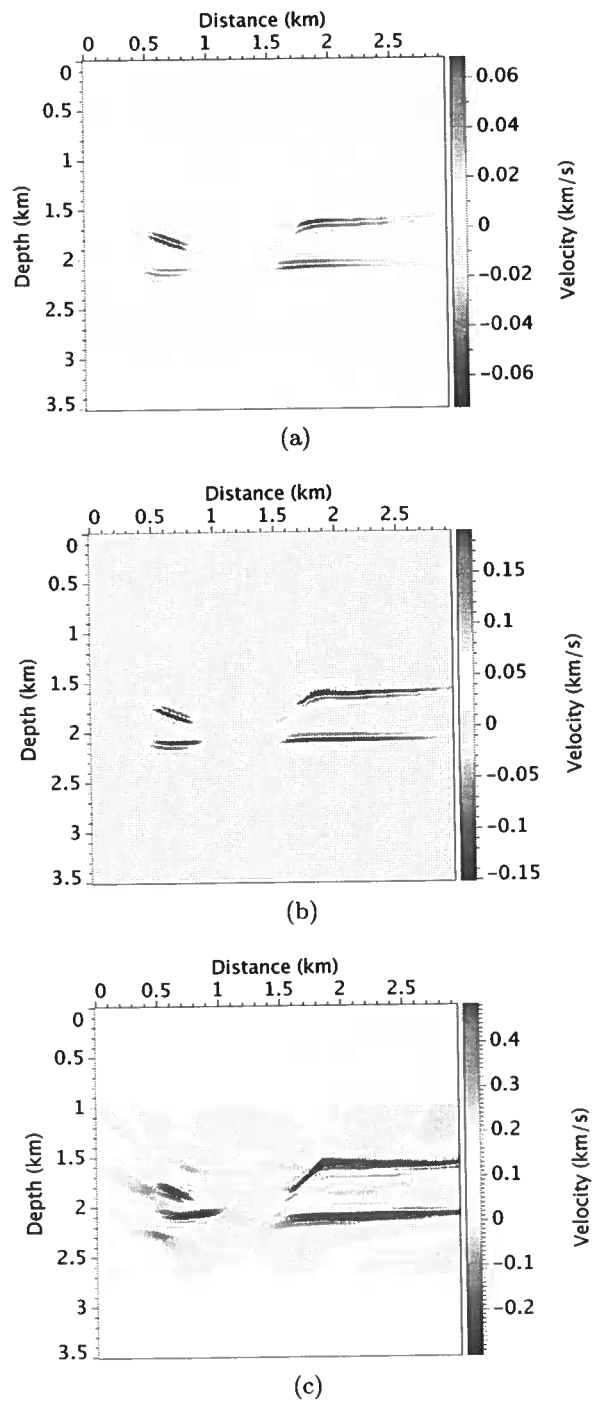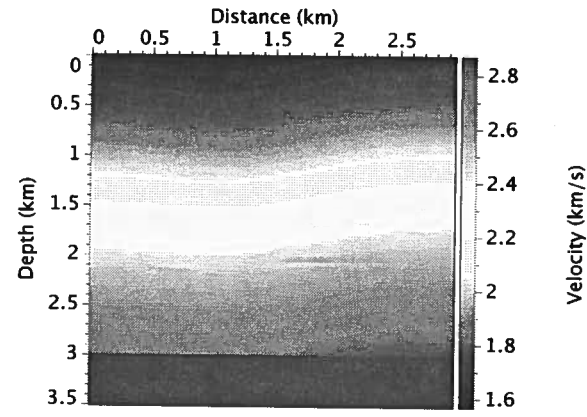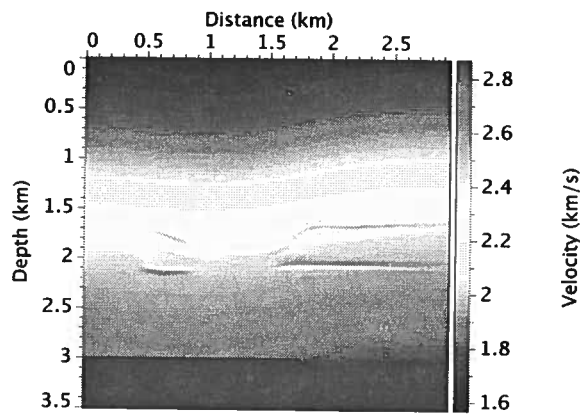
Figure 2.5: Accumulated velocity updates after (a) 1 iteration, (b) 2 iterations and (c) 5 iterations.

(a)

(b)

(c)
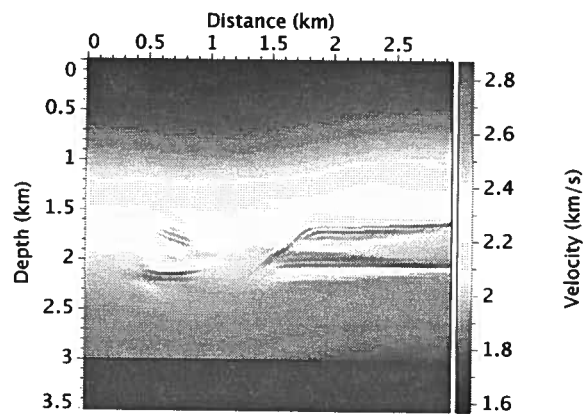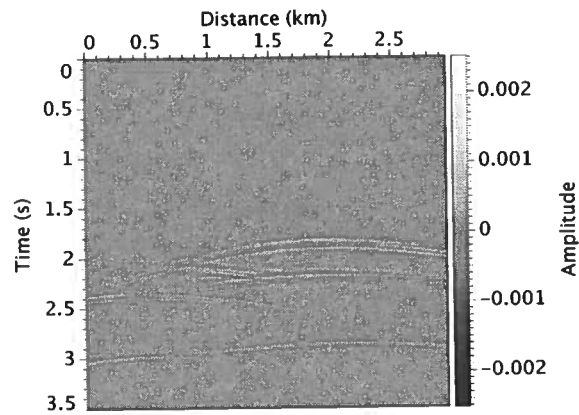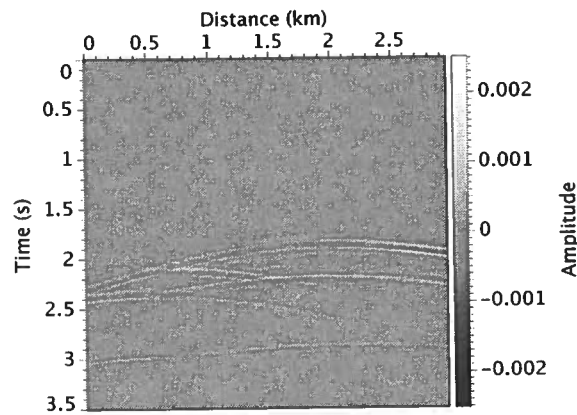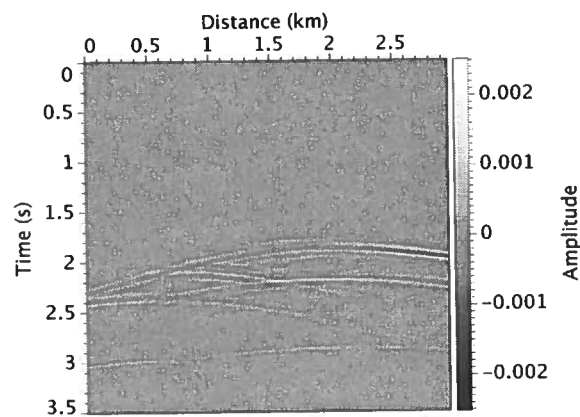
Figure 2.6: Updated velocity models after (a) 1 iteration, (b) 2 iterations and (c) 5 iterations.

Figure 2.7: Data residuals after (a) 1 iteration, (b) 2 iterations and (c) 5 iterations. These residuals are differences between data simulated in models shown Figure 2.6 and the recorded data shown in Figure 2.3(a), respectively.
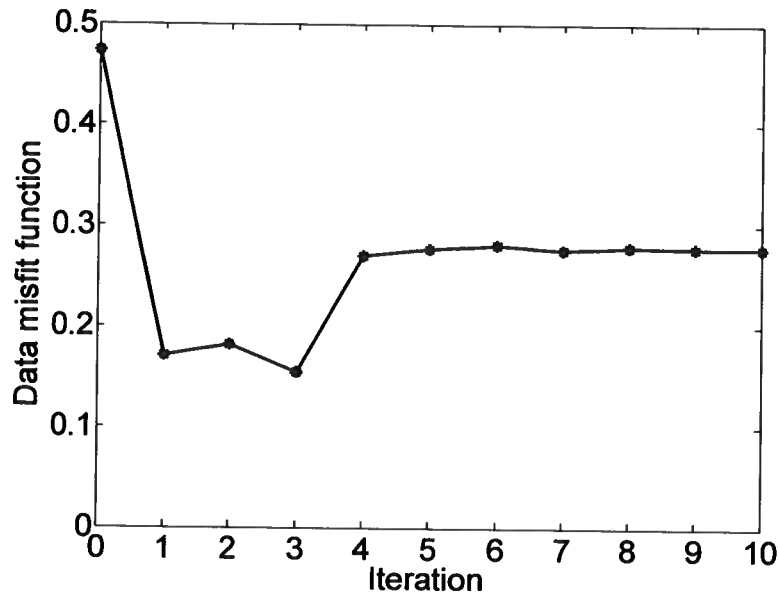
Figure 2.8: Variation of data misfit function versus iterations in FWI. The first 10 iterations can be summarized in this figure. Note the up-and-down changes for iteration< 5 and the flatness for iteration > 5.

2.6   Problems in FWI

Figure 2.6 depicts velocity models that take FWI updates into account after 1, 2 and 5 iterations, respectively, and Figure 2.7 shows the corresponding data residuals. By looking at Figure 2.6 and Figure 2.7, one may have a first impression that FWI fails to converge to a satisfactory result, at least within the first 5 iterations.

2.6.1   Slow Convergence

After the 1st iteration, the data residual corresponding to shot number 13, as shown in Figure 2.7(a), becomes significantly smaller than that in Figure 2.3(c). However, in subsequent iterations, the data residuals shown in Figure 2.7(b) and (c) increase. This nonmonotonic change of data residual in fact slows the convergence of FWI.

In principle, each iteration of FWI should reduce the data misfit $E(\mathbf{m})$, but in the search for a step length α_i, FWI risks producing unsatisfactory models with larger data residuals. Figure 2.8 plots the data misfit function $E(\mathbf{m})$ as a function of the number of iteration. For example, the data residual after the 2nd iteration of FWI is even larger than the residual of the 1st iteration; a similar case occurs in the 4th iteration. Such nonmonotonic relationship between $E(\mathbf{m})$ and the iteration number is mainly caused by

two reasons in the line search.

First, as noted in earlier section, the quadratic line search requires at least two trials of step-length. Unfortunately, within a limited number (e.g., 5 in this paper) of trials of gradient descent, FWI sometimes fails to find a step-length α_i that decreases the data misfit function $E(\mathbf{m})$. In this case, FWI cannot be simply terminated. In order to continue FWI, I keep the step-length that gives the smallest data misfit function $E(\mathbf{m})$ in those five trials, and hope that FWI can reduce $E(\mathbf{m})$ in subsequent iterations. FWI in fact reduces the data residual in the 3rd iteration, but we confront another increase of the data residual in the 4th iteration.

Second, FWI updates the model in the conjugate direction $\mathbf{h}_i$ instead of the gradient direction $\mathbf{g}_i$, which guarantees the descent of the data misfit function. In contrast, the conjugate direction may lead to temporal increase of the data residual. The well-known iterative conjugate-gradient method, in fact, does not monotonically decrease the residual.

Another proof of slow convergence of FWI in this example is that FWI nearly cannot reduce the data residual any more after the 5th iteration. Figure 2.9 shows data residuals after FWI updates the model in the 6th, 8th and 10th iteration, respectively. The data residuals in Figure 2.9 vary little, and this corresponds to the flatness of the objective function curve shown in Figure 2.8. Because this example illustrates a reflection FWI, it recovers only high-wavenumber components but fails in recovering the low-/mid-wavenumber components in the velocity model. Therefore, after the recovery of high-wavenumber components, FWI eventually stops reducing the data misfit function.

2.6.2   Band-limited Data

Recall the peak frequency (15 Hz) of the Ricker wavelet in the above example. Clearly the data used for FWI is highly band-limited. In other words, low-frequency and high-frequency information cannot be recorded in the data. Particularly, the low frequencies in data are crucial in FWI, because the absent low frequencies help resolve the intermediate- and low-wavenumber (large-scale) components of the subsurface. Due to the absence of low frequencies, FWI updates the velocity model only at locations of reflecting interfaces, as shown in the previous example (see Figure 2.5).

For the same reason, the accumulated velocity updates produced by FWI, as we observe from Figure 2.5, contain significant imprints of the seismic wavelet. Those wavelet imprints thereby make those velocity updates look more like migrated images rather than any reasonable perturbations to the initial velocity model.
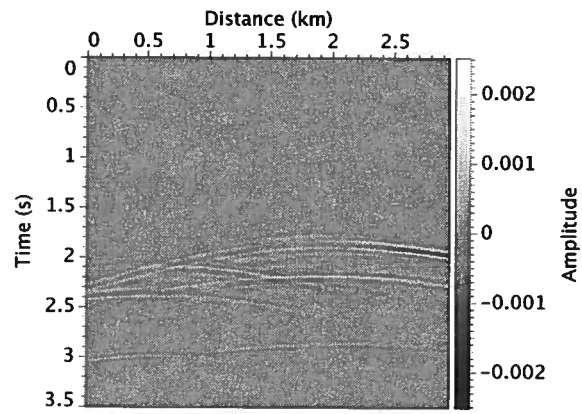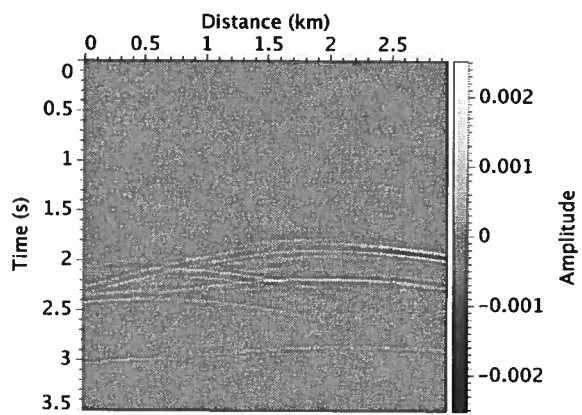
Figure 2.9: Data residual after (a) 6 iteration, (b) 8 iterations and (c) 10 iterations.

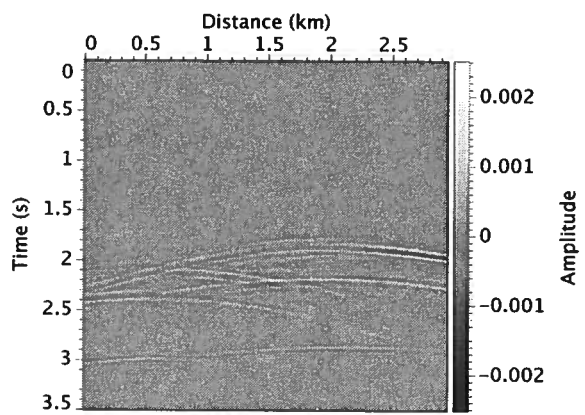# Chapter 3

Image-guided full waveform inversion

3.1   Summary

This chapter focuses on improving and speeding up convergence of FWI. I first explain the relationship between the convergence rate of FWI and model parameters. Then the inverse problem is posed with respect to fewer model parameters in the sparse domain. Through inverting fewer model parameters, FWI converges more quickly to a more accurate model. In particular, *image-guided interpolation* (Hale, 2009a) and its adjoint are employed to build *image-guided FWI* for a sparsely parameterized model.

3.2   Convergence Rate and Model Parameters

Conjugate-gradient methods are guaranteed to minimize the quadratic misfit function of a linear system within M iterations, where M is the number of model parameters in the solution vector $\mathbf{m}$ (Nocedal & Wright, 2000). Often the number of iterations required for convergence deceases with the number of model parameters to invert. More precisely, the convergence rate of a conjugate-gradient method depends on the condition number of the Hessian matrix $\mathbf{H}$ (Cohen, 1972; Wheeler & Wilton, 1988).

In practice, FWI is usually ill-posed due to a typically large condition number of the Hessian matrix. A large condition number tends to appear especially when an inverse problem has a large number of model parameters in $\mathbf{m}$. Since the Hessian matrix $\mathbf{H}$ measures the sensitivity of the data misfit function $E(\mathbf{m})$ with respect to model parameters, if the change of a model parameter in $\mathbf{m}$ cannot cause significant change in the data misfit function $E(\mathbf{m})$, the Hessian matrix $\mathbf{H}$ will have a small (or nearly zero) eigenvalue. As a consequence, the condition number of the Hessian matrix can be large enough such that the gradient-descent method used to solve a FWI problem converges slowly.

3.3   Fewer Model Parameters

Inspired by the intuitive relationship between the convergence rate and the number of model parameters, if we can pose a FWI problem that only needs to invert a few model parameters, to which the data misfit function is sensitive, we can reduce the condition number of the Hessian matrix and thereby the number of required iterations. Pratt et al. (1998, Appendix A) discuss a point collocation scheme to reparameterize the model space $\mathbf{m}$ for this purpose.

Similar to the point collocation scheme, subspace methods (Kennett *et al.*, 1988; Oldenburg *et al.*, 1993) reconstruct the finely- and uniformly-sampled (dense) model $\mathbf{m}$ from a sparse model $\mathbf{s}$ that contains a much smaller number of model parameters than does the dense model $\mathbf{m}$:

$$\mathbf{m} = \mathbf{Rs} \, , \tag{3.1}$$

where $\mathbf{R}$ denotes a linear operator that links model parameters in the sparse model and the dense model.

3.4   Inverse Problem in Sparse Domain

Differentiating both sides of equation 3.1, we have

$$\delta\mathbf{m} = \mathbf{R}\delta\mathbf{s} \, . \tag{3.2}$$

Then, substituting equation 3.2 into equation 2.5, we can reformulate the inverse problem posed in equation 2.5, with respect to a smaller number of model parameters in the sparse model $\mathbf{s}$, as

$$\mathbf{H}_i\mathbf{R}\delta\mathbf{s}_i = -\mathbf{g}_i \, . \tag{3.3}$$

Because $\mathbf{R}$ is not a square matrix, equation 3.3 is different from conventional preconditioning.

However, we cannot solve equation 3.3 with a solution like $\delta\mathbf{s}_i = -\left(\mathbf{H}_i\mathbf{R}\right)^{-1}\mathbf{g}_i$ in the sparse domain $\mathbf{s}$ because equation 3.3 is overdetermined; i.e., there are more equations than parameters. Alternatively, we obtain a solution for equation 3.3 in the sparse domain $\mathbf{s}$:

$$\delta\mathbf{s}_i = -\left(\mathbf{R}^T\mathbf{H}_i\mathbf{R}\right)^{-1}\mathbf{R}^T\mathbf{g}_i \, , \tag{3.4}$$

where $\mathbf{R}^T$ is the adjoint operator of $\mathbf{R}$. This adjoint operator projects model parameters from the dense model $\mathbf{m}$ to the sparse model $\mathbf{s}$.

Like equation 2.7, the model update $\delta \mathbf{s}$ can be iteratively approximated by replacing the inverse of the projected Hessian matrix $\left(\mathbf{R}^T \mathbf{H}_i \mathbf{R}\right)$ with a scalar step length α_i:

$$\mathbf{s}_{i+1} = \mathbf{s}_i - \alpha_i \mathbf{h}_i^{\mathbf{s}} \, , \tag{3.5}$$

where the conjugate direction $\mathbf{h}_i^{\mathbf{s}}$ is determined by

$$
\begin{aligned}
\mathbf{h}_0^{\mathbf{s}} &= \mathbf{R}^T \mathbf{g}_0 \, , \\
\beta_i &= \frac{\left(\mathbf{R}^T \mathbf{g}_i\right)^T \left(\mathbf{R}^T \mathbf{g}_i - \mathbf{R}^T \mathbf{g}_{i-1}\right)}{\left(\mathbf{R}^T \mathbf{g}_{i-1}\right)^T \mathbf{R}^T \mathbf{g}_{i-1}} \\
&= \frac{\mathbf{g}_i^T \mathbf{R} \mathbf{R}^T \left(\mathbf{g}_i - \mathbf{g}_{i-1}\right)}{\mathbf{g}_{i-1}^T \mathbf{R} \mathbf{R}^T \mathbf{g}_{i-1}} \, , \\
\mathbf{h}_i^{\mathbf{s}} &= \mathbf{R}^T \mathbf{g}_i + \beta_i \mathbf{h}_{i-1}^{\mathbf{s}} \, .
\end{aligned}
\tag{3.6}
$$

In equation 3.5, the step length can be again achieved with a quadratic line-search method. Equation 3.6 differs from equation 2.8 in that the gradient $\mathbf{g}_i$ is replaced by $\mathbf{R}^T \mathbf{g}_i$, which implies that equation 3.5 provides a solution for the FWI problem in the sparse domain $\mathbf{s}$. Because of fewer model parameters involved, the projected Hessian matrix $\left(\mathbf{R}^T \mathbf{H}_i \mathbf{R}\right)$ can become better-conditioned and thus equation 3.5 requires fewer iterations than equation 2.7 to converge to a solution model $\mathbf{s}$.

As noted in equation 3.1, we can apply the linear operator $\mathbf{R}$ to both sides of equation 3.5 and thereby project the sparse model update $\delta \mathbf{s}_i$ to obtain the dense model update $\delta \mathbf{m}_i$:

$$\mathbf{m}_{i+1} = \mathbf{m}_i - \alpha_i \mathbf{h}_i^{\mathbf{m}} \, , \tag{3.7}$$

where we compute the search direction $\mathbf{h}_i^{\mathbf{m}}$ by projecting the sparse conjugate direction $\mathbf{h}_i^{\mathbf{s}}$ to the dense domain:

$$\mathbf{h}_0^{\mathbf{m}} = \mathbf{R}\mathbf{h}_0^{\mathbf{s}} = \mathbf{R}\mathbf{R}^T\mathbf{g}_0 \, ,$$

$$\beta_i = \frac{\left(\mathbf{R}^T\mathbf{g}_i\right)^T \left(\mathbf{R}^T\mathbf{g}_i - \mathbf{R}^T\mathbf{g}_{i-1}\right)}{\left(\mathbf{R}^T\mathbf{g}_{i-1}\right)^T \mathbf{R}^T\mathbf{g}_{i-1}}$$

$$= \frac{\mathbf{g}_i^T\mathbf{R}\mathbf{R}^T \left(\mathbf{g}_i - \mathbf{g}_{i-1}\right)}{\mathbf{g}_{i-1}^T\mathbf{R}\mathbf{R}^T\mathbf{g}_{i-1}} \, ,$$

$$\mathbf{h}_i^{\mathbf{m}} = \mathbf{R}\mathbf{R}^T\mathbf{g}_i + \beta_i\mathbf{h}_{i-1}^{\mathbf{m}} \, . \qquad\qquad (3.8)$$

Equations 3.7 and 3.8 provide a solution $\mathbf{m}$ for FWI in the dense space with the advantage derived from the solution $\mathbf{s}$ in the sparse space.

3.5   Choice of R

The projection operator $\mathbf{R}$ can take different forms, including Fourier transform, wavelet transform, cubic splines, etc. Unfortunately, none of these forms accounts for the geological information of the subsurface. In this paper, we implement $\mathbf{R}$ with image-guided interpolation (IGI) (Hale, 2009a), which uses metric tensor fields to guide interpolation of a few sparsely scattered data points, making the interpolant conform to structural features in the gradient image $\mathbf{g}$.

3.5.1   Image-guided Interpolation

The input of IGI is a set of scattered data, a set

$$\mathcal{F} = \{f_1, f_2, ..., f_K\}$$

of K known sample values $f_k \in \mathbb{R}$ that correspond to a set

$$\chi = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_K\}$$

of K known sample points $\mathbf{x}_k \in \mathbb{R}^n$. Combining these two sets forms a space (e.g., the sparse model $\mathbf{s}$), in which $\mathcal{F}$ and χ denote sample values and coordinates, respectively. The result of the interpolation is a function $q\left(\mathbf{x}\right) : \mathbb{R}^n \rightarrow \mathbb{R}$, such that $q\left(\mathbf{x}_k\right) = f_k$. Here, the dense model $\mathbf{m}$ consists of uniformly sampled values $q\left(\mathbf{x}\right)$.

Image-guided interpolation includes the application of two linear operators:

$$\mathbf{R} = \mathbf{QP} \, , \tag{3.9}$$

where $\mathbf{P}$ and $\mathbf{Q}$ denote nearest neighbor interpolation and blended neighbor interpolation, respectively. We follow the steps in Hale (2009a) to describe details of $\mathbf{P}$ and $\mathbf{Q}$:

1. **P**: solve

$$\nabla t\left(\mathbf{x}\right) \cdot \mathbf{D}\left(\mathbf{x}\right) \nabla t\left(\mathbf{x}\right) = 1, \mathbf{x} \notin \chi \, ;$$
$$t\left(\mathbf{x}\right) = 0, \mathbf{x} \in \chi \tag{3.10}$$

   for

   $t\left(\mathbf{x}\right)$: the minimum time from $\mathbf{x}$ to the nearest known sample point $\mathbf{x}_k$, and

   $p\left(\mathbf{x}\right)$: the nearest neighbor interpolant corresponding to f_k, the value of the sample point $\mathbf{x}_k$ nearest to the point $\mathbf{x}$.

2. **Q**: for a specified constant $e \geq 2$ (e.g., $e = 4$ in this thesis), solve

$$q\left(\mathbf{x}\right) - \frac{1}{e}\nabla \cdot t^2\left(\mathbf{x}\right) \mathbf{D}\left(\mathbf{x}\right) \nabla q\left(\mathbf{x}\right) = p\left(\mathbf{x}\right) \tag{3.11}$$

   for the blended neighbor interpolant $q\left(\mathbf{x}\right)$.

In equation 3.10, the metric tensor field $\mathbf{D}\left(\mathbf{x}\right)$ (van Vliet & Verbeek, 1995; Fehmers & Höcker, 2003) represents structural features of the subsurface, such as structural orientation, coherence, and dimensionality, and therefore the image-guided interpolation result is geologically plausible. In n dimensions, each tensor $\mathbf{D}$ in the metric tensor field $\mathbf{D}\left(\mathbf{x}\right)$ is a symmetric positive-definite $n \times n$ matrix (Hale, 2009a). Letting $\mathbf{p}$ and $\mathbf{q}$ denote vectors that contain all elements in $p\left(\mathbf{x}\right)$ and $q\left(\mathbf{x}\right)$, respectively, we can rewrite equation 3.11 in a matrix-vector form:

$$\left(\mathbf{I} + \mathbf{B}^T \mathbf{D} \mathbf{B}\right) \mathbf{q} = \mathbf{p} \, , \tag{3.12}$$

where $\mathbf{B}$ corresponds to a finite-difference approximation of the gradient operator (Hale, 2009b). Therefore, $\mathbf{q} = \mathbf{Q}\mathbf{p}$, where

$$\mathbf{Q} = \left(\mathbf{I} + \mathbf{B}^T \mathbf{D} \mathbf{B}\right)^{-1} \, , \tag{3.13}$$

and this inverse can be efficiently approximated by conjugate-gradient iterations because
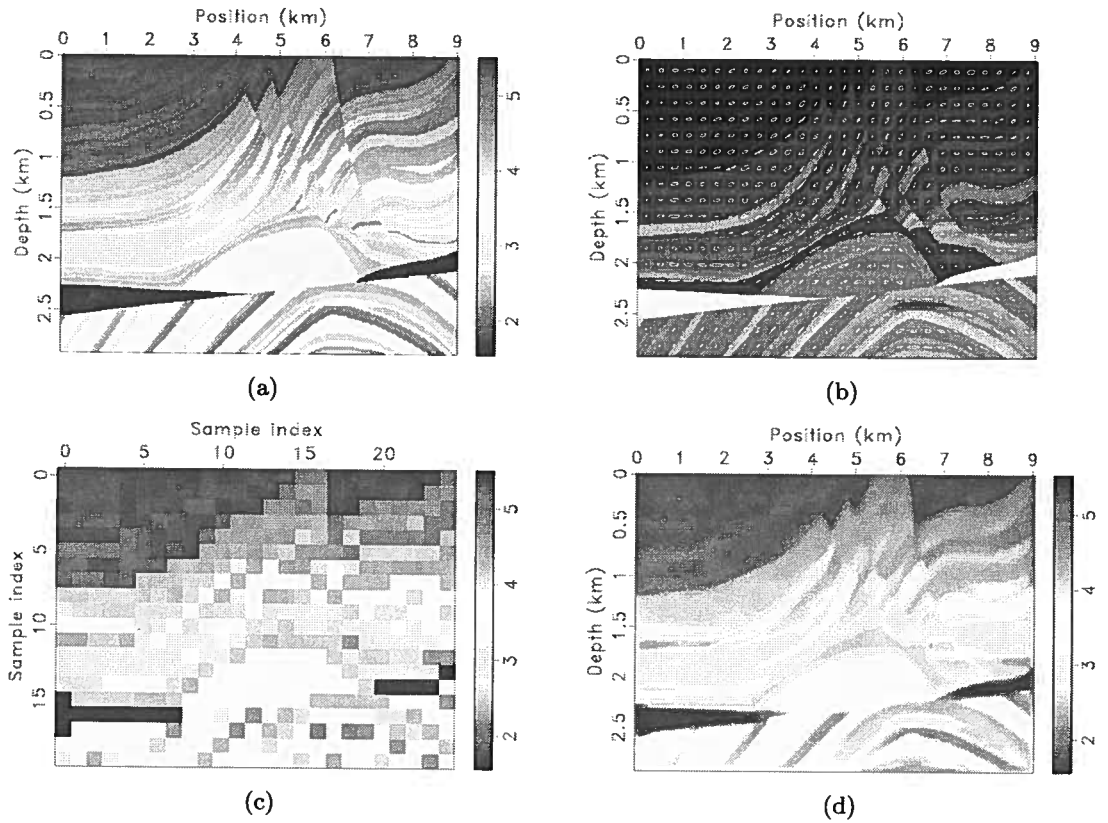
Figure 3.1: An example of image-guided interpolation. (a) The original Marmousi model. (b) The metric tensor field illustrated by ellipses. (c) A decimated Marmousi model, with only 0.2% samples remaining. (d) Image-guided interpolation of samples in (c).

$\mathbf{I} + \mathbf{B}^T \mathbf{D} \mathbf{B}$ is symmetric and positive-definite (SPD). Intuitively, the nearest neighbor interpolation operator $\mathbf{P}$ *scatters* values f_k from sample points $\mathbf{x}_k$ to the interpolation ponits $\mathbf{x}$, and $\mathbf{Q}$ *smooths* the nearest neighbor interpolant $\mathbf{p}$.

Figure 3.1 illustrates an example of image-guided interpolation with a Marmousi velocity model. This example demonstrates the potential of IGI to reduce the number of model parameters. Figure 3.1(a) shows the original Marmousi model with 400×500 samples; ellipses in Figure 3.1(b) depict the metric tensor field $\mathbf{D}(\mathbf{x})$ of the Marmousi model; Figure 3.1(c) represents an undersampled Marmousi model, with only 20×25 (0.2%) samples remaining; Figure 3.1(d) displays the image-guided interpolation result. With IGI, we can reconstruct the Marmousi model in great detail from only a sparsely-sampled model. In practice, we can compute the metric tensor field from migrated images.

3.5.2   Adjoint Image-guided Interpolation

Because $\mathbf{Q}^T = \mathbf{Q}$, we can write the adjoint image-guided interpolation as

$$\mathbf{R}^T = \mathbf{P}^T\mathbf{Q}^T = \mathbf{P}^T\mathbf{Q}\,. \tag{3.14}$$

The adjoint operator $\mathbf{Q}^T$ is again a two-step process:

1. $\mathbf{Q}^T(=\mathbf{Q})$: solve equation 3.11 again to *smooth* the input image;

2. $\mathbf{P}^T$: solve equation 3.10 for $t\,(\mathbf{x})$ and *gather* information from the interpolation points $\mathbf{x}$ to the sample points $\mathbf{x}_k$.

3.6   Implementation of Image-guided FWI

In this section, our scheme is to use image-guided interpolation (Hale, 2009a) to reduce the number of model parameters in the calculation of the gradient of the data misfit function. Because we choose image-guided interpolation as the operator to link the dense model $\mathbf{m}$ and the sparse model $\mathbf{s}$, we refer to the gradient $\mathbf{R}\mathbf{R}^T\mathbf{g}$ in equation 3.7 as the *image-guided gradient*. We also refer to implementation of FWI using the image-guided gradient as image-guided FWI, which again consists of four steps performed iteratively, beginning with an initial model $\mathbf{m}_0$:

(i) Compute the data difference $\mathbf{d} - \mathbf{F}\,(\mathbf{m}_i)$;

(ii) Compute the gradient $\mathbf{g}_i$, $\mathbf{R}^T\mathbf{g}_i$, and $\mathbf{R}\mathbf{R}^T\mathbf{g}_i$;

(iii) Search for a step length α_i in the conjugate direction $\mathbf{h}_i^{\mathbf{m}}$;

(iv) Compute the updated model $\mathbf{m}_{i+1}$ using equation 3.7.

Compared with the four steps of conventional FWI, the only significant difference is the calculation of image-guided gradient in step (ii), where we not only compute the original gradient $\mathbf{g}_i$, but also compute the projected gradient $\mathbf{R}^T\mathbf{g}_i$ and image-guided gradient $\mathbf{R}\mathbf{R}^T\mathbf{g}_i$. Although these two additional computations are required in image-guided FWI, the additional cost can be neglected because it is an insignificant part of the total cost of FWI, which includes much more expensive wavefield simulations.

3.7   Synthetic Example of Image-guided FWI

To illustrate the feasibility of image-guided FWI, we test this technique using the previous model with the same experimental settings, and compare the image-guided FWI results with conventional FWI results.

In step (i), we start with the same initial model $\mathbf{m}_0$ displayed in Figure 2.2(b), and so we obtain the same data residual $\mathbf{d} - \mathbf{F}(\mathbf{m}_0)$ displayed in Figure 2.3(c).

In step (ii), we first compute the gradient of the data misfit function as in step (ii) in the conventional FWI, and thereby obtain a gradient displayed in Figure 2.4(a) that corresponds to the data residual shown in Figure 2.3(c) and the current model $\mathbf{m}_0$ shown in Figure 2.2(b), respectively.

We then compute the image-guided gradient. To obtain this gradient, one must compute the metric tensor field $\mathbf{D}(\mathbf{x})$ that corresponds to the original gradient $\mathbf{g}_0$ of the data misfit function $E(\mathbf{m})$. Because of the structural coincidence between the migrated image and the gradient, we can obtain the metric tensor field $\mathbf{D}(\mathbf{x})$ from the migrated image. Figure 3.2(a) displays ellipses which correspond to the structural orientation of the subsurface over the migrated image. We must also choose several sparse sample points, as depicted by black dots in Figure 3.2(b). In this example, we select only six samples, two of which are located between reflectors. Figure 3.2(c) shows the image-guided gradient $\mathbf{R}\mathbf{R}^T\mathbf{g}_0$ computed in this way for the 1st iteration of image-guided FWI.

In step (iii), we use the same quadratic line-search algorithm to compute a step length α_0. The search direction $\mathbf{h}_0^m$ is determined by conjugate gradients in equation 3.8.

Finally, in step (iv), we update the current velocity model according to equation 3.7. Figure 3.3(a) is the change $\delta\mathbf{m}$ in the velocity model computed in the 1st iteration of image-guided FWI; this change is simply a scaled version of the image-guided gradient in step (ii).

With the image-guided gradient displayed in Figure 3.2(c), image-guided FWI can, even in the 1st iteration, recover most velocity anomalies, as indicated by Figure 3.3(a). Figure 3.5(a) depicts the data residual of shot number 13 after the 1st iteration. However, comparison between Figure 3.5(a) and the data residual displayed in Figure 2.7(a) shows that the 1st iteration of image-guided FWI does not reduce the data misfit as significantly as the conventional FWI does. This is because the image-guided gradient $\mathbf{R}\mathbf{R}^T\mathbf{g}_0$ employed in image-guided FWI cannot clearly resolve the boundaries of the velocity anomalies (see Figure 3.4(a)) due to the smoothing process $\mathbf{Q}$ embedded in the second step of the image-guided interpolation $\mathbf{R}$.
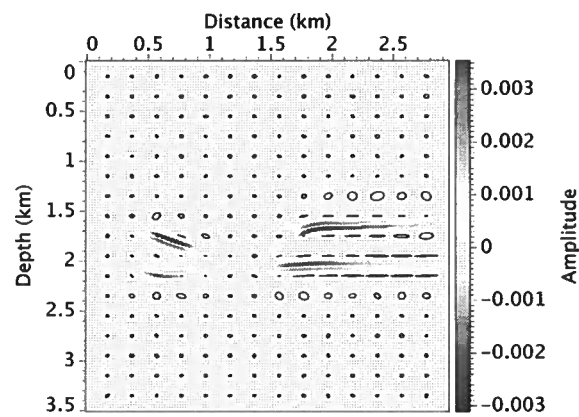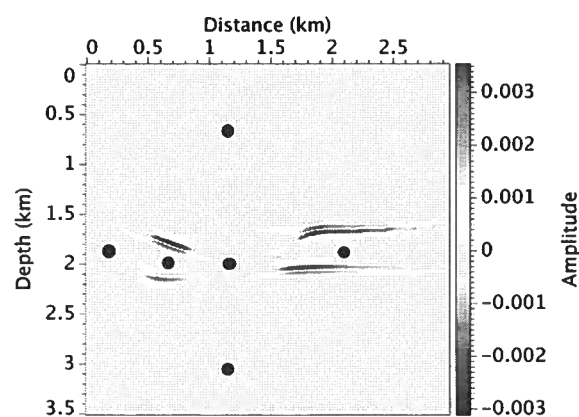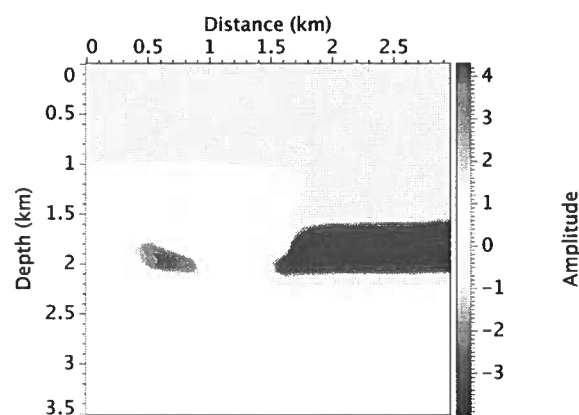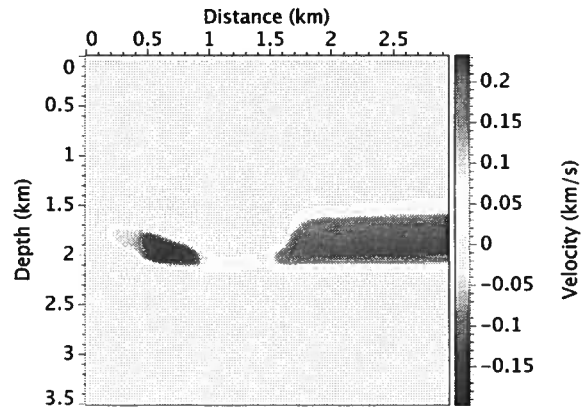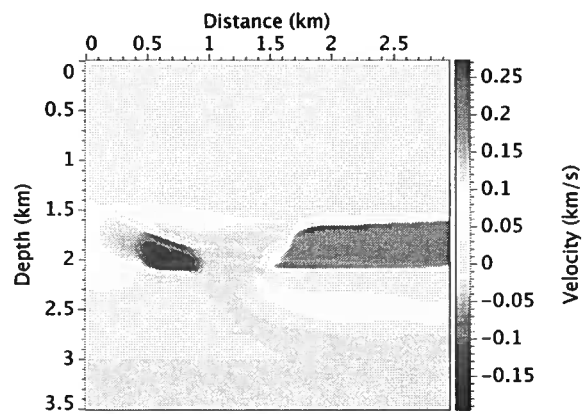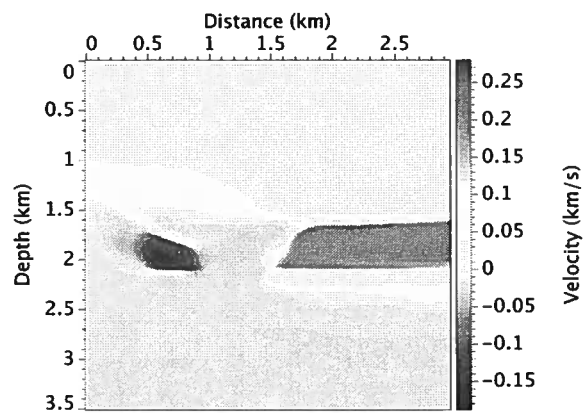
Figure 3.2: (a) The metric tensor field and (b) six selected samples (black dots) overlaid on the gradient image. (c) Image-guided gradient.
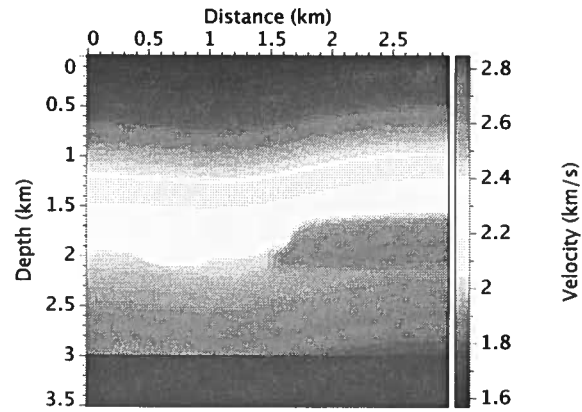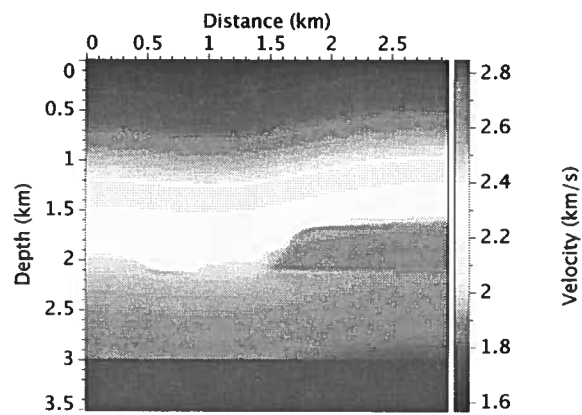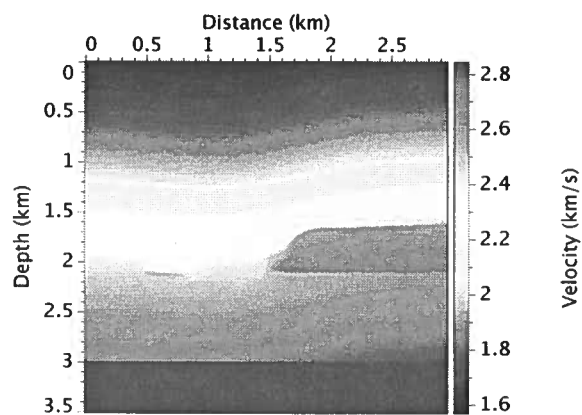
Figure 3.3: Accumulated velocity updates after (a) 1 iteration, (b) 2 iterations and (c) 5 iterations. The image-guided gradient is used only in the first iteration.

Figure 3.4: Updated velocity models after (a) 1 iteration, (b) 2 iterations and (c) 5 iterations.

Figure 3.5: Data residual after (a) 1 iteration, (b) 2 iterations and (c) 5 iterations. The image-guided gradient is used only in the first iteration.

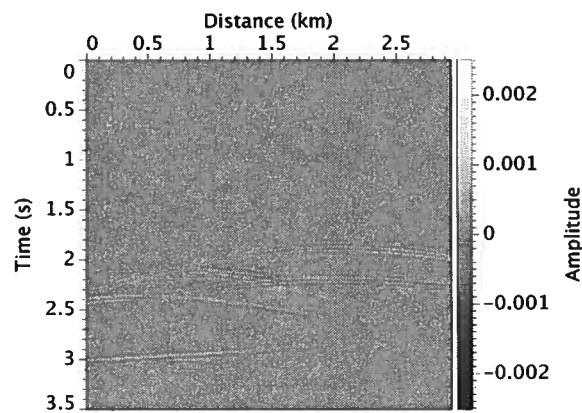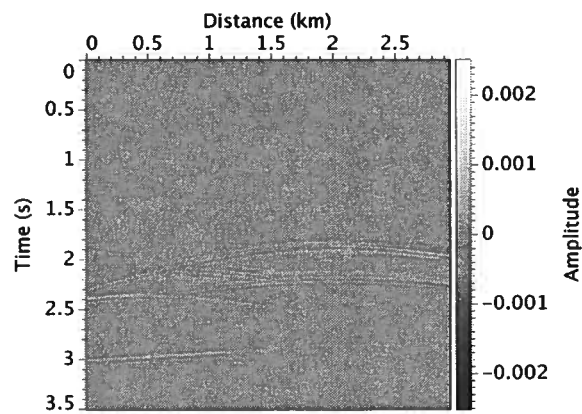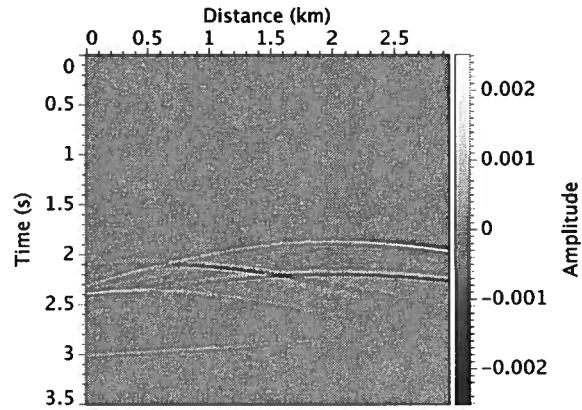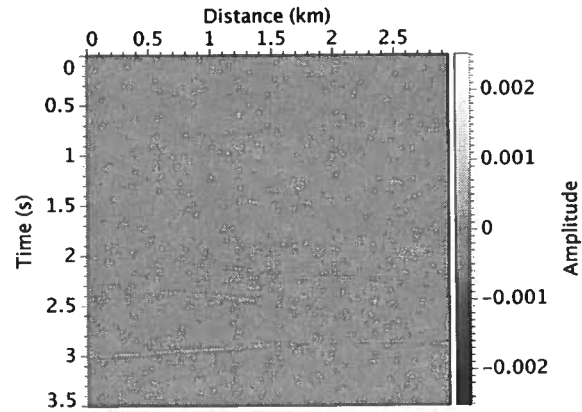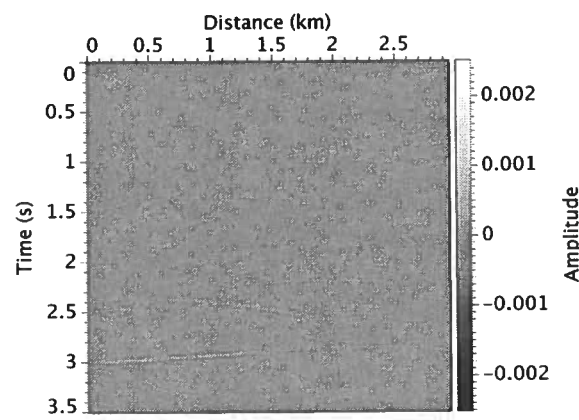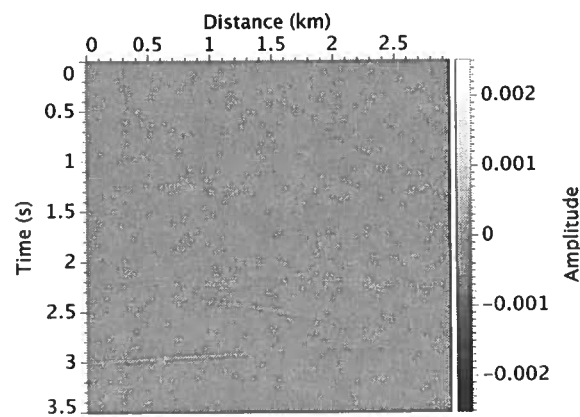Figure 3.6: Data residual after (a) 6 iteration, (b) 8 iterations and (c) 10 iterations. The image-guided gradient is used only in the first iteration.
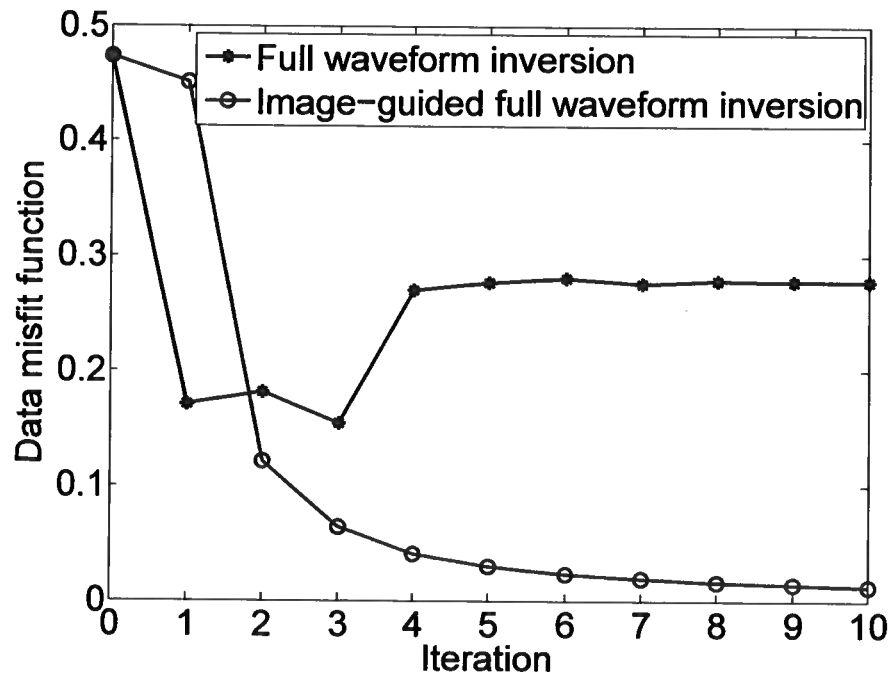
Figure 3.7: Variation of the data misfit function versus iterations in image-guided FWI. Compared with conventional FWI, image-guided FWI monotonically decreases the data misfit.

I solve this problem by using several iterations of conventional FWI to sharpen the boundaries of velocity anomalies. Figure 3.3(b) and (c) are accumulated velocity updates after the 2nd and 5th iterations, respectively. With sharper boundaries, the data simulated in updated velocity model shown in Figure 3.4(b) and 3.4(c) match well with waveforms in the recorded data, thereby decreasing corresponding data residuals significantly, as shown in Figure 3.5(b) and (c). Figure 3.6 shows data residuals after higher-iteration updates, and data residuals are further reduced. Figure 3.7 plots the data misfit objective function as a function of the iteration number. The objective function of image-guided FWI decreases monotonically in the first 10 iterations, unlike that of conventional FWI.

Chapter 4

Discussion

4.1 Summary

The synthetic example demonstrates the improvements of image-guided FWI, which changes only one step in the four-step implementation of conventional FWI. Using image-guided gradient, image-guided FWI speeds convergence by mitigating the limitations of line-search and the absence of low frequencies in recorded data. The effectiveness of image-guided FWI depends on sample selections.

4.2 Line Search

This thesis used a quadratic line-search method to seek a scalar step length that determines how much the velocity model can change. The best situation for this quadratic line search is that we need only two attempts of gradient descent to calculate a step length that decreases the data misfit function. Unfortunately, in some cases, even after many attempts of gradient descent, FWI cannot find a step length to decrease the data misfit function. Because each gradient descent requires a simulation of seismic wavefields of all sources in a dense model space, the line-search approach is quite expensive. Figure 3.7 clearly indicates the failure of the conventional FWI in searching for a proper step length in the 2nd and 4th iterations, using only five trial step lengths.

Although more sophisticated line-search methods may help overcome the limitation of the quadratic line search, I suggest the use of image-guided FWI to avoid the limitation, as indicated by the change of the data misfit function in Figure 3.7. Image-guided FWI successfully finds a step length to decrease the data misfit function in the first 10 iterations, again using only five trial step lengths.

4.3   Low Frequencies

As mentioned in the introduction, the absence of low frequencies in data is one of the major reasons for local minima and cycle-skipping, which prevent FWI from converging to a correct model. Multiscale approaches are proposed to solve the problem by gradually adding high-frequency details into inversion results obtained from low-frequency data. Although those multiscale approaches often start from impractical low frequencies, a question remains. Do low frequencies in data really help? As noted earlier, the velocity updated by FWI maintains imprints of the seismic wavelet because of band-limited data. For this reason, even though one can take advantage of low frequencies in data, wavelet imprints remain and degrade the velocity updates. This degradation is evident in migrated images.

Figure 4.1 compares migrated images with the initial model shown in Figure 2.2(b), the updated model with change shown in Figure 2.5(c), and the updated model with change shown in Figure 3.3(c), respectively. Because of velocity anomalies, deeper reflectors in Figure 4.1(a) do not appear at the correct depth; these deeper reflectors in Figure 4.1(b) appear at almost the same position as in Figure 4.1(a). This mislocation implies that the velocity updated by conventional FWI cannot correct the traveltime mismatch in the data set. One reason for this incorrect update is the wavelet imprint in the velocity updates apparent in Figure 2.5(c). Only the migrated image with the image-guided FWI model places these deeper reflectors at the correct depth, as indicated by Figure 4.1(c).

Rather than dealing with unreliable low frequencies in recorded data, image-guided FWI mitigates the absence of low frequencies in the data space by the image-guided interpolation in the model space. Comparison between the original gradient (Figure 2.4(a)) and the image-guided gradient (Figure 3.2(c)) illustrates the appearance of more low frequencies.

4.4   Limitation of Image-guided FWI

The effectiveness of image-guided FWI depends strongly on how sparse sample points are chosen. In this thesis, I manually picked several sample points, some of which are located between strong structural features, such as reflectors, which are outlined by structure tensor fields. In the example of image-guided FWI in this thesis, if those samples between reflectors were omitted, image-guided interpolation would fail to properly add low-frequency information to the model.
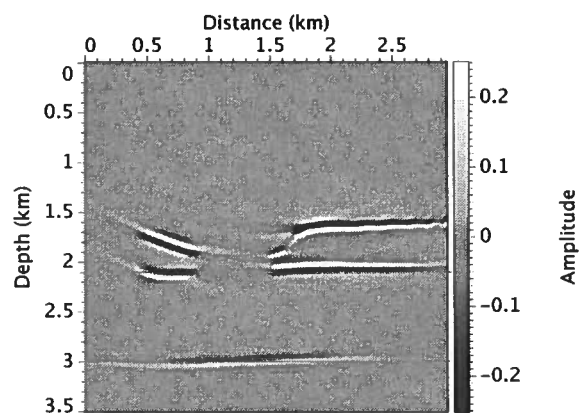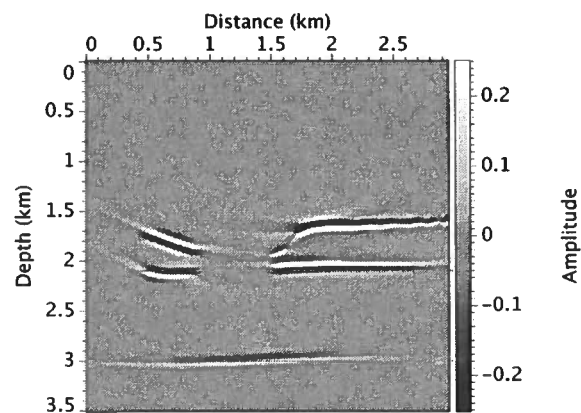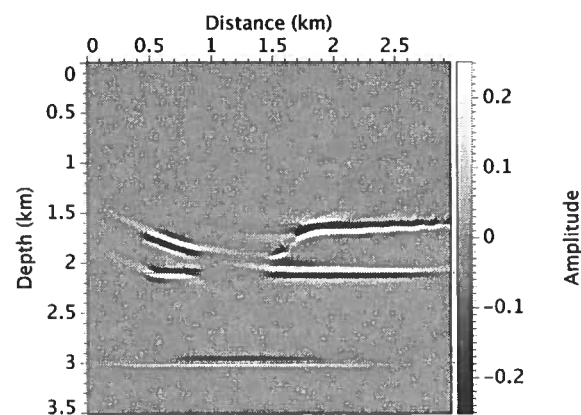
Figure 4.1: Migrated images with (a) the initial model, (b) the conventional FWI model after 5 iterations, and (c) the image-guided FWI model after 5 iterations. Two red lines in each figures indicate the correct depth of reflectors.

Chapter 5

Conclusions and future work

5.1   Conclusions

I have proposed image-guided FWI for speeding up the convergence and mitigating the absence of low frequencies. Different from taking advantage of unreliable low frequencies in the data, as do multiscale approaches, this method reduces the number of model parameters and yields low frequencies in the model space by computing the image-guided gradient with image-guided interpolation and its adjoint. The synthetic example shown in this thesis illustrates that image-guided FWI improves both inversion speed and quality without significant additional cost. Because the structural features in the subsurface are taken into consideration, models updated by image-guided FWI are geologically sensible. This thesis tests image-guided FWI with only a simple synthetic example, and further investigations of this technique with more realistic models and real data are necessary.

5.2   Sample Selection in IGI

As noted in the previous section, effectiveness of the image-guided FWI technique depends on sample selections. Further study on sample selection should address the following questions:

- How many samples are necessary and sufficient to build the image-guided gradient?

- Where are those samples best located?

A baseline criteria for selecting samples is that the Hessian matrix $\mathbf{H}$ (or the projected Hessian $\mathbf{R}^T \mathbf{H} \mathbf{R}$) should become better conditioned. For this to happen, samples should be located in areas where changes in model parameters will create significant changes in data.

5.3   Which Iteration Should Be Image-guided?

Image-guided FWI plays an important role in improving the convergence by reducing the number of model parameters as well as incorporating low frequencies in the model space. However, conventional FWI is still needed to enhance high-frequency components. Further investigation on the optimal integration of two kinds of FWI is necessary:

- What are the criteria for switching between conventional FWI and image-guided FWI?

5.4   FWI in Reduced Model Space

Image-guided FWI is simply FWI performed in a reduced model space. The feasibility of FWI in the reduced model space comes from the fact that a dense model used in simulation of seismic wave propagation has redundancy. For example, the grid size in the finite-difference approximation is usually small relative to the wavelengths of seismic waves. This redundancy in the model space is even more obvious when some kind of *a priori* knowledge (e.g., structural features) can be used to describe the model.

5.4.1   Direct Conjugate-gradient Solution

In the thesis, I designed the image-guided FWI by applying the image-guided interpolation operator $\mathbf{R}$ onto the gradient. With further investigation, $\mathbf{R}$ might be directly combined with the wave equation. For example, a linear system is formulated as

$$\mathbf{Fm} = \mathbf{d} . \tag{5.1}$$

With equation 3.1, we can rewrite equation 5.1 as

$$\mathbf{FRs} = \mathbf{d} . \tag{5.2}$$

The solution for equation 5.2 is equivalent to the solution for the following equation:

$$\mathbf{R}^T \mathbf{F}^T \mathbf{FRs} = \mathbf{R}^T \mathbf{F}^T \mathbf{d} , \tag{5.3}$$

where $\mathbf{F}^T$ is the adjoint operator of $\mathbf{F}$. In seismic migration language, $\mathbf{F}$ is the modeling operator and $\mathbf{F}^T$ is the migration operator. If $\mathbf{R}^T \mathbf{F}^T \mathbf{FR}$ remains sparse, symmetric and positive-definite, equation 5.3 can be efficiently solved using a direct conjugate-gradient

solver with respect to a reduced set of parameters **s**. In this case, we can avoid line-search approaches, which are necessary in this thesis.

5.4.2   Use $\mathbf{R}^T\mathbf{HR}$

Most of the computational cost of FWI lies in the search for a scalar step length in the gradient direction **g** or conjugate direction **h**. This scalar step-length is a global parameter that scales the whole model space uniformly. Intuitively, this global scaling could be unreliable, because the gradient image **g** might not proportionally and correctly measure the model update of the subsurface, especially in areas with poor illumination.

Instead, we might search for a vector (or a matrix); i.e, we need a set of distributed parameters to scale the gradient image **g** locally with different weights. One example of such a set of distributed parameters can be the inverse of the (approximated) Hessian matrix **H** in the Newton-like method used to solve FWI. However, the typically large number of model parameters makes the computation of the Hessian matrix and its inverse impractical. A more practical way is to compute the projected Hessian matrix $\mathbf{R}^T\mathbf{HR}$, which corresponds to a sparse model with a reduced number of parameters. An efficient algorithm to compute the projected Hessian matrix $\mathbf{R}^T\mathbf{HR}$ is critical.

5.4.3   Multiparameter FWI

FWI often focuses on inverting for a velocity model of the subsurface, and meanwhile it is also important to invert for a density model that in part accounts for the amplitude mismatch between recorded and synthetic data. Multiparameter FWI becomes more feasible in the reduced model space. To invert for velocity and density models simultaneously, redundancies of both velocity and density can be taken into account in the reduced model space method.

5.4.4   Application of Borehole Data

FWI should produce models that conform to borehole measurements. In the reduced model space, FWI may sacrifice accuracy for efficiency, and in this case, borehole data are useful to

- Constrain inversion in the data space and the model space;

- Estimate uncertainty of FWI.

Moreover, borehole data can also be incorporated into steps of A draft for image-guided FWI with borehole data is that, from an initial model $\mathbf{m}_0$, we can perform following steps iteratively:

- Obtain a migrated image with the current model;

- Apply image-guided interpolation to borehole data for an model for the next step;

- Update the model in the previous step using one iteration of image-guided FWI (or conventional FWI if necessary).

Here, the borehole data help build a better initial model for image-guided FWI.

References

Batzle, M., & Wang, Z.J. 1992. Seismic properties of pore fluids. *Geophysics*, **57**(11), 1396–1408.

Boonyasiriwat, C., Valasek, P., Routh, P., Cao, W., Schuster, G. T., & Macy, B. 2009. An efficient multiscale method for time-domain waveform tomography. *Geophysics*, **74**(6), WCC59–WCC68.

Bunks, C. 1995. Multiscale seismic waveform inversion. *Geophysics*, **60**(5), 1457.

Claerbout, J. F. 1985. *Imaging the earth's interior.* Cambridge, MA, USA: Blackwell Scientific Publications, Inc.

Cohen, A. I. 1972. Rate of Convergence of Several Conjugate Gradient Algorithms. *SIAM Journal on Numerical Analysis*, **9**(2), 248–259.

Etgen, J., Gray, S. H., & Zhang, Y. 2009. An overview of depth imaging in exploration geophysics. *Geophysics*, **74**(6), WCA5–WCA17.

Fehmers, G. C., & Höcker, C. F. W. 2003. Fast structural interpretation with structure-oriented filtering. *Geophysics*, **68**(4), 1286–1293.

Gong, B., Chen, G., Yingst, D., & Bloor, R. 2008. 3D waveform inversion based on reverse time migration engine. *SEG Technical Program Expanded Abstracts*, **27**(1), 1900–1903.

Hale, D. 2009a. Image-guided blended neighbor interpolation. *CWP Report*, **634**(634).

Hale, D. 2009b. Structure-oriented smoothing and semblance. *CWP Report*, **635**(635).

Jiao, J., Lowrey, D. R., Willis, J. F., & Martinez, R. D. 2008. Practical approaches for subsalt velocity model building. *Geophysics*, **73**(5), VE183–VE194.

Kennett, B.L.N., Sambridge, M.S., & Williamson, P.R. 1988. Subspace methods for large inverse problems with multiple parameter classes. *Geophysical Journal*, **94**(2), 237–247.

Krebs, J. R., Anderson, J. E., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A., & Lacasse, M. 2009. Fast full-wavefield seismic inversion using encoded sources. *Geophysics*, **74**(6), WCC177–WCC188.

Meng, Z. 2009. Dip guided full waveform inversion. *Patent*, 41279–USPRO.

Meng, Z., & Scales, J. A. 1996. 2D tomography in multi-resolution analysis model space. *SEG Technical Program Expanded Abstracts*, **15**(1), 1126–1129.

Mora, P. 1989. Inversion = migration + tomography. *Geophysics*, **54**(12), 1575.

Nocedal, J., & Wright, S. J. 2000. *Numerical Optimization.* Springer.

Oezsen, R. 2004. Velocity modelling and prestack depth imaging below complex salt structures: a case history from on-shore Germany. *Geophysical Prospecting*, **52**(6), 693–705.

Oldenburg, D.W., McGillvray, P.R., & Ellis, R.G. 1993. Generalized subspace methods for large-scale inverse problems. *Geophysical journal international*, **114**(1), 12–20.

Operto, S., Ravaut, C., Improta, L., Virieux, J., Herrero, A., & Dell'Aversana, P. 2004. Quantitative imaging of complex structures from dense wide-aperture seismic data by multiscale traveltime and waveform inversions: a case study. *Geophysical Prospecting*, **52**(6), 625–651.

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

Pratt, R.G., Shin, C., & Hicks, G.J. 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophysical Journal International*, **133**(2), 341–362.

Sava, P., & Biondi, B. 2004a. Wave-equation migration velocity analysis. I. Theory. *Geophysical Prospecting*, **52**(6), 593–606.

Sava, P., & Biondi, B. 2004b. Wave-equation migration velocity analysis. II. Subsalt imaging examples. *Geophysical Prospecting*, **52**(6), 607–623.

Shin, C., & Ha, W. 2008. A comparison between the behavior of objective functions for waveform inversion in the frequency and Laplace domains. *Geophysics*, **73**(5), VE119–VE133.

Shin, C., & Min, D. J. 2006. Waveform inversion using a logarithmic wavefield. *Geophysics*, **71**(3), R31–R42.

Sirgue, L., & Pratt, R. G. 2004. Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics*, **69**(1), 231.

Snieder, R. 1998. The role of nonlinearity in inverse problems. *Inverse Problems*, **14**(3), 387.

Stork, C. 1992. Reflection tomography in the postmigrated domain. *Geophysics*, **57**(5), 680–692.

Symes, W. W. 2008. Migration velocity analysis and waveform inversion. *Geophysical Prospecting*, **56**(6), 765–790.

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

Tarantola, A. 1986. A strategy for nonlinear elastic inversion of seismic reflection data. *Geophysics*, **51**(10), 1893—1903.

Tarantola, A. 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation.* Society for Industrial and Applied Mathematics.

Tarantola, A., & Valette, B. 1982. Generalized non-linear inverse problems solved using the least-squares criterion. *Reviews of Geophysics*, **20**(2), 219–232.

Tromp, J., Tape, C., & Liu, Q.Y. 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, **160**(1), 195–216.

van Vliet, L. J., & Verbeek, P. W. 1995. Estimators for Orientation and Anisotropy in Digitized Images. *Proceeding of the First Annual Conference of the Advanced School for Computing and Imaging*, 442–450.

Vasco, D.W., & Majer, E.L. 1993. Wavepath travel-time tomography. *Geophysical Journal International*, **115**(3), 1055–1069.

Vigh, D., & Starr, E.W. 2008. 3D prestack plane-wave, full-waveform inversion. *Geophysics*, **73**(5), VE135–VE144.

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

Wang, B., Kim, Y., Mason, C., & Zeng, X. 2008. Advances in velocity model-building technology for subsalt imaging. *Geophysics*, **73**(5), VE173–VE181.

Wheeler, J.E., & Wilton, D.R. 1988. Comparison of convergence rates of the conjugate gradient method applied to various integral equation formulations. *Pages 229–232 of: Antennas and Propagation Society International Symposium.*

Woodward, M. J. 1992. Wave-equation tomography. *Geophysics*, **57**(1), 15–26.

Yilmaz, O., & Chambers, R. 1984. Migration velocity analysis by wave-field extrapolation. *Geophysics*, **49**(10), 1664–1674.

Zelt, C. A., & Barton, P. J. 1998. Three-dimensional seismic refraction tomography: A comparison of two methods applied to data from the Faeroe Basin. *J. Geophys. Res.*, **103**(B4), 7187–7210.