SEISMIC WAVEFORM MODELING AND INVERSION
IN ACOUSTIC ORTHORHOMBIC MEDIA

by
Hui Wang
A thesis submitted to the Faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of Master of Science (Geophysics).

Golden, Colorado
Date _____________________

Signed: ___________________  Hui Wang

Signed: ___________________  Dr. Ilya Tsvankin
   Thesis Advisor

Golden, Colorado
Date _____________________

Signed: ___________________  Dr. Roel Snieder
   Professor and Head
   Department of Geophysics
ABSTRACT

Three-dimensional seismic waveform inversion (WI) for anisotropic media is highly challenging due to its computational cost, large number of model parameters, and parameter trade-offs. In this thesis, I explore 3D waveform inversion for orthorhombic media in the acoustic approximation. Two mixed-domain seismic wavefield simulators are implemented; one of them is based on low-rank decomposition and the other on the generalized pseudospectral method. Both methods can produce kinematically accurate pure-mode P-wavefields with an acceptable computational cost. The low-rank-decomposition-based method is used to simulate both state and adjoint wavefields due to its higher accuracy and stability. The wave equations from the pseudospectral method are employed to obtain the gradients of the WI objective functionals. To build the initial long-wavelength model for waveform inversion, I use an envelope-based misfit functional, which alleviates the reliance of WI on low-frequency data. The WI gradients are derived for both the conventional data-difference and the envelope-based objective functions. Numerical examples illustrate the performance of the developed wavefield-extrapolation and gradient-computation algorithms for orthorhombic media with realistic complexity. WI is conducted with the help of a limited-memory version of the quasi-Newton optimization algorithm. A test for a modified version of the SEG/EAGE overthrust model validates the proposed approach to waveform inversion in acoustic orthorhombic media.
# TABLE OF CONTENTS

ABSTRACT ................................................................. iii  
LIST OF FIGURES ......................................................... vi  
ACKNOWLEDGMENTS .................................................... viii  
CHAPTER 1 INTRODUCTION ................................................. 1  
CHAPTER 2 MIXED-DOMAIN WAVEFIELD SIMULATOR ................. 5  
  2.1 Wavefield Simulator based on Low-rank Decomposition ........ 9  
  2.2 Generalized Pseudospectral Wavefield Simulator .......... 10  
  2.3 Absorbing Boundary Condition ................................. 12  
  2.4 Numerical Examples ........................................... 12  
CHAPTER 3 WAVEFORM INVERSION ....................................... 16  
  3.1 Misfit Functionals .............................................. 16  
  3.2 Adjoint Wavefield Propagation ................................. 19  
  3.3 Medium Parameterization ..................................... 20  
  3.4 Waveform Inversion Gradients ................................. 23  
  3.5 Synthetic Example of Waveform Inversion ................. 26  
CHAPTER 4 CONCLUSIONS AND RECOMMENDATIONS .................... 33  
  4.1 Conclusions .................................................. 33  
  4.2 Recommendations ............................................ 34  
REFERENCES CITED .................................................. 36  
APPENDIX A - ANALYTIC SOLUTION FOR P-WAVE DISPERSION RELATION . 43
LIST OF FIGURES

Figure 2.1 Wavefield snapshots for a homogeneous orthorhombic model computed with: (a) the generalized pseudospectral method, and (b) low-rank decomposition. Red dotted lines correspond to the P-wave traveltimes obtained from the eikonal equation. ........................................ 14

Figure 2.2 Shot gathers computed at the surface of the model without absorbing boundary conditions. (a) The generalized pseudospectral method, and (b) low-rank decomposition. .................................................. 15

Figure 2.3 Shot gathers computed at the surface of the model with absorbing boundary conditions. (a) The generalized pseudospectral method, and (b) low-rank decomposition. .................................................. 15

Figure 3.1 Comparison of the predicted (red) and observed (green) traces: (a) raw data, (b) data envelopes, and (c) squared envelopes. ................................. 18

Figure 3.2 Spectra of the predicted (blue) and observed (red) raw data. The spectra of the predicted (green) and observed (yellow) envelope data are shifted toward lower frequencies. ................................. 19

Figure 3.3 Models for the six parameters of an acoustic orthorhombic medium. Gaussian anomalies are embedded in a background medium with linearly increasing velocities. .................................................. 22

Figure 3.4 (a) Six Gaussian anomalies from Figure 3.3 plotted together. (b) Projections of the anomalies from Figure 3.4(a) and the array of sources (red dots) at the surface. ........................................ 22

Figure 3.5 WI gradients for the model from Figure 3.3 computed with the \( \ell_2 \)-norm data difference. The gradients correspond to the background model in Figure 3.3 with the linearly increasing velocities. ........................................ 23

Figure 3.6 Orthorhombic medium obtained from the SEG/EAGE overthrust model. The velocities are scaled from the original P-wave isotropic velocity field. ................................................................. 25

Figure 3.7 Horizontal projection of the source locations (red dots), which are on the surface of the model from Figure 3.6. ........................................ 25
Figure 3.8  Initial parameter fields used to compute the WI gradients for the model in Figure 3.6. The initial velocities are smoothed and deviate significantly from the actual values. ........................................... 26

Figure 3.9  Gradients for the model from Figure 3.6 computed with the $\ell_2$-norm data difference functional. The gradients correspond to the initial model Figure 3.8. ................................................................. 27

Figure 3.10 Gradients for the model from Figure 3.6 computed with the $\ell_2$-norm squared envelope functional. The gradients correspond to the initial model Figure 3.8. ................................................................. 27

Figure 3.11 Orthorhombic medium modified from the SEG/EAGE overthrust model. The velocities are scaled from the original P-wave isotropic velocity field. ................................................................. 28

Figure 3.12 Horizontal projection of the source locations (red dots), which are on the surface of Figure 3.11. ................................................................. 28

Figure 3.13 Initial parameter fields used to perform WI for the model in Figure 3.11. The initial velocities are smoothed and deviate significantly from the actual values. ........................................... 30

Figure 3.14 Inverted parameter fields after 102 iterations of the inversion algorithm. . 30

Figure 3.15 Vertical profiles of the true (black), initial (red), and inverted (blue) velocity $V_{Pz}$ at locations shown in Figure 3.16. ........................................... 31

Figure 3.16 Locations of velocity profiles in Figure 3.15. ................................. 32

Figure 3.17 Data residuals using (a) the initial, and (b) the inverted models. .... 32
ACKNOWLEDGMENTS

This thesis is dedicated to my advisor, professors, and colleagues at Colorado School of Mines. I appreciate the financial support for my M.S. study from the Center for Wave Phenomena (CWP) and its sponsors.

I am deeply grateful to my advisor, Ilya Tsvankin, for his persistent guidance, continual patience, and insightful research advice throughout the research projects I have worked on. I am fortunate to take three seismology courses from Ilya, which I think are the best graduate courses in this department. Ilya is not only a great advisor, an excellent lecturer, but also a scientific role model to me. Ilya sets high scientific standards for his students, stays connected to the details of students’ research projects, and continually raises the bar and drives A-Team to deliver high-quality research products.

Many thanks to my professors, especially Paul Sava and Michael Wakin. Paul’s class on seismic imaging filled the gaps of my knowledge in seismic imaging. The sparse signal processing course I took from Michael was quite comprehensible and valuable to my research. I was amazed to find how well the digital signal processing (DSP) concepts were formulated in suitable mathematical settings. Roel Snieder and Yaoguo Li were kindly willing to be part of my committee and provided valuable feedback on this thesis.

I enjoyed my overlap with Pam Kraus, Shingo Sean Ishida, and Diane Witters. Pam handled all CWP students’ non-technical work, which could be a headache for me. Shingo compiled all research reports and posters for the group. Diane helped me a lot with drafting my technical papers and communicating with other people in general. Many thanks are due to the fellow students at CWP, who generously shared their time deriving equations on white boards and pair programming with me.

Finally, I am eternally grateful to my family for their lasting love and support. I owe them beyond what words can express.
CHAPTER 1
INTRODUCTION

Seismic wave simulation and parameter estimation in anisotropic media have become common practice in both academic and industrial applications. With the advances in seismic acquisition, which make it possible to record long-offset wide-azimuth and multicomponent data, exploration seismologists can no longer ignore the directional dependence of elastic properties. Last decades have witnessed extensive research on simulating seismic wave propagation in anisotropic media with high symmetries, especially transverse isotropy (Duveneck and Bakker, 2011; Fletcher et al., 2008, 2009; Fowler et al., 2010; Schleicher and Costa, 2015; Xu and Zhou, 2014; Zhang et al., 2011; Zhou et al., 2006a,b). The corresponding inverse problems for anisotropic parameter estimation utilizing these wavefield simulators are also being explored to some extent (Bakulin et al., 2010; Barnes et al., 2008; Bozdag and Trampert, 2008; Burridge et al., 1998; de Hoop et al., 1999; Debayle and Kennett, 2000; Ferreira et al., 2010; Marone et al., 2007; Prieux et al., 2011; Ursin, 2004; Warner et al., 2013b).

Transverse isotropy, however, cannot describe many subsurface formations that exhibit orthorhombic symmetry due to the influence of aligned fractures and nonhydrostatic stresses. Orthorhombic models have been successfully used in processing of wide-azimuth reflection and VSP data and fracture characterization (Tsvankin, 1997; Tsvankin and Grechka, 2011). In this thesis, I focus on acoustic orthorhombic models described by a simplified wave equation that preserves the P-wave kinematics (Alkhalifah, 1998, 2000). As shown by Tsvankin (1997), P-wave kinematic signatures in orthorhombic media are controlled by six parameters – the P-wave vertical velocity $V_{P0}$ and anisotropy coefficients $\varepsilon^{(1)}$, $\varepsilon^{(2)}$, $\delta^{(1)}$, $\delta^{(2)}$, and $\delta^{(3)}$ (assuming one of the symmetry planes to be horizontal). Other parameterizations for acoustic orthorhombic media (e.g., Masmoudi and Alkhalifah, 2016), include certain combinations of $V_{P0}$, $\varepsilon^{(1,2)}$, and $\delta^{(1,2,3)}$.  


Since the formal introduction by Lailly (1983) and Tarantola (1984), waveform inversion (WI) has been an active area of research in both exploration and global seismology (Brossier et al., 2009; Fichtner, 2010; Fichtner et al., 2008; Liu and Tromp, 2008; Plessix, 2009; Pratt et al., 1996; Pratt, 1999; Pratt et al., 1998; Pratt and Shipp, 1999; Sirgue et al., 2010, 2009, 2008; Sirgue and Pratt, 2004; Symes, 2010; Tromp et al., 2005; Vigh and Starr, 2008). However, most existing waveform-inversion techniques are designed to recover just P-wave velocity due to the high computational cost and the intrinsic nonlinearity of the inverse problem. Recently, WI has been extended to both acoustic and elastic transversely isotropic models with a vertical symmetry axis (VTI). Kamath and Tsvankin (2013) apply elastic WI to multicomponent reflection data for layer-cake VTI media to obtain the interval medium parameters. Gholami et al. (2011) present a case study for Valhall field using a 2D VTI acoustic WI algorithm. Kamath and Tsvankin (2016) develop elastic WI for 2D VTI media and apply it to transmission data for models with Gaussian anomalies in the Thomsen parameters. They also perform sensitivity analysis using the WI radiation patterns for parameter perturbations. A natural extension of the previous research is to explore waveform inversion in orthorhombic media.

Wavefield simulators are critically important for waveform inversion. Two categories of methods have been proposed to model P-wave propagation in anisotropic media: coupled systems and mixed-domain wavefield extrapolators. The coupled systems have been originally introduced for TI media (Duveneck and Bakker, 2011; Fletcher et al., 2008, 2009; Fowler et al., 2010; Zhou et al., 2006a,b) and can be extended to orthorhombic symmetry. However, the coupled systems produce shear-wave “artifacts” (Grechka et al., 2004) and suffer from the ambiguity in the physical interpretation of the auxiliary wavefield variables. The mixed-domain wavefield extrapolators, on the other hand, can simulate pure P-wave propagation. In this thesis, I implement two efficient mixed-domain wavefield extrapolators: those based on low-rank matrix decomposition (Fomel et al., 2013; Song and Alkhalifah, 2012) and generalized pseudospectral (Fowler and Lapilli, 2012) methods.
WI requires an accurate initial model for local optimization methods to converge to the global minimum of the objective function (Virieux and Operto, 2009; Warner et al., 2013a). Migration velocity-analysis methods (Biondi et al., 2012; Wang and Tsvankin, 2013; Yang and Sava, 2011) can produce long-wavelength parameter fields that accurately describe the kinematics of recorded arrivals. To improve long-wavelength models at early stages of waveform inversion, it is common to use multiscale methods (Bunks et al., 1995). The data are often divided into several frequency bands, and WI is performed sequentially starting with the lowest frequencies.

However, conventional seismic acquisition cannot provide ultra-low-frequency (down to 1-2 Hz) data, which are critical for constraining long-wavelength parameter fields. Designing suitable WI misfit functionals can help address this issue. In global seismology, Nolet et al. (1986) describe an envelope-based formalism for waveform fitting with surface waves. Snieder et al. (1989) show that misfit functionals operating with envelopes are smoother and more convex than the conventional $\ell_2$-norm objective function. More recently, Fichtner et al. (2008), Fichtner (2010), and Bozdağ et al. (2011) explore time-frequency and envelope information of waveform data to mitigate cycle-skipping issues. At the exploration scale, Wu et al. (2014) and Luo and Wu (2015) show that so-called envelope inversions can constrain long-wavelength models without such low frequencies. I employ such envelope-based misfit functionals to alleviate the reliance of WI on low-frequency data.

This thesis is divided into three parts. Chapter 2 introduces P-wave simulators based on the low-rank matrix decomposition and generalized pseudospectral mixed-domain operators. The corresponding numerical adjoint systems are presented, which are crucial to implementation of waveform inversion. To validate the numerical propagators, I test both methods using several orthorhombic models.

Chapter 3 is devoted to waveform inversion using the aforementioned wavefield simulators. I employ the adjoint technique (Chen, 2011; Fichtner et al., 2006a,b; Liu and Tromp, 2008; Plessix, 2006; Tarantola, 1988; Tromp et al., 2005) to derive and compute the gradi-
ents of the WI objective function with respect to the six orthorhombic parameters. Both the envelope-based and conventional objective functions are considered. The gradients are tested on synthetic data from 3D orthorhombic media. The multiparameter waveform inversion is then carried out using a nonlinear optimization algorithm (Benson and Moré, 2001; Kolda et al., 1998; Nocedal, 1992; Nocedal and Wright, 2006; Thiébaut, 2002). The inversion results show that the P-wave velocities along the coordinate axes are better constrained by surface data than the NMO velocities. There are significant parameter trade-offs among the NMO velocities during the WI process.

In chapter 4, I summarize the thesis results and provide recommendations for future work.
CHAPTER 2  
MIXED-DOMAIN WAVEFIELD SIMULATOR

The starting point for deriving pure-mode mixed-domain wavefield extrapolators is the dispersion relation of the corresponding wave mode. These extrapolators satisfy a general equation of the form

$$\partial_{tt} u(k,t) + \Phi(x,k) u(k,t) = 0,$$

(2.1)

where $u(k,t)$ denotes the scalar wavefield variable in the time-wavenumber domain, $k$ is the magnitude of the wave vector, $\partial_{tt}$ is the second time-derivative operator, and $\Phi(x,k)$ is a linear operator defined in the mixed (spatial and wavenumber) domain; the source term in equation 2.1 is ignored. In isotropic media, the mixed-domain operator $\Phi$ reduces to

$$\Phi(x,k) = v^2(x) |k|^2,$$

(2.2)

where $v(x)$ is the velocity. If the model is anisotropic, the mixed-domain operator for a certain mode can be obtained from the corresponding dispersion relation using the Christoffel equation. For VTI media, the P-wave mixed-domain operator in the acoustic approximation has the form (Alkhalifah, 1998):

$$\Phi(x,k) = \frac{1}{2} \left[ (1 + 2\varepsilon) V^2_{P0} k^2_r + V^2_{P0} k^2_z \right] + \frac{1}{2} \left[ (1 + 2\varepsilon) V^2_{P0} k^2_r + V^2_{P0} k^2_z \right] \sqrt{1 - \frac{8 (\varepsilon - \delta) k^2_r k^2_z}{(1 + 2\varepsilon) k^2_r + k^2_z} },$$

(2.3)

where $\varepsilon$ and $\delta$ are Thomsen parameters, $V_{P0}$ is the P-wave vertical velocity, and $k^2_r$ is the horizontal wavenumber ($k^2_r = k^2_x + k^2_y$). In the case of acoustic orthorhombic media, the Christoffel matrix can be written as

$$G = \begin{bmatrix}
  k^2_r V^2_{P0} (1 + 2\varepsilon^{(2)}) & k_x k_y V^2_{P0} (1 + 2\varepsilon^{(2)}) \sqrt{1 + 2\delta^{(3)}} & k_z k_y V^2_{P0} \sqrt{1 + 2\delta^{(2)}} \\
  k_x k_y V^2_{P0} (1 + 2\varepsilon^{(2)}) \sqrt{1 + 2\delta^{(3)}} & k^2_r V^2_{P0} (1 + 2\varepsilon^{(1)}) & k_y k_z V^2_{P0} \sqrt{1 + 2\delta^{(1)}} \\
  k_x k_z V^2_{P0} \sqrt{1 + 2\delta^{(2)}} & k_y k_z V^2_{P0} \sqrt{1 + 2\delta^{(1)}} & k^2_z V^2_{P0}
\end{bmatrix},$$

(2.4)
where the six independent Thomsen-style parameters (Tsvankin, 1997) represent the following combinations of the stiffness coefficients:

\[ V_{p0} = \sqrt{\frac{c_{33}}{\rho}} \]  

(2.5)

\[ \varepsilon^{(1)} = \frac{c_{22} - c_{33}}{2c_{33}} \]  

(2.6)

\[ \varepsilon^{(2)} = \frac{c_{11} - c_{33}}{2c_{33}} \]  

(2.7)

\[ \delta^{(1)} = \frac{(c_{23} + c_{44})^2 - (c_{33} - c_{44})^2}{2c_{33}(c_{33} - c_{44})} \]  

(2.8)

\[ \delta^{(2)} = \frac{(c_{13} + c_{55})^2 - (c_{33} - c_{55})^2}{2c_{33}(c_{33} - c_{55})} \]  

(2.9)

\[ \delta^{(3)} = \frac{(c_{12} + c_{66})^2 - (c_{11} - c_{66})^2}{2c_{11}(c_{11} - c_{66})} \]  

(2.10)

These parameters are defined similarly to the Thomsen parameters in the corresponding symmetry planes:

- \( V_{p0} \) – the vertical velocity of P-wave;
- \( \varepsilon^{(1)} \) – the VTI parameter \( \varepsilon \) in the \([x_2, x_3]\) plane normal to the \(x_1\)-axis;
- \( \varepsilon^{(2)} \) – the VTI parameter \( \varepsilon \) in the \([x_1, x_3]\) plane normal to the \(x_2\)-axis;
- \( \delta^{(1)} \) – the VTI parameter \( \delta \) in the \([x_2, x_3]\) plane normal to the \(x_1\)-axis;
- \( \delta^{(2)} \) – the VTI parameter \( \delta \) in the \([x_1, x_3]\) plane normal to the \(x_2\)-axis;
- \( \delta^{(3)} \) – the VTI parameter \( \delta \) in the \([x_1, x_2]\) plane normal to the \(x_3\)-axis.

An important advantage of Tsvankin’s (1997; 2011) notation is that it reduces the number of independent parameters responsible for P-wave velocity from nine to six. The P-wave dispersion relation is then obtained by solving the characteristic equation of the eigenvalue-eigenvector problem:

\[ \det (G - \omega^2 I) = 0, \]  

(2.11)
which results in a cubic equation in $\omega^2$:

$$\omega^6 + a_2 \omega^4 + a_1 \omega^2 + a_0 = 0,$$

(2.12)

where

$$a_2 = - \left[ (1 + \epsilon^{(2)})k_x^2 + (1 + \epsilon^{(1)})k_y^2 + k_z^2 \right] V_{P0}^2,$$

$$a_1 = \left\{ 2(\epsilon^{(2)} - \delta^{(2)})k_x^2k_z^2 + 2(\epsilon^{(1)} - \delta^{(1)})k_y^2k_z^2 
+ (1 + 2\epsilon^{(2)})(1 \epsilon^{(1)})k_x^2k_y^2 \right\} V_{P0}^4,$$

$$a_0 = \left[ (1 + 2\epsilon^{(2)})^2(1 + 2\delta^{(3)}) - 2(1 + 2\epsilon^{(2)})\sqrt{1 + 2\delta^{(2)}} \sqrt{1 + 2\delta^{(1)}} \right] k_x^2k_y^2k_z^2 V_{P0}^6.$$

(2.13)

Introducing another set of parameters,

$$V_{Pz} = V_{P0},$$

(2.14)

$$V_{Px} = V_{P0}\sqrt{1 + 2\epsilon^{(2)}},$$

(2.15)

$$V_{Py} = V_{P0}\sqrt{1 + 2\epsilon^{(1)}},$$

(2.16)

$$V_{nmo}^{(1)} = V_{P0}\sqrt{1 + 2\delta^{(1)}},$$

(2.17)

$$V_{nmo}^{(2)} = V_{P0}\sqrt{1 + 2\delta^{(2)}},$$

(2.18)

$$V_{nmo}^{(3)} = V_{P0}\sqrt{1 + 2\delta^{(3)}},$$

(2.19)

yields the coefficients of the cubic equation 2.12:

$$a_2 = - \left[ V_{Px}^2k_x^2 + V_{Py}^2k_y^2 + V_{Pz}^2k_z^2 \right],$$

$$a_1 = V_{Pz}^2[V_{Px}^2 - (V_{nmo}^{(2)})^2]k_x^2k_z^2 + V_{Pz}^2[V_{Py}^2 - (V_{nmo}^{(1)})^2]k_y^2k_z^2 + V_{Px}^2[V_{Py}^2 - (V_{nmo}^{(3)})^2]k_x^2k_y^2,$$

(2.20)

$$a_0 = \left\{ V_{Pz}^2V_{Px}^2 \left[ (V_{nmo}^{(3)})^2 + (V_{nmo}^{(1)})^2 \right] - 2V_{Pz}^2V_{Px}^2V_{nmo}^{(1)}V_{nmo}^{(2)}V_{nmo}^{(3)} 
+ V_{Pz}^2V_{Py}^2(V_{nmo}^{(2)})^2 - V_{Pz}^2V_{Py}^2V_{Pz}^2 \right\} k_x^2k_y^2k_z^2.$$

Notice that the P-wave dispersion relation corresponds to the largest real root of the cubic equation 2.12, which can be solved analytically (Appendix A). Once the dispersion relation
is solved, the mixed-domain operator is simply
\[ \Phi(x, k) = \omega^2. \tag{2.21} \]

The next step is to solve the generic-form mixed-domain wavefield extrapolator \(2.1\) numerically. Note that for a spatially invariant operator \(\Phi(x, k) = \Phi(k)\), equation \(2.1\) reduces to a system of ordinary differential equations with the time variable \(t\), which has the formal solution
\[ u(k, t \pm \Delta t) = e^{\pm i \sqrt{\Phi(k)} \Delta t} u(k, t). \tag{2.22} \]
Adding the outgoing and incoming solutions of equation \(2.22\), one arrives at the time-stepping formula:
\[ u(k, t + \Delta t) + u(k, t - \Delta t) = 2 \cos \left( \sqrt{\Phi(k)} \Delta t \right) u(k, t). \tag{2.23} \]
Applying the Fourier transforms to both sides of equation \(2.23\), we obtain the space-domain wavefields:
\[ u(x, t + \Delta t) + u(x, t - \Delta t) = \mathcal{F}^{-1}\left[ 2 \cos \left( \sqrt{\Phi(k)} \Delta t \right) \mathcal{F}[u(x, t)] \right], \tag{2.24} \]
where \(\mathcal{F}[\cdot]\) and \(\mathcal{F}^{-1}[\cdot]\) denote the forward and inverse Fourier transforms, respectively. When the mixed-domain operator \(\Phi(x, k)\) varies in space, the time-stepping formula \(2.24\) provides only an approximate solution to equation \(2.1\). Solving equation \(2.24\) is time-consuming because the number of inverse FFT’s is equal to the number of the spatial grid points. I use two techniques described below to speed up this computation: low-rank decomposition (Fomel et al., 2013) and the generalized pseudospectral method (Fowler et al., 2015; Fowler and Lapilli, 2012).
2.1 Wavefield Simulator based on Low-rank Decomposition

The low-rank decomposition approach first discretizes the cosine term of equation 2.24 into a matrix:

\[
\cos(\sqrt{\Phi} \Delta t) = \begin{bmatrix}
\cos(\sqrt{\Phi(x_1, k_1)} \Delta t) & \cdots & \cos(\sqrt{\Phi(x_1, k_N)} \Delta t) \\
\vdots & \ddots & \vdots \\
\cos(\sqrt{\Phi(x_M, k_1)} \Delta t) & \cdots & \cos(\sqrt{\Phi(x_M, k_N)} \Delta t)
\end{bmatrix}.
\] (2.25)

This matrix is called a “propagator” and is iteratively applied to the wavefield during wave propagation. The matrix 2.25 has a low-rank feature provided that \( \Delta t \) is sufficiently small. In other words, the discretized matrix and its Hermitian have a large null space. This sparsity feature makes it possible to represent its column and row spaces using a relatively small number of column and row vectors. Although singular value decomposition (SVD) is the standard choice to select those vectors, it is impractical because the dimension of the matrix for 3D problems is extremely large (typically with the number of rows and columns on the order of \( 10^9 \)). A cheaper way to obtain those vectors is based on a randomized algorithm, which performs sparse matrix decomposition by selecting certain columns and rows of the original matrix. Symbolically, the decomposition takes the form of

\[
\cos(\sqrt{\Phi} \Delta t) \equiv W = U \Lambda V,
\] (2.26)

where \( W \) is the \( M \times N \) propagator matrix, \( U \) is the \( M \times m \) matrix of selected columns, \( V \) is the \( n \times N \) matrix of selected rows, \( \Lambda \) is a \( m \times n \) full matrix of relatively small size, where \( m \) and \( n \) are called the approximate numerical row and column rank of the matrix \( W \). Here \( m \ll M \) and \( n \ll N \). The obvious differences between this decomposition and SVD are that the columns of \( U \) are a subset of the columns of \( W \) rather than the eigenvectors of \( WW^\dagger \), the rows of \( V \) are a subset of the rows of \( W \) rather than the eigenvectors of \( W^\dagger W \), and \( \Lambda \) is a small full matrix rather than a diagonal matrix consisting of the eigenvalues obtained by SVD.
In a typical implementation, one often multiplies the matrices $U$ and $\Lambda$:

$$L \equiv U \Lambda, \quad \text{(2.27)}$$

which yields the low-rank decomposition:

$$\cos(\sqrt{\Phi} \Delta t) \equiv W = L \mathbf{R}, \quad \text{(2.28)}$$

where $\mathbf{R}$ coincides with $\mathbf{V}$ (2.26). With the decomposition of the propagator matrix, one can iteratively propagate the wave field along the time axis:

$$u(x, t + \Delta t) + u(x, t - \Delta t) = L \mathcal{F}^{-1} \left[ \mathbf{R} \mathcal{F}[u(x, t)] \right], \quad \text{(2.29)}$$

where $\mathbf{R}$ is applied in the wavenumber domain and $L$ is applied in the spatial domain.

### 2.2 Generalized Pseudospectral Wavefield Simulator

Low-rank decomposition methods can accurately simulate wave propagation since they do not involve any approximations of the corresponding dispersion relations. However, the matrix decomposition is numerical and the decomposed matrices cannot be explicitly expressed in terms of the medium parameters. This causes a problem for adjoint-state techniques, where the wave equation needs to be differentiated with respect to the medium parameters.

The pseudospectral method (Kosloff and Baysal, 1982) provides an efficient way to simulate wavefields accurately while maintaining the explicit form of the wave equation. Its extension to orthorhombic media has been explored by Fowler and Lapilli (2012), who proposed the generalized pseudospectral method. That method approximates the derivatives using global basis functions, rather than local finite-differences. First, the cosine term in equation 2.24 is expanded in a linear Taylor series:

$$\cos \left( \sqrt{\Phi} \Delta t \right) \approx 1 - \frac{1}{2} (\Delta t)^2 \Phi. \quad \text{(2.30)}$$

The time-stepping formula then becomes:

$$u(x, t + \Delta t) + u(x, t - \Delta t) = 2u(x, t) - (\Delta t)^2 \mathcal{F}^{-1} \left[ \Phi \mathcal{F}[u(x, t)] \right]. \quad \text{(2.31)}$$
Because the mixed-domain operator $\Phi(x, k)$ involves a certain form of spatial derivatives, equation 2.31 implements the generalized pseudospectral method with Fourier basis functions. However, application of equation 2.31 is still hampered by the fact that the operator $\Phi(x, k)$ varies spatially and appears inside the inverse Fourier transform. To use fast Fourier transforms, the mixed-domain operator must be represented in separable form:

$$\Phi(x, k) = \sum_i f_i(x) g_i(k), \quad (2.32)$$

where $f_i(x)$ and $g_i(k)$ are the pure spatial- and wavenumber-domain operators, respectively.

If we consider an acoustic orthorhombic medium with the symmetry planes that coincide with the Cartesian coordinate planes, the separable mixed-domain operator takes the form (see Appendix B):

$$\Phi(x, k) \approx V_{Px}^2 k_x^2 + V_{Py}^2 k_y^2 + V_{Pz}^2 k_z^2 - \frac{V_{Px}^2 (V_{Py}^2 - (V_{nmo}^{(3)})^2) k_x^2 k_y^2}{V_t^2 k^2}$$

$$- \frac{V_{Px}^2 (V_{Py}^2 - (V_{nmo}^{(2)})^2) k_x^2 k_z^2}{V_t^2 k^2}$$

$$- \frac{V_{Pz}^2 (V_{Py}^2 - (V_{nmo}^{(1)})^2) k_x^2 k_z^2}{V_t^2 k^2}, \quad (2.33)$$

where $V_r$ is a reference velocity, $V_{Px}$, $V_{Py}$, and $V_{Pz}$ are the P-wave velocities in the coordinate directions, and $V_{nmo}^{(i)}$ ($i = 1, 2, 3$) are the P-wave NMO velocities. The velocities $V_{nmo}^{(1)}$ and $V_{nmo}^{(2)}$ are measured in the $x_1$- and $x_2$- directions, respectively, above a horizontal orthorhombic layer. The $V_{nmo}^{(3)}$ is defined by Fowler and Lapilli (2012) in a similar fashion (equations 2.14-2.19). Once the mixed-domain operator is separated into the pure spatial- and wavenumber-domain operators (equation 2.32), the corresponding time-stepping formula can be expressed as:

$$u(x, t + \Delta t) + u(x, t - \Delta t) = 2u(x, t) - (\Delta t)^2 \sum_i f_i(x) \mathcal{F}^{-1} \left[ g_i(k) \mathcal{F} [u(x, t)] \right]. \quad (2.34)$$
2.3 Absorbing Boundary Condition

An important component of the numerical simulation is boundary conditions. For low-rank-based and generalized pseudospectral simulators, the absorbing boundary condition can be implemented by adding an exponentially decaying term to the wavefield after applying the propagators:

\[ u(k, t \pm \Delta t) = e^{\mp \alpha(x)} e^{\pm i \sqrt{\Phi(k)} \Delta t} u(k, t), \quad (2.35) \]

where \( \alpha(x) \) is the damping profile with nonzero values on the boundary. Adding the outgoing and incoming solutions yields the two-step time extrapolation formula in the spatial domain:

\[ u(x, t + \Delta t) + e^{-2\alpha(x)} u(x, t - \Delta t) = e^{-\alpha(x)} \mathcal{F}^{-1} \left[ 2 \cos \left( \sqrt{\Phi} \Delta t \right) \mathcal{F}[u(x, t)] \right]. \quad (2.36) \]

Following the approach discussed in the last two sections, one can arrive at the two-step time-stepping formula for low-rank decomposition extrapolators:

\[ u(x, t + \Delta t) + e^{-2\alpha(x)} u(x, t - \Delta t) = e^{-\alpha(x)} \mathcal{L} \mathcal{F}^{-1} \left[ \mathcal{R} \mathcal{F}[u(x, t)] \right]. \quad (2.37) \]

For generalized pseudospectral extrapolators, the corresponding equation is:

\[ u(x, t + \Delta t) + e^{-2\alpha(x)} u(x, t - \Delta t) = e^{-\alpha(x)} \left\{ 2u(x, t) - (\Delta t)^2 \sum_i f_i(x) \mathcal{F}^{-1} \left[ g_i(k) \mathcal{F}[u(x, t)] \right] \right\}. \quad (2.38) \]

2.4 Numerical Examples

Software verification and validation is an important aspect of computation-related research. In order to verify the two wavefield simulators, I use a constant-parameter model. These results should be calibrated through comparisons with analytic solutions. However, it is difficult to derive analytic wavefield solutions in acoustic orthorhombic media. I verify the developed software by showing that the wavefield solutions do not contain shear modes, and that the two simulators produce similar wavefields withina relatively small error. Furthermore, the traveltimes of the modeled P-waves are compared with fast-marching solutions of
the P-wave eikonal equations.

I test the two wavefield simulators with and without absorbing boundaries on a homogeneous model. The model is described by the parameters: $V_{Pz} = 2.25$ km/s, $V_{Px} = 2.99$ km/s, $V_{Py} = 2.76$ km/s, $V_{nmo}^{(1)} = 2.59$ km/s, $V_{nmo}^{(2)} = 2.66$ km/s, and $V_{nmo}^{(3)} = 3.22$ km/s. The corresponding Tsvankin’s anisotropy parameters are $\varepsilon^{(1)} = 0.256$, $\varepsilon^{(2)} = 0.384$, $\delta^{(1)} = 0.16$, $\delta^{(2)} = 0.20$, and $\delta^{(3)} = 0.08$. For generalized pseudospectral simulators, the reference velocity is set to $V_r = \sqrt[3]{V_{Pz}V_{Px}V_{Py}} = 2.65$ km/s. The model has a grid dimension of $n_z \times n_x \times n_y = 200 \times 200 \times 200$, with a grid spacing of $dz = dx = dy = 0.01$ km. I use a source function with peak frequency of 20 Hz and time sampling of 1 ms. The source is located at the center of the model. Figure 2.1(a) and Figure 2.1(b) show the wavefield snapshots simulated using the generalized pseudospectral method and low-rank decomposition, respectively. Those wavefields only contain pure P waves, with no shear wave artifacts (Grechka et al., 2004). Both wavefield simulators generate kinematically accurate P-wavefields, as confirmed by comparison with the traveltimes computed from the eikonal equation (Figure 2.1).

Figure 2.2(a) and Figure 2.2(b) show shot gathers recorded at the surface. Since the absorbing boundary conditions are not applied, the recorded data after the first arrivals represent reflections from the boundaries. Figure 2.3(a) and Figure 2.3(b) show the same shot gathers, but obtained with absorbing boundary conditions. The wavefield snapshots and traces generated by generalized pseudospectral method and low-rank decomposition are close to one another within a small numerical error.
Figure 2.1: Wavefield snapshots for a homogeneous orthorhombic model computed with: (a) the generalized pseudospectral method, and (b) low-rank decomposition. Red dotted lines correspond to the P-wave traveltimes obtained from the eikonal equation.
Figure 2.2: Shot gathers computed at the surface of the model without absorbing boundary conditions. (a) The generalized pseudospectral method, and (b) low-rank decomposition.

Figure 2.3: Shot gathers computed at the surface of the model with absorbing boundary conditions. (a) The generalized pseudospectral method, and (b) low-rank decomposition.
CHAPTER 3
WAVEFORM INVERSION

Successful seismic waveform inversion requires accurate and efficient seismic wavefield simulators, proper medium parameterizations, well-designed misfit functionals, and efficient large-scale nonlinear optimization algorithms. In this chapter, I address these ingredients of waveform inversion in acoustic orthorhombic media with the exception of the wavefield simulators discussed in the previous chapter.

3.1 Misfit Functionals

Waveform inversion is performed by minimizing a certain misfit functional (objective function). The most commonly used choice is the \( \ell_2 \)-norm data difference. However, due to the high nonlinearity, the \( \ell_2 \)-norm data-difference misfit functionals are often minimized gradually with increasing frequency bandwidth. Such cascaded inversions assume the existence of ultra low-frequency data, which are often missing in seismic acquisition. Building background models for waveform inversion without ultra low-frequency data is challenging. Wu et al. (2014) and Luo and Wu (2015) present inversion with envelope-based misfit functionals, which successfully produce background velocities models without ultra low-frequency data.

In this thesis, I consider two types of misfit functionals: the classical \( \ell_2 \)-norm data difference and the \( \ell_2 \)-norm squared envelope difference. The \( \ell_2 \)-norm data difference is defined as

\[
J_{\text{dat}} = \frac{1}{2} \sum_{i \in \Gamma_x} \| d_i - d_{o,i} \|^2 ,
\]

where the subscript \( i \) denotes the data coordinate, \( \Gamma_x \) is an index set for the data coordinates, and \( d_i \) and \( d_{o,i} \) are the modeled and observed discrete-time data (respectively) for a given source-receiver pair that belongs to \( \mathbb{R}^{N_t} \). The data are obtained by applying a binning
operator $W$ to the wavefield:

$$d_i = \sum_j W_{ij} u_j .$$

(3.2)

The envelope-based functional is given by:

$$J_{\text{env}} = \frac{1}{2} \sum_{i \in \Gamma_x} ||e_i^2 - eo_i^2||^2 ,$$

(3.3)

where $e_i$ and $eo_i$ are the envelopes of the modeled and observed discrete data, respectively. The envelope of a continuous-time signal $d(t)$ is defined as:

$$e(t) = \sqrt{d^2(t) + \mathcal{H}[d]^2(t)} ,$$

(3.4)

where $\mathcal{H}[d](t)$ is the Hilbert transform of the signal. Different misfit functionals produce different adjoint-source functions used for modeling the adjoint variables. For the $\ell_2$-norm data difference, the adjoint source function is

$$f^a = d - do .$$

(3.5)

The adjoint source function for the envelope-based misfit functional is derived in Appendix C:

$$f^a = 2\{ \Delta e^2 \circ d - H (\Delta e^2 \circ Hd) \} ,$$

(3.6)

where

$$\Delta e^2 = e^2 - eo^2$$

(3.7)

is the squared envelope difference. The symbol “$\circ$” denotes the Hadamard (Schur) product (Davis, 1962) and $H$ is the Hilbert-transform matrix.

To demonstrate how the envelope-based misfit functional can help build macromodels for waveform inversion, I generate a shot gather for a single source using a modified version of the 3D SEG/EAGE overthrust model. Figure 3.1(a) displays a random selection of 16 traces from the predicted (red) and observed (green) data sets. Because the initial model significantly deviates from the actual one, the predicted synthetic data are cycle-skipped compared to the observed data, which implies that the objective function is highly nonlinear.
Direct use of this data difference as the residual would guide the optimization search toward a local minimum.

Figure 3.1: Comparison of the predicted (red) and observed (green) traces: (a) raw data, (b) data envelopes, and (c) squared envelopes.

For comparison, I compute the envelopes and squared envelopes of the predicted and observed data (Figure 3.1(b) and Figure 3.1(c)). Although the envelope functions seem to be cycle-skipped as well, they look much simpler than the original data, and the inversion
operating with the envelope functions should be better posed. Figure 3.2 shows the frequency content of the predicted and observed data, with the spectra averaged over the traces for all receivers. The lack of low frequencies in these spectra is one of the reasons for the wavenumber gaps in the inversion results. In contrast, the envelope functions shift the spectra toward low frequencies (yellow and green lines in Figure 3.2), which indicates that the $\ell_2$-norm envelope misfit functional could be used to either generate an accurate long-wavelength initial model for WI as well as to update the model during initial iterations.

Figure 3.2: Spectra of the predicted (blue) and observed (red) raw data. The spectra of the predicted (green) and observed (yellow) envelope data are shifted toward lower frequencies.

3.2 Adjoint Wavefield Propagation

Seismic waveform inversion is often performed using the adjoint methods because the cost of computing the Fréchet derivatives of the misfit functionals is prohibitively high. Such adjoint methods operate adjoint wavefield variables, which satisfy the so-called adjoint wave equations. If the wave equations used in the forward simulations are self-adjoint, the adjoint wave equations retain the same form. On the other hand, if the wave equations for forward simulation are not self-adjoint, such as the ones I use in this thesis (based on low-rank decomposition and generalized pseudospectral methods), the adjoint wave equations
have to be solved differently from the forward-modeling equations.

A common way to verify the correctness of forward and adjoint simulators is the dot-product test (Claerbout and Black, 2008). One approach to pass the dot-product test is by automatic differentiation (Griewank and Walther, 2008; Rall, 1981) that programs the adjoint simulator. The automatic differentiation is a powerful algorithm because it can deal with arbitrarily complicated forward simulators and there is no need to manually code up additional programs. However, state-of-art implementations of the automatic differentiation do not necessarily produce optimized codes and the resulting codes can sometimes be user-unfriendly. For the simple wavefield simulators I use in the thesis, the adjoint wave equation can be derived analytically and coded up in a straightforward way.

Since the forward wavefield simulators are basically successive matrix-vector multiplications applied to wavefield vectors, the numerical adjoint wavefield simulators represent the transposed matrices successively operating with wavefield vectors. The mixed-domain adjoint wave equations can be written in the following generic form:

$$
\partial_{tt}u(k, t) + \hat{\Phi}(x, k)u(k, t) = 0 ,
$$

(3.8)

where $\hat{\Phi}(x, k)$ is the numerical adjoint mixed-domain operator. In the case of the low-rank decomposition simulator, the adjoint mixed-domain operator is:

$$
\hat{\Phi}(x, k) = R^\top L^\top,
$$

(3.9)

where $^\top$ indicates the matrix transpose, and $L$ and $R$ are defined in equations 2.27 and 2.28. The corresponding time-stepping formula then becomes:

$$
u(x, t + \Delta t) + u(x, t - \Delta t) = F^{-1}[R^\top F[L^\top u(x, t)]] ,
$$

(3.10)

which is different from the forward time-stepping formula 2.29.

### 3.3 Medium Parameterization

The goal of seismic waveform inversion is to estimate the medium parameters by matching modeled and observed seismic data. Tsvankin (1997, 2012) shows that all kinematic signa-
tures of P-waves in orthorhombic media are fully controlled by six independent parameters, which were introduced in the previous chapter. However, one can use certain combinations of these parameters to facilitate the inversion. The optimal choice of parameterization is crucial in obtaining accurate inversion results (Kamath and Tsvankin, 2016; Masmoudi and Alkhalifah, 2016). Here, I perform the WI using the parameters

\[
V_{Pz}^2, V_{Px}^2, V_{Py}^2, V_{nmo1}^2, V_{nmo2}^2, V_{nmo3}^2
\]

(3.11)
defined in equations 2.14-2.19. This choice is based primarily on the convenience in computing the gradients. Analysis of the radiation patterns can help in choosing optimal parameterizations for specific acquisition geometries and inversion scenarios. The numerical example below illustrates how this parameterization addresses parameter trade-offs.

I employ relatively simple orthorhombic models with a Gaussian anomaly in each parameter (Figure 3.3). As illustrated in Figure 3.4(a), the anomalies for different parameter fields do not overlap. The wavefield is excited by an areal source array located at the surface (Figure 3.4(b)). To remove the influence of illumination on the inversion results, I put receivers at each grid point on all six faces of the model cube.

The WI gradients for the background (initial) model are computed using the adjoint-state method, as discussed above and in section 3.4. Because the gradients govern the spatial positions and relative magnitudes of model updates, they help evaluate the performance of the chosen parameterization. Ideally, the gradient for each parameter should be nonzero only in the area of its Gaussian anomaly, which would imply the absence of parameter trade-offs.

The gradients for the parameterization (equation 3.11) computed with the classical \(\ell_2\)-norm data difference are shown in Figure 3.5. Although the gradients generally concentrate near the Gaussian anomalies, they are somewhat smeared in space, especially for the parameters \(V_{nmo1}^2, V_{nmo2}^2,\) and \(V_{nmo3}^2\). Therefore, it is difficult to identify the precise locations of the anomalies for the three NMO velocities, which would create trade-offs in the inversion. This is an indication of coupling or trade-offs between the medium parameters. Fortunately, the gradients for the P-wave velocities \(V_{Pz}, V_{Px},\) and \(V_{Py}\) show clear Gaussian-shape patterns
Figure 3.3: Models for the six parameters of an acoustic orthorhombic medium. Gaussian anomalies are embedded in a background medium with linearly increasing velocities.

Figure 3.4: (a) Six Gaussian anomalies from Figure 3.3 plotted together. (b) Projections of the anomalies from Figure 3.4(a) and the array of sources (red dots) at the surface.

at the locations of the anomalies, which indicates that they are better constrained than the NMO velocities.
Figure 3.5: WI gradients for the model from Figure 3.3 computed with the $\ell_2$-norm data difference. The gradients correspond to the background model in Figure 3.3 with the linearly increasing velocities.

3.4 Waveform Inversion Gradients

To compute the gradients of the objective function using the adjoint-state method (Plessix, 2006; Tromp et al., 2005), one augments the misfit functional as:

$$\chi = \mathcal{J} - \langle \lambda, F \rangle ,$$

where the symbol $\langle \cdot, \cdot \rangle$ denotes the inner product in the $L_2$-space (to which the state and adjoint variables belong), and $F$ is the discretized state equation:

$$F = \partial_t u(k, t) + \Phi(x, k) u(k, t) - f^s = 0 ,$$

where $f^s$ is the source term. The adjoint variable $\lambda$ satisfies the discretized adjoint equation:

$$F^\top = \partial_t \lambda(k, t) + \Phi(x, k)\lambda(k, t) - f^a = 0 .$$
The gradients of the original misfit functional are then derived by setting the derivatives of the augmented misfit functional (equation 3.12) to zero:

$$\frac{\partial \chi}{\partial m} = \frac{\partial J}{\partial m} - \left\langle \lambda, \frac{\partial F}{\partial m} \right\rangle = 0. \quad (3.15)$$

Since the low-rank decomposition simulator for acoustic orthorhombic media does not produce a closed-form wave equation, I propose to use the wave equation obtained with the generalized pseudospectral method to compute the term $\frac{\partial F}{\partial m}$ in equation 3.15, which is required to obtain the gradients. With the separable mixed-domain operator (equation 2.33) derived from generalized pseudospectral method, substituting the parameters $V^2_{rPx}$, $V^2_{rPy}$, $V^2_{rPz}$, $V^2_{nmo1}$, $V^2_{nmo2}$, and $V^2_{nmo3}$ for $m$ yields

$$\frac{\partial J}{\partial (V^2_{rPx})} = \left\langle \lambda, \left( k_x^2 - \frac{V^2_{Py} - V^2_{nmo3}}{V^2_{r}} k_y^2 k_z^2 - \frac{V^2_{Pz}}{V^2_{r}} k_y^2 k_z^2 \right) \frac{V^2_{r}}{k^2} \right\rangle u, \quad (3.16)$$

$$\frac{\partial J}{\partial (V^2_{rPy})} = \left\langle \lambda, \left( k_y^2 - \frac{V^2_{Px} k_x^2 k_z^2}{V^2_{r}} - \frac{V^2_{Pz}}{V^2_{r}} k_y^2 k_z^2 \right) \frac{V^2_{r}}{k^2} \right\rangle u, \quad (3.17)$$

$$\frac{\partial J}{\partial (V^2_{rPz})} = \left\langle \lambda, \left( k_z^2 - \frac{V^2_{Px} - V^2_{nmo2}}{V^2_{r}} k_x^2 k_y^2 - \frac{V^2_{Py} - V^2_{nmo1}}{V^2_{r}} k_y^2 k_z^2 \right) \frac{V^2_{r}}{k^2} \right\rangle u, \quad (3.18)$$

$$\frac{\partial J}{\partial (V^2_{nmo1})} = \left\langle \lambda, \left( \frac{V^2_{Pz}}{V^2_{r}} k_y^2 k_z^2 \right) \frac{V^2_{r}}{k^2} \right\rangle u, \quad (3.19)$$

$$\frac{\partial J}{\partial (V^2_{nmo2})} = \left\langle \lambda, \left( \frac{V^2_{Px} k_z^2 k_y^2}{V^2_{r}} \right) \frac{V^2_{r}}{k^2} \right\rangle u, \quad (3.20)$$

$$\frac{\partial J}{\partial (V^2_{nmo3})} = \left\langle \lambda, \left( \frac{V^2_{Px} k_x^2 k_y^2}{V^2_{r}} \right) \frac{V^2_{r}}{k^2} \right\rangle u, \quad (3.21)$$

where $u$ is the wavefield variable found from the state equation 3.13, and $\lambda$ is the wavefield variable from the adjoint equation 3.14.

To verify the derived gradient formulae, I apply both the data-difference and envelop-based misfit functionals for the model in Figure 3.6. The data are generated for 16 shots (red dots) in Figure 3.7 at every grid point on the horizontal surface. Figure 3.8 shows the initial models used for computing the gradients. Figure 3.9 and Figure 3.10 display the gradients obtained using the classical data-difference and squared envelope misfit functionals. Both gradients have substantial values only in the shallow part of the model because the initial
Figure 3.6: Orthorhombic medium obtained from the SEG/EAGE overthrust model. The velocities are scaled from the original P-wave isotropic velocity field.

Figure 3.7: Horizontal projection of the source locations (red dots), which are on the surface of the model from Figure 3.6.
velocity fields are quite smooth and most of the modeled energy represents diving waves. The gradients computed with the data-difference functional contain higher-wavenumber information, which may cause problems during the early stages of WI. In contrast, the gradients produced by the squared envelope misfit functional are more smooth and have a lower-wavenumber content, which should help in updating long-wavelength macro models for later iterations of WI.

Figure 3.8: Initial parameter fields used to compute the WI gradients for the model in Figure 3.6. The initial velocities are smoothed and deviate significantly from the actual values.

3.5 Synthetic Example of Waveform Inversion

Here, WI is applied to synthetic data using an iterative gradient-based algorithm. I employ the limited memory variable metric method with box bounds (Benson and Moré, 2001; Thiébaut, 2002), which represents a version of the BFGS methods (Liu and Nocedal, 1989).
Figure 3.9: Gradients for the model from Figure 3.6 computed with the $\ell_2$-norm data difference functional. The gradients correspond to the initial model Figure 3.8.

Figure 3.10: Gradients for the model from Figure 3.6 computed with the $\ell_2$-norm squared envelope functional. The gradients correspond to the initial model Figure 3.8.
Figure 3.11: Orthorhombic medium modified from the SEG/EAGE overthrust model. The velocities are scaled from the original P-wave isotropic velocity field.

Figure 3.12: Horizontal projection of the source locations (red dots), which are on the surface of Figure 3.11.
For this test, I use a modified portion of the SEG/EAGE overthrust model (Figure 3.11). The observed data consist of 25 shot gathers, which are generated with the generalized pseudospectral method. To avoid the “inverse crime,” the low-rank decomposition simulator is used in the inversion process. Figure 3.12 shows the shot locations on the surface; the receivers are also located on the surface. I assume that the source function is known, which has peak frequency around 20 Hz and a frequency bandwidth of approximately 50 Hz.

The initial models for WI are obtained by smoothing the actual velocity fields (Figure 3.13). The WI is performed using the multifrequency-band approach with five ranges: 0–8 Hz, 0–16 Hz, 0–32 Hz, and 0–64 Hz. I stop the iterations for each frequency band when the misfit functionals no longer decrease significantly. The total number of iterations reached 102 with over 600 gradient evaluations. The inverted models (Figure 3.14) show significant improvement compared to the initial ones, with important geological structures such as faults, synclines, anticlines and low-velocity zones being better resolved.

The velocities $V_{Pz}$, $V_{Py}$, and $V_{Py}$ are better constrained than the NMO velocities for this particular configuration of surface data, which confirms the results of the gradient computation for the model with Gaussian-shape anomalies (Figure 3.5). I put masks around source positions to avoid large spurious updates in those areas, which is a typical practice in WI. The spurious updates are more obvious in the NMO velocities, especially around $z = 0.2$ km in the inverted $V_{nmo2}$ model.

To examine the inversion results, I plot vertical $V_{Pz}$-profiles from the actual, initial, and inverted models (Figure 3.15) at the locations marked red dots in Figure 3.15. Although not perfectly recovered, the inverted velocities are much closer to the actual values compared to the initial models. In field-data applications, one does not have access to the actual models, so it is important the verify the inversion results using other metrics. Since WI is a data-fitting problem, the inversion results can be evaluated by comparing data residuals before and after the inversion. Figure 3.17(a) and Figure 3.17(b) show the data residuals for a typical shot gather computed with the initial and inverted models. The most significant decrease
Figure 3.13: Initial parameter fields used to perform WI for the model in Figure 3.11. The initial velocities are smoothed and deviate significantly from the actual values.

Figure 3.14: Inverted parameter fields after 102 iterations of the inversion algorithm.
of the data residual appear in the far-offset traces (around $x = 7$ km). However, the data residuals at near-offset traces do not improve too much. This is because conventional WI mostly fits the diving wave energy. Quantitatively, the $\ell_2$-norm of data residuals decreases from 27.3 to 15.4 after the inversion.

Overall, the synthetic example demonstrates the potential of obtaining high-resolution anisotropy parameters from WI. The data fit is not perfect because of parameter trade-offs that produce a complicated, multimodal objective function. Nevertheless, there is a significant decrease in the data misfit after performing WI, and the inverted model parameters are much closer to the actual values compared to those from the initial model.

![Figure 3.15: Vertical profiles of the true (black), initial (red), and inverted (blue) velocity $V_{Pz}$ at locations shown in Figure 3.16.](image-url)
Figure 3.16: Locations of velocity profiles in Figure 3.15.

Figure 3.17: Data residuals using (a) the initial, and (b) the inverted models.
CHAPTER 4
CONCLUSIONS AND RECOMMENDATIONS

In this chapter I summarize the thesis work and provides some recommendations for future work.

4.1 Conclusions

I developed algorithms for modeling and inversion in acoustic orthorhombic media, and employed them to study the feasibility of 3D waveform inversion of wide-azimuth surface data. Wavefield simulations are carried out with the mixed-domain extrapolator using low-rank decomposition. The gradient computation, however, is based on the wave equation obtained by the generalized pseudospectral method. The choice of wavefield extrapolator is determined by the superior accuracy and stability of the method based on low-rank decomposition. I performed multiparameter inversion for acoustic orthorhombic models described by the P-wave velocities $V_{Pz}^2$, $V_{Px}^2$, and $V_{Py}^2$ in the coordinate directions and the NMO velocities $V_{nmo1}^2$, $V_{nmo2}^2$, and $V_{nmo3}^2$.

Synthetic examples reveal significant trade-offs among the parameters. For surface seismic data, $V_{Pz}$, $V_{Px}$, and $V_{Py}$ are generally better constrained by WI than the NMO velocities. I demonstrate that envelope-based inversion can potentially help reconstruct low-wavenumber model parameters when the initial model is highly inaccurate. The quasi-Newton gradient descent optimization methods perform reasonably well in the synthetic testing in terms of their convergence rate and computational cost.

A synthetic example for a realistic geological structure shows that WI generates high-resolution $V_{Pz}$, $V_{Px}$, and $V_{Py}$ models in acoustic orthorhombic media, provided that the initial velocity fields are sufficiently accurate. Although the low-wavelength NMO velocities are well-constrained by reflection traveltimes, their higher-frequency components are not
accurately resolved by WI. Overall, this work proves that there is great potential in improving velocity models by performing WI for orthorhombic media.

4.2 Recommendations

Depending on the type of data used in the inversion, the acoustic approximation may break down. Incorporating the shear-wave information from multicomponent data can help constrain the anisotropy parameters which do not influence P-wave kinematics. However, moving beyond the acoustic approximation for orthorhombic media requires solving the solving 3D elastic wave equation. Aside from the need for more computational power, the large model space may cause additional parameter trade-offs and numerical challenges even for state-of-the-art optimization algorithms.

As my synthetic WI examples showed, there are significant trade-offs among the parameters of acoustic orthorhombic media. To choose the optimal parameter set for a given acquisition geometry, one can analyze the sensitivity kernels (Liu and Tromp, 2008) and radiation (scattering) patterns (Alkhalifah and Plessix, 2014; Gholami et al., 2013) based on the Born approximation.

Comparison of the observed and synthetic seismograms remains a topic of ongoing research. The conventional $\ell_2$-norm data-difference misfit functionals are highly nonconvex, which causes cycle-skipping issues. Standard signal-processing techniques (e.g., time-frequency and wavelet analysis), and more advanced statistical approaches might help mitigate the cycle-skipping and obtain better inversion results.

The efficiency and stability of large-scale anisotropic WI depend on the performance of nonlinear optimization solvers. Current research shows that the limited-memory quasi-Newton methods generally outperform other gradient-based optimization algorithms in large-scale waveform inversion. It would be beneficial to incorporate efficient optimization algorithms with nonlinear constraints into seismic WI methods.

As a general inverse problem, WI should benefit from preconditioning and regularization. In particular, good preconditioners improve the convergence of numerical optimization. Di-
agonal scaling is often used to account for amplitude factors including geometrical spreading, source illumination, etc. Most preconditioners involve some information about the Hessian of the misfit functional. However, preconditioners are problem- and model-dependent, and need to be updated during WI iterations. Regularization explicitly adds a priori information into the misfit functionals, which improves the convexity and makes the inverse problem less nonunique. Standard techniques like Tikhonov and total-variation regularizations can be helpful in making WI results more robust.

Another important issue is how to quality-control the inversion results. Conventional WI techniques only generate one final model as the output, which makes it essentially a data-fitting problem. Data assimilation and uncertainty quantification are important in understanding and assessing the inversion results and, therefore, help lower the risk in exploration applications.

Finally, applying anisotropic WI to field data faces a number of challenges. Noise suppression and including available information about anisotropy parameters are especially important in obtaining reliable inversion results.
REFERENCES CITED


Sirgue, L., J. Etgen, and U. Albertin, 2008, 3D frequency domain waveform inversion using time domain finite difference methods: Presented at the 70th EAGE Conference and Exhibition.


———, 2006b, An anisotropic acoustic wave equation for VTI media: Presented at the 68th EAGE Conference & Exhibition.
Here, I use an analytic solution for the roots of a cubic polynomial equation:

\[ x^3 + ax^2 + bx + c = 0. \]  \hspace{1cm} (A.1)

A convenient method of solving equation \( A.1 \) was originally published by Gerolamo Cardano in the 16th century. It operates with the coefficients

\[ p = b - \frac{a^2}{3}, \]  \hspace{1cm} (A.2)

and \[ q = \frac{2a^3 - 9ab + 27c}{27}. \]  \hspace{1cm} (A.3)

The discriminant of the cubic equation is:

\[ \Delta = \left( \frac{p}{3} \right)^3 + \left( \frac{q}{2} \right)^2. \]  \hspace{1cm} (A.4)

If \( \Delta > 0 \), the cubic equation has one real and two imaginary roots:

\[ x_1 = m + n - \frac{a}{3}, \]
\[ x_{2,3} = -0.5(m + n) \pm (m - n)\frac{\sqrt{3}}{2}i - \frac{a}{3}, \]  \hspace{1cm} (A.5)

where

\[ m = \sqrt[3]{-q/2 + \sqrt{\Delta}}, \]  \hspace{1cm} (A.6)
\[ n = \sqrt[3]{-q/2 - \sqrt{\Delta}}. \]  \hspace{1cm} (A.7)

If the discriminant goes to zero, the cubic equation has three real roots, two of which are identical:

\[ x_1 = 2\sqrt[3]{-q/2} - \frac{a}{3}, \]
\[ x_{2,3} = \sqrt[3]{q/2} - \frac{a}{3}. \]  \hspace{1cm} (A.8)
If the discriminant is negative, the equation has three real roots:

\[ x_1 = 2\sqrt{-p/3} \cos\left(\frac{\phi}{3}\right) - \frac{a}{3}, \]
\[ x_2 = 2\sqrt{-p/3} \cos\left(\frac{\phi + 2\pi}{3}\right) - \frac{a}{3}, \]
\[ x_3 = 2\sqrt{-p/3} \cos\left(\frac{\phi + 4\pi}{3}\right) - \frac{a}{3}, \]

(A.9)

where

\[ \phi = \cos^{-1}\left(-\frac{q}{2\sqrt{(-p/3)^2}}\right). \]

(A.10)

The P-wave dispersion relation is obtained by finding the largest real root, which produces the highest velocity.
APPENDIX B - SEPARABLE MIXED-DOMAIN OPERATOR IN ACOUSTIC ORTHORHOMBIC MEDIA

I consider an orthorhombic model for which the Cartesian coordinate planes coincide with the planes of symmetry. The P-wave dispersion relation in orthorhombic media represents a cubic equation in $\omega^2$:

$$\omega^6 + a \omega^4 + b \omega^2 + c = 0,$$

where $a$, $b$, and $c$ depend on the elements of the Christoffel matrix. The analytic solutions for phase velocity given, for example, by Tsvankin (1997), yield the following dispersion relation for P-waves:

$$\omega_p^2 = -\frac{2}{3} \left( \mu \sqrt{1 - 3b/a^2} + \frac{1}{2} \right),$$

(B.2)

where

$$\mu = \cos \left\{ \frac{1}{3} \cos^{-1} \left[ -\frac{\sqrt{27}}{2} q(-d)^{-3/2} \right] \right\},$$

(B.3)

and

$$d = -\frac{a^2}{3} + b,$$

$$q = 2 \left( \frac{a}{3} \right)^3 - \frac{ab}{3} + c.$$  

(B.4)

To derive a separable form of the dispersion relation, I first expand $\mu$ and $\sqrt{1 - 3b/a^2}$ in a series with respect to the coefficient $a$ in equation B.2. Fowler and Lapilli (2012) and Fowler et al. (2015) present several approximations for such an expansion. Here, I use the following approximate dispersion relation:

$$\omega_p^2 = -a + b \frac{a}{a^2} - c \frac{a}{a^3} + \frac{b^2}{a^4} - \frac{3bc}{a^5} + \frac{2(b^3 + c^2)}{a^6} - \frac{10b^2c}{a^7} + \frac{5b(b^3 + 3c^2)}{a^8} - \frac{7c(5b^3 + c^2)}{a^9} + \frac{14b^2(b^3 + 6c^2)}{a^{10}} - \frac{42bc(3b^3 + 2c^2)}{a^{11}} - \frac{6(7b^6 + 70b^3c^2 + 5c^4)}{a^{12}} - \cdots.$$  

(B.5)
The term $a^{-1}$ has to be modified further for equation B.5 to be separable. I follow the procedure of Fowler and Lapilli (2012), who represent $a^{-1}$ as:

$$a^{-1} = -V_r^{-2}k^{-2}(1 - 2E)^{-1} = -V_r^{-2}k^{-2} \left(1 + 2E + 4E^2 + 8E^3 + \cdots\right),$$  \hfill (B.6)

where

$$E \equiv \frac{1}{2}(1 + \frac{a}{V_r^2k^2});$$  \hfill (B.7)

$V_r = \left(V_{P1}V_{P2}V_{P3}\right)^{1/3}$ is a reference velocity and $V_{Pi}$ ($i = 1, 2, 3$) are the P-wave velocities in the coordinate directions. Combining equations B.5 - B.7, substituting the expressions for $a, b, c$ in terms of the elements of the Christoffel matrix, and truncating the expansions after the linear terms, I arrive at the following separable form of the approximate P-wave dispersion relation for acoustic orthorhombic media:

$$\omega_p^2 = V_{P1}^2k_1^2 + V_{P2}^2k_2^2 + V_{P3}^2k_3^3 - \frac{V_{P3}^2(V_{P2}^2 - V_{nm01}^2)k_2^2k_3^2}{V_r^2}k^2 - \frac{V_{P3}^2(V_{P1}^2 - V_{nm02}^2)k_1^2k_3^2}{V_r^2}k^2 - \frac{V_{P1}^2(V_{P2}^2 - V_{nm03}^2)k_1^2k_2^2}{V_r^2}k^2,$$  \hfill (B.8)

where $V_{nm01}, V_{nm02},$ and $V_{nm03}$ are defined in the main text (equations 2.17 - 2.19). A more symmetric form of equation B.8 can be obtained by introducing new variables,

$$V_{p1}^{2} = V_{nm01}V_{P3},$$  \hfill (B.9)

$$V_{p2}^{2} = V_{nm02}V_{P3},$$

$$V_{p3}^{2} = V_{nm03}V_{P1}.$$  \hfill (B.9)

Substitution of the parameters from equation B.9 into equation B.8 leads to

$$\omega_p^2 = V_{P1}^2k_1^2 + V_{P2}^2k_2^2 + V_{P3}^2k_3^3 - \frac{V_{P2}^2V_{P3}^2 - V_{ps1}^2k_2^2k_3^2}{V_r^2}k^2 - \frac{V_{P1}^2V_{P3}^2 - V_{ps2}^2k_1^2k_3^2}{V_r^2}k^2 - \frac{V_{P1}^2V_{P2}^2 - V_{ps3}^2k_1^2k_2^2}{V_r^2}k^2.$$  \hfill (B.10)

The parameterization of acoustic orthorhombic media in terms of $V_{P1}, V_{P2}, V_{P3}, V_{p1}, V_{p2}, V_{p3}$ can be found in Chapman (2004) and Fowler et al. (2015). The associated mixed-domain operator $\Phi(x,k)$ can also be expressed in a separable form (see equation 2.32):

$$\Phi(x,k) = \sum_{i=1}^{3} \left( V_{P1}^2k_i^2 - \frac{V_{P2}^2V_{P3}^2}{V_r^2}k_i^2 \right).$$  \hfill (B.11)
APPENDIX C - ADJOINT SOURCE FOR ENVELOPE-BASED MISFIT FUNCTIONALS

Here, I derive the adjoint source for the squared-envelope-difference misfit functional. An analytic signal \(a(t)\) is defined as:

\[
a(t) = d(t) + i \mathcal{H}[d](t),
\]

where \(\mathcal{H}[\cdot]\) is the Hilbert transform, and \(d(t)\) is the original signal. The Hilbert transform of a time series with compact support can be written as the convolution of that series with the function \(h(t) = 1/(\pi t)\):

\[
\mathcal{H}[d](t) = h(t) \ast d(t).
\]

Because convolution is a linear operation, equation (C.1) can be discretized in time:

\[
a[l] = d[l] + i \sum_m H[l, m]d[m],
\]

where \(H[l, m]\) is the \(l\)-th row, \(m\)-th column element of the convolution matrix obtained from \(h[l - m]\). In vector notation,

\[
a = d + i \mathbf{H}d.
\]

The envelope of the original signal is given by:

\[
e(t) = \sqrt{d^2(t) + \mathcal{H}[d]^2(t)},
\]

and the discrete envelope function has the form:

\[
e[l] = \sqrt{d[l]d[l] + \sum_m H[l, m]d[m] \sum_n H[l, n]d[n]},
\]

where there is no summation over \(l\). The misfit functional based on \(\ell_2\)-norm squared envelope difference is:

\[
\mathcal{J} = \frac{1}{2} \sum_{i \in \Gamma_x} \| e_i^2 - e_{o_i}^2 \|^2,
\]
where $i$ is the index of spatial position and $\Gamma_x$ is the corresponding index set; $e$ and $eo$ are the envelopes of the synthetic and observed (recorded) data, respectively. To compute the adjoint sources used for adjoint modeling, the misfit functional is differentiated with respect to the wavefield variable $u$:

$$f^a[n] = \frac{\partial J}{\partial u[n]},$$

where $f^a[n]$ is the discrete-time adjoint source at the location of the receivers, and $u[n]$ is the discrete-time wavefield. Here, because I do not distinguish the wavefield $u[n]$ from the recorded data $d[n]$, the adjoint source can also be written as:

$$f^a[n] = \frac{\partial J}{\partial d[n]}.$$

Substituting the misfit functional C.7 into equation C.9 yields:

$$f^a[n] = \frac{\partial J}{\partial d[n]}$$

$$= \sum_l \frac{\partial J}{\partial e^2[l]} \frac{\partial e^2[l]}{\partial d[n]}$$

$$= \sum_l \Delta e^2[l] \frac{\partial e^2[l]}{\partial d[n]}$$

$$= \sum_l \Delta e^2[l] \left( 2d[l]\delta_n + 2H[l, q]\delta_q \sum_p H[l, p]d[p] \right)$$

$$= 2 \left[ \Delta e^2[n]d[n] + \sum_l H[l, n] \left( \Delta e^2[l] \sum_p H[l, p]d[p] \right) \right]$$

$$= 2 \left[ \Delta e^2[n]d[n] + \sum_l H^T[n, l] \left( \Delta e^2[l] \sum_p H[l, p]d[p] \right) \right],$$

where $\Delta e^2[l] = e^2[l] - eo^2[l]$ is the the squared envelope difference, and $\delta_n$ and $\delta_q$ are Kronecker’s symbolic deltas. The adjoint source can be written in vector notation:

$$f^a = 2 \left\{ \Delta e^2 \circ d + H^T (\Delta e^2 \circ Hd) \right\},$$

where $\Delta$ is the alternate vectorization operator.
where “⊙” denotes the Hadamard (Schur) product (Davis, 1962). The Hilbert transform is an anti-self-adjoint operator:

\[ H^\dagger[n,l] = -H[n,l], \tag{C.12} \]

and the adjoint source function becomes:

\[ f^a = 2 \left\{ \Delta e^2 \circ d - H(\Delta e^2 \circ H d) \right\}. \tag{C.13} \]

The corresponding continuous-time adjoint source function can be written as:

\[ f(t) = 2 \left\{ \Delta e^2(t) d(t) - \mathcal{H} \left[ \Delta e^2(t) \mathcal{H}[d(t)] \right] \right\}, \tag{C.14} \]

which is derived in a different way in Wu et al. (2014).