Multi-parameter Analysis and Inversion for Anisotropic Media Using the Scattering Integral Method

Dissertation by
Ramzi Djebbi

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

King Abdullah University of Science and Technology
Thuwal, Kingdom of Saudi Arabia

October, 2017
The dissertation of Ramzi Djebbi is approved by the examination committee.

Committee Chairperson: Prof. Tariq Alkhalifah
Committee Members: Prof. Ibrahim Hoteit, Prof. Taous Meriem Laleg-Kirati, Prof. Alexey Stovas
ABSTRACT

Multi-parameter Analysis and Inversion for Anisotropic Media Using the Scattering Integral Method

Ramzi Djebbi

The main goal in seismic exploration is to identify locations of hydrocarbons reservoirs and give insights on where to drill new wells. Therefore, estimating an Earth model that represents the right physics of the Earth’s subsurface is crucial in identifying these targets. Recent seismic data, with long offsets and wide azimuth features, are more sensitive to anisotropy. Accordingly, multiple anisotropic parameters need to be extracted from the recorded data on the surface to properly describe the model. I study the prospect of applying a scattering integral approach for multi-parameter inversion for a transversely isotropic model with a vertical axis of symmetry. I mainly analyze the sensitivity kernels to understand the sensitivity of seismic data to anisotropy parameters. Then, I use a frequency domain scattering integral approach to invert for the optimal parameterization.

The scattering integral approach is based on the explicit computation of the sensitivity kernels. I present a new method to compute the traveltime sensitivity kernels for wave equation tomography using the unwrapped phase. I show that the new kernels are a better alternative to conventional cross-correlation/Rytov kernels. I also derive and analyze the sensitivity kernels for a transversely isotropic model with a vertical axis of symmetry. The kernels structure, for various opening/scattering angles, highlights the trade-off regions between the parameters. For a surface recorded data, I show that the normal move-out velocity $v_n$, $\eta$ and $\delta$ parameterization is suitable for a simultaneous inversion of diving waves and reflections. Moreover, when seismic
data is inverted hierarchically, the horizontal velocity $v_h$, $\eta$ and $\epsilon$ is the parameterization with the least trade-off. In the frequency domain, the hierarchical inversion approach is naturally implemented using frequency continuation, which makes $v_h$, $\eta$ and $\epsilon$ parameterization attractive.

I formulate the multi-parameter inversion using the scattering integral method. Application to various synthetic and real data examples show accurate inversion results. I show that a good background $\eta$ model is required to accurately recover $v_h$. For 3-D problems, I promote a hybrid approach, where efficient ray tracing is used to compute the sensitivity kernels. The proposed method highly reduces the computational cost.
ACKNOWLEDGEMENTS

"I can no other answer make but thanks, and thanks, and ever thanks"~ William Shakespeare. To my supervisor, Prof. Tariq Alkhalifah, I would like to express my profound gratitude. His high-value expectations, enthusiasm, and motivation helped me improve my skills. His deep knowledge and understanding of seismic inversion helped me choose a research direction and accomplish the proposed tasks. I am also grateful to him for his financial support for many international conference trips. I hope that our scientific collaboration will continue during my professional career.

My gratitude goes to my dissertation committee members, Prof. Ibrahim Hoteit, Prof. Meriem Laleg-Kirati and Prof. Alexey Stovas. Their personal and technical advices helped me strengthen my research base and overcome many obstacles.

I am grateful to companions in my research group SWAG at KAUST for their help and support. Namely, Muhammad Zuberi, Yunseok Choi, Christos Saragiotis, Umair Bin Waheed, Babar Hasan Khan, Zedong Wu, Nabil Masmoudi, Mahesh Kalita, Ghada Sindi, Qiang Guo, Hanchen Wang, Zhendong Zhang and Christos Tzivanakis.

I would like to thank the developers and contributors to Madagascar open source software package (www.ahay.org). I also would like to extend my gratitude to CGG for making the real data available.

"Family is not an important thing. It’s everything."~Michael J. Fox. Nothing that could ever be said describes my gratitude to my wife, Khouloud. My gratitude to my brothers, Rami, Abir and Khalil for always being by my side. To my brother in law Khalil. I saved the most important persons for last. To my parents, Taher and Naima, all my achievements and successes would not have been possible without my parents’ love, prayers and endless support. I dedicate this dissertation and my success to them.
# TABLE OF CONTENTS

Examination Committee Page 2

Copyright 3

Abstract 4

Acknowledgements 6

List of Abbreviations 10

List of Figures 11

List of Tables 19

1 Introduction 20

1.1 Model parameter estimation ............................................ 21
    1.1.1 Traveltime tomography .......................................... 22
    1.1.2 Wave equation tomography ...................................... 22
    1.1.3 Full waveform inversion ......................................... 24

1.2 The inversion approach ................................................ 27
    1.2.1 Adjoint-state method ............................................ 27
    1.2.2 Scattering integral method ..................................... 28

1.3 Seismic Anisotropy ..................................................... 29

1.4 Multi-parameter anisotropic inversion .............................. 33

1.5 Thesis objectives and outline ....................................... 36

2 Traveltime sensitivity kernels for wave equation tomography using
   the unwrapped phase 40

2.1 Introduction ............................................................ 41

2.2 Theory ........................................................................ 44
    2.2.1 Cross-correlation and Rytov traveltime sensitivity kernels .. 46
    2.2.2 Unwrapped phase sensitivity kernels ............................ 48
    2.2.3 Relation between unwrapped phase and Rytov sensitivity kernels 50
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2 Theory</td>
<td>132</td>
</tr>
<tr>
<td>5.2.1 The scattering integral method</td>
<td>133</td>
</tr>
<tr>
<td>5.2.2 The hybrid inversion approach</td>
<td>134</td>
</tr>
<tr>
<td>5.3 Computational cost analysis</td>
<td>135</td>
</tr>
<tr>
<td>5.4 Numerical results</td>
<td>139</td>
</tr>
<tr>
<td>5.5 Conclusions</td>
<td>143</td>
</tr>
<tr>
<td>6 Application to real data</td>
<td>146</td>
</tr>
<tr>
<td>6.1 Data description</td>
<td>146</td>
</tr>
<tr>
<td>6.2 Multi-parameter inversion</td>
<td>146</td>
</tr>
<tr>
<td>7 Concluding Remarks and Future Work</td>
<td>157</td>
</tr>
<tr>
<td>7.1 Summary</td>
<td>157</td>
</tr>
<tr>
<td>7.1.1 Sensitivity kernels using the unwrapped phase</td>
<td>158</td>
</tr>
<tr>
<td>7.1.2 Analysis of the traveltime sensitivity kernels for VTI media</td>
<td>158</td>
</tr>
<tr>
<td>7.1.3 Frequency domain multi-parameter full waveform inversion for</td>
<td>159</td>
</tr>
<tr>
<td>acoustic VTI media</td>
<td>159</td>
</tr>
<tr>
<td>7.1.4 Hybrid frequency domain full waveform inversion using Born</td>
<td>161</td>
</tr>
<tr>
<td>sensitivity kernels</td>
<td>161</td>
</tr>
<tr>
<td>7.2 Future Research Work</td>
<td>161</td>
</tr>
<tr>
<td>References</td>
<td>164</td>
</tr>
<tr>
<td>Appendices</td>
<td>175</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS

CG            conjugate gradient
CIGs          common image gathers
FWI           full waveform inversion
HTI           transversely isotropic model with a horizontal axis of symmetry
l-BFGS        limited-memory BFGS
LU            lower-upper
MVA           migration velocity analysis
NMO           normal move-out
PML           perfectly matched layers
QC            quality control
RTM           reverse time migration
RWI           reflection waveform inversion
SI            scattering integral
TI            transversely isotropic
TOR           tilted orthorhombic
TTI           tilted transversely isotropic
VTI           transversely isotropic model with a vertical axis of symmetry
WEMVA         wave-equation migration velocity analysis
WET           wave equation tomography
<table>
<thead>
<tr>
<th>Number</th>
<th>Figure Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Sensitivity kernel for the traveltime tomography: the kernel is the ray-path (Figure from Rickett (2000)).</td>
<td>23</td>
</tr>
<tr>
<td>1.2</td>
<td>Finite frequency traveltime sensitivity kernel (Figure from Rickett (2000)).</td>
<td>24</td>
</tr>
<tr>
<td>1.3</td>
<td>Model inverted with FWI compared to traveltime tomography: (left) traveltime tomography model, (right) FWI model (Figure from Kapoor et al. (2012)).</td>
<td>25</td>
</tr>
<tr>
<td>1.4</td>
<td>Illustration of the cycle skipping problem (Figure from Virieux and Operto (2009)).</td>
<td>26</td>
</tr>
<tr>
<td>1.5</td>
<td>Illustration of different symmetries for anisotropic models: (a) VTI symmetry, (b) HTI symmetry, (c) TTI symmetry.</td>
<td>30</td>
</tr>
<tr>
<td>1.6</td>
<td>Illustration of orthorhombic anisotropic models. It has a VTI symmetry in one direction and HTI in the perpendicular direction (caused by fractures): (a) orthorhombic model, (b) rotated orthorhombic model.</td>
<td>31</td>
</tr>
<tr>
<td>1.7</td>
<td>Illustration of different symmetries for anisotropic models: (a) VTI symmetry, (b) HTI symmetry, (c) Elliptic symmetry, (d) TTI with tilt angle $\theta = 30^\circ$.</td>
<td>34</td>
</tr>
<tr>
<td>1.8</td>
<td>Radiation patterns for two VTI model parameterizations: (a) $(v_n, \eta, \delta)$ parameterization, $v_n$ in red, $\eta$ and $\delta$ in orange, (b) $(v_h, \epsilon, \eta)$ parameterization, $v_h$ in red, $\eta$ and $\epsilon$ in orange.</td>
<td>35</td>
</tr>
<tr>
<td>2.1</td>
<td>Two different scenarios. (a) Two Ricker wavelets, one of the two is advanced with $\delta t = 0.1$ s with no phase rotation. (b) Two Ricker wavelets, one of the two wavelets is advanced with $\delta t = 0.1$ s with a phase rotation $\pi/2$.</td>
<td>54</td>
</tr>
<tr>
<td>2.2</td>
<td>Crosscorrelation of the two scenarios depicted in figure 2.1. (a) No phase rotation, the maximum of the cross-correlation corresponds to $\delta t = 0.1$ s. (b) $\pi/2$ phase rotation, the maximum of the cross-correlation corresponds to $\delta t = 0.08$ s and cross-correlation is inaccurate.</td>
<td>55</td>
</tr>
</tbody>
</table>
2.3 Traveltime difference measurement using the instantaneous traveltime (unwrapped phase). (a) No phase rotation (b) \( \pi / 2 \) phase rotation. For both cases, the traveltime lag \( \delta t \) is given as a function of the frequency, for most of the frequencies the instantaneous traveltime is able to compute the exact traveltime difference. I pick the lag for the wavelet peak frequency \( f = 10 \text{Hz} \).

2.4 30 \text{Hz} single frequency traveltime sensitivity kernels. (a) Crosscorrelation. (b) Unwrapped phase (instantaneous traveltime).

2.5 30 \text{Hz} Broadband traveltime sensitivity kernels for \( v(z) \). (a) Crosscorrelation. (b) Unwrapped phase (instantaneous traveltime).

2.6 Slices at \( x = 4 \text{km} \) of the 30 \text{Hz} traveltime sensitivity kernels. (a) Crosscorrelation. (b) Unwrapped phase (instantaneous traveltime). The broadband slices in continuous blue line are overlaid to the monochromatic 30 \text{Hz} slices represented in dashed red line.

2.7 Smoothed Marmousi model.

2.8 30 \text{Hz} Broadband traveltime sensitivity kernels for the Marmousi model. (a) Crosscorrelation. (b) Unwrapped phase (instantaneous traveltime).

2.9 Slices at \( x = 4 \text{km} \) of the 30 \text{Hz} Broadband traveltime sensitivity kernels for the Marmousi model. (a) Crosscorrelation. (b) Unwrapped phase (instantaneous traveltime).

2.10 Traveltime sensitivity kernels using the unwrapped phase for a source-receiver offset \( h = 8.2 \text{km} \). (a) 30 \text{Hz} single frequency kernel (b) 30 \text{Hz} broadband kernel.

2.11 Velocity model slice at \( y = 0.5 \text{km} \). The anomaly is placed at a distance \( x_a = 0.5 \text{km} \) from the source. The source is shown with the blue star. The red lines show receivers’ location. I place receivers on one horizontal and two vertical lines.

2.12 Banana-doughnut traveltime sensitivity kernels slices overlaid to a small anomaly for different receiver locations. The anomaly diameter is \( d_a = \lambda = 0.2 \text{km} \), where \( \lambda \) is the wavelength corresponding to the peak frequency \( f_0 = 10 \text{Hz} \).
2.13 Measured traveltimes along the horizontal array of receivers. Black line represent traveltime difference computed using an Eikonal solver, blue is the cross-correlation traveltime and red is the traveltime measured using the unwrapped phase. Delay measured using the Eikonal solver is constant after the anomaly, but is over-estimated. For both cross-correlation and instantaneous traveltime, we observe the same decreasing tendency. Crosscorrelation traveltimes converge to zero for receivers far from the anomaly. However, the instantaneous traveltime is still detecting the effect of the anomaly. Notice that the inflection point near 1.8 km for the cross-correlation curve is a sampling artifact.

2.14 Recorded pressure seismograms along the horizontal array of receivers for different locations: (a) $x = 0.6 \text{ km}$ (b) $x = 1.0 \text{ km}$ (c) $x = 1.4 \text{ km}$ (d) $x = 1.8 \text{ km}$. The blue and red curves correspond to the homogeneous and the case with positive anomaly respectively. I show under each figure the measured traveltime shift using cross-correlation $\Delta t_{Cross}$, and using the instantaneous traveltime $\Delta t_{Inst}$.

2.15 Measured traveltimes along the vertical array of receivers placed at $x = 1.0 \text{ km}$. Black line represent traveltime computed using an Eikonal solver, blue is the cross-correlation traveltime and red is the traveltime measured using the unwrapped phase. Crosscorrelation traveltimes minimum is along the source-receiver axis. For the unwrapped phase, it predicts traveltimes with maximum values along $z = 0.5 \text{ km}$, the source-receiver axis.

2.16 Measured traveltimes along the vertical array of receivers placed at $x = 1.8 \text{ km}$. Black line represent traveltime computed using an Eikonal solver, blue is the cross-correlation traveltime and red is the traveltime measured using the unwrapped phase. Crosscorrelation traveltimes minimum is along the source-receiver direction and is almost zero. For the unwrapped phase, it predicts traveltimes with maximum values along $z = 0.5 \text{ km}$, the source-receiver axis.

3.1 Schematic representation of the experimental geometry. The traveltime kernels are studied for two configurations: the transmission and the reflection geometry. The transmission and kinematic reflection kernels are given by the first Fresnel zone around the transmission ray-path and the reflection ray-path respectively.
3.2 5 Hz bandlimited traveltime sensitivity kernels for the first parameterization: NMO velocity, \(\eta\) and \(\delta\) parameters. (a-c) NMO velocity kernels, (d-f) \(\eta\) parameter kernels (g-i) \(\delta\) parameter kernels. The effect of source-receiver orientation is shown by changing the source receiver angle. .......................... 82

3.3 5 Hz bandlimited sensitivity kernels for the second parameterization \((v_h, \eta, \delta)\). (a-c) show the \(\eta\) parameter kernels and (d-f) are the \(\delta\) parameter kernels. Since \(v_h\) kernels look similar to \(v_n\) kernels in the previous parameterization we omitted their plots. .......................... 83

3.4 5 Hz bandlimited sensitivity kernels for the third parameterization \((v_h, \eta, \epsilon)\). (a-c) show the \(\eta\) parameter kernels and (d-f) are the \(\epsilon\) parameter kernels. .......................... 85

3.5 5 Hz bandlimited diving wave sensitivity kernels, (a) NMO velocity kernel, (b) \(\eta\) parameter kernel (c) \(\delta\) parameter kernel. .......................... 86

3.6 A smoothed version of the 3-D SEG/EAGE salt body model. The three slices include a depth slice (top) a y-slice (bottom left) and an x-slice (bottom right) corresponding to \(z = 2000\) m, \(y = 6760\) m, and \(x = 6760\) m, respectively, as the white lines indicate. .......................... 87

3.7 5 Hz peak frequency sensitivity kernels for the SEG/EAGE salt model for the case of a diving wave. The source is located at \((x_s, y_s, z_s) = (1000, 6760, 0)\) m, the receiver is placed at \((x_r, y_r, z_r) = (12600, 6760, 0)\) m, corresponding to perturbations to (a) the NMO velocity, (b) the \(\eta\) parameter and (c) the \(\delta\) parameter. .......................... 88

3.8 5 Hz peak frequency sensitivity kernels for the SEG/EAGE salt model for the case of a reflected wave (with fixed reflection point). The source is located at \((x_s, y_s, z_s) = (1000, 6760, 0)\) m, the receiver is placed at \((x_r, y_r, z_r) = (12600, 6760, 0)\) m. The image point is located at \((x, y, z) = (6800, 6760, 4000)\) m, corresponding to perturbations of (a) NMO velocity, (b) \(\eta\) parameter and (c) \(\delta\) parameter. .......................... 89

4.1 Frequency domain modeled wavefield for an elliptic anisotropic medium, with \(v_h = 1.8\) km/s, \(\epsilon = 0.25\) and \(f = 5\) Hz: (a) the real part and (b) the imaginary part. (c) and (d) show the inverse Fourier transformed frequency domain wavefield to time domain at \(t = 0.2\) s and \(t = 0.5\) s, respectively. The wavefront is compared to the Eikonal traveltime solution given by the red curve. .......................... 100
4.2 Frequency domain modeled wavefield for a VTI anisotropic medium, with $v_h = 1.8$ km/s, $\epsilon = 0.25$, $\eta = 0.1$ and $f = 5$ Hz: (a) the real part and (b) the imaginary part. (c) and (d) show the inverse Fourier transformed frequency domain wavefield to time domain at $t = 0.2$ s and $t = 0.5$ s, respectively. The wavefront is compared to the Eikonal traveltime solution given by the red curve. The wavefront and the first arrival traveltime are matching each other. We can also notice the shear wave artifacts, caused by the pseudo-acoustic equations. Aliasing artifacts are also present as a result of the large frequency sampling.

4.3 The VTI Marmousi II model: (a) the horizontal velocity $v_h$, (b) $\epsilon$ parameter, (c) $\eta$ parameter.

4.4 Frequency domain modeling for the VTI Marmousi II model with $f = 5$ Hz: (a) the real part and (b) the imaginary part. (c) and (d) show the inverse Fourier transformed frequency domain wavefield to time domain at $t = 0.8$ s and $t = 1.2$ s, respectively. The wavefront is compared to the Eikonal traveltime solution given by the red curve. Here, there is no shear waves artifacts as the source is placed in the water (isotropic layer).

4.5 Single frequency Born sensitivity kernels for a VTI with increasing velocity with depth model: (a) $v_h$, (b) $\epsilon$ and (c) $\eta$.

4.6 The VTI initial models: (a) the horizontal velocity $v_h$, (b) $\epsilon$ parameter, (c) $\eta$ parameter.

4.7 FWI inverted models: (a) the horizontal velocity $v_h$, (b) $\epsilon$ parameter and (c) $\eta$ parameter. All three parameters are simultaneously inverted.

4.8 Vertical profiles for the simultaneous inversion at $x = 8$ km. (a) $v_h$, (b) $\epsilon$ parameter and (c) $\eta$ parameter.

4.9 Vertical profiles for the simultaneous inversion at $x = 12$ km. (a) $v_h$, (b) $\epsilon$ parameter and (c) $\eta$ parameter.

4.10 Convergence history for the simultaneous inversion. All frequencies are used as it is the measured misfit for time domain data.

4.11 FWI inverted models: (a) the horizontal velocity $v_h$, (b) $\epsilon$ parameter and (c) $\eta$ parameter. The three parameters are hierarchically inverted: for frequency band 1, only $v_h$ is inverted, then for frequency band 2, $v_h$ and $\epsilon$ are inverted. Finally all three parameters are inverted for frequency bands 3 and 4.
4.12 Vertical profiles for the hierarchical inversion at \( x = 8 \) km. (a) \( v_h \), (b) \( \epsilon \) parameter and (c) \( \eta \) parameter. 117

4.13 Vertical profiles for the hierarchical inversion at \( x = 12 \) km. (a) \( v_h \), (b) \( \epsilon \) parameter and (c) \( \eta \) parameter. 118

4.14 FWI inverted models: (a) the horizontal velocity \( v_h \), (b) \( \epsilon \) parameter. \( \eta \) parameter is fixed as the initial \( \eta_0 \). The two parameters are inverted simultaneously. 119

4.15 Vertical profiles for the inversion with \( \eta \) fixed as the initial \( \eta_0 \) at \( x = 8 \) km. (a) \( v_h \) and (b) \( \epsilon \) parameter. 120

4.16 Vertical profiles for the inversion with \( \eta \) fixed as the initial \( \eta_0 \) at \( x = 12 \) km. (a) \( v_h \) and (b) \( \epsilon \) parameter. 121

4.17 Realistic VTI initial models, corresponding to \( \delta \) parameter equals zero: (a) the horizontal velocity \( v_h \), (b) \( \epsilon \) parameter, (c) \( \eta \) parameter. 122

4.18 FWI inverted models for inversion with realistic initial models: (a) the horizontal velocity \( v_h \), (b) \( \epsilon \) parameter and (c) \( \eta \) parameter. 123

4.19 Vertical profiles for the inversion with realistic initial models at \( x = 8 \) km. (a) \( v_h \), (b) \( \epsilon \) parameter and (c) \( \eta \) parameter. 124

4.20 Vertical profiles for the inversion with realistic initial models at \( x = 12 \) km. (a) \( v_h \), (b) \( \epsilon \) parameter and (c) \( \eta \) parameter. 125

4.21 FWI inverted models for inversion without inverse crime: (a) the horizontal velocity \( v_h \), (b) \( \epsilon \) parameter and (c) \( \eta \) parameter. 126

4.22 Vertical profiles for the inversion without inverse crime at \( x = 8 \) km. (a) \( v_h \), (b) \( \epsilon \) parameter and (c) \( \eta \) parameter. 127

4.23 Vertical profiles for the inversion without inverse crime at \( x = 12 \) km. (a) \( v_h \), (b) \( \epsilon \) parameter and (c) \( \eta \) parameter. 128

5.1 Modeling computational cost for a 2-D problem. (a) frequency domain modeling with repeated LU decomposition, (b) frequency domain modeling with the same LU used, (c) time domain modeling (d) Hybrid frequency domain-ray tracing method. 137

5.2 Modeling computational cost for a 3-D problem. (a) frequency domain modeling, (b) time domain modeling (c) Hybrid frequency domain-ray tracing method. 138
5.3 Anomaly inversion using Born sensitivity kernels and truncated Gauss-Newton method. (a) is the exact model, (b) the inverted velocity using Helmholtz based scattering integral (SI) approach, (c) the inverted velocity using the hybrid implementation.

5.4 Vertical profiles for the anomaly inversion using Born sensitivity kernels and truncated Gauss-Newton method at $x = 2500$ m. (a) Helmholtz based SI and (b) Hybrid approach. The red line indicates the exact model and the blue line represents the inverted model.

5.5 The Marmousi model (a), and the initial model (b).

5.6 Inversion results for the Marmousi model for frequencies between 1 Hz and 3.5 Hz. (a) Helmholtz equation based inversion and (b) Hybrid inversion.

5.7 FWI inversion results for the Marmousi model using the models inverted for 1-3.5 Hz, shown Figure 5.6, as initial models. (a) Helmholtz equation based inversion and (b) Hybrid inversion.

5.8 Vertical profiles for the Marmousi model inversion. (a,b) Helmholtz equation based inversion vertical profiles at $x = 3000$ m and $x = 7500$ m, respectively. (c,d) Hybrid inversion vertical profiles at $x = 3000$ m and $x = 7500$ m, respectively. The red line stands for the exact model, the black is the initial model, the purple is the inverted model for 1-3.5 Hz, and the blue line is the final inverted models using FWI.

6.1 Acquisition location (Courtesy of CGG).

6.2 Shot gather example: (a) Shot gather for a source located at $(x, z) = (3.75, 0.005)$ km, (b) data frequency spectrum.

6.3 Bandpass filtered shot gather between 3 and 12 Hz: (a) Shot gather for a source located at $(x, z) = (3.75, 0.005)$ km, (b) data frequency spectrum.

6.4 Initial velocity model.

6.5 The source wavelet.

6.6 Inverted models: (a) $v_h$ model (b) $\epsilon$ parameter model and (c) $\eta$ parameter model.

6.7 Vertical profiles at $x = 10.5$ km compared with well-log.

6.8 Convergence history. The misfit is given by the phase difference.

6.9 Data comparison for shot gather 50: (a) initial model data (b) final model data.
6.10 Data comparison for shot gather 100: (a) initial model data (b) final model data.
........................................................................................................ 154

6.11 Image using reverse time migration (RTM) for (a) the initial model (b) the final model.
........................................................................................................ 155

6.12 Angle gathers for (a) the initial model (b) the final model. The angle gathers are shown for multiple lateral positions. The angles is between 0 and 45°. The green arrows show locations where there is improvement. The yellow arrows show some locations where the angles gathers become less flat.
........................................................................................................ 156
LIST OF TABLES

4.1 VTI Marmousi II inversion parameters. 4 bands of multiples frequencies are used. ........................................ 110

6.1 CGG real data inversion parameters. 2 bands of multiples frequencies are used. ........................................ 150
Chapter 1

Introduction

In seismic exploration, the main goal is to provide an Earth model that properly represents the true subsurface geological structure. This model is used to identify locations of hydrocarbon reservoirs and gives insights on where to drill new wells. The surface recorded seismic data, after being processed, is utilized to map the reflections to their exact positions in the Earth’s subsurface. This operation is called seismic migration or imaging. To accurately migrate seismic events, an accurate velocity, or Earth model is required. Therefore, the estimation of model parameters is an important and essential task in hydrocarbon exploration.

Recently, seismic exploration is deployed for deep complex structures where Earth’s model can be complicated, for instance, encounter large viscoelastic and anisotropic behavior. In these complex situations, multiple parameters need to be extracted from the recorded data on the surface. Understanding the sensitivity of seismic data to these parameters and developing new techniques handling the complexity of multi-parameter estimation is necessary.

Motivated by finding an accurate subsurface Earth model, this dissertation studies the prospect of applying a scattering integral approach for multi-parameter inversion for a transversely isotropic model with a vertical axis of symmetry (VTI). The scattering integral approach is based on the explicit computation of the sensitivity kernels. In the first two chapters of this dissertation, I describe the sensitivity kernels for wave equation tomography (WET), and show their usefulness in the sensitivity analysis of various VTI medium parameterizations. In the following chapters, I cover the in-
version implementation and the application to synthetic and real data sets in a full waveform inversion (FWI) framework.

1.1 Model parameter estimation

The advances in seismic data acquisition allowed recording seismic data with potentially more embedded information. The newly acquired data have wider frequency content (broadband seismic data) and a wide-azimuth 3-D coverage. These advances, accompanied with the improvement in computational capabilities motivated the development of new Earth parameters estimation methods. The goal is to find an accurate subsurface model that produces, with the proper physics simulation, data similar to the observed ones. This task can be performed in both image or data domains.

In the image domain, migrated images are used to evaluate the accuracy of the velocity model. This technique is called migration velocity analysis (MVA) (Al-Yahya, 1989; Symes and Carazzone, 1991). The coherence of the common image gathers (CIGs) is usually exploited to update the migration velocity, iteratively. A good inverted velocity model should flatten the CIGs. To account for the wave propagation complexity associated with complex velocity models, MVA can be implemented using the wave-equation, it is denoted as wave-equation migration velocity analysis (WEMVA) (Symes and Kern, 1994; Chavent and Jacewitz, 1995; Biondi and Sava, 1999). Performing velocity analysis in the image domain is less sensitive to cycle-skipping problems encountered in full waveform inversion (FWI); however, the resolution of the final model is limited compared to FWI.

In the data domain, the model parameters are estimated using a data fitting procedure. An initial velocity model is used to simulate seismic data. Then, the data residuals between the observed data and the modeled ones is computed at the receivers locations. Finally, these residuals are back-propagated, or smeared, along the
sensitivity kernels to update the model. The sensitivity kernels, or Fréchet derivatives (Woodward, 1992; Marquering et al., 1998; Dahlen et al., 2000; Hung et al., 2000; Tromp et al., 2005), are represented by the region in the model to which the recorded data might be sensitive. Depending on the wavefield modeling method, the inversion scheme can vary from traveltime tomography to full waveform inversion.

1.1.1 Traveltime tomography

For traveltime tomography, the high-frequency assumption is used to model the data. Ray theory is used to represent the data by only the kinematic part (the traveltimes). Thus, the misfit function is the difference between the arrival traveltimes (Dziewonski, 1984; Van Der Hilst et al., 1997; Nolet, 2012). The model is, then, updated by the back-projection of the traveltime residuals along the ray-path or using the adjoint-state. Figure 1.1 shows the ray-path which forms the sensitivity kernel for traveltime tomography. The advantage of traveltime tomography is the quasi-convex objective function resulting in the fast convergence to a global minimum. Nevertheless, the inverted models suffer from limited resolution caused by the smoothness constraints needed for ray theory.

1.1.2 Wave equation tomography

The wave equation tomography, or the finite frequency traveltime tomography, (Deveney, 1984; Pratt and Goulty, 1991; Luo and Schuster, 1991; Woodward, 1992; Van Leeuwen and Mulder, 2010) had been proposed to tackle the FWI nonlinearity problem. The misfit function for WET is based on the finite frequency traveltime, which results in a quasi-convex misfit function. The traveltime misfit is measured by identifying the maximum of the cross-correlation between two waveforms (observed and modeled) (Marquering et al., 1998; Dahlen et al., 2000; Hung et al., 2000; Tromp et al., 2005). Rytov approximation can also be used instead of cross-correlation to
compute the finite frequency traveltimes. It gives the phase perturbation as a function of Born wavefield perturbation (Snieder and Lomax 1996; Jocker et al. 2006; Liu et al. 2009).

Since the traveltime delays are extracted from seismic data, WET handles the finite frequency features for wave propagation and it gives better resolution than traveltime tomography. The finite frequency traveltimes are back-projected along the finite frequency traveltime sensitivity kernels to update the model. Figure 1.2 shows an example of the sensitivity kernels for a linearly increasing velocity model. A non-intuitive property of the finite frequency traveltime can be observed in this figure. Contrary to the ray based kernel, the finite frequency sensitivity to slowness perturbation is zero along the ray-path. This feature has been the subject of multiple research papers (Woodward 1992; Marquering et al. 1998, 1999; Dahlen et al. 2000; Hung et al. 2000; Tromp et al. 2005). In fact, computing the maximum of the cross-correlation is only valid under certain circumstances (Hörmann et al. 2002; De Hoop and Van Der Hilst 2005; Van Leeuwen and Mulder 2010). Cross-correlation computes the phase shift correctly only when the source spectrum of the modeled data and the
real data are identical. Also, the Born approximation yields poor approximation for transmitted waves as shown by Bleistein et al. (2000).

For the Rytov approximation, a linear relation between phase and traveltime is used to derive the sensitivity kernels. In fact, for real inversion problems, the Rytov approximation for phase perturbation should be unwrapped to compute the traveltime misfit accurately (Gélis et al., 2007). Thus, limitations in the cross-correlation, as well as, Born and Rytov approximations raise serious questions to the accuracy and applicability of the Born/Rytov kernels, also known as banana-doughnut kernels (Marquering et al., 1999). Motivated by these limitations, Chapter 2 of this dissertation will be dedicated to the analysis and derivation of new sensitivity kernels using the instantaneous traveltime.

Figure 1.2: Finite frequency traveltime sensitivity kernel (Figure from Rickett, 2000).

1.1.3 Full waveform inversion

Contrary to tomographic methods, full waveform inversion (Lailly, 1983; Tarantola, 1984; Mora, 1988; Tarantola, 2005; Brenders and Pratt, 2007; Virieux and Operto, 2009) uses the full data content, including diving waves, reflections and multiples,
in order to invert for high accuracy models. A local optimization technique is commonly used to update the model as the model size is too large for global optimization techniques. At each iteration, the misfit between observed and modeled data (using an approximate model), is used to compute the model update, composed of the gradient and the Hessian matrix or an approximate Hessian. FWI can be used to estimate various model parameters: for example P-wave velocity, S-wave velocity, density, anisotropy, and attenuation. The final inverted models admit high accuracy and resolution compared to tomographic models. Figure 1.3 shows that FWI inverted model contains details not available in the tomography model [Kapoor et al. 2012].

A considerable limitation of full waveform inversion is the high nonlinearity of the misfit function [Bunks et al. 1995; Virieux and Operto 2009]. Due to the oscillatory nature of the wavefield, the modeled data should lie within a half-cycle of the period of the observed data to guarantee the convergence to the global minimum. Figure 1.4 illustrates the cycle skipping problem [Virieux and Operto 2009]. A time difference larger than the T/2, where T is the signal period, results in convergence with one cycle mismatch. However, when the difference is smaller than T/2, the model will be accurately updated. As a result, an accurate initial model, low frequencies, and large
data aperture are essential for FWI.

Figure 1.4: Illustration of the cycle skipping problem (Figure from Virieux and Operto (2009)).

Several approaches are proposed to deal with the cycle skipping problem. For example, in the multiscale approach, subsets of the data are inverted sequentially from low to high frequencies (Bunks et al. 1995). As a result, the long wavelengths of the model are constructed first, then the shorter wavelengths are added throughout the inversion. Layer stripping techniques are also used as regularization methods to update the model from the shallow to deeper parts (Shipp and Singh 2002). Shin and Min (2006) proposed the logarithmic approach to separate the amplitude and the phase. They showed that phase inversion is more convenient than inverting the whole wavefield. Nevertheless, the cycle skipping problem is not resolved, and phase unwrapping is necessary (when the data is cycle-skipped, the phase wraps around $-\pi$ and $\pi$. Correcting for this problem is denoted as phase unwrapping). Choi and Alkhalifah (2011) proposed frequency domain inversion with phase unwrapping. The phase is unwrapped using a differential operator coupled with a damping factor applied to the wavefield (Choi and Alkhalifah 2011; Alkhalifah and Choi 2012; Choi
and Alkhalifah 2013, 2015). This approach is similar to the multiscale one since the damping factor is relaxed during the inversion to include more data and resolve more features of the model. Recently, reflection waveform inversion (RWI) is introduced to deal with the cycle skipping problem (Xu et al., 2012; Wu and Alkhalifah 2015b; Alkhalifah and Wu 2016). It inverts for both a background model and a model perturbation that represents the image. Alkhalifah and Wu (2016) described how an optimized version of RWI can represent a natural combination of FWI and MVA.

The complexity of FWI is magnified if we invert for more than one parameter. In fact, the problem becomes ill-posed, as we try to invert for multiple parameters at different scales. Understanding the behavior of each parameter and its sensitivity to the data is a fundamental task (Plessix and Cao, 2011; Gholami et al., 2013b; Alkhalifah and Plessix, 2014; Djebbi et al., 2017). In the third chapter of this dissertation, I study the seismic data sensitivity to the VTI parameters using the sensitivity kernels.

1.2 The inversion approach

As mentioned in section 1.1, the data domain inversion procedure is based on the iterative minimization of the misfit functional between the observed and modeled data. First, the wavefield is forward propagated using an initial model, and the residuals are computed at the receivers locations. Then, we estimate the perturbation in the model (the gradient) to update the model. The procedure is repeated iteratively until fitting is achieved. This update direction can be calculated using two different approaches: the adjoint-state method or the scattering integral method.

1.2.1 Adjoint-state method

The adjoint-state method is proposed to compute the gradient of the misfit functional without computing explicitly the Fréchet derivatives (Chavent, 1974). The idea is to formulate an adjoint-state problem where we solve for adjoint-state variables that
measure the perturbation with the forward problem state variables (Tromp et al., 2005; Plessix, 2006). The gradient of the misfit function requires only one forward modeling and another adjoint modeling per shot. Thus, the adjoint-state method is reasonably efficient to compute the first-order update direction (the gradient). On the other hand, the adjoint-state method lacks the flexibility of the scattering integral approach when computing the second-order update direction which requires the inverse of the Hessian/approximate Hessian matrix. Further modeling steps are needed for the adjoint-state method to solve for the Hessian matrix (Métiévier et al., 2013).

1.2.2 Scattering integral method

The gradient of the objective function can be directly computed using the scattering integral (SI) method. The sensitivity kernels represent the derivatives of the wavefield with respect to the model parameters. They show which region of the model the data is sensitive to. As a result, the data residuals multiplied with the sensitivity kernels represent the gradient of misfit function (Chen et al., 2007; Tao and Sen, 2013). The method is widely used in global seismology, where the number of recording stations is small compared to seismic exploration applications. Chen et al. (2007) presented a comparative study between the adjoint-state method and the SI method. Liu et al. (2015) introduced a method to avoid storing the Fréchet derivatives. The Green’s functions are stored in memory, and a vector matrix product is used to compute on the fly the gradient. The main advantage of the SI approach is the availability of the Fréchet derivatives (the sensitivity kernels). Therefore, additional quantities can be directly estimated, among them, the diagonal part of the Hessian, the Gauss Newton Hessian, as well as the gradient step length. In Chapter 4, I study in more details the SI method and present a multi-parameter anisotropic FWI implementation in the frequency domain.
1.3 Seismic Anisotropy

*Seismic anisotropy* is defined as the variation of the seismic wave velocity with respect to the direction of propagation. Anisotropic wave propagation arises in multiple situations: thin layering compared to the dominant seismic wavelength, grain preferred orientation (for shales), aligned fractures in fractured media and non-hydrostatic stresses causing directional micro-cracking (Levin 1979; Thomsen 1986).

Seismic anisotropy has been recognized many decades ago (McCollum and Snell 1932; Postma 1955). However, it was not incorporated in imaging and inversion workflows until recently. In fact, the recent attention given to anisotropy is mainly driven by the improvements in seismic data acquisition. The improved data quality and coverage (multi-component data with large aperture and wide/full azimuth) make it more sensitive to anisotropy. Also, the evolution of the computational capabilities has remarkably reduced the cost of handling complex anisotropic models. For seismic migration, incorporating anisotropy has become the industry standard, especially for complex deep structures. However, inverting for anisotropy parameters using surface recorded seismic data is still a challenge. In this dissertation, I tackle problems related to multi-parameter anisotropic inversion.

The simplest model that can physically represent the Earth’s layers is the transversely isotropic (TI) model (Tsvankin 2012). It has a single axis of rotational symmetry. It describes appropriately the intrinsic anisotropy of shales where plate-shaped clay particles have preferential alignment (Sayers 1994). Moreover, TI anisotropy describes very well the thin layering behavior which is caused by gravity, as well as the aligned fracturing resulting from regional stresses. So, transverse anisotropy is the most common anisotropic model in exploration seismology (Tsvankin 2012). When TI anisotropy is driven by gravity, the axis of symmetry will be vertical, and the resulting medium is transversely isotropic with a vertical axis of symmetry (VTI).
However, when regional stresses are the governing physical phenomena, vertical fractures appear in the medium, and the axis of symmetry will be horizontal. This is referred to as transversely isotropic model with a horizontal axis of symmetry (HTI). Another case of TI anisotropy is when the layers are titled because of internal tectonic forces (near salt flanks). The resulting medium is tilted transversely isotropic (TTI). Figure 1.5 illustrates the three TI symmetries.

Figure 1.5: Illustration of different symmetries for anisotropic models: (a) VTI symmetry, (b) HTI symmetry, (c) TTI symmetry.

In some situations, the subsurface layers contain both horizontal/titled layers and aligned fractures. This behavior is common in fractured reservoirs. In this situation, the TI symmetry is not sufficient to describe the anisotropic behavior of the Earth. In fact, these layers are described better with an orthorhombic model (Schoenberg and
Figure 1.6 shows an orthorhombic model illustration. When the layers are tilted, it becomes a tilted orthorhombic (TOR). The representation of an orthorhombic model requires 9 model parameters (for the elastic case) not counting density. Inverting for anisotropy is an open topic for research, where we aim to invert the maximum number of parameters. Thus, the first step towards inverting for complex anisotropic models is to study the VTI model. I choose to focus on VTI models in this dissertation. I will show in chapter 3 a study of the sensitivity of the seismic data to the VTI parameters. In chapters 4 and 6, I will present inversion applications to synthetic and real data sets.

Historically, anisotropic media are described using the stiffness coefficients $c_{ij}$. However, these parameters are not well suited for seismic data processing, imaging and inversion. The Earth model can be parameterized using Thomsen parameters (Thomsen, 1986). The motivation for this VTI parameterization is the separation of anisotropy into different parameters, which simplifies the quantification of the
anisotropy effects. Thomsen parameters are given by:

\[
\begin{align*}
    v_{P0} &= \sqrt{\frac{c_{33}}{\rho}}, \\
v_{S0} &= \sqrt{\frac{c_{55}}{\rho}}, \\
    \epsilon &= \frac{c_{11} - c_{33}}{2c_{33}}, \\
    \delta &= \frac{(c_{15} + c_{52})^2 - (c_{33} - c_{55})^2}{2c_{33}(c_{33} - c_{55})}, \\
    \gamma &= \frac{c_{66} - c_{55}}{2c_{55}}.
\end{align*}
\]  \tag{1.1}

\(v_{P0}\) and \(v_{S0}\) are the P- and S-waves velocities in the vertical direction. \(\epsilon, \delta\) and \(\gamma\) are three dimensionless parameters describing the VTI anisotropy.

Alkhalifah (1998) showed that for processing and imaging purposes, the acoustic approximation can be used to reduce the number of parameters to 3 parameters \(v_{P0}, \epsilon\) and \(\delta\) parameters. For the sake of notation simplicity, I denote the vertical P-wave velocity by \(v_v\) instead of \(v_{P0}\) in the remainder of this dissertation.

Alkhalifah and Tsvankin (1995) introduced a new parameter, the anellipticity parameter \(\eta\) that relates the normal move-out velocity \(v_n\) to the horizontal velocity \(v_h\). \(\eta\) parameter is given by:

\[
\eta = \frac{1}{2} \left( \frac{v_h^2}{v_n^2} - 1 \right) = \frac{\epsilon - \delta}{1 + 2\delta}.
\]  \tag{1.2}

So, I presented here three velocities, the vertical velocity \(v_v\), the horizontal velocity \(v_h\) and the normal move-out (NMO) velocity \(v_n\), and three anisotropy parameter \(\epsilon, \delta\) and \(\eta\). Only three of these six parameters are needed to characterize the acoustic VTI medium. These model parameters are related through equations \(1.3\):

\[
\begin{align*}
v_h^2 &= v_v^2(1 + 2\epsilon), \\
v_n^2 &= v_v^2(1 + 2\delta), \\
v_h^2 &= v_n^2(1 + 2\eta), \\
\eta &= \frac{\epsilon - \delta}{1 + 2\delta}.
\end{align*}
\]  \tag{1.3}
I show in Figure 1.7 the effect of anisotropy on wave propagation. Figure 1.7(a) shows the isotropic traveltime for a constant velocity equals to 1.8 km/s. An example of wavefront (traveltime contour) is shown with the black line. For an isotropic model, this wavefront is circular. Figure 1.7(a) shows the VTI case for a model parameterized using $v_h = 1.8$ km/s, $\eta = 0.2$ and $\epsilon = 0.45$. The contour is no longer circular. It has an ellipsoid shape but with compressions in the diagonal directions. This is caused by the anellipticity $\eta$. In Figure 1.7(c) I show the elliptic case where $\eta = 0$. The wavefront is a perfect ellipse. In the last example, Figure 1.7(d), I introduce a tilt angle of $\theta = 30^\circ$ to the VTI model, to visualize the behavior of seismic waves in TTI models.

1.4 Multi-parameter anisotropic inversion

The nonlinearity problem associated with FWI is magnified when we start inverting for multi-parameters. The inversion problem is an ill-posed problem when we try to resolve more than one parameter with different scattering properties. Seismic data can be sensitive, in the same way, to multiple parameters. Then, leakage between parameters can occur, and the inverted models will contain unwanted features.

The idea of multi-parameter inversion is to study the sensitivity of seismic data to different parameters. Then, depending on the properties of the data, as well as the experimental set-up, we can choose the best parameterization that minimizes the null space of the inversion. For VTI anisotropic media, Plessix and Cao (2011) showed, using an eigenvalue/eigenvector decomposition, which anisotropy parameters could be retrieved using FWI. Their focus was on diving waves, and in that case, a combination of NMO and horizontal velocities provided the most sensitivity. The radiation patterns were also proposed as an analysis tool to study the sensitivity (Gholami et al., 2013b; Operto et al., 2013; Alkhalifah and Plessix, 2014). The trade-off between model parameters exists when the radiation patterns overlap for certain scattering
Figure 1.7: Illustration of different symmetries for anisotropic models: (a) VTI symmetry, (b) HTI symmetry, (c) Elliptic symmetry, (d) TTI with tilt angle $\theta = 30^\circ$.

angles range. Figure 1.8 shows an example of radiation patterns for two different VTI parameterizations. The first parameterization, given by $(v_n, \eta, \delta)$, is adequate for simultaneously inverting diving waves and reflections [Plessix and Cao, 2011]. However, the second parameterization, $(v_h, \epsilon, \eta)$, is more suited to hierarchical inversion of data having both diving waves and reflections [Alkhalifah and Plessix, 2014].

Alkhalifah (2016) studied the same problem by comparing the anisotropic parameters influence on both the reflections and the transmissions (diving waves). In this dissertation, one objective is to analyze the sensitivity kernels for different VTI pa-
Figure 1.8: Radiation patterns for two VTI model parameterizations: (a) \((v_n, \eta, \delta)\) parameterization, \(v_n\) in red, \(\eta\) and \(\delta\) in orange, (b) \((v_h, \epsilon, \eta)\) parameterization, \(v_h\) in red, \(\eta\) and \(\epsilon\) in orange.

rameterizations. These sensitivity kernels give a more complete picture compared to radiations patterns. The radiation patterns show the scattering/reflection behavior at the scattering point level. However, the sensitivity kernels show the sensitivity along the wave-path. I present in Chapter 3 this analysis for various VTI model
parameterizations.

1.5 Thesis objectives and outline

The behavior of waves propagating in the subsurface is anisotropic, and the isotropic approximation can not handle such complexity. Anisotropy has been extensively studied for imaging purposes, however, estimating the anisotropy parameters from surface recorded data is still a research topic. Therefore, my primary objective is to understand wave phenomena for the band of frequencies we use in seismic exploration. I aim to improve the clarity of pictures of the Earth under more complicated physics where anisotropy is considered for inversion.

The limited progress in inverting multiple anisotropic parameters is caused by the ill-posedness of the inversion problem. In this dissertation, the first objective is to analyze the sensitivity of seismic data to the anisotropic parameters for the VTI symmetry. I start with a description of the sensitivity kernels using the unwrapped phase for wave equation tomography (WET). Then, I conduct a sensitivity analysis using these kernels for VTI media. Based on the analysis results, I choose an optimal parameterization to perform the inversion using a scattering integral approach in the frequency domain. The advantage of the scattering integral method is the availability of the sensitivity information which simplifies computing additional attributes necessary for the inversion. I apply the proposed method to both synthetic and real data sets, where I show accurate inversion results. Finally, as the major limitation associated with SI method is the considerable computational cost for 3D problems, I propose a hybrid inversion implementation. I combine ray theory with wavefield modeling to achieve better performance with acceptable accuracy.

I intend to tackle these issues in the following chapters:

- **Chapter 2**: In this chapter, I explain the physical meaning of the sensitivity kernels and their spatial distribution. Both the analysis of the anisotropy
behavior as well as the implementation of the scattering integral method are based on those kernels. Investigating the spatial distribution of the kernels was motivated by a counterintuitive observation, observed for the wave equation tomography: The finite frequency kernels suffer from a zero sensitivity along the ray-path. This is the result of using either cross-correlation/Born or Rytov approximation in calculating the kernels. Thus, I investigate the prospect of using the unwrapped phase as an alternative to Born and Rytov approximations to derive the sensitivity kernels. The analysis of the sensitivity kernels shows that the unwrapped phase is a generalized form of Rytov approximation. I also show, through numerical examples, that the new kernels are more sensitive to scatterers compared to conventional ones which makes them a good alternative for inversion. This work is the subject of a journal publication in Geophysical Journal International (Djebbi and Alkhalifah, 2014).

- **Chapter 3**: I derive the anisotropic sensitivity kernels for the VTI acoustic wave equation based on the unwrapped phase proposed in Chapter 2. Then, I analyze the behavior of the kernels for various situations depending on the direction of propagation (vertical, horizontal, 45° angle with respect to the horizontal direction). An overlap of the sensitivity between two different parameters shows a possible trade-off. The analysis is carried out for wave equation tomography kernels. However, the results of the analysis are also valid for any finite frequency inversion scheme, specifically for FWI.

The analysis is conducted for multiple parameterizations to show which parameterization is optimal. \((v_n, \eta, \delta)\) is shown to be more suited for data dominated by diving waves. This parameterization is appropriate for wave equation tomography type of applications. \((v_h, \epsilon, \eta)\) is the optimal parameterization for FWI applications where the data contain both reflections and diving waves. Diving waves can be used at an early stage to fit \(v_h\) as it is the primary parameter
sensitive to large scattering angles. $\epsilon$ can play the role of density and collect the amplitude inaccuracies during the inversion. As a result, the inverted $v_h$ will be more accurate than $v_n$ in the first parameterization. This work is the subject of a journal publication in Geophysical Prospecting (Djebbi et al., 2017).

- **Chapter 4**: I present in this chapter the scattering integral method for FWI. First, I start with a description of the forward modeling for VTI media. The modeling code is developed in the frequency domain. It is a direct solver based on lower-upper (LU) decomposition for sparse systems. Then, I describe the scattering integral method. The availability of the sensitivity information makes this approach attractive. In fact, we can estimate approximations of the inverse Hessian directly by matrix-vector multiplications. The diagonal of the approximate Hessian, as well as the gradient step length, can be computed on the fly. The truncated Gauss Newton direction can also be estimated using an inner conjugate gradient solution of a linear system of equations. The $(v_h, \epsilon, \eta)$ parameterization is utilized for the inversion. I apply the inversion for various numerical tests to explain, based on the sensitivity, the resolution of the recovered models. This work was presented during the EAGE conference & exhibition 2017 and is the subject of a journal paper that is submitted to Geophysics journal.

- **Chapter 5**: In this chapter, I describe the hybrid method for inversion. The scattering integral method’s cost is prohibitive for 3-D applications. It requires solving the wave equation from all the receivers locations. To deal with this issue, I propose a ray+Born method where I use ray theory to trace waves from the receivers locations. The method makes use of modeling from the source locations and ray tracing from the receivers locations. The residuals are first computed using the modeled wavefield, then, these residuals are back-
propagated using ray based sensitivity kernels. As a result, the computational cost is highly reduced. The hybrid method is tested for the 2-D case for both transmission (cross-well), and surface recorded data. The method can provide models with the necessary resolution for migration. Also, the resulting models can be used as good initial models for FWI. This work was presented during the EAGE conference & exhibition 2016 and is also the subject of a journal paper submitted to *Geophysical Prospecting* journal.

- **Chapter 6**: I present in this chapter the application of the proposed inversion method to a real data set. The data are acquired by CGG from North-Western offshore Australia. I invert for \((v_h, \epsilon, \eta)\) parameterization using a logarithmic phase only objective function. The accuracy of the inversion results is verified using data comparison and well-log information. Additionally, I use migration and angle gathers to quality control (QC) the inverted models. The tests show the usefulness of the proposed inversion approach for real data.

- **Chapter 7**: I present the main conclusions and recommendations for future research.
Chapter 2

Traveltime sensitivity kernels for wave equation tomography using the unwrapped phase

Summary

Wave equation tomography attempts to improve on traveltime tomography, by better adhering to the requirements of our finite-frequency data. Conventional wave equation tomography, based on the first-order Born approximation followed by cross-correlation traveltime lag measurement, or on the Rytov approximation for the phase, yields the standard hollow banana sensitivity kernel indicating that the measured traveltime at a point is insensitive to perturbations along the ray path at certain finite frequencies. Using the instantaneous traveltime, which can unwrap the phase of the signal, instead of the cross-correlation lag, I derive new finite-frequency traveltime sensitivity kernels. The kernel reflects more the model-data dependency we typically encounter in full waveform inversion. This result confirms that the hollow banana shape is borne of the cross-correlation lag measurement, which exposes the Born approximations weakness in representing transmitted waves. The instantaneous traveltime can, thus, mitigate the additional component of nonlinearity introduced by the hollow banana sensitivity kernels in finite frequency traveltime tomography. The instantaneous traveltime represents the unwrapped phase of Rytov approximation, therefore, it is a good alternative to Born and Rytov to compute the misfit function for wave equation tomography. I show the limitations of the cross-correlation

The work presented in this chapter is related to a publication:

associated with Born approximation for traveltime lag measurement when the source signatures of the measured and modeled data are different. The instantaneous traveltime is proven to be less sensitive to the distortions in the data signature. The unwrapped phase full banana shape of the sensitivity kernels shows smoother update compared to the banana-doughnut kernels. The measurement of the traveltime delay caused by a small spherical anomaly, embedded into a 3-D homogeneous model, supports the full banana sensitivity assertion for the unwrapped phase.

2.1 Introduction

In seismic inversion, the goal is to find an accurate subsurface velocity model that produces, with the proper physical simulation, data similar to the observed ones. For this task, many inversion schemes exist spanning the whole wavefield theoretical frequency spectrum from traveltime tomography to full waveform inversion. For traveltime tomography, a high-frequency assumption is typically employed, and the objective function is based on the misfit in arrival traveltimes (Dziewonski, 1984; Van Der Hilst et al., 1997). The model is, thus, updated by the back-projection of the traveltime residuals along rays or using the adjoint state. Nevertheless, ray tracing based methods fail to model the wave behavior in complex media and more attention, recently, has been paid to finite frequency based analysis and wavefield modeling (Woodward, 1992; Marquering et al., 1998; Dahlen et al., 2000; Montelli et al., 2004; Tromp et al., 2005) to overcome these shortcomings.

In full waveform inversion (FWI) (Tarantola, 1984; Virieux and Operto, 2009), the misfit represents the difference between the modeled and observed data. It takes advantage of the whole broadband data, which gives better resolution compared to traveltime tomography. However, the biggest limitation of FWI is the highly nonlinear nature of the misfit function (Bunks et al., 1995; Virieux and Operto, 2009).

Another approach referred to usually as wave equation tomography (WET) (De-
vaney 1984; Pratt and Goulty 1991; Luo and Schuster 1991; Woodward 1992) had been proposed in the past to tackle the same problem. Here, the misfit function for WET is given in traveltimes, but the modeling and model update is based on solving the wave equation and specifically on the wavefields. After generating synthetic data using a predicted model, the traveltime difference is generally measured by identifying the maximum of the cross-correlation between two waveforms (observed and modeled). Though this misfit function, like the conventional high frequency asymptotic based traveltime tomography, is quasi-convex, computing the maximum of the cross-correlation is only valid under certain circumstances (Hörmann et al. 2002; De Hoop and Van Der Hilst 2005; Van Leeuwen and Mulder 2010). Cross-correlation computes the phase shift correctly only when the source spectrum of the modeled data and the real data are in phase.

Since WET is based on the misfit in finite frequency traveltime, the inversion part relies on identifying regions in the model that could have contributed to this misfit. Mapping such areas is commonly referred to as the traveltime sensitivity kernel. The first step is to use first order Born approximation (Beydoun and Tarantola 1988; Woodward 1992; Bleistein et al. 2000; Aki and Richards 2002) to derive an equation for the wavefield perturbation (Woodward 1992; Marquering et al. 1998). Then, cross-correlation of the background wavefield with the total wavefield is applied to compute the traveltime sensitivity kernels (Marquering et al. 1998; Dahlen et al. 2000; Hung et al. 2000; Tromp et al. 2005). Another approach utilizes the Rytov approximation, which gives the phase perturbation as a function of the Born wavefield perturbation (Snieder and Lomax 1996; Jocker et al. 2006; Liu et al. 2009). For both methods, a non-intuitive observation was first made by Woodward (1992), that the finite frequency traveltime has zero sensitivity to slowness perturbation along the ray-path. Marquering et al. (1999) and many other authors confirmed the same observations. Such sensitivity kernels are based on the Born approximation, which
yields poor approximation for transmitted waves (Bleistein et al., 2000). For the Rytov approximation, a linear relation between phase and traveltime is used to derive the sensitivity kernels. In fact, for real inversion problems, the Rytov approximation for phase perturbation should be unwrapped to compute accurate data misfit (Géli et al., 2007). Thus, limitations in the Born and the Rytov approximations raise serious questions to the accuracy and applicability of the banana-doughnut kernels.

One can also extract the traveltime information from wavefield by using an equation dedicated to unwrapping the phase (Stoffa et al., 1974; Shin et al., 2003; Choi and Alkhalifah, 2011; Choi et al., 2011). Shin et al. (2003) used the derivative of the wavefield with respect to the angular frequency to estimate traveltimes. A similar idea was utilized by Choi et al. (2011) to automatically pick the first arrival events by applying a damping factor. I refer to the computed traveltime through measuring the frequency derivative of the wavefield as the instantaneous traveltime. It provides us with an alternative finite frequency traveltime measurement. The question I raise and address in this chapter is whether this is a good alternative measure of finite-frequency traveltime than those associated with the cross-correlation lag of the measured and simulated wavefields.

I investigate the prospect of using the instantaneous traveltime as a source for finite frequency traveltimes. I analyze the traveltime sensitivity kernels, and I compare them to the banana-doughnut kernels. In this chapter, I only consider the kernels for the direct P-waves. I begin by reviewing Born and Rytov approximations and the resulting sensitivity kernels. I also introduce the instantaneous traveltime and derive the associated sensitivity kernels. I show that the instantaneous traveltime reduces to the unwrapped phase version of Rytov approximation. Through numerical examples, I show how the instantaneous traveltime is more suitable to resolve small scatterers compared to the cross-correlation approach, even when the source spectrum of the modeled and observed data are not identical. I also present the sensitivity kernels
obtained for a simple velocity model, as well as a more complex one and I compare them with the banana-doughnut kernels. I show how the unwrapped phase is a good alternative approach for WET through analysis of the sensitivity kernels’ properties. Finally, I consider the effect of a small spherical anomaly embedded inside a 3-D homogeneous velocity model. According to the banana-doughnut kernel observations, such anomalies are hidden inside the banana-doughnut hole for certain frequencies. Based on the unwrapped phase, I will show that a small anomaly affects the wavefield and thus the measured traveltimes at the receivers. This final test confirms that instantaneous traveltime is a good alternative approach for wave equation tomography.

### 2.2 Theory

The traveltime sensitivity kernels or the Fréchet derivative are computed by mapping the region of the velocity model that contributes to the traveltime recorded at a specific receiver. Since the traveltime residual in wave equation tomography is back projected on that region, the kernels estimation affects the accuracy of the inversion. In this section, I review the theoretical background behind the finite frequency traveltime sensitivity kernels.

Assuming that the squared slowness is the sum of a smoothly varying background, $s_0$, and a weak perturbation, $\delta s$, the first-order Born approximation could be used to model the scattered wavefield. Thus, the total wavefield is the sum of a background wavefield, $U_0$, and a scattered wavefield, $\delta U_B$ as follows:

$$U(x_r, x_s, \omega) = U_0(x_r, x_s, \omega) + \delta U_B(x_r, x_s, \omega).$$ (2.1)

Snieder and Lomax (1996) is given by equation (2.2),

$$
\delta U_B (x_r, x_s, \omega) = \omega^2 \int_x G_0 (x_r, x, \omega) G_0 (x, x_s, \omega) \delta s (x) \, dx^3,
$$

where $G_0 (x, x_s, \omega)$ is the Green’s function for a source located at $x_s$ and recorded at the scatterer point $x$. $G_0 (x_r, x, \omega)$ is the Green’s function ignited at the scatterer $x$ and recorded at the receiver $x_r$. $\delta s (x)$ is the squared slowness perturbation.

The Born approximation can be written in the form (Woodward, 1992; Marquering et al., 1998),

$$
\delta U_B (x_r, x_s, \omega) = \int_x K_{U_B} (x_r, x, x_s, \omega) \delta s (x) \, dx^3,
$$

where $K_{U_B} (x_r, x, x_s, \omega) = \omega^2 G_0 (x_r, x, \omega) G_0 (x, x_s, \omega)$ is the waveform sensitivity kernel.

The Born approximation is a scattering location perturbation based on small scatterers in size and magnitude. This condition is more crucial when the Born approximation is used to represent transmitted waves, as any phase shift induced by the scatterer will not be reflected in the phase term. Considering a small scatterer located on the ray-path and Green’s functions of the form $G_0 (x, x_s, \omega) = A_{x_s x} e^{i\varphi_{x_s x}}$, the phase of Born perturbation equation (2.3) is $\varphi_{x_s x} + \varphi_{xx_r}$. The phase of the direct wave is exactly the same as the perturbation phase: $\varphi_{x_s x} = \varphi_{x_s x} + \varphi_{xx_r}$. As a result, the effect of the scatterer on the perturbed wavefield is visible on the amplitude and not on the phase. As such, Bleistein et al. (2000) demonstrated that Born approximation is inadequate for transmitted waves. Although the size of the effective scatterer can be larger for low frequencies, as Born is also inherently a low-frequency approximation, the problem still exists.
2.2.1 Cross-correlation and Rytov traveltime sensitivity kernels

The time domain cross-correlation, equation (2.4), is used to measure the similarity between the observed wavefield $U(t)$, and the modeled wavefield $U_0(t)$. The observed wavefield differs from the modeled wavefield by a small Born perturbation $\delta U_B(t)$.

$$C(t) = \int_{\tau_1}^{\tau_2} U_0(\tau-t) U(\tau) d\tau. \quad (2.4)$$

It typically has a maximum at $\delta t$ measured from the zero lag, given by

$$\delta t = \arg\max_t C(t). \quad (2.5)$$

Using the cross-correlation property equation (2.5), and based on the first order Born wavefield perturbation, Marquering et al. (1999) derived an equation to compute the traveltime perturbation $\delta t$,

$$\delta t = \frac{\int_{\tau_1}^{\tau_2} \dot{U}_0(\tau) \delta U_B(\tau) d\tau}{\int_{\tau_1}^{\tau_2} \ddot{U}_0(\tau) U_0(\tau) d\tau}. \quad (2.6)$$

where $\dot{U}_0(\tau)$ and $\ddot{U}_0(\tau)$ denote the first and second order derivatives with respect to time.

The cross-correlation based traveltime sensitivity kernel, or as Marquering et al. (1999) denoted the Born kernel, is given by,

$$K_t^{\text{Cross}}(x_r, x, x_s) = \frac{\int_{\tau_1}^{\tau_2} \dot{U}_0(\tau) K_{U_B}(x_r, x, x_s, \tau) d\tau}{\int_{\tau_1}^{\tau_2} \ddot{U}_0(\tau) U_0(\tau) d\tau}. \quad (2.7)$$

To compute the traveltime sensitivity kernel $K_t^{\text{Cross}}(x_r, x, x_s)$, Marquering et al. (1999) used surface mode summation to evaluate seismograms and the wavefield sensitivity kernel. Accordingly, the resulting kernel is bandlimited, and it includes the effect of all frequencies of the source wavelet.
To evaluate the monochromatic version of the cross-correlation sensitivity kernel, I compute the Green’s functions and the wavefield single frequency sensitivity kernel $K_{UB}(x, x, x_s, \omega)$ in the frequency domain. Then, I apply an inverse Fourier transform.

Considering $U_0^1(t)$ the inverse Fourier transform of the Green’s function $G_0(x, x_s, \omega)$, and the wavefield kernel $K_{UB}^1(x, x, x_s, t)$ the inverse Fourier transform of $K_{UB}^0(x, x, x_s, \omega)$, the monochromatic sensitivity kernel is given by,

$$K_t^{cross}(x_r, x, x_s, \omega) = \frac{\int_{\tau_1}^{\tau_2} U_0^1(\tau) K_{UB}^1(x, x, x_s, \tau) d\tau}{\int_{\tau_1}^{\tau_2} U_0^1(\tau) d\tau}. \quad (2.8)$$

I add a dependency on the angular frequency $\omega$ in the monochromatic sensitivity kernel’s notation to differentiate it from the bandlimited one. I will show later in this chapter, how to estimate the bandlimited kernels based on a weighted summation over monochromatic kernels [Rickett, 2000].

Alternatively, the Rytov approximation [Woodward, 1992; Snieder and Lomax 1996; Joker et al., 2006; Liu et al., 2009] can be used to derive the monochromatic traveltime sensitivity kernels. It is more adapted for the representation of transmitted waves than Born approximation. The first-order Rytov approximation provides a relationship between the slowness squared perturbation $\delta s(x)$ and the complex phase perturbation $\delta \phi_R(x_r, x_s, \omega)$,

$$\delta \phi_R(x_r, x_s, \omega) = \frac{\delta U_B(x_r, x_s, \omega)}{U_0(x_r, x_s, \omega)} = \frac{\omega^2}{U_0(x_r, x_s, \omega)} \int_x G_0(x_r, x, \omega) G_0(x, x_s, \omega) \delta s(x) d x^3. \quad (2.9)$$

Assuming that the phase is a linear function of traveltime (i.e. the high frequency approximation), $\phi_R = i\omega t$. The traveltime perturbation is the imaginary part of the phase perturbation divided by the angular frequency $\omega$ [Woodward 1992; Snieder]
\[ \delta t = \Im \left( \delta \phi_R (x_r, x_s, \omega) \right) = \Im \left( \frac{\delta U_B (x_r, x_s, \omega)}{U_0 (x_r, x_s, \omega)} \right). \] (2.10)

The symbol \( \Im (.) \) stands for the imaginary part.

Finally, the single frequency Rytov traveltime sensitivity kernel is computed as,

\[ K_{Rytov}^t (x_r, x, x_s, \omega) = \Im \left( \frac{K_{UB} (x_r, x, x_s, \omega)}{U_0 (x_r, x_s, \omega)} \right). \] (2.11)

### 2.2.2 Unwrapped phase sensitivity kernels

Traveltime measurement using cross-correlation is only valid under certain circumstances \( \text{(Hörmann et al., 2002)} \). In fact, \( \text{De Hoop and Van Der Hilst (2005)} \) showed that cross-correlation computes the traveltime difference correctly only when the source spectrum of the modeled data and the real data are identical. For the Rytov approximation, a high frequency assumption for the phase is used to separate the traveltime term. Thus, for both methods, there exist some limitations which may lead to inaccurate results when used for traveltime sensitivity kernels derivation and finite frequency traveltime inversion.

Alternatively, I use another approach to compute the traveltime misfit. Taking the derivative of the wavefield with respect to the angular frequency, dividing by the wavefield, and finally taking the imaginary part can be a good alternative to cross-correlation based traveltimes. The result is referred to as the instantaneous traveltime \( \text{(Choi and Alkhalifah, 2011; Choi et al., 2011)} \),

\[ t = \Im \left( \frac{\partial U (x_r, x_s, \omega)}{\partial \omega} \right). \] (2.12)

Instantaneous traveltime measurements, like the cross-correlation or the Rytov
approach, is based on wavefields. However, there is no high frequency assumption on seismic data which results in a more practical method. Using this approach, I compute traveltimes automatically without the need for picking. The instantaneous traveltime, able to unwrap the phase, gives accurate estimates of the traveltime delays. It is the unwrapped phase version of Rytov approximation, and could be used to derive the sensitivity kernels.

To compute the derivative of the wavefield, I utilize the Helmholtz wave equation,

$$S(\omega)U(\omega) = f,$$

(2.13)

where $S$ is the impedance matrix, $U$ is the wavefield and $f$ the source. Taking the derivative of equation (2.13) with respect to $\omega$ yields,

$$S(\omega) \frac{\partial U(\omega)}{\partial \omega} = - \frac{\partial S(\omega)}{\partial \omega} U(\omega).$$

(2.14)

Since I solve for the Green’s functions, the source $f$ is frequency independent and the term $\partial f / \partial \omega$ is not included in the right hand side of equation (2.14). Note that the derivative is obtained by solving the same Helmholtz equation, but with a virtual source $- \frac{\partial S(\omega)}{\partial \omega} U(\omega)$. Assuming I use a direct solver, a considerable part of the computational cost is for the impedance matrix decomposition. The same impedance matrix is used for both equations (2.13) and (2.14). As a result, the computational cost is not affected heavily by computing the wavefield derivative.

Considering a wavefield of the form of equation (2.1), the instantaneous traveltime is (Djebbi and Alkhalifah, 2012),

$$t = \Im \left( \frac{\partial U(x_r, x_s, \omega)}{\partial \omega} \right) = \Im \left( \frac{\partial U_0(x_r, x_s, \omega)}{\partial \omega} \right) + \Im \left( \frac{\partial (\delta U_B(x_r, x_s, \omega))}{\partial \omega} \right).$$

(2.15)

For small perturbations, I use the Born approximation giving $U(x_r, x_s, \omega) \approx U_0(x_r, x_s, \omega)$. 

The instantaneous traveltime becomes
\[ t \approx \Im \left( \frac{\partial U_0(x_r, x_s, \omega)}{U_0(x_r, x_s, \omega)} \right) + \Im \left( \frac{\partial U_B(x_r, x_s, \omega)}{U_0(x_r, x_s, \omega)} \right) = t_0 + \delta t. \] (2.16)

Finally, using the instantaneous traveltime I derive an equation for the traveltime perturbation for a weak wavefield perturbation,
\[ \delta t = \Im \left( \frac{\partial U_B(x_r, x_s, \omega)}{U_0(x_r, x_s, \omega)} \right). \] (2.17)

The monochromatic traveltime sensitivity kernel using the instantaneous traveltime is given by,
\[ K_{\text{Inst}}(x_r, x, x_s) = \Im \left( \frac{\partial (K_{U_B}(x_r, x_s, \omega))}{U_0(x_r, x_s, \omega)} \right). \] (2.18)

The derivative of the wavefield kernel with respect to \( \omega \) is,
\[ \frac{\partial (K_{U_B}(x_r, x_s, \omega))}{\partial \omega} = 2\omega G_0(x_r, x, \omega) G_0(x_s, \omega) \] (2.19)
\[ + \omega^2 \left[ \frac{\partial G_0(x_r, x, \omega)}{\partial \omega} G_0(x_s, \omega) + G_0(x_r, x, \omega) \frac{\partial G_0(x_s, \omega)}{\partial \omega} \right], \]
which, as I showed earlier, is computed using a Helmholtz solver for a virtual source as described in equation (2.14).

2.2.3 Relation between unwrapped phase and Rytov sensitivity kernels

The phase perturbation given by the Rytov approximation (equation (2.9)) is accurate if the traveltime residual is within a half cycle. Otherwise, we will need to unwrap the Rytov phase to obtain accurate traveltime differences. When deriving the sensitivity kernels based on the Rytov approximation, a linear relationship between the phase perturbation and traveltime perturbation is considered. The sensitivity kernels based on the instantaneous traveltime have similar forms as the Rytov ones, however, the
phase is now unwrapped.

If I consider the Rytov wavefield approximation

$$U_R(x_r, x_s, \omega) = U_0(x_r, x_s, \omega) e^{i\delta \phi_R(x_r, x_s, \omega)},$$

(2.20)

I can use the instantaneous traveltime given by equation (2.12) to derive an equation for the traveltime perturbation $\delta t$, as follows:

$$t = \Im \left( \frac{\partial U_0(x_r, x_s, \omega)}{\partial \omega} \right)$$

(2.21)

$$= \Im \left( \frac{\partial U_0(x_r, x_s, \omega) e^{i\delta \phi_R(x_r, x_s, \omega)}}{U(x_r, x_s, \omega)} \right)$$

$$= \Im \left( \frac{\partial U_0(x_r, x_s, \omega) e^{i\delta \phi_R(x_r, x_s, \omega)}}{U(x_r, x_s, \omega)} \right) + \Im \left( \frac{U_0(x_r, x_s, \omega) e^{i\delta \phi_R(x_r, x_s, \omega)} \partial \delta \phi_R(x_r, x_s, \omega)}{U(x_r, x_s, \omega)} \right)$$

$$= \Im \left( \frac{\partial U_0(x_r, x_s, \omega)}{U_0(x_r, x_s, \omega)} \right) + \Im \left( \frac{\partial \delta \phi_R(x_r, x_s, \omega)}{\partial \omega} \right)$$

$$= t_0 + \delta t.$$

The traveltime perturbation using the unwrapped phase is, then, given by

$$\delta t_{\text{inst}} = \Im \left( \frac{\partial \delta \phi_R(x_r, x_s, \omega)}{\partial \omega} \right).$$

(2.22)

In the Rytov traveltime perturbation equation (2.10) a linear relation between phase and traveltime is considered, $\delta \phi = i\omega \delta t$. Then, for this special case,

$$\delta t_{\text{Rytov}} = \delta t_{\text{inst}}.$$

(2.23)

As a result, the instantaneous traveltime sensitivity kernels are the unwrapped versions of the Rytov ones. For the general case, I consider an unknown relation between the phase and traveltime. Then, phase unwrapping through the instantaneous trav-
time is necessary to get the traveltime perturbation,

\[
\delta t_{\text{inst}} = \Im \left( \frac{\partial \phi_R (\mathbf{x}_r, \mathbf{x}_s, \omega)}{\partial \omega} \right) = \Im \left( \frac{\partial \left( \frac{\delta U_B (\mathbf{x}_r, \mathbf{x}_s, \omega)}{U_0 (\mathbf{x}_r, \mathbf{x}_s, \omega)} \right)}{\partial \omega} \right) = \Im \left( \frac{\partial \left( \frac{\delta U_B (\mathbf{x}_r, \mathbf{x}_s, \omega)}{U_0 (\mathbf{x}_r, \mathbf{x}_s, \omega)} \right)}{\partial \omega} \right). \tag{2.24}
\]

For weak phase perturbation, the contribution of the second term in equation (2.24) is negligible. This was verified numerically for a bandlimited kernel in a 3-D \( v(z) \) model. Accordingly, I neglect the term \( \frac{\delta U_B (\mathbf{x}_r, \mathbf{x}_s, \omega)}{U_0 (\mathbf{x}_r, \mathbf{x}_s, \omega)} \partial U_0 (\mathbf{x}_r, \mathbf{x}_s, \omega) \) and the instantaneous traveltime perturbation corresponds to the approximation (2.17),

\[
\delta t_{\text{inst}} \approx \Im \left( \frac{\partial \delta U_B (\mathbf{x}_r, \mathbf{x}_s, \omega)}{U_0 (\mathbf{x}_r, \mathbf{x}_s, \omega)} \right). \tag{2.25}
\]

Equation (2.25) is the general form of Rytov phase approximation. It unwraps the phase and is used to compute the sensitivity kernels given by equation (2.18).

### 2.2.4 Broadband kernels

The monochromatic sensitivity kernels, \( K_t^{\text{Cross}} (\mathbf{x}_r, \mathbf{x}, \mathbf{x}_s, \omega) \), \( K_t^{\text{Rytov}} (\mathbf{x}_r, \mathbf{x}, \mathbf{x}_s, \omega) \) and \( K_t^{\text{Inst}} (\mathbf{x}_r, \mathbf{x}, \mathbf{x}_s, \omega) \) are used to compute the band-limited kernels (Woodward 1992; Marquering et al., 1999; Snieder and Lomax, 1996; Djebbi and Alkhalifah, 2013b). A weighted summation over frequencies is applied. The weighting function can be a Gaussian distribution or the amplitude spectrum of the Ricker wavelet.

I choose the amplitude spectrum of a Ricker wavelet as the weighting function. The broadband sensitivity kernel is given by,

\[
K_t^i (\mathbf{x}_r, \mathbf{x}, \mathbf{x}_s) = \sum_{\omega} P(\omega) K_t^i (\mathbf{x}_r, \mathbf{x}, \mathbf{x}_s, \omega); \quad i = \text{Cross, Rytov, Inst.} \tag{2.26}
\]
The weighting function is:

\[
P(\omega) = 4\sqrt{\frac{\omega^2}{\omega_0^3}} e^{-\frac{\omega^2}{\omega_0^2}},
\]

(2.27)

where, \( \omega = 2\pi f \) is the angular frequency, \( \omega_0 = 2\pi f_0 \) and \( f_0 \) is the peak frequency of the Ricker wavelet.

Since both Rytov and the cross-correlation give the same type of sensitivity kernels (Rickett 2000; Liu et al. 2009), I will use the term banana-doughnut kernels instead of Rytov or cross-correlation in my analysis.

### 2.3 Cross-correlation limitations in traveltine delay measurement

In conventional WET implementations, cross-correlation is used to compute the traveltime lag \( \delta t \). Thus, finding \( \delta t \) measured from the zero lag to the maximum energy of the cross-correlation represents the shifts required for the synthetic seismogram to give the best match to the observed seismogram (equation (2.5)).

Hörmann et al. (2002) and De Hoop and Van Der Hilst (2005) showed that cross-correlation computes the traveltime difference correctly only when the source spectrum of the modeled data and the real data are in phase, which is not always the case even with wavelet estimation techniques. Van Leeuwen and Mulder (2010) derived a new misfit based on a weighted norm of the cross-correlation. Such regularization makes the misfit function less sensitive to the difference in wavelet spectrum.

Figure (2.1) shows two different scenarios for traveltime lag measurement. In Figure 2.1(a) the two wavelets have the same shape and time difference between the two is \( \delta t = 0.1 \) s. However, in Figure 2.1(b) one of the wavelets is subjected to a phase rotation equal to \( \pi/2 \). In this numerical test, I measure the traveltime difference by picking the lag corresponding to the maximum of the cross-correlation. Also, I use the instantaneous traveltime given by equation (2.12). The traveltime lag is basically
the difference between the measured instantaneous traveltimes for both.

Figure 2.1: Two different scenarios. (a) Two Ricker wavelets, one of the two is advanced with $\delta t = 0.1$ s with no phase rotation. (b) Two Ricker wavelets, one of the two wavelets is advanced with $\delta t = 0.1$ s with a phase rotation $\pi/2$.

Figure 2.2 shows the cross-correlation for the two scenarios presented in Figure 2.1. The maximum of the cross-correlation corresponds to the measured travelt ime lag. When the two wavelets have the same signature, the maximum of the cross-correlation is exactly the true traveltime time lag $\delta t = 0.1$ s. However, when one of the two wavelets is phase rotated, cross-correlation estimates an inaccurate traveltime delay. The maximum of the cross-correlation corresponds to $\delta t = 0.08$ s.

In Figure 2.3 I show the traveltime difference measured using the instantaneous traveltime. The two figures present the measured $\delta t$ as a function of the frequency.
Figure 2.2: Crosscorrelation of the two scenarios depicted in figure 2.1. (a) No phase rotation, the maximum of the cross-correlation corresponds to $\delta t = 0.1 \text{s}$. (b) $\pi/2$ phase rotation, the maximum of the cross-correlation corresponds to $\delta t = 0.08 \text{s}$ and cross-correlation is inaccurate.

For all the frequencies equal or larger than $2 \text{Hz}$ the instantaneous traveltime can predict the exact values of the delay for both cases. I conclude that the instantaneous traveltime is less sensitive to wavelet distortions. As a result, the instantaneous traveltime is an accurate alternative measure of traveltime misfit to the cross-correlation approach. Thus, it could be used to derive the traveltime sensitivity kernels instead of the conventional approaches.
Figure 2.3: Traveltime difference measurement using the instantaneous traveltime (unwrapped phase). (a) No phase rotation (b) $\pi/2$ phase rotation. For both cases, the traveltime lag $\delta t$ is given as a function of the frequency, for most of the frequencies the instantaneous traveltime is able to compute the exact traveltime difference. I pick the lag for the wavelet peak frequency $f = 10 \text{ Hz}$.

2.4 Traveltime sensitivity kernels

In this section, I compare the sensitivity kernels using the instantaneous traveltime and the banana-doughnut ones through numerical examples. I consider a simple 3-D model with an increasing velocity with depth and a more complex Marmousi 2-D model.
2.4.1 3-D simple model

I consider a linearly increasing velocity model with depth, \( v(z) = v_0 + \alpha z \), where \( v_0 = 2.0 \text{ km s}^{-1} \) and \( \alpha = 0.5 \text{ s}^{-1} \). For this type of velocity model, a simple two-point ray tracing can be used to compute approximate Green’s functions (Rickett, 2000). I consider a source and a receiver placed on the surface to mimic the behavior of diving waves, which are useful to invert for near surface structures in WET.

Figure 2.4 shows the single frequency kernels for \( f = 30 \text{ Hz} \). To compute the broadband kernels, I apply a weighted summation of the monochromatic kernels. The weighting function consists of the normalized amplitude spectrum of a Ricker wavelet with peak frequency equal 30 Hz (equation (2.27)).

![Figure 2.4: 30 Hz single frequency traveltime sensitivity kernels. (a) Crosscorrelation. (b) Unwrapped phase (instantaneous traveltime).](image)

I show in Figure 2.5 the resulting, broadband 30 Hz, traveltime sensitivity kernels based on cross-correlation and on the instantaneous traveltime. A constructive interference of the monochromatic kernels can be noticed in the banana shape zone around the ray-path. In the rest of the model, a destructive interference removes the effect of
any slowness perturbation on traveltimes. Figure 2.5(a) shows the sensitivity kernel obtained using the Born approximation with cross-correlation lags. We observe the same features obtained first by Woodward (1992) then confirmed by many other authors (Marquering et al., 1999; Dahlen et al., 2000; Hung et al., 2000; Rickett, 2000). The sensitivity kernel has a hollow banana shape. This is the case even if we use the Rytov approximation (Rickett, 2000; Snieder and Lomax, 1996; Joker et al., 2006; Liu et al., 2009). In the cross-section, it has a doughnut shape. The observation is counter intuitive since the traveltime is insensitive to velocity perturbations along the ray-path. In the high frequency limit, these sensitivity kernels are expected to thin out and reduce to the ray path. However, the doughnut effect collapse is not clear, which raises a possible singularity in this definition, suggesting the traveltime defined by the high frequency asymptotic approximation is different than that associated with correlation lag in the high frequency limit.

![Figure 2.5: 30 Hz Broadband traveltime sensitivity kernels for v(z). (a) Crosscorrelation. (b) Unwrapped phase (instantaneous traveltime).](image)

On the other hand, the instantaneous traveltime yields the kernel shown in Figure 2.5(b). The traveltime kernel has a banana shape, but now the banana is a
plain one. Traveltimes are sensitive to slowness perturbation along the geometric ray-path. The hollow banana shape is the direct result of using cross-correlation traveltimes. Alternative methods for traveltime measurement results in a different kernel shape. Van Leeuwen and Mulder (2010) developed a wave equation tomography by regularizing the cross-correlation. This regularization makes the sensitivity kernels more stable, and they have a plain banana shape. Here, in our instantaneous traveltime time method, we get a realistic plain kernel directly with lower sensitivity to the wavelet’s spectrum distortion. For the Rytov approximation, it has limitations caused by assuming a linear relationship between phase and traveltime. This approximation suffers from the phase wrapping phenomena. Thus the instantaneous traveltime is a good alternative since it unwraps the phase of the Rytov. For realistic models, the instantaneous traveltime includes a derivative operator. When applied to the wavefield, it can overcome some of the existing non-linearities in the phase. Figure 2.6 shows a slice at $x = 4\,km$ of the two kernels. The broadband slice is overlaid onto the monochromatic one to get a better understanding of the sensitivity kernels’ properties.

The instantaneous traveltime provides an excellent alternative to measure finite-frequency traveltime misfit. It is less sensitive to wavelet distortion and has smooth kernels. Thus, using the instantaneous traveltime for WET is a good option to examine.

2.4.2 Marmousi model

The Marmousi model has served the seismic exploration for a long time as a benchmark for imaging and inversion algorithms. Thus, I compute the sensitivity kernels for its smoothed version given by Figure 2.7. The original Marmousi model is smoothed using triangular smoothing with a radius of 40 samples in both directions. The source is located ($x_s = 3.0\,km$, $z_s = 0.1\,km$) and the receiver is located at
Figure 2.6: Slices at $x = 4\, km$ of the $30\, Hz$ traveltime sensitivity kernels. (a) Cross-correlation. (b) Unwrapped phase (instantaneous traveltime). The broadband slices in continuous blue line are overlaid to the monochromatic $30\, Hz$ slices represented in dashed red line.

$(x_r = 7.0\, km, z_r = 0.1\, km)$. The Green’s functions are computed using a Helmholtz equation solver based on LU decomposition.

Figure 2.8 shows the sensitivity kernels for the Marmousi model. The kernels using the unwrapped phase and cross-correlation cover the same parts of the model. They
Figure 2.7: Smoothed Marmousi model.

have a similar behavior with some differences in their amplitude. A closer look, given by Figure 2.9 reveals the smooth nature of the instantaneous traveltime sensitivity kernel.

Figure 2.8: 30 Hz Broadband traveltime sensitivity kernels for the Marmousi model. (a) Crosscorrelation. (b) Unwrapped phase (instantaneous traveltime).

Furthermore, I show the kernel for a larger offset. In this case, the kernel will
Figure 2.9: Slices at $x = 4 \, km$ of the 30 Hz Broadband traveltime sensitivity kernels for the Marmousi model. (a) Crosscorrelation. (b) Unwrapped phase (instantaneous traveltime).

cover a deeper area of the Marmousi model. Since the velocity in deep parts is more complex, the kernel has a more complicated shape. I show in Figure 2.10 the sensitivity kernel computed using the instantaneous traveltime. The figure shows the single 30 Hz frequency kernel and the broadband one obtained by the weighted summation over frequencies. The broadband kernel has fewer oscillations compared to the single frequency one. However, because of the complex velocity, oscillations in the area far from the ray path does not cancel perfectly. The sensitivity kernel shows the large area that could be updated using finite frequency traveltimes compared to
ray theory based updates.

![Figure 2.10: Traveltime sensitivity kernels using the unwrapped phase for a source-receiver offset \( h = 8.2 \text{ km} \). (a) 30 Hz single frequency kernel (b) 30 Hz broadband kernel.](image)

2.5 Traveltime measurement for a homogeneous background medium with a 3-D spherical anomaly

In this section, I consider a homogeneous background velocity model with a small 3-D smooth spherical anomaly. The background velocity is \( v_0 = 2.0 \text{ km/s} \). The model size is \((2.0 \times 1.0 \times 1.0) \text{ km}^3\). The anomaly is positive with 10% velocity perturbation. The diameter of the anomaly is \( d_a = \lambda = 0.2 \text{ km} \). The distance between the source and the anomaly center is \( x_a = 0.5 \text{ km} \). I show in Figure 2.11 a slice at \( y = 0.5 \text{ km} \) of the velocity model. The source is placed at \( \{x_s, y_s, z_s\} = \{0.0, 0.5, 0.5\} \) and the data is recorded along one horizontal and two vertical arrays of equidistant receivers, as shown with the red lines. I use a finite difference time domain scheme to solve
the wave equation and model the wavefields. The source function is a Ricker wavelet with a peak frequency $f_0 = 10 \text{ Hz}$.

![Figure 2.11: Velocity model slice at $y = 0.5 \text{ km}$. The anomaly is placed at a distance $x_a = 0.5 \text{ km}$ from the source. The source is shown with the blue star. The red lines show receivers’ location. I place receivers on one horizontal and two vertical lines.](image)

The objective of this simple test is the analysis of the anomaly’s effect on the measured traveltimes. According to the banana-doughnut theory, a small anomaly, compared to the wavelength, can not be detected using cross-correlation. A small anomaly can be hidden inside the hole in the banana as shown in Figure 2.12. Actually, for a short distance between the source and the receiver, the traveltime is sensitive to some parts of the velocity anomaly (Figure 2.12(a)). However, as long as we increase the source-receiver offset to $2.0 \text{ km}$ (Figure 2.12(c)) the anomaly is hidden and can not be resolved. For the last case, there is zero traveltime difference which explains the hollow banana shape of the sensitivity kernel.

Based on the instantaneous traveltime equation (2.12), I use the methodology described by Choi et al. (2011) to automatically compute traveltimes. A damping factor is applied to the instantaneous traveltime. The traveltime difference is the difference between the instantaneous traveltimes for the homogeneous model and the one with the anomaly. I compare the measured traveltimes with the ones computed using cross-correlation.
Figure 2.12: Banana-doughnut traveltime sensitivity kernels slices overlaid to a small anomaly for different receiver locations. The anomaly diameter is $d_a = \lambda = 0.2 \text{ km}$, where $\lambda$ is the wavelength corresponding to the peak frequency $f_0 = 10 \text{ Hz}$.

Figure 2.13 shows the traveltime delay measured along the horizontal array of receivers. The traveltime difference, measured using the Eikonal solver, in black, reflect the conventional ray time perturbation and has larger values compared to the two other methods. On the other hand, cross-correlation (blue) and unwrapped phase (red) methods show smaller traveltime perturbations. They both decrease with offset reflecting the finite frequency spreading effect (Fresnel zone), where the traveltime perturbation is spread over a region. However, when cross-correlation traveltime converges to zero delays for receivers far from the anomaly, the unwrapped phase is still detecting the effect of the anomaly.

I show in Figure 2.14 seismograms recorded for different locations along the horizontal array of receivers. I show seismograms for $x = \{0.6; 1.0; 1.4; 1.8\} \text{ km}$, the blue curve represent the seismogram for the homogeneous case whereas the red curve represents the case with an embedded anomaly. I also show the measured traveltime
Figure 2.13: Measured traveltimes along the horizontal array of receivers. Black line represents traveltime difference computed using an Eikonal solver, blue is the cross-correlation traveltime and red is the traveltime measured using the unwrapped phase. Delay measured using the Eikonal solver is constant after the anomaly, but is overestimated. For both cross-correlation and instantaneous traveltime, we observe the same decreasing tendency. Crosscorrelation traveltimes converge to zero for receivers far from the anomaly. However, the instantaneous traveltime is still detecting the effect of the anomaly. Notice that the inflection point near 1.8 km for the cross-correlation curve is a sampling artifact.

shifts using cross-correlation and the instantaneous traveltime. From Figure 2.14(a), the traveltime shift is visible, and it is well approximated using both methods. As long as we move along the $x$ direction, wavefront healing affects the seismograms and the difference decreases. The instantaneous traveltime, more sensitive to small scatterers, gives larger values of traveltime shifts compared to cross-correlation. I conclude that as predicted by the sensitivity kernels, the instantaneous traveltime can resolve small anomalies.

I display in Figures 2.15 and 2.16 traveltime lags computed for the vertical lines of receivers, at 1.0 km and 1.8 km respectively. Traveltime delays using the Eikonal, as shown before, give ray perturbation traveltimes. Traveltimes computed using cross-correlation have their minimum along the source-receiver horizontal axis, and converge to zero for receivers far from the anomaly. Though, like the cross-correlation, the
(a) $\Delta t_{\text{Cross}} = 7.8 \text{ ms}$; $\Delta t_{\text{Inst}} = 7.2 \text{ ms}$

(b) $\Delta t_{\text{Cross}} = 1.5 \text{ ms}$; $\Delta t_{\text{Inst}} = 2.7 \text{ ms}$

(c) $\Delta t_{\text{Cross}} = 0.6 \text{ ms}$; $\Delta t_{\text{Inst}} = 2.0 \text{ ms}$

(d) $\Delta t_{\text{Cross}} = 0.2 \text{ ms}$; $\Delta t_{\text{Inst}} = 1.8 \text{ ms}$

Figure 2.14: Recorded pressure seismograms along the horizontal array of receivers for different locations: (a) $x = 0.6 \text{ km}$ (b) $x = 1.0 \text{ km}$ (c) $x = 1.4 \text{ km}$ (d) $x = 1.8 \text{ km}$. The blue and red curves correspond to the homogeneous and the case with positive anomaly respectively. I show under each figure the measured traveltime shift using cross-correlation $\Delta t_{\text{Cross}}$, and using the instantaneous traveltime $\Delta t_{\text{Inst}}$.

In summary, in both finite frequency traveltime misfit measurements, the misfit stretches over a larger region courtesy of the Fresnel zone, reflecting the milder traveltime sensitivity of low frequencies to small anomalies. However, the instantaneous traveltime spreads the misfit keeping the maximum sensitivity along the ray path.

2.6 Conclusions

Wave equation tomography (WET) shares the same misfit measure as conventional traveltime tomography. However, the difference lies in the way traveltimes are com-
Figure 2.15: Measured traveltimes along the vertical array of receivers placed at $x = 1.0 \, \text{km}$. Black line represent traveltime computed using an Eikonal solver, blue is the cross-correlation traveltime and red is the traveltime measured using the unwrapped phase. Crosscorrelation traveltimes minimum is along the source-receiver axis. For the unwrapped phase, it predicts traveltimes with maximum values along $z = 0.5 \, \text{km}$, the source-receiver axis.

Figure 2.16: Measured traveltimes along the vertical array of receivers placed at $x = 1.8 \, \text{km}$. Black line represent traveltime computed using an Eikonal solver, blue is the cross-correlation traveltime and red is the traveltime measured using the unwrapped phase. Crosscorrelation traveltimes minimum is along the source-receiver direction and is almost zero. For the unwrapped phase, it predicts traveltimes with maximum values along $z = 0.5 \, \text{km}$, the source-receiver axis.
puted and the model is updated. Each traveltime measurement method has its own sensitivity kernel or Fréchet derivatives, which shows the parts of the model that will be updated during WET.

Conventionally, first order Born or Rytov approximation is used to derive these kernels and examine their spatial distribution. The Born or Rytov kernels have the same banana-doughnut shape with zero sensitivity along the geometrical ray-path.

In this chapter, I introduced an alternative approach to measure the traveltime misfit based on the unwrapped phase (or the instantaneous traveltime). This traveltime is a function of frequency, and thus, admits finite frequency features. I showed that the instantaneous traveltime is more sensitive to small anomalies compared to the cross-correlation based traveltimes. I also demonstrated that the instantaneous traveltime is the unwrapped phase version of the Rytov traveltime approximation. The unwrapped phase is used to derive the corresponding traveltime sensitivity kernels. The resulting kernels admit a plain banana shape with maximum sensitivity along the geometrical ray-path. Observations on a simple 3-D velocity model and the more complex Marmousi model shows smooth kernels. A simple test of traveltime lag measurement for a 3-D model with an embedded small spherical anomaly, confirms the plain sensitivity for the unwrapped phase. The unwrapped phase kernels are smoother and share with the high frequency asymptotic ray-path kernels the maximum sensitivity along the ray-path. Moreover, the instantaneous traveltime kernels account for real physical phenomena like dispersion and wavefront healing. The unwrapped phase forms a good alternative for traveltime misfit measurements with finite frequency information in WET.

In the next chapter, I aim to derive the unwrapped phase sensitivity kernels for transversely isotropic medium with a vertical axis of symmetry (VTI) and use them to analyze the seismic data sensitivity to anisotropic parameters.
Chapter 3

Analysis of the traveltime sensitivity kernels for an acoustic transversely isotropic medium with a vertical axis of symmetry

Summary

In anisotropic media, several parameters govern the propagation of the compres-
sional waves. To correctly invert surface recorded seismic data in anisotropic media, a multi-parameter inversion is required. However, the trade-off between parameters exists because several models can explain the same data set. To understand these trade-offs, diffraction/reflection and transmission type sensitivity kernels analyses are carried out. Such analyses can help choosing the appropriate parameterization for inversion. In tomography, the sensitivity kernels represent the effect of a parameter along the wave-path between a source and a receiver. At a given illumination angle, similarities between sensitivity kernels highlight the trade-off between the parameters. To discuss the parameterization choice in the context of finite-frequency tomography, I compute the sensitivity kernels of the instantaneous traveltimes derived from the seismic data traces. I consider the transmission case with no encounter of an inter-
face between a source and a receiver; with surface seismic data this corresponds to a diving wave path. I also consider the diffraction/reflection case when the wave path is formed by two parts one from the source to a sub-surface point and the other from

The work presented in this chapter is related to a publication:
the sub-surface point to the receiver. I illustrate the different parameter sensitivities for an acoustic transversely isotropic medium with a vertical axis of symmetry. The sensitivity kernels depend on the parameterization choice. By comparing different parameterizations, I explain why the parameterization with the normal move-out velocity, the anelliptic parameter $\eta$ and the $\delta$ parameter is attractive when we invert simultaneously diving and reflected events recorded in an active surface seismic experiment. Also, I show the advantages of using the horizontal velocity $v_h$, $\eta$ and $\epsilon$ parameters when the data allow hierarchical inversion by inverting first the diving waves, then including the reflections.

3.1 Introduction

In Chapter 2, I derived new sensitivity kernels using the unwrapped phase (or the instantaneous traveltime) for wave equation tomography (WET). In this chapter, I extend the derivation to a transversely isotropic medium with a vertical axis of symmetry (VTI). The kernels describe the sensitivity of seismic data to anisotropy parameters, therefore, they can be used as an analysis tool for inversion.

The recent improvements in seismic data acquisition, with large offsets and broad frequency content, and the resolution limitations associated with the high-frequency conventional traveltime tomography motivated the use of finite-frequency approaches in seismic inversion. Full waveform inversion (FWI) \cite{Lailly1983, Tarantola1984} corresponds to the most general inversion scheme. The data misfit function, however, suffers from many local minima due to cycle-skipping between modeled and observed data \cite{Virieux2009}. The nonlinearity of the inversion problem is less severe with transmissions than with reflections \cite{Gauthier1986, Mora1987, Snieder1989}. Inverting only the transmitted (diving) waves leads to waveform tomography \cite{Pratt1996}. Long-to-intermediate wavelength information can be retrieved with a frequency-continuation inversion scheme to mitigate the sensitivity
to the initial model. The nonlinearity can also be reduced by replacing the data misfit by the misfit between finite-frequency traveltimes extracted from the observed and modeled data. This gives the wave-equation tomography (WET) approach ([Cara and Lévêque 1987, Woodward 1992]). In the literature, several methods have been proposed to measure the traveltime misfit between the observed and synthetic data, notably the cross-correlation ([Marquering et al. 1999, Dahlen et al. 2000]). Here, we shall employ the instantaneous traveltime estimated by applying a differential operator on the wavefield ([Choi and Alkhalifah 2011, 2013]).

In exploration geophysics, most of the existing inversion schemes consider an acoustic Earth to reduce the computational cost and the number of unknowns to invert for. Acoustic anisotropic wave equations have been derived by zeroing the shear velocities ([Alkhalifah 2000, Duveneck et al. 2008]). These equations accurately model the kinematics of (quasi) P-waves in a transversely isotropic smooth medium with a vertical axis of symmetry (VTI medium) under the weak anisotropy approximation. At a discontinuity, the energy conversion to shear modes is neglected, which limits the validity of the approach to media with small shear velocity variations with respect to the wavelengths of the signal. The acoustic approximation is, therefore, apriori, more valid at low frequencies than for high frequencies. In this work, I’m interested in illustrating the sensitivity of the transmission data to the anisotropic parameters. These sensitivity kernels are the basis of tomography which aims in retrieving relatively smooth velocity fields, that is the long-to-intermediate wavelengths of velocity.

An outstanding issue with multi-parameter anisotropic inversion is the ambiguity or trade-off (also called cross-talk) between the model parameters ([Plessix and Cao 2011, Gholami et al. 2013b, a, Alkhalifah and Plessix 2014]). A trade-off occurs when we can explain, up to the noise level, a data set with a different set of parameters than the one used to generate it. Then, a trade-off occurs when the Hessian matrix
of the misfit function (at the minimum) has small eigenvalues compared to the noise level. Plessix and Cao (2011) used numerical eigenvalue and eigenvector decomposition of the Hessian of the least-squares misfit between the observed and modeled data to show that surface recorded diving waves are mainly sensitive to a combination of the normal-moveout (NMO) and horizontal velocities. Gholami et al. (2013b) and Alkhalifah and Plessix (2014) analyzed the trade-off between the parameters based on the radiation pattern of each anisotropy parameter. Since the radiation patterns describe the wave amplitude variation with respect to the scattering angle, an overlap of the patterns highlights a trade-off between the parameters for this particular angle range. The trade-off can be resolved when the angle coverage at the subsurface point is larger than the overlap between patterns. This analysis based on diffraction gives only information at a given sub-surface point and does not explain the possible trade-off between different sub-surface points, for instance along the wave path provided by the first Fresnel zone (Woodward, 1992).

The analysis based on radiation patterns can be complemented with a study of the sensitivity kernels. I qualitatively study the anisotropic sensitivity kernels for a single source-receiver pair (Sieminski et al., 2009; Zhou, 2009; Zhou and Greenhalgh, 2009, 2011; Djebbi and Alkhalifah, 2013a). The sensitivity kernels or Fréchet derivatives provide a map in the model space depicting the regions contributing to the data misfit for a specific source/receiver pair (Woodward, 1992; Snieder and Lomax, 1996, Marquering et al., 1998, 1999; Dahlen et al., 2000; Tromp et al., 2005). They can be evaluated by cross-correlating the source and receiver wavefield computed using the Born (Marquering et al., 1998, 1999; Dahlen et al., 2000) or Rytov approximation (Woodward, 1992; Snieder and Lomax, 1996). When using the instantaneous travel-times as input, I obtain sensitivity kernels similar to the ones obtained with a Rytov approximation, but in a more general form and with a maximum sensitivity along the ray-path (Djebbi and Alkhalifah, 2014). These sensitivity kernels are associated
Several studies focused on the analysis of anisotropic sensitivity kernels for perturbation of the stiffness coefficients in the elastic case (Zhou, 2009; Sieminski et al., 2009). Here, I focus the analysis on the sensitivity kernels for the acoustic VTI wave equation. This should allow discussing the wave equation parameterization for the inversion in this simple anisotropic case. I start by presenting the sensitivity kernels with a misfit function based on the instantaneous traveltime. I will consider several parameterizations of the VTI medium. Then, I perform numerical studies to analyze the transmission case in a homogeneous medium and a more complicated geological setting both with a surface acquisition. In exploration geophysics with surface seismic data, transmitted (diving) waves carry information of only the shallow part of the model and provide mainly wide angle information. Therefore, in order to study the deep part of the model, wave-equation based tomography can be adapted to reflection data. This can be achieved, at least in horizontally layered media, by reformulating the waveform inversion in vertical time instead of depth to avoid the depth/velocity ambiguities (Alkhalifah, 2003b; Plessix, 2012). Bakker and Gerritsen (2013) developed the angle domain reflection wave-path tomography to handle complex velocity models. Migration velocity analysis (MVA) can also be seen as an extension of wave equation tomography to reflection data (Prucha et al., 1999; Rickett and Sava, 2002; Xie and Yang, 2008; Fomel, 2011). In this chapter, to illustrate the kinematic reflection case, I consider a non-realistic experiment where I fix a sub-surface point and consider the kernel between the source point and the sub-surface point and the kernel between the sub-surface point and the receiver point. The kinematic diffraction/reflection kernel is the summation of these two parts. I compute the sensitivity kernels in the 3-D SEG/EAGE salt model. Ray theory (Červený, 2001) is used to evaluate both phase and amplitudes of the Green’s function as the wave-equation based approach is prohibitively expensive.
3.2 Theory

To reduce some of the non-linearities associated with full waveform inversion, I transform the observed data to the frequency domain and I use the instantaneous travel-times \(t(\omega) = \Im \left( \frac{\partial U(x_r, x_s, \omega)}{\partial \omega} \right)\),

\[
t(\omega) = \Im \left( \frac{\partial U(x_r, x_s, \omega)}{\partial \omega} \right),
\]

where \(\omega\) is the angular frequency, \(\Im\) denotes the imaginary part and \(U(x_r, x_s, \omega)\) is the pressure wavefield due to a shot at the source position \(x_s\) and recorded at the receiver position \(x_r\).

In a transversely isotropic medium with vertical axis of symmetry, the acoustic pressure field satisfies:

\[
-\frac{\omega^2}{v_n^2} U(x, x_s, \omega) - (1 + 2\eta) (\partial_{xx} U(x, x_s, \omega)) - \frac{1}{\sqrt{1+\delta^2}} \partial_{zz} \frac{U(x, x_s, \omega)}{\sqrt{1+\delta^2}}
+ \partial_{yy} U(x, x_s, \omega) = s(x_s, \omega)
\]

Following Plessix and Cao (2011), I parametrize the model using the normal move-out (NMO) velocity \(v_n\), \(\delta\) parameter and the anisotropic aneliptic parameter \(\eta = (\epsilon - \delta)/(1 + 2\delta)\), where \(\epsilon\) and \(\delta\) are the Thomsen’s VTI anisotropic parameters (Thomsen, 1986). The variable \(s(x_s, \omega)\) is the source term. In equation (3.2), I combined the two second-order equations (Duveneck et al., 2008) into one fourth-order wave equation.

In this work, I focus on five parameterizations of the VTI wave equation. These parameterizations are adequate for FWI when attempting to recover the long-to-intermediate wavelenghts of the Earth’s parameters (Alkhalifah and Plessix, 2014). I number the parameterizations as follows:
• **parameterization 1**: NMO velocity $v_n$, $\eta$ and $\delta$ parameters

• **parameterization 2**: Horizontal velocity $v_h$, $\eta$ and $\delta$ parameters

• **parameterization 3**: Horizontal velocity $v_h$, $\eta$ and $\epsilon$ parameters

• **parameterization 4**: NMO velocity $v_n$, horizontal velocity $v_h$ and $\delta$ parameter

• **parameterization 5**: NMO velocity $v_n$, vertical velocity $v_v$ and $\eta$ parameter

The sensitivity kernels are first derived for parameterization 1. Then, I deduce the kernels for the other parameterizations by simple chain rule relations.

The sensitivity kernels are obtained from the first-order Born approximation of $U = U_0 + \Delta U$, where $U_0$ is the wavefield in the background model and $\Delta U$ its first-order perturbation/scattered term. The wave equation for $\Delta U$ in parameterization 1 is given as:

$$
\begin{align*}
-\frac{\omega^2}{v_0^2} \Delta U(x, x_s, \omega) &- \partial_{xx} \Delta U(x, x_s, \omega) - \partial_{yy} \Delta U(x, x_s, \omega) \\
-\partial_{zz} \Delta U(x, x_s, \omega) &- \frac{\alpha_n}{v_0^2} \omega^2 U_0(x, x_s, \omega) \\
+2\Delta \eta (\partial_{xx} U_0(x, x_s, \omega) + \partial_{yy} U_0(x, x_s, \omega)) \\
+\frac{2}{\omega} \partial_{zz} v_0^2 \Delta \eta (\partial_{xx} U_0(x, x_s, \omega) + \partial_{yy} U_0(x, x_s, \omega)) \\
-\Delta \delta \partial_{zz} U_0(x, x_s, \omega) - \partial_{zz} \Delta \delta U_0(x, x_s, \omega).
\end{align*}
$$

(3.3)

Here, $\alpha_n v_0^2$, $\Delta \eta$ and $\Delta \delta$ are the perturbations of $v_n^2$, $\eta$ and $\delta$, respectively, and $v_0$ is the isotropic background velocity.

In this derivation, I perturb from an isotropic background/reference model for simplicity. This means that I assume weak anisotropy and neglect some anisotropic effects. For instance, I do not account for the differences between slowness and polarization vectors (Tsvankin, 2012). However, the conclusions of this work can be extended to a weak anisotropic background as the radiation patterns, and the sensitivity kernels are weakly dependent on the anisotropic parameters (Tsvankin, 2012). The pressure
$U_0$ then satisfies the isotropic wave equation:

$$-rac{\omega^2}{v_0^2} U_0(x, x_s, \omega) - \partial_{xx} U_0(x, x_s, \omega) - \partial_{yy} U_0(x, x_s, \omega) - \partial_{zz} U_0(x, x_s, \omega) = s(x_s, \omega).$$

(3.4)

When the source is considered as Dirac delta source term, $s(x_s, \omega) = \delta(x - x_s)$, the pressure field is given by the Green’s function $G$. The Green’s function for the background medium is given by $G_0$, and the Green’s function perturbation $\Delta G$ is given as:

$$\Delta G(x_r, x_s, \omega) = \int a_1 \cdot K_1 dx$$

$$= \int \left[ \begin{array}{c} \alpha_n v_0^2 \\ \Delta \eta \\ \Delta \delta \end{array} \right] \cdot \left[ \begin{array}{c} K_1^{(v_n^2)} \\ K_1^{(\eta)} \\ K_1^{(\delta)} \end{array} \right] dx,$$

(3.5)

where $a_1$ is the perturbation of the anisotropy parameters in parameterization 1 and $K_1$ the sensitivity kernel vector. The subscript 1 relates to parameterization 1. Here, I derived the sensitivity kernels for the pressure field.

Each element of $K_1$ represents the single frequency wavefield sensitivity kernel to one of the anisotropy parameters $\{v_n^2, \eta, \delta\}$. The single frequency wavefield sensitivity kernels are, thus, given as:

$$K_1^{(v_n^2)} = -\frac{\omega^2}{v_0^2} G_0(x, x_s, \omega) G_0(x, x_r, \omega);$$

$$K_1^{(\eta)} = -\frac{2v_0^2}{\omega^2} \left( \partial_{xx} G_0(x, x_s, \omega) + \partial_{yy} G_0(x, x_s, \omega) \right)$$

$$\left( \partial_{xx} G_0(x, x_r, \omega) + \partial_{yy} G_0(x, x_r, \omega) \right);$$

$$K_1^{(\delta)} = - (G_0(x, x_s, \omega) \partial_{zz} G_0(x, x_r, \omega)$$

$$+ G_0(x, x_r, \omega) \partial_{zz} G_0(x, x_s, \omega)).$$

(3.6)

Using the chain rule relations between the parameters [Alkhalifah and Plessix, 2014], the single frequency wavefield kernels for the second and third parameterizations,
denoted with the subscripts 2 and 3 respectively, are given as:

\[
K_2^{(v_h^2)} = -\frac{\omega^2}{v_0^2} G_0(x, x_s, \omega) G_0(x, x_r, \omega);
\]
\[
K_2^{(\eta)} = -2 \left( G_0(x, x_s, \omega) \partial_{zz} G_0(x, x_r, \omega) \right)
+ G_0(x, x_r, \omega) \partial_{zz} G_0(x, x_s, \omega))
- \frac{2v_0^2}{\omega^2} \partial_{zz} G_0(x, x_s, \omega) \partial_{zz} G_0(x, x_r, \omega);
\]
\[
K_2^{(\delta)} = - \left( G_0(x, x_s, \omega) \partial_{zz} G_0(x, x_r, \omega) \right)
+ G_0(x, x_r, \omega) \partial_{zz} G_0(x, x_s, \omega)) ,
\]

and

\[
K_3^{(v_n^2)} = -\frac{\omega^2}{v_0^2} G_0(x, x_s, \omega) G_0(x, x_r, \omega);
\]
\[
K_3^{(\eta)} = - \left( G_0(x, x_s, \omega) \partial_{zz} G_0(x, x_r, \omega) \right)
+ G_0(x, x_r, \omega) \partial_{zz} G_0(x, x_s, \omega))
- \frac{2v_0^2}{\omega^2} \partial_{zz} G_0(x, x_s, \omega) \partial_{zz} G_0(x, x_r, \omega);
\]
\[
K_3^{(\epsilon)} = - \left( G_0(x, x_s, \omega) \partial_{zz} G_0(x, x_r, \omega) \right)
+ G_0(x, x_r, \omega) \partial_{zz} G_0(x, x_s, \omega)) .
\]

The wavefield kernels for parameterizations 4 and 5 are shown in Appendix A.

Using equation (3.1), we obtain the sensitivity kernel related to the instantaneous traveltime by

\[
K_t^{(m)}(x_r, x, x_s, \omega) = \Im \left( \frac{\partial(K_t^{(m)}(x_r, x, x_s, \omega))}{\partial \omega} \right),
\]

with \( m \) corresponding to one of the anisotropy parameters: \( v_n^2, v_h^2, \eta, \delta \) and \( \epsilon \). Here, the subscripts corresponding to the used parameterization are ignored for simplicity.

The single frequency sensitivity kernel’s plot is composed of two regions: the first Fresnel zone corresponding to the transmission, and the higher order Fresnel zones which are ellipses with constant traveltimes similar to the reflection isochrones in prestack migration (Woodward 1992). In this chapter, I focus the analysis on the transmission part.
Band-limited kernels are obtained by applying a weighted summation over frequencies (Woodward, 1992; Marquering et al., 1999; Snieder and Lomax, 1996; Djebbi and Alkhalifah, 2014). The weighting function can be a Gaussian distribution (Liu et al., 2009) or the amplitude spectrum of the Ricker wavelet (Djebbi and Alkhalifah, 2014). The weighted summation results in a band-limited sensitivity kernel where only the first Fresnel zone corresponding to the transmitted wave is preserved.

I choose the amplitude spectrum of a Ricker wavelet as a weighting function. The band-limited sensitivity kernel is given by:

\[
K^{(m)}_{t}(x_r, x, x_s, \omega_0) = \sum_{\omega} A(\omega) K^{(m)}_{t}(x_r, x, x_s, \omega)
\]  

(3.10)

where \( A \) the amplitude spectrum of a Ricker wavelet with \( \omega_0 \) its peak angular frequency

\[
A(\omega) = 4\sqrt{\frac{\pi}{\omega_0^3}} e^{-\frac{\omega^2}{\omega_0^2}}.
\]  

(3.11)

The over-line in equation (3.10) is used to distinguish the band-limited kernel from a single frequency one.

For a given source-receiver pair, there exists a trade-off between parameters when the sensitivity kernels have similar shapes. The comparison of the absolute amplitude between the kernels is not relevant as the kernels are weighted by the perturbation strengths (see equation (3.5)).

3.3 Numerical results

I qualitatively analyze the sensitivity kernels through numerical examples. I consider a transmission case and a diffraction/reflection case, shown in Figure 3.1. In the diffraction/reflection case, the sensitivity kernel is composed of two branches: one from the source to the sub-surface point and the other from the sub-surface point to the receiver. I assume a scattering potential of 1 at the sub-surface point. This
is not realistic because it does not take into account the amplitude versus offset effects. Strictly speaking, this case is a diffraction case since we do not introduce the dip of the reflector. This case corresponds to the reflection only when the sum of the slowness vectors at the diffraction point is perpendicular to the reflector. In this chapter, I mainly focus on the transmission aspect. The reflection case allows illustrating what happens in reflection tomography when the waves mostly travel in the vertical direction. In practice, the scattering potential plays a significant role in FWI but not in wave path tomography as defined by Bakker and Gerritsen (2013).

![Schematic representation of the experimental geometry](image)

Figure 3.1: Schematic representation of the experimental geometry. The traveltime kernels are studied for two configurations: the transmission and the reflection geometry. The transmission and kinematic reflection kernels are given by the first Fresnel zone around the transmission ray-path and the reflection ray-path respectively.

I show the sensitivity kernels for a homogeneous medium, a $v(z)$ medium with increasing velocity with depth and the SEG/EAGE salt model. In the case of a homogeneous medium, I employ the analytical formula of the Green’s function $G_0$. I use the finite difference wave equation solution for the increasing velocity with depth model. For the inhomogeneous 3-D medium, computing the sensitivity kernels is very expensive using a finite-difference approximation of the wave equation. In this case, I approximate the Green’s functions with ray theory. I use the ray tracing algorithm
developed by Alkhalifah and Bednar (2000) for anisotropic Kirchhoff migration to compute the traveltimes and amplitudes of the Green’s functions in smooth VTI media. The first step is to shoot a fan of rays from the source or receiver locations, then the traveltime and amplitude are interpolated from the rays to the regular grid. Once I get the traveltime and amplitude maps on the regular grid for both source and receiver locations, the Green’s functions are computed at each grid point. Finally, I use equation (3.10) to calculate the band-limited sensitivity kernels.

3.3.1 Angular dependency

In an anisotropic medium, the influence of the parameters depends on the angle of propagation. To illustrate the angular dependence of the sensitivity kernels, I first consider a homogeneous isotropic medium with $v_0 = 2000$ m/s and different source and receiver orientations. I show the bandlimited kernels with a Ricker source wavelet centered around 5 Hz. The kernels are normalized to unit maximum amplitude.

In Figure 3.2, I plot the sensitivity kernels for parameterization 1 for three different source and receiver orientations corresponding to a 90 (horizontal), 45 and 0 (vertical) degree angle with respect to the vertical direction. The kernel for the NMO velocity is isotropic for this parameterization 1 (the amplitudes in the sensitivity kernel are distributed along the wave path independently of the direction of propagation). The sensitivity kernels of $\eta$ and $\delta$ parameters are angle dependent. The sensitivity of $\eta$ is maximal in the horizontal direction and weak in the vertical direction, whereas the sensitivity of $\delta$ has an opposite behavior, that is maximal in the vertical direction and minimal in the horizontal direction. Because the sensitivity of the NMO velocity is isotropic, we have a trade-off with $\eta$ in the horizontal direction and with $\delta$ in the vertical direction. At 45 degrees, the three parameters $v_n, \eta, \delta$ have a non zero sensitivity. This means that there would be a trade-off between all of the parameters when we have only a 45-degree angle information. However, in classical reflection
tomography from surface seismic data, we do not retrieve \( \delta \) (it is determined by well matching), and we invert traveltimes corresponding to a sufficiently large range of angles from near to far offsets. The differences in the sensitivity of \( v_n \) and \( \eta \) over angles allow resolving both \( v_n \) and \( \eta \).

![Figure 3.2: 5 Hz bandlimited travelttime sensitivity kernels for the first parameterization: NMO velocity, \( \eta \) and \( \delta \) parameters. (a-c) NMO velocity kernels, (d-f) \( \eta \) parameter kernels (g-i) \( \delta \) parameter kernels. The effect of source-receiver orientation is shown by changing the source receiver angle.](image)

I now consider parameterization 2 with the horizontal velocity \( v_h \), \( \eta \) and \( \delta \) pa-
rameters. The kernel of $v_h$ in parameterization 2, that is angle independent, is equal to the kernel of $v_n$ in parameterization 1 (see equations (3.6) and (3.7)). I show in Figure 3.3 the sensitivity kernels of $\eta$ and $\delta$. In parameterization 2, these parameters have no sensitivity at large angles corresponding to the horizontal direction. With this parameterization and with large offset surface data, $v_h$ should be well determined since there is no trade-off with the other two parameters. At small to intermediate angles, it would be hardly possible to resolve the trade-off between $\eta$ and $\delta$. When we choose to fix $\delta$, $\eta$ could be estimated with the small-to-intermediate angle data.

In Figure 3.3 I show the kernels of $\eta$ and $\epsilon$ for parameterization 3 ($v_h, \eta, \epsilon$). The sensitivity kernel of $v_h$ in this parameterization is again angle independent (see equation (3.8)). $\eta$ sensitivity is maximal at only intermediate angles whereas $\epsilon$ sensitivity
is maximal at small and intermediate angles. The $\epsilon$ kernels in parameterization 3 are in fact the same as the $\delta$ kernels in the two previous parameterizations (see equations (3.6), (3.7) and (3.8)). In this parameterization, $\epsilon$ can be determined by the well-to-seismic matching and not during the tomography inversion. Another observation is that the sensitivity magnitudes in the $\eta$ kernels are smaller than in the $\epsilon$ kernels. A re-scaling based on the Hessian of the least-squares misfit function (Virieux and Operto, 2009) may help during the inversion.

This numerical study illustrates that the shape of the parameter kernels is parameterization dependent. For instance, choosing parameterization 2 instead of parameterization 1 highly affects the $\eta$ kernels. We also notice that the kernels change when we use $\epsilon$ (parameterization 3) instead of $\delta$ (parameterization 2). In a nonlinear inversion, the choice of the parameterization is a critical step when we invert for more than one parameter. It is case dependent and also dependent on the workflow we choose. For instance, if we decide to apply a hierarchical data approach and invert first the long offsets of a surface data set, the parameterizations 2 and 3 may be preferable since there are no trade-offs between $v_h$ and the other parameters at large angles.

### 3.3.2 Diving waves: Offset and frequency dependency

I now focus on parameterization 1 with $v_n$, $\eta$ and $\delta$ parameters as it is a suitable parameterization for wave equation tomography (Plessix and Cao, 2011; Alkhalifah and Plessix, 2014). I compute the sensitivity kernels in a simple model: $v(z) = 2000 + 0.5z$ (m/s). The source and the receiver are at the surface. Therefore, the wave path between the source and the receiver corresponds to a diving wave. In Figure 3.5 I display the three band-limited sensitivity kernels normalized to unit maximum amplitude. It can be noticed that the NMO velocity sensitivity kernel is angle-independent since the amplitude is constant inside the first Fresnel zone. This is not the case with the $\eta$ and $\delta$ sensitivity kernels. For $\delta$ the amplitudes in the kernel
Figure 3.4: 5 Hz bandlimited sensitivity kernels for the third parameterization \((v_h, \eta, \epsilon)\). (a-c) show the \(\eta\) parameter kernels and (d-f) are the \(\epsilon\) parameter kernels.

are significant only where the angles with respect to the vertical direction are not too large. In this acquisition configuration, \(\delta\) can not be retrieved. As mentioned previously, \(\delta\) is generally kept fixed during tomography. Otherwise, the amplitudes of the \(\eta\) kernel is maximal when this angle is large; this is when the wave propagates in the horizontal direction.

When comparing the \(v_n\) and \(\eta\) kernels, I notice a large similarity in the wave path for the diving waves. I conclude that when inverting diving waves, the data is mainly sensitive to the horizontal velocity. If the acquisition is dense and contain large offsets, depending on the noise level, we may use the behavior differences at intermediate angles to discriminate between \(v_n\) and \(\eta\) (Stopin et al., 2014) although this remains very challenging without adding more information. Here, I take into account that the penetration depth of the diving waves depends on the offset between the source
and the receiver. Therefore, a shallow sub-surface point can be illuminated through different angles depending on the source and receiver locations and offsets. Increasing the frequency content helps to improve the resolution as the width of the first Fresnel zone is reduced.

![Figure 3.5: 5 Hz bandlimited diving wave sensitivity kernels, (a) NMO velocity kernel, (b) $\eta$ parameter kernel, (c) $\delta$ parameter kernel.](image)

### 3.3.3 3-D SEG/EAGE salt model

I finally compute the sensitivity kernels in a smooth version of the 3-D SEG/EAGE salt model, Figure 3.6. I plot the sensitivity kernels in the transmission case with
a source/receiver located at the surface and separated by 11600 m, Figure 3.7. In this case, the sensitivity kernels depend not only on anisotropy but also on the model inhomogeneity. The influence of the high-velocity salt body is observed in the non-symmetric shape of the kernels. Although we have a coupling between inhomogeneity and anisotropy, the conclusions from the homogeneous case and the \( v(z) \) model case hold. In the transmission mode, the \( \delta \) sensitivity kernel has weak amplitudes, whereas the \( v_n \) and \( \eta \) sensitivity kernels have a similar shape, with a slight difference at intermediate angles, illustrating the trade-off between these two parameters in this data configuration.

![Figure 3.6](image.png)

Figure 3.6: A smoothed version of the 3-D SEG/EAGE salt body model. The three slices include a depth slice (top) a y-slice (bottom left) and an x-slice (bottom right) corresponding to \( z = 2000 \) m, \( y = 6760 \) m, and \( x = 6760 \) m, respectively, as the white lines indicate.

Active surface seismic data with sufficient aperture contain both reflection and transmitted (diving) waves. Schematically, a reflection corresponds to a secondary source at the subsurface point. Therefore, the reflection allows having information from a wave traveling at small to intermediate angles. As discussed previously, I called sensitivity kernel in the diffraction/reflection case, the sum of the sensitivity
Figure 3.7: 5 Hz peak frequency sensitivity kernels for the SEG/EAGE salt model for the case of a diving wave. The source is located at \((x_s, y_s, z_s) = (1000, 6760, 0)\) m, the receiver is placed at \((x_r, y_r, z_r) = (12600, 6760, 0)\) m, corresponding to perturbations to (a) the NMO velocity, (b) the \(\eta\) parameter and (c) the \(\delta\) parameter.

In this experiment, I, however, ignore the angle dependency of the scattering potential (and I assume that the source/receiver pairs are connected through a reflection at the sub-surface point). In Figure 3.8 I display the three sensitivity kernels in the diffraction/reflection case. The source/receiver pair is at the surface with an offset of 11600 m as previously, and the sub-surface point is situated at mid-offset and 4000 m deep. This configuration corresponds to a relatively large reflection angle. As expected, the amplitudes in the \(\eta\) kernel diminish signifi-
Figure 3.8: 5 Hz peak frequency sensitivity kernels for the SEG/EAGE salt model for the case of a reflected wave (with fixed reflection point). The source is located at \((x_s, y_s, z_s) = (1000, 6760, 0)\) m, the receiver is placed at \((x_r, y_r, z_r) = (12600, 6760, 0)\) m. The image point is located at \((x, y, z) = (6800, 6760, 4000)\) m, corresponding to perturbations of (a) NMO velocity, (b) \(\eta\) parameter and (c) \(\delta\) parameter.

Significantly when the propagation is near vertical, whereas the amplitudes in the \(v_n\) kernel are relatively homogeneous. There are amplitude variations because the geometrical spreading depends on the Earth medium parameter variations. These relatively large amplitude differences along the sensitivity kernels mean that with enough angle coverage one might be able to invert both \(v_n\) and \(\eta\). This conclusion is expected because, with classic ray-based reflection tomography, we can invert both parameters when we have small and intermediate angle information. When the wave travels vertically,
the $\delta$ sensitivity kernel has a large amplitude. However, because of the depth/velocity ambiguity with only surface seismic data, it is not possible to retrieve the near vertical model information. As mentioned early, the $\delta$ parameter is generally kept fixed during tomography with the parameterization $v_n, \eta, \delta$ and determined through seismic-to-well ties.

### 3.4 Discussions

In the first parameterization, $v_n$ and $\eta$ suffer from trade-off for large opening angles (corresponding to diving waves). Also, $v_n$ and $\delta$ suffer from trade-off at small opening angles (the reflections). $\delta$ parameter is usually constrained from well ties and can be kept fixed during the inversion. So, inverting simultaneously diving and reflection waves should help to resolve the trade-off between $v_n$ and $\eta$ as we have a large angle coverage. This possibility is restricted to the part of the model that is illuminated by both reflection and diving wave paths, that is the shallow part of the model. If the data is dominated by diving waves, this inverted models using this parameterization will suffer from large trade-offs. As a result, leakage from one parameter to the other is expected. If the acoustic approximation is used for the inversion, the inaccuracy in estimating the waves amplitude behavior will affect the inversion. In this case, $\delta$ parameter should be considered during the inversion to collect this undesirable amplitude errors.

Parameterization 3 with $v_h, \eta, \epsilon$ is preferable when hierarchical inversion can be used. In fact, for large opening angles, the data is only sensitive to $v_h$ parameter. In this case, the $v_h$ long wavelength can be constructed without leakage from other parameters. $\eta$ parameter is sensitive to intermediate angles information. Then, including reflections in a second stage can help recovering this parameter. For acoustic approximation, $\epsilon$ can be included in the inversion, to play the role of garbage collector. This is the same role as density or $\delta$ for the first parameterization.
In the next chapter, I use the \( v_h, \eta, \epsilon \) parameterization for FWI. The implementation I consider is a frequency domain multi-stage implementation, where I start with low frequencies, then move towards higher frequencies during the inversion. Equation (3.12) shows the relation between the scattering angle (or the opening angle), frequency and model wavenumber update for a dip-oriented TI medium (Alkhalifah and Sava, 2010),

\[ k_m = \frac{\omega}{v_p \left( \frac{\theta_s}{2} \right)} \cos \frac{\theta_s}{2} n, \]

where \( k_m \) is the model’s wavenumber update, \( \theta_s \) the scattering angle, \( v_p \) the phase velocity and \( n \) the vector normal to the reflector. For general TI models, the relation is more complicated and can not be written explicitly, however, the conclusions driven by equation (3.12) can be generalized. According to this equation, updating the low model wavenumbers requires large opening angles (corresponding to diving waves or large offset data), or low frequencies. Then, the frequency continuation method I promote in the next chapter, fall into the hierarchical type of implementations. In this situation, the \( v_h, \eta, \epsilon \) parameterization will be the optimal parameterization with a minimal trade-off.

### 3.5 Conclusions

In order to build a model that represents well the real Earth, anisotropy should be considered in modeling and inversion, meaning that the model contains several parameters. This raises the question of the choice of the parameterization. In a transversely isotropic medium with a vertical axis of symmetry and active surface seismic data, I investigated the multi-parameter tomography by numerically analyzing the sensitivity kernels when the input data are the instantaneous traveltimes. The similarity between sensitivity kernels of the different parameters of a given parameterization helps understanding the trade-off between parameters at particular illumination an-
gles. In fact, the parameterization choice depends on the information content of the data. The sensitivity to the $\delta$ parameter is reduced because of the depth/velocity ambiguity with surface seismic data. With surface seismic data, we can simultaneously invert diving and reflection data. At a sub-surface point that is illuminated by both reflection and diving waves, the angle coverage is large and the choice of parameterizations between $v_n, \eta$ or $v_h, \eta$ is less crucial because for both should be able to retrieve the two parameters. The trade-off between those parameters exists only at certain angles, and therefore we can resolve the ambiguity due to the large angle coverage. When the sub-surface point is illuminated only by reflection wave paths with a limited angle coverage, corresponding to a sub-surface point in the deep part of the model, the trade-off between parameters may hamper the inversion.

The parameterization $v_h, \eta, \delta$ is not attractive because of the trade-off between the three parameters at small angles. The parameterization $v_n, \eta, \delta$ is preferable when inverting both diving waves and reflections simultaneously. However, the inversion results might be affected by trade-off artifacts caused by the same sensitivity of $v_n$ and $\eta$ for diving waves.

$v_h, \eta, \epsilon$ parameterization is preferable when we can access the data in a hierarchical approach. Starting with only large offsets (giving rise to large scattering angles), the data is only sensitive to $v_h$. The long wavelength component of $v_h$ is constrained without trade-off. Including the reflections, $\eta$ parameter can be recovered and $\epsilon$ plays the role of garbage collector for the acoustic approximation. Due to its sensitivity to small scattering, $\epsilon$ absorbs the inaccuracies in amplitude caused by using the acoustic approximation.
Chapter 4

Frequency domain multi-parameter full waveform inversion for acoustic VTI media

Summary
Multi-parameter full waveform inversion (FWI) for transversely isotropic (TI) media with a vertical axis of symmetry (VTI) suffers from the trade-off between the parameters. The trade-off results in the leakage of one parameter’s update into the other. It affects the accuracy and convergence of the inversion. The sensitivity analyses suggested a parameterization using the horizontal velocity $v_h$, $\epsilon$ and $\eta$ to reduce the trade-off for surface recorded seismic data.

In this chapter, I invert for this parameterization using the scattering integral (SI) method. The available Born sensitivity kernels, within this approach, can be used to calculate additional inversion information. I mainly compute the diagonal of the approximate Hessian, used as a conjugate gradient preconditioner, and the gradient step length. I consider modeling in the frequency domain. In fact, the large computational cost of the scattering integral method can be avoided with direct Helmholtz equation solvers. I apply the proposed method to the VTI Marmousi II model for different inversion strategies. The accuracy of the inversion results shows that this parameterization, as well as the scattering integral method, are suitable for multi-parameter VTI acoustic inversion.
4.1 Introduction

In Chapters 2 and 3, I introduced the sensitivity kernels using the unwrapped phase for isotropic and transversely isotropic medium with a vertical axis of symmetry (VTI). I analyzed the VTI model’s parameterizations using the sensitivity kernels and showed that $v_h$, $\eta$ and $\epsilon$ parameterization is suitable for a multiscale inversion, where the data can be accessed hierarchically. In this chapter, I use this parameterization for frequency domain full waveform inversion (FWI) using the scattering integral (SI) method.

Full waveform inversion aims to invert the full recorded data content to recover an accurate Earth model (Virieux and Operto, 2009). It is based on the minimization of data residuals, measured between the real data and the data generated numerically using an approximate Earth model. As the seismic data are highly nonlinear, FWI can suffer from cycle skipping. In this situation, the minimization problem may converge to a local minimum resulting in a wrong inverted model. FWI can be implemented in time domain (Tarantola, 1984; Mora, 1988; Bunks et al., 1995). Bunks et al. (1995), proposed a multiscale method to reduce the FWI nonlinearity. The inversion problem is decomposed into scales starting from a long scale to shorter ones. The low-frequency content of data, at the first stages, permits converging to the global minimum. Another approach to deal with the cycle-skipping problem is to window the data, so we focus the inversion on specific arrivals (Shipp and Singh, 2002). A more natural multiscale inversion scheme is realized in frequency domain (Pratt and Worthington, 1990; Zhou et al., 1995; Liao and McMechan, 1996) where a discrete number of increasing frequencies are selected. Sirgue and Pratt (2004) proposed a frequency selection algorithm that uses the redundancy criterion of seismic data. In this case, the frequency sampling is enlarged with the increasing frequency; therefore, few frequencies can be selected. Moreover, inverting a group of frequencies
increases the signal-to-noise ratio and results in better inversion resolution (Virieux and Operto, 2009; Brossier et al., 2009). In this case, the maximum frequency per group of frequencies controls the cycle-skipping limit.

Conventionally, the adjoint-state method (Tromp et al., 2005; Plessix, 2006) is used for FWI. The main advantage is the reasonable efficiency in the gradient computation. However, it does not provide the Fréchet derivatives. Therefore, it is not flexible when computing additional inversion attributes: the gradient step length and the Hessian matrix approximations, for instance. The scattering integral (SI) method is an alternative method, where the gradient is computed explicitly using the sensitivity kernels (Chen et al., 2007; Tao and Sen, 2013). The gradient is given by the product of the Fréchet derivatives matrix with the complex conjugate of the residual vector. The large computational cost and memory requirements to compute the gradient made the SI method less popular than the adjoint-state in seismic exploration. Chen et al. (2007) compared the two methods, and showed situations where the SI method outperforms the adjoint-state approach. Liu et al. (2015) introduced a method to avoid storing the Fréchet derivatives. The Green’s functions are stored in memory, and a vector matrix product is used to compute the gradient on the fly. Furthermore, when modeling is performed in the frequency domain, direct solvers can be used. The main cost of the modeling is in the decomposition of the impedance matrix into a lower-upper (LU) structure. Then, the back-substitution cost, and accordingly the SI integral cost, become trivial. In this chapter, I consider a frequency domain scattering integral implementation of FWI.

Most of the existing inversion schemes consider only an acoustic isotropic Earth model to reduce the computational cost. However, the recent improvements in data acquisition made anisotropic effects more prominent in the data. Inverting for anisotropy is a complicated task. The complication comes from the additional problem of trade-off between the parameters (Gholami et al., 2013b; Alkhalifah and
Therefore, understanding the sensitivity of surface recorded data to the anisotropy parameters is substantial to choose a parameterization with a minimum trade-off. For transversely isotropic (TI) media with a vertical axis of symmetry (VTI), Plessix and Cao (2011) studied the trade-off problem using an eigenvalue-eigenvector decomposition of the Hessian matrix. Gholami et al. (2013b) and Alkhalifah and Plessix (2014) analyzed the radiation patterns for several parameterizations to understand the sensitivity to anisotropy parameters at the scattering point level. Moreover, Alkhalifah (2016) studied the short and long wavelength influences using perturbations in the model parameters. He perturbed the model parameters and investigated the effects on the recorded data. In a more qualitative study, shown in the previous chapter, I used the sensitivity kernels for anisotropy parameter as a tool of analysis (Djebbi et al., 2017). Compared to the radiation patterns, the kernels show additional trade-off information along the wave-path.

The VTI wave equation parameterization using the normal move-out velocity $v_n$, $\eta$ and $\delta$ parameters is adapted for an inversion that includes simultaneously reflections and diving waves (Plessix and Cao, 2011; Alkhalifah and Plessix, 2014). For this parameterization, the $\delta$ parameter is considered as a secondary parameter, which is used to fit the inaccuracy in the amplitude caused by the acoustic assumption. However, the trade-off between $v_n$ and $\eta$ parameter for large scattering angles (diving waves) is significant and results in the leakage of $\eta$ parameter into $v_n$ during the inversion. Another widely used parameterization is based on the vertical velocity $v_v$, $\delta$, and $\epsilon$ parameters (Vigh et al., 2014). It suffers from the same type of trade-off for large scattering angles between $v_v$ and $\epsilon$ parameter. Alkhalifah (2016) suggested to use the horizontal velocity $v_h$, $\eta$ and $\epsilon$. This parameterization is considered optimal for conventional surface seismic experiments with the least trade-off. Guitton and Alkhalifah (2016) compared the two last mentioned parameterizations for elastic
FWI. The authors showed that using $v_h$, $\eta$ and $\epsilon$ parameterization the leakage from density and shear waves into $v_h$ is reduced. Wu and Alkhalifah (2016) used the same parameterization for reflection waveform inversion (RWI). They assumed $\epsilon = \eta$ to reduce the number of variables and used $\epsilon$ as a perturbation parameter. The inversion strategy performed well, and both the background and perturbation models are inverted accurately.

In this chapter, I present a multi-parameter acoustic VTI inversion using $v_h$, $\eta$ and $\epsilon$ parameterization. The conjugate gradient (CG) method is used for the optimization within the scattering integral framework. The time gradient step length as well as the diagonal approximate Hessian, which is used as CG preconditioner, are directly estimated using the sensitivity kernels. The first section of this chapter presents the frequency domain modeling approach. I show wavefields computed for a simple homogeneous model and the Marmousi II model. Then, I review the scattering integral approach for FWI. In the numerical examples section, I compare various multiscale inversion strategies. Then, I show a more realistic example, where the source signature is unknown, and the observed (real) data are modeled with an independent modeling code. I show that we can achieve reasonable inversion results using this parameterization.

### 4.2 Frequency domain VTI modeling

The key element in inversion is to efficiently solve the wave equation that represents the correct physics of wave propagation. I consider the VTI system of wave equations (Duveneck et al., 2008) in frequency domain,

\[
\begin{align*}
\frac{\omega^2}{v_h^2} p + (1 + 2\epsilon) \frac{\partial^2 p}{\partial x^2} + \sqrt{1 + 2\delta} \frac{\partial^2 q}{\partial z^2} &= f_x, \\
\frac{\omega^2}{v_h^2} q + \sqrt{1 + 2\delta} \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 q}{\partial z^2} &= f_z,
\end{align*}
\]

(4.1)
where \( p \) and \( q \) are the wavefield components, \( f_x \) and \( f_z \) are the source components and \( \omega \) is the angular frequency. The wave equations are parameterized using the vertical velocity \( v_v \) and Thomsen anisotropy parameters \( \epsilon \) and \( \delta \). I denote the total wavefield by \( U(x, x_s, \omega) \) and it can be calculated as \( U(x, x_s, \omega) = \frac{q+2p}{3} \).

Considering that parameterization \( v_h, \eta \) and \( \epsilon \) is used for the inversion, I re-parameterize the system of wave equations as,

\[
\frac{\omega^2(1+2\epsilon)}{v_h^2} p + (1 + 2\epsilon) \frac{\partial^2 p}{\partial x^2} + \sqrt{\frac{1+2\epsilon}{1+2\eta}} \frac{\partial^2 q}{\partial z^2} = f_x, \\
\frac{\omega^2(1+2\epsilon)}{v_h^2} q + \sqrt{\frac{1+2\epsilon}{1+2\eta}} \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 q}{\partial z^2} = f_z.
\]

(4.2)

In a compact form, this system can be written as,

\[
B(x, \omega)u(x, x_s, \omega) = f(x_s, \omega),
\]

(4.3)

where \( B(x, \omega) \) is the impedance matrix, \( u(x, x_s, \omega) \) the wavefield vector and \( f(x_s, \omega) \) is the source function vector.

For 2-D problems, solving equation (4.3) is generally performed using direct solvers. The sparse impedance matrix \( B(x, \omega) \) is decomposed into a LU decomposition. The main advantage of this approach is the non-dependence of \( B(x, \omega) \) on the source location. Thus, the same LU can be used for all sources. Considering the low computational cost of the forward and backward substitutions, compared to the LU decomposition, the wave equation can be solved efficiently for a large number of source locations. This approach can not be extended to 3-D problems as the decomposition will require large prohibitive memory, also the scalability issue when dealing with parallel computing implementations.

I consider 2nd-order and 4th-order finite-difference stencils to discretize the VTI wave equations system. I also use SuiteSparse package to perform the LU decomposition efficiently (Davis, 2004). I consider perfectly matched layers (PML) boundary
conditions to absorb the wavefields on the lateral and bottom boundaries. Additionally, the code permits modeling with an absorbing or a free surface boundary condition.

Figure 4.1 shows the modeling results for a homogeneous elliptic medium, with $v_h = 1.8\ \text{km/s}$ and $\epsilon = 0.25$. In Figures 4.1(a) and 4.1(b) I show the real and imaginary parts of the frequency domain wavefield for a frequency $f = 5\ \text{Hz}$. The source is placed at the center of the model and equals 1. The surface boundary is absorbing. In Figures 4.1(c) and 4.1(d) I show the wavefront for $t = 0.2\ \text{s}$ and $t = 0.5\ \text{s}$, respectively. The time domain wavefield solution is obtained by modeling multiple frequencies using a Ricker wavelet source function. Then, the frequency domain wavefields are inverse Fourier transformed to time domain. I plot, with a red curve, the first arrival traveltimes computed using an Eikonal solver. The wavefront is perfectly matching the first arrival traveltime.

In Figure 4.2 I show the modeling results for a homogeneous VTI medium, with $v_h = 1.8\ \text{km/s}$ and $\epsilon = 0.25$ and $\eta = 0.1$. Figures 4.2(a) and 4.2(b) show the real and imaginary parts of the frequency domain wavefield for a frequency $f = 5\ \text{Hz}$. As in the previous example, I show in Figures 4.2(c) and 4.2(d) the wavefront for $t = 0.2\ \text{s}$ and $t = 0.5\ \text{s}$. The traveltime Eikonal solution and the modeled wavefield are also matching. However, shear waves artifacts are contaminating the wavefields. This is an inherent problem with the pseudo-acoustic wave equations. This can be solved by placing the sources in a small elliptic medium box around the source (Alkhalifah, 2000). In most of FWI applications, sources are placed in water. Then, the solution will be free of these artifacts. As I consider only few frequencies, the inverse Fourier transform suffers from aliasing and this appears as artifacts near the corners of the model.

The modeling algorithm is also tested on a more complicated model. I consider the VTI Marmousi II shown in Figure 4.3. The model contains complex anisotropic
Figure 4.1: Frequency domain modeled wavefield for an elliptic anisotropic medium, with $v_h = 1.8$ km/s, $\epsilon = 0.25$ and $f = 5$ Hz: (a) the real part and (b) the imaginary part. (c) and (d) show the inverse Fourier transformed frequency domain wavefield to time domain at $t = 0.2$ s and $t = 0.5$ s, respectively. The wavefront is compared to the Eikonal traveltime solution given by the red curve. 

structures. This model will be used later for inversion. Figure 4.4 shows the frequency domain modeled wavefields. Figures 4.4(a) and 4.4(b) show the real and imaginary parts of the wave equation solution at 5 Hz. The wavefield is free of shear waves artifact as the source is placed in the water layer (isotropic). In Figures 4.4(c) and 4.4(d) we show the wavefronts, overlaid with the Eikonal traveltime solution, at $t = 0.8$ s
Figure 4.2: Frequency domain modeled wavefield for a VTI anisotropic medium, with $v_h = 1.8$ km/s, $\epsilon = 0.25$, $\eta = 0.1$ and $f = 5$ Hz: (a) the real part and (b) the imaginary part. (c) and (d) show the inverse Fourier transformed frequency domain wavefield to time domain at $t = 0.2$ s and $t = 0.5$ s, respectively. The wavefront is compared to the Eikonal traveltime solution given by the red curve. The wavefront and the first arrival traveltime are matching each other. We can also notice the shear wave artifacts, caused by the pseudo-acoustic equations. Aliasing artifacts are also present as a result of the large frequency sampling.

and $t = 1.2$ s, respectively. The modeling results are accurate and the VTI Helmholtz solver can, therefore, be used for inversion.
4.3 Full waveform inversion using the scattering integral approach

Full waveform inversion algorithm is based on the minimization of the residuals between the modeled data $U(m)$ using an approximate initial model and the observed
data $\mathbf{d}$. The data residual vector is denoted as $\Delta \mathbf{d}$ and is given by,

$$\Delta \mathbf{d} = \mathbf{U}(\mathbf{m}) - \mathbf{d}. \quad (4.4)$$

The modeled data and residual vectors depend on the model $\mathbf{m}$ used for modeling. I consider $n$ and $m$ as the data and model sizes, respectively. To minimize $\Delta \mathbf{d}$, the $l$-2 norm misfit function is used

$$E(\mathbf{m}) = \frac{1}{2} \Delta \mathbf{d}^T \Delta \mathbf{d}^*, \quad (4.5)$$

where the superscripts $^T$ and $^*$ represents the transpose and complex conjugate, respectively.

Minimizing the misfit function, the model update is given as,
\[ \Delta m = - \left( \frac{\partial^2 E(m)}{\partial m^2} \right)^{-1} \frac{\partial E(m)}{\partial m} = -H(m)^{-1}g(m), \] (4.6)

where \( g(m) \) and \( H(m) \) are the gradient vector and the Hessian matrix, respectively. Calculating the exact inverse of the Hessian matrix is usually impractical. Methods, like the limited-memory BFGS (l-BFGS) or truncated Gauss-Newton/Newton methods, are used to efficiently approximate the inverse of the Hessian.

Another approach is the conjugate gradient method (CG), where only the gradient information is used to update the model. Generally, a gradient step length estimation is required for this algorithm. Preconditioning the gradient with an approximate inverse Hessian can improve the convergence properties. The update is given as,

\[ \Delta m_{CG} = -\gamma P(m) \frac{\partial E(m)}{\partial m} = -\gamma p_2, \] (4.7)

where \( \gamma \) is the step size and \( P(m) \) is the preconditioning operator. I denote the preconditioned CG direction as \( p_2 \). Taking the derivative of the misfit function with respect to the model parameters, the gradient reads,

\[ g(m) = p_1 = \frac{\partial E(m)}{\partial m} = \Re \left( F^T \Delta d^* \right), \] (4.8)

where \( \Re \) denotes the real part. I denote the gradient update direction as \( p_1 \). \( F \) is the \((n \times m)\) matrix, representing the Fréchet derivatives. The Fréchet derivatives show the sensitivity of the data to the model parameters. The elements of this matrix are given as,

\[ k_{ij} = \frac{\partial U_i}{\partial m_j} \quad i = (1, \ldots, n); \quad j = (1, \ldots, m). \] (4.9)
4.3.1 The scattering integral method

I consider the scattering integral (SI) method to compute the gradient. The gradient is given by the product of the Fréchet derivatives matrix with the complex conjugate of the residual vector as shown with equation (4.8). For a single source and receiver, each row of the Fréchet derivatives matrix $F$ corresponds to the Born wavefield sensitivity kernel $K(x_r, x, x_s, \omega)$. For the isotropic case, the Born kernel for the velocity model is given as,

$$K^{(v)}(x_r, x, x_s, \omega) = \frac{2\omega^2}{v_0^2} U_0(x, x_s, \omega) G_0(x, x_r, \omega),$$

where $U_0(x, x_s, \omega)$ is the source wavefield and $G_0(x, x_r, \omega)$ is the receiver’s Green’s function in the background medium $v_0$.

The explicit computation of the Fréchet derivatives requires wavefields computed at the shot locations, as well as, the Green’s functions from the receivers locations, which makes it computationally expensive. In this situation, the features of direct Helmholtz solvers can be utilized. The LU decomposition can be used for all sources and receivers with a little additional cost. An additional difficulty associated with the SI approach is the large memory required to handle the Fréchet derivative (sensitivity) matrix. (Liu et al., 2015) proposed to save the Green’s functions into memory and use matrix decomposition to compute the gradient. This approach reduces the memory requirements significantly. Equation (4.11) shows the gradient computation procedure,

$$g(m) = \begin{pmatrix} k_{11} & \ldots & k_{i1} & \ldots & k_{n1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ k_{1j} & \ldots & k_{ij} & \ldots & k_{nj} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ k_{1m} & \ldots & k_{im} & \ldots & k_{nm} \end{pmatrix} \begin{pmatrix} \Delta d_1^* \\ \vdots \\ \Delta d_j^* \\ \vdots \\ \Delta d_n^* \end{pmatrix} = \sum_{i=0}^{n} \begin{pmatrix} k_{i1} \\ \vdots \\ k_{ij} \\ \vdots \\ k_{im} \end{pmatrix} \Delta d_i^*. \quad (4.11)$$
Furthermore, I use the inverse of the approximate Hessian diagonal as a CG preconditioning operator,

$$P(m) = \frac{1}{H_d(m)},$$  \hspace{1cm} (4.12)$$

where the approximate Hessian diagonal is given as,

$$H_d(m) = \text{diag}(F^{T^*} F) = \sum_{i=0}^{n} \begin{pmatrix} k_{i1}^* k_{i1} & \ldots & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots & \ddots \\ 0 & \ldots & k_{ij}^* k_{ij} & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots & \ddots \\ 0 & \ldots & 0 & \ldots & k_{im}^* k_{im} \end{pmatrix}.$$  \hspace{1cm} (4.13)$$

$F^{T^*}$ stands for the complex conjugate transpose of the Fréchet derivatives matrix.

Considering a second order approximation of the objective function $E_1(\gamma) = E(m + \gamma p)$, the time step size can be directly computed using the sensitivity kernels as $\text{[Hu et al., 2011]}$,

$$\gamma = \frac{p^{T^*} p_1}{p^{T^*} H_a p_1} = \frac{p^{T^*} p_1}{(Fp)^{T^*} (Fp)}.$$  \hspace{1cm} (4.14)$$

$p$ can be either the gradient direction $p_1$ or the preconditioned gradient direction $p_2$. $H_a = F^{T^*} F$ is the approximate Hessian. The decomposition of the denominator in equation (4.14) simplifies the gradient step length calculation.

### 4.3.2 The Born sensitivity kernels for VTI media

For a VTI medium parameterized using $v_h$, $\epsilon$ and $\eta$, the Fréchet derivatives matrix is given as:

$$F = [F_{v_h} \hspace{0.5cm} F_{\epsilon} \hspace{0.5cm} F_{\eta}].$$  \hspace{1cm} (4.15)$$

The total size of the Fréchet matrix is $3 \times (n \times m)$. Each row of the three sub-matrices, given in equation (4.15), is the sensitivity kernel for $v_h$, $\epsilon$ and $\eta$, respectively, for a
specific source-receiver pair.

The derivation of the sensitivity kernels has been performed in Chapter 3. The VTI system of equations (4.2) is combined into a single 4th order equation. Then, the three anisotropy parameters are perturbed to obtain the corresponding single frequency Born sensitivity kernels. Equation (4.16) shows the resulting kernels,

\[
\begin{align*}
K^{(vh)}(x_r, x_s, x, \omega) &= \frac{2\omega^2}{v_0^3} U_0(x, x_s, \omega) G_0(x, x_r, \omega); \\
K^{(\epsilon)}(x_r, x_s, x, \omega) &= -\left( U_0(x, x_s, \omega) \partial_{zz} G_0(x, x_r, \omega) + \partial_{zz} U_0(x, x_s, \omega) G_0(x, x_r, \omega) \right); \\
K^{(\eta)}(x_r, x_s, x, \omega) &= -\frac{v_0^2}{\omega^2} \left( \partial_{xx} U_0(x, x_s, \omega) \partial_{zz} G_0(x, x_r, \omega) + \partial_{zz} U_0(x, x_s, \omega) \partial_{xx} G_0(x, x_r, \omega) \right); \\
\end{align*}
\]

(4.16)

where \( U_0(x, x_s, \omega) \) and \( G_0(x, x_r, \omega) \) are the source wavefield and the receiver Green’s function using the background model, respectively.

Figure 4.5 shows the Born sensitivity kernels for an increasing velocity with depth model for a source located at 0.5 km and a receiver located at 7.5 km. The sensitivity kernel is composed of a central region which corresponds to the diving wave update and the isochrones corresponding to reflections information. It can be observed, for this conventional seismic experiment, that the horizontal velocity sensitivity is large for both diving waves and reflections. The main contribution for the \( \epsilon \) parameter sensitivity is due to reflections. \( \eta \) kernel looks similar to \( \epsilon \) in this set-up, although it is scattering at intermediate angles only.

\[ \text{4.3.3 Multi-parameter step estimation} \]

The step size is estimated for each parameter using a second order approximation of the objective function \( E_1(\gamma) = E(m + \gamma^T p_2) \). For a VTI medium, the vector \( m \) is composed of three vectors corresponding to: \( v_h, \epsilon \) and \( \eta \). The vector \( p_2 \) is the preconditioned CG update. Finally, \( \gamma \) is the step size vector given as: \( \gamma = \begin{bmatrix} \gamma_{v_h} & \gamma_{\epsilon} & \gamma_{\eta} \end{bmatrix}^T \). The step size vector elements are obtained by solving the \( 3 \times 3 \)
Figure 4.5: Single frequency Born sensitivity kernels for a VTI with increasing velocity with depth model: (a) $v_h$, (b) $\epsilon$ and (c) $\eta$.

For the sake of notation simplicity, $^T$ denotes the complex conjugate transpose.

The resulting 3 step lengths take into account a simultaneous perturbation of the three anisotropy parameters. Although this approach, in considering the trade-off, is simple compared to the full Hessian, the accurate inversion results, in the next section, show the applicability of the method. For the adjoint-state based FWI,
the elements of the matrix on the left-hand side require additional modeling steps (needed to approximate the matrix elements by finite-differences). For the scattering integral (SI) approach only vector-matrix multiplications are required. As a result, no additional modeling steps are needed, and the main computation occurs during the construction of the sensitivity matrix and its decomposition.

4.4 Numerical results

Based on the sensitivity analysis, I consider $v_h$, $\epsilon$ and $\eta$ parameterization for the inversion. I invert for the VTI Marmousi II model, shown in Figure 4.3. I investigate the scattering integral inversion with various strategies and initial models.

4.4.1 Inversion strategy

The initial model, shown in Figure 4.6, is obtained by directly smoothing the exact anisotropy parameters. I smooth the exact models using a triangular smoother with 1.5 km smoothing radius. The modeling is performed in the frequency domain using the acoustic VTI Helmholtz solver. In the following tests, the same modeling is used for both the observed and the modeled data. As my goal is to analyze the inversion strategy, using the same solver for both observed and modeled data is appropriate. I consider a multiscale inversion approach where 4 frequency bands are considered. Table 4.1 shows the inversion parameters for each frequency band. The minimum frequency for the inversion is 2 Hz, which is acceptable for recently acquired real data. I use the preconditioned CG gradient method to update the models. Also, I estimate the gradient step lengths using the sensitivity kernels as described in the theory section.

4.4.1.1 Simultaneous inversion of $v_h$, $\epsilon$ and $\eta$

In this test, $v_h$, $\eta$ and $\epsilon$ are updated simultaneously. I aim to invert $v_h$ with a minimum trade-off as it is the main parameter. Seismic data are sensitive to all
Figure 4.6: The VTI initial models: (a) the horizontal velocity $v_h$, (b) $\epsilon$ parameter, (c) $\eta$ parameter.

<table>
<thead>
<tr>
<th>Frequency Band</th>
<th>$f_{min}$ (Hz)</th>
<th>$f_{max}$ (Hz)</th>
<th>$\Delta f$ (Hz)</th>
<th>Number of iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.0</td>
<td>4.0</td>
<td>0.25</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>3.5</td>
<td>6.0</td>
<td>0.5</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>5.5</td>
<td>9.0</td>
<td>0.5</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>8.5</td>
<td>12.0</td>
<td>0.5</td>
<td>40</td>
</tr>
</tbody>
</table>

Table 4.1: VTI Marmousi II inversion parameters. 4 bands of multiples frequencies are used.

scattering angles for $v_h$ perturbation. Therefore, the resulting inverted $v_h$ model should include both the long and the short wavelengths. The final $v_h$ model, shown
in Figure 4.7(a), confirms the sensitivity analysis. Most of the features are well recovered. The reservoir area around \( x = 8 \) km and \( z = 3 \) km location is well resolved. \( \epsilon \) scatters data only at small scattering angles. Therefore, it plays the secondary role of fitting the amplitudes. \( \epsilon \) absorbs the inaccuracies in the inversion of short wavelengths due to the acoustic approximation. In the final model, as shown in Figure 4.7(b), only the model short wavelengths are updated. The recorded data are sensitive to \( \eta \) only for mid-range scattering angles (around 45\(^\circ\)) and the sensitivity magnitude is small compared to \( v_h \) and \( \epsilon \). The final \( \eta \) model, Figure 4.7(c) contains some features of the exact model. However, the trade-off with the other parameters affects the inversion in some areas of the model.

Figures 4.8 and 4.9 show the vertical profiles at 8 and 12 km. The horizontal velocity is well inverted and the main features are recovered. Furthermore, the inverted \( \epsilon \) parameter is reasonable and the short wavelength features are well recovered for parts of the model deeper than 1.5 km. From the \( \epsilon \) profiles, the variations in the shallow part of \( \epsilon \) model are not well recovered. In fact, in the shallow parts of the model, the scattering angles are larger than the deeper parts. The same observation holds for the \( \eta \) parameter. To improve these models, higher frequencies should be used.

Using the frequency continuation (multi-scale) approach, the difference in the sensitivity between \( v_h \), \( \epsilon \) and \( \eta \) is used to reduce the trade-off. For low frequencies, the data are mainly sensitive to \( v_h \). As a result, the \( v_h \) model is well recovered. When increasing the frequency, the data become more sensitive to the remaining parameters. Only short wavelengths for \( \epsilon \) and \( \eta \) can be inverted. I conclude that the acoustic frequency domain multi-parameter inversion using a parameterization based on \( v_h \), \( \epsilon \) and \( \eta \) can provide inverted models with a minimum trade-off.

Figure 4.10 shows the data misfit history for the 4 frequency bands. The misfit is the time domain data misfit. The misfit is largely reduced, which confirms the
convergence of the inversion.

Figure 4.7: FWI inverted models: (a) the horizontal velocity $v_h$, (b) $\epsilon$ parameter and (c) $\eta$ parameter. All three parameters are simultaneously inverted.

4.4.1.2 Hierarchical inversion of $v_h$, $\epsilon$ and $\eta$

In this example, I consider inverting hierarchically for $v_h$, $\epsilon$ and $\eta$. For low frequencies, given by the first frequency band, I only invert for $v_h$. Long wavelengths of the horizontal velocity are updated without trade-off from the other parameters. Starting from the second frequency band, I include $\epsilon$ parameter for the inversion. Finally, for frequency bands 3 and 4, I invert the three parameters simultaneously.
Figure 4.8: Vertical profiles for the simultaneous inversion at $x = 8$ km. (a) $v_h$, (b) $\epsilon$ parameter and (c) $\eta$ parameter.

The final inverted models are shown in Figure 4.11. $v_h$ model is well recovered. Also, short wavelengths for $\epsilon$ and $\eta$ are well inverted in the deep parts of the model. In the shallow parts of the model, $\epsilon$ and $\eta$ parameters are not recovered well due to sensitivity properties of these parameters. $\eta$ parameter is affected by the leakage...
from the other parameters. Therefore, many artifacts are present. For instance, the low velocity layers around 2.2 and 3 km deep are mainly caused by a leakage from $\epsilon$ parameter.

Figures 4.12 and 4.13 show the vertical profiles at 8 and 12 km. The quality of
the inverted horizontal velocity is comparable to the previous example. The main features are recovered. The inverted $\epsilon$ parameter is reasonable and the short wavelength features are well inverted. The long wavelengths features are missing in the $\eta$ parameter inversion. Also, it is contaminated with trade-off artifacts from the other anisotropy parameters.

4.4.1.3 Simultaneous inversion of $v_h$ and $\epsilon$, $\eta$ maintained as the initial $\eta_0$

As the sensitivity of the recorded data to $\eta$ parameter is weak, Alkhalifah (2016) proposed to invert only for $v_h$ and $\epsilon$. With this approach, the main objective is to recover a good horizontal velocity model. The $\epsilon$ parameter is useful for inversion of elastic data using the acoustic assumption. As $\epsilon$ radiation pattern is similar to the radiation patterns of density and shear waves, it can absorb the errors in amplitude caused by ignoring them during an acoustic inversion.

I tested this inversion strategy for the same initial model. $\eta$ parameter is fixed as the initial $\eta_0$ model. I perform a simultaneous inversion of both $v_h$ and $\epsilon$ parameter. Figure 4.14 shows the inverted models. The vertical profiles for $v_h$ and $\epsilon$ are shown in Figures 4.15 and 4.16. The horizontal velocity is accurately inverted. The short wavelengths of $\epsilon$ parameter are also recovered, especially in the deep parts of the model. In
Figure 4.11: FWI inverted models: (a) the horizontal velocity $v_h$, (b) $\epsilon$ parameter and (c) $\eta$ parameter. The three parameters are hierarchically inverted: for frequency band 1, only $v_h$ is inverted, then for frequency band 2, $v_h$ and $\epsilon$ are inverted. Finally all three parameters are inverted for frequency bands 3 and 4.

In these areas, the data are mainly governed by reflections with small scattering angles. Therefore, it is sensitive to $\epsilon$ parameter. A good background $\eta$ model is required to minimize the errors in $v_h$ inversion. Such models can be recovered using tomographic methods. Thus, ignoring $\eta$ short wavelengths, will not affect the inversion results.
Figure 4.12: Vertical profiles for the hierarchical inversion at $x = 8$ km. (a) $v_h$, (b) $\epsilon$ parameter and (c) $\eta$ parameter.

4.4.2 Realistic initial models

In this example, I consider a more realistic initial models. Thomsen parameter $\delta$, relating $\epsilon$ and $\eta$, is not inverted using surface recorded seismic data. It is obtained by matching well logs. As a result, when well information is not available, a realistic
initial model is to consider $\delta$ equals zero. Conventionally, the initial NMO velocity $v_{n0}$ and $\eta_0$ models are obtained using tomographic methods. Therefore, I construct the initial models for $v_h$, $\epsilon$ and $\eta$ parameters as follows: I smooth the exact $v_h$ and $\eta$ parameters using 1.5 km radius triangular smoothing. This step is equivalent to the
The inversion parameters are the same as in the previous examples. I use a frequency continuation approach with the same frequency parameters listed in Table 4.1. The inversion results are shown in Figure 4.18. The $v_h$ model contains the exact model structures. However, looking into the vertical profiles in Figures 4.19 and 4.20, the constructed model is shifted in depth. This shift is caused by ambiguity between $\delta$ parameter and the depth. In imaging, inaccurate $\delta$ models produce a vertical misplacement of the reflectors. For $\epsilon$ parameter, only the short wavelength is recovered. The difference between the initial and exact models is large. Due to the scattering properties of $\epsilon$, this large difference can not be retrieved. $\eta$ parameter is also affected by $\epsilon$ model errors. Leakage from $v_h$ and $\epsilon$ affects the inverted $\eta$ parameter.
4.4.3 An independent modeling code

I model the observed data with a time domain acoustic VTI solver. I consider a free-surface boundary condition on the surface. Also, the absorbing boundary conditions are implemented differently for the two solvers. To further complicate the inversion task and mimic a real data inversion situation, the source wavelet is unknown. I use a source estimation procedure during the inversion. The source wavelet is corrected after every iteration. $l$-2 misfit function is used for the inversion. In this example, I consider initial models obtained by smoothing the exact ones.

Figure 4.21 shows the inverted models. Figures 4.22 and 4.23 show the vertical profiles at $x = 8$ km and $x = 12$ km. Most of the features are well recovered,
Figure 4.16: Vertical profiles for the inversion with $\eta$ fixed as the initial $\eta_0$ at $x = 12$ km. (a) $v_h$ and (b) $\epsilon$ parameter.

especially in the shallow parts of the model. Details in the deep parts of the model are not well inverted. For $\epsilon$ and $\eta$ parameters, the short wavelengths are well resolved. Considering the complexity of the Marmousi II model, and the problems associated with multi-parameter inversion, the final inversion result is acceptable.

### 4.5 Conclusions

I applied a frequency domain FWI using the scattering integral approach to invert for a transverse isotropic with a vertical axis of symmetry model. The VTI wave equation is parameterized using the horizontal velocity $v_h$, $\epsilon$ and $\eta$ anisotropy parameters to reduce the trade-off between different parameters. I used a preconditioned conjugate gradient method to update the model. The step lengths are estimated through a
second order approximation of the objective function. It takes into account perturbations in the three anisotropy parameters and can reduce the inversion errors caused by the trade-off between the parameters.

I tested various inversion strategies: first, $v_h$, $\epsilon$ and $\eta$ are inverted simultaneously. In a second test, I considered a hierarchical inversion approach. Finally, I kept $\eta$ fixed as the initial model and inverted for $v_h$ and $\epsilon$. I showed that the inversion gives accurate inverted models for all three tests. Ignoring $\eta$ in the inversion reduces the cost
and the accuracy of $v_h$ inversion is not affected. However, good initial $\eta$ models are required. $\epsilon$ parameter absorbs the short wavelengths and amplitude errors associated with the acoustic approximation. Therefore, it should be included in the inversion, especially for elastic data.

I also considered a realistic initial model where $\epsilon_0 = \eta_0$. In this situation, the difference between the exact and initial $\epsilon$ models is large. The structure of the inverted models is accurate; however, the positions of the velocity layers are shifted.
In a final example, I tested the inversion on data generated with an independent modeling solver. The inversion results are also acceptable. To conclude, $v_h$, $\epsilon$ and $\eta$ parameterization is suitable for multi-parameter inversion within a frequency domain multi-scale framework. High-resolution horizontal velocity models can be obtained.
Figure 4.20: Vertical profiles for the inversion with realistic initial models at $x = 12$ km. (a) $v_h$, (b) $\epsilon$ parameter and (c) $\eta$ parameter.

However, good background $\eta$ is necessary to ensure accurate models inversion.
Figure 4.21: FWI inverted models for inversion without inverse crime: (a) the horizontal velocity $v_h$, (b) $\epsilon$ parameter and (c) $\eta$ parameter.
Figure 4.22: Vertical profiles for the inversion without inverse crime at $x = 8$ km. (a) $v_h$, (b) $\epsilon$ parameter and (c) $\eta$ parameter.
Figure 4.23: Vertical profiles for the inversion without inverse crime at $x = 12$ km. (a) $v_h$, (b) $\epsilon$ parameter and (c) $\eta$ parameter.
Chapter 5

Hybrid frequency domain full waveform inversion using Born sensitivity kernels

Summary
Full waveform inversion (FWI) using the scattering integral (SI) approach is an explicit formulation of the inversion optimization problem. The inversion equations are straightforward, and the dependence on the data residuals and model parameters is clear. However, the biggest limitation with this approach is the huge computational cost for exploration seismology applications, especially for 3-D problems. To deal with this issue, I propose a hybrid implementation of the frequency domain FWI using the SI method. Specifically, I use the sensitivity kernels computed from dynamic ray-tracing to build the gradient. With ray theory, the Green’s functions can be approximated using a coarser grid compared to wave equation modeling. Therefore, the memory requirements, as well as, the computational cost are largely reduced. Considering that in FWI long-to-intermediate wavelengths are updated during the first iterations, using a transmission experiment I obtain accurate inverted models. The inversion managed to develop the anomaly embedded in the homogeneous background medium. For more complex models, the hybrid inversion method helps improving the initial model with little cost compared to conventional SI inversion. The accuracy of the inversion results shows usefulness of the hybrid approach for 3-D realistic problems.
5.1 Introduction

In full waveform inversion (FWI) (Lailly, 1983; Tarantola, 1984), the data misfits between the observed and the modeled data are minimized to update the Earth’s model. However, this misfit suffers from many local minima due to cycle-skipping between modeled and observed data (Virieux and Operto, 2009). To deal with this issue, a common approach is to consider a frequency-continuation inversion scheme to mitigate sensitivity to the initial model and recover the long-to-intermediate wavelengths. Conventionally, FWI is implemented using the adjoint-state method (Tarantola, 1984). It requires an equal number of forward and adjoint modeling steps to compute the gradient. The convergence properties of the least-squares inversion problem can be improved using the Hessian information. Métivier et al. (2013) promoted the use of the second order adjoint-state equations to approximate the Hessian matrix in a truncated Newton/Gauss-Newton scheme.

Another FWI approach is the scattering integral (SI) method where the gradient is computed explicitly by multiplying the Fréchet derivatives matrix (or the sensitivity kernels matrix) with the complex conjugate of data residuals (Chen et al., 2007; Liu et al., 2015). The gradient of the misfit function requires more modeling steps compared to the adjoint-state method. A total number of modeling steps corresponding to the number of sources and the non-repeated receivers is required. Thus, for seismic exploration applications, the SI is less relevant compared to the adjoint-state method. However, when dealing with approximations of the Hessian matrix, the scattering integral approach cost is comparable to the second order adjoint-state. Chen et al. (2007) compared the two inversion approaches and described situations where the SI outperforms the adjoint-state method.

Liu et al. (2015) proposed to compute the scattering integral gradient by a matrix decomposition method in which only the wavefields are stored. This method permits
using the SI method with a reasonable memory storage. This approach has been used in Chapter 4 for 2-D anisotropic multi-parameter inversion. I used a direct Helmholtz solver to reduce the computational cost considerably. Actually, for direct wave equation solvers, the computational cost is mainly in lower upper (LU) decomposition of the impedance matrix. Also, the impedance matrix is source independent. Therefore, the same matrix decomposition can be used for all sources and receivers with a reasonable computational cost. For 3-D problems, however, LU decomposition is problematic. The computational cost, as well as the memory requirements, increase exponentially for large scale problems. Iterative or hybrid direct-iterative solvers are used, instead, to solve the frequency domain wave equation (Virieux and Operto 2009). Therefore, for 3-D applications solving the wave equation for large number of sources is expensive, which limits the applicability of the SI method.

To deal with the large computational cost for 3-D problems, ray theory (Červený 2001) has been widely used for imaging and inversion. Traveltimes and amplitudes, computed using dynamic ray tracing (Červený and Hron 1980), were used for 3-D true amplitude Kirchhoff migration (Bleistein 1987; Schleicher et al. 1993; Alkhalifah and Bednar 2000), as well as ray+Born migration/inversion (Thierry et al. 1998; Lambaré et al. 2003). The high-frequency approximation of the Green’s functions for a smooth background model is used to invert for the image or the velocity perturbations. Also, ray theory has been used for traveltime tomography (Dziewonski 1984; Van Der Hilst et al. 1997; Nolet 2012) and for wave equation tomography (Hung et al. 2000; Montelli et al. 2004; Tian et al. 2007), where the finite frequency sensitivity is considered. Liu et al. (2009) proposed a combination of wavefield modeling and ray tracing for Fresnel zone tomography. The sensitivity kernels are calculated using frequency domain modeling, and the traveltimes residuals are computed using ray tracing. With this approach, the resolution of the inverted models is improved compared to traveltime tomography.
In this chapter, I propose a hybrid inversion method, for which I combine the efficiency of ray theory (Cerveny 2001) and the flexibility of the scattering integral method. As the model is usually updated from low to high wavenumbers (relying on smooth models in the beginning), I use ray tracing as an effective alternative to wavefield modeling to compute Born kernels. This approach is proposed to deal with the enormous computational cost and memory requirements for SI inversion in the 3-D case. Generally, with ray theory, a coarse grid can be used to compute and save in memory the Green’s functions. Then, during the calculation of the sensitivity kernels, traveltimes and amplitudes can be interpolated to the inversion fine grid. An additional benefit is the ray direction information which can be used to compute the wavefield derivatives in the framework of anisotropic multi-parameter inversion (Alkhalifah 2016; Djebbi et al. 2017). The resulting models are limited by the smoothness requirements (for ray tracing) of the background models. Therefore, the inversion results, can be used as good initial models for a subsequent FWI, or directly utilized for migration.

First, I briefly describe the scattering integral approach for inversion as well as the proposed hybrid approach. Then, I analyze the computational cost of the proposed method compared to conventional SI. Finally, I invert for 2-D models to show the accuracy of the proposed method.

5.2 Theory

The sensitivity kernels or the Fréchet derivatives relate a perturbation in the recorded data to a perturbation in the Earth model. The frequency domain Born approximation of the perturbed wavefield, due to a velocity perturbation of the form: $v = v_0 + \Delta v$, is given as,

$$\Delta U(x_r, x, x_s, \omega) = \int \frac{2\omega^2}{v_0^3} G_0(x, x_r, \omega) U_0(x, x_s, \omega) \Delta v dx = \int K(v)(x_r, x, x_s, \omega) \Delta v dx, \quad (5.1)$$
where \( v_0 \) is the background velocity, \( \Delta v \) is the velocity perturbation. \( G_0(x, x, \omega) \) and \( U_0(x, x_s, \omega) \) are the receiver’s Green’s function and the source wavefield for the background velocity \( v_0 \), respectively. The velocity Born kernel is denoted as \( K^{(v)}(x_r, x, x_s, \omega) \).

In order to compute a single Born sensitivity kernel, two modeling steps are needed, one from the source and another from the receiver’s location. The kernel’s dimension equals the model dimension: \( m = n_x \times n_z \). Each row of the Fréchet derivatives matrix \( F \) corresponds to a source-receiver pair. The Fréchet derivatives matrix size is \( n \times m \) where \( n \) is the data size. Therefore, the storage of sensitivity kernels matrix in memory requires large resources, especially for 3-D problems.

### 5.2.1 The scattering integral method

The gradient of the least-squares misfit function between the modeled and observed data equation (4.5), \( E(m) \) is given as,

\[
g(m) = \frac{\partial E(m)}{\partial m} = \Re \left( F^T \Delta d^* \right), \tag{5.2}
\]

where \( \Delta d(x_s, x_r, \omega) \) is the data residual vector and \( F \) is the Fréchet derivatives (sensitivity kernels) matrix. \( .^T \) and \( .^* \) denote the transpose and complex conjugate respectively.

Liu et al. (2015) proposed to compute and save in memory the sources and receivers Green’s functions. Then, a multiplication of the Born kernels with the
residuals and accumulation is used to get the gradient, as shown in equation (5.3):

$$
g(m) = \begin{pmatrix} 
    k_{11} & \ldots & k_{i1} & \ldots & k_{1n_l} \\
    \vdots & \ddots & \vdots & \ddots & \vdots \\
    k_{1j} & \ldots & k_{ij} & \ldots & k_{nj} \\
    \vdots & \ddots & \vdots & \ddots & \vdots \\
    k_{1m} & \ldots & k_{im} & \ldots & k_{nm} 
\end{pmatrix} \begin{pmatrix} 
    \Delta d_{1}^* \\
    \vdots \\
    \Delta d_{j}^* \\
    \vdots \\
    \Delta d_{n}^* 
\end{pmatrix} = \sum_{i=0}^{n} \begin{pmatrix} 
    k_{i1} \\
    \vdots \\
    k_{ij} \\
    \vdots \\
    k_{im} 
\end{pmatrix} \Delta d_i^*.
$$

(5.3)

The steepest descent update direction $p_1$ is obtained after multiplication with the step length $\gamma$ estimated using a second order approximation of the objective function $E_1(\gamma) = E(v + \gamma p_1)$. Since the approximate Hessian matrix $H_a$ is given by $H_a = K^T K$, the truncated Gauss-Newton update direction $p_2$ can also be estimated using the matrix-vector multiplication approach. In fact, after computing the steepest descent update direction $p_1$, the system $H_a p_2 = p_1$ is solved using the conjugate gradient method to estimate $p_2$ [Liu et al., 2015]. Therefore, the kernels based inversion allows the computation of the truncated Gauss-Newton updates without additional modeling steps.

5.2.2 The hybrid inversion approach

Ray theory has been used in a wide range of applications from Kirchhoff migration to Ray+Born inversion. The advantage of using ray theory is its high efficiency compared to wavefield modeling [Cerveny, 2001]. For 3-D problems, the main limitation with the scattering integral approach for FWI is the high computational cost. Therefore, I propose to reduce the cost, and memory requirements to store the wavefields, by considering a hybrid implementation.

The gradient is composed of two components: the data residuals and the Born sensitivity kernels. In the proposed approach, I solve the Helmholtz equation for the source wavefield. Thus, the resulting residuals at the receivers locations contain the
exact residuals information. For receiver locations, I use ray theory to approximate the Green’s functions. I compute traveltime and amplitude maps; then these maps are saved into memory and used in gradient computation. Ray tracing does not require fine spatial sampling. Therefore, traveltime and amplitude maps can be computed for a coarse grid to further reduce the cost and memory requirements.

The proposed method can be used for the first FWI iterations. Indeed, at that stage only long-to-intermediate wavelengths of the model are updated, and the update is smooth which justifies the use of ray theory. The resulting smooth model can be used as a good initial model for FWI, or for imaging. For anisotropy, as shown by Alkhalifah (2016) and Djebbi et al. (2017), the kernels are angle dependent. Hence, dynamic ray-tracing can handle this dependency explicitly using the calculated ray-parameter information.

5.3 Computational cost analysis

In this section, I analyze the computational cost for the proposed hybrid inversion approach. The total number of modeling steps for the scattering integral based inversion is \( n_s + n_r \), where \( n_s \) and \( n_r \) denote the number of sources and receivers, respectively. The cost of calculating the gradient by matrix vector multiplication is the same for the conventional SI and the proposed approach. Thus, it is ignored in this analysis. I assume that the computational cost for one modeling in the frequency domain is given by \( t_f = 1 \). In general, time domain modeling costs more than frequency domain modeling. Therefore, I also assume that the computational cost for the time domain modeling \( t_t \) is given by \( t_t = 2 \times t_f \). The cost of one ray tracing is assumed to be much smaller than frequency domain modeling. In fact, a coarser grid can be used, which reduces enormously the computational cost. I consider the ray tracing cost equals \( t_r = 0.1 \times t_f \). This cost can be even smaller for 3-D problems. Multiple frequencies are usually used for frequency domain inversion. The number of frequencies is denoted
as $n_f$.

For 2-D problems, I fix the number of shots to $n_s = 100$ and consider variable number of frequencies $n_f$ and receivers $n_r$. Figure 5.1 shows the gradient computational cost for various modeling approaches. Figure 5.1(a) shows the computational cost for frequency domain modeling. The multi-source feature of the Helmholtz solver is not used here, and LU decomposition is repeated for every source and receiver. This situation is also equivalent to solving the Helmholtz equation using iterative methods. The total cost is given by $C_{2D}^f = n_f \times (n_s + n_r) \times t_f$. In Figure 5.1(b) I show the computational cost when a single LU decomposition is used to solve for all the sources and receivers. I assume here that solving for a single source costs 1% of the LU decomposition. The cost is $C_{2D}^f = n_f \times [t_f + 0.01 \times (n_s - 1 + n_r) \times t_f]$.

For time domain modeling, Figure 5.1(c), the wavefield is Fourier transformed to the frequency domain. The Fourier transform cost is negligible compared to the modeling cost. The scattering integral using time domain is independent of the number of frequencies, and the cost is given by $T_{2D}^t = (n_s + n_r) \times t_t$. The SI inversion using time domain modeling is interesting, only, when the number of sources and receivers is small. Finally, using ray tracing, shown in Figure 5.1(c) the modeling cost is given as $C_{2D}^r = n_s \times t_f + n_r \times t_r$. I conclude that for 2-D problems, the LU based approach for modeling outperforms time domain as well as the proposed hybrid approach. However, when the number of frequencies and receivers are large, the hybrid approach becomes interesting.

For 3-D problems, the model size and the number of sources and receivers become larger. The computational cost for full wavefield modeling methods is expensive. Here, I consider that the number of shots is $n_s = 1000$. Direct methods are not applicable for three dimensional problems because of the large memory requirements. Therefore, iterative methods and hybrid direct-iterative methods can be used to solve the Helmholtz equation. I consider the same assumptions as the 2-D case. Figure 5.2
Figure 5.1: Modeling computational cost for a 2-D problem. (a) frequency domain modeling with repeated LU decomposition, (b) frequency domain modeling with the same LU used, (c) time domain modeling (d) Hybrid frequency domain-ray tracing method.

shows the computational cost for frequency domain, time domain and the hybrid approach. The hybrid approach is, by far, the best approach for modeling. It is independent of the number of frequencies. Also, the proposed method is more efficient for large number of receivers. For instance, using 4000 receivers, and 6 frequencies, the computational cost is $C^r_{3D} = 1400$ for the Hybrid approach, compared to $C^f_{3D} = 30000$ for frequency domain and $C^t_{3D} = 10000$ for time domain modeling. These improvements in the computational cost are obtained assuming that the ray tracing cost is $t_r = 0.1 \times t_f$. However, in real applications, ray tracing cost can be much smaller, which results in a better efficiency.
Figure 5.2: Modeling computational cost for a 3-D problem. (a) frequency domain modeling, (b) time domain modeling (c) Hybrid frequency domain-ray tracing method.

The memory requirements to save the Green’s functions in memory is $n_s \times n_r \times m$, where $m$ is the model size. For three dimensional problems, I assume that the model size is $\mathcal{O}(N^3)$ where $N$ is the dimension in one direction. Also, assuming $n_s \sim n_r \sim N^2$, the total memory requirement is $\mathcal{O}(N^5)$ (Liu et al., 2015). With the proposed hybrid approach, a coarse grid can be used to save the ray tracing traveltimes and the amplitudes. One order of magnitude can be reduced if ray tracing attributes are saved every 10 space samples. Then, during the kernels calculation, the traveltimes and amplitudes are interpolated to the original grid (Thierry et al., 1998). The final memory requirements will be $\mathcal{O}(N^4)$. Also, for real applications, the number
of sources and receivers is smaller than $O(N^2)$. Additional cost reduction can also be achieved with saving in memory only the area covered by each seismic shot (Liu et al., 2015). Therefore, for 3-D problems, the memory requirements for the proposed approach can be appropriately handled even with a large memory workstation.

5.4 Numerical results

In this section, I present numerical results using the proposed hybrid approach. The method is applied for 2-D problems to show the accuracy of the inversion results.

First, I consider a circular anomaly embedded in a homogeneous background medium. The background model velocity is $v_0 = 3500$ m/s. The anomaly perturbation is +200 m/s with respect to the background velocity. The anomaly is located at the center of the model ($x_a = 2500$ m, $z_a = 2500$ m), as shown in Figure 5.3(a). Here, a transmission experiment is considered. 21 sources are located on the left and top sides at a distance of 500 m from the edges. The receivers are located on the right and bottom sides with the same distance (500 m) from the edges. I invert 8 frequencies from 2 to 16 Hz with 5 iterations per frequency. In this example, I use the truncated Gauss-Newton updates. The inner conjugate gradient loop (to solve for the truncated Gauss-Newton update) is fixed to 10 iterations.

Figure 5.3(b) shows the conventional scattering integral (SI) result using the Green’s functions computed with a Helmholtz equation solver. The anomaly is fully recovered as shown in the slice plot Figure 5.4(a). Figure 5.3(c) shows the inversion result using the hybrid implementation. The same inversion parameters as the conventional case are used. Ray tracing is used to compute the traveltimes, and the Green’s functions amplitudes are kept fixed. Therefore, I use only the half of the needed memory to store the wavefields in the conventional SI inversion. The computational cost of a high accuracy traveltime maps using ray tracing is negligible. Additional cost reductions can be achieved as ray-tracing does not require fine grid
sampling. Traveltimes can be computed for coarse grid then interpolated to the finer grid during the update calculation. The inversion results given by Figures 5.3(c) and 5.4(b) show high accuracy. It can be concluded that the hybrid method performs well for transmission geometry with limited memory requirements (only traveltime maps are saved in memory).

![Anomaly inversion using Born sensitivity kernels and truncated Gauss-Newton method.](image)

Figure 5.3: Anomaly inversion using Born sensitivity kernels and truncated Gauss-Newton method. (a) is the exact model, (b) the inverted velocity using Helmholtz based scattering integral (SI) approach, (c) the inverted velocity using the hybrid implementation.

The second example is the Marmousi model, shown in Figure 5.5(a). The model is modified by including a small water layer. This layer is fixed during the inversion
Figure 5.4: Vertical profiles for the anomaly inversion using Born sensitivity kernels and truncated Gauss-Newton method at $x = 2500$ m. (a) Helmholtz based SI and (b) Hybrid approach. The red line indicates the exact model and the blue line represents the inverted model.

to avoid ray tracing artifacts near the sources. I use 57 shots with 160 m sampling. The data are recorded on 225 receivers with an interval of 40 m. The initial model is an increasing velocity model with depth as shown in Figure 5.5(b). Here, I update the model using the conjugate gradient method preconditioned with the diagonal of the approximate Hessian. I also calculate the step length using the Born sensitivity kernels. I consider 10 frequencies from 1 Hz to 10 Hz with 10 iterations per frequency for this example. The frequency sampling is variable and is enlarged with higher inverted frequencies (Sirgue and Pratt 2004).
I start by inverting for frequencies from 1 to 3.5 Hz. In this stage, the proposed hybrid approach is used to reduce the computational cost. The inversion results are shown in Figures 5.6(a) and 5.6(b) for the conventional scattering integral and the hybrid approach respectively. The model recovered using the hybrid approach is smoother compared to the conventional Helmholtz based scattering integral method. However, from the vertical profiles shown in Figure 5.8, the long wavelengths features are well inverted.

The models inverted using 1-3.5 Hz frequencies are used as initial models for conventional SI inversion. Figure 5.7 shows the inversion results. The inversion results are comparable. I show the vertical profiles at $x = 3000$ m and $x = 7500$ m for both inversions in Figures 5.8. The inversion accurately recovered most of the features. I conclude that using the hybrid approach the long wavelengths are well inverted. The method is a good tool to provide accurate initial models for conventional FWI with a reduced computational cost.
In this chapter, I proposed a hybrid implementation for full waveform inversion within the scattering integral framework. The method is based on dynamic ray tracing for the computation of the sensitivity kernels. The residuals are computed using wavefield modeling in the frequency domain. The method has the advantage of lower memory requirements as well as the lower computational cost. Through numerical examples, I showed that the hybrid inversion approach performs well in transmission inversion set-up. However, for the case of complicated velocity models, it can only be used for the early iterations, as ray tracing can not handle non-smooth models. The smooth models resulting from the hybrid approach can be used for migration or as good initial models for FWI.

The proposed method is more advantageous for 3-D anisotropic inversion where
Figure 5.7: FWI inversion results for the Marmousi model using the models inverted for 1-3.5 Hz, shown Figure 5.6 as initial models. (a) Helmholtz equation based inversion and (b) Hybrid inversion.

wavefield modeling requires large computational resources. I showed through cost analysis the computational cost reduction using the hybrid approach for 3-D inversion. Also, when inverting for anisotropy parameters, the ray parameter (direction) can be used instead of finite differences to approximate the wavefield derivatives required for anisotropy updates.
Figure 5.8: Vertical profiles for the Marmousi model inversion. (a,b) Helmholtz equation based inversion vertical profiles at $x = 3000$ m and $x = 7500$ m, respectively. (c,d) Hybrid inversion vertical profiles at $x = 3000$ m and $x = 7500$ m, respectively. The red line stands for the exact model, the black is the initial model, the purple is the inverted model for 1-3.5 Hz, and the blue line is the final inverted models using FWI.
Chapter 6

Application to real data

In this chapter, I apply the proposed scattering integral inversion to a real data set. I consider inverting for a transversely isotropic medium with a vertical axis of symmetry (VTI) using $v_h$, $\eta$, and $\epsilon$ model parameterization.

6.1 Data description

I consider a 2-D marine data set acquired by CGG from the North-Western Australia Continental shelf (Figure 6.1). The data are a BroadSeis data acquired with variable depth streamers. This type of acquisition improves the signal to noise ratio for low frequencies (Soubaras and Dowle, 2010). The data set consists of 1824 shots with 648 receivers per shot. I display one shot gather and its frequency content in Figures 6.2(a) and 6.2(b) respectively. The data minimum frequency is 2.5 Hz and the maximum recording offset is 8.3 km. The number of receivers is reduced to 324 receivers per shot, covering the full offset, to reduce the computational cost.

6.2 Multi-parameter inversion

The noise prior to the first arrivals is muted and the data are filtered to low frequencies appropriate for full waveform inversion (Figure 6.3). I consider only 300 shots covering a part of the model to reduce the computational cost. The lateral extent of the considered model is 12.5 km.

The initial model is estimated using reflection waveform inversion (RWI) (Wu
Figure 6.1: Acquisition location (Courtesy of CGG).

Figure 6.2: Shot gather example: (a) Shot gather for a source located at \((x, z) = (3.75, 0.005)\) km, (b) data frequency spectrum.

The initial velocity model is smooth and represents well the kinematics of wave propagation. Figure 6.4 shows the initial model. The water depth is estimated using the initial model by picking the water bottom. The water velocity is kept fixed during the inversion. The source wavelet, as shown in Figure 6.5, is estimated using the direct arrivals in the water layer. As the water velocity is known,
the source signature can be well recovered (Kim et al., 2011; Kalita and Alkhalifah, 2017).

Figure 6.3: Bandpass filtered shot gather between 3 and 12 Hz: (a) Shot gather for a source located at \((x, z) = (3.75, 0.005)\) km, (b) data frequency spectrum.

Figure 6.4: Initial velocity model.

The acoustic approximation, used to model the synthetic data, suffers from inaccurate estimation of the data amplitudes. Therefore, the conventional least squares misfit function fails to properly invert for the Earth’s model. To deal with this problem, I use the logarithmic misfit function (Shin and Min, 2006) to invert for the phase only. The data residuals for the logarithmic misfit function are given as,
\[ \Delta d = \Delta \Phi = \Im \left( \log \left( \frac{U(m)}{d} \right) \right), \] (6.1)

where \( \Delta \Phi \) is the phase difference, \( \log \) is the complex logarithm, \( U(m) \) is the modeled data and \( d \) is the observed data. \( \Im \) denotes the imaginary part.

The logarithmic misfit function reads,

\[ E(m) = \frac{1}{2} \Im \left( \log \left( \frac{U(m)}{d} \right) \right)^2. \] (6.2)

The gradient is given as,

\[ g(m) = p_1 = \frac{\partial E(m)}{\partial m} = \Im \left( F^T \Delta d_b \right), \] (6.3)

where \( \Delta d_b = \Im \left( \log \frac{U(m)}{d} \right) \) is the complex back-propagated residual.

I consider a preconditioned conjugate gradient method for the inversion. The preconditioning operator is the diagonal of the approximate Hessian matrix. The frequency domain scattering integral method described in Chapter 4 is used. I invert for two frequency bands, given in table 6.1 to update the model in a multiscale approach [Bunks et al., 1995].

The inverted anisotropic models are shown in Figure 6.6. Two high-velocity zones
Table 6.1: CGG real data inversion parameters. 2 bands of multiples frequencies are used.

<table>
<thead>
<tr>
<th>Frequency Band</th>
<th>$f_{\text{min}}$ (Hz)</th>
<th>$f_{\text{max}}$ (Hz)</th>
<th>$\Delta f$ (Hz)</th>
<th>Number of iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.0</td>
<td>5.5</td>
<td>0.125</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>5.0</td>
<td>7.5</td>
<td>0.25</td>
<td>40</td>
</tr>
</tbody>
</table>

are recovered. The high-velocity zone observed at $z = 2$ km in the inverted model is validated by the well-log shown in Figure 6.7. Vertical profiles of the initial model and the final inverted model at $x = 10.5$ km are also shown. The inverted horizontal velocity model contains most of the features observed in the well-log but with a low resolution. Data with higher frequencies need to be inverted to recover the high-resolution features. The inverted $\epsilon$ and $\eta$ parameters contain several high amplitude zones. As the background model is not available, only high wavenumber features are recovered.

Next, I show the phase misfit for the two frequency bands in Figure 6.8. The data misfit is largely reduced especially for the first frequency band with low frequencies. To confirm the accuracy of the inverted models, I generate the time domain modeled data using the initial and final anisotropic models. The data modeled using the initial and inverted models are compared to the filtered real data. I show two shot gathers in Figures 6.9 and 6.10. The data generated with the inverted models contain additional reflections which are matching the real data. The data analysis confirms the accuracy of the inverted models.

Furthermore, I apply reverse time migration (RTM) using the initial and inverted models. I show the images, and the resulting angle gathers in Figures 6.11 and 6.12 respectively. As the initial model already contains the required low wavenumbers for migration, slight changes are observed in the RTM image with the final model. In the angle gathers, multiple locations are improved, and the gathers are flattened. These locations are indicated by the green arrows.
Figure 6.6: Inverted models: (a) $v_h$ model (b) $\epsilon$ parameter model and (c) $\eta$ parameter model.
Figure 6.7: Vertical profiles at $x = 10.5$ km compared with well-log.

Figure 6.8: Convergence history. The misfit is given by the phase difference.
Figure 6.9: Data comparison for shot gather 50: (a) initial model data (b) final model data.
Figure 6.10: Data comparison for shot gather 100: (a) initial model data (b) final model data.
Figure 6.11: Image using reverse time migration (RTM) for (a) the initial model (b) the final model.
Figure 6.12: Angle gathers for (a) the initial model (b) the final model. The angle gathers are shown for multiple lateral positions. The angles is between 0 and 45°. The green arrows show locations where there is improvement. The yellow arrows show some locations where the angles gathers become less flat.
Chapter 7

Concluding Remarks and Future Work

7.1 Summary

In this dissertation, I studied full waveform inversion (FWI) for transversely isotropic media with a vertical axis of symmetry (VTI) within the scattering integral framework. First, I analyzed the behavior of the sensitivity kernels. I developed new wave equation tomography kernels using the unwrapped phase. The kernels, with unwrapped phase, avoid the nonlinearities associated with the conventional cross-correlation ones. I also used sensitivity kernels to analyze sensitivity of seismic data to anisotropy parameters. The spatial distribution of the kernels highlights the trade-off between the parameters. Then, an optimal parameterization, based on the analysis results, is used for inversion. The frequency domain multi-scale inversion creates a framework to access the data hierarchically, which minimizes the trade-off effects. Furthermore, I proposed a hybrid scattering inversion approach to deal with the enormous computational cost for 3-D problems. The efficiency of ray tracing and the flexibility of the scattering integral method are combined to reduce the computational cost without affecting the inversion accuracy. Finally, I applied the frequency domain multi-parameter inversion to a real data set. I verified the accuracy of the inversion results with data comparison, as well as imaging.
7.1.1 Sensitivity kernels using the unwrapped phase

The sensitivity kernels or Fréchet derivatives are the essential elements in computing the gradient. Motivated by the analysis of these kernels, I studied in Chapter 2 the sensitivity kernels for wave equation tomography. Conventionally, they are calculated using cross-correlation or Rytov approximation. They usually suffer from the zero sensitivity along the central ray-path. I proposed an alternative approach using the unwrapped phase (or the instantaneous traveltime). Unwrapping the phase using a differential operator permits resolving the non-linearity associated with the standard kernels. I showed that the instantaneous traveltime is more sensitive to small anomalies compared to the cross-correlation. I also demonstrated that the instantaneous traveltime is the unwrapped phase version of the Rytov traveltime approximation. Observations on a simple 3-D velocity model and the more complex Marmousi model shows smooth kernels. With a simple test of traveltime lag measurement for a 3-D model with an embedded small spherical anomaly, I confirm the plain sensitivity for the unwrapped phase. The unwrapped phase kernels share with the high frequency asymptotic ray-path kernels the maximum sensitivity along the ray-path. They, also, account for real physical phenomena like dispersion and wavefront healing. The unwrapped phase, therefore, forms a good alternative for traveltime misfit measurements with finite frequency information in WET.

7.1.2 Analysis of the traveltime sensitivity kernels for VTI media

Inverting for anisotropy parameters is an ill-posed inversion problem, where multiple inverted parameters can resolve the same data. Aiming to reduce the null space of the inversion, and accordingly the trade-off between the parameters, I studied in Chapter 4 the VTI parameterization choice. I use the traveltime sensitivity kernels to carry out the analysis. Unlike radiation patterns, that describe the sensitivity at the scattering point level, kernels represent the sensitivity along the wave-path. There-
fore, it is a more comprehensive analysis. The similarity between sensitivity kernels
of the different parameters helps understanding the trade-off between parameters at
particular illumination angles.

The parameterization $v_h, \eta, \delta$ is not attractive because of the trade-off between the
three parameters at small angles. The parameterization $v_n, \eta, \delta$ is preferable when
inverting both diving waves and reflections simultaneously. However, the inversion
results may be affected by trade-off artifacts caused by the same sensitivity of $v_n$ and
$\eta$ for diving waves. $v_h, \eta, \epsilon$ parameterization is preferable when we can access the
data in a hierarchical approach. Starting with only large offsets (giving rise to large
scattering angles), the data is only sensitive to $v_h$. The long wavelength component
of $v_h$ is constrained without trade-off. Including the reflections, $\eta$ parameter can be
recovered and $\epsilon$ plays the role of garbage collector for acoustic media. Due to its
sensitivity to small scattering, $\epsilon$ absorbs the inaccuracies in amplitudes caused by
using the acoustic approximation.

7.1.3 Frequency domain multi-parameter full waveform inversion for acous-
tic VTI media

Following the analysis results, I considered $v_h, \eta, \epsilon$ parameterization for anisotropic
inversion in Chapter 4. I used the scattering integral approach. The proposed method
allows computing the CG preconditioning operator and the gradient step lengths with
simple vector matrix multiplications. The inversion errors caused by the trade-off
between the parameters are, also, reduced with the multi-parameter step estimation
procedure. Nonetheless, the SI suffers from a large computational cost compared to
the adjoint state method. I used a direct frequency domain wave equation solver to
reduce the computational cost. A single LU decomposition is required to solve for
all non-repeated sources and receivers. I showed examples on the accuracy of the
modeling approach.
The horizontal velocity $v_h$, $\epsilon$ and $\eta$ parameterization gives a minimum trade-off between the parameters. It is useful when the seismic data can be inverted hierarchically. In frequency domain, starting with low to high frequencies permits inverting the long wavelengths of the VTI model first, then the short wavelengths. I tested various inversion strategies: first, $v_h$, $\epsilon$ and $\eta$ are inverted simultaneously. In a second test, I considered a hierarchical inversion approach. Finally, I kept $\eta$ fixed as the initial model and inverted for $v_h$ and $\epsilon$. I showed that the inversion gives accurate inverted models for all three tests. Ignoring $\eta$ in the inversion reduces the cost and the accuracy of $v_h$ inversion is not affected. However, a good initial $\eta$ model is required. $\epsilon$ parameter absorbs the short wavelengths and amplitude errors associated with the acoustic approximation. Therefore, it should be included in the inversion, especially for elastic data.

I also considered a realistic initial model where $\epsilon_0 = \eta_0$. In this situation, the difference between the exact and initial $\epsilon$ models is large. The structure of the inverted models is accurate; however, the position of the layers is shifted. These shifts are caused by the missing $\epsilon$ long wavelengths. In a final example, I tested the inversion on data generated with a different solver.

In Chapter 6, I applied the inversion method to a real data set. I inverted simultaneously the horizontal velocity $v_h$, $\epsilon$ and $\eta$ parameters. The accuracy of the inversion is verified using multiple QC methods. I compared a vertical profile with the available well log and most of the features are recovered. Moreover, data modeled with the initial and the final models are compared with the real data. There is large improvement in data quality and multiple reflection events emerged. Finally, I generated the RTM image and the angle gathers, for which some improvements are observed.

In general, the inversion results are accurate, and $v_h$, $\epsilon$ and $\eta$ parameterization is suitable for multi-parameter inversion within a frequency domain inversion multi-scale
framework.

7.1.4 Hybrid frequency domain full waveform inversion using Born sensitivity kernels

Extending the proposed multi-parameter inversion using the scattering integral method to 3-D is problematic. The huge computational cost and large memory requirements, restrict the SI method to 2-D problems, where direct solvers can be used. To deal with this issue, I propose a hybrid implementation of the frequency domain FWI using the Born sensitivity kernels.

The method is based on dynamic ray tracing for the computation of the sensitivity kernels. The residuals are computed using wavefield modeling in the frequency domain. The method has the advantage of lower memory requirements as well as the lower computational cost. With the hybrid inversion approach, I accurately recovered the velocity anomaly for a transmission experiment. For complex velocity models, the method works well for the first stages when the model is smooth. The model long wavelengths are perfectly recovered, and the resulting models can be used as good initial models for conventional FWI.

The proposed method is advantageous for 3-D anisotropic inversion where wavefield modeling requires huge computational resources. Also, when inverting for anisotropy parameters, the ray direction can be used to compute the wavefield derivatives, necessary for the gradients calculation.

7.2 Future Research Work

The proposed frequency domain scattering integral method showed accurate results for VTI inversion. The work presented can be extended in multiples future research directions.

Understanding the robustness of the method requires additional testing for more
realistic tests. Elastic data with variable density, where the acoustic approximation fails in approximating the amplitude variations, should be tested. In this situation, the $v_h, \eta, \epsilon$ parameterization should take care of the errors caused by the acoustic approximation. Furthermore, testing the method on additional real data sets, especially for 3-D cases. The proposed hybrid approach can be used to reduce the computational cost and recover 3-D anisotropic models with the the resolution levels needed for imaging.

Another research direction is to improve on the update direction. Approximations of the Hessian matrix, namely the truncated Gauss-Newton, without extra modeling steps. The truncated Gauss-Newton was tested for the transmission example in this dissertation. However, more testing is required to ensure the convergence of the truncated Gauss-Newton inner-loop (a conjugate gradient inner loop at each iteration FWI to solve for the inverse approximate Hessian) for multiple parameters inversion. The truncated Gauss-Newton update would improve the inversion results and reduce the trade-off artifacts.

The multi-parameter inversion approach can also be extended to more complex types of anisotropy, for instance, HTI and orthorhombic models. The inversion framework is the same but with additional kernels variables, which will increase the memory requirements. To deal with this issue, the proposed hybrid approach can be used. Alternatively, data compression methods can be investigated to improve the method. Wavelets, seislets or curvelets can be used to compress the Green’s functions, as well as the sensitivity kernels, for 3-D problems. In frequency domain, frequency dependent attenuation Q factor can also be included for the inversion. The difficulty stems from the sensitivity of seismic data to the Q factor, which is similar to the sensitivity to velocity. The trade-off between the parameters will be affecting the Q factor inversion in this case. The truncated Gauss-Newton update direction, estimated within the scattering integral method, can be used to reduce such trade-off.
The work in this dissertation has been published in peer-reviewed Journals, and presented at multiple international conferences. Appendix B contains a list of these publications.
REFERENCES


Society of America, 70, 47–77.


——–, 2015, Unwrapped phase inversion with an exponential damping: Geophysics, 80, R251–R264.


——, 2013b, Wave equation tomography using the unwrapped phase: Analysis of the traveltime sensitivity kernels: EAGE Expanded Abstracts, TU0707.


Nolet, G., 2012, Seismic tomography: with applications in global seismology and


APPENDICES

A Sensitivity kernels for additional parameterizations

We present the wavefield sensitivity kernels for the remaining two parameterizations: parameterization 4 given by the NMO velocity $v_n$, the horizontal velocity $v_h$ and $\delta$ and parameterization 5 given by the NMO velocity $v_n$, the vertical velocity $v_v$ and $\eta$. Unlike the previously discussed parameterizations, here we use two velocities and one anisotropy parameter to describe the VTI medium.

Parameterization 4: NMO velocity $v_n$, horizontal velocity $v_h$ and $\delta$

We use the relation between the horizontal and NMO velocity $v_h^2 = (1 + 2\eta)v_n^2$ to write Born approximation as a function of $v_n^2$, $v_h^2$ and $\delta$ perturbations. Using the relation $\alpha_h = \alpha_n + 2\Delta\eta$ the sensitivity kernels for this parameterization are given as:

$$K_4(v_n^2) = \frac{1}{\omega^2} (G_0(x, x_s, \omega) \partial_{zz} G_0(x, x_r, \omega) + G_0(x, x_r, \omega) \partial_{zz} G_0(x, x_s, \omega)) + \frac{1}{\omega^2} \partial_{zz} G_0(x, x_s, \omega) \partial_{zz} G_0(x, x_r, \omega);$$

$$K_4(v_h^2) = -\frac{1}{\omega^2} (\partial_{xx} G_0(x, x_s, \omega) + \partial_{yy} G_0(x, x_s, \omega)) (\partial_{xx} G_0(x, x_r, \omega) + \partial_{yy} G_0(x, x_r, \omega));$$

$$K_4(\delta) = -(G_0(x, x_s, \omega) \partial_{zz} G_0(x, x_r, \omega) + G_0(x, x_r, \omega) \partial_{zz} G_0(x, x_s, \omega)).$$

The NMO velocity kernel is angular dependent for this case. The maximum sensitivity
is observed for the vertical orientation. For the horizontal orientation, the sensitivity is minimal but does not go to zero. For \( v_h \), the maximum sensitivity is observed for the horizontal orientation. The \( \delta \) parameter sensitivity kernels look similar to \( \delta \) kernels in parameterization 1 with a maximum sensitivity in the vertical direction and a zero sensitivity in the horizontal one. The main observation is the angular dependency of the NMO velocity sensitivity kernel. The sensitivity in the horizontal direction is absorbed into \( v_h \) kernel. Thus, the NMO velocity sensitivity becomes complementary to the \( v_h \) sensitivity and the maximum sensitivity is observed in the vertical direction.

**Parameterization 5: NMO velocity \( v_n \), vertical velocity \( v_v \) and \( \eta \)**

We use the relation between the horizontal and NMO velocity \( v_n^2 = (1 + 2\delta)v_v^2 \) to write Born approximation as a function of \( v_n^2 \), \( v_v^2 \) and \( \eta \) perturbations. Using this relation between the parameters leads to the new wavefield sensitivity kernels:

\[
K^{(v_n^2)}_\delta = \frac{1}{2v_0^2} \left( G_0(\mathbf{x}, \mathbf{x}_s, \omega) \left( \partial_{xx} G_0(\mathbf{x}, \mathbf{x}_r, \omega) + \frac{1}{2v_0} G_0(\mathbf{x}, \mathbf{x}_r, \omega) \right) + \partial_{yy} G_0(\mathbf{x}, \mathbf{x}_r, \omega) \right) ; \\
K^{(v_v^2)}_\delta = \frac{1}{2v_0^2} \left( G_0(\mathbf{x}, \mathbf{x}_s, \omega) \partial_{zz} G_0(\mathbf{x}, \mathbf{x}_r, \omega) + G_0(\mathbf{x}, \mathbf{x}_r, \omega) \partial_{zz} G_0(\mathbf{x}, \mathbf{x}_s, \omega) \right) ; \\
K^{(\eta)}_\delta = -\frac{2v_0^2}{\omega^2} \left( \partial_{xx} G_0(\mathbf{x}, \mathbf{x}_s, \omega) + \partial_{yy} G_0(\mathbf{x}, \mathbf{x}_s, \omega) \right) \left( \partial_{xx} G_0(\mathbf{x}, \mathbf{x}_r, \omega) + \partial_{yy} G_0(\mathbf{x}, \mathbf{x}_r, \omega) \right) .
\]

(A.2)

For this case, the NMO velocity kernel has a maximum sensitivity for the horizontal orientation. For the vertical velocity \( v_v \), the maximum sensitivity is for the vertical orientation. The \( \eta \) parameter sensitivity kernels behave similarly to \( \eta \) kernels in parameterization 1 with maximum sensitivity in the horizontal direction and zero sensitivity in the vertical one.
B Papers Submitted and Under Preparation

Journal publications


- Djebbi R., Plessix R-É. and Alkhalifah T., [2017], Analysis of the traveltime sensitivity kernels for the acoustic vertical transverse isotropic medium. Geophysical Prospecting, 65, 22-34.


Conference proceedings

- Djebbi R. and Alkhalifah T., [2017], Frequency domain multi-parameter full waveform inversion for acoustic VTI media. 79th Annual EAGE meeting, EAGE Expanded Abstracts..

- Djebbi R. and Alkhalifah T., [2016], Hybrid frequency domain full waveform inversion using Born sensitivity kernels. 78th Annual EAGE meeting, EAGE Expanded Abstracts..

• Djebbi R. and Alkhalifah T., [2014], Analysis of the multi-component pseudo-pure-mode qP-wave inversion in vertical transverse isotropic (VTI) media. 84th Annual SEG Meeting, SEG Expanded Abstracts, 394-399.

• Djebbi R. and Alkhalifah T., [2014], 3-D waveform tomography sensitivity kernels for anisotropic media. 76th Annual EAGE meeting, EAGE Expanded Abstracts.


• Djebbi R. and Alkhalifah T., [2013], Wave equation tomography using the unwrapped phase: Analysis of the traveltime sensitivity kernels. 75th Annual EAGE meeting, EAGE Expanded Abstracts.

• Djebbi R. and Alkhalifah T., [2012], It Is Only a Banana- Traveltime Sensitivity Kernels Using the Unwrapped Phase. 74th Annual EAGE meeting, EAGE Expanded Abstracts.